The Geography of the Internet Infrastructure in Europe

Emmanouil Tranos

School of Geography, Politics and Sociology

Newcastle University

Thesis submitted for the degree of Doctor of Philosophy

October, 2009

To the memory of my father, who taught me to enquire

Abstract

This doctoral thesis is concerned with the geographic analysis of the Internet infrastructure and its impacts on the economic development of the city-regions. The starting point for this research is the infrastructural attributes of the Internet which enables it to facilitate the modern – and rapidly expanded – digital economy by transporting its informational goods and services. In order to approach this research subject a wide range of quantitative methods is employed: from network analysis and complex network theory to principal components and cluster analysis as well as panel data analysis and Granger causality test.

The empirical research is firstly focused on analysing the urban economic geography of the Internet backbone network in Europe. In order to better understand the geography and the topology of the Internet backbone network, a structural comparison with the aviation networks in Europe also takes place. Secondly, effort is spent in highlighting the determinant geographic and socio-economic factors behind the distribution of this Internet infrastructure across the European city-regions. Thirdly, this study examines the impact of the – unevenly distributed – Internet infrastructure on the economic development of the European city-regions.

The above empirical analysis highlighted the unequal distribution of the Internet infrastructure and mostly the Internet capacity across the European cities. Different roles were identified for different cities, but over time the golden diamond of London, Paris, Amsterdam and Frankfurt appears to be the core of the European Internet backbone network, with London being the dominant hub. However, no clear evidence for scale free attributes was identified. Moreover, the analysis demonstrated that the level of development, the services and the knowledge economy, the spatial structure as well as the physical transport and accessibility level are significant predictors of the distribution of the Internet infrastructure. In addition, the econometric modelling concluded that the Internet infrastructure is a significant predictor of the economic development of the city-regions and that the causality runs from the Internet infrastructure to the regional economic development. Even more interesting is the geographic analysis of the causality direction as an almost north-south pattern emerged, with the northern city-regions in Europe being more efficient

in exploiting the installed Internet infrastructure. The latter can be used as an evidence for the inclusion of the Internet infrastructure in the local and regional economic development agenda. However, a set of other framework condition should be also present in order for the Internet infrastructure to have a positive impact on the regional economic development.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, my supervisor Andy Gillespie for inspiring me, supporting me and for his continuous positive attitude. My co-supervisors Seraphim Alvanides and Vassilis Tselios for guiding my quantitative journey. ERSA-PREPARE summer school organisers and participants for their constructive comments and tacit knowledge. Thomas de Graaff for his suggestions about the comparison of the Internet backbone with the aviation network. Trey Hood III for providing the STATA code for the Granger causality tests. The Greek Scholarship Foundation and the School of GPS, Newcastle University for funding this PhD research. My family for supporting my decisions and always standing behind me. Drew Gertner for loving me.

Chapter 1

Introduction

“The change from atoms to bits is irrevocable and unstoppable” (Negroponte 1995, 4)

1.1 Aims and research questions

The main aim of this doctoral thesis is to study the geography of the Internet infrastructure in Europe. Using Castells’ (1996) *space of flows* as the main theoretical vehicle and drawing upon his seminal work, effort is spent in order to understand and explain the geography and highlight the regional economic impacts of the Internet infrastructure in Europe. The infrastructural element which is under study here is the international backbone network in Europe, aggregated at the level of the city-region. This backbone network consists of the long-haul links, which connect long distance destinations and are responsible for the global character of the Internet (e.g. Malecki 2004). The resulting outcome is a study of the participation of the European city-regions in this global infrastructural network.

Epistemologically, this thesis is placed in the emerging field of the *Internet geography* or *cybergeography*, which is a branch of the field of *communications* geography focused on the geographical aspects of the Internet. It feeds the discussion about the relationship between the physical geographic space – the city-regions in this case – and this supporting layer of the cyberspace identified as the *cyberplace* (Batty 1997). Using Castells’ (1996) framework, this study is concerned with the 1st – supporting – layer of the space of flows.

More specifically, three research questions (RQ) have been set up for this doctoral study:

RQ1: How is the Internet infrastructure allocated across European city-regions?

RQ2: Which are the geographic and socio-economic factors that shape the distribution of the Internet infrastructure across European city-regions?

RQ3: What are the impacts that the Internet infrastructure can generate on the development of city-regions in Europe?

The first research question is a clear geographic one and aims to explore the geographic pattern of the Internet backbone linkages in Europe. As the Internet backbone is firstly a network, the topology of this network is important. However, because this study has a clear geographic starting point, what is more important is to link the topology of the Internet backbone network with the geography of the city-regions. This difficulty is raised by the fact that the Internet and consequently its underpinning infrastructural layer were designed to support the Internet function which is rather a-geographic and mostly topologically based.

The second research question follows from the results of the first one and intends to explain the geography of the Internet backbone links in Europe. As will be explained later in this thesis, these long-haul Internet links are mostly privately developed and owned. Consequently, the location decisions behind the installation of this infrastructural layer reflect the perceptions of the telecommunications companies (known as *telcos*) about the demand for such facilities in order to maximise the returns of their investments (Gillespie and Robins 1989). Based on this, effort is spent in this thesis in finding these socio-economic and geographic factors that explain the geography of this infrastructure in Europe and consequently the perceptions of the telcos for higher demand for their networks.

The third question goes one step further and seeks to examine whether the Internet infrastructure generates economic development impacts at the regional level. As will be discussed in Chapter 3, research has been concerned about the impacts of the expansion of Internet usage in the economy through productivity growth. However, research has not yet focused on the localised economic impacts of the supporting layer of the Internet infrastructure. This doctoral thesis aims to research if this infrastructure can generate such impacts and also address the issue of the direction of causality between the infrastructure and regional economic development. The latter is a well known problem in regional science and it will be extensively discussed.

In order to address the above research questions, this thesis draws upon three different research areas. Firstly, as mentioned above, the basis of this study is the *Internet geography* field. It provides the fundamental theoretical and empirical background in order to pursue the above research questions. However, because of the importance of this infrastructural layer in the post-modern economy and society, there is a need to also employ theoretical and methodological tools from other fields to approach the current research questions.

The *world city* literature is the second theoretical pillar that this study is based on. Telecommunications (just like transportation) are *friction reducing technologies* because of their ability to reduce the cost of distance (Cohen et al 2002, Cohen-Blankshtain and Nijkamp 2004). Because of this attribute, the Internet and Internet infrastructure enable global interaction and facilitate global economic activity (Malecki and Wei 2009), supporting the emergence of a *world cities network*. This theoretical pillar will provide the necessary input in order to understand the importance of this infrastructure from a global (inter)urban perspective.

Lastly, this thesis also draws upon the literatures of the *digital economy* and of *regional science* as it attempts to map the regional economic impacts of the Internet infrastructure. The digital economy is the main theoretical framework used here in order to explain the link between the new technological paradigm reflected in the expansion of the Internet and economic development through productivity gains. However, as will be highlighted in Chapter 2, this link mostly refers to the scale of the national economies. In order to transfer this argument to the scale of the analysis used in this study, economic geography and regional science literatures are employed.

In order to approach the above themes, secondary data about the Internet backbone links in Europe and quantitative data analysis methods are utilized. Briefly, the main dataset used in this study contains data about the international intercity Internet backbone links and their capacity, which are present in European cities for the 6 year period 2001-2006 (Telegeography 2007). In order to fully exploit the structure of the data, methods from (social) network analysis and complex network theory have been utilized. At a first level, the results (global statistics) of the network analysis comment on the topology of the backbone network and effort is spent to incorporate geography in this as well. At a second level, the local level results of the network analysis are translated to attributes for the city-regions which participate in this global network. The geographic and socio-economic factors behind the distributions of these local level statistics are explained with the use of statistical techniques such as *principal component analysis* and statistical modelling. Lastly, these city-level attributes, which reflect the Internet infrastructural capital, are used in order to model the impact of the Internet infrastructure on regional economic development. For the latter econometric modelling is employed and more specifically panel data analysis and Granger causality tests for panel data.

1.2 Rationale for this study

The main motivation for choosing this research subject is the growing importance of Information and Communication Technologies (ICTs) in the economy. ICTs, which include the Internet and its backbone networks, are the backbone of the new – *digital* – economy (Antonelli 2003), with processes of production, distribution and exchange increasingly reliant on them. Thus, the Internet is the most essential development as regards the distribution and exchange of information after the telephone (Moss and Townsend 2000). Shiller (1999, cited in Warf 2001) even goes further by suggesting that the Internet might be the most rapidly spreading technology in human history.

From a macro perspective, it is established nowadays that the Internet and the ICTs affect the economy by improving its productivity (Atkinson and McKay 2007; Cairncross 2001). Additionally, ICTs and the Internet along with the aviation network can be said to be the supporting layer of the globalization, as they are responsible for the transportation of the weightless goods and the main actors of the global economy, but also for the transportation of the ideas which underpin this global process (Taylor 2004; Graham and Marvin 2001; Rimmer 1998; Cieslik and Kaniewska 2004). In such a global economy, a country's importance depends upon the cities located within it, the importance of which depends in turn upon the multinational firms located within these cities (McCann and Acs forthcoming). The function and the global extent of these firms is supported and enabled by the ICTs.

However, ICTs and consequently the Internet are more than just a new technology, despite the rapid pace of their expansion and the wide range of the impacts they generate. Hence, the wide adoption of these technologies appears to create a new *technological paradigm* (Perez 1983), which affects not only production and the economy, but society in general. Upon this element Castells built his theoretical work about the *network society* (Castells 1996).

In terms of geography, ICTs and the Internet are not a homogenous system equally spread around places (Gorman and Malecki 2000). From an analytical point of view, despite what the average users experience as a *placeless cyberspace*, the latter depends on *real world's fixities*, which are found on cyberplace (Kitchin 1998a and 1998b). From a more economic perspective, neither the outcomes that ICTs and the Internet generate are homogenous in space. On the contrary, it seems that ICTs can generate different impacts on different regions. And this differentiation is not only due to the different level of ICTs infrastructure installation or ICT diffusion, but also because of the different capability of each region to exploit benefits from them (Antonelli 2003).

More specifically, the backbone network is one of the most interesting elements of the Internet infrastructure from the geographical point of view as it is responsible for the Internet's global reach. From an urban viewpoint, the structure of the backbone network can potentially provide information about the intensity of the participation of cities in the digital economy. From an analytical perspective, the geography of the backbone networks can provide insights about the determinants of these networks. From a more policy oriented standpoint, the geography of the backbone links but also (and maybe more importantly) their capacity might have an impact on local economic activities as it can directly affect firms which are highly dependent on global Internet communications (Greenstein 2004). Hence, the Internet's performance and efficiency between any two places is not dependent on the physical distance between them but mostly on the capacity of the backbone connections – known as bandwidth – between them (Gorman and Malecki 2000).

The last argument is fundamental for understanding the importance of the Internet in the frame of the digital economy. However, such argumentation has created misconceptions about the impacts of ICTs and the Internet on spatial structure: early commentators have expressed positions according to which these technologies will result in the *death of cities* (Gilder 1995; Drucker 1989 cited in Kolko 1999), the *death of distance* (Cairncross 1997), the emergence of *electronic cottages* (Toffler 1981) and in general to the *end of geography*. All the above rather deterministic approaches foresee the devitalisation of centralizing forces and the growing dominance of centrifugal forces, which will eventually result in a decentralized spatial pattern of economic activities. However, as has been proved, the Internet is an urban phenomenon (Rutherford et al 2004) and consequently the same applies to the Internet backbone network, which is characterized as an *urban infrastructure* (Moss and Townsend 2000).

The above preliminary discussion verifies the importance of ICTs and the Internet from a geography stand point. However, in spite of their established importance, it seems that there is still a long way to go in order to further comprehend this new technological paradigm from a geographical point of view. As will be extensively discussed in Chapter 2, ICTs have not been among the leading research subjects among geographers, planners and regional scientists, mainly due to the inherent technical complexity of the actual subject (Bakis 1991, Hepworth 1989, Kellerman 1993). As a result, regardless of the various theoretical geographical approaches concerning ICTs, which emerged even prior to the establishment of a digital economy, it seems that there is a scarcity of empirical studies researching the geography of ICTs: such a study can shed light on the geographical distributions of ICTs, can explain the factors behind these (centralised) distributions, and also explain the impacts that this new technological paradigm can generate on local economic activity. The latter apart from its academic importance, can provide valuable insights to the local and regional development policy agenda.

1.3 Structure of the study

The structure of the thesis goes as follows: Chapter 2 provides the necessary literature background in order to reach the research questions. It starts with a brief technical description of the Internet. In spite of the geographical starting point of this doctoral research, it is necessary to have some understanding of the Internet function from a technical point of view. Then, the three main theoretical pillars of this study are critically presented. First, the literature of the emerging field of *Internet geography* is analysed. This is the core and the most influential part of the literature for this study. The main subject of this study, the Internet infrastructure, is defined here. Then, the *world cities* literature is critically presented and the importance of ICTs is highlighted. Thirdly, a theoretical framework is built in order to comprehend the regional economic impacts of the Internet infrastructure.

Chapter 3 is dedicated in analysing the data and the methods used for this study. As mentioned above, in order to approach the research questions, secondary data about the Internet backbone network and quantitative methods have been employed. The rationale of choosing the specific methods is illustrated here as well as the main methodological points. This chapter ends with the construction of the thesis's research framework, where all the research questions, methods, data sources and theoretical pillars are schematically presented together.

The empirical research takes place in the next four chapters (4-7). Chapter 4 presents a descriptive network analysis. It is the chapter where the fundamental analysis of the Internet backbone network takes place. The network topology is built and both global and local statistics are calculated. Additionally, complex network methods are also introduced. This chapter concludes with some initial results from the exploratory analysis both for the global level of the whole network and for the local level of city-region attributes.

Chapter 5 uses the same set of methods in order to perform a structural comparison between the Internet backbone and the aviation network in Europe. As will be highlighted in Chapter 2, the first layer of the space of flows (Castells 1996) is formed by these two infrastructural networks (Taylor 2004). The comparison takes place both at the global level of the whole network structure but also – and probably more interesting from the geography point of view – at the level of city attributes. The latter indicates the different roles different cities perform in these two networks.

Chapter 6 is dedicated to the explanatory analysis of the Internet backbone network. Principal component analysis and statistical modelling are exploited in order to explain the geographic and socio-economic reasons behind the spatial distribution of this infrastructure. The chapter concludes with a set of new components which resulted from the analysis as significant predictors of the distribution of the Internet backbone networks.

Chapter 7 is the final empirical chapter of this doctoral study and is dedicated in identifying the regional economic impacts that the Internet infrastructure can generate. Apart from building an empirical econometric model which tests the impact of this infrastructure on the economic development level of the city-regions in Europe, the analysis goes a step forward by identifying the direction of causality between the Internet infrastructure and economic development with the use of appropriate econometric tests. At the end of each one of the four empirical chapters (4-7), a separate annex is provided, which supports the quantitative analysis.

This thesis concludes by providing the empirical answers to the research questions stated in this first chapter. Additionally, in this last chapter the further contributions of this research to the relevant literature are highlighted. Drawing on the empirical results some policy recommendations are also stated in order to promote the inclusion of Internet infrastructure in the local and regional development agenda. The thesis ends with the identification of some limitations and suggestions for further research.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews the relevant literature and provides the necessary theoretical framework to further investigate the research questions illustrated in the previous chapter. The starting point for approaching the Internet infrastructure is an analysis of the Internet’s architecture. This provides all the necessary technical knowledge in order to understand how this complex system works. Such an understanding will be fundamental for approaching the research questions of this doctoral thesis.

The main core of the relevant literature used here can be grouped in three main pillars: (a) the Internet geography, (b) the world cities literature, and (c) the regional economic development impacts of Internet infrastructure. These pillars are necessary in order to approach the three research questions presented in Chapter 1. More specifically, the first element identified as the Internet geography is directly related with all the three research questions of this study. As will be further analysed in the relevant section, this study is part of this emerging research field of the Internet geography. The second pillar, the world cities literature, provides a wider theoretical framework for this study and for the Internet geography field as it maps and analyses the increased interaction and interdependence among cities, the importance of which is highlighted in the frame of the global economy. Lastly, the third pillar of this literature review is mostly related with the third research question about the impacts of the Internet infrastructure on regional economic development. The current economic framework is analysed as well as how this is facilitated by the Internet infrastructure especially at the level of city-regions.

The structure of this chapter reflects the above three-pillar segmentation. It starts with the technical analysis of the Internet function and it then continues with the three theoretical pillars: the Internet geography, the world cities literature and the economic developmental impacts of the Internet’s infrastructure. The chapter ends with an epilogue which also links the above three pillar theoretical structure with the empirical research of this study.

2.2 Technical analysis of the Internet

2.2.1 Introduction

This section provides a technical analysis of the Internet in order to facilitate the further research on the geography of the Internet infrastructure. Despite the fact that this doctoral thesis is not concerned with the engineering side of the Internet, such knowledge is important in order to investigate and comprehend the geography of the Internet infrastructure and the impacts that this might generate. Because of the Internet’s strong technical character, any attempt to approach it from the social sciences point of view, would be ineffective without considering its primary technical nature.

Broadly speaking, it could be said that the Internet consists of two layers: a technical layer, and a content layer, with the latter overlaid on the former. The key characteristic of the Internet’s technical layer is its network topology. It consists of edges and nodes, which have a specific physical location and structure and can be approached as such (Gorman and Kulkarni, 2004). In fact, it is not a single network neither one specific system, as many of its users think. The Internet, as *the network of networks*, consists of several interconnected small, medium and large networks (Gorman and Malecki, 2000). Because of this complexity, these networks should be characterized by a specific and predefined hierarchy in order to be functional (Malecki and Gorman, 2002). A schematic approach of the Internet function can be found at the end of this section in Figure 2.1a and 2.1b (page 37).

2.2.2 Internet Service Providers

The Internet’s networks can be approached from different points of view. From the business perspective, the interconnected networks can be identified as being associated with the Internet Service Providers (ISPs). The latter refers to companies or organizations which maintain one or more interconnected networks and through these provide Internet access. Usually, an ISP in order to achieve the desirable universal connectivity (i.e. connectivity with the rest of the Internet’s networks and through them with all the interconnected computers) needs to cooperate, interconnect and exchange data with other ISPs. This can happen in various ways, as described in the next section. The distinction among ISPs follows the Internet’s rigorous structure. On top of it, the Tier-1 ISPs can be found which are characterized by extensive global networks and which are able to achieve global connectivity without buying any Internet connectivity from another ISP (i.e. exchange data with an upstream provider under a fiscal agreement, known as IP transit). Only a few Tier-1 ISPs exist in the world (Telegeography 2007) and they are usually part of global telecommunications companies (telcos) which maintain large capacity backbone networks around the world. Tier-1 ISPs exchange data with each other (known as peering) and sell Internet connectivity (known as IP transit) to lower ranking ISPs, called Tier-2, which also have their own but less extended networks. While Tier-1 networks can ensure their connectivity with the rest of the world without buying any upstream connectivity, this is not the case with the Tier-2 networks, which on top of their own connectivity also need to purchase Internet access from the upstream Tier-1 ISPs. Tier-3 and Tier-4 ISPs are lower scale providers, who mainly act at national and local levels (Telegeography 2007).

2.2.3 Peering and nodal locations of the Internet

Peering is an essential process of the Internet function because it integrates different networks by giving them access to each other (Gorman and Malecki 2000). “Only through peering do two networks interconnect to form what we know as the Internet” (Malecki and Gorman 2001, 93). The latter takes place in specific nodal locations, such as Internet Exchange Points (IXPs), which can be identified as the nodes of the Internet. An IXP is a facility where different ISPs connect their networks to and place their dedicated routers and through them interconnect with some or all of the ISPs present in this IXP. Industry defines an IXP as:

> “a physical network infrastructure operated by a single entity with the purpose to facilitate the exchange of Internet traffic between Internet Service Providers. The number of Internet Service Providers connected should at least be three and there must be a clear and open policy for others to join” (Euro-IX 2006, 4).

This part of the Internet’s topology was first introduced in 1991 in USA, when a number of commercial backbone carriers founded the Commercial Internet

Exchange (CIX) in Santa Clara, California (Kende, 2000). Later, in 1995, the National Science Foundation, having as an objective the commercialization of the NSFNET (i.e. the ancestor of today's Internet) introduced the four privately managed Network Access Points (NAP) located in San Francisco (operated by PacBell), Chicago (BellCore and Ameritech), Washington, DC (MFS) and Pennsauken, NJ (New York operated by SprintLink). Similar to CIX, they enabled the ISPs to peer, leading to the emergence of a 'national commercial Internet' (Kende, 2000, Grubesic and O'Kelly, 2002). Nowadays the term IXP is more commonly used rather than NAP, especially in Europe.

Peering can be distinguished in two different sub-categories, public and private peering. The former takes place in IXPs, where all the interconnected ISPs freely interchange data in order to achieve the goal of universal connectivity. The main characteristics of public peering are the following (Kende, 2000):

- Peering ISPs only interchange data when the origin and the destination are parts (customers) of these two networks. Otherwise, transit data cannot be transported through peering agreements.
- Peering agreements are not fiscal agreements. The only costs for peering are logistic costs, such as ISP's interconnection to the IXP, its' routers and also any fees that the IXP may charge. Most of the IXPs in Europe are at this time non-commercial facilities, financed by membership fees paid by the interconnected ISPs. In 2006, 54 non-profit and 30 profit IXPs were present in Europe (Euro-IX, 2006).
- The third characteristic refers to the 'hot-potato' routing policy. Because usually the peering takes place to scattered locations, ISPs agree in passing the data from one network to another at the earliest peering point.
- The last characteristic refers to the quality of service. The ISPs, which receive data from another ISP, are not obligated to guarantee any specific level of quality rather than committing in undertaking 'best effort'.

ISPs choose to participate in IXPs in order to achieve the desired global connectivity through public peering with other ISPs. Public peering at IXPs seems to be the cheapest choice for ISPs to interchange any volume of data because they are only charged (if any) the IXPs' fees and the logistic costs. In addition, by interconnecting to an IXP, which has a national or even a local reach, data-packets with national or local origin and destination can remain at this level, avoiding long data transport only for peering reasons. Such practise results in time and money gains as the unnecessary use of the expensive long-haul links is avoided (Paltridge 2002). However, public peering is not always the case. Because of the rapidly increased Internet traffic by the end of 1990s', many US IXPs (i.e. NAPs) became bottlenecks for Internet data traffic. And this is why the major ISPs started implementing 'private peering', which refers to bilateral peering agreements between any two ISPs using direct connections, in order to bypass the congested routers of the IXPs (Telegeography 2007; Kende 2000). Private peering takes place either at IXPs if the two ISPs are already present there, but without using IXPs' routers, or directly at ISPs' Points of Presence (POPs), which are the nodes where the end-users are connected with the ISPs and which are further analyzed below. The above IXPs' problems resulted in diminishing IXPs' role in USA by the late 1990s (Telegeography, 2006, Kende, 2000). However, recent technological advances have enabled IXPs to strengthen their role and this probably explains the steady increase of the number of IXPs especially in Europe (Euro-IX 2006).

Comparing IP transit with peering agreements from the business point of view, the obvious difference is the fiscal character of the former and the free willing base of the latter. In addition under peering agreements, the different ISPs are equal members of an agreement and their relationship cannot be approached by the customer-consumer relation. Furthermore, the ISP which sells transit connectivity to another ISP will transmit traffic from its customer to its peering partners (Kende 2000). To sum up, ISPs in order to achieve universal connectivity, use different combinations of public and private peering as well as IP transit. The above peering choices are related with an ISPs' customers, its business plan, its location etc.

From the topological point of view, IXPs can be regarded as the nodes of the Internet, since they represent the locations where the different edges of the network are switched. However, IXPs are not the only Internet nodes. Points of Presence (POPs) are also considered as Internet nodes. Their role is to enable end-users to connect with the ISPs. End users connect via the *local loop* or the *last mile* (i.e. the link between the end-user and the ISPs) with ISPs' routers, which are located at the POPs and through them to the rest of the world. So, it could be roughly said that

POPs are responsible for the end-users connectivity with the ISPs while IXPs are responsible for ISPs universal connectivity. From the hardware point of view, the nodes of the Internet are either routers or switches. IXPs and POPs are equipped with both of them. Their main role is to send the Internet data packets to specific locations, but their difference will be highlighted below in section 2.2.5.

In reality, the distinction between IXPs and POPs is sometimes quite vague, since peering can also take place at the latter under private peering agreements. POPs are usually owned by ISPs since they connect the end-users with the ISP's network. Usually they are located in specific establishments, which are known under various names such as *data centres*, *telecom hotels*, *data warehouses*, *colocation*, *colo centre*, *server farms* etc. and provide a wide range of services (Evans-Cowley et al. 2002, Townsend 2003). They include colocation facilities, servers hosting, data archives, hardware management etc. in a controlled environment for climate conditions and physical disasters. These facilities are characterised by great Internet connectivity with access to backbone networks and this is why low-rank ISPs are located there or rent racks to place their routers in order to connect with higher tier ISPs. These facilities are usually found in the wider metropolitan areas, employing redundant buildings such as warehouses and department stores with high ceilings and high capacity power supply: they are found in locations which combine both access to high capacity backbone networks and closeness to customers in order for them to have physical access to their equipment. However, usually such facilities can neither afford the cost of nor can find buildings with proper specifications in central locations (Evans-Cowley et al 2002; Townsend 2003)[1]. Nowadays, it is also common to find colocation facilities in remote areas which combine access to backbone networks and low cost electric power. The discussion has also emerged for locating such facilities in areas where renewable energy is available as a low carbon-dioxide emission measure, by exploiting the vast installed bandwidth (for this discussion see Arnaud 2009). To sum up, IXPs are differentiated from POPs and telecom hotels because the main objective of the former is ISPs' interconnection and not the provision of any other services. Additionally, it is common for IXPs to be owned by

---

[1] In USA, the average size of such facilities is 43,700 square feet with the NAP of the Americas in Miami being the biggest with 761,000 square feet (Evans-Cowley et al. 2002)

non-commercial bodies, especially in Europe, contrary to the POPs and the telecom hotels (Evans-Cowley et al., 2002).

2.2.4 ISPs – further classifications

Telegeography (2007) suggests another distinction among the ISPs, which takes into consideration their geographical reach. According to this, ISPs can be differentiated between global, regional and national. The former refers to IP (Internet Protocol) carriers which sell wholesale Internet connectivity in at least two different regions, and in at least two countries in those regions[2] (ibid). Regional carriers sell Internet connectivity in only one region, while the national IP backbone providers are only focused on a single country. In 2007 only 21 IP backbone providers out of the 530 within Telegeography's database are characterized as global. However, they own 62% of the total bandwidth, revealing how concentrated is the Internet in a few ISPs (Telegeography 2007).

In terms of terminology, there seems to be a blurred picture of what an ISP exactly is. According to Moss and Mitra (1998, 25) "an ISP is the consumer's gateway to the Internet". And they continue by drawing parallels an ISP with a telephone company, which instead of providing voice connections, provides the necessary data connections for using the various Internet services, such as web browsing, email etc. Gorman and Malecki (2000) characterise all the networks which comprise the Internet as ISPs, no matter their differences in size and function. Cukier (1998 cited in Gorman and Malecki 2000) classify them into four categories: transit backbone ISPs, downstream ISPs, online service providers, and web hosting specialized firms. The backbone providers in this case are actually the Tier-1 ISPs and downstream providers refers to the lower tier ISPs. Online service providers are those ISPs which are focused to end users and could be linked with the Tier-4 ISPs. They are also called virtual ISPs, because actually they do not own any physical links for carrying data but instead they lease IP connectivity from upstream providers. They have presence at their upstream ISPs POPs, where their end-users are connected to. So, online service providers, Tier-4 ISPs or virtual ISPs are in reality Internet connectivity retailers, whose main advantage is their penetration among

---

[2] At this case, the term region refers mainly to continents. Telegeography's (2007) regions are USA and Canada, Latin America and Caribbean, Europe, Asia and Africa.

residential end-users. Malecki (2004) refers to ISPs as the companies which provide access to the backbones, distinguishing this way the ISPs from the Internet Backbone Providers. The last term is used by various researchers (Evans-Cowley et al, 2002, Telegeography, 2007, Kende, 2000) in order to differentiate the long haul data network carriers from national, regional and local ISPs. Apparently, there are minor differences among these terms. For simplicity reasons, the term ISP for this study refers to all the interconnected networks, since all of them provide Internet services by exchanging data with other networks.

2.2.5 Internet architecture

From the Internet architecture point of view, the Internet's numerous networks can be characterized as Autonomous Systems (AS). The latter refers to "a connected group of one or more IP prefixes run by one or more network operators, which has a single and clearly defined routing policy" (Hawkinson and Bates 1996, 2). The term AS refers then to one or more physical networks which operate under the same administration, which is responsible for choosing the peer networks they will interchange data with. IP is the communication protocol which determines the Internet's function by enabling data packet interchange among the Internet's different sites using switches and the routers which run under the Transmission Control Protocol / Internet Protocol (TCP/IP) switching technology (Gorman and Malecki 2000). In order for data exchange to take place, all data is fragmented in data packets labelled with their origin and destination address, as well as with the order in which the data is to be rebuilt, and those packets are transported through the different interconnected nodes (UN 2006). Each destination on the Internet, that is an interconnected computer, has been given a unique IP address in order to be reachable from the rest of the world. So, an AS could be read as a network which interconnects a set of IP addresses. Each AS is named with a unique number from a 16 bit pool (2^{16}), which results in 65,536 unique values. This number is the exclusive identification of each AS, upon which is based the data interchange system among the AS, know as Border Gateway Protocol (BGP-4). In 2007, around 27,000 AS were found to be active in the world (Telegeography, 2007). Similar to this is the procedure for obtaining an IP address. The IP version 4 protocol (IPv4), which is currently the most widely used, is limited to a pool of 2^{32} unique addresses while the next version extends the IP space to 2^{128} unique addresses. So, it is obvious that the AS numbers and the IP addresses are a scarce resource. Unlike the ISPs classification, the IP and AS one seems to be more robust with Internet Assigned Number Authority (IANA, 2007) being the responsible organization for the allocation of these numbers.

The TCP/IP follows the guidelines of a wider protocol, which is developed in a model form and governs not only the function of the Internet, but also the global networking process. It is known as the Open Systems Interconnection (OSI) model and it was introduced by the International Standards Organisation (ISO) in 1984 and updated in 1994 (UN, 2006). It is built on a 7 layers form: the lower layers are dedicated to basic technical tasks while the upper, which rely on lower layers' efficient function, are closer to the end-user and include more sophisticated functions. The first layer is called the *physical* layer and consists of the wires, the fibre, the wireless links and the physical elements in general, which are responsible for the data transmission following precisely the directives from the upper layers. The *data* layer feeds the physical layer with error-free flows, which are transmitted by this lowest layer between two adjacent nodes. The third layer, identified as the *network* layer, is the first layer where a complete origin-destination route is setup, using the IP addresses. While the switches, which function at the data layer, are able to switch data packets only between the adjacent nodes of the complete route between the origin and destination which usually consists of numerous intermediate nodes, the routers, which function at the network layer, are responsible for setting up and managing the complete routes of the data packets. The importance of this layer is that it defines the whole network: if a site is not visible by a router, it is not part of the Internet (Gilder 2000). The next layer is the *transport* layer, the main protocol of which is the TCP, which certifies that the data packets are received correctly and in the right order. The applications' announcement to senders' and recipients' computers takes place on this layer (UN 2006). However, the control of the dialog between the sender and the recipient is responsibility of the fifth layer, the *session* layer. The seventh layer is the *application* layer, where the most common Internet applications such as File Transfer Protocol (FTP) and Hypertext Transfer Protocol (HTTP) function (UN 2006). Between the application layer, which is the nearest to the user layer, and the session layer, there is another layer called the *presentation* layer, which is the interface between these two layers. Gilder (2000, 63) describes the OSI model very efficiently using the telephone analogy:

> “Pick up the handset and listen for a dial tone (*physical layer*); dial up a number (every digit moves the call another *link* closer to the destination); listen for the ring (signifying a *network* connection and *transport* of signals). Getting someone on the line, you may be said to have completed the first four layers of the OSI stack. Then your hello begins a *session*, the choice of English defines *presentation*, the conversation constitutes the *application* layer. The hangup ends the *session*”.

2.2.6 The physical layer and its metrics

The Internet’s performance is highly related with its physical layer. There are two main metrics for approaching a computer network’s performance: bandwidth and latency. The former simply refers to the “number of bits that can be transmitted over the network in a certain amount of time” (Peterson and Davie 2003, 40) while the latter refers to “the time (measured in milliseconds) that it takes to transport and receive data between two nodes on the Internet” (Dodge and Zook 2009, 2). For example, a modern transatlantic circuit can have a bandwidth of 10Gbps, which means that it can transport 10 x 10^9 bits every second (8 bit = 1 byte = one typed character). The bandwidth is mainly defined by the physical means which transports the data, with fibre optic cables providing today the greatest bandwidth. Latency on the other hand is a more complicated metric, which is measured in time units and refers to round-trip time or otherwise the time that a data packet needs in order to reach its destination and return back to its origin . Latency may be affected by three sources (Peterson and Davie 2003):

- Propagation delay. This is the time the data needs in order to travel the length of the line. It is related with the distance of the link and the speed that the data travels in the link. For the case of fibre optic links, light travels at 200 x 10^6 m/s, less than it travels in a vacuum (300 x 10^6 m/s) (Peterson and Davie 2003).

- Transmission delay. This is the time needed to move a data packet across the network media. It is related with the size of the data packet and the bandwidth of the internet medium.
- Processing delay or queue. This occurs for various reasons such as the time take to establish the route, the switching tasks, etc. This type of delay is highly correlated with the number of different nodes that a data packet needs to pass through (number of hops using the Internet architecture terminology) in order to reach it's final destination (Obraczka and Silva 2000).

Peterson and Davie (2003, 42) express latency as follows:

Latency = Propagation + Transmission + Queue, or

Latency = Distance/Speed of Light + Data packet size/Bandwidth + Queue

So, if a data packet of 1 bit is sent from one node to another without having any intermediate nodes, latency will only occur because of propagation delay. And despite the dramatic growth in networks' bandwidth no one can assume that latency is decreasing (Peterson and Davie 2003).

What is also important in order to comprehend the way the Internet functions and the way its different elements are scattered among and inside cities, is to understand the nature of its physical layer. The edges of the Internet are certainly the most expensive and extensive component of an ISP's investment. There are three main media types that facilitate data transmission (Tanenbaum, 2003). The oldest one is the *twisted pair*, which consists of two insulated copper wires, twisted together in order to avoid antenna phenomenon created by two parallel wires. Public Switched Telephone Networks (PSTN) are still largely based on twisted pair wires. They can achieve several Mbps for a few kilometres. The next category is the *coaxial cable*, which is also built on copper and it was firstly widely used for television transmission and then for telephone long haul links. Nowadays, the long haul links are exclusively based on *fibre optic cables*. Their main difference is that instead of transmitting electrical pulses, fibre optics transmit light pulses through the fibre, which is generated by a light source (usually LED) placed at one end and recognized by the detector at the other end. The absence of light is recognized as 0 while the light as 1, just like the electricity over copper cables. Nowadays, the commercially used fibre optic links can achieve a bandwidth of 10Gbps, with the detectors being the main obstacle for going further, while much greater bandwidth has been reached in research labs.

Apart from the bandwidth, there are a few more differences between fibre optic and copper based links. First, the low attenuation of the former and consequently the low needs for repeaters, which are used in order to enhance the signal, make fibre much more suitable for long haul links. In addition, fibre is not affected by external electromagnetic interference, and is less sensitive to environmental conditions. What is interesting is that telcos also prefer fibre because it is much lighter and has lower installation cost than copper wires. Furthermore, it consumes less space in the already narrow and filled ducts and provides greater capacity than copper. The fibre optic cables, just like the copper wires, are placed in pipes, which are installed either next to pre-existing network infrastructure (motorways, roads, railways etc.) or in pipes that are not used any more such as sewer networks. So, by replacing the oversized copper wires with the smaller in volume but higher capacity fibre, there is a potential gain for carriers. However, fibre's installation and maintenance needs special skills from the engineers and it is very sensitive in bending. Moreover, the cost of the optical interface is quite high and it is higher than the equivalent for copper wires (Tanenbaum, 2003). Nowadays, the extended interregional links are built on fibre optics, while the last mile is still based on copper wires. However, a growing discussion is taking place nowadays about the implementation of fibre optic technology in the local loop (Fibre To The x, where x represents Home, Building, Premises, and Cabinet – FTTH, FTTB, FTTP, and FTTC respectively). For example, OECD (2006) states that "fibre to the home is becoming increasingly important for broadband access, particularly in countries with high broadband penetration." Despite these advantages, installation costs are still too high-priced for extensive use of fibre in the local loop. The local loop, which is always related with excavations in heavily populated and urbanized areas with high-priced land cost, is supposed to be the most expensive element of a network's roll out, reaching 80% of the total cost (Graham 1999).

2.2.7 A little geography

From the geography point of view, the POPs and IXPs are part of the cities Internet hardware, which are interconnected through the backbone networks or otherwise the *first mile* (Grubesic and O'Kelly, 2002). What is also interesting is the physical location of the rest of the Internet's elements stated above, like the IP addresses or the AS, which is also vital for the Internet's function, no matter that they are less visible than the former. However, it is not always that easy to identify their physical location. The reason for this is that the Internet is developed based not on the geographical but rather on the topological location of its components; or otherwise the Internet recognizes the location of its elements only in relation to other Internet components (Dodge and Zook 2009).

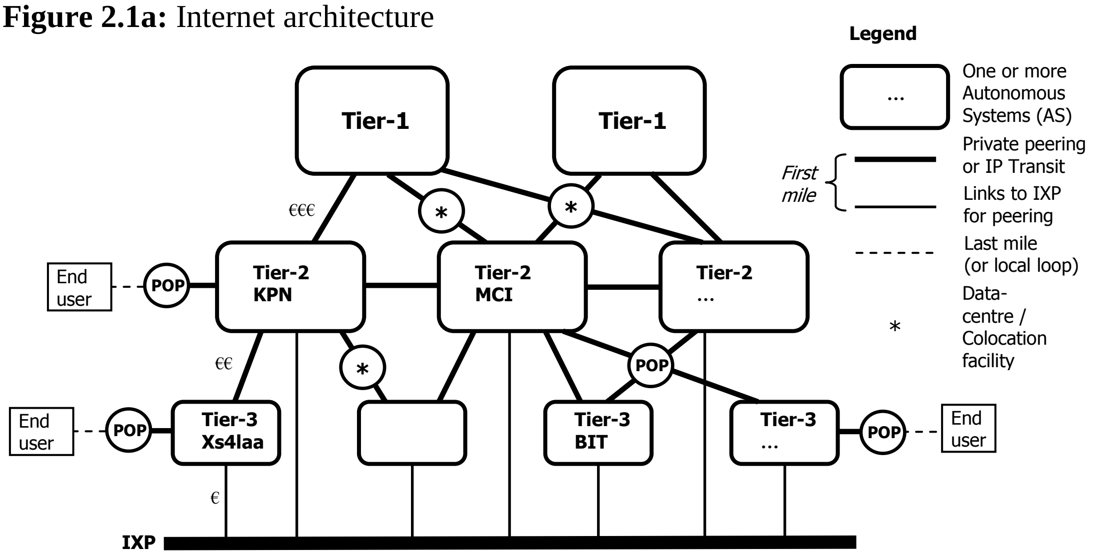
All the above seem to be important when the discussion goes to the urban geography, because the allocation of this Internet hardware determines the cities' Internet capacity. The number of links a city shares with the rest of the world, as well as the bandwidth of those links, reflect the city's capacity in data exchanging with the rest of the world. However, no matter the number of the links coming through a city and the capacity of those networks, a city could not benefit unless there is a node linking its Metropolitan Area Network (MAN) and through this the local loops and the end users, with the backbone networks. Otherwise, the city would be bypassed by those networks without gaining access to the rest of the world, just like small towns are bypassed by motorways and high speed rail, resulting in what is known in the literature as a *tunnel effect*. So, in order for a city to benefit from the Internet infrastructure, it is not enough to be near to the high capacity backbone networks, but it needs to be connected to them with multiple nodes, which enable its fast and secure interconnection with those networks. In addition, at the intra-urban level, in order for the urban area to benefit by the interconnection to the inter-urban networks, extended intra-urban hardware is needed, such as the MAN, the local loops and the POPs, which are fundamental in order for its end users to be connected with the rest of the world. The importance of the intra-city Internet infrastructure is reflected by the fact that nowadays the main Internet bottleneck is not the backbone connections nor the IXPs, but the last mile, which is still not facilitated by fiber optic technology (Pelletiere and Rodrigo 2001; Blum and Goldfarb 2006).

In terms of quality, in order for a city to develop its competitive advantage, it is not enough just to be connected to a backbone network. The quality and the capacity of its interconnection with the rest of the world determine its Internet capacity. The greater the city's Internet bandwidth is, ceteris paribus, the greater end user's bandwidth and the faster the data transport with the rest of the world will be. Additionally, the more the links a city shares with the rest of the world, or in other words the greater the number of the backbone networks are interconnected at city's POPs, the greater the reliability of city's communications with the rest of the world will be, in case one or more links go down.

If the above could reflects the city's aggregated Internet capacity, then the geography of IP addresses could indicate something more tangible, that is the location of the interconnected computers. If the links and the nodes define the Internet infrastructure supply at city level, then the geographical location of IP addresses could indicate the agglomeration of the Internet data origins and destinations at city level. However, it should be underlined that IP addresses are not related with the Internet's content but with its hardware, indicating the location of the computers which host the Internet content (a website for example) and not the location where this content is produced (Dodge and Zook 2009). And even this is not very accurate since the only available geographic information for the IP location is the registered postal address of the IP and is very common that this address is different from the actual geographical location of the IP.

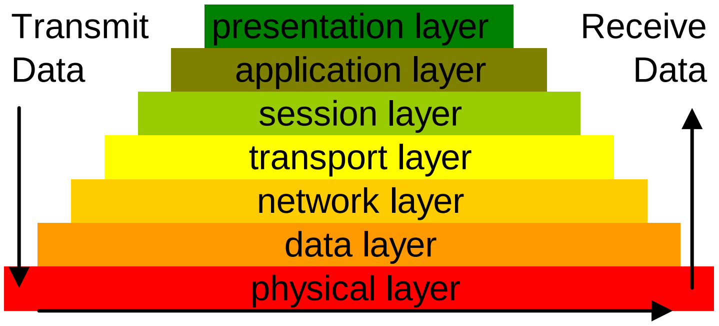
To sum up, this section has not only provided a technical analysis of the Internet but also a glimpse of the underlying geography of the Internet's infrastructure. Most of the technical elements presented in this section are graphically represented in Figures 2.1a and 2.1b below. The analysis presented here is crucial for supporting the main focus of this study: the geography of the Internet infrastructure. This technical analysis supports both the literature review presented here but it also helps to better understand the empirical analysis of this doctoral thesis.

Figure 2.1a: Internet architecture



Source: Adapted from AMS-IX, 2009

Figure 2.1b: OSI reference model



2.3 The Internet geographies

2.3.1 The Internet geography – an epistemological discussion

The main theoretical pillar of the present study is the rather new branch of communications geography identified as *cybergeography* or *Internet geography*. The first term, which is the older one, is based on the novel term *cyberspace*. Indeed, the term was introduced by William Gibson in his novel Neuromancer (1984, 51) in order to describe a virtual conceptual space, existing within ICTs (Dodge and Kitchin 2000). The etymology of the word goes back to the ancient Greek word *kyber* which means to navigate (ibid). This term has been much used by Martin Dodge and it was the title of his extensive and seminal research project about the mapping of cyberspace (for a synopsis of his work see Dodge 2008). The second term appears to be more generic and less connected with the novel discussion about the cyberspace. Recently, it appears more often in the relevant literature (for example Townsend 2003; Zook 2006) and this is the one which is used for this study.

Regardless of their etymological differences though, both the terms focus on the same problem: the geographical representation and analysis of this new virtual space, cyberspace. Dodge and Kitchin (2000, 1) illustrate this new form of space:

> "At present, cyberspace does not consist of one homogenous space; it is a myriad of rapidly expanding cyberspaces, each providing a different form of digital interaction and communication. In general, these spaces can be categorised into those existing within the technologies of the Internet, those within virtual reality, and conventional telecommunications such as the phone and the fax, although because there is a rapid convergence of technologies new hybrid spaces are emerging".

In his previous work Kitchin (1998b) approached cyberspaces as a multiple layer formation, which provides new virtual sites. Such sites are superimposed over and coexisting with traditional geographical spaces.

For Batty (1997), cyberspace is one of the four elements of what he identifies as *virtual geography,* which is the result of technology changes and usage on the traditional geography. More specifically, his typology of virtual geography consists of: (1) the *place/space*, which refers to the ordinary geographical domain; (2) the *cspace* or *computer space*, which is the space inside the computers; (3) the *cyberspace*, which is the new emerging space produced by the use of computers; and (4) the *cyberplace*, which refers to the impact of the infrastructure of cyberspace on the infrastructure of the traditional place.

The latter is the connecting point with what Castells identified as the *space of flows*. In his work about the *network society* (Castells 1996), he illustrated the emergence of a new spatial form, because of the structural transformation that our society is undergoing after the rapid changes and the extensive use of ICTs. He calls this new spatial form the *space of flows* and he defines it as the "managerial organization of time-sharing social practices that work through flows" (ibid, 442). He continues by defining these flows as "purposeful, repetitive, programmable sequences of exchange and interaction between physically disjoint positions held by social actors in the economic, political, and symbolic structures of society" (ibid, 442). In order to better describe this new spatial form, Castells further analyses the space of flows into its components, illustrated as a three layer-based system. The first layer can be parallelized with Batty's cyberplace (Malecki 2002a) and it consists of the technical network infrastructure, upon which the flows of Castells' network society are transported. This infrastructural layer of communications is the "fundamental spatial configuration […] and defines the new space, very much like railways defined *economic regions* and *national markets* in the industrial economy" (Castells 1996, 433). The spatial configuration of this first layer of the space of flows is the focus of this doctoral thesis.

The second layer refers to the actual hubs and nodes of the space of flows. These are the actual places with "well-defined social, cultural, physical, and functional characteristics" (ibid, 443) which are interlinked through the technical layer of the space of flows. An example for this layer is the global financial network, which consists of specific places around the world where global financial markets are located. Lastly, the third layer of the space of flows refers to "the spatial organization of the dominant managerial elites" (ibid, 433). It describes the networked layer of these elites, who increasingly locate themselves in isolated communities but in highly connected places.

While Castells highlighted the importance of the first layer as an underpinning layer of the space of flows, his analysis was mostly focused on the upper layers. The reasons for this can be found in the next section and summarised as lack of data and technical complexity. Building upon his seminal work here, the focus of the analysis is the first layer of the space of flows, overcoming the data availability and analysis related difficulties.

The above two theoretical approaches can be easily considered as the theoretical fundamentals of this emerging field of the Internet geography. Zook (2006, 69) trying to define this field comments:

> "Thus, just as geographers view the recursive link between nature and society as the source of the variation of human experience over the Earth's surface, Internet geographers look to the complexity of the interaction between electronic technology and human use as the origin of the multiplicity of Internet geographies".

Zook's thesis is that there is not only one Geography (with capital G) of the Internet but instead there are many different geographies (ibid). Zook's thesis builds upon Townsend's primary lessons from the research in cybergeography and furthermore his taxonomy fits with these lessons as presented below (Townsend 2003, 30):

- "The internal structure of digital networks is complex and often chaotic but understandable".
- "There is structure to the relationships between virtual and physical places".
- "Cybermaps, like maps of physical space, can provide useful metaphors for clarifying or obscuring our understanding of the structure of cyberspace".

Thus, Zook (2006) in his taxonomy identifies three main categories of the Internet geographies:

- the *technical* geographies of the Internet, which focus on the spatial aspect of the physical infrastructure of the Internet and was identified before as the first layer of the space of flows or as the cyberplace.
- the *human* geographies of the Internet, which are further divided into *political and cultural* geographies and *economic* geographies of the Internet. The former examines the social nature of the Internet and the impact that the Internet's extensive and diverse usage has on places. Furthermore, the Internet economic geographies are subdivided into *urban economic* geographies and *e-commerce* geographies. The latter refers to the re-organization of the geographies of production and consumption because of electronic commerce. The urban economic geographies are approached by Zook (ibid) as the study of the impact that telecommunications and more specifically the Internet have on urban development. However, most of the empirical studies identified by Zook as Internet urban economic geography studies are mostly based on the study of the physical infrastructure of the Internet, signifying the importance of this layer.
- the *visualized* geographies of the Internet. This last division focuses on visualizing and mapping the topology and even the physical location of the technical, political, cultural and economic layers of the Internet.

In order to bypass the complexity of the above taxonomy and mainly to avoid the confusion of the vague borders between urban economic geographies and the technical geographies of the Internet we will try to merge these two. Such an approach is not new in the field. Indeed, Malecki (2002a, 401) in his description of the cyberplace, recognizes that this element of the virtual geography fits best with the research questions that economic geography focuses on as the cyberspace relies on cyberplaces' real world's fixities (Kitchin 1998a and 1998b).

In accordance with Malecki's approach, Greenstein (2004) further defined this field as the *economic geography of the Internet infrastructure*. Firstly, he defines the Internet infrastructure as the (ibid, 5):

> "durable investments in software, communication and computing equipment, and related activities associated with operating information technology. This common and broad definition of Internet infrastructure encompasses quite a lot: capital equipment – such as mainframes, minicomputers, PCs (personal computers), LANs (local area networks), WANs (wide area networks), local and long-distance telephone equipment, private and quasi-public switching equipment, wireless networks for data transmission – and software – both packaged and customized. Notice that it also incorporates human capital, a key (and often local) input along any value chain for Internet services".

Gorman and Malecki (2002, 391) are more explicit and define as the Internet infrastructure only the Internet's physical layer: "the twisted pair wires to the house, the fibre lines to the central office, the switch or router, the 28,000km fibre trunks that connect Japan to England and all the gear in between" (Stephenson 1996; Gilder 2000). For the needs of this study, the Internet infrastructure is directly linked with the Internet's physical layer.

In his study, Greenstein (2004) uses industrial economics in order to explain the location of the Internet infrastructure in the USA. He elucidates the geographic properties of the various elements of the infrastructural layer of the Internet, such as the POPs, the backbone networks, the domain names, the broadband connections etc., using the "economies of density and scale in operation, the economies of entry into services with high sunk costs, and the economics of retrofitting technical upgrades on existing infrastructure [… and] economics of competitive behaviour for growing markets" (ibid, 2).

However, regardless of the economic geography starting point of the study about the Internet infrastructure, because of the strong urban orientation of both the cyberplace and cyberspace, it is unavoidable for such studies not to cross the borders with the field of urban studies. It is common for such studies to use tools such as urban hierarchies and urban networks in order to approach the geography of the Internet infrastructure, as will be illustrated below.

To sum up the above epistemological discussion and to further define the focus of the present doctoral thesis, in the last 10 years we are able to talk about this new emerging field of the *Internet Geography*, which is concerned with the geographical analysis of cyberspace and cyberplace. The different layers of the Internet phenomenon, which are included in the above terms, are characterised by very diverse content – from the technical infrastructure, to the users, the content, the social activism and interpersonal relations. Therefore there is a need for using very different theoretical and methodological tools in order to approach them. Just like traditional geography, the diversity of the Internet itself and mostly the diversity of the Internet geographies require from the researcher the usage of a variety of theoretical and methodological tools regarding the very specific subject area. In this frame, the present study is located in this new field of the Internet geography, focuses on the infrastructural layer of the Internet and uses mostly elements from economic geography and urban studies, as they are illustrated in the following sections of this chapter. In a few words, this study's focus is the *urban economic geography of the Internet infrastructure*.

2.3.2 An urban economic geography of the Internet infrastructure – a general approach

There are a few points which need to be highlighted in regards to the Internet infrastructure from the economic geography and urban studies points of view. Firstly, the geography of telecommunications attracts limited interest from geographers. The technical and intangible nature of telecommunications is the main reason why geographers tend to ignore this subject. Indeed, economic and urban geography usually deals with tangible objects, contrary to the elusive nature of telecommunications and specifically the Internet (Bakis 1991; Hepworth 1989; Kellerman 1993). Telecommunication infrastructure, just like any other network infrastructure, is fairly invisible when it works properly. It only becomes visible when it stops working (Star 1999). In addition, the complex technical structure of the telecommunications infrastructure prevents geographers, planners and regional scientists from considering and researching the topology, structure and design of such networks (Kellerman 1993). Another factor which is also responsible for the lack of interest from geographers is the deficiencies in relevant and accurate data for telecommunications usage and telecommunications infrastructure supply and most of all for the Internet. Secondary data for existing backbone networks is only available nowadays through Telegeography (2007) and only at an aggregated level. This is the data source used for this study. In addition, data for intercity data flows is not available. Despite the fact that ISPs collect such data for managing their networks, such data is not published for competition reasons. The above situation is not new. Batty in the early ‘90s declared that there is no interest in the impact of information flows on cities (1990), Moss in the late ‘80s characterised telecommunications infrastructural networks as a mystery to most of the cities (1987) and Graham and Marvin (1996) admitted that many city planners were not aware of the telecommunications infrastructure supply in their cities. What is more, the privatization of the telecommunications networks which took place across Europe in the late ‘90s reduced the data availability because of the heightened competition (ibid). Nonetheless, the rapid expansion of the Internet in the late ‘90s and the technologies convergence led to the emergence of the field of the Internet geography. But the above pre-Internet era comments are to a degree still valid, signifying the difficulties that the Internet geography field is asked to overcome.

Apart from data availability and the field’s popularity issues, there is a growing discussion about the implications of the Internet for the broader geography of activities and especially for the centralization or decentralization impacts on spatial structure that the Internet may generate. The Internet appears to promote a *double edge effect*, that is of simultaneously stimulating both centrifugal and centripetal forces. If we take the example of rural areas, the former can be identified as the benefits that people in rural areas gain from investments in ICTs, such as the access to cheaper and better quality services, the diffusion of knowledge etc. The centripetal forces have the same source; investments in ICTs in rural areas make local markets more accessible to larger, external businesses (Gorman and Malecki 2000), thereby increasing competition. This may result in diminishing local production because small businesses are unable to compete with larger ones (Richardson and Gillespie 2000). So, not only do ICTs “not automatically result in the decentralization of economic activity” (Richardson and Gillespie 2000, 201) but they can and do have both centralizing and decentralizing effects, contrary to the early *death of distance*

(Cairncross 2001) conceptualisations of their impacts, which saw only their decentralizing potential (Malecki and Gorman 2001). Without compensating public policy actions, ICTs may result in increasing the gap between urban (core) and rural (peripheral) areas (Richardson and Gillespie 2000).

Another important issue is the impact of the extensive use of ICTs and the Internet on cities. Although ICTs have managed to remove some of the geographical barriers that remote locations face, this evolution has not weakened people's and economic activities' tendency to cluster in urban areas (Moss and Townsend 2000). However, opposing opinions have also been stated. One of the most pessimistic views about the future of cities in the post-Internet era was introduced by the US National Research Council (National Research Council 1998):

> "One can anticipate a shift of population away from the metropolitan areas to bucolic agricultural settings (rural Vermont, the California wine country, fishing villages), to resort areas (Aspen, Monterey, Sedona), and to the sunbelt and beachfront. Just as the automobile, superhighways, and trucking helped shift population out of the central city to the suburbs in the 1950's, the computer, the information superhighway, and modems will help shift population from the suburbs to more remote areas."

Contrary to these early arguments that the developments in telecommunications will result in diminishing the importance of cities, population and economic activities tend more and more to agglomerate in core metropolitan regions, leading Malecki (2002a, 419) to conclude that "world cities are alive and well". Even after the rapid technological changes of the 1990s and the 2000s, cities proved not to be the "leftover baggage from the industrial era" (Gilder 1995 cited in Moss and Townsend 2000, 36). And most importantly, the *death of cities* never occurred and the Internet proved to act more as a complement rather than a substitute for face-to-face communications, which is facilitated by cities, and in general by urban agglomeration (Gillespie et al 2001; Kolko 1999).

ICTs and the Internet infrastructure more specifically are not an exception in this centripetal tendency since they concentrate in the important nodes of the world urban network (Sassen 1991 and 2000). Indeed, the Internet and the backbone networks which underpin it, is actually an "urban technology" (Rutherford et al 2004, 1) in as much as it is located primarily in cities, where demand is concentrated (Gorman and Malecki 2000). "The Internet cannot bypass mega-cities: it depends on the telecommunications [the technical layer of which is concentrated there] and on the *telecommunicators* located in those centres” (Castells 1996, 440).

The relation between telecommunications and cities is not unidirectional though. Not only do cities have the ability to shape the spatial structure of ICTs, but also telecommunications play a role in the urban development process. Moss and Townsend (2000, p. 38) illustrate the Internet backbone networks’ spatial pattern among urban areas:

> “Just as the geographic structure of these earlier infrastructure networks [highways and railways], both reflected and influenced existing and desired settlement patterns, the geography of the backbone systems has in part been shaped by the economic and social realities of the late 20th-century America and the specific properties of the technology”.

Graham and Marvin (2001, 15) appear to be positive on the urbanization impact of telecommunications:

> “New highly polarized urban landscapes are emerging where ‘premium’ infrastructure networks – high-speed telecommunications, ‘smart’ highways, global airline networks – selectively connect together the most favoured users and places, both within and between cities”.

Apart from the bidirectional relationship between infrastructural networks and urbanization, the above metaphor shifts the focus to another subject: the well established parallel between telecommunications and transportation networks. Even though this parallelism is commonly observed in the literature, the effort committed for the study of the Internet and generally ICTs as network infrastructure’s cannot be compared with the much greater interest for transportation networks (Moss and Townsend 2000).

This parallel takes place at two levels: economy and topology. From the economy point of view, the Internet, just like transportation networks, is an infrastructure as they both serve the production process. “Similar to the transportation networks of the past two centuries (rail, road, air, water), the Internet transports the valuable weightless goods of the digital economy: information, knowledge and communication” (O’Kelly and Grubesic 2002, 537). The higher the dependence of the economy on electronic transactions, the more the value of the Internet as an infrastructure will be acknowledged. Borland & Hu (2004, 2) highlight this point by arguing that broadband connections are fundamental for the future of the economy just like road and train networks during the past two centuries. And they continue:

> "[Railway and highways] transformed the way people lived and worked, irrevocably changing human conceptions of distance, speed and time. Even in its relative infancy, broadband is already having much the same effect".

From a more geographical point of view, transportation and telecommunications networks have strong physical links, since the latter are usually found superimposed on the former. Telecommunication carriers, in order to roll out their intra- and inter-city networks, use pre-existing infrastructural networks such as sewer systems and transportation networks (Graham and Marvin 1996, 282). It is common for backbone networks to be embedded by motorways or railway lines and for MANs to be installed underneath streets and even inside old sewer pipes.

Another link between transport and telecommunications infrastructure is the commercial partnerships between their operators. Because of the privatization of the infrastructural networks and their splintering character (Graham and Marvin 2001), it is common for older infrastructure network operators (i.e. transport or energy) to establish commercial partnerships with telecoms or even to start providing telecommunications services. By such initiatives, the new provider is directly benefitted by the economies of scale arising because of the use of the old infrastructural networks and avoiding the high sunk costs related with excavations (Graham and Marvin 1996, 282).

Another commonality is the regulatory status (ibid). Historically, network infrastructure was developed in a natural monopoly framework because of the market failure in infrastructure provision (Banister and Berechman 2003). This monopolistic framework was accompanied by a regulatory framework in order to prevent customers' over-exploitation (Graham and Marvin 1996). Recent technological developments as well as changes in dominant political economy views resulted in more liberal regulatory frameworks. Despite of any differences in the liberalisation process that the two infrastructural networks underwent, the development and the function of both networks was always and still is related with some kind of regulation.

It should be noted here that the relationship between telecommunications and transportation has been questioned by the scientific research of the last twenty years.

Hence, intensive debates took place in the literature in the 1980s and 1990s (for a detailed review see Graham and Marvin 1996), introducing sometimes rather deterministic argumentations. In addition to other similarities described in this section, both of them are *friction reducing technologies* as they reduce the cost of distance (Cohen et al 2002; Cohen-Blankshtain and Nijkamp 2004). Because of this similarity, effort has been spent in research in order to define the relation between them. The literature (Salomon 1986; Banister and Stead 2004; Mokhtarian, 1990 and 2002) suggests four possible types of interaction between them: "substitution (reduction, elimination), complementarity (stimulation, generation), modification (change time, mode, destination, and so on with respect to a trip or communication that would have occurred otherwise), and neutrality (no impact of one medium on the other, e.g. as many e-mail messages have no impact on travel and conversely)" (Cho and Mokhtarian 2007, 5). Early argumentation was in favor of vast substitution effects on transport because of the ICT's expansion. In reality, these effects were never observed and nowadays it cannot be claimed that the telecommunications infrastructure has a substitution effect on the demand for physical transportation (Black and Nijkamp 2006). On the contrary, complementarities and synergies have been developed between the two infrastructural networks as the demand and the supply side of both of them have met considerable increase (Gillespie et al 2001; Banister and Stead 2004). The extensive use of ICTs and the Internet affected the pattern of transportation of goods and people. While the use of teleconferencing was always seen as a substitute for traveling to business meetings, at the same time teleconferencing can result in more social contacts and accordingly in more trips in the long term (Geels and Smit 2000).

From the topological point of view, both transport infrastructure and the Internet backbone are rolled out as networks (for example: Gorman and Malecki 2000; O'Kelly and Grubesic 2002; Wheeler and O'Kelly 1999). Both consist of nodes and edges and both of them can be analyzed using network techniques (Malecki and Gorman 2001). Grubesic and O'Kelly (2002, 264-65) illustrate this similarity:

> […] imagine a series of interstate highways (fibre-optic backbones) converging at a single cloverleaf junction. At this location, cars (data packets) are allowed to continue their journey on the existing highway

(backbone 1), or they may switch highways (backbone 2, 3… n) using directional information provided by the interstate signs (routers).

Table 2.1 presents this analogy: if the backbone links symbolize the motorways, the IXPs and POPs represent the transport nodes (interchanges and access nodes) and the MANs and the local loops the intra-city roads, then the IP address stand for the numerous final destinations in the cities – the “Internet real estate” according to Dodge and Shiode (2000).

Table 2.1: The parallel between the Internet physical infrastructure and the road infrastructure

Importance at	The Internet infrastructure		Road infrastructure
Inter-city level	Backbone networks	↔	Motorways
	IXPs / private peering points	↔	Interchanges
Intra-city level	POPs	↔	Access nodes
	MANs / local loops	↔	Intra-city road networks
	IP addresses	↔	Premises

The Internet backbone network is the most interesting part of the infrastructural layer from the geography point of view, because it enables the interconnection of remote places, almost by surpassing the friction of distance. Backbone networks can be regarded as the infrastructural underpinning that enables the Internet to function, seamlessly and apparently place-lessly from the viewpoint of the user. According to Malecki (2004, p. 24):

> “The backbone networks […] are the core of the Internet and are essential for all but the most local of interactions. Although there is no consensus as to which networks are backbones, the following applies: A backbone is a set of paths that local area networks (LANs) connect to for long-distance connection. A backbone employs the highest-speed transmission paths in the network.”

One of the basic attributes of the Internet considered above is that it interconnects numerous different and widely dispersed networks. This attribute, which is responsible for its global character, only occurs because of the existence of backbone networks. In reality, backbone networks are extensive interregional networks, built on fibre optic cables, which are interconnected at the main nodes of the Internet, where data peering between them takes place. Batty (1991, 142) defines them as “a kind of electronic superhighway which enables networks at the next level of hierarchy down to be interconnected”.

So, it would not be an exaggeration to suggest that the study of the geography of Internet backbone networks is synonymous with the study of the Internet’s spatial dimensions or with the geography of the 1st layer of the space of flows. Indeed, the Internet backbone network with the global aviation network are the main elements of Castells’s first layer of the space of flows (Taylor 2004).

To sum up, despite the lack of data and the technical complexity of the Internet infrastructure, there is a growing discussion in the literature about the geography of this infrastructural network, as a representation of the first layer of the space of flows. It seems that the early arguments about the end of cities and the death of distance proved to be overly futuristic. On the contrary, the result of the extensive use of ICTs and the Internet more specifically is a double edge effect, with both the cities and the urban network being affected by but also affecting the spatial structure of the Internet infrastructure. The next section will shed more light on this by reviewing empirical studies about the geography of the Internet infrastructure.

2.3.3 Review of empirical studies on the geography of the Internet infrastructure

In this section a review of empirical studies about the geography of the Internet infrastructure takes place. Table 2.2 presents the majority of the papers that have been published from the early days of this emerging field (late 1990s) until recently. It is divided in three sections. The first one contains studies about the Internet edges and mainly the backbone networks. The second section includes the papers which are focused on the nodes of the Internet such as IXPs, the POPs, the colocation facilities and the towers for wireless telecommunications. The studies of the third section focus on the geographical analysis of domain names. The common characteristic of all of these papers reviewed here is that the Internet infrastructure is examined from the geography point of view. In other words, it is not the topology of the Internet infrastructure that is under question but rather the reflection of this topology on the physical world and mostly on cities and the urban network.

Table 2.2: Empirical studies on the geography of the Internet Infrastructure

Network element	Study	Region	Network element	Spatial unit	Indicator	Time	Methodology
edges	Wheeler and O'Kelly 1999	USA	backbone	city, backbone networks	tc	1997	graph theory
	Gorman and Malecki 2000	USA	backbone	city	tc, tb, network distance	1998	descriptive statistical analysis, graph theory
	Moss and Townsend 2000	USA	backbone	city	tb	1997-1999	descriptive statistical analysis
	Malecki and Gorman 2001	USA	backbone	city	tc, tb number of hops	1998	descriptive statistical analysis
	Townsend 2001a	World	backbone	city	tb	2000	descriptive analysis
	Townsend 2001b	USA	backbone	city	tc, tb, domains	1997, 1999	descriptive, statistical analysis
	Malecki 2002a	Europe	backbone, colocation facilities	city	tc, tb, colocation points	2000	descriptive statistical analysis, OLS for explaining city bandwidth distribution using only population
		Europe, Asia, Africa, Americas	IXP	continent	peering points	2000	
		USA	backbone, IXP, colocation facilities	city	tc, tb, b, colocation points	1997-2000	
	O'Kelly and Grubesic 2002	USA	backbone	backbone networks, city	c, tc	1997-2000	graph theory, descriptive statistical analysis
	Gorman and Kulkarni 2004	USA	backbone	city	tb, tc, c	1997-2000	network analysis, complex networks

	Malecki 2004	USA	backbone	city	tb, b	1997-2000	descriptive statistical analysis, OLS for explaining city bandwidth distribution and web design firms location
	Rutherford et al. 2004	Europe	backbone	city	b, tc, tb	2001	descriptive statistical analysis, rank plots
	Schintler et al. 2004	Europe, USA	backbone	city	tc	2001, 2003	complex networks
	Rutherford et al. 2005	Europe	backbone	city	c, tc, tb	2001, 2003	descriptive statistical analysis, rank plots
	Devriendt et al 2008	Europe	IXPs	city	intercity links based on IXPs presense and google.com	2001, 2006	
	Rutherford forthcoming	Europe	backbone	city	c, tc, tb	2001, 2004	descriptive statistical analysis, rank plots
nodes	Evans_Cowley et al 2002	USA	Telecom Hotels	city	number of Telecom Hotels	2001	descriptive statistical analysis, planning authorities responses
	Grubesic and O'Kelly 2002	USA	POP	city	number of POPs	1997-2000	descriptive statistical analysis
	Gorman and McIntee 2003	USA	Personal Communication Service Towers (wireless)	city	number of PCS towers	-	descriptive statistical analysis, OLS for PCS towers location

	D'Ignazio and Giovannetti 2007	World	IXPs	IXPs	c	2004-5	econometric analysis
domains	Moss and Townsend 1997	USA	domain names	city, intra-city	numbers of domains, domain density	1993-1997	descriptive statistical analysis
	Dodge and Shiode 2000	UK	domain names	city	numbers of domains, domain density	1997	descriptive statistical analysis
	Zook 2000	USA	domain names	city, intra-city	numbers of domains, domain density	1998	descriptive statistical analysis
	Zook 2001	World	domain names	country, city	numbers of domains, domain density	1999	descriptive statistical analysis

c = connections, tc = total connections, b = bandwidth, tb = total bandwidth

The first observation from this table is the rather small number of studies: only 23 papers. However, these 23 papers were published during a 12-years period (1997-2009) with the majority of them being published in the early 2000s. Indeed, it seems that there was a peak at this point and since then there is one paper per year published. The second point that should be highlighted is the geographical focus of the studies. Most of them are concerned with US cities. Such a bias was expected not only because the Internet itself originated in the USA and that ARPANET, today's Internet ancestor, was rolled out among a few American cities, but also because of the US leadership in telecommunications after the cold-war era (Kellerman 2002).

Another general comment is that most of the studies are concerned with the edges of the Internet rather than its nodes. As a result, the importance of the nodes is underplayed both from the Internet function but also from the urban and economic geography point of view.

With respect to the studies' main indicator, it seems that the papers which are concerned with the backbone networks are mostly focused on the number of different links terminating in each city and the total bandwidth accumulated at city level. These indicators highlight the city's capacity in cyberplace and it could be said that they represent the infrastructural capacity. What these indicators cannot do is to examine the one to one relations between cities. The total number of connections and bandwidth between any two cities may reflect the data that these two cities can potentially interchange and in some way the volume of the interactions that might take place between these two cities[3].

Additionally, most of the studies use data which usually refers to late 1990s – early 2000s. Data from this period though is unlikely to reflect current conditions in the geography of the backbone networks for two reasons. Firstly, regardless of the high sunk cost of the backbone networks, their upgrade (i.e. lighting up the fibre) is easy and for this reason the spatial distribution of bandwidth capacity can change dramatically. Secondly, conditions in the telecommunications industry have changed significantly since then. The dotcom bubble of the early 2000 was followed by the telecommunications crash. According to the Economist (2002) the latter was some ten times bigger than the better-known dotcom crash: "the rise and fall of telecoms may indeed qualify as the largest bubble in history". Indeed, one of the reasons of

---

[3] Nevertheless, the above hypothesis has some limitations, which will be discussed in detail in Chapter 3, where this study's main data is presented.

this bust was the “unrealistic expectation of demand for network capacity” which resulted in overbuilt backbone networks (Kam 2006, 508). All in all, the conclusion of the papers presented here only reflect the conditions of reference time point and need to be used carefully for other time periods.

A common point for all the papers is the strong urban character of the Internet infrastructure. Because of the private character of this infrastructure, the Internet’s physical layer is located where the demand puts it; and the demand for such infrastructure is concentrated in large urban areas (Malecki 2002b; Priemus 2007). “The Internet is not a utopian public good available to everyone, whether core or periphery and, furthermore, it is not available at the same level of technology and service to all locations (Gorman and Malecki 2000, 132; Fortune 1999).

At a wider scale, it could be said that the Internet on the one hand reinforces existing globalization patterns and on the other results in the emergence of new clusters (Malecki 2002a). The global cities are always in the first tier of the most connected cities, but this tier is not anymore a monopoly of the handful of well established global cities. Both in Europe and USA, the new urban hierarchy resulting from the agglomeration of Internet infrastructure appears to be notably different from the traditional urban geography.

For the case of the US, both old and new geographies coexist. The group of the most connected cities on the US commercial Internet changed very little between 1997 and 2000. New York, Chicago, Washington, D.C., San Francisco, Dallas, Atlanta, Los Angeles accumulated the most bandwidth in 2000. Four years earlier, the same seven cities were in the top-tier, but with different order and with New York being the fourth city (Malecki 2002a; Grubesic and O'Kelly 2002; O’Kelly and Grubesic 2002). Yet, cities which are traditionally significant in transportation networks and information flows such as Washington, D.C., Dallas and Atlanta (Wheeler and O'Kelly 1999), accumulated more bandwidth than Los Angeles, one of the most important nodes of the US urban network and even New York was not for the first three years of the study period served by the highest capacity links. The above led Townsend (2001b) to conclude that for the case of the US the scatter of the Internet infrastructure is wider than the world cities hypothesis would have predicted.

Indeed, apart from the first tier cities, the main changes in the Internet infrastructure based US urban hierarchy appeared in the lower tier cities. Portland, Kansas City, St Louis, and Salt Lake City became important nodes of the US backbone network either because of their location near existing transportation corridors or because of a strong local information technology economic base – the case of Portland (O'Kelly and Grubesic 2002). "Although results indicate that the *big 7* will probably continue their dominance in network accessibility, it is possible for smaller cities to make significant jumps in the rankings" (ibid, 548).

Similar spatial patterns, but slightly more dispersed, have been exhibited for Europe. Apart from the two dominant European world cities, i.e. London and Paris, significant bandwidth and backbone links are concentrated in other cities such as Amsterdam, Brussels, Lyon, Milan, and four or five German cities, highlighting the more diffused spatial pattern of the European fraction of the Internet backbone network, when compared with the US one (Rutherford et al 2004). Also important is the role of some *gateway* cities, such as Copenhagen, Vienna, and Prague, which act as hubs for peripheral regions – Nordic countries and Eastern Europe respectively (ibid). Rutherford et al (2004, 29) conclude by saying that:

> "at the end of the day, and even taking the recent market restructuring and consolidations into account, we can suggest that the major European Internet backbones rely on a minimum of 12–15 cities to deliver high-bandwidth networks and services across Europe."

Devriendt et al's (2008, 25) findings about the European cyberplace are slightly more differentiated. Based on European IXPs data they suggest that "Amsterdam, London, and Frankfurt are far more important in their gateway functions than Paris, Brussels, and Dusseldorf." However, their content-based analysis of the results of web-searches showed that Paris, London and Berlin have the most important links and cities such as Amsterdam, Rome and Frankfurt are secondary (ibid).

Regarding the Internet real estate approach (Dodge and Shiode 2000), which is focused on the geo-location of Internet domain names, all the relevant studies recognize again the impact of agglomeration forces on the spatial pattern. Moss and Townsend in their pioneering paper (1997) highlighted the dominance of New York City and mostly of Manhattan in US domain names production. In addition, Dodge and Shiode (2000) also identified the concentration of domain names around London. In a wider study, Zook (2001) recognizes the US dominance in domain names allocation, regardless of their global diffusion. According to his research, global cities are still important nodes of Internet content production, but at the same time, other cities such as San Francisco, San Diego and Austin in the US, and Zurich, Vancouver and Oslo globally emerge as major Internet content producers (ibid).

Another important point that some of the above papers highlighted, is the explanatory analysis of the spatial distribution of the Internet infrastructure. Two of the papers went a step beyond using population as the main explanatory variable for the Internet infrastructure. Indeed both of them (Malecki 2004; O'Kelly and Grubesic 2002) recognized the explanatory value of knowledge related variables in bandwidth accumulation. Variables such as the numbers of doctoral-granting institutions and patents proved to be better regressors than population in explaining bandwidth distribution.

From the methodology point of view, most of the studies use descriptive statistical analysis and mapping. Nonetheless, a few papers exploit more advanced methods to approach their research questions such as: graph theory, network analysis, rank plots and complex networks in order to better explain the network structure; OLS for the explanatory analysis of the infrastructure's spatial distribution; and econometrics for the effect of distance on ISPs interconnection.

With respect to the nodal infrastructure, only a few studies focused on them. Yet, the spatial pattern of this physical element of the Internet infrastructure is not much differentiated with the edges elements. Grubesic and O'Kelly (2002) concluded that POPs are unevenly spread in the US and Gorman and McIntee (2003) highlighted the fact that just as with backbone networks, the wireless infrastructure, in the form of nodes of the wireless networks, followed the diversely located demand. Evans et al (2002, 16) went a step further and classified the city-planners responses to this new privately driven infrastructure:

> "Some cities have been pro-technology and tried to assist telecom hotels in their development. A second set of cities responded after seeing telecom hotels enter an area of the city where planners did not believe they were a good fit. The third group simply decided that telecom hotels were similar to other uses already existing in the city"

To sum up, after ten years of empirical research on this rather narrow but still emerging field of the urban economic geography of the Internet infrastructure, some first results can be drawn. Based mainly on the research for US cities and also on the few studies about Europe, it seems that the implementation of this new infrastructural layer results in an urban geography which is partially new and partially based on the traditional urban hierarchies. This conclusion is mostly based on research focused on cities' infrastructural supply, as it concerns the urban accumulation of Internet backbone connections and bandwidth.

2.3.4 The contribution to the emerging field of Internet geography

In short, this doctoral thesis, which is situated in the emerging field of urban economic geography of the Internet infrastructure, will contribute to this field in the following ways:

- Firstly, by focusing on Europe, it is intended that this study will broaden the knowledge of how European cities are interconnected through the Internet backbone network. As mentioned above, some research on Europe has already taken place, but it is rather limited in comparison with the research about the US cities, and it is also out of date. Therefore, there is a need for further work exploring the way the European urban network is interconnected through this new infrastructural layer. This doctoral thesis will bring to light up to date results based on a dynamic analysis. The latter will enable us to eliminate the impact of specific events such as the early 2000s telecoms crash.
- In addition, this research will not only focus on the infrastructural capital approach and on the way the Internet infrastructure is distributed across European cities, but it will go a step further and include in the analysis a more relational approach, in order to identify the different roles the cities perform in the European part of the first layer of the space of flows.
- Furthermore, based on the analytical similarities between the Internet backbone networks and the transport infrastructure, but also based on the fact that the first layer of the space of flows mostly consists of the Internet backbone and aviation networks, effort is spent to explore how these two different infrastructural layers are deployed across the European cities, and the synergies and the complementarities between them.
- Moreover, this doctoral thesis research will also focus on identifying the factors that shape the spatial distribution of the Internet backbone network across European cities. Such research is limited, and has only taken place for US cities. Nevertheless, the results of these studies cannot be directly applied to Europe because of the obvious geographical and socio-economic differences.

- Apart from the above direct contributions to the emerging field of the urban economic geography of the Internet infrastructure, this study will also contribute to the field of the world cities research and to regional development studies. The inter-urban relational nature of the Internet backbone network and its identification as the main component of the first layer of the space of flows facilitate the global city process and the world city network. Moreover, the infrastructural character of the Internet backbone network in addition to its tendency to accumulate in specific nodes of the urban network might result in spatially differentiated developmental results. The above literatures are examined in the following two sections.
- In order to approach all the above, this doctoral thesis will use a wide range of quantitative methods and will attempt to go a step further than the descriptive approach of most of the existing studies in this field.

2.4 World cities

2.4.1 Introduction

The second theoretical pillar of this study is the *world city* literature. Different terms have been used in order to describe this contemporary phenomenon which is related with the growing interaction and interdependence among a selected set of cities, the importance of which emerges not only within the border of their national economy, but also in the frame of the globalized economy. Among others, Friedmann (1986) refers to the *world city hypothesis*, Sassen (1991) recognizes *global cities*, Castells highlights the *global city process* and Taylor (2004) analyzes the *world city network*. Peter Hall almost 30 years ago approached world cities as entities which perform multiple roles (Hall 1966, see also Hall 1998, 17): they are national and international centres of political power, centres of trade, banking, insurance and related financial services, centres of advanced professional activity of all kind, centres of knowledge and technology, information gathering and diffusion, centres of consumption, centres of arts, culture and entertainment, and of the ancillary activities that cater for them.

2.4.2 World city hypothesis

However, it was not until the mid-1980s when the discussion about cities with global reach was materialized to something more concrete; and that was the *world city hypothesis* of the “spatial organization of the new international division of labour” suggested by John Friedmann (1986, 69). As he admits, this hypothesis was not a robust theory which links urbanization with the global economy but rather a starting point for research (ibid). Some ten years after his path-breaking work, Friedmann (1995, 22) returned with a revised version of his hypothesis, approaching cities as “spatially organized socio-economic systems”. According to this work, the world city hypothesis can be summarized in the following five points (ibid, 25):

- “World cities articulate regional, national, and international economies into a global economy”. The main role that these cities perform is to act as the key nodes of the global economic system. Over the last 30 years, this discrete role the world cities carry out has increased because of the economic transformation that took place during this period: as Amin and Thrift (1992) claim, between the 1970s and 1980s the universal economic system shifted from an international to a global economy. The intensity of the global economic interactions and their importance for the global economic system and the globally (inter)linked national (sub-global) systems resulted in empowering the world cities.

- “A space of global capital accumulation exists, but it is smaller than the world as a whole”. Regardless of the growing interaction of the world cities because of the globalisation of the economy, not every corner of the world is included, at least with the same intensity, in this planetary economic system. As Sassen (1991) highlights, the degree of globalisation goes hand in hand with the increase of central functions’ concentration in a few locations, know as the global cities.

- “World cities are large urbanized spaces of intensive economic and social interaction”. Undoubtedly, world cities are extensive metropolitan areas, with large pools of labour power and high densities of economic and social activities.

- “World cities can be arranged hierarchically, roughly in accord with the economic power they command”. This hierarchical structure of world cities is one the main points of Friedmann’s hypothesis. Based on this he created a taxonomy of world cities distinguishing them as primary and secondary in core and semi-periphery countries (Friedann 1986). In 1995 he revised this taxonomy. The new “hierarchy of spatial articulations” was based on “global financial articulations”, “multinational articulations”, “important national articulations” and “subnational/regional articulations” (Friedmann 1995, 23-4), probably influenced by Sassen’s *global city* approach, the main element of which is the financial power of the global cities (Sassen 1991). Such hierarchical relations can be read as relations of power and competition, for instance in attracting Foreign Direct Investments (FDI) or headquarters of important multinational firms. What is interesting though, is that such a hierarchy cannot be stable over time (Friedmann 1995); the complexity of the globalized economy does not allow the standardization of a global hierarchy. The main exemption to this is probably the highest tier of cities, which consists of New York, London and Tokyo. These cities are identified as global cities by Sassen (1991) and through the short history of the world city research their dominance remained unchallenged.

- “The controlling world city strata constitute a social class that has been called the international capitalist class”. The main characteristic of this class is its cosmopolitan view of the world, extensive usage of the English language and its consumerist ideology (Friedmann 1995, 23).

Following Taylor’s (2004) comments on Friedmann’s work, two points need to be highlighted. Firstly, the world cities hypothesis gave the necessary push to include the world cities and their links within the urban research agenda. It was the first time that a theoretical framework for this field has been introduced, even in the form of a hypothesis. From the geography point of view though, Friedmann’s main contribution was that he set up the frame of the world city links at a global level, surpassing state boundaries and the already existing and extensive literature about national urban systems (ibid). It should be noted though that despite the importance of the hierarchical relations among the world cities in the world cities hypothesis, the criteria and the methodology behind this ranking are not explicitly specified, especially in the revised version (Friedmann 1995, 24).

2.4.3 Global cities

The real growth though of the world city studies field took place in the early 1990s, after Saskia Sassen's study of the global city (Sassen 1991). Her starting point is the opposed forces of spatial dispersal and global interaction. Building on previous approaches for world cities as international centres of trade and banking, she identifies four new functions and roles for such cities (ibid, 3-4): firstly, they host the decision centres of the world economy; secondly, the financial and specialised service firms, which have followed manufacture as the leading sector, tend to concentrate in these locations; thirdly, these cities are important for the global production in the leading sectors, including the production of innovations; and fourthly, the consumption of the products of the leading sectors – including innovations – mostly takes place there. In a few words, "the things that a global city makes are services and financial goods" (Sassen 1991, 4). All the above resulted in control and power concentration in specific cities, the leading league of which consists of New York, London and Tokyo.

Sassen (1991) did not extend her research outside of the leading league of global cities and consequently does not provide empirical results for a wider set of world cities. As Taylor (2004) highlights, she is not much concerned with the relations among the world cities – no matter that this is slightly changed in the revised version of her study (Sassen 2004) – and she rather performs a comparative study of the three global cities. Interestingly enough she does not adopt the most common term world cities, but she chooses to use the term global city to differentiate her leading league of cities from similar past studies such as the world cities hypothesis (ibid).

Nonetheless, the most valuable input of Sassen's study in this doctoral thesis is the comprehension of ICTs role in supporting the global cities. ICTs are essential in the two main processes which aid the spatial concentration of control and ownership: both the spatial dispersion of economic activity and the reorganization of the financial industry are strongly based on ICTs. Such infrastructure enables the long distance management of production and instant financial transactions, regardless of the physical distance. In addition, Sassen (1991) highlights the agglomerative character of ICTs as well as their developmental impact: the high entry cost for providing extensive ICTs infrastructure is an agglomerative factor itself since not all the cities can afford such an investment; yet, she continues, the established provision of high quality ICTs equals to “an almost absolute advantage” for a city (ibid, 19).

2.4.4 Global city process

A different approach, and rather more influential for this doctoral study about the world or global cities, has been presented by Manuel Castells. According to him (1996, 417):

> "the global city is not a place but a process. A process by which centres of production and consumption of advanced services, and their ancillary local societies, are connected in a global network, while simultaneously downplaying the linkages with their hinterlands, on the basis of informational flows".

This process leads to the concentration of economic activities in some global nodes. Castells (ibid, 415) identifies three reasons among others for the continuous and growing concentration: (1) world cities are mostly “information-based, value-production complexes”. The main elements of advanced service production, that is highly skilled labour and suppliers, can be found in these locales; (2) such cities are linked in networks of production and management. The flexibility of such networks enables the advanced service producers to gain access to labour and suppliers when necessary and in the needed quantities, using *a just in time* concept, avoiding the costly internalization of the above inputs of production; (3) such a flexible production model is facilitated by the concentration of production and management networks in specific core cities and the global networking of these core cities and their hinterlands. This networking is dependent on infrastructural networks, such as telecommunications and air-transportation (Castells 1996, 415).

Indeed, Castells adopts a more dynamic approach for the world city phenomenon compared to the rather static view of Sassen. He diffuses the global city phenomenon by accepting that the global city process cannot be limited only to “a few urban cores of the hierarchy” (ibid, 411). On the contrary, this networking architecture can be identified even at much lower scales; regional and local centres at national level, or to be more accurate parts of these centres' socio-economic system, appeared to be linked with the *global economy*. Castells (1996, 77) identifies a global economy as one of the three elements of what he calls the *new economy*. The new economy is this world scale economic system, which appeared in the last quarter of the twentieth century and is "informational, global and networked". The informational character of the new economy is analyzed in the next section, but it can be briefly mentioned here that the new economy is informational and not just information based, just like the industrial economy was something more than just an economy based on the manufacturing. Indeed, the emergence of the industrial economy was accompanied by the emergence of a broader social culture, the industrial culture (ibid).

The second element of Castells' view of the current economic system is its scale. He recognizes scale not as world neither as universal but rather as global and he defines the global economy as "an economy whose core components have the institutional, organizational, and technological capacity to work as a unit in real time, or in a chosen time, on a planetary scale" (Castells 1996, 101-2). This global economy is a further evolution of the *world economy*, which has existed at least since the sixteen century [i.e. the Mediterranean world economy as described by Braudel (1984, 22), see also (Wallerstein 2004)], and which only refers to capital accumulation throughout the world (Castells 1996, 101). Conversely, the global economy refers to the global integration of the actors of capital accumulation.

The integration element leads us to the third characteristic of Castells' approach to the new economy: its network character. The new economy is networked because "under the new historical conditions, productivity is generated through and competition is played out in a global network of interaction between business networks" (Castells 1996, 75). Such networks are global but not universal, meaning that they are spread around the world but they do not include every settlement on earth. On the contrary, they are very selective on which nodes of the world cities network they include.

To wrap up, Castells defines the new economy in accordance with the world city phenomenon and he strongly interrelates these two notions. Regardless of the universal impacts it generates, the new economy is mostly apparent in the cities which experience the global city process. This process though, which is typified by a global scale and a network topology, is mainly based on the recent advances in telecommunications and computing. As Castells states "without new information technology global capitalism would have been a much limited reality" (1996, 19). Throughout his work, Castells highlights the importance of ICTs in supporting the global city process and the new economy and he concludes by including them in the first layer of the space of flows.

2.4.5 World city network

Following up on Castells' work, Peter Taylor (2004) presents his own approach to the world city phenomenon identified as the *world city network*. His main point is the relational thinking about cities. Indeed, his book starts by explaining his relational view on cities: he is not concerned with the relations within the city or even with its hinterland; on the contrary he is focused on the relations between the cities, their "dependencies and interdependencies" (ibid, 1). In this frame, he is focused on how cities, through the networks they form, work together as economic entities (ibid, 1). He suggests that

> "concepts such as space of flows and cities as networked entities are transferable across different historical specificities. Thus, what [he is] basically taking from the above [different approaches of the world city phenomenon] is the necessity to think of cities relationally, as the product of networking activities" (ibid, 27).

Taylor not only understands the emergence of the network logic in the world city phenomenon, but he moves one step further by empirically testing his model. He creates a three level interlocking world city network. Usually, networks are two level entities: they consist of the links and the nodes. Taylor, instead of directly using the cities as the nodal level of his network, creates a third sub-nodal level, in order to include in his analysis the agents which "taken together, are primarily responsible for shaping the world city network"; these are service firms, city governments, service-sector institutions and nation-states (ibid, 58). From these four, he recognises service firms and more specifically Sassen's advanced producer services as the main agent for world city formation. Based on relational data[4] from about 100 multinational

---

[4] According to Taylor (1997), the two different types of relational data that can be found in the literature are the flows between cities and the organisational links between them. For his seminal work presented here (Taylor 2004), the second type of relational data is used.

firms which can be identified as advanced producer services, he empirically analysed the world city network, based not only on hierarchies but mostly on relational data.

Apart from the theoretical value of Taylor's research in highlighting the relational nature of the world cities, his main contribution is the empirical testing of the world city phenomenon. To my knowledge, it is the first such extensive empirical intercity research in the field of world city research. From such an extensive analysis, a wide range of conclusions can be drawn. Taylor (2004, 197) highlights three main issues worth our attention. Firstly, regardless of the network topology of his world city system, there is still a core-periphery geography: "while command power remains resolutely in core-located cities, the creation of a worldwide network of cities diffuses another sort of power. This network power is found in non-core cities that have been integral and essential to the servicing of global capital" (ibid, 199).

Secondly, Taylor recognizes the impact that globalization can have on cities' *independence* from the state economy. Network structures enable world cities to function outside state borders, creating their own *hinterworlds*. The latter stands for a city's "global distribution of service connections that lies behind its world city formation" (ibid, 102). Operating on such networks and mainly interacting with other world cities, the classic approach of the national urban hierarchy's goal to "spatially integrate the national economy" appears as an oversimplification (ibid, 200).

Thirdly, Taylor underlines his findings about US cities. Throughout his analysis, US cities appear to be relatively separated from the rest of the world city network. He identifies two possible reasons for this: (1) the *shadow effect*, because in the frame of their global strategy, non-US firms only locate in New York (and possibly Chicago and Los Angeles) avoiding other cities; (2) the *comfort effect*, according to which US firms avoid global strategy and only focus on the US market (ibid, 204). In conclusion, Taylor in his extensive empirical research manages to create an evidential base for what Castells describes as the space of flows.

2.4.6 World cities, world city-regions and some scalar issues

Regardless of the above rigorous analysis of the world city network, what has not been incorporated yet in the discussion about globalisation and the city network is the notion of scale: it is well accepted that London is a global city, but how is the global city of London defined in geographical terms?

Such a debate is much wider than the current discussion about the world city networks and is inherent in the field of urban studies and planning. Different terms have been introduced among others such as the *conurbation* (Geddes 1915), *megalopolis* (Gottman 1961), the *mega-city region* (Hall and Pain 2006), the *Functional Urban Areas* (FUA – Cheshire 1990; Cheshire et al 1986; ESPON 2005a), and the American *Metropolitan Statistical Area* (MSA – Hall 2009) in order to approach different urban agglomerations and their adjacent areas. It is out of the scope of this study to further facilitate this ongoing discussion. However, there is a need to adopt a definition for the urban unit upon which this study will focus. As such the use of the *city-region* is preferred, a notion which was introduced almost 60 years ago (Dickinson 1947). Despite the wide range of different definitions for this term as they are highlighted by Rodríguez-Pose (2008), the notion of city-region has been increasingly used lately (Parr 2005).

The first reason for adopting this concept here is its strong urban and metropolitan character. Based on Rodríguez-Pose's (2008) meta-analysis of the different definitions of this concept, the main criteria for defining a city-region is the existence of a highly urbanised metropolitan area. Indeed, Ache (2000, 704-5) highlights that from the spatial perspective a city-region is very similar to a conurbation or to a metropolitan area. Others also add to this concept the core urban area's hinterland (Scott 2001) or otherwise the urban core's surrounding territory (Parr 2005). However, the main determinant for a core region is the existence of a highly urbanised area. Charles et al (1999, 1) consider the above and define the city-region as "a functionally inter-related geographical area comprising a central, or core, city with a hinterland of smaller urban centres and rural areas, which are socially and economically interdependent". This urban orientation of the notion of the city-region is strongly interrelated with the urban character of the Internet and the Internet infrastructure as analysed in section 2.3.

Additionally, the concept of the city-region is linked with the notion of a regional economy. The scale of the city-region (above the local level) and the regional character of this concept incorporate some degree of *functionality* in this term, which signifies the existence of an integrated regional economy in the city-region. This is why, according to Davoudi (2003) the city-region concept is in accordance with the metropolitan economy and Scott and Storper (2003, 581) identify city-regions as the "locomotives of the national economy". This element is also very helpful for the third research question of this study as it enables the use of the city-region concept as the study unit for the regional economic impacts of the Internet infrastructure.

Furthermore, the city-region concept appears to be related with the world city literature. Indeed, when Kunzmann (1998) refers to the world city phenomenon, he prefers the use of the *world city-region* term in order to incorporate a more functional approach to the nodes of the global urban network. Similarly, Scott (1998) introduces the notion of the *global city-region*. The above approaches though are not inconsistent with Friedmann's (1995, 21–6) thesis about the contemporary role of cities in the world economy, according to which cities are the "organizing nodes" of world capitalism and the "articulations" of regional, national and global commodity flows (Brenner 1998a). On the contrary, this functional approach of the world city-regions seems to incorporate the space of flows concept in the definition of the city. Maybe from a world scale perspective cities can be seen as the nodes of the planetary economic system, in reality though cities are not homogenous spatial entities, but rather heterogeneous and even non-continuous territorial formations with an internal network structure.

From a more theoretical perspective, Brenner (1998a, 3) adds to the above discussion that the "global city formation" is related both to the "globalization of capital" but also to the "regionalization/localization of state territorial organization". However, he recognises that large urbanised regions rather than territorial economies of scale are the basic units of global capitalism. In a similar way, Harvey (1982) links globalisation with capitalism and more specifically with capital's tendency to remove spatial barriers to its circulation and to accelerate its turnover time, which result in the formation of fixed and immobile spatial configurations (Brenner 1998a). This spatial configuration is identified by Harvey (1982) as capital's *spatial fix*.

The above are highly related with the focus of this doctoral study on the Internet infrastructure. Harvey's spatial fix includes among others the use of the investments in infrastructural capital such as transportation and telecommunications networks as a tool for territorial organisation. Such networks and particularly the telecommunications networks are multi-scalar developments in a world city framework: they both influence and are influenced by "national territorial cohesion, urban regional cohesion, and local territorial cohesion on parallel" (Rutherford 2004, 55). They are *glocal* networks as they can interlink localities at different scales: the central business district (CBD) with the new developments at the *edge* of the city and both of these local sites with another world city thousand of miles away. Brenner (1998b) identified these networks as *glocal scalar fixes*, as instead of homogenising space at national level as infrastructural networks used to do, they result in increasing capital’s uneven geographic distribution.

The above discussion can further support the following section, where the world city literature is examined from the telecommunications perspective.

2.4.7 World cities and telecommunications

All the above signify the main contributors in the world city literature and form the theoretical background for any study which concerns the global urban network. Nonetheless, apart from these approaches, more specialized studies have also taken place. Such studies focus more on specific aspects of the globalization process and the urban function. From the point of this research, it is interesting to review such specialised studies of the world city literature, which focus more on telecommunications.

One such example is Kellerman’s (1993) work about telecommunications and geography. In his book, Kellerman among other subjects, studies the *global system of cities* having as a base the *transactional city* approach (Corey 1982; Gottmann 1983). According to this, the “transactional city specialises in the generation, processing, management and transmission of information, knowledge and decisions, rather than in the production of tangible goods” (Kellerman 1993, 98). Apparently, the main means for the realization of these transactions is the telecommunications infrastructure. And here is the importance of Kellerman’s research: although all the above urban researchers included the importance of telecommunications in their analysis of the world city phenomenon, Kellerman uses telecommunications as one of his structural elements for shaping the global urban hierarchy (Kellerman 1993).

Based on the above, he produced a rather descriptive four-tier global urban hierarchy, which is presented below emphasizing the telecommunications related characteristics of each tier. The first tier refers to *domestic cities*. Among other characteristics such as the strong manufacturing and/or tourism sectors, these cities “make use of telecommunications, but in a more limited sense, as far as the international and business controlling components are concerned” (ibid, 99). The second tier consists of the *world cities*. This term highlights here that a significant part of these cities' economies is internationalized. In regards to telecommunications, world cities offer sophisticated services and act as national and regional hubs (ibid, 100). *Regional hubs* are identified as the third tier of Kellerman's global hierarchy. Only a few cities around the world can fit this category, because not many cities have the capacity to functionally serve more than one country. As far as concerns the telecommunications infrastructure, such cities provide services for not only their own country but for neighbouring countries as well. On the top of the hierarchy, the *global hubs* are found. Kellerman follows Sassen's approach for the global role of the leading league of New York, London and Tokyo and he identifies these cities as the top of the pyramid. Although his global urban hierarchy cannot be characterized as relational – at least following Taylor's thesis (2004) – but chiefly as hierarchical, in this top tier of cities the relational element is present. These three cities because of their unique roles and functions are tightly linked together and this tightness is reflected in the telecommunications infrastructure as well.

To sum up, despite the fact that Kellerman is not usually included in the rather narrow group of researchers of the world city phenomenon, his research about the global urban hierarchy has a twofold value added for this doctoral thesis. Firstly, the role of telecommunications in defining the global urban hierarchy appears to be more prominent in comparison to other global urban studies. Kellerman successfully highlights the fact that his global urban hierarchy is not explicitly based on the urban distribution of the telecommunications systems, but interestingly enough this infrastructural layer is taken more into consideration here than in other studies (Kellerman 1993, 99). Secondly, Kellerman extends his global urban hierarchy to cities of lower importance– the so called domestic ones. This agrees with Castells' approach about the extent of the global city process and fits with the needs of this doctoral study: because of the rather smaller scale of this research (mainly European instead of global) more cities of lower ranking can be included in the analysis of the world city phenomenon.

2.4.8 The link with and the contribution to the world cities field

This section explains how the world city phenomenon is related with this doctoral thesis and the contribution of the present study to this research field. First of all, it should be noted that this research does not attempt to present a new world urban hierarchy based exclusively on the urban allocation of the Internet infrastructure. The complexity of the cities themselves and the complexity of the relational character of what was identified here as world city prevent such a unilateral approach. However, all the above theorists of the urban phenomenon of the global cities agree on one point:

> "Transport and communication have played a critical role in shaping the evolving world city system. In turn, world cities have been instrumental in shaping global, regional, and local transport and communication networks" (Keeling 1995, 128).

Indeed, both telecommunications and transportation have facilitated the world city phenomenon by decreasing and even extinguishing communications costs, enhancing the interaction of the globally spread actors of capital accumulation and supporting their integration. Both the Internet and more specifically the Internet backbone network as well as the aviation network carry a significant part of this interaction (Taylor 2004) facilitating not only the world city phenomenon but also globalisation itself (Graham and Marvin 2001, 8). Information is distributed around the world settlements through what is known as the "*information highways*" (Gore 1993). In the same way, people are being brought together via the aviation network in order to interact and acquire complex knowledge (Rimmer 1998). These processes enable Smith and Timberlake (2002, 139) to recognize world cities as the "spatial articulations of the global flows that constitute the world economy" and Rimmer (1998, 439) to identify them "as junctions in flows of goods, information and people rather than as fixed locations for the production of goods and services". However, these flows are not transported in the abstract space, but rather on this specific infrastructural layer, identified by Castells as the first layer of the space of flows, which is (unequally) spread around the cities in a network topology. This infrastructural layer is the necessary means for the circulation of flows and further for the emergence of the world cities phenomenon.

On the other hand, the infrastructural layer is also structurally affected by the shape of the world city network. Because of the private character of the telecommunications and aviation industries, the spatial distribution of their infrastructural networks is mostly shaped by the spatially differentiated demand for such services. Taking into consideration that the demand for communications –

electronic or air transportation – is maximized among the world cities and their hinterworlds, it is economically rationale for the carriers to focus and heavily invest in locating their networks among such locations. This demand, which is reflected in the fibre-filled corridors, enables Graham and Marvin (1996, 3) to announce cities as the "power houses of communications".

To sum up, there is a twofold contribution of this doctoral thesis in the field of world cities research. Firstly, despite the fact that this research will not propose a new global urban hierarchy, it will shed light on this "symbiotic" relation between the communications infrastructure and world city formation (Keeling 1995, 129). In more detail, this study's main contribution in the field of world city research is the geographical analysis of the main facilitator of this global urban phenomenon, the Internet infrastructure. In addition, a topological and geographical comparison with the other facilitator of the global city process, the aviation network, will also take place in order to investigate the degree of synergy between the two infrastructural networks in facilitating the global city process (Choi et al 2006).

Additionally, this doctoral thesis also attempts to bridge the gap between the theoretical sophistication in the work of Sassen (2000), Friedmann (1986 and 1995) and Castells (1996) and the lack of empirical evidence to back up their claims concerning an emerging networks of flows. The above is illustrated by Peter Taylor (1999, 1904) as an "evidential crisis" in the burgeoning field of world cities research. In particular, Taylor highlights the surprisingly limited use of relational data in the key studies in the field, given that it is precisely relations between cities that constitute the key to understanding the new world city networks that analysts contend are emerging. Although this study is not based on relational data for the actual flows between the cities, it exploits relational data for the supporting infrastructural layer of this interaction.

2.5 The Internet infrastructure and regional development

2.5.1 Introduction

The last section of the literature review explores the link between the Internet infrastructure and regional economic development, the third pillar of this doctoral thesis. Firstly the general economic framework, under which the Internet appears to be a valuable means for economic growth, is analysed. Secondly, the role of the Internet in the production process is highlighted. Thirdly, based on the above as well as on relevant theoretical approaches for regional economic development, a conceptual research framework for the economic development impacts of the Internet infrastructure at regional level is suggested. Lastly, empirical studies concerned with the direction of causality between ICTs and (regional) economic development are reviewed.

2.5.2 General economic framework

The massive technological improvements which took place in the post-industrial era, apart from having wider social impacts, resulted in structural changes in the economy. *Soft factors* such as information, knowledge and technology became fundamental factors in the production process and in the related policy agenda. Nevertheless, there is not a single conceptual framework which encompasses these changes in the post-modern economy (Cohen et al 2000). Among others, the most widely used concepts describing this post-industrial economy are the *information economy*, the *knowledge economy* and the *digital economy*[5]. The rest of this section is spent in critically analysing these approaches.

The concept of the information economy is the oldest one of the above. In 1977, Porat (204 cited in Hepworth 1989, 7) recognised that:

> "[We] are entering another phase of economic history. We are just on the edge of becoming an information economy. The information technologies – computers and telecommunications – are the engines of this transformation. And we are now seeing the growth of new information industries, products, services and occupations, which presage new work styles and lifestyles based on intensive use of information processing and communication technologies."

Hepworth (1989, 7), building on Porat's definition, further explained the information economy as a "new phase of economic development", the main characteristic of which is the dominance of information in goods and services production and in growth in general. The further enlargement of the information economy will result in

---

[5] Other terms used to approach this phenomenon and not analysed here are: the weightless economy (Quah 1996), virtual economy, e-economy etc.

wider transformation of economic products, activities and actors, which will affect not only the information intensive sectors, but will also lead to the emergence of a wider techno-economic change (Miles and Mathews 1992). In addition, Castells (1996) identified the new economy as informational rather than information based, as analysed in the previous section. From an economic stand point, although information is the main input for productivity growth and competitiveness, the effect of the information economy is wider than this production-input change. This is why the informational economy is not just the product of the technological improvement in the fields of computing and telecommunications, but instead is the result of the change of the technological – or to be more accurate the change of the techno-economic – paradigm, a term which is traced back to Perez (1983). The latter refers to the “combination of interrelated product and process, technical, organisational and managerial innovations, embodying a quantum jump in potential productivity for all or most of the economy and opening up an unusually wide range of investment and profit opportunities” (Freeman and Perez 1988, 47-8).

From a historical perspective, this new techno-economic paradigm, which is materialized by the information economy, can be also approached using Nikolai Kondratieff’s *economic waves* (1926)[6]. He identified long phases of development – almost half a century long – based on the shift of technological paradigms. Starting from the industrial revolution and *early mechanisation* Kondratieff, he continued with the *steam power and railway* Kondratieff and the *electrical and heavy engineering* Kondratieff. After his death, the *Fordist mass production* Kondratieff was introduced and according to Freeman (1987) we currently experience the fifth Kondratieff of *information and communication*.

Kellerman (2002), in his attempt to further explain the information economy, identified its structural elements. According to him the information economy consists of (ibid, 16):

- Infrastructure: the technical layer of the information economy, where the ICTs are located.
- Information: all kinds of information – personal, business, educational etc. – the delivery of which to customers (users) is based on the infrastructural layer.

---

[6] It should be noted that the long cycle theory is not originated to him but rather to other earlier economists such as the Dutch Marxist Van Gelderen, who introduced this idea in 1913, thirteen years before Kondratieff (Freeman and Soete 1997, Mumtaz 2003).

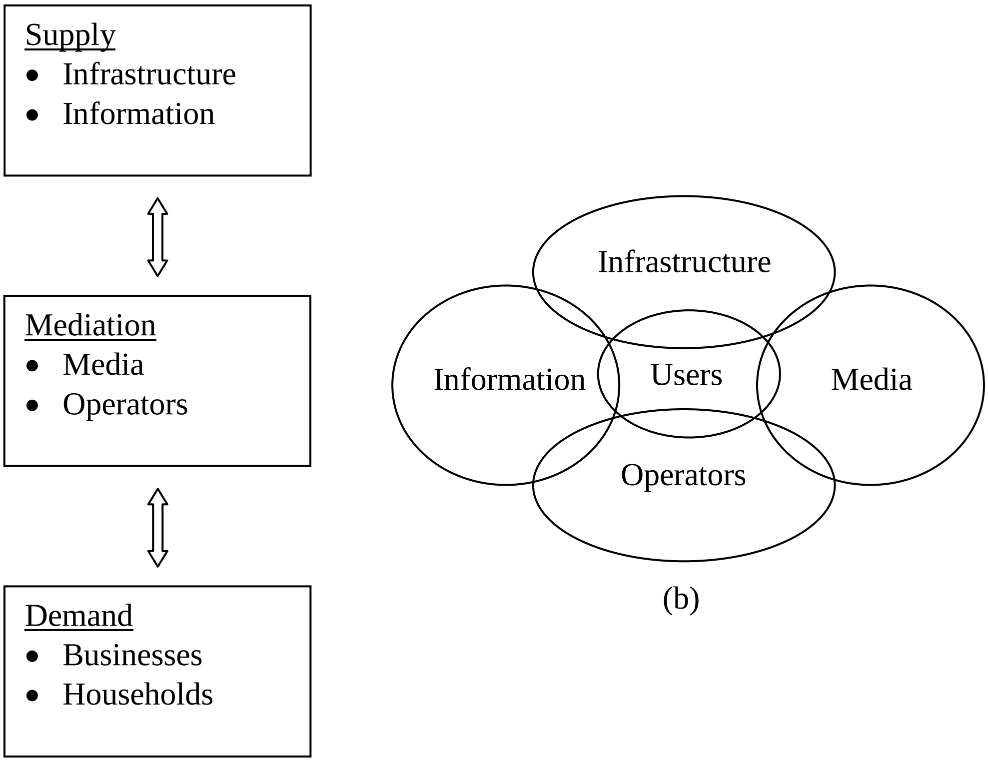
- Media: they are responsible for the consumption of the various types of information (TV, radio, the Internet etc.)
- Operators: it refers to companies which deal with the operation of business for all the above layers (i.e. production and servicing of the infrastructure, information and the media).
- Users: otherwise the customers who consume the infrastructure, the information and the media. Users can be both households and businesses.

The above five elements of the information economy aggregate together and form the three economic functions as are illustrated in Figure 2.2 (a). All the five elements of the information economy are merged together in order to facilitate the needs of the final user-customer as illustrated in Figure 2.2 (b).

The second concept analysed here is the *knowledge economy*. Knowledge is directly linked to information as "knowledge is more than information as information is more than simply data" (Mallecki and Moriset 2008, 29). The relation between these notions is hierarchical as one step higher in the hierarchy equals to higher level of sophistication, codification and consequently value. The relation between data, information and knowledge is very successfully illustrated in Figure 2.3. Nijkamp and Johnkhoff (2001, 2) identify knowledge as the "accumulated stock of information based on synergies" contrary to these "structured flows of data", which form information. The adding, restructuring, editing and other operative changes of information result to the formation of knowledge (ibid). Leydesdorff (2006, 17; original emphasis) further explains the notion of knowledge and distinguishes it from information:

> "Knowledge enables us to codify the *meaning* of information. Information can be more or less meaningful given a perspective. However, meaning is provided from a system's perspective and with hindsight. Providing meaning to an uncertainty […] can be considered as a first codification. Knowledge enables us to discard some meanings and retain some others in a second layer of codifications. In other words, knowledge can be considered *as a meaning which makes a difference*. Knowledge itself can also be codified, and codified knowledge can, for example, be commercialised".

Figure 2.2: Major elements of the information economy



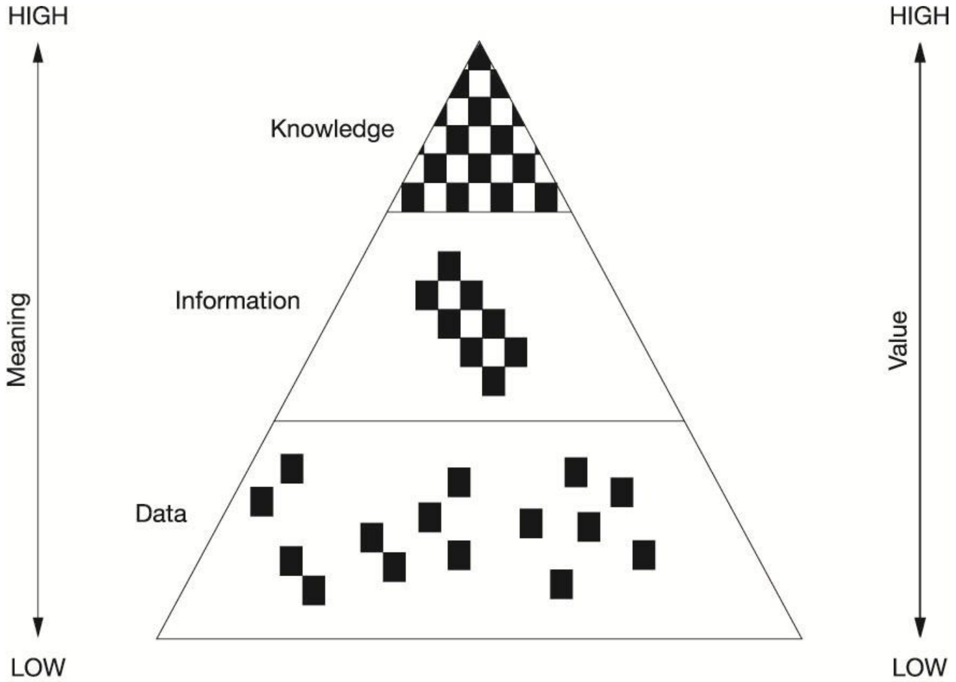
(b)

(a)

Source: Kellerman 2002, 17

This last point is the key characteristic of the knowledge economy: knowledge, as a commercialised entity, has become one of the factors of production, in advance of capital and labour (Drucker 1998). According to the OECD’s (1996, 7) definition, knowledge based economies are economies “which are directly based on the production, distribution and use of knowledge and information”. However, the economy was always dependent on knowledge, even in ancient times (Quah 1998). Despite the fact that steam engines and clay tablets are physical products, they embody knowledge, which was used for their production; however they are not knowledge themselves (Maignan et al 2003). This is the key characteristic of the knowledge economy: knowledge becomes a stand-alone product (Quah 1998). Additionally, investments in knowledge and knowledge products can have horizontal effects, affecting all the factors of production and even transforming them to new products and services. “And since these knowledge investments are characterised by increasing […] returns, they are the key to long-term economic growth (Stevens 1998).

Figure 2.3: Data, information and knowledge



Source: Burton-Jones (1996, 6) cited in Pike et al (2006)

But how is the knowledge economy linked to the information economy? Maignan et al (2003, 4) differentiate the above notions as follows:

> “Knowledge represents the capacities or capabilities of an individual or a social group [...] associated with meaning and understanding, as well as the abilities to organise, interpret and assess information” (Cohendet and Stainmueller 2000), while information is “knowledge reduced to messages that can be transmitted to decision agent” (Dasgupta and David 1994).

The notion of knowledge is tied with the notion of learning. The latter, as a collective ability of a society or a locale, appears to be central in the development process (Pike et al 2006, Lundvall 1992). Additionally, the advances in ICTs – identified as the infrastructural layer in Kellerman’s model of the information economy – managed to speed up the codification of this part of knowledge which is able to be codified[7]. This rapid change in the knowledge codification process resulted in the transformation of knowledge into a market commodity: “large chunks of knowledge can be codified and transmitted over computer and communication networks” (Stevens 1998, 90).

---

[7] The part of knowledge which is not codifiable is identified as tacit knowledge and is embodied in practices, people, and networks (Maignan et al 2003).

Such technological improvements are integral elements of the knowledge economy as technology can be approached as applied, routinised and transferable knowledge (Landes 1998).

The third and most recent theoretical attempt to describe the current economy is summarised under the term *digital economy*. The digital economy is linked in most people's minds with the economic transactions taking place in the Internet (Atkinson and Mckay 2007). However this is only part of what the digital economy really is. Atkinson and Mckay (2007, 7) define it as follows:

> "The digital economy represents the pervasive use of IT (hardware, software, applications and telecommunications) in all aspects of the economy, including internal operations of organizations (business, government and non-profit); transactions between organizations; and transactions between individuals, acting both as consumers and citizens, and organizations. Just as 100 years ago the development of cheap, hardened steel enabled a host of tools to be made that drove economic growth, today information technology enables the creation of a host of tools to create, manipulate, organize, transmit, store and act on information in digital form in new ways and through new organizational forms (Cohen et al 2001).

The key point of this theoretical concept is the pervasive character of ICTs in all sectors of the economy. The information economy concept is more or less linked with specific sectors of the economy. Porrat 1977 (cited in Hepworth 1989, 15) identified the *informational worker* and he developed a register with 422 information occupations based on the US Census of Population workforce classification. Additionally, concepts such as *quaternary employment*, which refers to services "closely related to the production, processing and distribution of information (Gottman 1983, 66) and the *informational sector* were introduced to frame the information economy (Hepworth 1989). Moreover, the concept of the knowledge economy appears to be more widely defined: no explicit knowledge sector was and the definition of knowledge-based occupations was also extended out of the service sector (Neef 1998).

However, such a sector-based approach does not apply to the digital economy. Indeed, the concept of the digital economy is by definition horizontal and refers to the impacts that the whole economy can enjoy – mostly through productivity gains –

because of the extensive use of ICTs in all aspects of the economy (Atkinson and McKay 2007). Simply put, computers, telecommunications and their combination function because of the impact of one to the other known as *infocommunications*, serving downstream industries in all sectors of the economy (Malecki and Moriset 2008, 4, 39). This process results in productivity effects, which according to Atkinson and McKay (2007, 20) can be distinguished in *capital deepening* and *total factor productivity* gains. While the former refers to the fact that more capital results in more productive labour, the latter refers to productivity increases when the same amount of capital is used more efficiently. In addition, OECD (2003) suggests a third path for the expansion of the productivity gains: the productivity acceleration in the ICTs-producing sector and the expansion of the ICTs-producing sector in the economy. All in all, such productivity gains can affect economic growth and result in increasing the level of economic development.

However, much discussion has taken place in the relevant literature with regards to the productivity gains because of the use of ICTs. This debate is known as the *Solow productivity paradox* because of his 1987 quote: “You can see the computer age everywhere but in the productivity statistics” (cited in Triplett 1998, 1). Regardless of the commonly accepted view that the use of ICTs will result in productivity gains, productivity indicators remained stable until the mid 1990s. Cairncross (2001, 193) provides three explanations for this paradox. First, there was a waste of resources, especially in the first stages of the diffusion of computer usage. Many office-workers did not have the knowledge to use this new device in a productive and efficient way. Second, economies usually do not immediately take advantage of new technologies. A couple of generations are needed for the economies in order to learn how to use proficiently new technologies. And third, productivity itself is difficult to measure, especially in the service industry, where the impacts of ICTs are likely to be greatest.

It was not until the end of 1990s when the first evidence of productivity growth appeared. Two thirds of the productivity growth in USA between the first and the second half of the 90’s were due to the investment in or the production of computers (Cairncross 2001, 195). In Europe, where the ICTs sector is smaller, more than half of productivity growth emerged because of ICTs. Productivity growth between 1995 and 2000 was around of 1.4% and over 0.7% was owing to ICTs. Those gains of productivity were widely explained because of the production of high-value goods based on ICTs and because of the adoption of ICTs in the production procedure (EC 2004).

To sum up, a common characteristic of all the above three concepts is that they cannot be limited to the Internet-based new economy (Malecki and Morisset 2008). As explained above, the changes the post-modern economic system is going through are wider than this. Additionally, despite the different starting point of the three concepts described above, there are overlaps between them since they describe the same phenomenon from different perspectives: the new techno-economic paradigm of the post-industrial economy. While the first two approaches mostly focus on the soft factors of this paradigm (i.e. information, knowledge and the learning process), the digital economy framework mostly emphasizes the hard factors (i.e. ICTs). However, all the three theoretical concepts agree on the central role of the ICTs in this new paradigm. This led Antonelli (2003, 197) to characterise advanced telecommunications services as the backbone of the new economy; this is the focus of the next section.

2.5.3 The infrastructural character of the Internet

This section attempts to illustrate the economic function of ICTs in general and the Internet more specifically. Effort is spent to identify and analyze these economic characteristics of the Internet which enable it to affect the production process.

The starting point for such an attempt is the notion of *general purpose technologies* (GPT). The latter is part of the wider notion of *drastic innovations*, which refer to innovations that create discontinuity because they result in radical change of the used technological means and even to the replacement of old technologies (Helpman 1998). GPT are part of this wider array of drastic innovations. The term is traced back to Bresnahan and Trajtenberg (1995). This first approach identifies GPT as those technologies which have the:

> “potential for pervasive use in a wide range of sectors […] As a GPT evolves and advances it spreads throughout the economy, bringing about and fostering generalized productivity gains. Most GPT play the role of ‘enabling technologies’, opening up new opportunities rather than offering complete, final solutions. For example […] the users of micro-

> electronics […] benefit from the surging power of silicon by wrapping around the integrated circuits their own technical advances" (ibid, 84)."

Lipsey et al (2005, 94-99) further analyzed this rather new concept and they identified the following key characteristics:

- Firstly, GPT are generic products, processes or organizational forms, that regardless of their evolution over time, are widely recognised as such. For instance, major technological advances have been applied to the first PC, but it is still recognised as a PC.
- Secondly, GPT can be both exogenous and endogenous production factors. The electronic computer for example was developed in universities and private firms labs with military funding in order to meet the World War II needs for a machine able to break the enemy's codes and conduct complicated calculations of ballistics. So, the early electronic computer was exogenous to the economic system because it lacked economic applications but endogenous to the military one.
- Additionally, GPT are not usually technology-radical but use-radical. This means that the expansion of a GPT is gradual over time, but once it reaches the GPT threshold it expands radically and generates impacts on its users.
- The next characteristic is the scope for improvement. Any technology in order to become a GPT by definition needs to go through an evolution process.
- GPT are characterised by the variety of different applications that this technology can have. This is different than just being widely used because the latter does not include the variety in usages. For instance, the electric bulb is widely used across the economy, but its only use is to produce light.
- GPT are valuable from the economy point of view because they create spillovers. The complex technological interrelations of a developed economy spread the effect of a GPT beyond the initial users.
- Lastly, many non-GPT might have some of the above characteristics and even to a greater degree than a GPT, without being a GPT.

Based on the above detailed approach of Lipsey et al (2005) and in accordance with Harris (1998), Malecki (2002a) and Atkinson and McKay (2007), there is little doubt that the Internet is a GPT: it is a generic technology, which was gradually developed, but once it reached a specific threshold – privatization in this case – was radically expanded across the economy with a huge variety of different applications, creating spillovers which enable the emergence of the digital economy. From the economy point of view, such spillovers represent productivity increases in downstream sectors (Helpman 1998; Malecki 2002a) which result in economic growth and development. In simple words, even the least sophisticated aspects of the Internet are essential to the production process. For example, albeit that e-mail technology is more than 20 years old[8] it is still the most broadly used information technology and its significance to the production procedure is doubtless (Batty 1997).

However, in order for a GPT to start having such impacts on the economy, investment in infrastructure is needed. To put this simply, electricity required huge investments in production and distribution systems in order for the society to benefit from this GPT (Lipsey et al 2005). In the same way, in order for consumers (users) to take advantage of the Internet a whole new infrastructure was developed, the Internet infrastructure.

Before analyzing further the Internet infrastructure, infrastructure as a generic notion is discussed. Infrastructure or otherwise infrastructural capital is identified by neo-classical economics as part of the overall capital stock and is characterised by a blend of publicness and capitalness (Biehl 1991). While the former is linked with nonrivalness and nonexcludability – the two goods' properties that cause market failure (Musgrave and Musgrave 1984; Biehl 1991), the latter highlights the significance of infrastructure as a factor of production, admeasured in the capital[9]. Traditionally, infrastructure is mostly related with transport infrastructure. Banister and Berechman (2003, 35) defined infrastructure as the "durable capital of the city, region and the country and its location is fixed". Jochimsen (1966 cited in Biehl 1991) suggested a broader definition and also includes all types of public services and institutional infrastructure. Hirschman (1958, 83) introduced the notion of *social overhead capital*, which is defined as

> "comprising those basic services without which primary, secondary, and tertiary productive activities cannot function. In its wider sense, it includes all public services from law and order through education and

---

[8] It was developed by Ray Tomlinson in July 1970 (Castells 2001).

[9] Apart from capital, classical economics identify land (or natural resources) and labour as the three basic resource categories. The appropriate combination of them leads to the production output and income (Biehl 1991).

public health to transportation, communications, and power and water drainage systems. The hard core of the concept can probably be related to transportation and power".

Kay (1993, 55) attached five characteristics to infrastructure:

- It is usually developed in network structure and can be approached as a delivery system. In order to achieve this there is a need for considerable interactions among the different infrastructural networks.
- Infrastructure results in the reduction of the production cost for a wide range of products, the production and distribution of which takes advantage of the infrastructural networks. The losses because of infrastructure failure are much higher than the gains in production cost reduction.
- It is very common for infrastructural networks to have characteristics of natural monopolies as infrastructure provision under market economy rules and open competition is costly.
- The necessary capital for infrastructure development is larger than the running cost.
- Infrastructure provision is linked with high sunk cost as most of the cost has been occurred before the provision of any kind of services based on the infrastructure.

Reviewing carefully the above approaches, it seems that the Internet infrastructure can fit with all the above points but one, the *publicness*. Although this characteristic is not as strong as it used to be after the extensive implementation of public-private joint projects for infrastructure provision, infrastructure still has elements of natural monopoly. For example, transportation networks are developed today as joint projects privately (co-)funded, but there are still monopolistic elements as these networks will be the only ones providing this service in this specific area. However, this is not the case with the Internet. Regardless of the different regulatory frameworks that affect the Internet function (IP addressing, access to the local loop etc.) most of its hardware is privately developed. Users (consumers) have the ability to choose between different hardware owned by different carriers showing that the market does not fail in Internet infrastructure provision.

But what exactly is the Internet infrastructure? In section 2.3.1 two approaches were described: while Gorman and Malecki (2002a), Stephenson (1996) and Gilder (2000) strictly link the Internet infrastructure with the Internet's physical layer, Greenstein (2004) also includes in his definition soft elements such as software and human capital. In order to better define the Internet infrastructure, we will use again the OSI, presented previously in section 2.2.5. As mentioned there while the lower part of the OSI model is related with the physical infrastructure or hardware, the upper part can be identified as the soft infrastructure. Indeed, as mentioned in section 2.2.5, the routing of the IP data packets takes place in the third layer of the OSI model. Additionally, this is the layer where the complete origin-destination paths are formed – contrary to the one-hop links which take place on the second layer. Switches, backbone networks and routers are part of the physical layer of the OSI (layer 1), but function at level 2 (switches) and 3 (backbone and routers) (Gilder 2000). So, it can be said that the physical infrastructure of the Internet infrastructure consists of the physical layer of the OSI model plus the two higher layers, which shape the network structure of the Internet.

The remaining four layers of the OSI model consist only of soft elements and they form the *infratechnologies* of the Internet infrastructure. This term refers to these technology elements usually identified as industrial standards which include research tools (measurement and test methods), scientific and engineering data, and the technical basis for both physical and functional interface standards such as factory automation and communications (Tassey 1992 and 2008). Interestingly enough, these technologies have a quasi-public character, but they are not provided by the state and are usually used freely in order to support the industry. Yet, the four higher layers of the OSI model support the Internet function not by supplying the physical links, but by standardising the Internet function.

The above physical and non-physical infrastructural elements, which can be alternatively identified as the hardware and the software of the Internet, constitute the Internet infrastructure (Table 2.3). From the geography point of view, what is interesting is the study of the physical layer of the Internet infrastructure, because its distribution can be spatially differentiated contrary to the infratechnologies and similarly to more conventional forms of infrastructure such as transport infrastructure. From the regional development point of view, it is interesting to examine if this spatially differentiated distribution of the infrastructure of the digital economy can result in localized developmental impacts. The next sections of the literature review provide the theoretical background for such a research investigation.

Table 2.3: The infrastructural approach of the OSI model

OSI	
7. Application	Soft infrastructure / infratechnologies
6. Presentation	
5. Session	
4. Transport	
3. Network	Physical infrastructure
2. Data link	
1. Physical	

2.5.4 Regional development and the Internet infrastructure

This section does not intend to provide an extensive review of the numerous regional development theories, but rather to highlight the way some of the main regional economic development theoretical approaches feature technology and infrastructure. This review will be the supporting theoretical fabric for the empirical investigation for the impacts of the Internet infrastructure on regional economic development, which is presented in Chapter 7.

The starting point for such a review is the neo-classical model growth introduced by Solow (1956). The basic characteristic of this model is the free-market approach, according to which the convergence of regional disparities will happen despite any policy interventions; the latter will only accelerate or decelerate the convergence process (Pike et al 2006). The main methodological tool is the aggregated production function, which identifies three sources of output growth: the capital stock, the labour force and technology. The underlying assumptions of unobstructed inter-regional factor mobility and perfect knowledge about factor prices in all regions explain long term convergence: capital and labour will migrate towards those regions where their marginal returns are higher (Armstrong and Taylor 2000). For example, a region with high capital stock, because of the assumption of constant returns to scale, will experience an outflow of capital towards regions with lower capital stock, where the capital will be facilitated by higher marginal output. The same adjustment mechanism applies also to the labour force.

In regards to technology, as well as other factors such as human capital and population growth, they appear to be disembodied with capital and labour (Pike et al 2006). This is the reason why the neo-classical model is also known as the model of exogenous growth (Aghio and Howitt 1998). Regardless of the fact that knowledge in the form of technology was included in the model as one of the three inputs for production, this factor was approached as an exogenous and residual:

> "Solow discovered that the contributions (the inputs) from the production factors labour and capital to the production process could explain less than 50 per cent of economic growth. The rest *had to be* explained by technology" (Lambooy 2002, 1022; emphasis added).

Apart from the above critique about the exogenous character of the technology, the neo-classic model was also criticised for the unrealistic assumption about constant returns to scale. Mydral (1957) introduced the notion of cumulative causation according to which growth is a circular and cumulative process. Increasing returns to scale and agglomeration are strong explanatory factors for the development level. Infrastructure provision is included in the cumulative process since it tends to agglomerate in the already developed regions, enhancing in this way regional inequalities and polarization in space. Later, other researchers also built upon Mydral's work such as Kaldor (1970) and Dixon and Thrilwall (1975). The latter introduced the homonymic model according to which regional output growth is affected by two factors: growth in the capital/labour ratio and the rate of technological change (Pike et al 2006).

In regards to the impact of technology on the growth process, the most important progress was the e*ndogenous growth theory*, firstly introduced by Romer (1986 and 1990). Recognizing the value of technological progress in the growth process, the main idea of this approach is the endogenous character of technological change, which is treated as a positive externality (Button 2000). Such models also accept the assumptions for increasing returns to scale and cumulative causation, but they follow a more sophisticated way than Mydral and Kaldor, by accepting the formalities of the neo-classical model such as the production function and the general equilibrium framework. According to this theory, technological knowledge will increase over time with a rate of change shaped by: (a) the volume of the workforce in the knowledge-producing sectors; and (b) by the existing stock of knowledge (Armstrong and Taylor 2000).

Finally, one of the latest advances in the field of economic geography, the *new economic geography* (NEG) introduced by Krugman (1991a and 1991b) is also linked with the infrastructural layer of the digital economy. NEG, or according to others *new geographical economics* (Martin 1999) "might best be described as a 'genre': a style of economic analysis which tries to explain the spatial structure of the economy using certain technical tricks to produce models in which there are increasing returns and markets are characterized by imperfect competition" (Krugman 1998, 10). The focal question of NEG is to explain the formation of agglomerations in space (Fujita and Krugman 2004). The central argument behind the NEG is that its main features, which are increasing returns to scale and imperfect competition, are more important factors for trade and spatial specialization than perfect competition and comparative advantage, which are basic elements of neo-classic economics. Most importantly, NEG, following the Marshallian external economies, recognizes spatial structure as a result of the simultaneous act of centripetal and centrifugal forces (Krugman 1998). But what is the role of ICTs in the balance between these opposing forces? According to Maignan et al (2003) the establishment of the digital economy results in dramatic transport and communications costs reduction, which can lead in a change of the current equilibrium of centrifugal and centripetal forces. In other words, the digital economy can affect the existing spatial structure and economic landscape.

In more details, it is proposed that ICTs can affect the existing spatial structure by reducing the (Venables 2001; Maignan et al 2003):

- search and matching cost for trading partners
- shipping cost of *weightless* products, which can be codified and digitized
- control and management costs
- cost of time in transit because of shipping and communications with distant locations
- cost of personal interaction[10] and promoting knowledge spillovers
- cost of commuting and moving within the agglomeration (teleworking, teleshopping)

---

[10] Face-to-face communication can be divided to two components: the conversation and the handshake, with the former being the "metaphor for simultaneous real-time interactive visual and oral messages" while the latter for the physical co-presence (Leamer and Storper 2001, 4). ICTs can only lower the cost of the conversation component of the face-to-face communication.

- cost of replicating products
- cost of relocation

However, as mentioned in section 2.4.2, all the above cost reductions will not lead to a new spatial equilibrium, in which geography and distance, agglomeration and increasing returns to scale and the distinction between core and peripheral regions do not matter. The above described transaction cost reductions can lead to further strapping in the well known dominant agglomerations economic of activities which can be characterized as complex, knowledge intensive and in need for face-to-face communications. On the other hand, more transportable and less dependent on (the handshake component of the) face-to-face communications might migrate and create new clusters specialized in such back-office activities (Venables 2001).

All in all, all the above theoretical approaches agree on the importance of infrastructure and technology on regional economic development. What they do not explain though is how exactly the infrastructure of the digital economy can affect the development level at regional level. The next section deals with this matter.

2.5.5 Conceptual framework for the research on the regional development impacts of the Internet infrastructure

Based on the above analysis the following points can be drawn. Firstly, in the frame of the digital economy and at the national scale, the Internet affects the economy through productivity increase, because of its attribute as GPT and through the required Internet infrastructure. Secondly, at a parallel regional level, all the reviewed theories – regardless of their different approaches – seem to agree on the positive impact that infrastructure and technology can have on the development level. And mostly drawing upon the core of the endogenous growth theory, it seems that technology is a growth driver even at the local level.

The emerging question is whether the physical layer of the infrastructure of the digital economy, the allocation of which is spatially differentiated, can affect economic development at the regional level. It would be easier to comprehend such a question if the discussion was about transportation networks (see for instance Biehl 1991) instead of Internet infrastructure, because of the more tangible nature of transportation than Internet infrastructure. Indeed, while transportation infrastructure reduces transaction costs on trade in goods, telecommunications infrastructure lowers transaction costs of trading ideas (Cieslin and Kaniewska 2004). And from the technical point of view, despite what the average Internet user thinks, the Internet is not a unique system evenly scattered across the globe, regardless of core or periphery (Gorman and Malecki 2000). And despite being a fairly young Large Technical System (LTS), at least for commercial usage, users consider it as a black box, something which is usually related with other older urban infrastructure networks such as water, sewerage etc (Graham and Marvin 2001). In reality, geographic location affects Internet connectivity and the speed at which data can be transmitted and received. The latter is the result of the uneven spatial allocation of the Internet's physical infrastructure (routers, switches, IXP, POP, cables, fibre optic links, etc.) across space (Malecki and Moriset 2008). However, the above differences in the quality of Internet connectivity are mostly visible not to end-users, but to large corporations. For instance, the DSL population coverage reached almost 90% in Europe in 2006 – a type of Internet connection which is considered as broadband nowadays. Such a connection will hardly dissatisfy any average user. However, a Trans-National Corporation (TNC) in order to locate a branch in a city will need a different type of physical infrastructure (you cannot squeeze 500 employees in a domestic usage DSL connection!). In order to accommodate such users, a city needs to be served by the highest rank of the Internet's physical infrastructure: not only backbone networks need to have a node in the city-region, but additionally the city-region needs to benefit by direct end-to-end links with the main world cities where the TNCs are agglomerated, providing secure, fast and low latency connections. Moriset (2003) using Lyon as case study and after applying a survey of 92 multimedia firms, among other factors identified the value of the installed Internet infrastructure for the firms' location decisions. Cushman & Wakefield (2008) identified a city's quality of telecommunications as the fourth most essential factor for locating a business in Europe, one place higher than the transport links. Graham (2004, 140) highlights the above locational logic by saying that the focus of real estate has changed from "location, location, location" to "location, bandwidth, location" (Malecki and Moriset 2008). And even at a lower spatial level, office buildings need to combine the physical qualities of high ceiling height, high power and back-up electricity supplies with nodal positions on fibre networks (Graham 2004). In general, it seems that ICTs not only stimulate the development of urban networks and network cities, but they also strengthen the urban economy (Lambooy et al 2000; Louter 2001; and van Oort et al 2003 cited in Priemus 2007).

Going back to the above question, the answer is yes: there is a rationale in investigating the localized economic impacts of the physical infrastructure of the Internet, because the concentration of this infrastructure in specific locations may affect the economic development of these areas as it will provide better access to the backbone of the digital economy. The concentration of infrastructure such as Internet backbone networks and IXPs in a city-region can affect the competitiveness at micro and territorial level: through efficiency and effectiveness effects, Internet infrastructure can result in cost reduction and revenue increase for corporations; and through connectivity effects and the endowment of location factors it can impact the accessibility and the attractiveness of territories (Camagni and Capello 2005). Put simply, Internet infrastructure can both result in attracting new firms (Cornford and Gillespie 1993) in a city-region which can exploit such infrastructure (financial firms, back-office activities, creative industries) and increase the productivity of the existing firms. Additionally, such infrastructure might also result in higher quality digital services for end users.

On the other hand though, it needs to be highlighted here that the reverse relation might also exist on space: instead of the Internet infrastructure affecting regional economic development, the regional economic development level might also be a pull factor for the distribution of Internet infrastructure. This *causality problem* is common in regional science and especially in the discussion about the relation between infrastructure and regional economic development. For instance, Banister and Berechman (2003) in their research about transport infrastructure noted that the empirical evidences are mixed. While some researchers conclude that increases in productivity may result in increases of infrastructural capital, others argue the opposite direction in this causal relation. Interestingly enough both causal relations might also exist at the same time and for different places. The next section presents evidence from the literature on this issue.

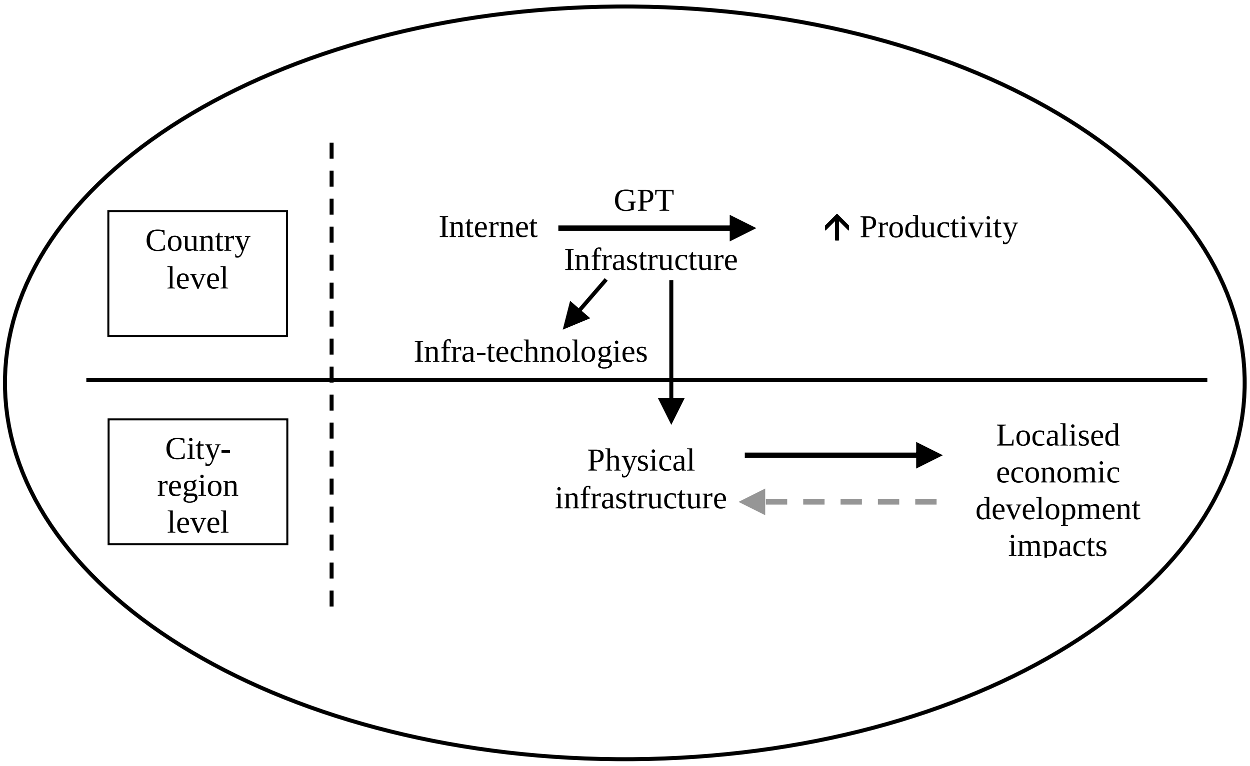
Parenthetically, it should be mentioned here that usually infrastructure deployment is related with short-term increase in employment due to the necessary large scale civil engineering works (*short-term construction employment* – Banister and Berechman 2003). This is also the case for the Internet networks. Indicatively, Liebenau et al (2009) suggest that 50% of broadband networks deployment cost is due to the labour cost. However, this is not the case of the international Internet backbone links because such networks are not intra-urban, where the main excavation and civil engineering costs are located. Backbone links are developed across existing infrastructural corridors (i.e. motorways and railways) or as submarine cables and consequently the excavation costs as well as the short term employment impacts are minor and not included in the analysis here.

Schematically, the above discussion is illustrated in Figure 2.4, which forms the conceptual research framework for the regional economic development impacts of the Internet infrastructure. The empirical research based on this framework takes place in Chapter 7.

2.5.6 A review of empirical studies on the causal relationship between ICTs and economic development.

Indeed, the direction of causality between infrastructure provision and economic development was always a complicated problem in regional science. Despite the rather limited interest of geographers, urban planners and regional

Figure 2.4: Conceptual research framework for the regional economic development impacts of Internet infrastructure



Digital Economy

scientists on telecommunications for reasons highlighted in section 2.3.2, there is disproportional interest in studying the direction of causality between technology, telecommunications and ICTs. The main motivation for such research is the policy implications of a causal relationship. In more detail, if causality runs from ICTs to economic development, then investments in ICTs can be used as a policy tool for economic activity stimulation. On the contrary, a unidirectional causal relation with causality running from economic development to ICTs, will vitiate the efficiency of such policies. On the other hand, a bi-directional causal relationship will result in a cyclical phenomenon: policies for ICTs stimulation will also result in economic development, which in turn will result in further stimulation of ICTs demand and supply. However, the lack of a significant causal relationship between ICTs and economic development prevents policy makers for using one of them as a stimulating tool for the increase of the other (Wolde-Rufael 2007).

Early work by Hardy (1980) tried to identify the role of the telephone on economic development using panel data[11] for 60 countries. He concluded that the telephone contributes to economic development, but he did not analyse further the causal relation between these two elements. The main barrier in analysing the direction of causality between ICTs and economic development is the methodological difficulties. This issue is discussed in detail in section 3.3.3, but it can be briefly mentioned here that simple regression analysis cannot identify the direction of causality between the two geographical phenomena. For such an analysis more advanced econometric methods are necessary such as Granger causality test (see section 3.3.3). Regardless of the complexity of the methods and the rather limited number of scholars in this research area, Table 2.4 reviews a number of studies which dealt with this problem.

Most of the studies presented here focus on national economic development. However, two papers are concerned with regional economic development and ICTs. Additionally, none of the above studies is concerned with the Internet infrastructure. On the contrary, most of them use as proxy for ICTs provision more traditional indicators such as fixed telephone lines per 100 habitants etc. Both the above observations indicate the importance of the current study in identifying the regional

---

[11] The term panel data is analysed in detail in Chapter 3. Briefly, it can be mentioned here that a panel dataset contains variables for a set of countries or regions (cross-section units) over time (time series).

economic development impacts of Internet infrastructure, as such research questions are still emerging in the literature.

Nonetheless, the most important finding of the table below is the opposing results of the different studies. Almost half of them concluded a bidirectional relationship between the different measures of ICTs and economic development. Slightly less but still a significant number of studies found a unidirectional relationship with ICTs causing economic development. Interestingly enough, only a few studies came up with a reverse causal relationship where ICTs are pulled by economic development. This wide range of results about the causal relationship between ICTs and economic development signifies the need for further exploring this research area in regards to the Internet infrastructure.

Table 2.4: Causality studies on ICTs and economic growth based on Granger causality tests

Reference	Geographic/time extent	Direction of causality
Cronin et al. (1991)	USA; 1958–1988	Telecom investment ←→ GDP
Cronin et al. (1993a)	Pennsylvania, USA; 1965–1991	Telecom investment ←→ employment
Cronin et al. (1993b)	USA: 1958–1990	Telecom investment → aggregate and sectoral productivity growth
Dutta (2001)	15 developing and 15 industrialized countries; 1960–1993	Teledensity → per capita GDP Telephones → GDP
Chakraborty and Nandi (2003)	12 Asian countries; 1975–2000	Degree of privatization: High: teledensity ←→ GDP Low: teledensity → GDP
Cieslik and Kaniewsk (2004)	Regional panel data, Poland; 1989–1998	Teledensity → retail sales per worker
Shinjo and Zhang (2004)	Japan: 38 industries USA: 31 industries	Japan: productivity growth → ICTs investment
		USA: ICTs investment ←→ productivity growth
Yoo and Kwak (2004)	Korea; 1965–1998	Information technology investment ←→ GDP
Beil et al (2005)	USA; 1947–1996	GDP → Telecommunications investments
Chu et al (2005)	New Zeland, 1987-2003	ICTs → GDP
Wolde-Rufael (2007)	USA; 1947–1996	GDP ←→ Telecommunications investments
Shiu and Lam (2008)	Regional panel data, China; 1978–2004	GDP → Teledensity/penetration rate
		High income/teledensity/penetration: Teledensity/penetration → GDP

Adapted by Shiu and Lam (2008, 707) and further extended by the author

2.6 Epilogue

This chapter's main objective is to critically present and analyse the relevant literature in order to create the theoretical framework for approaching the research questions of this doctoral thesis. First of all, a brief technical description of the Internet was provided. Such technical knowledge is vital for this research: despite the geographic orientation of this study's research questions, there is a need for the researcher to familiarise himself / herself with the technical nature of the Internet. This need is designated by the strong technical character of the main research object.

After this, the three main theoretical pillars of this thesis were explored. Firstly, this new emerging field of the Internet geography was introduced. Regardless of the wide perspective of this study, this is the field which this doctoral study best fits in. Secondly, the world city literature was analysed. This research field is highly related with the research subject as the Internet infrastructure facilitates the global city process. Thirdly, the infrastructural character of the Internet as well as a conceptual framework for analysing the regional economic development impacts of the Internet infrastructure was introduced.

The above three theoretical pillars support and are directly linked with the empirical aspects of this research. While the Internet geography pillar is linked with all of the four empirical chapters, the world cities literature is mostly linked with the comparison between the Internet backbone and the aviation network (Chapter 5) and with the descriptive network analysis of the Internet backbone network (Chapter 4); the explanatory analysis of the socio-economic factors that shape the Internet backbone distribution in Europe (Chapter 6) is directly linked to the Internet geography field, but also to the third pillar of this doctoral study; and finally the last empirical chapter (Chapter 7), where the regional economic impacts of the Internet infrastructure are evaluated, mostly draws upon the Internet infrastructure and regional economic development, but also upon the Internet geography as well. Before moving on to these research chapters, the main methodological issues are discussed in Chapter 3.

Table 2.5: Links between theoretical pillars and research elements of this doctoral thesis

Theoretical pillars	**Chapter 4** Descriptive (network) analysis	**Chapter 5** Comparison with aviation network	**Chapter 6** Geographical analysis of the socio-economic factors that shape the backbone provision in Europe	**Chapter 7** Internet infrastructure and regional economic development
Internet geography	++	+	++	+
World cities	++	++		
Internet infrastructure and regional development			+	++

Chapter 3

Methodology

3.1 Introduction

The aim of this chapter is to describe and justify the utilisation of specific methods and data sources used in this doctoral study in order to address the research questions defined in Chapter 1. Firstly, the lack of data in the field of the Internet geography is addressed as well as the research choices made in order to overcome this problem. Additionally, in section 3.2 the other data sources used in this study are also described. Section 3.3 describes the quantitative methods used for approaching the research questions. While section 3.3.2 focuses on the methods which deal with the network topology of the research subject, section 3.3.3 describes and justifies the utilisation of statistical and econometrical modelling. Lastly, section 3.4 summarises the discussion of this chapter and presents the research framework of this study, which is the fundamental for the empirical analysis which takes place in the following four chapters.

3.2 Data

As identified in Chapter 2, there are two main difficulties in approaching the physical layer of the Internet infrastructure: firstly, the invisible character of the infrastructural network, and secondly the lack of data because the confidentiality reasons. Various approaches have been introduced in the literature in order to overcome the latter and analyse the physical layer of the Internet infrastructure and more specifically the main core of this infrastructure, the Internet backbone networks. In the early stages of the field of Internet geography, researchers combined the few free sources available at this time and they constructed databases for the US intercity Internet backbone links. The main sources were the Boardwatch Magazine and its Boardwatch Directory of Internet Service Providers and also the Cooperative Association for Internet Data Analysis (CAIDA 2009 – for example Moss and Townsend 2000, Malecki 2002a, Gorman and Malecki 2002). However, the former ceased publication in early the 2000s and the latter stopped updating the relevant research project called *Mapnet* (CAIDA 2009).

Apart from these freely available data sources, two other data sources for the physical infrastructure of the Internet were used over the (short) history of the emerging field of the Internet geography and both of them were private consultants: Telegeography (2009) and KMI Research Group (2001). The former provides telecommunications analyses and specialised datasets over the last 30 years (Telegeography 2009). It has been used in various papers and it can be said that it is the *classic* data source when the discussion goes to data about Internet backbone networks (for example Devriendt et al 2008; Evans-Cowley et al 2002; Gorman and Kulkarni 2004; Gorman and Malecki 2002; Malecki 2002a and 2004; Prufer and Jahn 2007; Townsend 2003). The second consultant provided similar data and has been used as a source for ESPON (2005b), Rutherford et al (2004 and 2005), Rutherford (forthcoming) and Tranos and Gillespie (2009). However, KMI Research Group no longer exists as a firm.

Thus, nowadays the only available source for secondary data in regards to the Internet backbone networks is Telegeography. Among others, Telegeography provides data for international intercity backbone links, which function at the level 3 of the OSI model. This data refers to the Internet bandwidth (capacity), but not to the Internet traffic (capacity usage) and represents point-to-point rather than end-to-end relationships. As analysed in section 2.2 Internet data may travel between any two points anywhere on the network, but the Internet bandwidth shows only the routes available for each individual hop between those points. As Telegeography highlights (2007, 1) “end-to-end traffic data are based on an aggregate of individual usage, while point-to-point bandwidth indicators aggregate IP capacity logically provisioned over a physical network”. Briefly, it can be said that Telegeography dataset has the following advantages: (a) it is more recent – up to 2006, (b) it includes data for network capacity, and (c) it includes extra-European links.

As expected, the network capacity and usage are not unrelated. By assuming that the market economy works efficiently, supply meets demand and consumers (users in this case) act rationally, then from the supply side carriers would install as much bandwidth as needed in order to meet the demand for this infrastructure and from the demand side, users would pay to use as much bandwidth as they would really need. So, all the installed bandwidth would be used and as long as the data interchange reflects to some degree the interaction between two cities, then the installed bandwidth would indicate the interaction between these two cities. But in fact, most installed bandwidth is unused. For example, in 2004 only 3% of the total bandwidth capacity in Frankfurt was lit (Rutherford forthcoming) and by the end of 2006 only 14% of the total capacity of major submarine cables was being used (Roberts 2006). Before rejecting the previous argumentation as a market failure, the special characteristics of this infrastructure should be considered. What really costs in backbone networks is digging trenches in order to install ducts, rather then laying fiber once the ducts are installed. So, unlike the case for transport infrastructure, what really costs is the network's first installation and not its expansion in terms of extra fiber and bandwidth. Therefore, the volume of the lit or the unlit 'dark fiber' and the ISPs' philosophy of "build it and they will come" (Youtie 2000) indicate the expectations about the current and the future demand for this intercity linkage.

In order to obtain the above data, Telegeography integrated three research methods:

> "confidential surveys, informal discussions, and follow-up interviews were conducted with network engineering and planning staff of major backbones. [Additionally] standard and slightly modified network discovery tools were deployed from a large number of locations to gather an extensive data set on network topology, which was then parsed for identifiable characteristics, including geographic location. Finally, additional public and private information sources were consulted to verify and add to the data already in place" (ibid).

Based on the above but mostly on the long history and reputation of this consultant in providing services in this field, a strategic decision was taken to purchase from Telegeography the dataset for the international intercity Internet backbone links for the period 2001-2006. Currently, Telegeography is the only provider for such data. Additionally, an individual researcher could not build such a dataset on his/her own mostly because of lack of trust on behalf of the carriers. As mentioned above, the main reason for the lack of data in this field is the confidentiality of such data. Therefore, the strategic advantage of Telegeography is that it managed to overcome this difficulty over the years and now can accumulate and provide such data. It should be noted here that Telegeography's data is aggregated so the least amount of details possible is revealed for individual carriers. This specification is not a problem at all for the needs of this study as our focus here is not the individual Internet backbone providers, but rather their aggregated capacity in order to examine the urban economic geography of these networks.

While a more detailed description of the dataset takes place in Chapter 4 with the first analysis of the data, the aim here is to provide a brief description of the data and mostly justify its value for approaching the research questions of this study. Initially, the data was provided in the form of a table with the international intercity links and the total capacity of these links for the years 2001-2006. The first step was to link the European cities with NUTS3[12] regions. In the cases where more than one city of a NUTS3 region were connected with at least one Internet backbone network, these cities were summarised in a way that no NUTS3 region has more than one representation in the database. This *database cleaning* process underlies the assumption that a NUTS3 region is the nearest statistical unit to the notion of city-region described in section 2.4.6 and it means that all the Internet backbone nodes in this region facilitate the same urban area. The choice of NUTS3 instead of NUTS2 was made as the former has a stronger urban character since the latter includes a more extended area. Two processes took place after this. The first one was to *build* a network topology based on the above links. With the use the specialised software UCINET (Borgatti et al 2002) a network of the international Internet backbone links of European cities for the years 2001-2006 was built. This process included intensive data manipulation in order to bridge the tabular form of the initial data with the needs of a network topology. The latter is the medium to apply network analysis. A basic description of the main theoretical elements of network analysis and complex network analysis is presented in section 3.3.2.

The second process was to build a database linking the results of the network analysis which refer to the nodes of the network (i.e. city-regions) with socio-economic indicators, using again the NUTS3 regions as the spatial unit. The main sources for such socio-economic variables were Eurostat (2009) and ESPON (2006). Such a database is the fundamental for the analysis of the regional economic impacts of the Internet infrastructure, which takes place in Chapter 7. However, in order to exploit in full capacity the existing data and more importantly to obtain as robust results as possible, instead of creating a simple cross-sectional database, a panel

---

[12] NUTS stands for Nomenclature of Territorial Units for Statistics and is the official territorial units for the provision of regional statistics (EC 2009). NUTS2 regions have a more regional character, while NUTS3 are closer to metropolitan area or a city-region.

dataset was built. A brief methodological description of this takes place in section 3.3.3.

Additionally, two more other datasets are used in this study. The first represents aviation data and it is used in Chapter 5, where a topological comparison of the Internet backbone and the aviation network in Europe takes place following the argument presented in Chapter 2 that Internet infrastructure and aviation comprise the 1st layer of the space of flows. The aviation data comes from the International Civil Aviation Organization (ICAO), which is a specialised agency of the UN setting "standards and regulations necessary for aviation safety, security, efficiency and regularity, as well as for aviation environmental protection" (ICAO 2008). This dataset also refers to the capacity of international intercity links, but the main measure here is annual intercity passenger flows. The aviation data was initially provided by ICAO in a semi-tabular form (html webpages). So, the first step was to create actual tables with bilateral links and then based on this and in a way similar to the Telegeography data, the network topology of the international intercity aviation links of the European cities was built. Effort was spent to maintain the two different networks as compatibly as possible in order to facilitate the topological comparison.

The third main dataset used in the study is another earlier version of the Internet backbone data and was provided to CURDS by KMI Research Group for the needs of ESPON 1.2.2 Project (ESPON 2005b). The data was initially provided to the researchers of that study in a map form, which represents the international Internet backbone networks planned or existing in Europe during the third quarter of 2001. From that map, the ESPON research group extracted data, which includes measures about the Internet connectivity of European cities as well as the number of different backbone providers for each city. This data was available for the needs of this doctoral thesis. Again, the same cleaning process as before took place and all the cities were linked and for some cases aggregated to NUTS3 regions. This dataset is used only in Chapter 6 for the needs of the explanatory analysis of the spatial distribution of Internet backbone networks in Europe. This specific dataset was used because when this explanatory analysis was conducted, Telegeography data was not then available. Regardless of any differences between the two different datasets, which are extensively described in section 6.2, both represent the same phenomenon: the long-haul Internet links in Europe.

3.3 Methodological issues

3.3.1 General

In order to analyse the above diverse datasets to answer this study's research questions a variety of quantitative methods has been applied. Briefly, these methods can be separated in two large blocks: the first corresponds to the network topology of the Internet backbone and that is network analysis and complex network theory. The outcome of this analysis bridges the topological space with the geographical one by feeding the second block of the analysis, which uses statistical and econometric techniques to explain the spatial distribution and the regional economic impacts that this infrastructure might generate. As mentioned above, for the network analysis the software UCINET (Borgatti et al 2002, version 6.206) was employed, while the statistical and econometric analysis was based on SPSS (version 15) and STATA (version MP10). The rest of this section is dedicated to further analyse these two methodological blocks.

3.3.2 Network analysis and complex network theory: a brief review

Because of their complex network topology, Internet backbone and aviation networks can be approached using *complex network* theory. Their complex networks nature is well established in the literature (Gorman and Kulkarni 2004; Schintler et al 2004; Faloutsos et al 2009; Gastner and Newman, 2005; Guimera et al, 2005, Amaral et al, 2000). The latter examines large scale networks with complicated and at any case not easy to understand at a glance topology (Fosco 2004). More specifically, in complex network theory what is important is not the behaviour of the single actor but the information of "who is connected to who" (Crucitti et al 2003, 2). The main elements of networks are the nodes (vertices) and the links (edges). Networks were always part of graph theory, whose origins can be traced back to the eighteenth century and to Leonhard Euler's work for small graphs with high degrees of regularity. Thus, initially graph theory was focused on graphs in which all the network nodes have the same degree, or in other words all the vertices are connected with the same number of other vertices. Later, twentieth century graph theory was influenced by the advances in mathematics and statistics and became more algorithmic (Albert and Barabási 2002, Bonarich 2007). Network analysis went a step further with the introduction of *random networks* (RN) by two Hungarian mathematicians Paul Erdös and Alfréd Rényi, which refer to large scale networks with no obvious structure (Erdös and Rényi 1959). The distribution of vertices degree follows a Poisson distribution, which means that the majority of the vertices on the network have the same number of links and they are found nearby the average degree <k>; vertices that deviate from this are rare (Reggiani and Vinciguerra 2007, Albert and Barabási 2002, Fosco et al 2004). RN received a lot of attention for decades after their introduction. The main motivation of evolving further the complex networks theory though, was the question of whether this model can represent the real world networks such as the Internet or the cell (Albert and Barabási 2002).

In the following decades, complex networks played a more important role in different fields, from social science to biology. Albert and Barabási (2002) indicate four different reasons for this: the digitization of data in many different fields and the appearance of large databases enabled scientists to approach different real world systems from the network analysis point of view; second, advances in computer science and in computing power enabled scientists to handle very large databases, which represent better real world systems; third, the looseness between different disciplinary boundaries gave network analysts the opportunity to use a real world network databases from many different fields; and finally (and maybe because of all the above) reductionist approaches lose ground in favour of holistic research approaches, which try to understand the system as a whole (Albert and Barabási 2002).

The second milestone in the evolution of complex network theories is the *small world effect* and the Watts and Strogatz (1998) *small world networks* (SW), no matter that in order to move from the former to the latter thirty years have passed. The small world effect refers to the well known study of Milgram (1967), according to which there is an average distance[13] of 6 degrees between most pairs of people in

---

[13] Just to clarify here that in network analysis terminology distance does not refer to Euclidean distance but to the number of nodes that separate any two nodes. And because usually there are plenty of different ways to connect any two given nodes (also known as a *walk*), we usually focus on the shortest distance, known as *geodesic distance* (Nooy et al 2005). Some studies use the term *diameter* to name the above measure. For the needs of the present study, the term diameter is used to express the longest distance in the network.

the United States. In other worlds, the small world effect is a characteristic of numerous networks and identifies the short average distance among networks vertices, enabling in this way the actors of a network to reach all the rest within a few steps (Reggiani and Vinciguerra 2007). It has been found that most real world networks like the Internet, the actors in Hollywood, the chemicals in a cell etc. are characterized by short average distances. However, the small world effect is a structural characteristic rather than an organizing principle and even RN networks are characterized by short average distances (Albert and Barabási 2002). Watts and Strogatz (1998) developed further this attribute introducing the SW model. The basic feature of this model is the coexistence of short average distance with high clustering coefficient (Fosco 2004). The latter, as we will see later, is a measure of a node's cliquishness (Latora and Marchiori 2001). In fact, SW networks are located between regular and random networks; they are highly clustered like regular lattices, but they also have small distances like random networks (Latora and Marchiori 2002). In addition, their degree distribution is quite similar with the RN networks with a peak value <k> which decays exponentially for large k (Albert and Barabási 2002). So, a SW network can be approached as a set of clusters of nodes, which are highly connected at a local level, but in which there are also some links which span the entire network, linking the furthest clusters. In other words, an actor in such a network can benefit from the high local connectivity but can also be easily transferred to a remote cluster using the intra-cluster links and then take advantage of the high local connectivity in this domain (Batty 2001).

A common characteristic of the RN and SW models is that the probability of finding a highly connected vertex decreases exponentially. This means that the highly connected vertices, which are known as hubs, are practically absent in RN and SW models. And here lies the third milestone of the complex network theory; the introduction of the *scale free* (SF) networks, which are characterized by the existence of a few highly connected hubs and a vast majority of less connected vertices (Barabási and Albert 1999). The term scale free refers to the fact that their vertex degree distribution follows a power law distribution no matter what the observation scale is (Reggianni and Vinciguerra 2007). Such networks are being formed according to two mechanisms, growth and preferential attachment (Albert and Barabási 2002). The former refers to an attribute which is common in many real world networks, that is the actual expansion of the networks through time by the vertices' and edges' increase. The second mechanism refers to the fact that this growth is not equally dispersed across the network's vertices. On the contrary, highly connected vertices are more likely to be preferred by the new vertices (Reggianni and Vinciguerra 2007). Because of the preferential attachment, a vertex that acquires more connections than another one will increase its connectivity at a higher rate; thus, an initial difference in the connectivity between two vertices will increase further as the network grows, indicating a *rich get richer* phenomenon. The probability Π that a new vertex will be connected to a vertex *i* depends on the degree k_i of the vertex *i*,

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \text{ (Barabási and Albert 1999)}$$

The above mechanisms resulted in networks with vertex degree distribution which are governed by power law. So, the probability P(k) that a vertex has a degree k, or in other words interacts with k other vertices, decays with a power law

$P(k) \approx k^{-\gamma}$, with usually $2 < \gamma < 3$ (Barabási and Albert 1999).

In order to replicate the power-law distribution present in many real world networks both the above two mechanisms should be present simultaneously (Albert and Barabási 2002). It should be highlighted here that initially the SF model included only the above mechanisms for networks evolving. The importance of this model in the multidisciplinary field of network analysis is indicated by the numerous studies published later on SF networks. As a result, more realistic approaches regarding networks evolution were introduced. According to them, a network can be changed by any combination among the following 4 events: addition or removal of a vertex and addition or removal of an edge. Nevertheless, in real life networks the above happen simultaneously resulting in a phenomenon called *re-wiring*, which is part of many models which followed the initial BA model (see Albert and Barabási 2002 for an extensive review).

Table 3.1 summarizes the basic characteristic of the above three network models. If we were to compare them, it could be said that both RN and SW have short average distances, but RN cannot be included in SW because they lack the high clustering coefficient (Regianni and Vinciguerra 2007). In addition, SF networks share the short average distance and the high cluster coefficient of SW ones, but the SW are not characterised by the scale-free distribution (Gorman and Kulkarni 2004). Or in other words all scale free networks are believed to display small world properties while all small-world networks are not necessarily scale free (Sen et al 2003). Amaral et al (2000), analyzing further the above, suggested that because of SF networks' small world properties, SF networks are also part of SW. And they continue by distinguishing SW in three sub-categories: (a) SF networks with a power law degree distribution; (b) *broad-scale* networks, which can be recognized as truncated SF networks, which have a power law regime in their degree distribution followed by a sharp cut-off, such as an exponential or a Gaussian tail decay; (c) single-scale networks with an exponential or Gaussian degree distribution.

Table 3.1: Overview of main complex network models characteristics

		RN	**SW**	**SF**
Physical Measures	Average short path	Short	Short (scales as L~lnL)	Very short (scales as L~lnlnN)
	Clustering Coefficient	Low	High	High, but it decreases with the increasing of the network size N
Statistical Measures	Vertex connectivity degree distribution	Poisson $P(k) \propto e^{-k\frac{k^k}{k!}}$	Similar to RN, decaying exponentially for large set of vertices	Power law $P(k) \propto k^{-\gamma}$
	Exponent degree			$2 < \gamma < 3$

Source: Regianni and Vinciguerra 2007, 151

Another important element of the above network models is their tolerance of faults and attacks. According to relevant studies (Albert at al 2000, Albert and Barabási 2002, Crucitti et al 2004, Li et al 2005, Audestad 2007), SF networks are characterized by high tolerance in randomly connected nodes failure. On the contrary, such networks are vulnerable in attacks to specific vertices. More specifically, Albert et al (2000) showed that the average short distance of a SF network, which is a proxy of network's efficiency, remains the same even if a randomly selected 5% of vertices fail. This happens because of SF networks' severe inhomogeneous connectivity distribution, which occur because of the power law degree distribution. In plain English, because only few of the nodes in a SF network enjoy high connectivity and the rest are less connected, there is a small probability that among the nodes that fail, which presumably are randomly chosen, some of the highly connected ones are included. This is not the case though for random networks. Because of their homogeneity, which means that most of the vertices have similar degree, random nodes' failures have significant effects on networks' efficiency. On the other hand, when a SF network is under attack, which usually means that its most connected nodes are targeted, it is more vulnerable than homogenous random networks because of its heterogeneity. If 5% of its most connected vertices are removed, SF network's average short path is doubled. However, RN networks have the same behaviour both when randomly selected nodes are down and when the network is under attack because of their homogeneity (Albert et al 2000).

Nonetheless, recent studies reject the above argumentation as oversimplifying. Usually, the empirical verification of a SF network is limited by studying its vertex degree distribution and assuming that because it follows a power law the rest of SF networks attributes are present. However, Li et al (2005) proved that there is a great variety of networks whose vertex degree distributions follow the same power law but they are characterized by different quantitative and qualitative attributes. In order to better define the SF character, Li et al (2005) introduced the *s* metric:

$$s(g) = \sum_{(i,j)\in\varepsilon} d_i d_j ,$$

where d_i and d_j are the degrees of any nodes *i* and *j* respectively, which are connected by a link and ε is the array of all the links present in the graph *s*. In order to compare the *s* metric of different networks, the indicator is normalized by the s_{max} value, as it is described in Li et al (2005). This indicator measures the hub structure of the network and the higher the value of the *s* indicator is, the more common it is for highly connected nodes to be connected with similar highly connected ones. So, in order to confirm the *robust yet fragile* structure of the SF networks it is not enough to identify the existence of some highly connected nodes and a vast majority of less connected ones, as it is reflected by the power law degree distribution. The importance of the highly connected nodes in *holding the network together* is identified by the interconnection of the highly connected nodes together. For instance, if a high degree vertex, which is only connected with low degree vertices, is removed from the network then the worst case scenario is that some of the low degree nodes, which enjoyed connectivity only through the removed hub, will be disconnected. On the contrary, if a highly connected node, which is connected with a similar degree vertex, is removed from the network, the results might be more severe and affect the efficiency of the whole graph because of the higher structural importance of such a node. Accordingly, the power law vertex degree distribution cannot prove on its own the existence of highly connected hubs, which play the role of the Achilles heel for SF networks. On the contrary, such nodes appear in a network with a power law vertex degree distribution, only when this network is characterized by high *s* value.

To sum up, Albert and Barabási (2002) suggested three key elements of complex networks: the small average distances and the high cluster coefficient, which are related with the small world phenomenon, and the vertices degree distribution. In addition, s metric is equally important in order to identify the existence of any hubs. The above are empirically studied in Chapters 4 and 5 for the Internet backbone and aviation networks.

3.3.3 Statistical and econometric techniques

The second block of methodologies used in the analysis presented in the following chapters is based on modelling techniques and multivariate analysis. This analysis is directly linked with the topological analysis of the network space as the main focus is the derivatives of the network analysis which are allocated at the city-region level.

There are two main themes for the analysis of which such methodologies were adopted: (a) the explanatory analysis of the spatial distribution of the Internet backbone networks across European city-regions which is presented in Chapter 6; and (b) the analysis of the impacts that the Internet infrastructure, as it is reflected in the Internet backbone networks, can generate on the economic development of the city-regions in Europe which is presented in Chapter 7. In order to approach the above, specific statistical and econometric techniques have been adopted.

Firstly, the explanatory analysis of the spatial distribution of the Internet backbone networks is based on a combination of p*rincipal components analysis*

(PCA) and regression analysis, which is know in the literature as *principal components regression* (PCR) (for example Massy 1965). In order to explain the socio-economic factors behind the allocation of the Internet backbone networks, a large dataset of socio-economic variables was formed for European city-regions. The usual methodological choice would be to include all these variables as independent variables in a regression model, the dependent variable of which would be a proxy for the Internet infrastructure at the city-region level. However, such a choice would be problematic because of the multicollinearity problems that occur when a large set of socio-economic variables is used as independent variables in a regression model. One of the methods suggested by the literature in order to overcome this problem is to *group* the independent variables using PCA and use the new components as the regressors of the model. Because of the orthogonal transformation that takes place the new components are not collinear, eliminating multicollinearity problems which might appear if the initial large dataset of socio-economic variables is used as the regressors. Additionally, such a data reduction method could also result in better understanding of a large dataset of explanatory variables without loosing the explanatory value that a large dataset can provide. For the above reasons, PCR was selected in order to explain the socio-economic factors that shape the spatial distribution of the Internet backbone networks across European city-regions.

Briefly, three models are developed for the explanatory analysis, based on the KMI Research Group (2001) data. For all three a notion of the Internet backbone *connectivity* is used as the dependent variable. In simple words, all three models attempt to explain which socio-economic factors shape the distribution of these different notions of connectivity. The first one is based on *logistic regression* of the new components occurred by PCA and predicts the likelihood that a city-region is connected at least to one Internet backbone network. For this model all the NUTS3 regions of Europe are included. The second model focuses only on those connected city-regions. The dependent variable is the number of connections a city shares with the rest of the European city-regions, which is a continuous variable. In order to explain this notion of connectivity, *ordinary least squares* (OLS) regression is used on the new components derived from the PCA. This model explains why some city-regions are better connected than others in terms of the Internet backbone. Thirdly, another notion of connectivity was modelled using OLS: the number of Internet backbone providers present in a NUTS 2 region. The main change here is the spatial level. NUTS2 regions were adopted because at this level more socio-economic variables regarding the knowledge economy are available following what was highlighted in the previous chapter about the link between the knowledge economy intensity and the agglomeration of Internet infrastructure (Malecki 2004). Again a regression model based on OLS is used having as regressors the new components resulting from the PCA.

Secondly, *econometrics* has been used in order to facilitate the analysis of Chapter 7 about the impacts of Internet infrastructure on the economic development of city-regions in Europe. According to Wooldridge (2003, 1) “econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories and evaluating and implementing government and business policy”. In other words, econometrics is the social science which combines tools from economic theory, mathematics and statistics in order analyse economic phenomena (Goldberger 1964). Such tools are exploited in Chapter 7 in order to model and empirically assess the relation between the internet infrastructure and regional economic development. More specifically, in order to approach this, *panel data* analysis is used. The latter refers to those datasets which apart from having a cross-section dimension, also included time as the second dimension of the dataset. In simple words, while a cross-section dataset – as the one used for the explanatory analysis described above – only includes data for the different cross-section units – city-regions for our case – for a number of variables, a panel dataset also includes different observations for the same variables for these cross-section units over time (Wooldridge 2003).

For the case of this doctoral research, the panel data specification is chosen contrary to a cross-sectional approach as it enables us to assess the impacts of the Internet infrastructure in a dynamic framework. As was described in Chapter 3 in the discussion about the productivity paradox, sometimes it takes time for the impacts of a specific technology to become visible. The panel data specification provides the needed framework in order to assess the impact of the infrastructural capital installed in year t-n on the economic development level of year t across a set of city-regions. What is more, panel data also enables the researcher to control for omitted variable bias and therefore panel data is a better methodological choice in approaching the impact of Internet infrastructure on the economic development of city-regions. In order to do so panel data regression models are developed in Chapter 7. These models are directly linked with the network analysis which takes place in Chapter 4 and 5 as the main independent variables for these models are the different centrality measures arising from this analysis. The dependent variable for these models is always the economic development level and in addition a number of control socioeconomic variables are also included in the models.

However, as was illustrated in Figure 2.4, establishing the link between infrastructure provision and regional economic development usually raises a *causality* problem. In other words, it is not always clear which the direction of causality is and whether the infrastructural capital is the cause or the result of economic development. The panel data specification contributes in addressing this phenomenon, something which is not possible with simple cross-section data. More specifically, recent developments in panel data analysis enable the application of *Granger causality test* (Granger 1969), which were initially introduced for time-series, on panel data (Hoffmann at al 2005).

The former is differentiated by panel data because it lacks the cross-section dimension (Maddala 2001). Such a test enables the researcher to investigate the direction of causality between two variables. Briefly, the Granger test is based on a model where the dependent variable y is regressed against k lagged values of y and k lagged values of x. Based on such a model the null hypothesis can be tested according to which x does *not* cause y. If the test proved to be significant then the null hypothesis can be rejected and then it could be concluded that x causes y (Hood III et al 2008). This latter means that y is better predicted if all the information (i.e. both the lagged values of y and x) is included in the model than when the lagged values of x are excluded (Hurlin and Venet 2003). In order to evaluate both of the directions of causality, the above model takes place twice interchanging the dependent with the independent variable in order to evaluate the impact of Internet centrality on economic development, but also the impact of economic development on the centrality of the city-regions. The value of using panel data for Granger causality test instead of time series is twofold in addition to the omitted variable bias described above (Shiu and Lam 2008): not only does panel data provide more degrees of freedom than conventional time series, but it also takes account of the heterogeneity among the cross-section units. The latter is very important for a study in the field of geography because the direction of causality can be differentiated across the cross-section units. Put simply, the Internet backbone centrality might have a significant impact on economic development in some city-regions while in others it might not or even, for some cases, the Internet backbone centrality might rather be the result of economic development.

In general, the Granger causality test appears to be the most widely used method for empirically assessing causal relationships in the field of regional science, but also in econometrics (Erdil and Yetkiner 2009; Bronzini and Piselli 2009; Hoffmann et al 2005; Chamberlain, 1982; Florens and Mouchart 1982; and Hood III et al 2008). Even more specifically, it has been widely used in defining the causality between telecommunications and (regional) economic development (see Table 2.4). It is preferred by econometricians as an empirical method which can be easily implemented at least for time-series, where commercial econometric packages include specific ready-made routines (Hoover 2001). However, the implementation of the Granger causality test for panel data is a rather more complicated process as no commercial econometric package to date includes such a routine.

3.4 Research framework

The aim of the last section of this chapter is to review the above discussion about the methods and the data used in order to answer the research question stated in Chapter 1. In order to address the three research questions of this study two diverse blocks of quantitative methods (network analysis and complex network theory on the one hand and statistic and econometric modelling on the other hand) and five data sources (Telegeography 2007; KMI Research Group 2001; ICAO 2008; Eurostat 2009 and ESPON 2006) are utilised in the following four chapters. This research process is supported by the three theoretical pillars of this doctoral study as discussed in Chapter 2. The above elements are summarised in Figure 3.1, which presents the research framework of this thesis.

Figure 3.1: Research framework

Research question	**RQ1**	**RQ2**	**RQ3**
Methods	network analysis, complex network theory	principal component regression (PCA + logistic / OLS regression)	panel data analysis / Granger causality test
Data sources	Telegeography (2007) ICAO (2008)	KMI Research Group (2001), Eurostat (2009), ESPON (2006)	Telegeography (2007), Eurostat (2009), ESPON (2006)
Theoretical pillars (from Chapter 2)	IG, WC	IG, WC	IIRD, IG, WC
Chapters	4, 5	6	7

RQ1: How is the Internet infrastructure allocated across the European city-regions?

RQ2: Which are the geographic and socio-economic factors that shape the distribution of the Internet infrastructure across the European city-regions?

RQ3: What are the impacts that the Internet infrastructure can generate on the development of the city-regions in Europe?

IG: Internet geography; **WC:** World cities; **IIRD:** Internet infrastructure and regional development

Chapter 4

Descriptive (network) analysis

4.1 Introduction

The main objective of this chapter is to explore and analyze the dataset that is used for the needs of this study. More specifically, in this chapter the initial exploration and mapping of the international Internet backbone links, which enable the IP data transfer among European cities, takes place. For an empirical research approach, the results of this part are essential in designing the research process and consequently for defining the direction of the whole research.

This chapter starts with a shot description of the dataset. As the general data description took place in section 3.2, a brief but more detailed and targeted for the needs of this chapter description is presented here. Then, the basic descriptive statistics and the links mapping take place. Afterwards, basic network analysis methods are applied in order to better analyse the role of the European cities in the Internet backbone network and to take a first glance at the geography of the Internet infrastructure. Different centralization and centrality measures are applied in order to explore the cities' role as nodes of such networks. The synopsis of the above indicators is made using cluster analysis, resulting in a taxonomy of European cities regarding their role in the network. Then, complex network theory is used in order to explore whether the Internet backbone network fits with the well known theoretical network models. This chapter ends with some conclusions.

4.2 Data description

The initial Internet backbone dataset refers to all the international backbone connections present in European cities for each year of the period 2001-2006. It represents symmetrical Internet links, which follow the Internet Protocol (IP) and they are characterised by capacity counted in Megabits per second (Mbps). In reality, the links included in the data base are aggregations of the different fibre links installed and managed by different Internet Backbone Providers for each pair of cities. For example, the 58 fibre optic circuits that connected London with Paris in 2006 were managed by 35 Internet backbone providers but they are represented in the database by a unique link, the capacity of which is equal to the sum of all the different backbone links (Telegeography 2007). Comparing the above network with the overall global network, what is missing are the domestic connections as well as the non-European connections. For example, connections between London and Manchester and Tokyo and New York are missing, while connections between London and New York and London and Paris are present in the dataset. This limitation is due to data unavailability. In terms of geography, the absence of non-European international links prevents us from discussing the importance of cities outside of Europe because we are only aware of a fraction of their total connectivity (i.e. their connectivity with Europe and not with the rest of the world). In addition, the absence of domestic connections prevents us from looking into the cities which only have domestic roles and direct the analyses on cities with international importance in the Internet backbone network. However, because of structure of this network, the cities which are important at an international level are also important at the domestic level because they act as gateways for the whole country, enabling the rest of a country's cities to obtain universal Internet connectivity through them.

In order to further analyze the data, a sub-network of the backbone connections including only the intra-European links for the six years was subtracted from the initial one. For example, links like London – New York were removed resulting to networks with only intra-European links[14]. The reason behind such an extraction is to focus on cities' importance at the European level, without taking into consideration their out of Europe connections. However, it should be highlighted that the interpretation of such a subtracted network is not always straightforward. For instance, it is well known that because of the dominance of the USA in the development of the Internet, a significant part of the intra-European Internet traffic was routed through USA (Townsend 2003). By extracting the links that connect Europe with USA, the infrastructure which facilitated this transatlantic Internet packet flows is missed, diminishing in this way the importance of the US cities in global (and even in European) Internet function and on the other hand overestimating the importance and the autonomy of European cities. However, taking the above scalar limitations into consideration we can study the European part of the above global infrastructural network and the way it interconnects the European cities. And

---

[14] For the needs of this analysis, Europe's east borders are defined as the borders of European Union. In addition, West Balkans are also included.

by adding the extra-European links in the analysis when need be, the big, universal picture of this infrastructural network can be approached.

Another point is whether the network is weighted or not. For the case of the Internet backbone network, the edges are valued with the actual bandwidth of the link. However, it is very common in network analysis to use binary connections instead of using weighted ones. In this case, the value 1 for an edge points out that this link is active while a value 0 indicates the absence of connection between two nodes. The reason for this is to simplify the network and to highlight its structural characteristics. For the needs of this study, both network versions have been developed and are used when appropriate.

4.3 Descriptive statistical analysis

The above process resulted in two different extractions of the same network (for all the cities and only for the European ones) and in two different versions of them (binary and weighted) for each year of the period 2001-2006. The two weighted versions of the two extractions are presented in the following maps. Figures 4.1 and 4.2 present the Internet backbone links among the European cities for 2001 and 2006 respectively. The links are classified regarding how many standard deviations above the mean of all the links the capacity is. In 2001, the links with the highest capacity (bandwidth greater than 2.5 standard deviations above the mean) connected London with Paris, Brussels and Amsterdam, Frankfurt with Paris and Amsterdam and Amsterdam with Brussels. It is not coincidental that these links, which are the peaks or otherwise the outliers of the backbone links distribution in Europe, are concentrated in Europe's pentagon[15]. The only backbone link, whose capacity was more than 1.5 standard deviations greater than the mean and was outside of Europe's core, was the link between Stockholm and Copenhagen. Light green colour depicts the links which are between 0.5 and 1.5 standard deviations above the average. Again, most of them are found in Europe's pentagon but also in the two corridors connecting the Scandinavian countries with west Europe. This class's links towards Central and Eastern Europe are rare and only a few of them terminate in Vienna,

---

[15] Pentagon is the core area of the EU and is defined as "a geographical zone of global economic integration" (EC 1999). It is encompassed by London, Paris, Milan, Munich and Hamburg.

Budapest and Prague. Interesting is the vast amount of low backbone links that cross central Europe and mainly Germany and terminate in Eastern Europe and mainly Vienna. Regarding Europe’s west edge, Madrid is the main gateway city, since is the only city in the area which has at least a link of higher than the average capacity.

In 2006 the spatial allocation of the intra-European backbone links is rather changed. The outlier backbone links are still focused in the Pentagon connecting London with Paris, Amsterdam and Frankfurt, Paris with Madrid and Amsterdam with Frankfurt. The capacity of these links is higher than 2.5 standard deviations above the mean. Comparing with the 2001 allocation, the main change is the upgrade of the link between Madrid and Paris as well as Brussels’ absence in the cities which are served by the highest capacity backbone links. In addition, the link between Stockholm and Helsinki was also upgraded. However, the main differentiation with 2001 is the expansion towards the East. More specifically, many more backbone links with capacity between 0.5 and 1.5 standard deviations above the mean terminate in Central, Eastern and South-eastern European cities. Vienna, Bratislava and Prague seem to be better connected with the Pan-European Internet backbone networks. In addition, even more remote links such as these between Frankfurt and Warsaw as well as Milan and Athens are characterised by a capacity greater than the average. Furthermore, Madrid’s monopoly in high capacity backbone links in the Iberian Peninsula does not exist any more since both Lisbon’s and Barcelona’s backbone links are above Europe’s average in 2006.

In order to wrap up the above, Figure 4.3 presents the box-plots, Figure 4.4 the frequency distributions for both versions of the Internet backbone networks in 2001 and 2006 and for both geographical extents, and Table 4.1 presents some basic descriptive statistics. The shrink boxes and the dispread extreme values in the box-plots indicate the nature of our data set: there is a significant number of backbone links the capacity of which is far away from the median capacity[16]. Obviously, these links play a structural role in the Internet function but also indicate the volume of the potential interaction of the cities they interconnect. Additionally, as it can be observed in Figure 4.4, the frequencies of lower capacity are much greater than the frequencies of links with great capacity, indicating a highly skewed distribution.

---

[16] The points marked with the star symbol represent backbone links with capacity greater than 1.5 inter-quartile ranges.

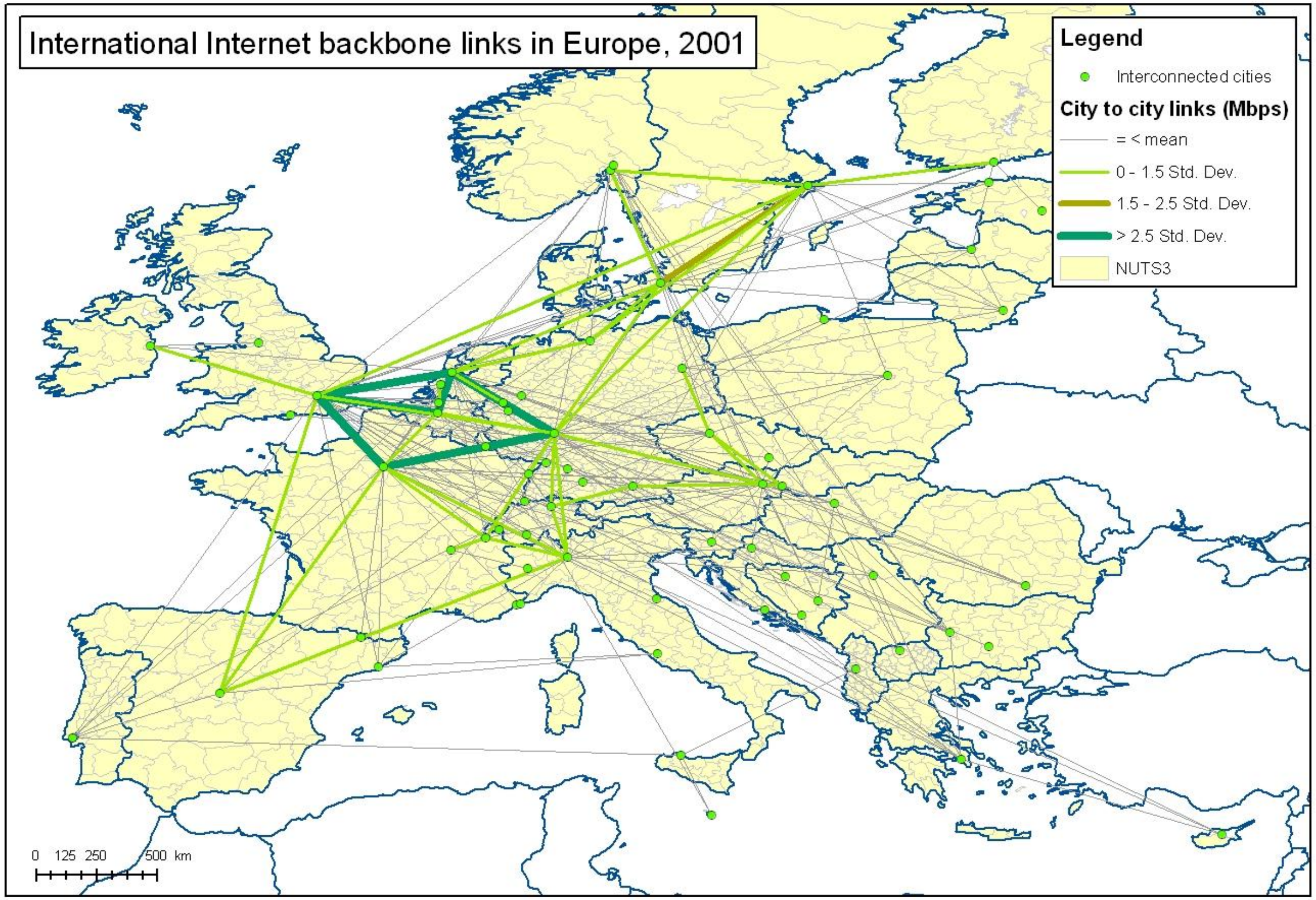


Figure 4.1: International Internet backbone links in Europe, 2001

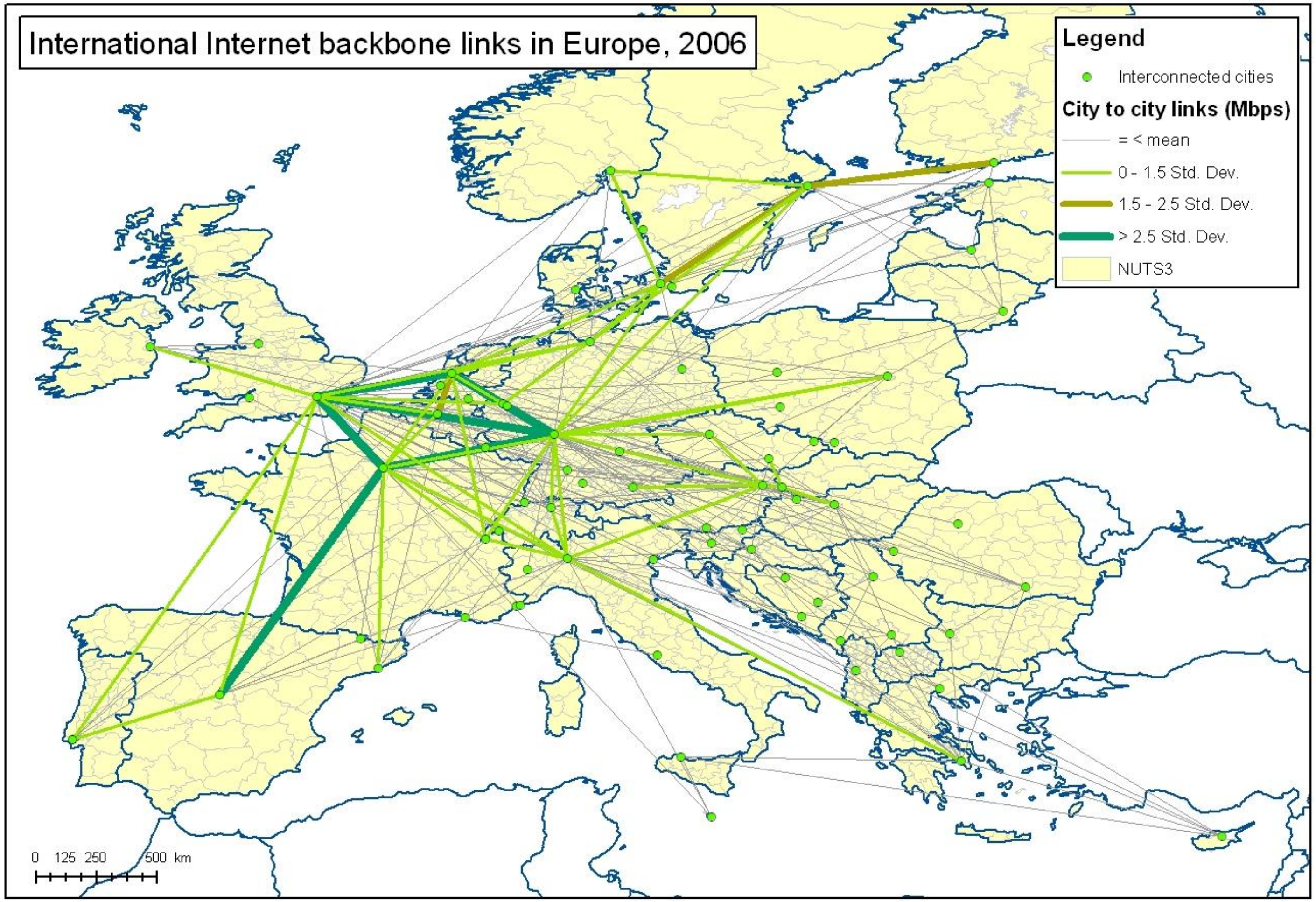
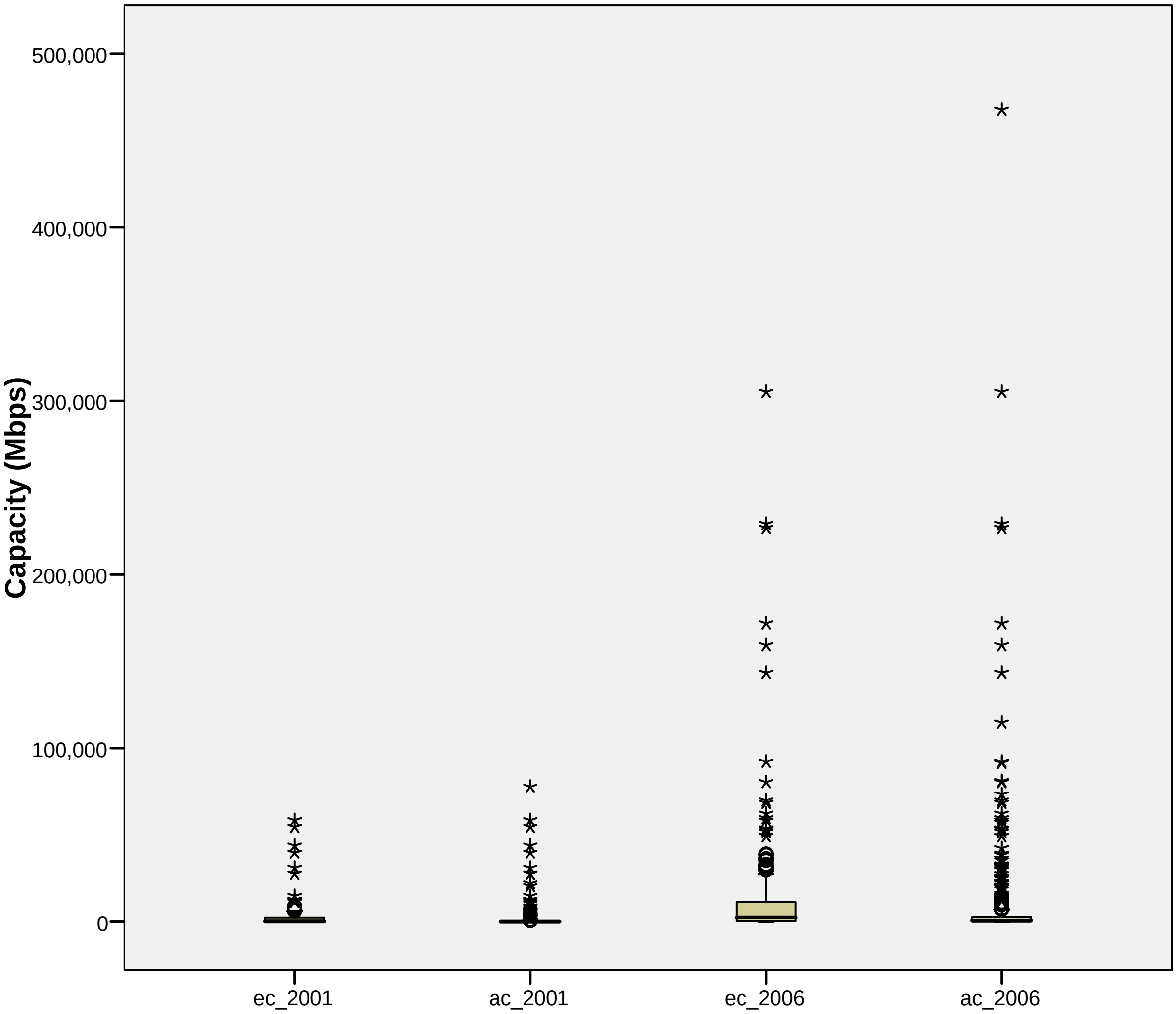


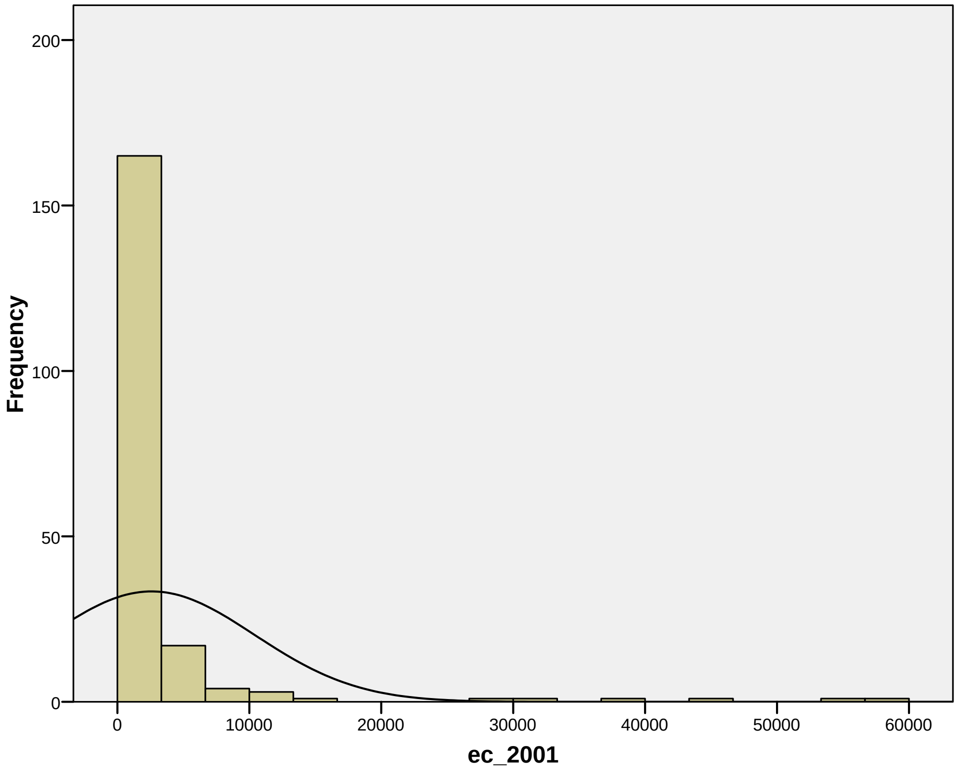
Figure 4.2: International Internet backbone links in Europe, 2006

Figure 4.3: Box-plots for backbone links



ec = European cities, ac = all the cities

Figure 4.4: Frequency distributions for the backbone links



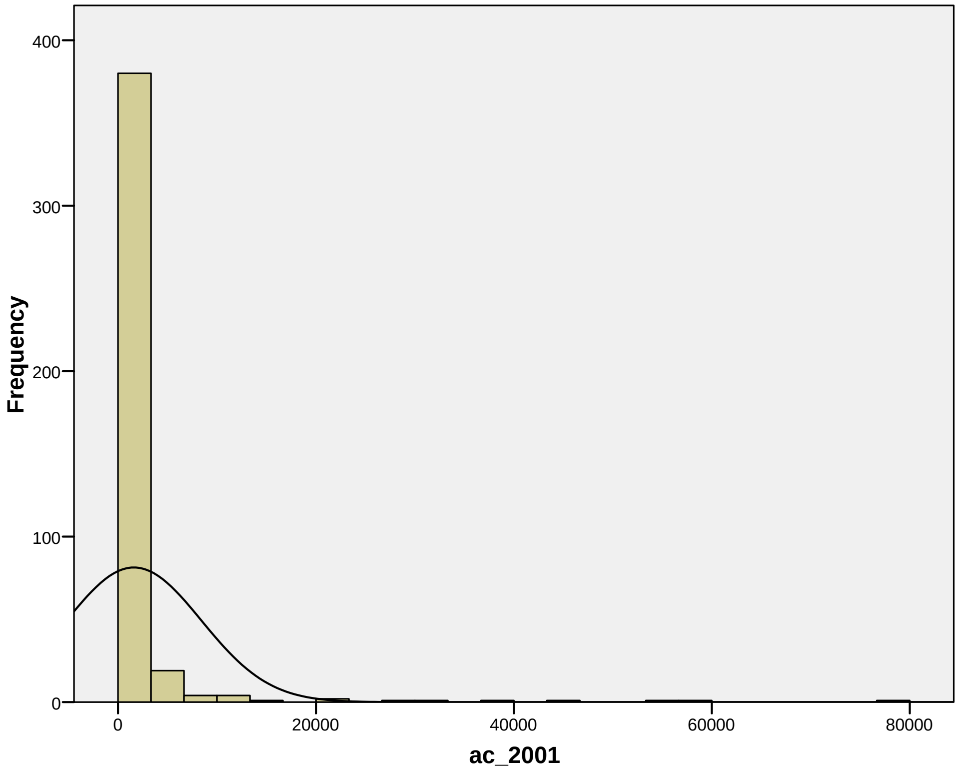
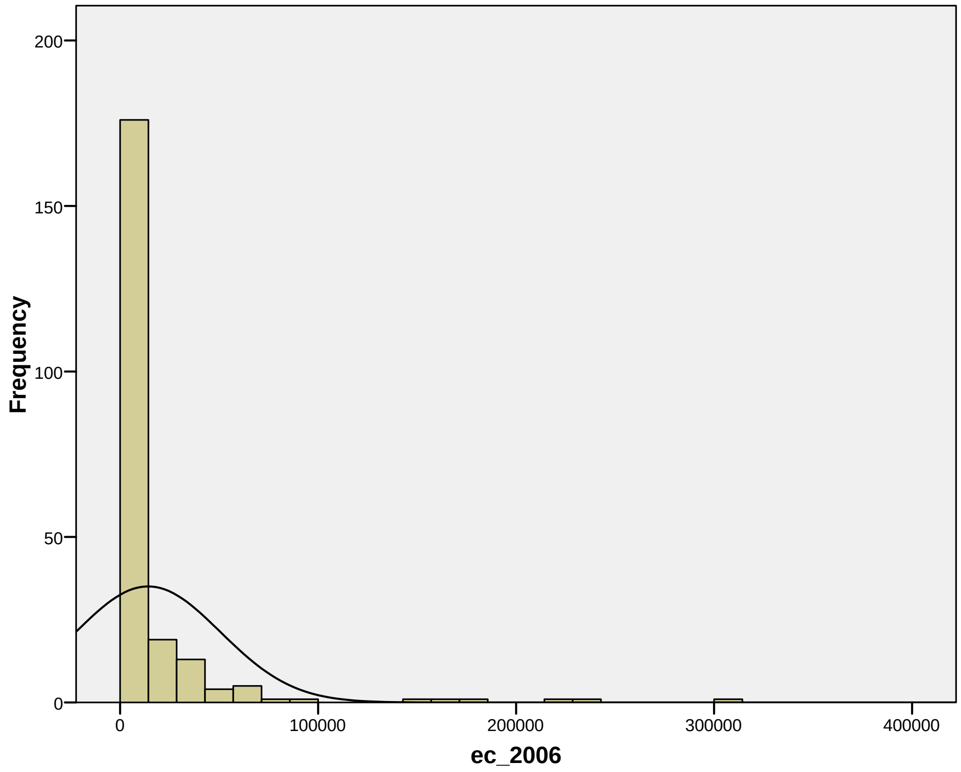
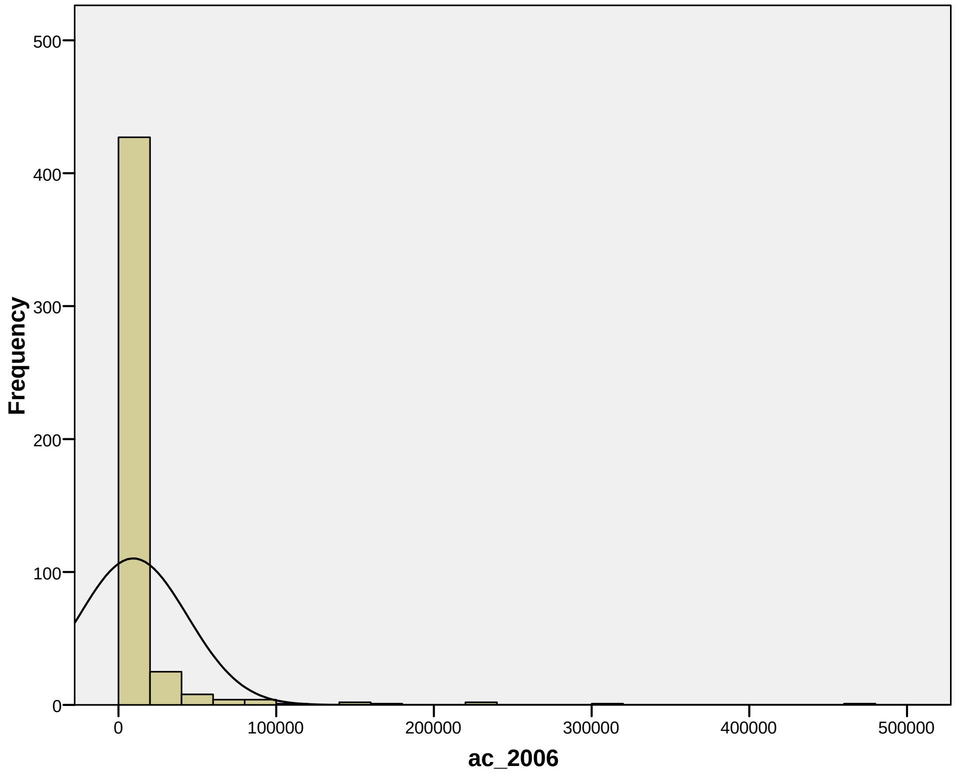


Figure 4.4: Continued





According to Table 4.1, the maximum capacity has been continuously increased over the six year time period resulting in a 5 fold increase. The same applies to the average capacity as well. This increase has a two fold explanation; part of it is because of the overall and geographically even technology change and the newly introduced bandwidth-demanding Internet applications. However, technological change has not equally distributed across space. This in demonstrated by the minimum capacity which has only slightly increased during the six year period. Even in 2005 there were still extra-European links with capacity less than 1 Mbps. The above argument reflects the second element of the backbone link capacity change, which is the localised capacity demand change. And this is the interesting part from the geography point of view since a spatial differentiation emerges. In order to eliminate the geographically even technology change and bring forward the localised capacity change, the capacity classes in all the maps presented here are based on standard deviations above the mean. However, because of the technology change and the overall bandwidth increase, standard deviation between the two different time points cannot be compared in order to draw conclusions for capacity dispersion. In order to overcome this, the coefficient of variation is introduced and presented in Table 4.1. The coefficient of variation is a normalized version of the standard deviation and is defined as the ratio of the standard deviation to the mean (Rogerson 2006). From this statistic it could be said that the capacity of the intra-European international backbone network is more dispersed through time, resulting in more cities being served by relative high capacity links.

Table 4.1: Descriptive statistics for the capacity of backbone links in Mbps

Extent	Year	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
E.c.	2001	0.128	58,641	2,589	7,808	3.016
	2002	2.000	65,041	3,433	8,667	2.525
	2003	2.000	96,870	4,850	12,575	2.593
	2004	2.000	153,529	7,223	19,156	2.652
	2005	2.000	240,952	10,724	27,568	2.571
	2006	2.000	305,169	14,150	36,565	2.584
a.c.	2001	0.064	77,768	1,609	6,813	4.234
	2002	0.190	95,665	2,318	8,218	3.545
	2003	0.256	165,760	3,558	12,969	3.645
	2004	0.260	241,391	5,423	19,496	3.595
	2005	0.256	398,149	7,495	28,281	3.773
	2006	1.500	467,671	9,347	34,453	3.686

E.c. = European cities, a.c. = all cities

Figure 4.5 and 4.6 present the overall network (i.e. it includes the cities out of Europe) are included for 2001 and 2006 respectively. Obviously, the scale of the map is too big in order to focus on our main study area, Europe. Nonetheless, the value added of these maps is the visualisation of Europe’s backbone links with the rest of the world and the differentiation of the capacity of these links. More specifically, for both years a great amount of links below the average capacity mainly occurs with Africa but also with Asia. On the contrary high capacity links connect Europe with cities on the east coast of the USA, such as New York, Boston, Washington DC and Miami. Interestingly enough, the capacity of these links is as high as the capacity of links in Europe’s pentagon, demonstrating the importance of these connections. In addition, backbone links with more remote areas such as Australia and New Zealand as well as with the USA’s west coast are also observed. Regarding the statistics in Table 4.1, the overall network was increased both in terms of links but also in terms of capacity and what was mentioned above about bandwidth demand applies here as well. Also, the coefficient of variation is decreased through time, indicating a more balanced network. The extra-European links are further analysed below.

Table 4.2 presents the basic characteristics of these networks; those are the number of vertices, the number of edges and the network density. The latter refers to the number of edges present in the networks expressed as fraction of the number of all the possible edges. This indicator is also known as γ in *graph theory* and for the case of the non-planar networks is (Taaffe et al 1996, 254):

$$\gamma = \frac{E}{{}^{1}\!/_{2}\,V(V-1)}$$

where E is the number of edges and V the number of vertices.

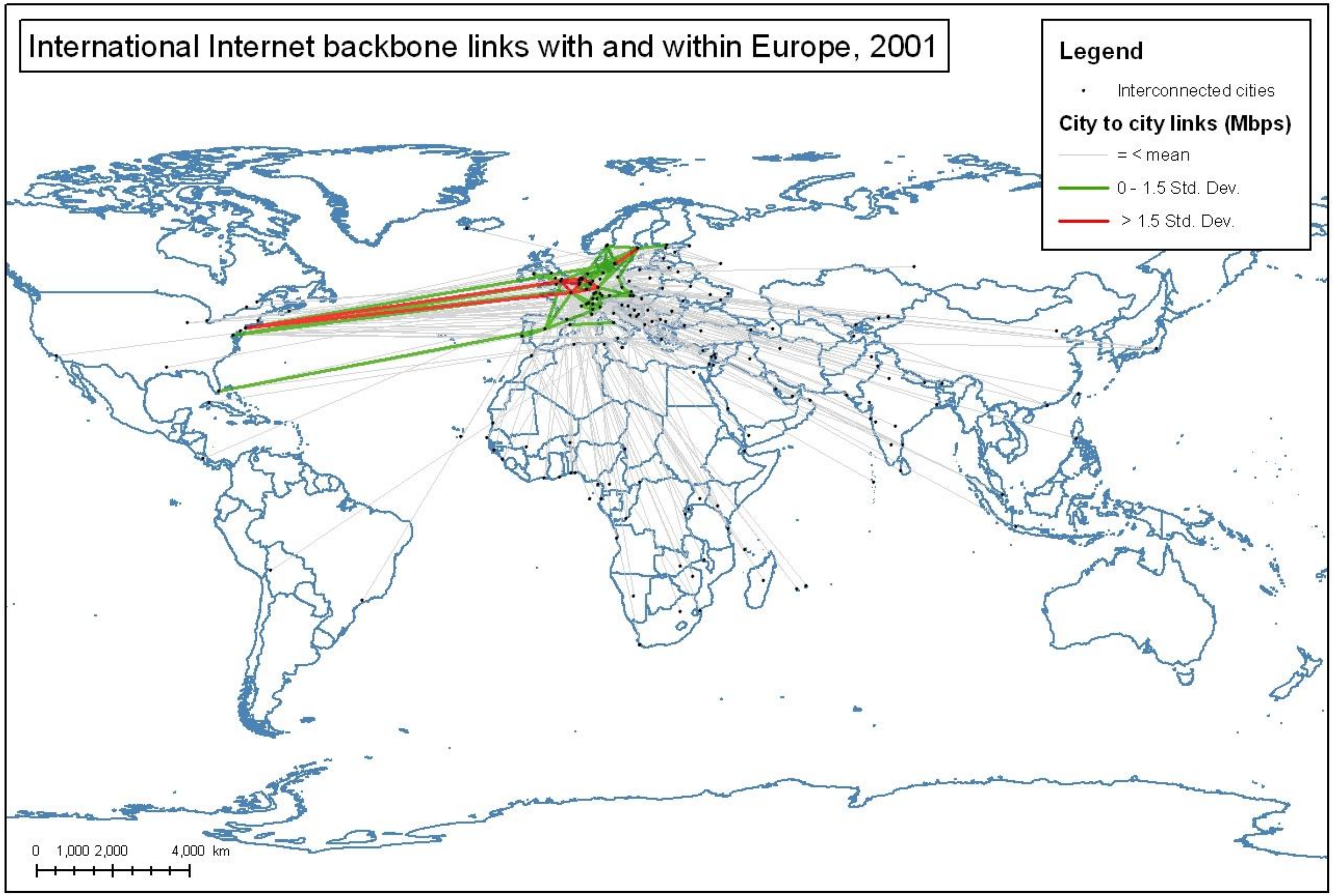


Figure 4.5: International Internet backbone links with and within Europe, 2001

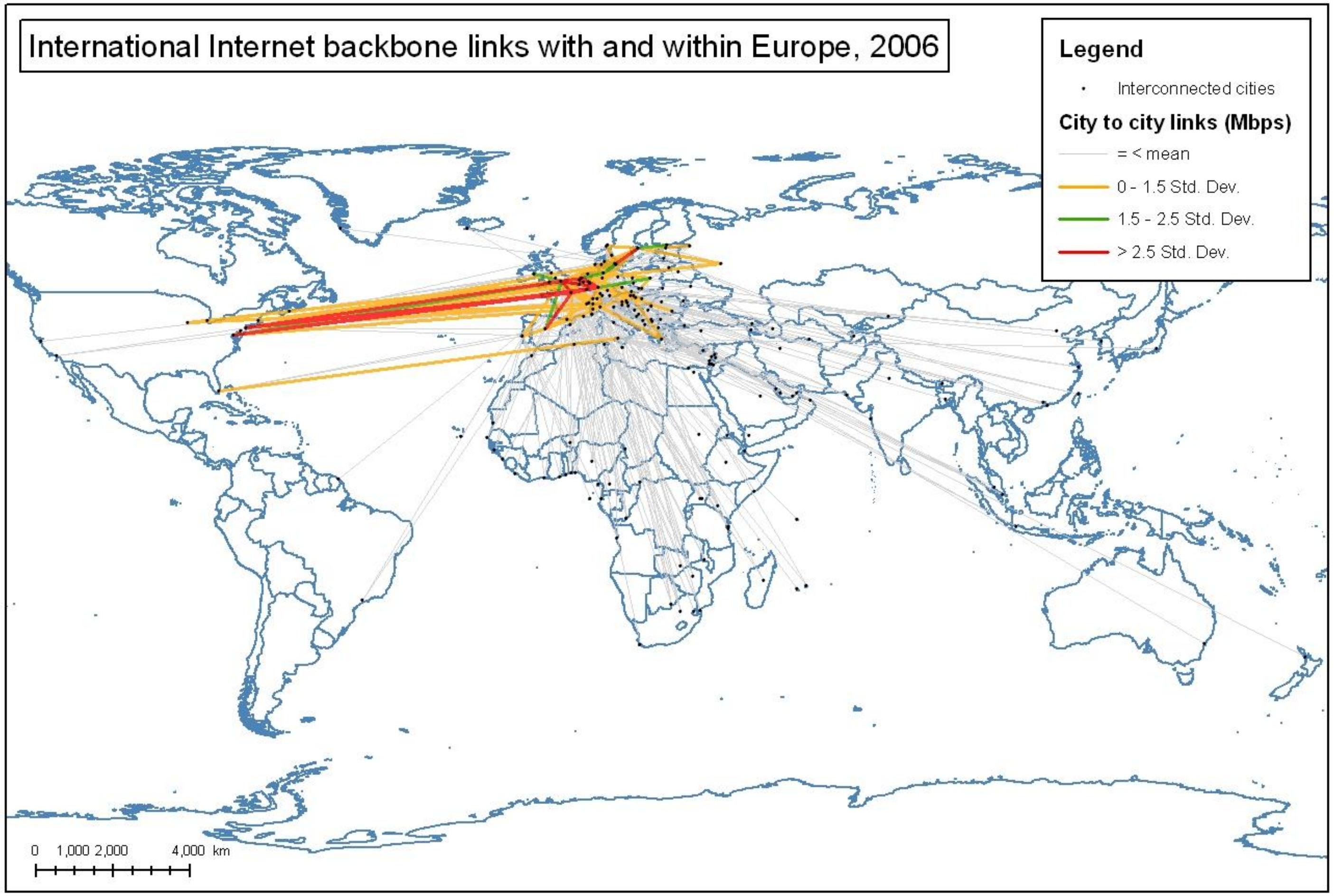


Figure 4.6: International Internet backbone links with and within Europe, 2006

Not surprisingly, the size of the network for all the cities is bigger than the intra-European one in terms of the number of vertices by a factor of 2.3-2.7 for the six years. However, in terms of the number of edges the factor varies from 1.8 to 2.1, indicating a denser intra-European network. Indeed, according to γ the intra-European network is denser than the overall one by a factor of 2.7-3.4 in 2001-2006. This difference in density is not surprising, since the overall network lacks a (significant) proportion of the global links; that is the non-European links (e.g. New York to Tokyo according to the previous example).

In more detail, the change of the main network characteristics through time is also of interest here. Regarding the overall network, an important decrease of 8.4% in the number of vertices and 10.6% in the number of edges took place in 2001-2002, reflecting the *dotcom* crash of the 2000 which was followed by the telecoms crash (Kam 2006). Nonetheless, these changes led to an 8% denser network. This decrease though only concerns the overall backbone network, because the number of intra-

Table 4.2: Basic network statistics

		Vertices		Edges		Density	
			change (%)		*change (%)*		*change (%)*
Internet backbone networks	2001 a.c.	184		417		0.025	
	2002 a.c.	168	*-8.7%*	373	-10.6%	0.027	8.0%
	2003 a.c.	173	*3.0%*	399	7.0%	0.027	0.0%
	2004 a.c.	169	*-2.3%*	399	0.0%	0.028	3.7%
	2005 a.c.	181	*7.1%*	438	9.8%	0.027	-3.6%
	2006 a.c.	194	*7.2%*	476	8.7%	0.025	-7.4%
	annual average change		*1.1%*		2.7%		0.0%
	% point change		*5.4%*		14.1%		0.0%
	2001 E.c.	69		196		0.084	
	2002 E.c.	71	*2.9%*	194	-1.0%	0.078	-7.1%
	2003 E.c.	76	*7.0%*	211	8.8%	0.074	-5.1%
	2004 E.c.	73	*-3.9%*	218	3.3%	0.083	12.2%
	2005 E.c.	72	*-1.4%*	215	-1.4%	0.084	1.2%
	2006 E.c.	76	*5.6%*	225	4.7%	0.079	-6.0%
	annual average change		*2.0%*		2.8%		-1.2%
	% point change		*10.1%*		14.8%		-6.0%

a.c. = all cities, E.c. = European cities
Data source: Telegeography 2007, Author's calculations

European nodes was increased by almost 2.9% in the same time. During the next year, the global backbone network was increased both in terms of interconnected vertices and edges among them, but its density remained steady. The unstable character of the backbone network is reflected by a decrease of 2.3% in the number of vertices in 2003-2004. During the same period the number of edges remained the same and as a result density was increased by 3.7%. After this and for the next two-year periods 2004-05 and 2005-6 a stable increase of 7.1-7.2% and 9.8-8.7% took place in the number of vertices and edges correspondingly. However, the increase in the number of edges was not big enough to prevent a decrease in the network's density. The last periods' network expansion might signal a new era of (stable) increase in the Internet infrastructure development.

Regarding the intra-European network, the downturn periods do not exactly fit with those for the overall network. The intra-European networks kept increasing until 2003, faced a decrease in 2003-04 and 2004-05 and grew again in 2005-06. Interestingly enough, the decrease in the number of intra-European nodes of 2003-4 was accompanied by an increase in the number of edges, resulting in a 12.2% density increase. Similar to the overall network, the increase of 2005-06 in the number of vertices and edges resulted in a less dense network.

All in all, the intra-European network grew more and faster than the overall network during the period 2001-2006. The overall increase in the number of intra-European vertices and edges was 10.1% and 14.8% respectively while the same figures for the overall network were 5.4% and 14.1%. Additionally, the annual average increase for the intra-European network was 2% while for the overall one was 1.1%. These differences in increase rates resulted in an overall global network, the intra-European nodes of which increased from 37.5% in 2001 to 39.2% in 2006. No matter how small that difference seems to be, it reflects a growing participation of the European cities in such a global infrastructural network. Analysis presented later will shed light on the geography of the Internet backbone network in Europe.

4.4 Network centralization measures

The next two sections are dedicated to the centrality and centralization measures. Such measures are always essential to network analysis because they comment both on how centralized the whole network is (centralization), but mainly on how central each node is (centrality). And this is important in order to approach network analysis' main objective, which is the analysis of the ties among the actors. The above are distinguished in a way equivalent to the global and local statistics. Another distinction which is usually made is whether the above indicators take into consideration the edges' weights. Centrality measures can be vastly differentiated when the network is weighted. However, just like network analysis in total, centrality indicators have been mainly developed for binary networks. In this section various centrality and centralization indicators are presented both for the weighted and the binary Internet backbone network for its global extent, but also for its European extraction.

Table 4.3 presents the centralization indicators for the 6 year time period. The upper part of the table refers to the global extent of the network and the lower to the intra-European connections. Four different centralization indicators are presented here, degree, betweenness and eigenvector for the binary (b) and the weighted (w) network. The common characteristic of the above is that they compare the centralization of the current network with the centralization of the most centralized network, which is a network with a star topology (Figure 4.7). In such a network all the nodes are only connected with the central node, the star. And all the above

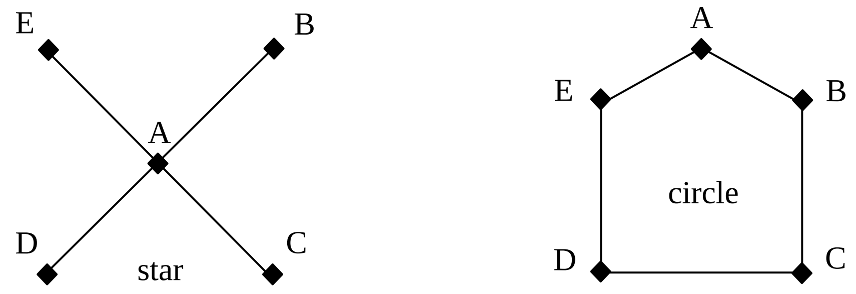
Table 4.3: Centralization indicators

	Degree (b)	Betweenness (b)	Eigenvector (b)	Eigenvector (w)
2001 a.c.	43.4%	46.4%	54.5%	83.6%
2002 a.c.	43.4%	48.1%	55.3%	87.4%
2003 a.c.	47.3%	50.7%	57.4%	89.8%
2004 a.c.	49.0%	46.9%	59.1%	91.3%
2005 a.c.	50.1%	46.1%	59.5%	93.3%
2006 a.c.	50.3%	45.6%	60.1%	89.4%
% point change	*16.1%*	*-1.7%*	*10.2%*	*7.0%*
2001 E.c.	33.8%	21.2%	43.4%	76.0%
2002 E.c.	34.6%	24.9%	46.7%	81.0%
2003 E.c.	36.2%	24.8%	45.5%	79.8%
2004 E.c.	45.7%	27.1%	49.6%	85.6%
2005 E.c.	39.2%	20.8%	47.2%	86.4%
2006 E.c.	41.2%	29.4%	47.2%	84.1%
% point change	*21.9%*	*38.7%*	*8.9%*	*10.6%*

a.c. = all cities, E.c. = European cities

Data source: Telegeography 2007, Author's calculations centralization measures are equal to 1 or (100%). On the contrary, for a circular network centralization indicators are equal to 0.

Figure 4.7: Star and circle networks



The degree centralization, which is the simplest measure, is based on the degree centrality which is nothing more than the sum of all the edges starting or ending to this node. In Figure 4.7, node A of the star-like network has a degree centrality equal to 4 and the rest of the nodes equal to 1. The degree centralization is the "variation in the degrees of vertices divided by the maximum degree variation which is possible in a network of the same size" (Nooy et all 2005, 126). According to Table 4.3, the Internet backbone network seems to be quite centralized, since in 2006 its degree centralization reached half of the maximal centralized network. However, the intra-European subtraction is less centralized. This is not surprising because of the nature of the initial data. As mentioned before, the overall network does not include the links among the extra-European cities. Consequently, this results in a more centralized network with a bunch of peripheral cities (i.e. the cities out of Europe) only connected with the European ones. So, the intra-European network represents reality better. However, even this network enjoys around 40% of the maximum centralization, indicating the structural importance of its main hubs. What is also interesting here is the change of the centralization over time. The centrality of both networks grew over time, but when only the intra-European links are included in the analysis, the centrality of the Internet backbone network grew faster. In simple terms, it could be said that degree centralization is a measure of infrastructural supply because it is based on the actual count of backbone links at city level. All in all, the Internet backbone network is moderate centralized and over time the importance of the Internet backbone hubs in Europe grew faster than their importance at the global scale.

The second measure presented here is the betweenness centralization, which is based on the homonymic measure of centrality. The latter defines the centrality of a vertex as an indication of being between other vertices. This notion of centrality fits better with the technical aspects of the Internet and the structure of the Internet data packet transport system, which is based on packet switching. So, some nodes are central in the Internet because they are in-between numerous origin and destination nodes and the Internet function is based on the efficiency of these hubs to transport data packets. Before defining further the notion of betweenness centrality, the notion of geodesics should be explained first. “A geodesic path is the shortest path, in terms of number of edges traversed, between a specified pair of vertices” (Newman 2008, 5). Of course, the geodesic paths between any two vertices might not be unique. Based on the above, the betweenness centrality of a vertex is defined as the “proportion of all geodesics between pairs of other vertices that include this vertex” (Nooy et al 2005, 131). In the case of the star network the central node has betweenness centrality equal to 6 because 6 geodesics pass by it (D – E, E – B, B – C, C – D, E – C and D – B) while the other nodes have betweenness centrality equal to 1. Just like the degree centralization, the betweenness centralization is defined as the “variation in the betweenness centrality of the vertices divided by the maximum variation in between centrality scores possible in a network of a same size” (Nooy et al 2005, 131).

According to the betweenness centralization the Internet backbone network seems to be less centralized in comparison to the degree centralization measures. This is not coincidental because this measure reflects better the Internet’s function since the Internet was initially designed as a decentralised network (Townsend 2003). The centralization of the overall network was slightly decreased during the six year period contrary to the intra-European one the centralization of which increased almost 39%. So, no matter that the Internet backbone network appeared to be less centralized, it is moving fast towards a more centralized structure in functional terms, at least its European subtraction. In simple terms, according to this measure the Internet backbone network appears less centralized than the infrastructural capital, which is indicated by the degree centralization, and the role of the hub cities in the IP data packets movements is increasing over time in Europe.

Contrary to the above indicators, eigenvector centrality does not consider all the connections being equal: the connections of a node to more central vertices are more important than the connections to less central ones. It is important to have a large number of connections, but a node with fewer but more important connections is more central than a node with more but less important connections (Newman 2008). Eigenvector centrality not only considers the direct links but also the indirect or in other words is more focused on the global structure of the network than the on the local structures, which are the main focus of the degree centrality (Bonarich 2007; Hanneman and Riddle 2005). The above considerations are important for the analysis of the Internet backbone network, the function of which is based on the indirect connections, as they are represented by peering agreements. Eigenvector centrality is widely used with the most well known application being Google's web pages rank (Newman 2008). Eigenvector centrality's calculation is based on factor analysis, which identifies new factors based on the distances among the vertices. Eigenvalue is the location of each node regarding the new factors produced and the collection of such values is called eigenvector. The first factor resulted from factor analysis reflects the global aspects of the distances among vertices while the second and the remaining factors reflect local structures (Hanneman and Riddle 2005). The eigenvector centralization is defined as the variation in the vertices' eigenvector centrality divided by the maximum eigenvector centrality variation which is possible in a network of the same size (Borgatti et al 2002).

The Internet backbone network is becoming more centralized according to eigenvector measures, both for its global extent but also for its European subtraction. For 2006 and for all the cities, the centralization score reached 60%, which is the highest of all the different measures. At the same time, the intra-European Internet backbone network centralization was 47% which is again the highest value. The differences between the degree and eigenvector centralization indicate that the network is more centralized when the indirect links are taken into consideration. Furthermore, the small decrease in centralization in 2005, which is also present in the other centralization measures as well, echoes the decrease in the number of intra-European vertices and edges in the same year, as was illustrated in Table 4.2. Over time, eigenvector centralization is increasing almost equally for both networks. In simple terms, the European hub-cities appear to have a more central role in the

Internet backbone network, which is also increasing over time, when not only the infrastructural supply but also the indirect links are taken into consideration.

The above three centralization indicators presented in Table 4.3 refer only to the binary network, without taking into consideration the actual weights of the connections. For the case of the Internet backbone network the weights of the links represent the bandwidth of the line connecting the two cities. Considering how important is the bandwidth in order to better understand the structure of the Internet backbone network, it is worth trying to include the links' weights in the calculation of the network's centralization. Borgatti et al (2002) in their widely used Social Network Analysis computer program UCINET, enable users to calculate the eigenvector centralization and centrality using a weighted network. The results are presented in Table 4.3. When the bandwidth is included in the analysis, then the network seems to be much more centralized. Especially for the intra European network, eigenvector centralization is more than double the degree centralization and it is very close to the centralization value for the whole network with the global extent. Over time, the centralization is increasing, indicating again an increase in the importance of the European hub cities. In short, when the diffusion of the technology is taken into consideration, that is the roll-out and the exploitation of high capacity fibre links for long-haul backbone Internet connections, then the network is more centralized, resembling better the star like topology.

To sum up, it could be said that the degree centralization provides a measure of how centralized is the distribution of the Internet infrastructure (for this case the Internet backbone connections) at the city level. The other two centralization measures, are more related with the function and the technical nature of the network; betweenness centralization provides a view on how vertices act like hubs and eigenvector comments on centralization based on indirect connections taking also into consideration the weights of the links. All in all, the European subtraction of global Internet backbone network is moderately centralized when the infrastructural supply only is taken into consideration. When the indirect communications are taken into consideration, the network appears to be more centralized. But the main difference emerges when bandwidth in included in the analysis; in this case the highly centralized character of the network is revealed. Lastly, over time and according to almost all the measures, the importance of the hub cities is increasing.

4.5 Cities’ centrality indicators

4.5.1 Degree centrality

What is also interesting apart from the overall network measures is to analyse the local statistics, that is the centrality indicators at the city level. Table 4.4 presents the degree centrality for the 30 most central European cities and for the binary network. The whole table can be found in the Annex. Centralities have been calculated both for the overall network and for the intra-European one for the years 2001 and 2006. For the needs of this table the results have been normalized and for

Table 4.4: Binary degree centrality, 2001-2006

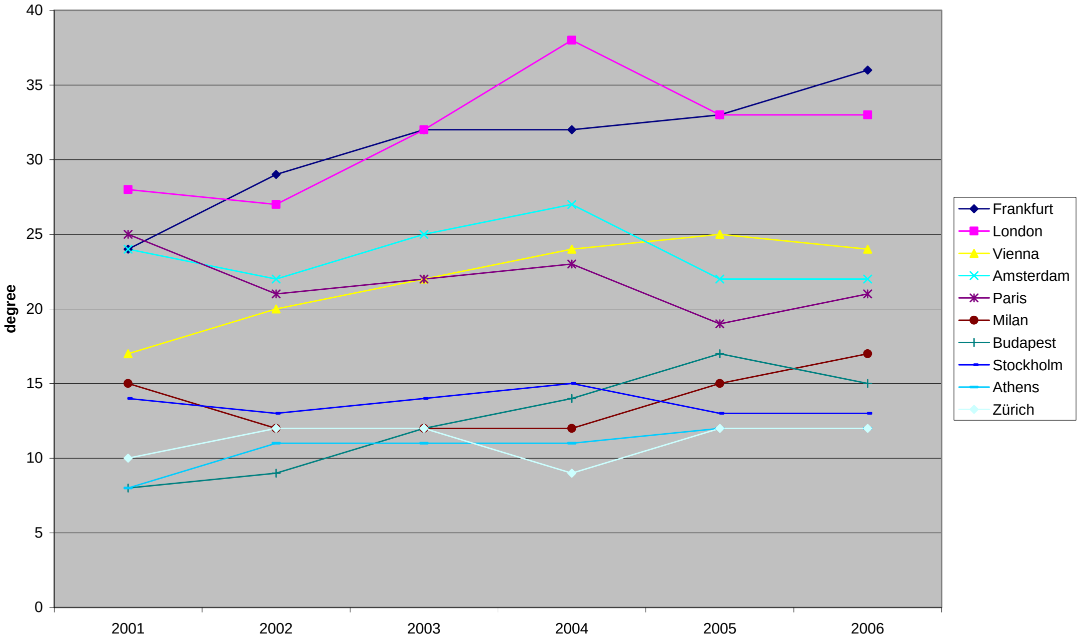
	2001					**2006**					**Change 2001-06 for E.c. (%)**	**Change 2001-06 for a.c. (%)**
	E.c.		**a.c.**		**%Eur. links**	**E.c.**		**a.c.**		**%Eur. links**		
Frankfurt	85.7	*3*	44.6	*3*	*64.9*	100.0	*1*	70.3	*2*	*50.7*	16.7	57.7
London	100.0	*1*	100.0	*1*	*33.7*	91.7	*2*	100.0	*1*	*32.7*	-8.3	0.0
Vienna	60.7	*5*	24.1	*5*	*85.0*	66.7	*3*	27.7	*5*	*85.7*	9.8	15.0
Amsterdam	85.7	*3*	39.8	*4*	*72.7*	61.1	*4*	31.7	*4*	*68.8*	-28.7	-20.3
Paris	89.3	*2*	66.3	*2*	*45.5*	58.3	*5*	48.5	*3*	*42.9*	-34.7	-26.8
Milan	53.6	*6*	24.1	*5*	*75.0*	47.2	*6*	24.8	*6*	*68.0*	-11.9	2.7
Budapest	28.6	*14*	10.8	*20*	*88.9*	41.7	*7*	17.8	*8*	*83.3*	45.8	64.4
Stockholm	50.0	*7*	22.9	*7*	*73.7*	36.1	*8*	18.8	*7*	*68.4*	-27.8	-17.8
Athens	28.6	*14*	12.0	*16*	*80.0*	33.3	*9*	15.8	*10*	*75.0*	16.7	31.5
Zürich	35.7	*11*	13.3	*14*	*90.9*	33.3	*9*	13.9	*14*	*85.7*	-6.7	4.6
Copenhagen	42.9	*8*	16.9	*13*	*85.7*	30.6	*11*	16.8	*9*	*64.7*	-28.7	-0.2
Zagreb	10.7	*34*	4.8	*35*	*75.0*	27.8	*12*	9.9	*17*	*100.0*	159.3	105.4
Hamburg	17.9	*22*	12.0	*16*	*60.0*	25.0	*13*	11.9	*15*	*75.0*	40.0	-1.4
Prague	39.3	*9*	13.3	*14*	*100.0*	25.0	*13*	8.9	*21*	*100.0*	-36.4	-32.8
Brussels	39.3	*9*	21.7	*8*	*61.1*	22.2	*15*	11.9	*15*	*66.7*	-43.4	-45.2
Madrid	28.6	*14*	10.8	*20*	*88.9*	22.2	*15*	14.9	*12*	*53.3*	-22.2	37.0
Warsaw	17.9	*22*	9.6	*24*	*62.5*	22.2	*15*	9.9	*17*	*80.0*	24.4	2.7
Stuttgart	3.6	*56*	1.2	*60*	*100.0*	22.2	*15*	8.9	*21*	*88.9*	522.2	639.6
Geneva	25.0	*18*	10.8	*20*	*77.8*	19.4	*19*	9.9	*17*	*70.0*	-22.2	-8.7
Barcelona	17.9	*22*	6.0	*31*	*100.0*	19.4	*19*	6.9	*24*	*100.0*	8.9	15.0
Tallinn	10.7	*34*	3.6	*40*	*100.0*	19.4	*19*	7.9	*23*	*87.5*	81.5	119.1
Bratislava	21.4	*19*	8.4	*25*	*85.7*	16.7	*22*	5.9	*27*	*100.0*	-22.2	-29.6
Bucharest	14.3	*30*	7.2	*28*	*66.7*	16.7	*22*	6.9	*24*	*85.7*	16.7	-4.1
Düsseldorf	10.7	*34*	3.6	*40*	*100.0*	13.9	*24*	5.9	*27*	*83.3*	29.6	64.4
Lisbon	21.4	*19*	18.1	*11*	*40.0*	13.9	*24*	15.8	*10*	*31.3*	-35.2	-12.3
Oslo	35.7	*11*	21.7	*8*	*55.6*	13.9	*24*	6.9	*24*	*71.4*	-61.1	-68.0
Belgrade	10.7	*34*	4.8	*35*	*75.0*	13.9	*24*	5.0	*31*	*100.0*	29.6	2.7
Nicosia	10.7	*34*	3.6	*40*	*100.0*	13.9	*24*	5.0	*31*	*100.0*	29.6	37.0
Sofia	28.6	*14*	12.0	*16*	*80.0*	13.9	*24*	5.0	*31*	*100.0*	-51.4	-58.9
Ljubljana	17.9	*22*	6.0	*31*	*100.0*	13.9	*24*	5.0	*31*	*100.0*	-22.2	-17.8

E.c. = European cities, a.c. = all cities, * based on normalized centralities
Data sources: Telegeography 2007, Author's calculations

each different case the maximum degree centrality or in other words the maximum number of the backbone connections that a city shares with the rest of Europe or the rest of the world is equal to 100. In addition the ranks for all the cities and for both cases are also presented as well as the overall change through time. The blank cells indicate the absence of any connections for the particular city for this year.

The table is ranked according to the degree centralization of 2006 for the intra-European network. For presentation reasons, the cities' centralities will be analyzed in blocks of tens following the rank of Table 4.4. The non-normalized degree centralities 2001-2006 for the 10 most central cites in 2006 are presented in Figure 4.8. In 2006 Frankfurt is the most centralized city since it shares 36 connections with the other 75 interconnected European cities in 2006. Very close is London with 91.7% of Frankfurt's connections. Interestingly enough, some cities which are located in the periphery of Europe are also found among the first ten. Vienna is the most characteristic case since it shares 24 backbone connections with the rest of Europe cities in 2006, reaching 66.7% of Frankfurt's centrality, which enables it to be the 3rd most central city according to this indicator. Along with Budapest, these two cities seem to play the role of gateway cities for Eastern Europe (Rutherford et al 2004). Apart from them, Athens and Milan represent the south of Europe, Stockholm seems to be the main node for the north and Zurich, with only 12 European connections corresponds to the centre of Europe. Taking into consideration the global connections, the picture is slightly different. The 5 most connected cities change their positions among them. London is the most connected city in 2006 with 101 connections, while Frankfurt falls back to the second position with only 70.3% of London's degree centrality and it is followed by Paris, Amsterdam and Vienna. This big difference between the first and the second indicates London's importance for the global Internet backbone network. The main change in the first tier of the 10 most connected cities is Zurich, which falls to the 14th position and is replaced by Copenhagen when the global connections are taken into consideration. The above differences in centralities can be further explained by the relevant column in Table 4.4, which presents the percentage of the intra-European connections of each city. The global character of London, in particular, but also Paris is highlighted by the fact that only 32.7% and 42.9% of their total connections are towards Europe contrary to the high European orientation of the peripheral gateway cites such as Vienna, Budapest but also Zurich. Somewhere in between the two extremes of the spectrum, the northern and southern Europe's hubs can be found. Looking retrospectively and in terms of the actual degree centrality, most of the cities presented in Figure 4.8 gain in actual connections. It is important though to highlight the fact that Amsterdam, Paris and Stockholm lost connections over time. Looking at the change of the normalized centrality Budapest and Frankfurt gain most in this time period since none of them was part of the first tier of cities in 2001. In particular they replaced Prague and Brussels, which in 2006 are located in the second tier of cities. The binary centrality indicator for 2006 both for the intra- and extra-European backbone network is also presented in Figure 4.9.

Figure 4.8: Degree centrality 2001-2006 for the 10 most central cities in 2006



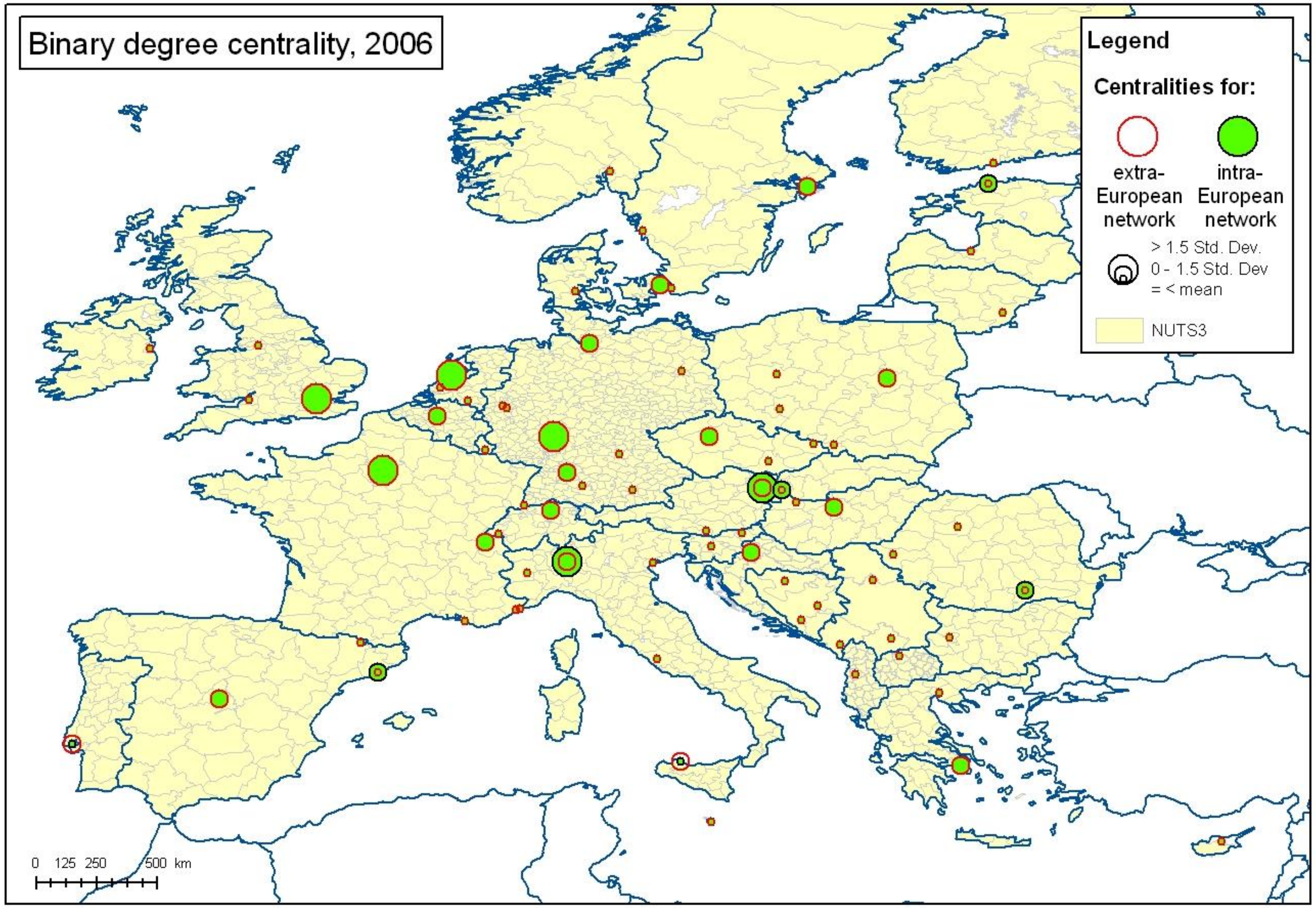
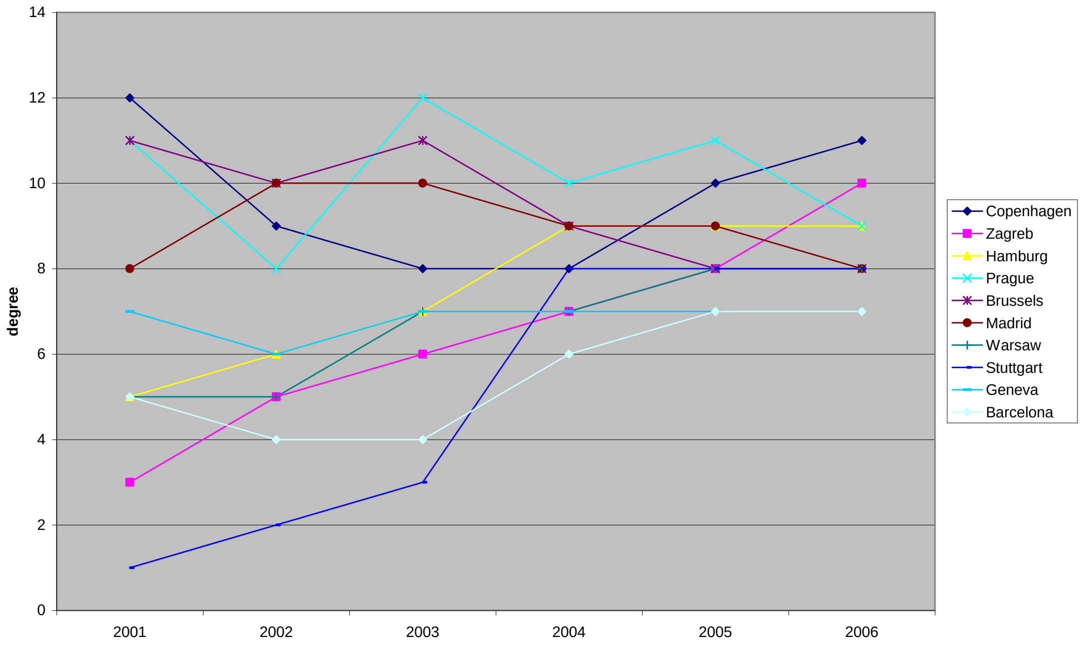


Figure 4.9: Binary degree centrality, 2006

Analyzing further the second tier of cities regarding their degree centralities, both cities of the centre and the periphery of Europe can be found (Figure 4.10). Brussels, no matter its growing importance because of the European Union's integration process and its central location, lost 6 places in connectivity ranking. The majority of its connections are with other European cities, without though missing the valuable global links. On the contrary, Stuttgart's normalized degree centrality increased by 523% for the European links and 640% for the global network in 2001-2006 and these are highest changes for this period. In actual connections, Stuttgart had just 1 in 2001 and grew steadily until 2006, when it reached 8 intra-European and 9 in total links in 2006. Prague seems to be overwhelmed by Vienna as the former performed a role of a gateway city for Eastern Europe in 2001. During the six years period, it lost 36.4% of its degree centrality and its actual European connections decreased from 11 to 9. During the whole time period, its degree centrality was not stable at all, as can be seen in Figure 4.10. It should be also highlighted here that Prague never managed to attract a direct extra-European connection contrary to its competitor in serving Eastern Europe, Vienna, a fact which might be related with the reason why the latter grows in terms of the number of connections while the former does not.

Figure 4.10: Degree centrality 2001-2006 for the 11th-20th most central cities in 2006



The pattern of the way the cities are being served by the international backbone networks can be highlighted by the Spanish example. Both the two most important cities of Spain, Madrid and Barcelona, can be found in this second tier of European cities. Apart from the fact that the capital city is the best connected one and the relation between them seems to be competitive as it is presented in Figure 4.10, Barcelona is not being served both in 2001 and 2006 by any extra-European backbone link, while 47.7% of Madrid’s connections are with non-European cities, reflecting Madrid’s dominant role in Spain’s extra-European IP communications. Comparing the above with the fact that 17 German cities[17] had (at least for one year in the study period) one international backbone connection and the fact that the three most central had a percentage of extra-European connections varying from 50.7%-88.9%, we can identify a similarity between the typology of the urban networks and spatial structure of the international IP backbone networks. In simple terms, the polycentric German urban development pattern and the dominance of Madrid and secondary Barcelona in Spain are reflected by the allocation of the Internet backbone nodes (Rutherford et al 2004).

What is also interesting in this second tier of cities is the position of Zagreb. The Croatian capital experienced also a huge growth in the European degree centrality of 159.3% in 2001-2006 and managed to be the 12th most connected city in terms of binary connections in Europe. Following the above argumentation, Zagreb started playing the role of a gateway city in the Balkan region, with Athens though being the main player, not only because of its greater centrality over time but mainly because of the higher bandwidth accumulation, as we will see later in the weighted degree centrality.

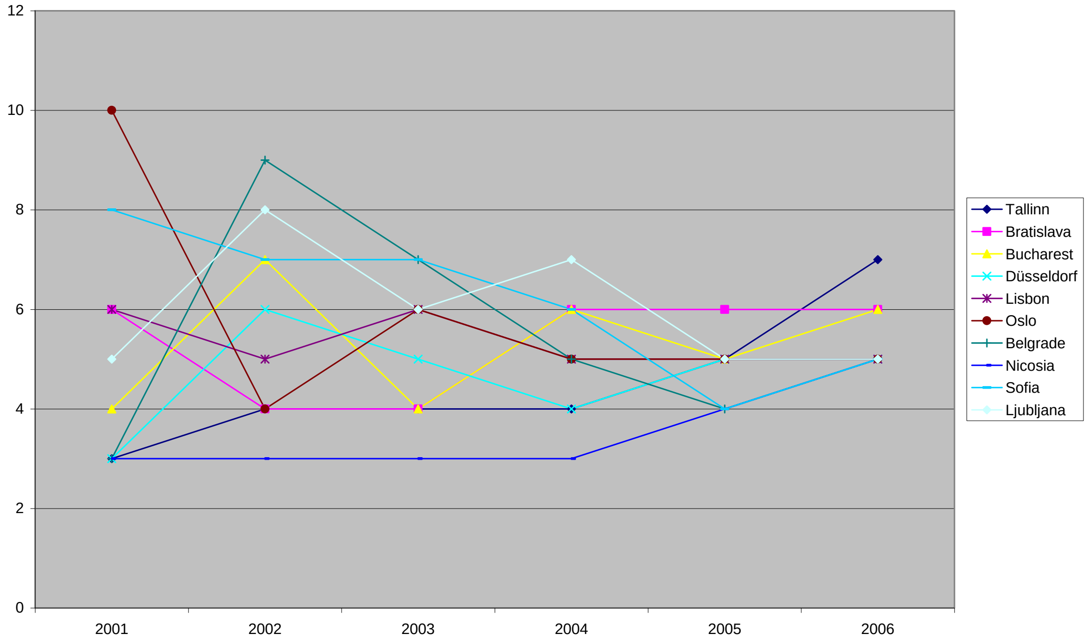
The third tier of cities mainly consisted of peripheral ones. We can identify a cluster of Eastern and South-Eastern European cities such as Bratislava, Bucharest, Belgrade, Sofia, Ljubljana and also Nicosia. Some of them lost in degree centrality measures through time (Bratislava, Sofia and Ljubljana) and some of them gained (Bucharest, Belgrade and Nicosia). Nonetheless, comparing Figure 4.11, which presents the degree centrality 2001-2006 of the third tier of cities according to 2006 ranking, with Figure 4.8 it is obvious how unstable becomes the degree evolution

---

[17] Those are: Frankfurt, Hamburg, Stuttgart, Karlsruhe, Dusseldorf, Munich, Nuremberg, Berlin, Hameln, Hannover, Cologne, Saarbrucken, Krefeld, Dortmund, Dresden, Ehingen and Hilden

through time when the analysis moves away from the most centralized nodes, indicating this way the existence and up to a degree the easiness of the rewiring or the (re)lighting of such networks in order to meet changing demand needs (Gorman and Kulkarni 2004). Apart from Dusseldorf, some of the most remote cities of Europe are also included in this cluster: Oslo and Tallinn represent Europe's northern frontier and Lisbon its western edge. The peripheral location of these cities is probably one of the reasons explaining the high percentage of their extra-European connections, with Lisbon being the most interesting case since in 2006 68.7% of its links were with cities out of Europe. The geography and the rationalization of the European cities' extra-European links are analyzed further below.

Figure 4.11: Degree centrality 2001-2006 for the 21st-30th most central cities in 2006



Regarding the rest of the cities it should be mentioned the very low position of Rome and Berlin, no matter their importance in the European urban and administrative system. They shared the 60th position in 2006 and both of them were served only by 1 international Internet backbone connection. This means that for some reasons illustrated later these cities were not chosen to be the national hubs for the global Internet interconnection and their local demand for IP connections is being served by national links with other cities, which as was explained earlier are not included in the current database.

The degree centrality illustrates the significance of each city taking into consideration only the number of the different cities which it can exchange IP data with directly. What is not included in this indicator is the volume of the data that can be exchanged. And because of the absence of information for volume of the actual IP flows, as it was earlier explained the best proxy in order to approach the importance of each link is the bandwidth. In order to take this into consideration, degree centrality for the weighted network and for the 30 most central cities is presented in Table 4.5 below. The whole table can be found in the Annex.

Table 4.5: Weighted degree centrality, 2001-2006

	2001					**2006**					**Change 2001-06 for**	**Change 2001-06 for**
	E.c.		**a.c.**		**%Eur. links**	**E.c. (%)**		**a.c.**		**%Eur. links**	**E.c. (%)**	**a.c. (%)**
London	96.8	*2*	100.0	*1*	*64.0*	100.0	*1*	100.0	*1*	*63.7*	3.3	0.0
Paris	100.0	*1*	75.6	*2*	*87.4*	84.9	*2*	66.4	*2*	*81.4*	-15.1	-12.1
Frankfurt	96.5	*3*	67.8	*4*	*94.1*	81.0	*3*	58.2	*3*	*88.6*	-16.0	-14.1
Amsterdam	93.2	*4*	71.9	*3*	*85.6*	64.7	*4*	49.6	*4*	*83.0*	-30.6	-31.0
Stockholm	35.7	*6*	24.5	*6*	*96.4*	29.2	*5*	20.2	*5*	*92.3*	-18.2	-17.7
Madrid	12.3	*10*	9.2	*9*	*88.5*	24.5	*6*	16.2	*7*	*96.1*	99.5	77.1
Copenhagen	21.1	*7*	18.4	*7*	*75.9*	24.2	*7*	18.1	*6*	*85.4*	14.9	-1.6
Vienna	11.3	*12*	7.5	*13*	*99.9*	21.4	*8*	13.7	*8*	*99.7*	89.2	82.7
Hamburg	11.0	*14*	7.6	*12*	*95.4*	19.4	*9*	12.7	*10*	*97.3*	75.9	66.3
Milan	18.2	*8*	12.5	*8*	*95.6*	19.0	*10*	12.8	*9*	*94.1*	4.4	2.3
Brussels	48.2	*5*	32.2	*5*	*99.0*	18.5	*11*	11.8	*11*	*100.0*	-61.6	-63.3
Düsseldorf	8.3	*15*	5.5	*15*	*100.0*	11.6	*12*	8.0	*12*	*92.5*	40.7	46.7
Geneva	11.2	*13*	7.5	*14*	*99.1*	10.2	*13*	7.9	*13*	*82.8*	-8.7	5.3
Zürich	12.2	*11*	8.3	*11*	*97.2*	10.2	*14*	6.7	*14*	*96.8*	-16.6	-19.2
Warsaw	1.4	*32*	0.9	*32*	*96.0*	8.7	*15*	5.6	*15*	*99.1*	531.3	489.5
Bratislava	4.4	*20*	2.9	*20*	*99.9*	7.7	*16*	4.9	*16*	*100.0*	77.3	70.8
Prague	7.3	*16*	4.8	*16*	*100.0*	7.4	*17*	4.7	*18*	*100.0*	0.6	-3.0
Helsinki	4.8	*19*	3.3	*18*	*95.4*	7.3	*18*	4.7	*19*	*100.0*	53.6	41.3
Oslo	13.7	*9*	9.1	*10*	*99.0*	6.9	*19*	4.9	*17*	*90.7*	-49.2	-46.5
Dublin	2.5	*24*	1.8	*24*	*92.6*	6.3	*20*	4.2	*20*	*96.4*	152.7	134.1
Budapest	1.9	*27*	1.3	*27*	*98.5*	5.5	*21*	3.5	*21*	*98.9*	191.2	179.8
Munich	6.0	*17*	4.0	*17*	*98.6*	4.4	*22*	2.8	*22*	*100.0*	-26.7	-30.3
Barcelona	2.4	*25*	1.6	*25*	*100.0*	3.9	*23*	2.5	*24*	*100.0*	64.7	58.8
Athens	0.6	*35*	0.5	*35*	*73.2*	3.4	*24*	2.8	*23*	*79.2*	510.9	444.2
Lisbon	2.0	*26*	1.5	*26*	*90.8*	3.3	*25*	2.2	*25*	*96.0*	62.4	48.0
Brno	0.0	*63*	0.0	*64*	*100.0*	2.8	*26*	1.8	*26*	*100.0*	109309.7	105388.1
Tallinn	0.3	*37*	0.2	*38*	*100.0*	1.7	*27*	1.1	*29*	*98.3*	431.2	421.0
Bucharest	0.9	*33*	0.6	*33*	*99.6*	1.7	*28*	1.1	*30*	*99.6*	88.8	82.0
Ljubljana	0.2	*39*	0.1	*40*	*100.0*	1.6	*29*	1.0	*31*	*100.0*	800.1	767.8
Marseille						1.2	*30*	0.8	*32*	*100.0*		

E.c. = European cities, a.c. = all cities
Data sources: Telegeography 2007, Author's calculations

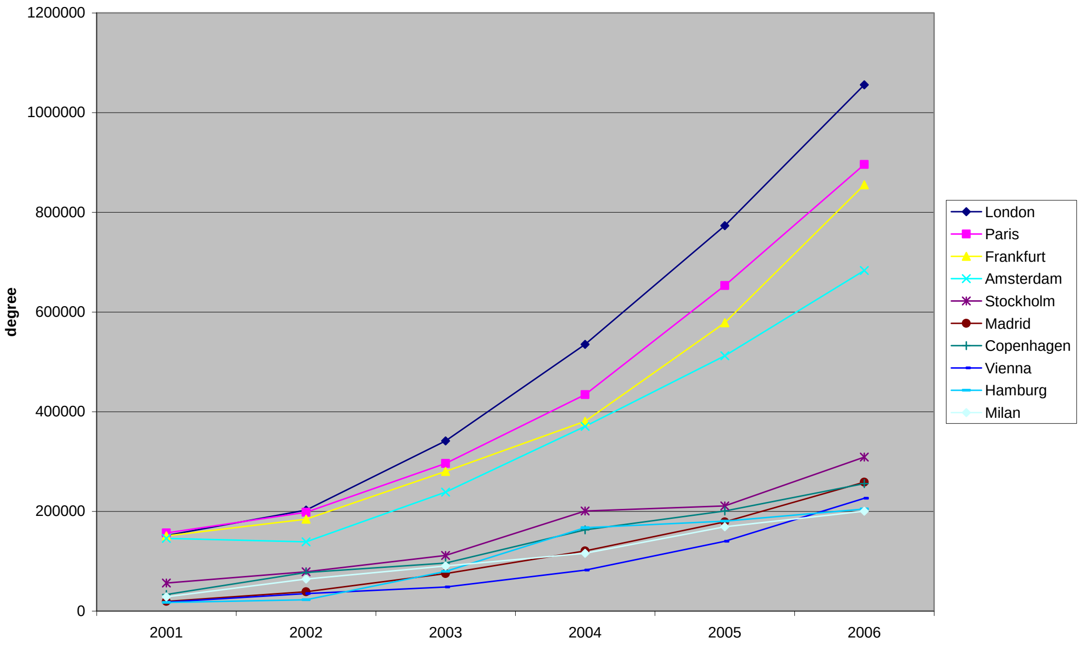
For presentation reasons, just as before, the cities' centralities will be analyzed in blocks of tens following the rank of Table 4.5. Taking into consideration the capacity of the links, the first tier of the most central cities is slightly changed in comparison with the centrality measures for the binary network. In 2006 the four most central cities remain the same with the binary degree centrality, but in this case London instead of Frankfurt is the most central city, followed by Paris, with Amsterdam remaining in fourth position. No matter that Frankfurt has more connections with European cities, London attracted higher capacity links, reflecting the higher demand for exchanging data with London than with Frankfurt and/or the higher demand for using London as an intermediate node for IP transporting. Analyzing further the first tier of central cities, peripheral gateway cites such as Athens and Budapest as well as Zurich appear to be less central when the capacity of the links is taken into consideration. The above cities have been replaced in the first tier by Copenhagen, Hamburg and Madrid. So, if the analysis of the centrality of the first tier of cities based on the binary network brought into sight a Pan-European hub and spoke pattern, then the analysis of the same tier of cities based on the capacity better reflects a more centralized pattern of the bandwidth allocation in Europe.

From a retrospective point of view, in 2001 only two cites were not part of the first tier that is Vienna and Hamburg. These cities experienced significant increase in centrality through the six year period (89.2% and 75.9% respectively), while the increase of the non normalized centrality, or in other words the increase of the aggregated bandwidth at city level, reached 1177% and 1088% correspondingly[18]. Nonetheless, Madrid experienced the higher increase in centrality during the time period (99.5%). Interesting is also the fact that for some of the most central cities, such as Paris, Frankfurt, Amsterdam and Stockholm, a decrease in normalised centrality was observed, but of course not in the absolute bandwidth aggregation. London managed to increase slightly its normalised degree centrality and this is why from 2002 onwards it is the most centralised city. Figure 4.12 presents the change in bandwidth (non-normalised centrality) in 2001-2006. It can be seen that bandwidth allocation has increased exponentially through time. What is also important is the unprompted division of the cities in two clusters: the four most connected ones and the remaining six. The change of each city’s bandwidth is highly related to the cluster it belongs to.

---

[18] It should be highlighted here that the increase in the non-normalised centrality is not independent from the technology improvement and this is the reason why the normalized one is usually used.

Figure 4.12: Degree centrality (w) 2001-2006 for the 10 most central cities in 2006



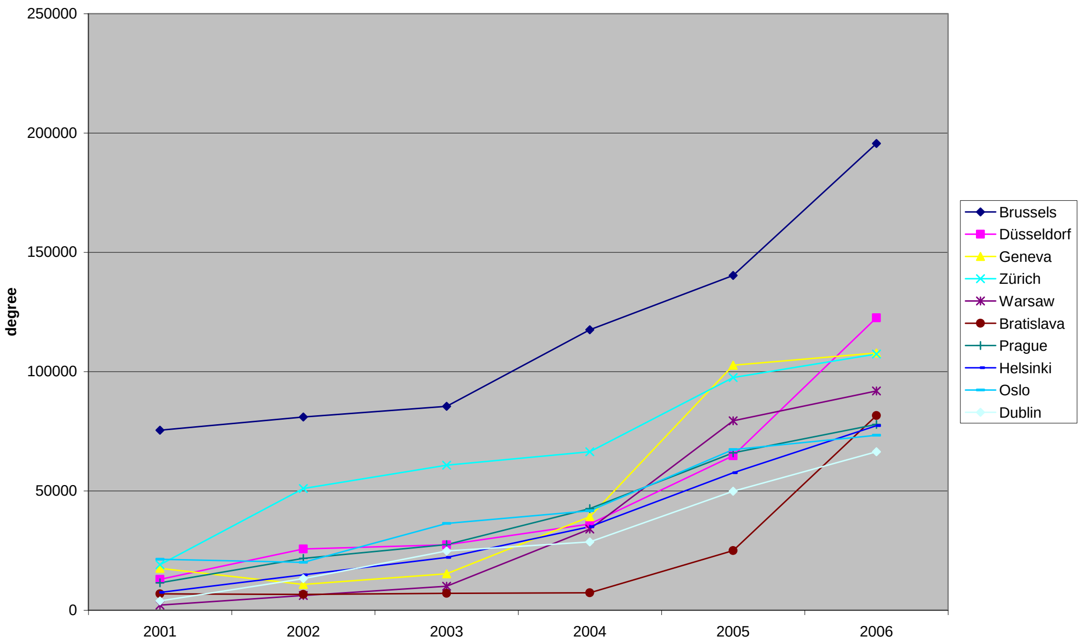
The ranking of the first tier cities according to the weighted network centralities do not change when the extra-European connections are taken into consideration, as is also demonstrated in Figure 4.14. What changes though is London’s importance. Paris, which is the second most central city, has a degree centrality equal to 84.9% of London’s when only the intra-European links are included. However, when the extra-European links are taken into consideration, then Paris reaches only 66.4% of London’s aggregated bandwidth. This difference between London and the second city is greater for the weighted network than the binary one, indicating a greater dominance of London in bandwidth allocation. The above conclusion is in accordance with the percentage of the bandwidth dedicated to European links. 36.3% of London’s aggregated bandwidth serves extra-European connections while only 18.6% of Paris bandwidth is dedicated to such links. Close to Paris is Amsterdam, indicating again its international role, while the rest of the first tier cities use less than 15% of their bandwidth for extra-European communications. Regarding 2001, the main difference was Copenhagen, which at this point appeared to be more globalised, since 24.1% of its aggregated bandwidth was due to extra-European links. During the six year period, Copenhagen’s bandwidth dedicated to intra-European links increased proportionally more than the bandwidth for extra-European.

In the second tier of cities, the differences between the two distinctive geographies emerge; the binary degree centralities geography and the weighted degree centralities geography. Only 4 out of the 10 cities of the second tier are located in the same tier in the two different centrality measures. Dusseldorf and Dublin are the cities which manage to increase their relative position 12 and 15 places respectively while Zurich and Prague lose 4 and 3 places. Dublin’s different performance in the two indicators reflects the fundamental difference between the two centrality measures; while the degree centrality based on the binary network reflects the hub and spoke structure and up to a certain degree the geography of the roll-out of these networks (e.g. the gateway cities), the aggregated bandwidth at city level or in other words the degree centrality of the weighted network seems to be related more with economic attributes. This is why the capital city of the Irish Tiger accumulates in relative terms much more bandwidth than links to other cities.

Going further, Copenhagen used to be part of the second tier of cities when the binary degree centralities were taken into consideration. However, as was stated above, Copenhagen is part of the first tier for this statistic. In addition, two cities from Scandinavia are included in this tier; Helsinki and Oslo, increasing the centrality of the northern cities and gaining 6 and 5 places. Bratislava also gained 6 places in the relative ranking because of the inclusion of bandwidth in the centrality measure. It could be said that Bratislava has a competitive role against Budapest; the former is more central when the actual weights are taken into consideration while the latter is more central for the binary network.

Figure 4.13 presents the evolution of the degree centrality in real terms (bandwidth) in the period 2001-2006. The picture is not as clear as it was in Figure 4.11 for the first tier because the hierarchies change through time. What is obvious though is the fact that Brussels seem to fit better to the first tier than to the second one, because of the high agglomeration of bandwidth. Through time, the European Union’s headquarters, the two main cities of Switzerland and the Norwegian capital lost in centrality relative terms. On the contrary, Warsaw had a massive increase in 2001-2006 of 531% which enabled it to increase its relative position from 32nd position to 15th. Dublin also experienced a great increase of 153% during the same time.

Figure 4.13: Degree centrality (w) 2001-2006 for the 11th-20th most central cities in 2006



Looking to the centrality ranking when the extra-European connections are included in the analysis, the two ranks seem to be almost identical with minor differences between them. However, this was not the case for the centralities of the binary networks because at this case there were significant differences between the overall and the European network. This does not occur at this case because the bulk of the bandwidth which is aggregated in the European cities is dedicated to intra-European connections with a very few exceptions such as London and, up to a degree, Paris and Amsterdam from the first tier. In the second tier, Geneva is the only city in which more than 10% of its aggregated bandwidth is due to extra-European links. It should be noted here that Geneva's global character can also be identified by its binary connections, 30% of which were with cities out of Europe. What is interesting here is the different connectivity profiles Switzerland's two main cities: Zurich is more central in terms of actual connections, while Geneva agglomerates slightly more bandwidth and is the most extroversive.

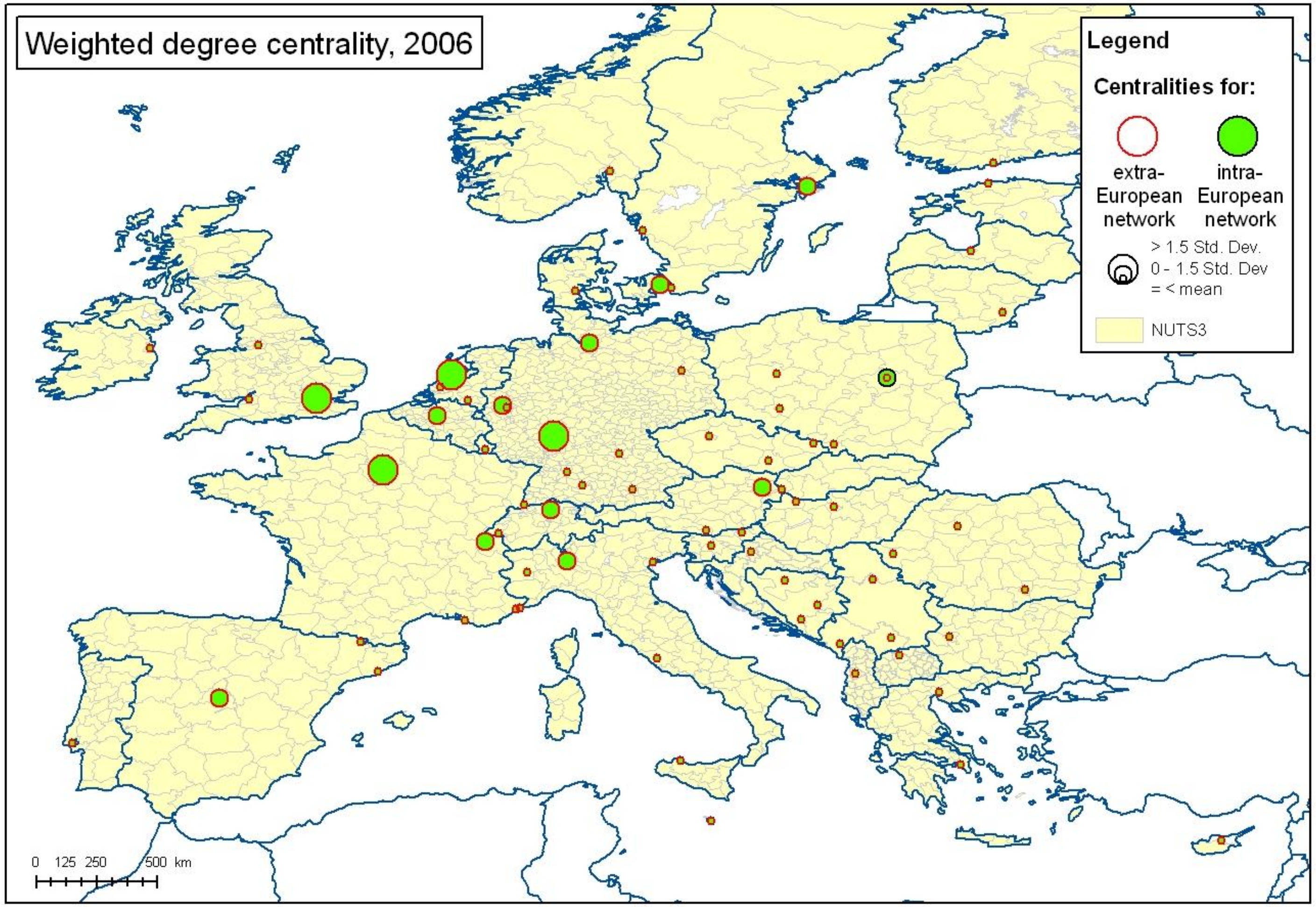
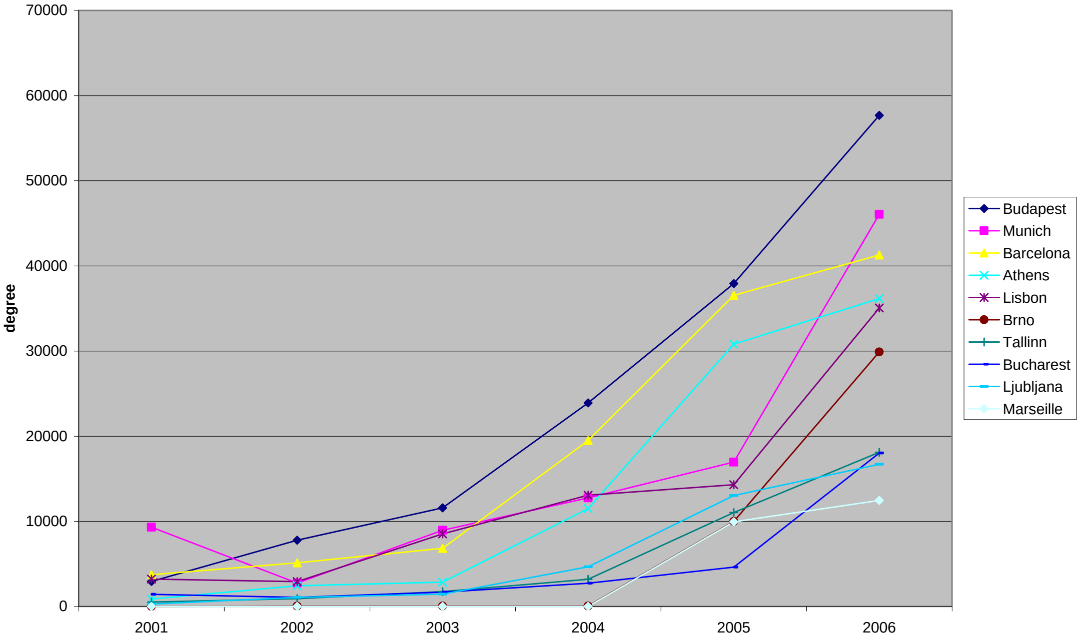


Figure 4.14: Weighted degree centrality, 2006

The third tier of cities, which is presented in Figure 4.15, mainly consists of cities which are located in the periphery of Europe. There is a cluster of cities from Central, Southeast and East Europe such as Budapest, Athens, Brno, Tallinn, Bucharest and Ljubljana and a cluster of cities from the Southwest such as Barcelona, Lisbon and Marseilles. Munich is the only city from Europe's pentagon. The third tier of cities for the case of the binary connections degree centrality also mainly consisted of peripheral cities. However, again only 4 out of 10 cities are the same in the third tier of the of two different centrality measures. Budapest is the most central city of this tier. It is notable though that when the bandwidth is taken into consideration, Budapest gains 14 places in the rank. Athens is also very low ranked comparing the binary connectivity sine it lost 15 places. On the contrary, the Czech city of Brno gained 18 places.

Figure 4.15 presents the evolution of the degree centrality of the third tier of cities for the weighted network. It could be said that it is more homogenous than the previous tier. What is interesting is the steady growth of Budapest through the six years time period, which was followed by Barcelona until 2005. Munich is another interesting case. In 2002 the aggregated bandwidth was decreased, but this decline only lasts for a year. From 2003 onwards, Munich's degree centrality keeps on growing. Brno and Marseille followed very similar growth patterns. The former only had two backbone networks passing through with total bandwidth of 4.048 Mbps and

Figure 4.15: Degree centrality (w) 2001-2006 for the 21st-30th most central cities in 2006



a degree centrality equal to 0.003 of the maximum in 2001 while the latter had no such links. Through time both of them grew similarly and Brno appears to have an enormous increase through the study period because of its very low initial degree, resulting in it gaining 38 places in the ranking. Ljubljana, Athens, and Tallinn also experienced a great increase in the accumulated bandwidth, contrary to Munich which lost in terms of relative centrality.

In terms of the percentage of the bandwidth that is accumulated because of the extra-European links, the only interesting case is Athens. 21.8% of the agglomerated bandwidth in the Greek capital in 2006 was because of the 16 extra-European backbone links that were present at this year, justifying its role as a gateway city. Apart from this, Lisbon is also interesting since in 2001 9.2% of its weighted degree centrality was due to the connections with non-European cities.

However, Lisbon is a very interesting case when the discussion goes to the extra-European links as it is analyzed below. Table 4.6 and 4.7 present the percentage of each city’s degree centrality because of its connections with extra-European cities, based on the binary and weighted (bandwidth) network. The above tables identify the most extroversive cities. However, the definition of extroversion is rather arbitrary. For the case of the binary network and because the greater diffusion of the connections among the cities, cities are defined as extroversive when at least 20% of their backbone connections are with non-European cites. On the contrary, as was mentioned above, the weighted backbone network is much more centralized, or in simpler terms the bulk of bandwidth is concentrated in a small number of cities. Because of the above attribute, cities with more than 5% of their accumulated bandwidth bound to extra-European links are also defined as extroversive. The cities included in the Table 4.6 and 4.7 satisfy at least one of these two attributes.

Two different percentages are presented for each city and for each continent; the columns with the italics font type (numbered with odd numbers) refer to the importance of each continent for each city’s connectivity and the sum of these columns is equal to 100% for each city and for each year. The remaining columns (numbered with even numbers) refer to the city’s importance in Europe’s total connectivity with each continent and the sum of each column is equal to 100%. Both of the tables are ranked according to 2006 degree centrality.

In general, cities perform differently in these two measures of extroversion. London for instance seems to be more globalised according the binary connections than the weighted ones. The reason behind this lies in the diffusion of new technology high capacity links and the demand for telecommunications services and interaction between any two cites. In 2006, apart from the 33% of intra-European links, only 6% of all London’s connections terminate in US and Canadian cities while 31% terminated in Africa and 28% into the Asia and Pacific region. On the contrary, 34% of its accumulated bandwidth in 2006 is because of links which terminate in the USA and Canada, while less than 1% of bandwidth terminates in Africa and 1% terminates with the Asia and Pacific region. This reflects the fact that the links with Africa and Asia are characterized by very low capacity, contrary to very high bandwidth (submarine) links with US and Canadian cities. So, on the one hand there is a demand for a few but high capacity transatlantic links and on the other hand there is a demand for a lot of low capacity links with numerous cities in Africa and the Asia and Pacific region. As a result, there is a high demand for data exchange between London and a few specific US and Canadian cities (the link between London – New York is for every year of the study period the highest capacity link) and a much lower demand between London and several cities in Africa and Asia and Pacific. What is interesting though is to try and explain these different demands. The links between London and North America are more straightforward to explain since they interconnect the two most developed regions of the planet, Europe and North America, and they represent the interaction that takes place between them and up to a point the interaction that takes place between the two global cities of London and New York. In addition, it highlights the role of London as a gateway to North America for the rest of European cities and its great importance in the geography of the Internet. The links with Africa and Asia though are more complicated. Firstly, they represent the global need for universal connectivity. The lack of Internet infrastructure and particularly IXP in Africa make the use of international (and expensive) backbone links essential even for intra-Africa data exchange (ITU 2004). And because of London’s importance in the Internet geography, cities from Africa and Asia are connected to London in order to achieve global and regional Internet connectivity. A second reason though lies in London’s role as a global city. Its predominant position in the world cities hierarchy is a significant pull factor in attracting backbone links from remote cities, underlining the extensive and geographically dispersed demand for data exchange and consequently interaction with one of the main nodes of the globalised informational economy. Last but not least, it could also be observed that London’s backbone links with Africa and the Asia and the Pacific region also reflect Britain’s past Empire.

At the same time, irrespective of the very small share of African and Asian connections to London’s total bandwidth, London remains the city with the greatest capacity towards Africa (Table 4.7, column 14) and Asia and Pacific region (column 16) and of course towards USA (column 23) and the rest of the European cities (column 18). In other words, because of London’s vast accumulated bandwidth, even the routes which represent a very small share of its total degree centrality are enough to provide London the role of the dominant city in the IP communications with other continents. The above highlights once more London’s importance in the geography of the Internet.

Table 4.6: Geographic allocation of backbone connections (%), binary links

	2001											
	Africa		**Asia and Pacific**		**Europe**		**Latin America and Caribbean**		**Rest of Europe**		**USA and Canada**	
	(1)	*(2)*	(3)	*(4)*	(5)	*(6)*	(7)	*(8)*	(9)	*(10)*	(11)	*(12)*
London	22	***30***	36	***45***	34	*7*	0	*0*	5	***15***	4	*6*
Paris	33	***30***	18	***15***	45	*6*	0	*0*	0	*0*	4	*4*
Frankfurt	3	*2*	11	*6*	65	*6*	0	*0*	8	***11***	14	*9*
Amsterdam	9	*5*	9	*4*	73	*6*	0	*0*	3	*4*	6	*4*
Milan	0	*0*	15	*4*	75	*4*	0	*0*	0	*0*	10	*4*
Stockholm	0	*0*	5	*1*	74	*4*	0	*0*	16	***11***	5	*2*
Brussels	17	*5*	0	*0*	61	*3*	0	*0*	6	*4*	17	*6*
Oslo	17	*5*	11	*3*	56	*3*	0	*0*	11	*7*	6	*2*
Palermo	19	*5*	31	*7*	19	*1*	13	***67***	13	*7*	6	*2*
Lisbon	33	*8*	0	*0*	40	*2*	7	***33***	0	*0*	20	*6*
Munich	7	*2*	20	*4*	60	*2*	0	*0*	0	*0*	13	*4*
Copenhagen	0	*0*	0	*0*	86	*3*	0	*0*	7	*4*	7	*2*
Athens	0	*0*	0	*0*	80	*2*	0	*0*	0	*0*	20	*4*
Hamburg	0	*0*	10	*1*	60	*2*	0	*0*	10	*4*	20	*4*
Leuk	50	*8*	30	*4*	10	*0*	0	*0*	0	*0*	10	*2*
Sofia	0	*0*	0	*0*	80	*2*	0	*0*	0	*0*	20	*4*
Madrid	0	*0*	0	*0*	89	*2*	0	*0*	0	*0*	11	*2*
Geneva	0	*0*	0	*0*	78	*2*	0	*0*	0	*0*	22	*4*
Helsinki	0	*0*	0	*0*	67	*2*	0	*0*	22	*7*	11	*2*
Warsaw	0	*0*	0	*0*	63	*1*	0	*0*	13	*4*	25	*4*
Nittedal	0	*0*	0	*0*	57	*1*	0	*0*	43	***11***	0	*0*
Luxembourg	14	*2*	0	*0*	71	*1*	0	*0*	0	*0*	14	*2*
Bucharest	0	*0*	0	*0*	67	*1*	0	*0*	0	*0*	33	*4*
Zagreb	0	*0*	0	*0*	75	*1*	0	*0*	0	*0*	25	*2*
Dublin	0	*0*	0	*0*	75	*1*	0	*0*	0	*0*	25	*2*
Belgrade	0	*0*	0	*0*	75	*1*	0	*0*	0	*0*	25	*2*
Lausanne	0	*0*	0	*0*	75	*1*	0	*0*	0	*0*	25	*2*
Rotterdam	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Tirane	0	*0*	0	*0*	33	*0*	0	*0*	0	*0*	67	*4*
Düsseldorf	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Gdansk	0	*0*	0	*0*	50	*0*	0	*0*	0	*0*	50	*2*
Tartu	0	*0*	0	*0*	50	*0*	0	*0*	50	*4*	0	*0*
Thessaloniki												
Graz												
Hannover												
Total	10	*100*	11	*97*	65	*67*	0	*100*	4	*93*	9	*91*

Table 4.6: (continue)

	2006											
	Africa		**Asia and Pacific**		**Europe**		**Latin America and Caribbean**		**Rest of Europe**		**USA and Canada**	
	(13)	*(14)*	(15)	*(16)*	(17)	*(18)*	(19)	*(20)*	(21)	*(22)*	(23)	*(24)*
London	31	***31***	28	***35***	33	*7*	0	*0*	3	***13***	6	***16***
Paris	37	***18***	14	*9*	43	*5*	2	***33***	0	*0*	4	*5*
Frankfurt	18	***13***	20	***18***	51	*8*	0	*0*	7	***22***	4	*8*
Amsterdam	3	*1*	9	*4*	69	*5*	0	*0*	6	*9*	13	***11***
Milan	12	*3*	12	*4*	68	*4*	0	*0*	0	*0*	8	*5*
Stockholm	5	*1*	16	*4*	68	*3*	0	*0*	11	*9*	0	*0*
Brussels	17	*2*	8	*1*	67	*2*	0	*0*	0	*0*	8	*3*
Oslo	14	*1*	0	*0*	71	*1*	0	*0*	0	*0*	14	*3*
Palermo	20	*3*	53	***10***	20	*1*	0	*0*	0	*0*	7	*3*
Lisbon	56	*9*	0	*0*	31	*1*	6	***33***	0	*0*	6	*3*
Munich	17	*1*	0	*0*	83	*1*	0	*0*	0	*0*	0	*0*
Copenhagen	6	*1*	0	*0*	65	*2*	0	*0*	18	***13***	12	*5*
Athens	0	*0*	19	*4*	75	*3*	0	*0*	0	*0*	6	*3*
Hamburg	0	*0*	8	*1*	75	*2*	0	*0*	0	*0*	17	*5*
Leuk	50	*1*	50	*1*	0	*0*	0	*0*	0	*0*	0	*0*
Sofia	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Madrid	20	*3*	7	*1*	53	*2*	7	***33***	0	*0*	13	*5*
Geneva	0	*0*	0	*0*	70	*2*	0	*0*	0	*0*	30	*8*
Helsinki	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Warsaw	0	*0*	0	*0*	80	*2*	0	*0*	20	*9*	0	*0*
Nittedal	100	***10***	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Luxembourg	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Bucharest	0	*0*	0	*0*	86	*1*	0	*0*	14	*4*	0	*0*
Zagreb	0	*0*	0	*0*	100	*2*	0	*0*	0	*0*	0	*0*
Dublin	0	*0*	0	*0*	80	*1*	0	*0*	0	*0*	20	*3*
Belgrade	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Lausanne	0	*0*	0	*0*	75	*1*	0	*0*	0	*0*	25	*3*
Rotterdam	0	*0*	0	*0*	50	*0*	0	*0*	0	*0*	50	*3*
Tirane	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Düsseldorf	0	*0*	0	*0*	83	*1*	0	*0*	0	*0*	17	*3*
Gdansk												
Tartu												
Thessaloniki	0	*0*	25	*1*	75	*1*	0	*0*	0	*0*	0	*0*
Graz	0	*0*	0	*0*	0	*0*	0	*0*	100	*4*	0	*0*
Hannover	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	100	*3*
Total	14	*98*	11	*94*	65	*62*	0	*100*	3	*83*	5	*95*

(1) + (3) + (5) + (7) + (9) + (11) = (13) + (15) + (17) + (19) + (21) + (23) = 100%

The cities included here fulfil at least one of the following criteria: % of extra-European links > = 20%, % of total bandwidth because of extra-European links >= 5%

Data sources: Telegeography 2007, Author's calculations

Table 4.7: Geographic allocation of backbone connections (%), bandwidth

2001

	Africa		**Asia and Pacific**		**Europe**		**Latin America and Caribbean**		**Rest of Europe**		**USA and Canada**	
	(1)	*(2)*	(3)	*(4)*	(5)	*(6)*	(7)	*(8)*	(9)	*(10)*	(11)	*(12)*
London	0	*18*	0	*54*	64	*15*	0	*0*	0	*10*	36	*53*
Paris	0	*48*	0	*29*	87	*15*	0	*0*	0	*0*	12	*14*
Frankfurt	0	*0*	0	*3*	94	*15*	0	*0*	0	*13*	6	*6*
Amsterdam	0	*3*	0	*3*	86	*14*	0	*0*	0	*0*	14	*15*
Milan	0	*0*	0	*0*	96	*3*	0	*0*	0	*0*	4	*1*
Stockholm	0	*0*	0	*0*	96	*6*	0	*0*	1	*52*	2	*1*
Brussels	0	*1*	0	*0*	99	*7*	0	*0*	0	*1*	1	*0*
Oslo	0	*1*	0	*3*	99	*2*	0	*0*	0	*1*	1	*0*
Palermo	13	*25*	4	*3*	9	*0*	2	*67*	1	*0*	71	*0*
Lisbon	0	*1*	0	*0*	91	*0*	0	*33*	0	*0*	9	*0*
Munich	0	*0*	0	*3*	99	*1*	0	*0*	0	*0*	1	*0*
Copenhagen	0	*0*	0	*0*	76	*3*	0	*0*	0	*3*	24	*6*
Athens	0	*0*	0	*0*	73	*0*	0	*0*	0	*0*	27	*0*
Hamburg	0	*0*	0	*0*	95	*2*	0	*0*	0	*0*	5	*1*
Leuk	11	*2*	4	*0*	42	*0*	0	*0*	0	*0*	42	*0*
Sofia	0	*0*	0	*0*	94	*0*	0	*0*	0	*0*	6	*0*
Madrid	0	*0*	0	*0*	89	*2*	0	*0*	0	*0*	11	*2*
Geneva	0	*0*	0	*0*	99	*2*	0	*0*	0	*0*	1	*0*
Helsinki	0	*0*	0	*0*	95	*1*	0	*0*	3	*15*	2	*0*
Warsaw	0	*0*	0	*0*	96	*0*	0	*0*	0	*0*	4	*0*
Nittedal	0	*0*	0	*0*	48	*0*	0	*0*	52	*2*	0	*0*
Luxembourg	0	*0*	0	*0*	96	*0*	0	*0*	0	*0*	3	*0*
Bucharest	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Zagreb	0	*0*	0	*0*	26	*0*	0	*0*	0	*0*	74	*0*
Dublin	0	*0*	0	*0*	93	*0*	0	*0*	0	*0*	7	*0*
Belgrade	0	*0*	0	*0*	54	*0*	0	*0*	0	*0*	46	*0*
Lausanne	0	*0*	0	*0*	75	*0*	0	*0*	0	*0*	25	*0*
Rotterdam	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Tirane	0	*0*	0	*0*	40	*0*	0	*0*	0	*0*	60	*0*
Düsseldorf	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Gdansk	0	*0*	0	*0*	50	*0*	0	*0*	0	*0*	50	*0*
Tartu	0	*0*	0	*0*	67	*0*	0	*0*	33	*0*	0	*0*
Thessaloniki	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Graz	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Hannover	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Total	0	*100*	0	*99*	86	*91*	0	*100*	0	*98*	14	*100*

Table 4.7: (continue)

	2006											
	Africa		**Asia and Pacific**		**Europe**		**Latin America and Caribbean**		**Rest of Europe**		**USA and Canada**	
	(13)	*(14)*	(15)	*(16)*	(17)	*(18)*	(19)	*(20)*	(21)	*(22)*	(23)	*(24)*
London	0	*41*	1	*41*	64	*17*	0	*0*	1	*24*	34	*49*
Paris	0	*22*	0	*9*	81	*14*	0	*1*	0	*0*	18	*17*
Frankfurt	0	*1*	1	*12*	89	*13*	0	*0*	1	*17*	10	*8*
Amsterdam	0	*2*	0	*4*	83	*11*	0	*0*	0	*1*	17	*12*
Milan	0	*1*	1	*3*	94	*3*	0	*0*	0	*0*	5	*1*
Stockholm	0	*0*	0	*0*	92	*5*	0	*0*	8	*51*	0	*0*
Brussels	0	*0*	0	*0*	100	*3*	0	*0*	0	*0*	0	*0*
Oslo	0	*0*	0	*0*	91	*1*	0	*0*	0	*0*	9	*1*
Palermo	14	*19*	43	*18*	4	*0*	0	*0*	0	*0*	39	*1*
Lisbon	1	*2*	0	*0*	96	*1*	0	*1*	0	*0*	3	*0*
Munich	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Copenhagen	0	*0*	0	*0*	85	*4*	0	*0*	0	*2*	14	*4*
Athens	0	*0*	15	*12*	79	*1*	0	*0*	0	*0*	5	*0*
Hamburg	0	*0*	0	*0*	97	*3*	0	*0*	0	*0*	3	*0*
Leuk	17	*0*	83	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Sofia	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Madrid	1	*11*	0	*0*	96	*4*	0	*98*	0	*0*	3	*1*
Geneva	0	*0*	0	*0*	83	*2*	0	*0*	0	*0*	17	*2*
Helsinki	0	*0*	0	*0*	100	*1*	0	*0*	0	*0*	0	*0*
Warsaw	0	*0*	0	*0*	99	*1*	0	*0*	1	*2*	0	*0*
Nittedal	100	*1*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Luxembourg	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Bucharest	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Zagreb	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Dublin	0	*0*	0	*0*	96	*1*	0	*0*	0	*0*	4	*0*
Belgrade	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Lausanne	0	*0*	0	*0*	75	*0*	0	*0*	0	*0*	25	*0*
Rotterdam	0	*0*	0	*0*	50	*0*	0	*0*	0	*0*	50	*0*
Tirane	0	*0*	0	*0*	100	*0*	0	*0*	0	*0*	0	*0*
Düsseldorf	0	*0*	0	*0*	92	*2*	0	*0*	0	*0*	8	*1*
Gdansk	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Tartu	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*
Thessaloniki	0	*0*	7	*0*	93	*0*	0	*0*	0	*0*	0	*0*
Graz	0	*0*	0	*0*	0	*0*	0	*0*	100	*1*	0	*0*
Hannover	0	*0*	0	*0*	0	*0*	0	*0*	0	*0*	100	*2*
Total	0	*100*	1	*99*	83	*88*	0	*100*	1	*99*	15	*100*

(1) + (3) + (5) + (7) + (9) + (11) = (13) + (15) + (17) + (19) + (21) + (23) = 100%
The cities included here fulfil at least one of the following creteria: % of extra-European links > = 20%, % of total bandwidth because of extra-European links >= 5%
Data sources: Telegeography 2007, Author's calculations

However, its dominance in Africa’s Internet connectivity was not constant through time. In 2001 and in terms of binary links, both London and Paris had a share of 30% each in Europe’s total links with Africa (Column 2 in table 4.6). What is interesting though is that Paris’ links with Africa were at higher capacity than London’s and as a result Paris was the dominant city in connecting with Africa in terms of capacity, since it represented 48% of Europe’s total capacity towards Africa (column 2 in Table 4.7). The above can be interpreted as an antagonism over time between London and Paris for the dominant city in Internet connectivity rank and also as a competition for being the gateway city for the communications with other continents such as Africa; or, from the ISP perspective, as a competition for selling universal (and regional) connectivity to a continent with strong colonial relations both with London and Paris. Through time though, London seems to gain the role as a gateway city both for Africa and for the Asia and Pacific region.

In general, Paris’ extra-European links seem to have a similar spatial structure to London, but because of their lower capacity, Paris’s importance at European level is lower than London’s. Apart from its links with Africa, 14% of its binary degree centrality is due to connections with Asia and the Pacific region, reflecting again the demand for extensive – in geographical range – but not intensive – in terms of bandwidth – post colonial informational links. It should be noted here that 18% of the accumulated bandwidth in Paris is bound for communications with USA and Canada, almost half of London’s percentage, indicating again London’s dominance in Europe’s Internet geography.

Regarding the other two highly central cities, Frankfurt and Amsterdam, their global character is mainly reflected in the binary degree centrality. In general, Frankfurt has a more balanced division of its extra-European links and Amsterdam is more tied with USA and Canada. A significant share of Frankfurt’s binary degree connectivity is because of its links with Asia and Pacific and Africa. In addition, Frankfurt is the dominant gateway city in terms of binary backbone links with the rest of Europe (17% of Europe’s total links with the rest of Europe region), which includes countries such as Russia, Ukraine, Belarus and also Iceland and Greenland. However, Frankfurt’s links towards the cities of the rest of Europe are rather low capacity, since Frankfurt is only responsible for 17% of the Europe’s total bandwidth towards this region. The dominant city in terms of bandwidth bound for Internet links with the rest of the Europe is Stockholm, which carries 51% of Europe's total bandwidth towards the cities of the rest of Europe. In more details, the above high capacity backbone links, which enable Stockholm to have such a dominant role in 2006 are towards Moscow and St. Petersburg while Frankfurt's links are also with Moscow and St. Petersburg and in addition with Kiev, Lvov and Minsk. Looking retrospectively, Frankfurt also managed to increase its importance over time in providing connectivity to Africa and Asia and Pacific region. Amsterdam on the other hand is more focused to European Internet backbone links. The main share of Amsterdam's links with non-European cities is cities from North America. More specifically, 13% and 17% of its degree centrality (binary and weighted) are due to US and Canadian links. The above result in Amsterdam controlling of 11% and 12% of Europe's binary links and bandwidth respectively to North America.

Going further down in Table 4.6, apart from Stockholm, Copenhagen and still further below Warsaw and Bucharest are also interesting cases because of their large share of links with the rest of Europe. What is also interesting is that there seems to be complementarity between them; while Stockholm decreased its share of connections with these countries, Copenhagen increased it through time. However, its links are not of high capacity as it can be seen in Table 4.7. Interesting enough though, for the case of Warsaw and Bucharest, the extra-European links refer only to cities in the rest of Europe. The above highlights the fact that the distinction between Europe and rest of Europe is practical and made only for assisting the current analysis. The geographic continuity between the cities of Central and Eastern Europe explains why the above cities, which do not have other extra-European links, are well connected with the cities of what is defined here as 'rest of Europe'.

Brussels is also a unique case. It is the only city which can be found at such a high place in both centrality measures but in which almost 100% of its already accumulated bandwidth is due to intra-European links. More specifically, 33% of Brussels binary degree centrality is because of its backbone links with non-European cities, these links are of low capacity and this is the reason why they result in an insignificant share of external links in terms of bandwidth. It could be said that the above reflects the city's unique role: hosting the European Union's headquarters makes it important, but its importance is almost exclusively intra-European.

The next interesting case is Lisbon. 56% of its total binary connectivity is towards Africa and 6% towards Latin America. The above shares are equal to 9% and 33% of Europe's total connectivity with Africa and Latin America and Caribbean region which enables Lisbon to be the 5th dominant city for African connection and to share the 1st place for Latin America and region. However, in Europe in total there are only 3 links with the latter region, one of which terminates in Lisbon in 2001. This is not surprisingly though, because Latin America and Caribbean Region mainly gain universal IP connectivity through the USA and more specifically through their gateway city, Miami (Garcia 2000, Grubesic and O'Kelly 2002). And this is obviously the reason why Lisbon (along with Madrid, Milan and Palermo) has a direct link with Miami.

The capacity of the links with Africa and Latin America though is very low as can be seen in Table 4.7. This reflects both the low demand for Internet services in these regions, but at the same time the need for universal interconnectivity. In addition, the importance of geography in submarine backbone networks roll-out is also reflected in these links. The location of Lisbon facing the Atlantic is a competitive advantage for setting up backbone fibre links with Africa because of the way these links are rolled-out; the fibre is dug by a special ship, which follows a route near the coast of Portugal and Africa for this case. No matter that Lisbon is a very suitable location for setting up such links, it cannot be claimed that this is the main reason why Lisbon is highly connected with Africa. Lisbon is also a suitable location for setting up link with USA, but its share of connectivity with USA is very low (6%), highlighting the multidimensional interpretation of the backbone Internet geography. The main reason behind the extensive links with Africa is Portugal's colonial past, which is reflected once more in the geography of the backbone networks.

A similar case to Lisbon is that of Madrid. Almost half of the Spanish capital's connections in 2006 were with non-European cities. More specifically 20% of its binary degree centrality was due to links with Africa and 7% with Asia and Pacific and 7% with Latin America, having one of the three backbone links that Europe shares with this continent. In terms of capacity though, Madrid is the main nodal point for links with Latin America and the Caribbean since 98% of Europe's dedicated bandwidth for this continent passes through Madrid. However, this is a recent development because in 2001 Madrid had only links with other European cities and with North America. Madrid is also an important node for communications with Africa; 11% of Europe's bandwidth to this continent comes through the Spanish capital. Once again, a multilevel interpretation of the geography of the Internet backbone networks can be observed; Madrid's extra-European backbone links seem to be defined both by the city's location in the south-west tip of Europe but also by its historical relations with Latin America.

Palermo is another interesting case: 80% of its binary degree connectivity results from backbone links which terminate in cities outside of Europe. More specifically, in 2006 20% of the city's 13 links were towards Africa and 53% towards Asia and Pacific Region. Even more interesting is its share of links with Latin America and Caribbean region in 2001, which reached 13% and enabled it to be the most important European node for backbone connections with Latin America. Its global character is even more obvious when the discussion goes to the weighted degree centrality. In 2006 14% of its accumulated bandwidth was because of its links with African cities, 43% because of the Asia and Pacific backbone networks and only 4% because of the intra-European links. From the dominance point of view, 14% of Europe's bandwidth to Africa and 18% towards Asia and Pacific region pass through Palermo. In addition in 2001 67% of Europe's capacity towards Latin America and Caribbean was passing through Palermo. In addition, 39% of its total bandwidth terminates in Northern America. All in all, Palermo is an Internet node of global importance. The reason for this lies on its geographical location. It is located in Sicily, in the middle of the Mediterranean Sea, which was always an area with great trade activity between Europe, Middle East and Africa. It seems that nowadays the trade activity in the area not only takes part on the sea's surface but also on the sea-floor; the trade past of the Mediterranean Sea is replicated in the numerous fibre optic cables of the Internet backbone networks, which are laid on the floor of the Mediterranean Sea and Palermo is a nodal point for these corridors of the new economy.

When the discussion focuses on the determinant role of geography and physical distance in the allocation of the extra-European links, then Athens is also a good example. 19% of its binary and 15% of its weighted degree centrality is because of its links with the Asia and Pacific region and more specifically with the neighbouring countries of Turkey and Israel. The above capacity represents 12% of Europe's total capacity towards this continent.

On the other hand, for some cases geographic proximity is not a determinant factor at all. Leuk (Switzerland) and Nittedal (Norway) for instance do not have any intra-European links. Leuk is only served by two international backbone links which terminate to Africa and Asia and Pacific region. Nittedal has a degree centrality of 10 in 2006 and all of the links are with Africa. So, for these cities other factors different than distance determine their Internet backbone connectivity.

A different case is the Swiss cities of Geneva and Lausanne as well as Dublin. They share respectively 30%, 25% and 20% of their backbone links, and 17%, 25% and 4% of their accumulated bandwidth, with USA and Canada in 2006 while the rest of their connections are only with European cities. The above pattern highlights the unique international character of Switzerland and the demand for interaction with specific cities; the link between the New York and Geneva might reflect to a degree the location of United Nations' (UN) headquarters in the former and the concentration of UN's agencies in the latter (Sassen 2008). In addition, Dublin has well established socio-economic relations with the USA with significant bidirectional migration flows (for instance Walsh 2007) and in addition is located on the west tip of Europe.

Contrary to the interpretation of the above cities' Internet backbone links with North America, back in 2001 a different group of cities used to be highly connected with the US cities, and for very different reasons. Cities such as Bucharest, Zagreb, Belgrade, Tirane and up to a degree Athens used to be highly dependent upon the USA in order to gain universal connectivity. Even at this time a significant part of the intra-European data traffic was through the US (Townsend 2003). Countries with low connectivity did not have any other choice but of using the expensive transatlantic backbone networks even for short intra-European data transfers. US took advantage of its primacy in the development of the Internet to vend such services across the globe in ways similar to London's and to a degree Lisbon's attempt to provide global connectivity to their ex-colonies.

Looking retrospectively, the allocation of the connections from Europe towards the rest of the world seems to be quite stable, with the only difference to appear on the connections with Africa and USA and Canada. Regarding the allocation of the capacity the changes through time are minor and the vast majority of the extra-European bandwidth is because of links with North America.

4.5.2 Betweenness and eigenvector centrality

Continuing the analysis of centrality scores according to different measures, Table 4.8 presents the ‘betweenness centrality’ scores for the 30 most central European cities. The whole table can be found in the Annex of this chapter. As was stated above, this centrality measure is based on the binary network and it is a representation of how common it is to find a city as an internal (not origin or destination) part of a geodesic; or using the Internet terminology, how common it is for a city to be one of the different hops which consist a data packet route between any potential origin and destination, without taking into consideration though the importance of the origin and the destination and how likely is to appear on such a route. Apparently the number of connections that each city has does not affect this measure. Betweenness centrality is not an approach of a city’s infrastructural supply, as it could be assumed for the degree centrality, which measures the accumulated binary connections or the accumulated bandwidth, but rather an indicator of how important a city can potentially be for the Internet data transport system. As a result of this structural difference between the two indicators the cities’ hierarchy presented in Table 4.8 is very different from the one which emerged from the degree centrality.

For presentation reasons, the same method as above is adopted here and the cities’ centralities will be analyzed in blocks of tens following the centrality ranking. The first tier of the ten most central cities has been changed with respect to the degree centrality measures. The most important change refers to the well-established, up to now, four most central cities. According to this measure the cluster of London, Paris, Frankfurt and Amsterdam does not exist anymore. Just like the binary degree centrality measure, Frankfurt is the most central city according to betweenness centrality of the intra-European network and is followed by London, Vienna and Milan, with the latter being the one which is mostly relatively favoured by this measure. Further below, Paris and Amsterdam are found on the fifth and sixth place and interestingly enough the rest of this tier’s cities are located in the Eastern and

South-eastern part of Europe: Zagreb, Budapest, Warsaw and Athens, with the Croatian and Polish capitals being favoured by this measure. Looking at betweenness centrality retrospectively, notable changes took place during the six year study period. In 2001, apart from the changes in the top rank positions with London being the most central city followed by Frankfurt and Amsterdam, the most notable changes were for the cities of Eastern and South-eastern Europe: Budapest, Warsaw and Athens were replaced by Stockholm, Geneva and Lisbon. In terms of centrality change, apart from Frankfurt, those only were the only cities which experienced centrality increase in 2001 – 2006 period, with Warsaw having the greatest one. It can be easily observed that through time there is a noteworthy geographical pattern of change in Europe: because of the development of the Internet infrastructure in the

Table 4.8: Betweenness centrality, 2001-2006

	2001				**2006**				**E.c. change 2001-06**	**a.c. change 2001-06**
	E.c.		**a.c.**		**E.c.**		**a.c.**			
Frankfurt	73.1	*2*	22.0	*3*	100.0	*1*	63.7	*2*	36.8%	189.8%
London	100.0	*1*	100.0	*1*	62.5	*2*	100.0	*1*	-37.5%	0.0%
Voesendorf	59.4	*4*	12.0	*9*	49.2	*3*	14.0	*4*	-17.2%	16.2%
Milan	45.7	*6*	12.4	*8*	39.6	*4*	12.2	*6*	-13.4%	-1.1%
Paris	57.6	*5*	49.3	*2*	36.4	*5*	34.8	*3*	-36.9%	-29.3%
Amsterdam	72.7	*3*	17.3	*4*	30.3	*6*	12.5	*5*	-58.4%	-27.6%
Zagreb	24.6	*9*	4.6	*18*	30.1	*7*	7.5	*10*	22.5%	62.3%
Budapest	5.9	*28*	0.9	*37*	19.3	*8*	6.3	*12*	227.1%	583.0%
Warsaw	1.4	*35*	2.5	*21*	18.6	*9*	5.6	*13*	1193.3%	120.6%
Athens	5.0	*29*	1.5	*33*	13.7	*10*	5.4	*14*	171.0%	269.5%
Copenhagen	18.0	*12*	5.4	*13*	11.7	*11*	7.7	*9*	-34.7%	42.4%
Hamburg	0.0	*48*	3.6	*19*	10.8	*12*	3.1	*17*		-13.1%
Prague	14.4	*15*	2.3	*24*	10.6	*13*	2.9	*18*	-26.2%	23.2%
Marseille					8.6	*14*	2.2	*19*		
Dublin	12.5	*20*	2.3	*26*	8.6	*15*	2.2	*20*	-31.5%	-4.3%
Monaco	1.0	*37*	0.2	*44*	8.6	*15*	2.2	*20*	746.8%	857.4%
Ostrava					8.6	*15*	2.2	*20*		
Zürich	22.6	*11*	4.6	*17*	7.9	*18*	1.9	*23*	-64.8%	-59.1%
Belgrade	0.2	*43*	1.4	*34*	6.1	*19*	1.6	*26*	2608.7%	13.6%
Ljubljana	12.8	*17*	2.0	*32*	4.2	*20*	1.7	*25*	-66.8%	-11.1%
Stockholm	26.9	*7*	9.1	*10*	3.8	*21*	1.8	*24*	-85.8%	-80.5%
Madrid	10.9	*26*	2.4	*23*	3.8	*22*	4.4	*15*	-65.3%	85.0%
Barcelona	4.8	*30*	1.0	*36*	2.6	*23*	0.8	*29*	-46.1%	-14.9%
Stuttgart	0.0	*48*	0.0	*52*	2.3	*24*	1.1	*27*		
Nicosia	0.0	*48*	0.0	*52*	2.2	*25*	1.0	*28*		
Skopje	1.8	*33*	0.4	*39*	1.1	*26*	0.1	*41*	-37.3%	-59.7%
Pristina					1.1	*27*	0.2	*38*		
Tirane	0.0	*48*	0.3	*41*	1.1	*28*	0.3	*33*		-4.4%
Bratislava	0.3	*42*	0.4	*40*	1.0	*29*	0.3	*36*	311.5%	-21.4%
Podgorica					1.0	*30*	0.1	*42*		

E.c. = European cities, a.c. = all cities

Data sources: Telegeography 2007, Author's calculations

cities of Eastern and South-eastern Europe and their steady disengagement from the US in order to gain universal Internet connectivity, cities from this area managed to integrate in the complex network of Internet backbone links in Europe. They also integrated well enough to appear more central than some of the cities of the West, with greater presence in the informational economy.

What is interesting though is to verify the above for the overall network and not only for its intra-European extraction. When all the links are included in the analysis, in 2006 London is the most central city, once again demonstrating its global importance in the geography of the Internet. Paris and Amsterdam overpass Milan, and Vienna remained in the same position. The three cities from Eastern Europe, that is Budapest, Warsaw and Athens, are replaced by Copenhagen, Palermo and Lisbon because of their geographically extensive linkages with non European cities, which enable them to be part of many different geodesics. The differences because of the inclusion of the extra-European links can also be observed in Figure 4.16. Back in 2001, betweenness centrality for the whole network was even more biased toward the west part of Europe. London, Paris, Frankfurt and Amsterdam were the three most central cities. Vienna was the only gateway city for Eastern Europe and the rest of the South-eastern countries were replaced by Stockholm, Brussels, Palermo and Lisbon, reflecting to some extent the degree centrality for all the cities in Table 4.4.

The second tier of cities seems to be much differentiated by the same tier of Table 4.4; only three of ten cities are common in the second tiers of the two binary centrality measures, that is Copenhagen, Hamburg and Prague. Apart from the Czech capital, Ostrava, Belgrade and Ljubljana constitute a cluster of Eastern and Southern cities. Cities such as Ostrava and also Dublin and Monaco climbed in the hierarchy while Zurich’s centrality is undermined by this measure. If the overall network was taken into consideration, a few changes would have taken place in this tier. Apart from the cities of South-eastern Europe mentioned above, Madrid would also be part of this tier, again because of its extroversive connectivity profile.

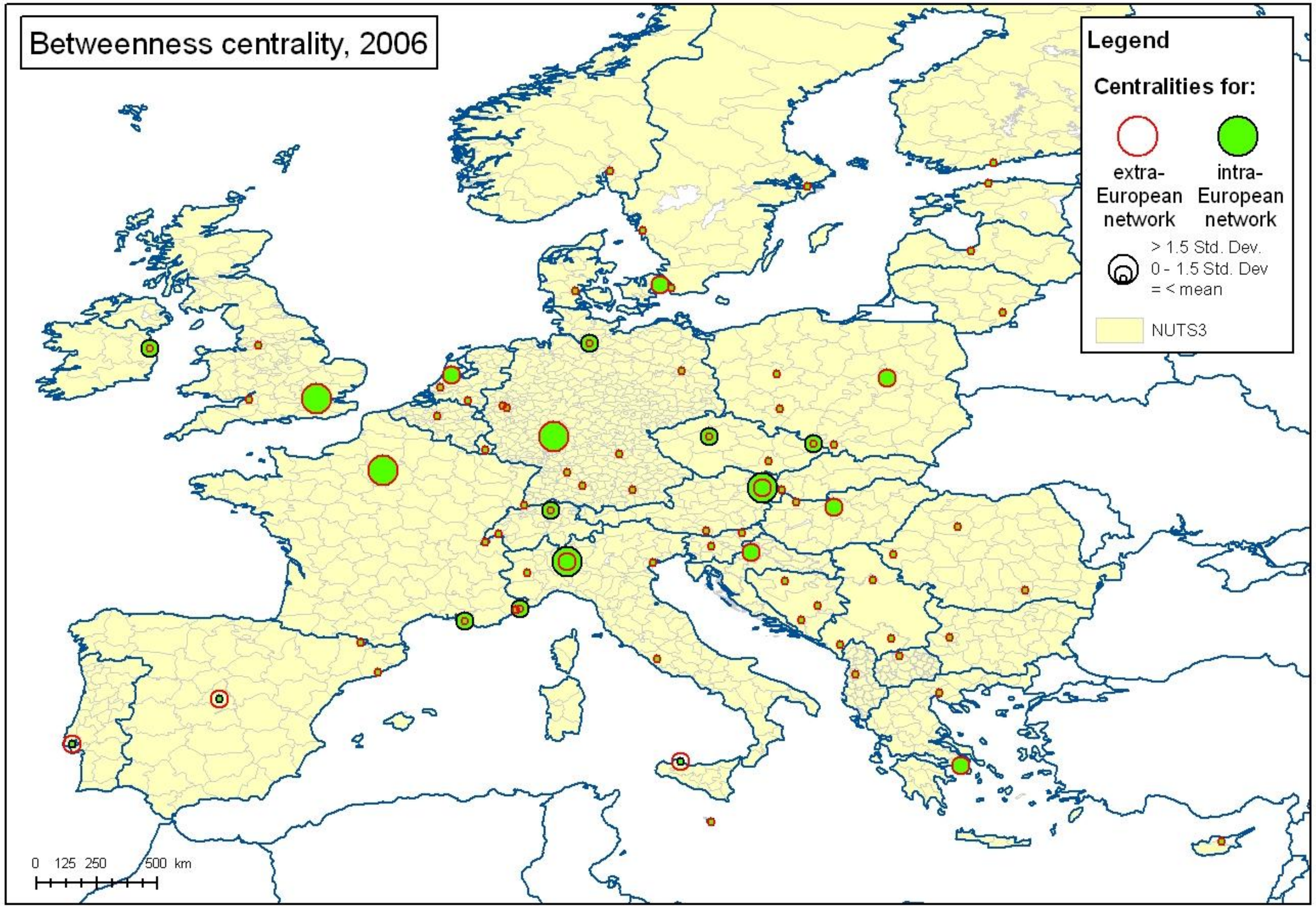


Figure 4.16: Betweenness centrality, 2006

Looking backwards, many changes took place in the six year study period. Belgrade and Monaco gained a lot in terms of betweenness centrality. In general, only 4 out of 10 cities were of this tier back in 2001. The changes are even more radical when the overall network is taken into consideration. Cities such as Marseille and Ostrava would be further below sharing the 52nd place and cities such Geneva, Oslo, Nittedal and Leuk would be part of this tier because of their external links.

Looking at the rest of Table 4.8 cities, many differences with the degree centrality can be observed. Important cities in terms of degree centrality such as Stockholm, Brussels, Geneva and Lisbon lost positions in betweenness rank while a cluster of cities from South-eastern Europe, consisted of Skopje, Pristina, Tirane, Podgorica and Thessaloniki, appears to be more important using this measure. This happens because in order to approach one of these quite peripheral in network distance as well as in real geography cities, the geodesic will necessarily pass by some of the neighbouring cities because they are connected together following almost a serial pattern. And this is the reason why the centrality of these cities appears to be higher according to this measure.

Betweenness centrality provides another view of how cities act as nodes in the Internet backbone. If it can be compared with a previous measure, this is the binary degree centrality, because both of them are based on the binary network. While degree centrality provides the valuable information on the number of the cities that a city is directly connected with, the betweenness centrality is an indicator of how important the location of city is, not in geographical but in the network space terms.

The last centrality measure analysed here is the eigenvector centrality for the weighted network, which is presented in Table 4.9. As it was stated above, this centrality measure is useful because the centrality score of a city is based on the centrality of its neighbours. Again just for comparison reasons the last column of the table presents the rank of the degree centrality for the weighted network of the intra-European international links.

The same strategy applies here and the eigenvector centrality analysis is based on city blocks of tens. Just like the weighted degree centrality, the cluster of the first 4 highly important cities emerged again; London is the most central node and is followed by Paris, Frankfurt and Amsterdam. What is interesting here is the fact that the difference between the first and second city in terms of the normalised score is the smallest comparing all the other centrality measures. Obviously, because this indicator takes into consideration the importance of the neighbours, the differences in centrality measures are smoother than the more clear-cut approach of degree centrality. The cities which are included in the first tier of cities on this measure, but were not according to the degree centrality, are Brussels, Dusseldorf and Geneva. The reason why these cities were upgraded under this centrality measure is because of their intensive linkages with the cluster of the four highly centralised cities. The explanation for the above intensive connectivity is twofold: all of the above mentioned cities are included in Europe’s pentagon, so not only the small physical distance but also their intensive socio-economic links define this connectivity pattern. This is the reason why cities such as Copenhagen, Stockholm and Vienna, which accumulated enough bandwidth to be part of the first tier of cities regarding the weighted degree centrality, ended up being part of the second tier of cities for this measure. Apparently their connectivity pattern with the highly centralised cluster of cities was not intensive enough to enable them to remain in their positions according to the accumulated bandwidth, which is due to their roles as gateway cities to the north and east of Europe.

Changes because of the inclusion of the extra-European links as well changes through time seem to be minor for this indicator (Figure 4.17). When the backbone connections with non-European cities are taken into consideration the main change is that Copenhagen and Dublin appear to be more central, obviously because of their high capacity links with New York. In general, changes because of the external links are minor because almost all of the non European cities appeared to have very low centrality because only their backbone links with European cities are included[19]. Looking retrospectively, Stockholm and Copenhagen used to be more central for both versions of the network. Madrid and Hamburg experienced the highest centrality increase while Brussels the highest decrease through the six year study period.

---

[19] New York is the main exception of this. Even if only its backbone links with Europe are taken into consideration, its weighted degree centrality is equal to 46% of London centrality. This would enable New York to take the fifth position in the relevant rank for the European cities.

Table 4.9: Weighted eigenvector centrality, 2001-2006

	2001				**2006**				**E.c. change 2001-06**	**a.c. change 2001-06**
	E.c.		**a.c.**		**E.c.**		**a.c.**			
London	98.4	*2*	100.0	*1*	100.0	*1*	100.0	*1*	1.6%	0.0%
Paris	100.0	*1*	77.4	*2*	92.3	*2*	70.4	*2*	-7.7%	-9.0%
Frankfurt	88.5	*3*	58.7	*4*	79.9	*3*	55.3	*3*	-9.7%	-5.8%
Amsterdam	85.0	*4*	66.3	*3*	65.1	*4*	52.0	*4*	-23.4%	-21.6%
Madrid	11.7	*8*	7.8	*9*	33.5	*5*	21.6	*5*	186.8%	178.7%
Brussels	59.0	*5*	40.9	*5*	24.9	*6*	16.7	*6*	-57.8%	-59.2%
Milan	14.1	*6*	8.6	*8*	15.1	*7*	9.4	*8*	7.1%	9.7%
Düsseldorf	9.3	*11*	5.8	*10*	12.0	*8*	8.9	*9*	29.0%	53.6%
Geneva	9.7	*10*	5.8	*11*	11.3	*9*	7.4	*12*	16.4%	28.4%
Hamburg	5.3	*13*	3.9	*13*	10.9	*10*	7.6	*11*	104.5%	98.0%
Copenhagen	9.9	*9*	11.1	*6*	10.6	*11*	10.0	*7*	6.8%	-9.7%
Dublin	2.8	*16*	2.3	*15*	10.2	*12*	7.9	*10*	268.5%	244.6%
Zürich	8.1	*12*	4.8	*12*	8.8	*13*	5.2	*14*	8.2%	9.0%
Warsaw	0.4	*26*	0.2	*27*	8.8	*13*	4.7	*15*	2117.3%	1868.2%
Stockholm	13.7	*7*	9.6	*7*	8.6	*15*	5.4	*13*	-37.0%	-44.1%
Voesendorf	2.6	*17*	1.4	*18*	7.3	*16*	3.9	*16*	185.4%	186.3%
Prague	1.0	*19*	0.5	*25*	5.4	*17*	3.0	*18*	443.0%	473.1%
Barcelona	1.0	*19*	0.7	*19*	5.0	*18*	3.1	*17*	406.8%	325.7%
Lisbon	0.8	*24*	0.6	*24*	3.9	*19*	2.7	*19*	397.8%	391.7%
Oslo	4.6	*14*	3.0	*14*	2.9	*20*	2.6	*20*	-37.0%	-13.3%
Budapest	0.4	*26*	0.3	*26*	2.2	*21*	1.1	*22*	443.0%	313.3%
Athens	0.6	*25*	0.7	*20*	2.0	*22*	1.5	*21*	231.8%	125.3%
Bucharest	0.2	*29*	0.1	*34*	1.6	*23*	0.8	*24*	714.5%	931.9%
Bristol					1.4	*24*	0.9	*23*		
Helsinki	1.0	*19*	0.6	*21*	1.1	*25*	0.5	*27*	8.6%	-18.5%
Munich	1.0	*19*	0.6	*23*	0.9	*26*	0.5	*28*	-9.5%	-11.7%
Bratislava	0.2	*29*	0.1	*29*	0.9	*26*	0.4	*30*	352.5%	227.6%
Tallinn	0.2	*29*	0.1	*30*	0.7	*28*	0.5	*29*	262.0%	365.2%
Basel	0.0	*34*	0.0	*39*	0.7	*28*	0.4	*31*		1082.7%
Kolding					0.5	*30*	0.3	*33*		

E.c. = European cities, a.c. = all cities

Data sources: Telegeography 2007, Author's calculations

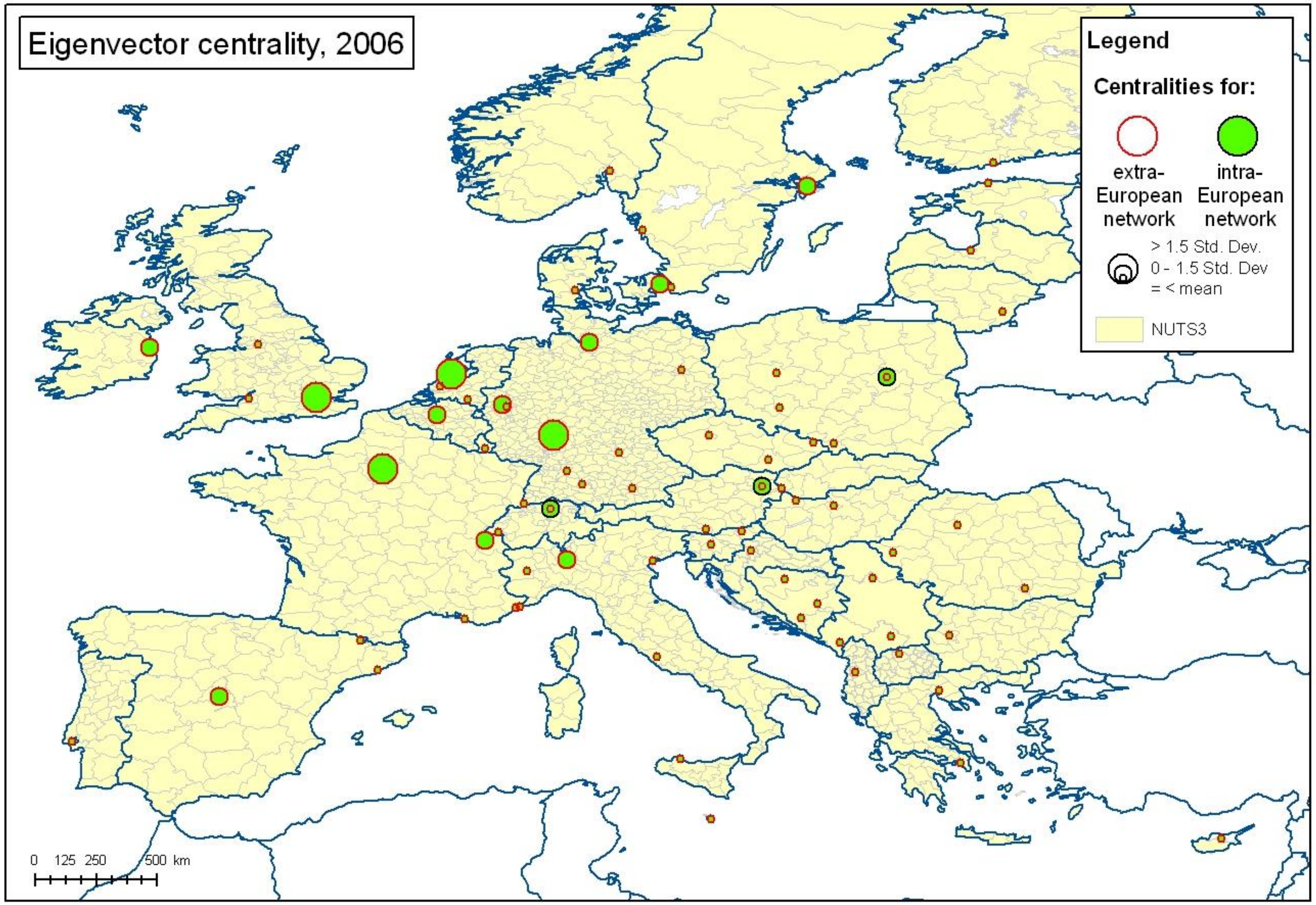


Figure 4.17: Eigenvector centrality, 2006

The second tier of cities consists of the more peripheral (in geographical terms) cities, which play a significant role in the geography of the Internet in Europe. The gateway cities of North Europe, Copenhagen, Stockholm and Oslo, as well as the gateways of the Eastern and South-eastern Europe such as Warsaw, Vienna and Prague are included here. In addition, Dublin and Barcelona because of their high capacity links with London and Paris respectively are found in this tier. Finally Zurich, appears to be a more peripheral city in the geography of the Internet, despite being in Europe's pentagon.

Again the changes because of the extra-European links as well as the changes through time are minor with Warsaw gaining the most centrality increase through time. Interestingly enough, despite Lisbon's external links, its centrality because of the inclusion of extra-European linkages did not change simply because its links are low bandwidth and with cities which appear to be of low importance.

The third tier contains even more peripheral cities. From the Eastern and South-Eastern part of Europe Budapest, Athens, Bucharest and Bratislava are part of this tier while from the North, Helsinki, Tallinn and Kolding can be found here. From the centre of Europe, Munich and Basel no matter their geographic proximity to the well connected cluster, they appeared to be low in the centrality rank. Differences through time and through different versions of networks are again minor.

All the above centrality measures provide different but equally valuable understandings about the distinctive roles of the European cities in the Internet backbone network, helping us to better understand the geography of the Internet infrastructure in Europe. Binary and weighted degree centrality represent the infrastructural supply, with the latter better highlighting the economic geography of Europe while the former being also associated with the political as well as the physical geography. On the contrary, the other two centrality measures are more related with the technological nature of the Internet backbone: betweenness centrality highlights the potential utility of the city in the IP data transfer because of its location in the network space, while eigenvector centrality highlights cities' importance based on its indirect weighted links, in a way similar to Internet's function.

4.5.3 Centrality indicators summary

In order to summarise the above different centrality measures, cluster analysis is applied. As the main goal here is to create clusters of cities based on their performance on the above different centrality indicators, cluster analysis is an appropriate method as its usefulness in classifying relatively ‘raw’ data in an exploratory comparative analysis is well established (Nijkamp et al 1999). Based on the above analysis, the resulting clusters, apart from distinguishing the obvious most centralized cities, should also provide some insights concerning the least central cities. In order to achieve this, the empirical method of *k-means* is selected. In simple terms, this non-hierarchical method results in *k* new clusters, with *k* being a-priori defined (Rogerson 2006). All the centrality indicators were included in this analysis: the binary and weighted degree centrality, the betweenness and eigenvector centrality, with and without the extra-European links. This method was applied both for 2001 and 2006. In order to achieve the above goal for avoiding a two-cluster solution with the most central cities (usually London, Frankfurt, Paris and Amsterdam) forming one cluster and having the rest of the cities crowded in a second cluster, some calibration tests were initially applied in order to select a *k* suitable for our analysis. According to the tests, the most suitable *k* for better explaining the pattern of centrality was equal to 7. Table 4.10 presents the allocation of the European cities to the new clusters. The clusters are also graphically presented in Figures 4.18 and 4.19. According to this, London and Paris have a unique character and for both years they shape individual clusters. This is the case for the other two most central cities, that is Amsterdam and Frankfurt, but only for 2006, indicating the increasingly dissimilarity between these two cities over time regarding the centrality measures.

The distinctive character of the four most central cities was relatively apparent from the analysis of the different centrality measures. Therefore, cluster analysis’ value added is the further classification of the less central cities in order to summarize the analysis of all the different centrality measures. More specifically, the first cluster after the four most central cities is formed by Milan and Vienna for both 2001 and 2006. For the latest year it can be more safely stated that these two cities have quite similar performance in all centrality measures and they play the role of a link between the most central and the moderate central cities. However, it is not that clear for 2001. Table 4.11 presents the final centres of the 7 clusters, as they were produced by the analysis. For the case of 2001 it seems that the 4th cluster is more central when the centralities of the binary connections are taken into consideration while the 5th cluster, which only consists of Brussels, is more central when the weights of the links are included. The above indicate a supplementary relation between Milan and Vienna from the one side and Brussels from the other.

The 6th cluster consists for both years of moderate central cities but still important for Europe's Internet backbone geography. In 2006, 17% of the cities interconnected with at least one backbone network in Europe were part of this cluster (13 out 76), while in 2001 14% (10 out of 69). This cluster includes the gateway cities of north Europe, Stockholm and Copenhagen (as well as Oslo in 2001), Madrid (and Lisbon in 2001) which represents the west border of Europe and the secondary – after Vienna – gateway cities of the east and southeast Europe such as Athens, Budapest, Prague, Warsaw and Zagreb. In 2001 Prague and Sofia were the only cities in this cluster from this area. This change (from 20% to 38% of the cluster) of the cluster indicates the radical change of the Internet backbone connectivity of the cities of the eastern and South-eastern Europe during the six year study period. In addition, this cluster also includes some central (in geographical terms) cities, such as Brussels, Geneva and Zurich (Munich in 2001), which no matter that they are part of Europe's pentagon their Internet backbone centrality measures are not high enough to enable them to be part of Europe's first tier cities. Finally, for both years the most extensive cluster is the one which refers to the least central cities: 75% of both years' interconnected cities are located in this cluster indicating this way the centralized character of the Internet backbone network in Europe.

Table 4.10: Cluster analysis based on the centrality measures for 2001 and 2006

	1	**2**	**3**	**4**	**5**	**6**	**7**
2001	London	Paris	Amsterdam, Frankfurt	Milan, Vienna	Brussels	Copenhagen, Geneva, Lisbon, Madrid, Munich, Oslo, Prague, Sofia, Stockholm, Zürich	Andorra, Antwerp, Athens, Banja Luka, Barcelona, Basel, Belgrade, Berlin, Bratislava, Brno, Bucharest, Budapest, Cologne, Dortmund, Dublin, Düsseldorf, Ehingen, Gdansk, Hamburg, Helsinki, Karlsruhe, Lausanne, Leuk, Ljubljana, Luxembourg, Lyon, Manchester, Monaco, Mostar, Msida, Nice, Nicosia, Nittedal, Palermo, Plovdiv, Portsmouth, Riga, Rome, Rotterdam, San Marino, Sarajevo, Skopje, Split, Strasbourg, Stuttgart, Tallinn, Tartu, Tirane, Turin, Vilnius, Warsaw, Zagreb
2006	London	Paris	Amsterdam	Frankfurt	Milan, Vienna	Athens, Brussels, Budapest, Copenhagen, Düsseldorf, Geneva, Hamburg, Madrid, Prague, Stockholm, Warsaw, Zagreb, Zürich	Andorra, Banja Luka, Barcelona, Basel, Belgrade, Berlin, Bielsko-Biala, Bratislava, Bristol, Brno, Bucharest, Cluj, Dublin, Ehingen, Eindhoven, Gothenburg, Gyor, Helsinki, Hilden, Klagenfurt, Kolding, Lausanne, Lisbon, Ljubljana, Luxembourg, Malmö, Manchester, Maribor, Marseille, Monaco, Mostar, Msida, Munich, Nice, Nicosia, Nuremberg, Oslo, Ostrava, Palermo, Podgorica, Poznan, Pristina, Riga, Rome, Rotterdam, Sarajevo, Skopje, Sofia, Stuttgart, Tallinn, Thessaloniki, Timisoara, Tirane, Turin, Venice, Vilnius, Wroclaw

Table 4.11: Cluster's centres

		1	**2**	**3**	**4**	**5**	**6**	**7**	
2001	degree	100	89	86	57	39	34	11	
	degree (w.)	97	100	95	15	48	12	1	
	degree (ac)	100	66	42	24	22	16	5	
	degree (w.-a.c.)	100	76	70	10	32	9	1	
	betweenness	100	58	73	53	17	17	3	
	betweenness (a.c.)	100	49	20	12	13	6	1	
	eigenvector (w.)	98	100	87	8	59	6	1	
	eigenvector (w.-a.c.)	100	77	63	5	41	4	0	
	N	*1*	*1*	*2*	*2*	*1*	*10*	*52*	*69*
2006	degree	92	58	61	100	57	27	8	
	degree (w.)	100	85	65	81	20	13	1	
	degree (ac)	100	49	32	70	26	13	4	
	degree (w.-a.c.)	100	66	50	58	13	9	1	
	betweenness	63	36	30	100	44	10	1	
	betweenness (a.c.)	100	35	13	64	13	4	1	
	eigenvector (w.)	100	92	65	80	11	11	1	
	eigenvector (w.-a.c.)	100	70	52	55	7	7	0	
	N	*1*	*1*	*1*	*1*	*2*	*13*	*57*	*76*

w. = weighted, a.c. = all the cities

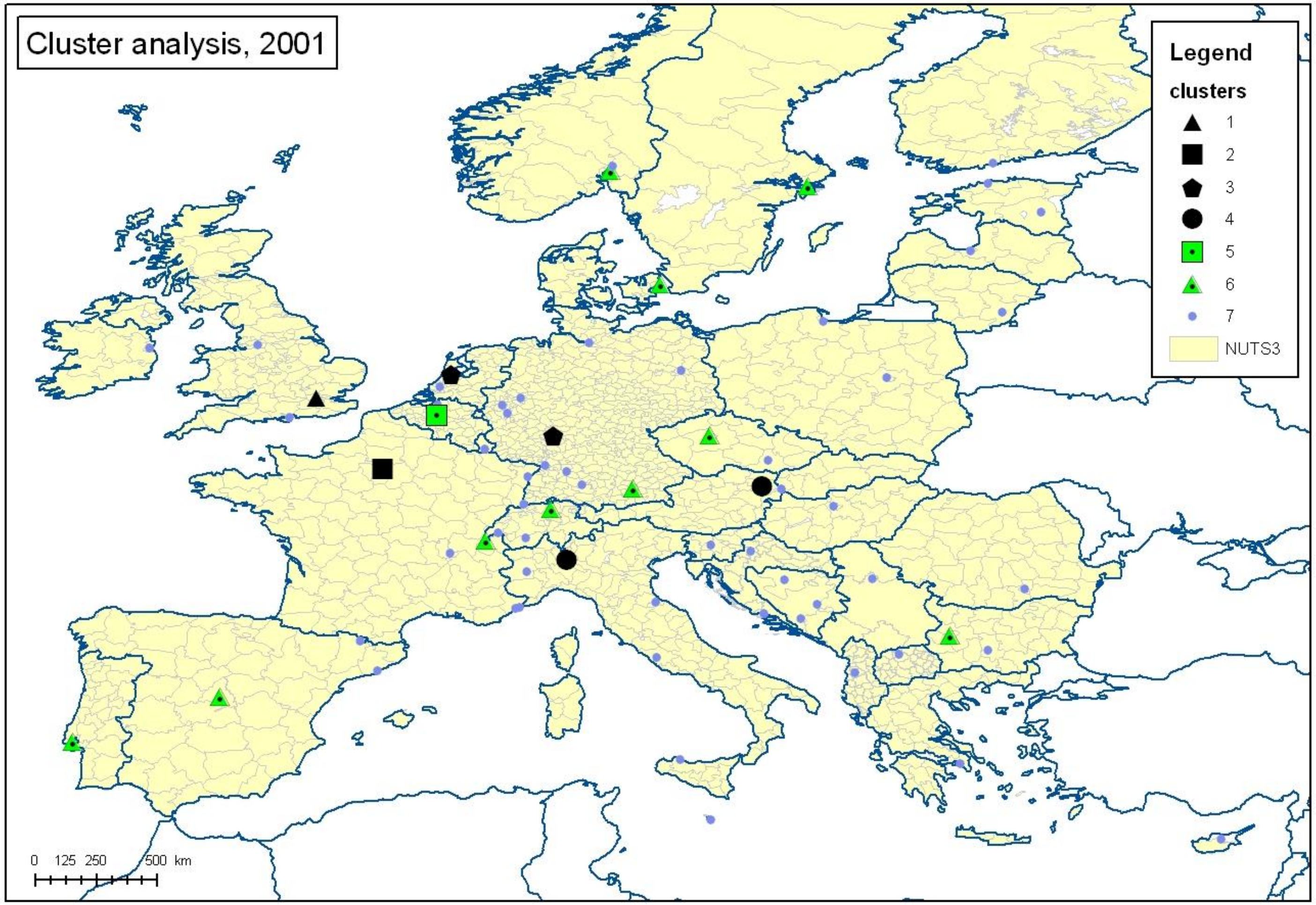


Figure 4.18: Cluster analysis, 2001

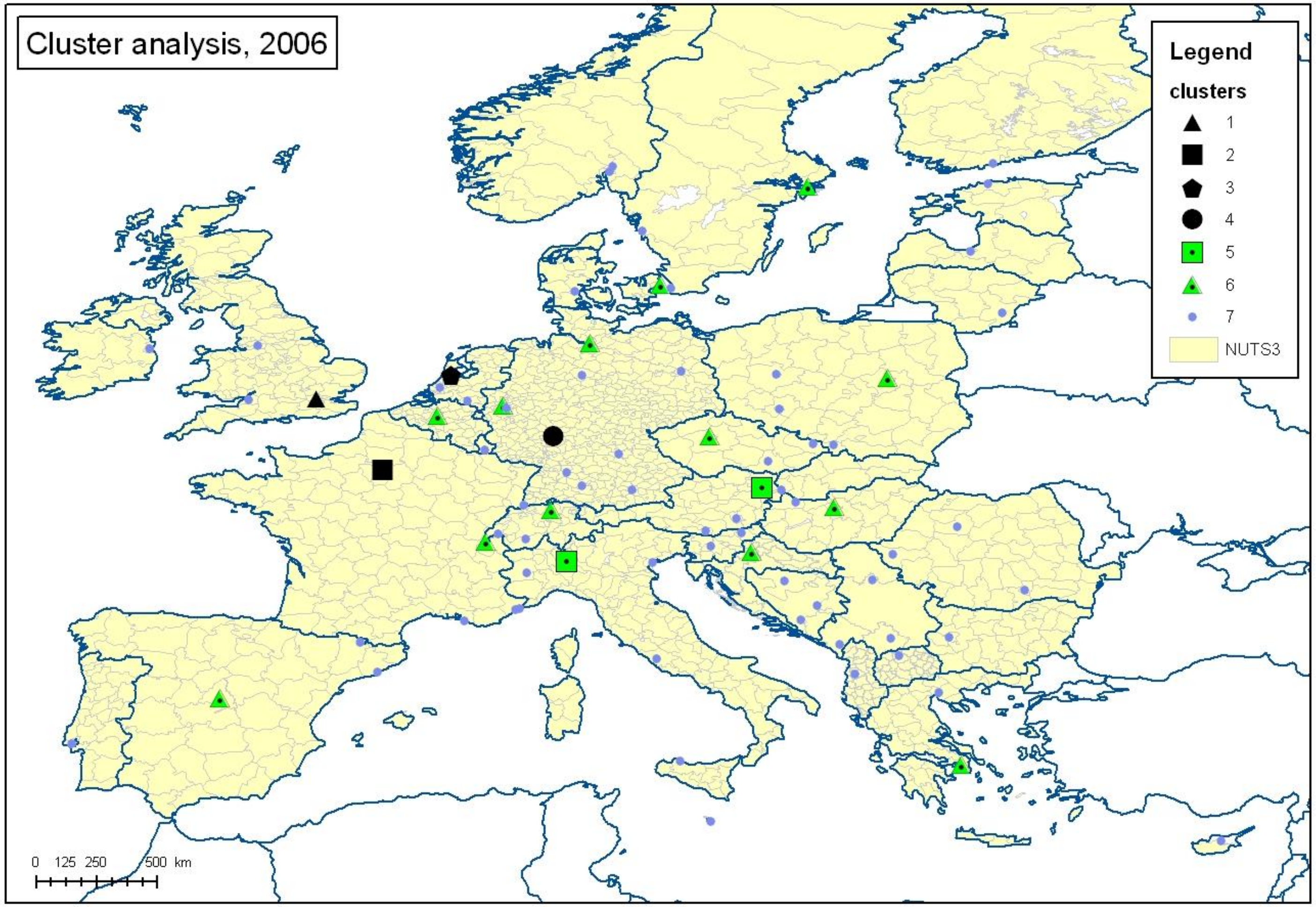


Figure 4.19: Cluster analysis, 2006

4.6 Complex networks analysis

In the following section, an analysis of the Internet backbone network from *complex network* theory is presented. As was explained in Chapter 3, because of the rather complicated topology of the Internet backbone network, the use of complex network theory seems to be necessary in order to comprehend its structure. More specifically, what takes place here is the comparison of the Internet backbone network with some theoretical (and well established in the literature) network models, in order to identify any topological similarities and more important common attributes. Linking with the theoretical research presented in the previous chapter, the main network elements that are under research here is the (small) average distances and the (high) cluster coefficient in order to identify the existence of the small world phenomenon, the vertices degree distribution to identify SF properties (Albert and Barabási 2002) and the *s* metric to diagnose the existence of hubs which hold the network together (Li et al 2005).

Table 4.12 presents some network statistics for the international Internet backbone network for both the European and the global extent for the whole study period 2001-2006. These statistics refer only to the binary version of the network. The first column presents the average distance of the network, which is the mean of all the shortest distances between any given pair of nodes. At Internet jargon the analogy would be the number of hops that a packet should go through in order to reach its final destination. At any case, the average number of nodes that an Internet data packet needs to go through in order to reach any given destination is less than three, indicating a small world phenomenon. In addition, the diameter of the networks, which refers to the longest distance between all the interconnected nodes, is equal or less than 7 for all cases. As it was expected, the average distances for the global extent are slightly higher than those for the intra-European network, not only because the former is larger in extent but also because the extra-European nodes are only connected with European cities and not among them. In order to highlight the small world effect, the average distance of a same size random network is also presented. For almost all the cases, the real networks' average distances and diameters are smaller than the ones for the random networks and for the rest of the cases are slightly longer. To sum up, shorter average distances and smaller diameters are equal with networks' greater efficiency. For the case of the Internet backbone networks, the number of hops that a packet needs to pass through is proportional to processing delays and packet queues, also known as latency (Obraczka and Silva, 2000). It is important to highlight that through time average distances generally decreased, indicating a network efficiency improvement.

Table 4.12: Network statistics

Internet backbone network	**Average Distance (1)**	**Average Distance RN (2)**	**Diameter (3)**	**Diameter RN (4)**	**CC (5)**	**CC RN (6)**
2001 a.c.	2.861	3.600	6	*7*	0.478	0.025
2002 a.c.	2.789	3.565	6	*8*	0.457	0.042
2003 a.c.	2.716	3.538	6	*7*	0.522	0.022
2004 a.c.	2.669	3.378	6	*7*	0.597	0.024
2005 a.c.	2.611	3.413	5	*7*	0.581	0.039
2006 a.c.	2.725	3.457	7	7	0.574	0.033
2001 E.c.	2.762	2.579	7	5	0.424	0.082
2002 E.c.	2.641	2.719	5	6	0.524	0.114
2003 E.c.	2.555	2.653	5	5	0.539	0.075
2004 E.c.	2.495	2.549	6	5	0.562	0.063
2005 E.c.	2.477	2.570	5	5	0.563	0.086
2006 E.c.	2.549	2.631	6	5	0.571	0.112

a.c. = all cities, E.c. = European cities

Data sources: Telegeography 2007, Author's calculations

The next column presents the *clustering coefficients* (CC), which distinguish the SW networks from the small world phenomenon as was explained in the methodology chapter. The latter measures the average cliquishness of a node (Latora and Marchiori 2001), using the following formula:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

So, the clustering coefficient of a node *i* is the ratio between the number of edges *Ei* that exist among its nearest neighbours (nodes which are directly connected with node *i*) and the maximum number of these edges, which is equal to $\frac{k_i(k_k - 1)}{2}$ (Albert and Barabási 2002). According to the SW model, the clustering coefficient should be high and at any case higher than a random network of the same size. As is illustrated in Table 4.12, the clustering coefficient is always much higher comparing to ones occurred in random networks. So, it could be said that according to this indicator, the Internet backbone network for all the different versions and for the whole time period seem to fit with the *small world* (SW) networks model.

Taking the analysis a step further, we empirically explore the networks' vertex degree distribution in order to identify whether they fit with the SW model or if they have *scale free* (SF) attributes. As it was mentioned before, SF networks are related with power law vertex degree distributions in a way that the probability *P(k)* that a vertex in a network interacts with k other vertices or in other words the *probability distribution function* (PDF) decays as a power law, following $P(x) \approx x^{-a}$ (Barabási and Albert 1999). On the contrary, SW networks are characterized by exponential degree distribution.

Most of the network analysis studies which have as a starting point the statistical physics are based on a the stochastic approach, which assumes an underlying probability model as the mechanism for generating the power law distribution, which is responsible for denoting the distribution function. Consequently, the main objective is to describe the PDF by calculating the exponent. However, this research approach includes the danger that the distribution might not have been emerged by a power law mechanism (Li et al 2005). Thus, for the needs of this paper, we try to empirically test whether the degree distribution follows a power or an exponential law, using a non-stochastic and fairly simple approach. Instead of using the PDF for exploring SF properties, the use of the *cumulative distribution function* (CDF) is preferred (Li et al 2005). The latter indicates the probability (or the frequency) that a vertex interacts with *x* or more other vertices. The advantage of CDF is its ability to minimise the statistical noise usually present in the tail when the distribution is plotted (Newman 2005). CDF was introduced by Vilfredo Pareto for his work on income distribution, named after him. According to this, income distribution follows a power law in a way that a person's income is greater than or equal to x when $P(X \geq x) = (m/x)^k$, with $m>0$, $k>0$ and $x>m$ where m is the minimum income. Its CDF will be $P(X < x) = 1 - (m/x)^k$ and the PDF $P(X = x) = km^k x^{-(k+1)}$ (Adamic 2000; Adamic and Huberman 2002; Fosco 2004). CDF's and PDF's exponents are related as $a = k + 1$, which means that if PDF follows a power law then CDF also follows a power law but in this case the straight line in a log-log graph would be steeper, indicating a less homogenous distribution.

Pareto's distributions and CDF are also related with *Zipf's* law and *rank/frequency plots*. Zipf's law was suggested by George Kingsley Zipf in order to explain the size of the r'th largest occurrence of an event *y* and he concluded that the latter is inversely proportional to its rank, with $y \sim r^{-\beta}$, β close to 1. Zipf's law initially described the English language's most common words, but it has been used widely in explaining many social phenomena, including the rank size of cities. Rank/frequency plots are easy to be constructed. It is the plot of the event's occurrences as a function of the rank of these occurrences in descending order. For the case of a binary network, a rank/frequency plot can be interpreted by saying that the *r-th* most connected vertex has *n* connections with other vertices. However, this is equivalent to saying that r nodes have *n* or more connections, which is exactly the same with CDF or Pareto's distribution except for the fact that axis x and y are flipped in a way that in a CDF plot the horizontal axis represents the actual event (the number of connections for a binary network) and the vertical axis the rank (or the frequency) of the node. So, the easiest way to create CDF plots is to transpose the axis of a rank/frequency plot (Adamic 2000, Newman 2005). The above method is used here in order to create CDF plots, which are analyzed below.

The next step after creating the CDF is to identify the law that the distribution follows. The most straightforward way to identify a power law distribution is to present its CDF in a log-log plot. In this case, a power law will form a straight line because the initial equation $f(x) = cx^{-a}$ will be transformed to $\log f(x) = \log(c) - a\log(x)$, which represents a straight line in a log-log space. In order to calculate the fit of the power law and its exponent or to try additional laws, such as exponential, OLS is the simplest method. R-square of the fit line has been widely used in various studies as the determinant of whether the vertex degree distribution follows a power or exponential law (Faloutsos et al 1999, Gorman and Kulkarni 2004, Schintler et al 2004, Patuelli et al 2007). The same studies also calculated distributions exponents using OLS.

However, there is an ongoing debate in the literature regarding the accuracy of OLS in empirically exploring distributions. Newman (2005) suggested using a *maximum likelihood (ML)* approach in estimating the scaling factor. Clauset et al (2008) suggested the use of *Kolmogorov – Smirnov (KS)* statistic for testing the power law hypothesis and the likelihood test for the comparison of different models. Russo et al (2007) in order to identify the distribution used *J* test, *KS* test and the *encompassing test*. However, the common consensus is that the appearance of the

CDF's plot as a straight line in a log-log plot seems to be the necessary but not a sufficient factor in order to conclude about the power law (Clauset et al 2008). In addition, after testing the accuracy of OLS on PDF and CDF and CDF occurred from rank/frequency plots against ML, Clauset et al (2008, 6) found that OLS on CDF based on rank/frequency plots "do reasonably well" for continuous data. For the needs of this section and for simplicity reasons, OLS is used in order to test whether power or an exponential law fits better, identifying this way whether Internet backbone appear SF or just SW properties.

Figure 4.20 presents the CDF plots of the weighted and binary version of the Internet backbone network both for the global and the European extent. Table 4.13 presents the R-square and the exponents delivered by OLS. Nevertheless, the latter should be treated carefully because of the above discussion. The first observation that can be made is that there is no graph with a perfect straight line. However, some of the CDF almost form straight lines, which is a first indication of an SF structure. In terms of R square, for most of the cases the dominant law seems to fit quite well, with R square being higher than 0.95 for some cases. Another common characteristic is the fact that the plots for valued Internet networks seem to be more differentiated through time, reflecting in this way the differences in technology such as the diachronic bandwidth upgrade in backbone links. On the contrary, the binary network does not demonstrate such differences through time because the number the Internet backbone links remain relative stable through time.

In more detail, the Internet backbone network for all the cities appear to have SF properties both for weighted and binary versions according to their degree distribution. R square for power law fit is above 0.90 for binary versions and around 0.90 for the weighted Internet backbone network. What is interesting is that R square for the power law fit is decreasing through time for the weighted Internet network for all the cities while the R square for the exponential law is increasing. The latter reflects the tendency towards a more homogenous bandwidth distribution across cities.

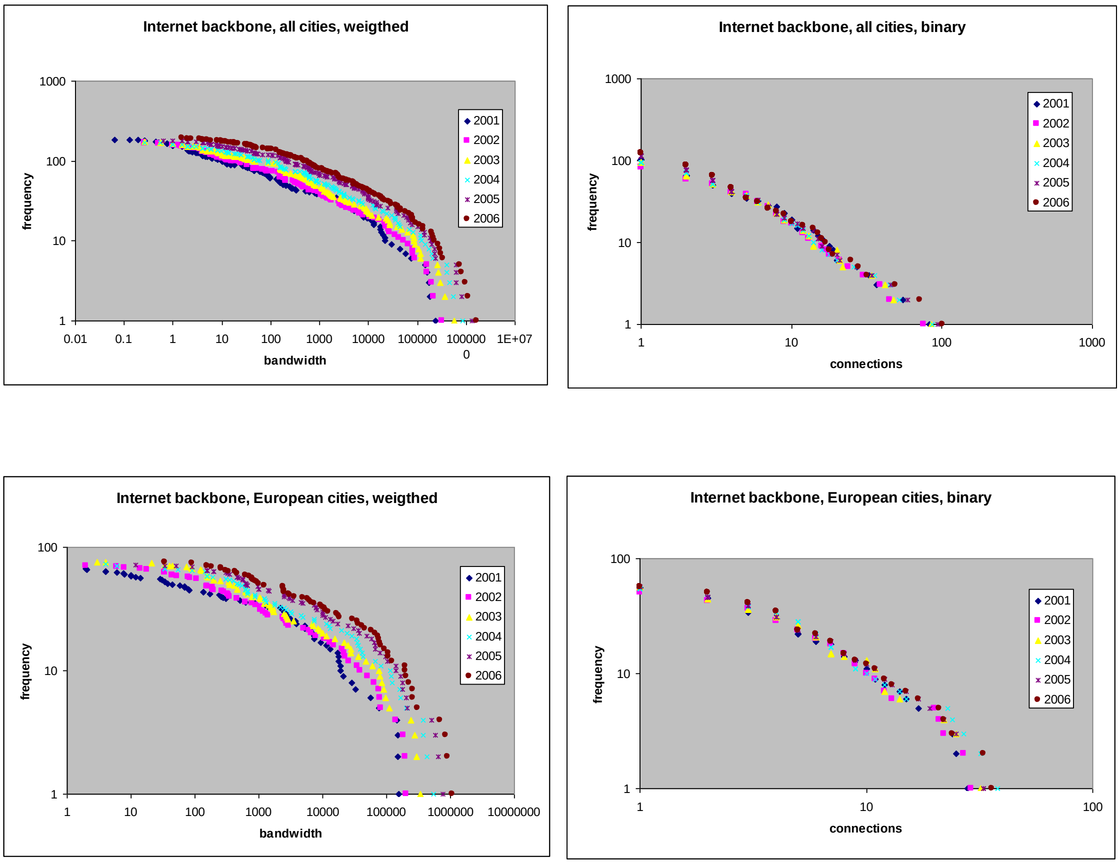
Table 4.13: Power and exponential law fit (OLS)

	Weighted			Binary		
	Exp R^2	Power R^2		Exp R^2	Power R^2	
2001 a.c.	0.525	**0.921**	*SF*	0.762	**0.966**	*SF*
2002 a.c.	0.578	**0.919**		0.810	**0.934**	
2003 a.c.	0.557	**0.892**		0.760	**0.949**	
2004 a.c.	0.571	**0.886**		0.750	**0.947**	
2005 a.c.	0.546	**0.874**		0.703	**0.968**	
2006 a.c.	0.543	**0.876**		0.671	**0.964**	
2001 E.c.	0.727	**0.790**	*SF*	**0.962**	0.894	*SW*
2002 E.c.	0.785	**0.855**		**0.965**	0.859	
2003 E.c.	0.755	**0.869**		**0.950**	0.881	
2004 E.c.	0.782	**0.839**		**0.924**	0.876	
2005 E.c.	0.776	**0.837**		**0.967**	0.873	
2006 E.c.	0.758	**0.844**		**0.941**	0.873	

a.c. = all cities, E.c. = European cities
Data sources: Telegeography 2007, Author's calculations

What is interesting is to analyze further the structure of this network exploring the structural importance of these highly connected nodes using the s metric. As is presented in Table 4.14, the s metric, which takes values between 0 and 1 is very low

Figure 4.20: Internet backbone network's degree distribution



for the global extent of Internet backbone network for the whole time period. Over time, the s metric for the Internet backbone steadily decreases, indicating an increasing tendency of the Internet backbone hubs to be connected with less connected vertices. At any case, s metric is very low and prevents from recognizing the global extent of the Internet backbone networks as SF, no matter their power law fit. However, it should be highlighted here that distributions close to power law and hubs with neighbours of low connectivity were expected for the network of all the cities. As it was noted before, for the non-European cities, only their edges with the European ones are included in the analysis while their links with the non-European cities are missed. As a result, these cities appeared to be poorly connected.

Table 4.14: *s* metric

	a.c.	E.c.
2001	0.233	0.606
2002	0.242	0.571
2003	0.216	0.513
2004	0.216	0.482
2005	0.196	0.528
2006	0.178	0.511

a.c. = all cities; E.c. = European cities
Data sources: Telegeography 2007, Author's calculations

In order to overcome the above, the analysis focuses on the networks consisted only by European cities. From a first look at the graph and R square, it seems that the weighted version better fits to a power law, while the binary one is more homogenous since it better fits to exponential law for the whole time period. The fit for the binary network over the 6 year period is higher than 0.94 and also higher than the ones for the valued networks.

It should be noted here that it is common for vertex degree CDF to follow a power law distribution not for all the vertices but only for a part of them and usually the most connected one, indicating a truncated power law distribution. Graphically, this would be translated to a curve with two different slopes: one which follows a power law and appears as a straight line in log-log space and another one different than a straight line. No matter that this is a common case (Amaral et al 2000) it is not the case for the Internet backbone network. A number of tests took place in order to identify sets of vertices (or otherwise parts of the CDF curve) which fit to a power law distribution (and appear as a straight line in CDF). However, no truncated power law can be identified for the Internet backbone network.

All in all, the Internet backbone binary network seems to be more homogenous comparing with the previous networks. However, even for this scale, the bandwidth's distributions better fit with power laws indicating SF properties, which are related with the existence of some very-well connected (in terms of bandwidth and passenger loads) hubs and a bulk of less connected cites. However, it should be highlighted here the sharp cut-off which appears for the four most connected cities in the weighted Internet backbone network, which are the main hubs of a SF network. In order to better fit within such a network structure and a SF distribution, these cities should have been characterized by greater bandwidth.

Going back to Table 4.14, although this indicator is slightly higher for the intra-European Internet backbone network than for its global extent it is still low, indicating the lack of hubs which hold the network together. So, it could be said that there is quite an uneven bandwidth distribution across the European cities, which is a characteristic of SF networks, but at any case it cannot be said that those networks follow SF models.

To sum up, the Internet backbone network does not fit with a SF model. In spite of the highly connected hubs, it cannot be claimed that these nodes can hold the network together. When the 5 most central nodes of the intra-European network (7% of all European nodes) were removed, then the average distance among all reachable European cities would be increased only by 22%. However, for a SF network, as it was mentioned before, the increase would be more than 100% when the 5% of the most connected nodes were removed according to Albert et al (2000).

It seems that there is a less homogenous distribution of the technology (i.e. bandwidth), which might indicate the existence of SF properties, but at the same time the distribution of the actual connections is more homogenous, better fitting with SW model. Last but not least, the low s metric values for almost all the cases indicate structures which do not fit with highly connected super-hubs, which hold the networks together.

4.7 Conclusions

The main objective of this chapter has been to present and explore the Internet backbone network. This initial analysis is fundamental in understanding the actual data. Further analysis and modelling techniques application, which will assist in approaching the research questions, would be nonsensical without first exploring and mapping the quantitative dataset.

Summing up the above analysis, the first point which should be mentioned is the distribution of the backbone links when their capacity is taken into consideration, which is far from a normal distribution. There is a small number of links which can be characterised as outliers because their capacity is much greater than the average one. What is interesting from the geography point of view is that these outliers are mainly concentrated in Europe’s core area, known as the pentagon. Additionally, over time there is a trend towards a decrease in the variation of the capacity of the backbone links. Technological improvements and cost reductions enabled this cohesion trend in the capacity distribution: while the maximum capacity link was increased by a factor of 6 (from 77,768 to 467,671 Mbps) during 2001-2006, the lowest capacity link was increased by a factor of 31 (from 0.064 to 2Mbps). Apart from this, specific reasons should explain the skewed distribution of the high capacity backbone links among cities such as London, Paris, Amsterdam and Frankfurt. An attempt to explain the above takes place in the following chapters of this thesis.

Taking the analysis a step further, all the backbone links together form the network of the international Internet backbone links. And this is the reason why network analysis is the appropriate methodological approach to further analyse this infrastructure. The basic networks statistics demonstrated that both the intra-European but also the global extent of the international Internet backbone links grew in terms of nodes and edges, with the intra-European ones growing faster. And this difference in development can be translated to the better participation of the European cities in this universal infrastructural network.

Also interesting are the results of the centralisation measures in order to better realize the network’s big picture. According to the degree centralisation, no matter that the intra-European network appears to be less centralised than the global one, there is a clear tendency for a more centralised network to emerge through the six year study period. The trend is similar when the betweenness centralisation is calculated. However, according to this measure, which is a better proxy of the IP data packet transport system, the European international Internet backbone network is less centralised compared to the previous measure, which is more related with the infrastructural supply. The centralisation of the network though appears to be higher regarding the eigenvector centrality, which also takes into consideration the indirect links. But the network seems even more centralised when the capacity of the backbone links is included in eigenvector centrality calculation. All in all, it could be said that the European extraction of the international Internet backbone network is in general moderately centralised, but when the capacity is included in the analysis the network appears to be more centralised.

Nonetheless, what is more interesting from the geography perspective, are the centrality measures. These local statistics comment on the distinctive roles of the cities as nodes of the backbone network. Different geographies emerged due to these measures. First of all, London’s dominance is more than obvious. Different cities have different and significant roles in the European Internet backbone network, but London is beyond that. London is a global hub for the Internet backbone network in a way that its hinterland is not limited inside Europe’s border. Its giant capacity links with New York both reflect its position in the global urban hierarchy, but also determine its role in the geography of the Internet backbone networks.

However, the economic geography is not limited to explaining London’s dominance. The spatial allocation of the capacity of the backbone links as well as the cities’ centrality when the capacity is taken into consideration reflect both the potential interaction between cities but also the city’s economic role. And this is why 55% of the intra-European bandwidth and 59% of the whole bandwidth of the European cities is allocated among London (17% and 22% respectively), Paris (14% and 14%), Frankfurt (13% and 13%) and Amsterdam (11% and 11%), forming in this way the main core of the European IP backbone network.

The picture is slightly differentiated when the binary links are studied. Because of the absence of the unequally distributed capacity, the hub and spoke structure of this infrastructural network and the distinctive role of some cities as gateways for their hinterland emerge. Among other reasons, the importance of the physical location is highlighted in order to explain such connectivities. Notable is also the political geography which is reflected on the binary extra-European links, since some post-colonial relations seem to affect the connectivity patterns with out of Europe regions.

Interesting also are the results from the cluster analysis, which summarise the importance and the distinctive roles of cities outside of Europe's main core. Milan and Vienna but also cities such as Athens, Brussels, Budapest, Copenhagen, Dusseldorf, Geneva, Hamburg, Madrid, Prague, Stockholm, Warsaw, Zagreb and Zurich proved to be significant nodes of the European international Internet backbone network.

However, despite the moderate centralised character of the network and the dominance of some cities, it cannot be claimed that this network fits with the scale-free model, at least as it has been approached by the recent literature. Europe's main hubs are important, but not important enough to hold the network together. Apart from this, small-world properties such as low average distance and high clustering coefficient, which can confirm its efficiency, are also present.

All in all, despite the fact that the network has been expanded mainly towards the East during the 6 year period and some cities out of the core gained in terms of centrality, its core remained strong and the overall network appeared to be slightly more centralised.

Chapter 4

Annex

Table A4.1: Binary degree centrality, 2001-2006

	2001					2006					Change 2001-06 for	
	E.c.		**a.c.**		**%Eur. links**	**E.c. (%)**		**a.c.**		**%Eur. links**	**E.c. (%)**	**a.c. (%)**
Frankfurt	85.7	*3*	44.6	*3*	*64.9*	100.0	*1*	70.3	*2*	*50.7*	16.7	57.7
London	100.0	*1*	100.0	*1*	*33.7*	91.7	*2*	100.0	*1*	*32.7*	-8.3	0.0
Vienna	60.7	*5*	24.1	*5*	*85.0*	66.7	*3*	27.7	*5*	*85.7*	9.8	15.0
Amsterdam	85.7	*3*	39.8	*4*	*72.7*	61.1	*4*	31.7	*4*	*68.8*	-28.7	-20.3
Paris	89.3	*2*	66.3	*2*	*45.5*	58.3	*5*	48.5	*3*	*42.9*	-34.7	-26.8
Milan	53.6	*6*	24.1	*5*	*75.0*	47.2	*6*	24.8	*6*	*68.0*	-11.9	2.7
Budapest	28.6	*14*	10.8	*20*	*88.9*	41.7	*7*	17.8	*8*	*83.3*	45.8	64.4
Stockholm	50.0	*7*	22.9	*7*	*73.7*	36.1	*8*	18.8	*7*	*68.4*	-27.8	-17.8
Athens	28.6	*14*	12.0	*16*	*80.0*	33.3	*9*	15.8	*10*	*75.0*	16.7	31.5
Zürich	35.7	*11*	13.3	*14*	*90.9*	33.3	*9*	13.9	*14*	*85.7*	-6.7	4.6
Copenhagen	42.9	*8*	16.9	*13*	*85.7*	30.6	*11*	16.8	*9*	*64.7*	-28.7	-0.2
Zagreb	10.7	*34*	4.8	*35*	*75.0*	27.8	*12*	9.9	*17*	*100.0*	159.3	105.4
Hamburg	17.9	*22*	12.0	*16*	*60.0*	25.0	*13*	11.9	*15*	*75.0*	40.0	-1.4
Prague	39.3	*9*	13.3	*14*	*100.0*	25.0	*13*	8.9	*21*	*100.0*	-36.4	-32.8
Brussels	39.3	*9*	21.7	*8*	*61.1*	22.2	*15*	11.9	*15*	*66.7*	-43.4	-45.2
Madrid	28.6	*14*	10.8	*20*	*88.9*	22.2	*15*	14.9	*12*	*53.3*	-22.2	37.0
Warsaw	17.9	*22*	9.6	*24*	*62.5*	22.2	*15*	9.9	*17*	*80.0*	24.4	2.7
Stuttgart	3.6	*56*	1.2	*60*	*100.0*	22.2	*15*	8.9	*21*	*88.9*	522.2	639.6
Geneva	25.0	*18*	10.8	*20*	*77.8*	19.4	*19*	9.9	*17*	*70.0*	-22.2	-8.7
Barcelona	17.9	*22*	6.0	*31*	*100.0*	19.4	*19*	6.9	*24*	*100.0*	8.9	15.0
Tallinn	10.7	*34*	3.6	*40*	*100.0*	19.4	*19*	7.9	*23*	*87.5*	81.5	119.1
Bratislava	21.4	*19*	8.4	*25*	*85.7*	16.7	*22*	5.9	*27*	*100.0*	-22.2	-29.6
Bucharest	14.3	*30*	7.2	*28*	*66.7*	16.7	*22*	6.9	*24*	*85.7*	16.7	-4.1
Düsseldorf	10.7	*34*	3.6	*40*	*100.0*	13.9	*24*	5.9	*27*	*83.3*	29.6	64.4
Lisbon	21.4	*19*	18.1	*11*	*40.0*	13.9	*24*	15.8	*10*	*31.3*	-35.2	-12.3
Oslo	35.7	*11*	21.7	*8*	*55.6*	13.9	*24*	6.9	*24*	*71.4*	-61.1	-68.0
Belgrade	10.7	*34*	4.8	*35*	*75.0*	13.9	*24*	5.0	*31*	*100.0*	29.6	2.7
Nicosia	10.7	*34*	3.6	*40*	*100.0*	13.9	*24*	5.0	*31*	*100.0*	29.6	37.0
Sofia	28.6	*14*	12.0	*16*	*80.0*	13.9	*24*	5.0	*31*	*100.0*	-51.4	-58.9
Ljubljana	17.9	*22*	6.0	*31*	*100.0*	13.9	*24*	5.0	*31*	*100.0*	-22.2	-17.8
Munich	32.1	*13*	18.1	*11*	*60.0*	13.9	*24*	5.9	*27*	*83.3*	-56.8	-67.1
Vilnius	17.9	*22*	7.2	*28*	*83.3*	13.9	*24*	5.9	*27*	*83.3*	-22.2	-17.8
Helsinki	21.4	*19*	10.8	*20*	*66.7*	13.9	*24*	5.0	*31*	*100.0*	-35.2	-54.3
Marseille						13.9	*24*	5.0	*31*	*100.0*		
Dublin	10.7	*34*	4.8	*35*	*75.0*	11.1	*35*	5.0	*31*	*80.0*	3.7	2.7
Skopje	10.7	*34*	3.6	*40*	*100.0*	11.1	*35*	4.0	*39*	*100.0*	3.7	9.6
Riga	17.9	*22*	6.0	*31*	*100.0*	11.1	*35*	5.0	*31*	*80.0*	-37.8	-17.8
Tirane	3.6	*56*	3.6	*40*	*33.3*	11.1	*35*	4.0	*39*	*100.0*	211.1	9.6
Pristina						11.1	*35*	4.0	*39*	*100.0*		
Podgorica						11.1	*35*	4.0	*39*	*100.0*		
Msida	7.1	*45*	2.4	*47*	*100.0*	8.3	*41*	3.0	*45*	*100.0*	16.7	23.3
Luxembourg	17.9	*22*	8.4	*25*	*71.4*	8.3	*41*	3.0	*45*	*100.0*	-53.3	-64.8
Basel	3.6	*56*	1.2	*60*	*100.0*	8.3	*41*	3.0	*45*	*100.0*	133.3	146.5
Brno	7.1	*45*	2.4	*47*	*100.0*	8.3	*41*	3.0	*45*	*100.0*	16.7	23.3
Lausanne	10.7	*34*	4.8	*35*	*75.0*	8.3	*41*	4.0	*39*	*75.0*	-22.2	-17.8
Palermo	10.7	*34*	19.3	*10*	*18.8*	8.3	*41*	14.9	*12*	*20.0*	-22.2	-23.0
Monaco	7.1	*45*	2.4	*47*	*100.0*	8.3	*41*	3.0	*45*	*100.0*	16.7	23.3
Hilden						8.3	*41*	3.0	*45*	*100.0*		
Timisoara						8.3	*41*	3.0	*45*	*100.0*		

Thessaloniki						8.3	*41*	4.0	*39*	*75.0*		
Andorra	10.7	*34*	3.6	*40*	*100.0*	5.6	*51*	2.0	*52*	*100.0*	-48.1	-45.2
Sarajevo	14.3	*30*	6.0	*31*	*80.0*	5.6	*51*	2.0	*52*	*100.0*	-61.1	-67.1
Malmö						5.6	*51*	2.0	*52*	*100.0*		
Kolding						5.6	*51*	2.0	*52*	*100.0*		
Klagenfurt						5.6	*51*	2.0	*52*	*100.0*		
Ostrava						5.6	*51*	2.0	*52*	*100.0*		
Rotterdam	14.3	*30*	4.8	*35*	*100.0*	2.8	*57*	2.0	*52*	*50.0*	-80.6	-58.9
Banja Luka	7.1	*45*	2.4	*47*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-61.1	-58.9
Rome	7.1	*45*	2.4	*47*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-61.1	-58.9
Berlin	17.9	*22*	7.2	*28*	*83.3*	2.8	*57*	1.0	*60*	*100.0*	-84.4	-86.3
Manchester	3.6	*56*	1.2	*60*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-22.2	-17.8
Ehingen	7.1	*45*	2.4	*47*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-61.1	-58.9
Turin	3.6	*56*	1.2	*60*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-22.2	-17.8
Nice	7.1	*45*	2.4	*47*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-61.1	-58.9
Mostar	7.1	*45*	2.4	*47*	*100.0*	2.8	*57*	1.0	*60*	*100.0*	-61.1	-58.9
Maribor						2.8	*57*	1.0	*60*	*100.0*		
Nuremberg						2.8	*57*	1.0	*60*	*100.0*		
Gothenburg						2.8	*57*	1.0	*60*	*100.0*		
Bielsko-Biala						2.8	*57*	1.0	*60*	*100.0*		
Bristol						2.8	*57*	1.0	*60*	*100.0*		
Gyor						2.8	*57*	1.0	*60*	*100.0*		
Venice						2.8	*57*	1.0	*60*	*100.0*		
Cluj						2.8	*57*	1.0	*60*	*100.0*		
Eindhoven						2.8	*57*	1.0	*60*	*100.0*		
Wroclaw						2.8	*57*	1.0	*60*	*100.0*		
Poznan						2.8	*57*	1.0	*60*	*100.0*		
Cologne	7.1	*45*	2.4	*47*	*100.0*						-100.0	-100.0
Dortmund	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Antwerp	7.1	*45*	2.4	*47*	*100.0*						-100.0	-100.0
Nittedal	14.3	*30*	8.4	*25*	*57.1*			9.9	*17*	*0.0*	-100.0	17.4
Strasbourg	10.7	*34*	3.6	*40*	*100.0*						-100.0	-100.0
Portsmouth	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Plovdiv	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Gdansk	3.6	*56*	2.4	*47*	*50.0*						-100.0	-100.0
Lyon	7.1	*45*	2.4	*47*	*100.0*						-100.0	-100.0
Tartu	3.6	*56*	2.4	*47*	*50.0*						-100.0	-100.0
Karlsruhe	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Leuk	3.6	*56*	12.0	*16*	*10.0*			2.0	*52*	*0.0*	-100.0	-83.6
San Marino	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Split	3.6	*56*	1.2	*60*	*100.0*						-100.0	-100.0
Bijeljina												
Dresden												
Lille												
Graz								1.0	*60*	*0.0*		
Hannover								1.0	*60*	*0.0*		
Innsbruck												
Katowice												
Lodz												
Salzburg												
Varna												
Oradea												
Rijeka												

E.c. = European cities, a.c. = all cities, Data sources: Telegeography 2007, Author's calculations

Table A4.2: Weighted degree centrality, 2001-2006

	2001					**2006**					**Change 2001-06 for**	
	E.c.		**a.c.**		**%Eur. links**	**E.c.**		**a.c.**		**%Eur. links**	**E.c. (%)**	**a.c. (%)**
London	96.8	*2*	100.0	*1*	*64.0*	100.0	*1*	100.0	*1*	*63.7*	3.3	0.0
Paris	100.0	*1*	75.6	*2*	*87.4*	84.9	*2*	66.4	*2*	*81.4*	-15.1	-12.1
Frankfurt	96.5	*3*	67.8	*4*	*94.1*	81.0	*3*	58.2	*3*	*88.6*	-16.0	-14.1
Amsterdam	93.2	*4*	71.9	*3*	*85.6*	64.7	*4*	49.6	*4*	*83.0*	-30.6	-31.0
Stockholm	35.7	*6*	24.5	*6*	*96.4*	29.2	*5*	20.2	*5*	*92.3*	-18.2	-17.7
Madrid	12.3	*10*	9.2	*9*	*88.5*	24.5	*6*	16.2	*7*	*96.1*	99.5	77.1
Copenhagen	21.1	*7*	18.4	*7*	*75.9*	24.2	*7*	18.1	*6*	*85.4*	14.9	-1.6
Vienna	11.3	*12*	7.5	*13*	*99.9*	21.4	*8*	13.7	*8*	*99.7*	89.2	82.7
Hamburg	11.0	*14*	7.6	*12*	*95.4*	19.4	*9*	12.7	*10*	*97.3*	75.9	66.3
Milan	18.2	*8*	12.5	*8*	*95.6*	19.0	*10*	12.8	*9*	*94.1*	4.4	2.3
Brussels	48.2	*5*	32.2	*5*	*99.0*	18.5	*11*	11.8	*11*	*100.0*	-61.6	-63.3
Düsseldorf	8.3	*15*	5.5	*15*	*100.0*	11.6	*12*	8.0	*12*	*92.5*	40.7	46.7
Geneva	11.2	*13*	7.5	*14*	*99.1*	10.2	*13*	7.9	*13*	*82.8*	-8.7	5.3
Zürich	12.2	*11*	8.3	*11*	*97.2*	10.2	*14*	6.7	*14*	*96.8*	-16.6	-19.2
Warsaw	1.4	*32*	0.9	*32*	*96.0*	8.7	*15*	5.6	*15*	*99.1*	531.3	489.5
Bratislava	4.4	*20*	2.9	*20*	*99.9*	7.7	*16*	4.9	*16*	*100.0*	77.3	70.8
Prague	7.3	*16*	4.8	*16*	*100.0*	7.4	*17*	4.7	*18*	*100.0*	0.6	-3.0
Helsinki	4.8	*19*	3.3	*18*	*95.4*	7.3	*18*	4.7	*19*	*100.0*	53.6	41.3
Oslo	13.7	*9*	9.1	*10*	*99.0*	6.9	*19*	4.9	*17*	*90.7*	-49.2	-46.5
Dublin	2.5	*24*	1.8	*24*	*92.6*	6.3	*20*	4.2	*20*	*96.4*	152.7	134.1
Budapest	1.9	*27*	1.3	*27*	*98.5*	5.5	*21*	3.5	*21*	*98.9*	191.2	179.8
Munich	6.0	*17*	4.0	*17*	*98.6*	4.4	*22*	2.8	*22*	*100.0*	-26.7	-30.3
Barcelona	2.4	*25*	1.6	*25*	*100.0*	3.9	*23*	2.5	*24*	*100.0*	64.7	58.8
Athens	0.6	*35*	0.5	*35*	*73.2*	3.4	*24*	2.8	*23*	*79.2*	510.9	444.2
Lisbon	2.0	*26*	1.5	*26*	*90.8*	3.3	*25*	2.2	*25*	*96.0*	62.4	48.0
Brno	0.0	*63*	0.0	*64*	*100.0*	2.8	*26*	1.8	*26*	*100.0*	109309.7	105388.1
Tallinn	0.3	*37*	0.2	*38*	*100.0*	1.7	*27*	1.1	*29*	*98.3*	431.2	421.0
Bucharest	0.9	*33*	0.6	*33*	*99.6*	1.7	*28*	1.1	*30*	*99.6*	88.8	82.0
Ljubljana	0.2	*39*	0.1	*40*	*100.0*	1.6	*29*	1.0	*31*	*100.0*	800.1	767.8
Marseille						1.2	*30*	0.8	*32*	*100.0*		
Ostrava						1.0	*31*	0.7	*33*	*100.0*		
Vilnius	0.0	*49*	0.0	*47*	*62.1*	1.0	*32*	0.7	*34*	*100.0*	2606.8	1521.4
Bristol						0.9	*33*	0.6	*35*	*100.0*		
Riga	0.2	*40*	0.1	*41*		0.9	*34*	0.6	*36*	*98.4*	432.0	421.4
Basel	1.6	*30*	1.1	*30*	*100.0*	0.7	*35*	0.5	*37*	*100.0*	-55.6	-57.1
Luxembourg	0.8	*34*	0.6	*34*	*96.4*	0.6	*36*	0.4	*38*	*100.0*	-21.8	-27.3
Zagreb	0.0	*51*	0.1	*45*	*25.7*	0.6	*37*	0.4	*39*	*100.0*	2372.2	512.5
Stuttgart	0.2	*38*	0.1	*39*	*100.0*	0.5	*38*	0.4	*40*	*100.0*	177.3	167.4
Kolding						0.5	*39*	0.3	*42*	*100.0*		
Hilden						0.3	*40*	0.2	*44*	*100.0*		
Sofia	0.1	*42*	0.1	*43*	*94.4*	0.3	*41*	0.2	*45*	*100.0*	170.8	146.6
Lausanne	0.1	*43*	0.1	*43*	*75.0*	0.3	*42*	0.2	*43*	*75.0*	229.2	217.4
Rotterdam	1.8	*28*	1.2	*28*	*100.0*	0.2	*43*	0.3	*41*	*50.3*	-86.6	-74.3
Venice						0.2	*44*	0.2	*46*	*100.0*		
Rome	1.7	*29*	1.1	*29*	*100.0*	0.2	*45*	0.2	*47*	*100.0*	-86.1	-86.6
Manchester	0.4	*36*	0.3	*37*	*100.0*	0.2	*45*	0.2	*47*	*100.0*	-40.7	-42.9
Timisoara						0.2	*47*	0.1	*49*	*100.0*		
Belgrade	0.0	*56*	0.0	*56*		0.2	*48*	0.1	*50*	*100.0*	2416.6	1206.5
Berlin	4.1	*21*	2.7	*21*		0.1	*49*	0.1	*51*	*100.0*	-97.1	-97.2

Poznan						0.1	*49*	0.1	*51*	*100.0*		
Bielsko-Biala						0.1	*51*	0.1	*53*	*100.0*		
Palermo	0.1	*45*	0.4	*36*	*9.4*	0.1	*52*	1.5	*27*	*3.6*	68.4	320.2
Skopje	0.0	*57*	0.0	*57*		0.1	*53*	0.1	*54*	*100.0*	982.8	944.0
Nicosia	0.0	*58*	0.0	*59*		0.1	*54*	0.0	*55*	*100.0*	1108.3	1065.0
Msida	0.0	*46*	0.0	*50*		0.1	*55*	0.0	*56*	*100.0*	51.4	46.0
Malmö						0.1	*56*	0.0	*57*	*100.0*		
Nuremberg						0.1	*56*	0.0	*57*	*100.0*		
Monaco	0.1	*43*	0.1	*46*	*100.0*	0.1	*58*	0.0	*59*	*100.0*	-21.0	-23.8
Gothenburg						0.1	*59*	0.0	*60*	*100.0*		
Wroclaw						0.1	*59*	0.0	*60*	*100.0*		
Klagenfurt						0.0	*61*	0.0	*64*	*100.0*		
Thessaloniki						0.0	*61*	0.0	*63*	*93.2*		
Tirane	0.0	*67*	0.0	*63*	*40.0*	0.0	*63*	0.0	*65*	*100.0*	3078.0	1125.6
Sarajevo	0.0	*54*	0.0	*55*	*98.3*	0.0	*64*	0.0	*66*	*100.0*	67.0	58.4
Banja Luka	0.0	*62*	0.0	*62*	*100.0*	0.0	*65*	0.0	*67*	*100.0*	665.5	638.0
Mostar	0.0	*64*	0.0	*65*		0.0	*65*	0.0	*67*	*100.0*	1048.2	1007.1
Pristina						0.0	*67*	0.0	*69*	*100.0*		
Nice	0.2	*41*	0.1	*42*		0.0	*68*	0.0	*70*	*100.0*	-84.6	-85.1
Podgorica						0.0	*69*	0.0	*71*	*100.0*		
Turin	1.6	*30*	1.1	*30*		0.0	*70*	0.0	*72*	*100.0*	-99.1	-99.1
Gyor						0.0	*70*	0.0	*72*	*100.0*		
Cluj						0.0	*70*	0.0	*72*	*100.0*		
Eindhoven						0.0	*70*	0.0	*72*	*100.0*		
Andorra	0.0	*47*	0.0	*51*		0.0	*74*	0.0	*76*	*100.0*	-80.7	-81.4
Ehingen	0.0	*61*	0.0	*61*		0.0	*75*	0.0	*77*	*100.0*	-33.6	-36.0
Maribor						0.0	*75*	0.0	*77*	*100.0*		
Strasbourg	4.8	*18*	3.2	*19*	*100.0*						-100.0	-100.0
Lyon	3.3	*22*	2.2	*22*	*100.0*						-100.0	-100.0
Antwerp	3.3	*23*	2.2	*23*	*100.0*						-100.0	-100.0
Cologne	0.0	*48*	0.0	*52*	*100.0*						-100.0	-100.0
Gdansk	0.0	*50*	0.0	*48*	*50.0*						-100.0	-100.0
Dortmund	0.0	*52*	0.0	*54*	*100.0*						-100.0	-100.0
Leuk	0.0	*52*	0.0	*49*	*42.1*			0.0	*80*	*0.0*	-100.0	-100.0
Nittedal	0.0	*55*	0.0	*53*	*48.3*			0.0	*69*	*0.0*	-100.0	-100.0
San Marino	0.0	*59*	0.0	*60*	*100.0*						-100.0	-100.0
Tartu	0.0	*60*	0.0	*57*	*66.7*						-100.0	-100.0
Plovdiv	0.0	*64*	0.0	*65*	*100.0*						-100.0	-100.0
Portsmouth	0.0	*66*	0.0	*67*	*100.0*						-100.0	-100.0
Karlsruhe	0.0	*67*	0.0	*68*	*100.0*						-100.0	-100.0
Split	0.0	*67*	0.0	*68*	*100.0*						-100.0	-100.0
Bijeljina												
Dresden												
Lille												
Graz								0.0	*60*	*0.0*		
Hannover								1.2	*28*	*0.0*		
Innsbruck												
Katowice												
Lodz												
Salzburg												
Varna												
Oradea												
Rijeka												

E.c. = European cities, a.c. = all cities; Data sources: Telegeography 2007, Author's calculations

Table A4.3: Betweenness centrality, 2001-2006

	2001				**2006**				**Change 2001-06 for**	
	E.c.		**a.c.**		**E.c.**		**a.c.**		**E.c.**	**a.c.**
Frankfurt	73.1	*2*	22.0	*3*	100.0	*1*	63.7	*2*	36.8%	189.8%
London	100.0	*1*	100.0	*1*	62.5	*2*	100.0	*1*	-37.5%	0.0%
Voesendorf	59.4	*4*	12.0	*9*	49.2	*3*	14.0	*4*	-17.2%	16.2%
Milan	45.7	*6*	12.4	*8*	39.6	*4*	12.2	*6*	-13.4%	-1.1%
Paris	57.6	*5*	49.3	*2*	36.4	*5*	34.8	*3*	-36.9%	-29.3%
Amsterdam	72.7	*3*	17.3	*4*	30.3	*6*	12.5	*5*	-58.4%	-27.6%
Zagreb	24.6	*9*	4.6	*18*	30.1	*7*	7.5	*10*	22.5%	62.3%
Budapest	5.9	*28*	0.9	*37*	19.3	*8*	6.3	*12*	227.1%	583.0%
Warsaw	1.4	*35*	2.5	*21*	18.6	*9*	5.6	*13*	1193.3%	120.6%
Athens	5.0	*29*	1.5	*33*	13.7	*10*	5.4	*14*	171.0%	269.5%
Copenhagen	18.0	*12*	5.4	*13*	11.7	*11*	7.7	*9*	-34.7%	42.4%
Hamburg	0.0	*48*	3.6	*19*	10.8	*12*	3.1	*17*		-13.1%
Prague	14.4	*15*	2.3	*24*	10.6	*13*	2.9	*18*	-26.2%	23.2%
Marseille					8.6	*14*	2.2	*19*		
Dublin	12.5	*20*	2.3	*26*	8.6	*15*	2.2	*20*	-31.5%	-4.3%
Monaco	1.0	*37*	0.2	*44*	8.6	*15*	2.2	*20*	746.8%	857.4%
Ostrava					8.6	*15*	2.2	*20*		
Zürich	22.6	*11*	4.6	*17*	7.9	*18*	1.9	*23*	-64.8%	-59.1%
Belgrade	0.2	*43*	1.4	*34*	6.1	*19*	1.6	*26*	2608.7%	13.6%
Ljubljana	12.8	*17*	2.0	*32*	4.2	*20*	1.7	*25*	-66.8%	-11.1%
Stockholm	26.9	*7*	9.1	*10*	3.8	*21*	1.8	*24*	-85.8%	-80.5%
Madrid	10.9	*26*	2.4	*23*	3.8	*22*	4.4	*15*	-65.3%	85.0%
Barcelona	4.8	*30*	1.0	*36*	2.6	*23*	0.8	*29*	-46.1%	-14.9%
Stuttgart	0.0	*48*	0.0	*52*	2.3	*24*	1.1	*27*		
Nicosia	0.0	*48*	0.0	*52*	2.2	*25*	1.0	*28*		
Skopje	1.8	*33*	0.4	*39*	1.1	*26*	0.1	*41*	-37.3%	-59.7%
Pristina					1.1	*27*	0.2	*38*		
Tirane	0.0	*48*	0.3	*41*	1.1	*28*	0.3	*33*		-4.4%
Bratislava	0.3	*42*	0.4	*40*	1.0	*29*	0.3	*36*	311.5%	-21.4%
Podgorica					1.0	*30*	0.1	*42*		
Thessaloniki					1.0	*31*	0.5	*30*		
Brussels	17.2	*13*	13.4	*6*	0.9	*32*	0.4	*32*	-94.7%	-97.0%
Msida	4.4	*31*	0.3	*43*	0.8	*33*	0.3	*34*	-81.7%	18.4%
Vilnius	0.8	*38*	2.5	*22*	0.7	*34*	0.3	*37*	-17.0%	-89.0%
Tallinn	0.0	*48*	0.0	*52*	0.7	*35*	0.2	*39*		
Munich	6.4	*27*	3.0	*20*	0.4	*36*	0.1	*44*	-94.1%	-97.5%
Geneva	24.9	*8*	5.1	*16*	0.3	*37*	0.3	*35*	-99.0%	-94.4%
Bucharest	0.3	*41*	0.1	*46*	0.2	*38*	0.4	*31*	-50.3%	258.7%
Palermo	12.7	*18*	12.9	*7*	0.2	*39*	9.2	*7*	-98.8%	-28.7%
Lisbon	23.3	*10*	15.2	*5*	0.1	*40*	8.8	*8*	-99.4%	-42.4%
Sarajevo	1.7	*34*	1.2	*35*	0.1	*41*	0.0	*46*	-95.0%	-99.3%
Sofia	13.2	*16*	5.6	*12*	0.1	*42*	0.0	*47*	-99.4%	-99.9%
Düsseldorf	0.0	*48*	0.0	*52*	0.1	*43*	0.0	*45*		
Riga	12.6	*19*	2.3	*25*	0.1	*44*	0.1	*43*	-99.4%	-96.4%
Oslo	12.3	*25*	8.5	*11*	0.0	*45*	0.2	*40*	-100.0%	-98.2%
Helsinki	14.7	*14*	2.1	*31*	0.0	*45*	0.0	*48*	-100.0%	-100.0%
Luxembourg	0.0	*48*	0.3	*42*	0.0	*45*	0.0	*48*		-100.0%
Basel	0.0	*48*	0.0	*52*	0.0	*45*	0.0	*48*		
Brno	12.5	*20*	2.3	*26*	0.0	*45*	0.0	*48*	-100.0%	-100.0%
Lausanne	0.1	*46*	0.1	*47*	0.0	*45*	0.0	*48*	-100.0%	-100.0%

Hilden			0.0 *45*	0.0 *48*		
Timisoara			0.0 *45*	0.0 *48*		
Andorra	0.0 *48*	0.0 *52*	0.0 *45*	0.0 *48*		
Malmö			0.0 *45*	0.0 *48*		
Kolding			0.0 *45*	0.0 *48*		
Klagenfurt			0.0 *45*	0.0 *48*		
Rotterdam	0.6 *39*	0.2 *45*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Banja Luka	0.2 *45*	0.0 *50*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Rome	0.0 *47*	0.0 *51*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Berlin	1.3 *36*	0.4 *38*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Manchester	0.0 *48*	0.0 *52*	0.0 *45*	0.0 *48*		
Ehingen	0.2 *44*	0.0 *49*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Turin	0.0 *48*	0.0 *52*	0.0 *45*	0.0 *48*		
Nice	0.4 *40*	0.0 *48*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Mostar	12.5 *20*	2.3 *26*	0.0 *45*	0.0 *48*	-100.0%	-100.0%
Maribor			0.0 *45*	0.0 *48*		
Nuremberg			0.0 *45*	0.0 *48*		
Gothenburg			0.0 *45*	0.0 *48*		
Bielsko-Biala			0.0 *45*	0.0 *48*		
Bristol			0.0 *45*	0.0 *48*		
Gyor			0.0 *45*	0.0 *48*		
Venice			0.0 *45*	0.0 *48*		
Cluj			0.0 *45*	0.0 *48*		
Eindhoven			0.0 *45*	0.0 *48*		
Wroclaw			0.0 *45*	0.0 *48*		
Poznan			0.0 *45*	0.0 *48*		
Cologne	0.0 *48*	0.0 *52*				
Dortmund	0.0 *48*	0.0 *52*				
Antwerp	0.0 *48*	0.0 *52*				
Nittedal	1.9 *32*	5.4 *14*		6.8 *11*	-100.0%	26.6%
Strasbourg	12.5 *20*	2.3 *26*			-100.0%	-100.0%
Portsmouth	0.0 *48*	0.0 *52*				
Plovdiv	0.0 *48*	0.0 *52*				
Gdansk	0.0 *48*	0.0 *52*				
Lyon	12.5 *20*	2.3 *26*			-100.0%	-100.0%
Tartu	0.0 *48*	0.0 *52*				
Karlsruhe	0.0 *48*	0.0 *52*				
Leuk	0.0 *48*	5.3 *15*		4.4 *16*		-16.8%
San Marino	0.0 *48*	0.0 *52*				
Split	0.0 *48*	0.0 *52*				
Bijeljina						
Dresden						
Lille						
Graz				0.0 *48*		
Hannover				0.0 *48*		
Innsbruck						
Katowice						
Lodz						
Salzburg						
Varna						
Oradea						
Rijeka						

E.c. = European cities, a.c. = all cities

Data sources: Telegeography 2007, Author's calculations

Table A4.4: Weighted eigenvector centrality, 2001-2006

	2001				**2006**				**Change 2001-06 for**	
	E.c.		**a.c.**		**E.c.**		**a.c.**		**E.c.**	**a.c.**
London	98.4	*2*	100.0	*1*	100.0	1	100.0	*1*	1.6%	0.0%
Paris	100.0	*1*	77.4	*2*	92.3	2	70.4	*2*	-7.7%	-9.0%
Frankfurt	88.5	*3*	58.7	*4*	79.9	3	55.3	*3*	-9.7%	-5.8%
Amsterdam	85.0	*4*	66.3	*3*	65.1	4	52.0	*4*	-23.4%	-21.6%
Madrid	11.7	*8*	7.8	*9*	33.5	5	21.6	*5*	186.8%	178.7%
Brussels	59.0	*5*	40.9	*5*	24.9	6	16.7	*6*	-57.8%	-59.2%
Milan	14.1	*6*	8.6	*8*	15.1	7	9.4	*8*	7.1%	9.7%
Düsseldorf	9.3	*11*	5.8	*10*	12.0	8	8.9	*9*	29.0%	53.6%
Geneva	9.7	*10*	5.8	*11*	11.3	9	7.4	*12*	16.4%	28.4%
Hamburg	5.3	*13*	3.9	*13*	10.9	10	7.6	*11*	104.5%	98.0%
Copenhagen	9.9	*9*	11.1	*6*	10.6	11	10.0	*7*	6.8%	-9.7%
Dublin	2.8	*16*	2.3	*15*	10.2	12	7.9	*10*	268.5%	244.6%
Zürich	8.1	*12*	4.8	*12*	8.8	13	5.2	*14*	8.2%	9.0%
Warsaw	0.4	*26*	0.2	*27*	8.8	13	4.7	*15*	2117.3%	1868.2%
Stockholm	13.7	*7*	9.6	*7*	8.6	*15*	5.4	*13*	-37.0%	-44.1%
Voesendorf	2.6	*17*	1.4	*18*	7.3	*16*	3.9	*16*	185.4%	186.3%
Prague	1.0	*19*	0.5	*25*	5.4	*17*	3.0	*18*	443.0%	473.1%
Barcelona	1.0	*19*	0.7	*19*	5.0	*18*	3.1	*17*	406.8%	325.7%
Lisbon	0.8	*24*	0.6	*24*	3.9	*19*	2.7	*19*	397.8%	391.7%
Oslo	4.6	*14*	3.0	*14*	2.9	*20*	2.6	*20*	-37.0%	-13.3%
Budapest	0.4	*26*	0.3	*26*	2.2	*21*	1.1	*22*	443.0%	313.3%
Athens	0.6	*25*	0.7	*20*	2.0	*22*	1.5	*21*	231.8%	125.3%
Bucharest	0.2	*29*	0.1	*34*	1.6	*23*	0.8	*24*	714.5%	931.9%
Bristol					1.4	*24*	0.9	*23*		
Helsinki	1.0	*19*	0.6	*21*	1.1	*25*	0.5	*27*	8.6%	-18.5%
Munich	1.0	*19*	0.6	*23*	0.9	*26*	0.5	*28*	-9.5%	-11.7%
Bratislava	0.2	*29*	0.1	*29*	0.9	*26*	0.4	*30*	352.5%	227.6%
Tallinn	0.2	*29*	0.1	*30*	0.7	*28*	0.5	*29*	262.0%	365.2%
Basel	0.0	*34*	0.0	*39*	0.7	*28*	0.4	*31*		1082.7%
Kolding					0.5	*30*	0.3	*33*		
Luxembourg	1.0	*19*	0.6	*22*	0.5	*30*	0.3	*34*	-45.7%	-53.8%
Ljubljana	0.0	*34*	0.0	*41*	0.5	*30*	0.2	*35*		1057.8%
Rotterdam	0.2	*29*	0.1	*32*	0.4	*33*	0.5	*26*	81.0%	443.2%
Lausanne	0.0	*34*	0.1	*33*	0.4	*33*	0.4	*32*		345.4%
Vilnius	0.0	*34*	0.0	*51*	0.4	*33*	0.2	*36*		6119.3%
Riga	0.0	*34*	0.0	*45*	0.2	*36*	0.1	*37*		852.0%
Sofia	0.0	*34*	0.0	*36*	0.2	*36*	0.1	*38*		159.6%
Marseille					0.2	*36*	0.1	*39*		
Timisoara					0.2	*36*	0.1	*40*		
Zagreb	0.0	*34*	0.1	*35*	0.2	*36*	0.1	*41*		27.3%
Ostrava					0.2	*36*	0.1	*42*		
Brno	0.0	*34*	0.0	*60*	0.2	*36*	0.1	*42*		
Belgrade	0.0	*34*	0.0	*58*	0.0	*43*	0.0	*44*		8358.3%
Stuttgart	0.0	*34*	0.0	*46*	0.0	*43*	0.0	*45*		324.8%
Wroclaw					0.0	*43*	0.0	*46*		
Malmö					0.0	*43*	0.0	*47*		
Gothenburg					0.0	*43*	0.0	*48*		
Nicosia	0.0	*34*	0.0	*50*	0.0	*43*	0.0	*49*		580.1%
Thessaloniki					0.0	*43*	0.0	*50*		
Manchester	0.0	*34*	0.0	*47*	0.0	*43*	0.0	*51*		148.8%

Monaco	0.0	*34*	0.0	*40*	0.0	*43*	0.0	*52*		-28.6%
Berlin	0.4	*26*	0.2	*28*	0.0	*43*	0.0	*53*	-100.0%	-91.3%
Msida	0.0	*34*	0.0	*55*	0.0	*43*	0.0	*54*		1411.3%
Hilden					0.0	*43*	0.0	*55*		
Palermo	0.0	*34*	0.0	*57*	0.0	*43*	0.0	*56*		1695.8%
Venice					0.0	*43*	0.0	*57*		
Poznan					0.0	*43*	0.0	*58*		
Turin	0.0	*34*	0.0	*52*	0.0	*43*	0.0	*59*		487.7%
Cluj					0.0	*43*	0.0	*59*		
Eindhoven					0.0	*43*	0.0	*59*		
Andorra	0.0	*34*	0.0	*48*	0.0	*43*	0.0	*62*		-46.2%
Tirane	0.0	*34*	0.0	*53*	0.0	*43*	0.0	*63*		273.2%
Banja Luka	0.0	*34*	0.0	*60*	0.0	*43*	0.0	*64*		
Nuremberg					0.0	*43*	0.0	*65*		
Skopje	0.0	*34*	0.0	*55*	0.0	*43*	0.0	*66*		-6.7%
Gyor					0.0	*43*	0.0	*67*		
Pristina					0.0	*43*	0.0	*68*		
Rome	0.0	*34*	0.0	*42*	0.0	*43*	0.0	*68*		-98.3%
Klagenfurt					0.0	*43*	0.0	*70*		
Podgorica					0.0	*43*	0.0	*72*		
Sarajevo	0.0	*34*	0.0	*60*	0.0	*43*	0.0	*72*		
Ehingen	0.0	*34*	0.0	*60*	0.0	*43*	0.0	*72*		
Nice	0.0	*34*	0.0	*49*	0.0	*43*	0.0	*72*		-100.0%
Mostar	0.0	*34*	0.0	*60*	0.0	*43*	0.0	*72*		
Maribor					0.0	*43*	0.0	*72*		
Bielsko-Biala					0.0	*43*	0.0	*72*		
Hannover							0.6	*25*		
Graz							0.0	*70*		
Nittedal	0.0	*34*	0.0	*60*			0.0	*72*		
Leuk	0.0	*34*	0.0	*38*			0.0	*72*		-100.0%
Cologne	0.0	*34*	0.0	*43*						-100.0%
Dortmund	0.0	*34*	0.0	*44*						-100.0%
Antwerp	2.2	*18*	1.7	*17*					-100.0%	-100.0%
Strasbourg	3.8	*15*	2.0	*16*					-100.0%	-100.0%
Portsmouth	0.0	*34*	0.0	*60*						
Plovdiv	0.0	*34*	0.0	*53*						-100.0%
Gdansk	0.0	*34*	0.0	*37*						-100.0%
Lyon	0.2	*29*	0.1	*31*					-100.0%	-100.0%
Tartu	0.0	*34*	0.0	*60*						
Karlsruhe	0.0	*34*	0.0	*60*						
San Marino	0.0	*34*	0.0	*58*						-100.0%
Split	0.0	*34*	0.0	*60*						
Bijeljina										
Dresden										
Lille										
Innsbruck										
Katowice										
Lodz										
Salzburg										
Varna										
Oradea										
Rijeka										

E.c. = European cities, a.c. = all cities

Data sources: Telegeography 2007, Author's calculations

Chapter 5

Internet backbone networks and aviation networks: a comparative study

5.1 Introduction

This chapter compares the topology and the emerging geography of two infrastructural networks: the Internet backbone and the aviation network across European cities. The aim for such an analysis is not only to explore these networks' topological and spatial pattern in Europe, but also to investigate the existence of any similarities in the way the Internet and the aviation network interconnect the nodes of the European urban network.

Such a comparative analysis is crucial in the context of this doctoral research for two reasons. Firstly, as was discussed in Chapter 2, the Internet is an infrastructural network itself, which also shares analytical similarities with other infrastructural networks, such as the aviation network. Both of them facilitate the modern economy in a similar way: the Internet backbone network transports the informational goods of the modern economy while the aviation network transports the physical products but mostly the main actors of the knowledge economy, the people, across the distributed centers of production and consumption. In addition, they share topological similarities. Both of them are rolled out as spatial nonplanar networks. This refers to networks with specific physical footprints whose main attribute is that their edges can cross without forming a node. On the contrary, for the case of planar networks such as the motorway network, the crossing point of any two edges becomes automatically a new network node (Gorman and Kulkarni 2004).

Secondly, from a more conceptual perspective, both of these networks facilitate and enhance interaction among cities around the globe. Castells' (1996, 417) space of flows reflects on the importance of these infrastructural networks as a structural element of the global city:

> "The global city is not a place but a process. A process by which centres of production and consumption of advanced services, and their ancillary local societies, are connected in a global network, while simultaneously downplaying the linkages with their hinterlands, on the basis of informational flows."

Global cities mainly exist because technology has enabled interaction among the remote centres of production and consumption. Both the Internet backbone and the aviation networks carry a significant part of this interaction. Information as well as knowledge is being distributed around the world settlements through what are known as digital highways. In the same way, people are being brought together via the aviation network in order to interact and acquire complex knowledge (Rimmer 1998). Both networks diminish the importance of traditional barriers such as national borders, but at the same time highlight the locational advantage of being part of these networks.

Apart from the above twofold similarity between the two infrastructural networks, this study is also justified as an attempt to bridge the gap between the theoretical sophistication in the work of Sassen (2000) and Castells (1996) and others in the field of the world cities research and the lack of empirical evidence to back up their claims concerning emerging networks of flows. The above is illustrated by Peter Taylor (2004) as an ‘evidential crisis’ in the burgeoning field of world cities research. In particular, Taylor highlights the surprisingly limited use of relational data in the key studies in the field, given that it is precisely relations between cities that constitute the key to understanding the new world city networks that analysts contend are emerging. In recent years, attempts to tackle this gap have been advanced, including the work of Taylor and others in the Globalisation and World Cities (GaWC) network (see Taylor, 2004, for an account of this work), in which inter-locking networks of advanced producer service firms constitute the relational data. In addition there are a number of studies which attempt to identify the relations between cities based on airline networks (e.g. Derudder and Witlox 2005), but also based on Internet networks (e.g. Moss and Townsend 2000, Townsend 2001a and 2001b, Rutherford et al 2004, Rutherford forthcoming). Despite the above analogies, there are no comparative studies on the way these networks facilitate the world city phenomenon, with Choi et al (2006) being the only exception. However, because of the global scale of this study, its aim is rather different than the one adopted here. More specifically, while Choi et al (2006) focused on the global extent of these networks, this doctoral thesis is focused on the European part of these networks, while taking into consideration their global connections. This choice of scale enables us to draw more detailed geographical conclusions.

For this comparative analysis a variety of methods are used: from simple statistical analysis to social network analysis, complex network theory and Quadratic Assignment Process (QAP). What is also important is that this comparison does not take place on an abstract topological space, but on real geographical space, using the

European cities as the networks' nodes. For the needs of this analysis the Internet backbone capacity and air passengers have been aggregated at the city level.

The chapter goes as follows: the next section is dedicated to analysing the data which is used and then a descriptive statistical analysis takes place; after that network statistics such as centralization and centrality indicators are analyzed in a comparative way; in the next part the two infrastructural networks are analyzed using the complex network approach; the next section attempts to further explain the results of the comparative analysis using network economics and engineering argumentation and focusing on the disaggregated-micro level of network carriers; and this chapter concludes by highlighting the difference in the way the two infrastructural networks interconnect the European cities.

5.2 Data description

In order to compare the structure of the infrastructural networks, two different datasets are used in this chapter. The first one represents the Internet backbone networks in Europe (Telegeography 2007) and was described in detail in the previous chapter. The second dataset refers to the international airline connections for European cities (ICAO 2008). To give an example, just like the Internet backbone data, links between London and Paris as well as links between London and New York are included in the analysis, but links between London and Manchester or links between New York and Tokyo are excluded. Again, two versions of both networks are included in the analysis: the networks of binary links and the networks of weighted links, with the former being a derivative of the latter for the needs of the analysis. Just like in Chapter 4, the binary networks represent only the existence of a link between any two cities, both in the Internet backbone and in the aviation network. What is slightly more complicated for this comparative analysis is the weighted versions of the networks. Regarding the Internet backbone network, the weights represent the capacity of the intercity linkages as analyzed in Chapter 4. In order to enter weights in the aviation edges, the annual intercity passenger flows for the years 2001-2006 are used. This indicator is widely used in aviation literature in general and more specifically in the part of this literature which uses the aviation networks in order to illustrate the world cities network (Choi et al 2006, Derudder and Wiltox 2008 and 2005, Matsumoto 2007, Lee 2008). In addition, this measure is also widely used in the part of the complex network literature which deals with aviation networks (Gastner and Newman 2005, Guimera et al 2005, Amaral et al 2000). This indicator fits well with the needs of this research because it represents the intercity flows that support the global city. However, there is a significant conceptual difference between the two metrics used here for the comparison of the weighted versions of the two networks. While bandwidth represents the capacity of the installed infrastructure, the passenger volume represents the usage of the service. In other words, while the former represents the supply, the latter represents the demand for a specific service. Nonetheless, it could not be otherwise simply because of data availability problems. More specifically, it is very difficult to obtain data for intercity Internet packet volumes especially at this scale, in order to compare the usage of both infrastructural networks[20]. The reason for this is the ISPs' reluctance of publish such data for competition reasons, although they collect it in order to manage their own networks. An alternative way of approaching the intercity Internet data links but not the actual intercity flows is by surpassing the ISPs with the use of specific programs called traceroutes, which map the route that a data packet travels through the different nodes in order to reach its final destination (Dodge and Kitchin 2000). However, the main drawback of such a process is that it usually results to an ego-centric network consisting of the routes that a data packet travels to reach a number of destinations only from a specific location, which is usually the traceroute's host location ignoring the overall intercity data flows. Such an approach does not fit with the needs of this research because the overall image of the European cities Internet connectivity could not be approached by such a process.

Another difference between the aviation network and the Internet backbone is that the former is directional while the latter is not. This means that between London and Paris there are two different edges for the aviation network, London-Paris and Paris-London, contrary to what applies for the Internet backbone network. Or in simple terms, the passenger flows between the links London-Paris and Paris-London are different. In order to eliminate this difficulty and enable the comparison with the undirected network of the Internet backbone links, the directed aviation network was converted to an undirected one by symmetrising its edges using the maximum value, following Choi's et al (2006) methodological choice. This means the passenger flows

---

[20] One example of the use of such data is the NYTE project (Ratti et al 2008). However, it only includes the Internet data of one specific ISP between New York and the rest of the world.

for the links London-Paris and Paris-London were defined to be equal to the highest value between the two opposed edges and only one of those edges was included in the network. In addition, in order to ‘clean’ the initial dataset, intercity links with less than 1000 passengers were excluded from the analysis because such data entries represent non-commercial and non-scheduled flights.

Just like for the analysis of the Internet backbone network, the intra-European aviation network was extracted from the initial dataset. This includes the international edges only among the European cities. What is more, in order to further compare the two networks, a third extraction for both of them was created. This network contains only the links among the 62 European cities which are present in both networks. So, this third extraction enables us to study how these two networks are deployed among the same cities and monitor the network characteristics of the 62 common cities through the six year period. However, just as explained in Chapter 4, the process of extracting a subset from a network imposes some limitations on the analysis. For both networks the subset of the 62 common cities do not represent an independent standalone network but a theoretical assignment in order to better compare the structure of the two networks.

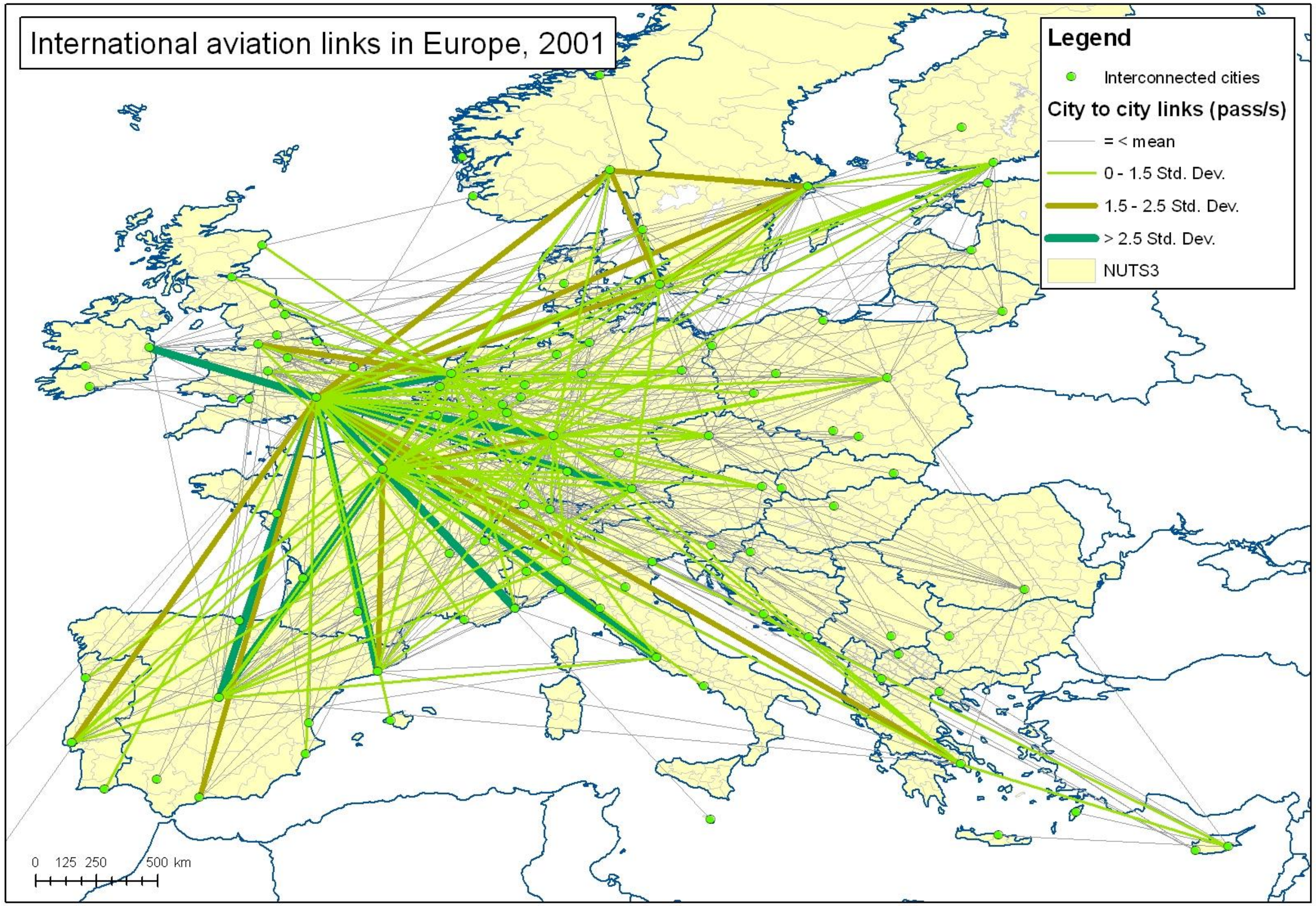
5.3 Descriptive statistical analysis

The first step in the process of the networks’ structural comparison is the visualization of the actual networks and the descriptive analysis of their edges in comparison with the Internet backbone network, as it was presented in Chapter 4. Figures 5.1 and 5.2 present the intra-European intercity aviation links for the years 2001 and 2006. From a first glance it is obvious that there is a different structure in comparison with the Internet backbone network. While the most important IP edges (capacity greater than 2.5 standard deviations above the mean) formed a ring among London, Paris, Amsterdam and Frankfurt, for the aviation network no such pattern emerges. On the contrary, the busiest links (passenger volumes greater than 2.5 standard deviations above the mean) are displayed in the form of star networks around nodal cities such as London and Paris. Furthermore, the main difference between the Internet backbone network between 2001 and 2006 was the expansion of the high capacity links towards Central and Eastern Europe, signaling these countries’ late entrance in these infrastructural networks. However, this is not the case for the aviation networks; between 2001 and 2006 the airline network did not change rapidly. Even at this scale, there are some visible differences in density and in the diffusion of the high volume edges, but from these maps no geographical pattern emerges, indicating the maturity of this network contrary to the still evolving character both in geographical and technological terms of the Internet backbone network.

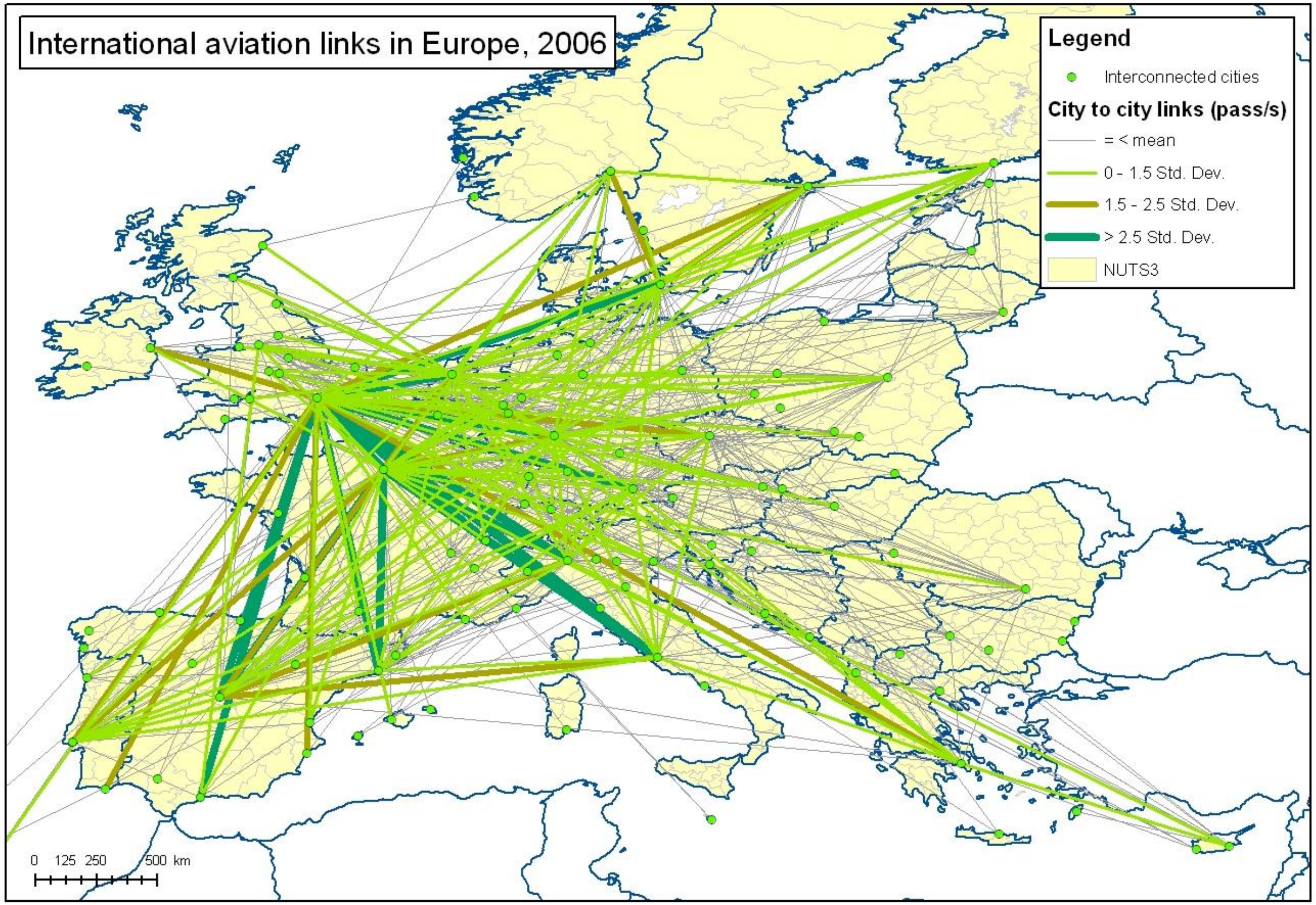
The above comments are also demonstrated by Table 5.1, which presents the descriptive statistics of the edges of the intra-European and global network. Contrary to the case of the Internet backbone links, the weights of the aviation network edges remain rather stable during the six year period. More specifically neither the maximum value nor the mean changed significantly through time. The only exception is the years 2001 and 2002, when the volumes were decreased probably because of the 11/9/2001 terrorist attacks against the World Trade Centre and the Pentagon in USA.

Regarding the standard deviation, just like the Internet backbone network, the passenger flows are more dispersed among the network's edges over time. In addition, the use of the coefficient of variation enables us to directly compare the two infrastructural networks. By this indicator it can be said that the network weights are more dispersed in the aviation network than in the Internet backbone one. In other words, the passenger flows are more equally allocated across the network's edges for the case of the aviation network than the capacity for the Internet backbone network. This however, has nothing to do with the role of the cities in the networks; this is the subject of the centrality indicators which are presented later in this chapter.

The comparison is even more observable in the lower part of Table 5.1, where the descriptive statistics of the edges among the 62 common cities for the two infrastructural networks are presented. What we gain by focusing on this extraction is that the number of nodes is the same between the two networks so the comparison between two same size networks is more legible. Again, while the average bandwidth was steadily increased through time, the change in the average passenger volume is not increased in a stable way and in any case the increase is much smaller. Regarding the coefficient of variation the decrease is fairly stable for the aviation network but not for the Internet backbone network.



Figures 5.1: International aviation links in Europe, 2001



Figures 5.2: International aviation links in Europe, 2006

Table 5.1: Descriptive statistics for the passenger volumes of the aviation links

Extent	Year	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
Aviation network	2001 a.c.	1124	1870166	104017	154631	1.49
	2002 a.c.	1020	1919814	98178	151768	1.55
	2003 a.c.	1015	1716223	99392	147432	1.48
	2004 a.c.	1077	1773816	106502	151980	1.43
	2005 a.c.	1027	1896055	117954	158142	1.34
	2006 a.c.	1021	1785174	123335	163163	1.32
	2001 E.c.	1132	1870166	113457	172651	1.52
	2002 E.c.	1212	1919814	107452	170465	1.59
	2003 E.c.	1015	1659316	105903	161812	1.53
	2004 E.c.	1077	1679225	110401	163074	1.48
	2005 E.c.	1039	1571293	117756	162109	1.38
	2006 E.c.	1021	1566286	129265	171475	1.33
	2001 c.c.	1132	1870166	128566	192126	1.49
	2002 c.c.	1212	1919814	120112	187546	1.56
	2003 c.c.	1015	1659316	118569	176454	1.49
	2004 c.c.	1077	1679225	121654	175448	1.44
	2005 c.c.	1039	1571293	134230	176480	1.31
	2006 c.c.	1021	1566286	146592	186202	1.27
Internet backbone network	2001 c.c.	1.5	58641	3270	8718	2.67
	2002 c.c.	2	65041	4578	9809	2.14
	2003 c.c.	4	96870	6481	14279	2.20
	2004 c.c.	2	153529	9073	21196	2.34
	2005 c.c.	2	240952	13089	30167	2.30
	2006 c.c.	2	305169	17611	40523	2.30

a.c. = all cities; E.c. = European cities; c.c = common cities
Data sources: Telegeography 2007, ICAO 2008

Even more interesting are the descriptive statistics not only for the edges but for the whole networks. Table 5.2 presents the basic network statistics for the aviation network for all the different extents but also for the 62 common cities in order to compare them with the Internet backbone. Through time, the aviation network has slowly increased and at any case much slower than the Internet backbone network. Again, the increase is not stable and between some years there is a decrease in the number of nodes or edges. In general, the aviation network is more expanded than the Internet backbone one. The number of its nodes and its edges is almost double the number of the nodes and edges of the backbone network for all the years and both for the intra-European network as well as for the network of all the cities. Regarding density though, the aviation network appears to be denser for the global extent while the Internet backbone density is greater for the intra-European network.

Table 5.2: Basic network statistics

Network	**Year**	**Vertices**		**Edges**		**Density**	
			change (%)		***change (%)***		***change (%)***
Aviation network	2001 a.c.	228		812		0.031	
	2002 a.c.	243	*6.6%*	883	*8.7%*	0.030	*-4.3%*
	2003 a.c.	238	*-2.1%*	881	*-0.2%*	0.031	*4.0%*
	2004 a.c.	239	*0.4%*	900	*2.2%*	0.032	*1.3%*
	2005 a.c.	225	*-5.9%*	842	*-6.4%*	0.033	*5.6%*
	2006 a.c.	230	*2.2%*	856	*1.7%*	0.033	*-2.7%*
	overall change		*0.9%*		*5.4%*		*3.6%*
	2001 E.c.	105		415		0.0760	
	2002 E.c.	107	*1.9%*	435	*4.8%*	0.0767	*0.9%*
	2003 E.c.	107	*0.0%*	445	*2.3%*	0.0785	*2.3%*
	2004 E.c.	105	*-1.9%*	456	*2.5%*	0.0835	*6.4%*
	2005 E.c.	107	*1.9%*	441	*-3.3%*	0.0778	*-6.9%*
	2006 E.c.	110	*2.8%*	452	*2.5%*	0.0754	*-3.0%*
	overall change		*4.8%*		*8.9%*		*-0.8%*
	2001 c.c.	62		308		0.1629	
	2002 c.c.	62	*0.0%*	323	*4.9%*	0.1708	*4.9%*
	2003 c.c.	62	*0.0%*	331	*2.5%*	0.1750	*2.5%*
	2004 c.c.	62	*0.0%*	344	*3.9%*	0.1819	*3.9%*
	2005 c.c.	62	*0.0%*	322	*-6.4%*	0.1703	*-6.4%*
	2006 c.c.	62	*0.0%*	335	*4.0%*	0.1772	*4.0%*
	overall change		*0.0%*		*8.8%*		*8.8%*
Internet backbone network	2001 c.c.	62		153		0.081	
	2002 c.c.	62	*0.0%*	144	*-5.9%*	0.076	*-5.9%*
	2003 c.c.	62	*0.0%*	156	*8.3%*	0.082	*8.3%*
	2004 c.c.	62	*0.0%*	172	*10.3%*	0.091	*10.3%*
	2005 c.c.	62	*0.0%*	174	*1.2%*	0.092	*1.2%*
	2006 c.c.	62	*0.0%*	177	*1.7%*	0.094	*1.7%*
	overall change		*0.0%*		*15.7%*		*15.7%*

a.c. = all cities; E.c. = European cities; c.c = common cities
Data sources: Telegeography 2007, ICAO 2008

In order to better compare the two infrastructural networks, the focus now is on the networks between the 62 common cities. In general, the aviation network for these cities has almost double the number of edges in comparison with the Internet backbone network. However, the latter is being increased faster than the former. In the six year period the number of backbone edges increased by 15.7% while for the aviation network the increase only reached 8.8%. Of course, the same pattern applies to the density as well, simply because density is only affected by the number of the edges and nodes, as was explained in Chapter 4. Through time the increase in the number of edges and in density is almost stable with the exception of the decrease between the years 2001-2002 for the aviation network, which apparently reflects the 9/11 terrorist attack in US.

5.4 Network centralization

In this section the overall network centralization measures are presented. Just like in Chapter 4, four network centralization indicators are included in this analysis: degree centralization, betweenness centralization for the binary network and eigenvector centralization both for the weighted and the binary networks. Table 5.3 presents these measures both for the aviation and the Internet backbone network in order to compare the different infrastructural networks.

Table 5.3: Centralization indicators

Year	**Degree (b)**		**Betweenness (b)**		**Eigenvector (b)**		**Eigenvector (w)**	
	Av.	**I.b.**	**Av.**	**I.b.**	**Av.**	**I.b.**	**Av.**	**I.b.**
2001 a.c.	65%	43%	52%	46%	44%	54%	81%	84%
2002 a.c.	61%	43%	46%	48%	43%	55%	81%	87%
2003 a.c.	62%	47%	45%	51%	43%	57%	81%	90%
2004 a.c.	62%	49%	46%	47%	42%	59%	80%	91%
2005 a.c.	61%	50%	41%	46%	42%	59%	79%	93%
2006 a.c.	61%	50%	41%	46%	42%	60%	76%	89%
overall change	*-5%*	*7%*	*-11%*	*-1%*	*-2%*	*6%*	*-5%*	*6%*
2001 E.c.	56%	34%	37%	21%	38%	43%	81%	76%
2002 E.c.	54%	35%	36%	25%	36%	47%	81%	81%
2003 E.c.	54%	36%	35%	25%	36%	45%	81%	80%
2004 E.c.	56%	46%	37%	27%	36%	50%	80%	86%
2005 E.c.	54%	39%	33%	21%	36%	47%	77%	86%
2006 E.c.	53%	41%	32%	29%	36%	47%	74%	84%
overall change	*-3%*	*7%*	*-5%*	*8%*	*-2%*	*4%*	*-7%*	*8%*
2001 c.c.	53%	31%	16%	12%	33%	42%	81%	76%
2002 c.c.	50%	34%	13%	13%	31%	46%	80%	81%
2003 c.c.	50%	37%	12%	17%	31%	46%	79%	80%
2004 c.c.	52%	47%	13%	17%	31%	50%	78%	86%
2005 c.c.	48%	41%	14%	15%	30%	47%	75%	87%
2006 c.c.	48%	43%	15%	17%	29%	48%	72%	84%
overall change	*-5%*	*12%*	*-2%*	*6%*	*-4%*	*6%*	*-9%*	*8%*

a.c. = all cities; E.c. = European cities; c.c = common cities
Av. = Aviation network; I.b. = Internet backbone network; b = binary; w = weighted
Data sources: Telegeography 2007, ICAO 2008

In regards to the degree centralization, the aviation network is for all the different extents and for the whole six year period more centralized than the Internet backbone one. This is not the case though for the other centralization measures. According to most of them the Internet backbone network is clearly more centralized. In order to explain this, we need to recall the definitions of the different centralization indicators. Degree centralization is the only indicator presented here which is based only on direct connections. On the contrary, the betweenness and eigenvector centralization measures take also into consideration the indirect connections and this is why they are suitable for approaching the centralization of a network such as the Internet backbone, which is based on data packet movements among different nodes, as was explained in Chapter 4. Regarding the aviation network, indirect connections also play a significant role, but in reality geodesics longer than 3 are not efficient at all because they mean that in order to fly from a specific origin to the final destination, a passenger should fly through more than two airports! So, the aviation network is more centralized than the Internet backbone one according to the degree centralization because of the greater importance of direct connections for airline transportation, while the Internet backbone is more centralized according to the betweenness and the eigenvector centralization because it is built on a topology convenient to indirect communications. The only exception on the above is that the aviation network appears to be more centralized for the case of the intra-European network, indicating a more dispersed structure for the Internet backbone network for this extent, which is not the case of the network of the 62 common cities.

The Internet backbone network appears to be more centralized than the aviation one when the weights are taken into consideration. For almost all the cases, eigenvector centralization for the IP network is greater than the same measure for the aviation network. This can also be visually explained by the maps presented above (Figures 4.1, 4.2, 5.1 and 5.2). The two infrastructural networks studied here are aggregations of hundreds of different usually private owned and developed but also interconnected networks. For the case of the aviation network, the result of the aggregation process is a summary of different usually star-like networks, which indicate the existence of different hubs as well as the hub and spoke structure of these networks. So, the highest flows, as they are presented in Figures 5.1 and 5.2 are allocated among a few hub cities (mainly London and Paris) and their spoke cities. The result of this process is that a number of cities (i.e. the main hubs and their spokes) are served by high passenger flows. On the contrary, the links with the highest capacity of the Internet backbone network are allocated among only a handful of cities, creating a ring between them as can be seen by the Figures 4.1 and 4.2. And this is an indication of a more centralized network, as can be verified by the eigenvector centralization for the weighted network. However, at a disaggregated level as is analysed later, the network of an Internet backbone provider would be less centralized than the hub and spoke network of an air-carrier.

Looking at the changes through time, the two infrastructural networks seem to follow completely different trends. For almost all the cases the centralization indicators are being decreased through time for the aviation network and increased for the Internet backbone one. In other words, the aggregated network of the different Internet backbone providers is moving towards a more centralized structure, which means that regardless of the increase of the interconnected cities, fewer cities through time perform a central role in this network. However, the aviation network appears to be less centralized by the end of the six year study period. The new destinations which were added during the six year period resulted in a decrease in the overall centralization.

To sum up, the aviation network appears to be more centralized as regards the degree distribution, but when the indirect links as well as the actual weights are taken into consideration the Internet backbone network appears to be more centralized. In addition, this tendency is being increased through time contrary to what applies for the aviation network. However, it should be noted here that the above findings are only valid for the aggregated networks, which are the summary of all the different interconnected networks. At a disaggregated level the different networks might perform differently.

5.5 Cities’ centrality indicators

No matter how important is the network’s overall centralization even more important are the centrality measures for individual cities. The value of such measures enable us to distinguish the different roles that cities play in such infrastructural networks. For instance, the analysis of the centrality indicators of the previous chapter brought out the gateway roles of some cities, the intra- or extra-European importance of some other cities etc. The comparative analysis of the centrality measures in this section has as its main goal the identification of the varying roles that different cities play in those two infrastructural networks. In other words what is expected to come out from this analysis is whether the European cities perform the same way or not with respect to the two different infrastructural networks.

5.5.1 Degree centrality

The first measure analyzed here is the degree centrality. As mentioned before, degree centrality both for the binary and the weighted network is a measure of infrastructural capital. Regarding the former, degree centrality represents the amount of edges (backbone links – airline connections) through which a city is served while regarding the latter it represents the aggregated volume (passenger flows) or the aggregated capacity (bandwidth) of the edges linked to a city. Table 5.4 presents the degree centralities for the aviation and the Internet backbone network, both for the weighted and the binary networks and for the years 2001 and 2006. The centralities here have been calculated taking into consideration the edges between all cites (i.e. the extra-European links are included). However, in order to enable the comparison, the degree centralities are presented only for the 62 common European cities present in both networks. The whole table can be found in the Annex of this chapter and Table 5.4 only focuses on the 30 most central cities. In addition Figures 5.3 and 5.4 present comparatively the centralities for the two infrastructural networks both for the binary and the weighted versions. In order to present centralities for both networks in the same map, they are presented as standard deviations from the mean centrality just as in the previous chapter.

The first observation from this table is that the cluster of the four main cities performs similarly in the two different infrastructural networks. No matter if the weighted or the binary links are taken into consideration, London, Paris, Frankfurt and Amsterdam are the most central nodes in both networks. What is interesting though is London's greater dominance in the weighted networks. While the second most central cities in both binary networks have almost similar centrality measures (66-65 and 70-72), the differences between London and Paris, which is always the second most central city in the weighted networks, is much greater (76-50 and 66-57). As was highlighted in the previous chapter, while the geography of the binary links can be explained by proximity, physical and political geography, the spatial pattern of the weighted links seems to be more related with economic geography and for this case with London’s superiority as a global city and world financial centre.

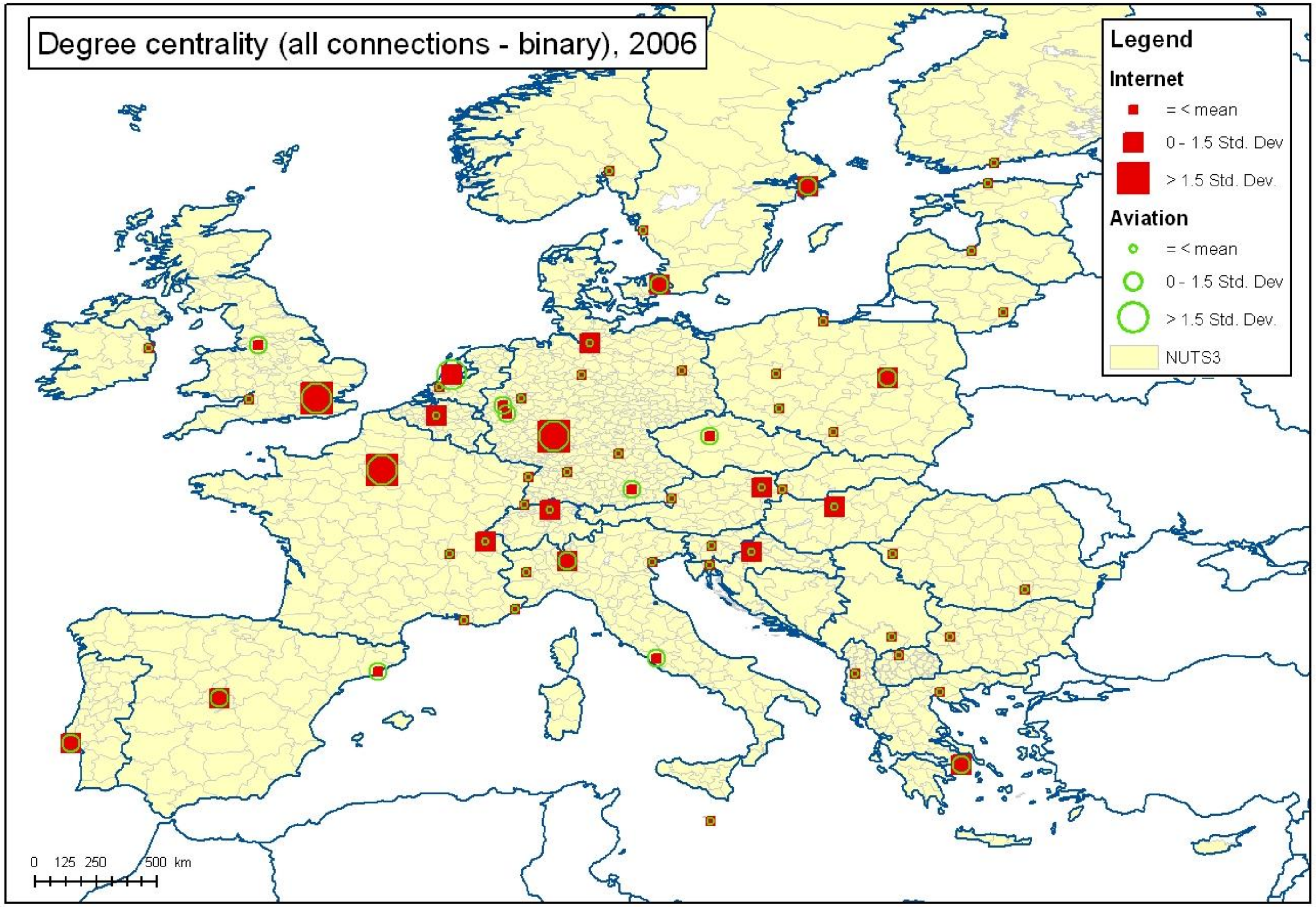
Table 5.4: Degree centrality

	Binary networks								**Weighted networks**							
	Internet 2001		Aviation 2001		Internet 2006		Aviation 2006		Internet 2001		Aviation 2001		Internet 2006		Aviation 2006	
London	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*
Paris	66	*2*	65	*2*	49	*3*	72	*2*	76	*2*	50	*2*	66	*2*	57	*2*
Frankfurt	45	*3*	58	*3*	70	*2*	71	*3*	68	*4*	37	*4*	58	*3*	43	*3*
Amsterdam	40	*4*	51	*4*	32	*4*	54	*4*	72	*3*	38	*3*	50	*4*	38	*4*
Madrid	11	*18*	25	*7*	15	*12*	37	*6*	9	*9*	19	*5*	16	*7*	24	*5*
Milan	24	*5*	10	*18*	25	*6*	39	*5*	13	*8*	5	*17*	13	*9*	19	*6*
Copenhagen	17	*12*	27	*6*	17	*9*	26	*9*	18	*7*	14	*6*	18	*6*	15	*7*
Munich	18	*10*	21	*8*	6	*25*	31	*7*	4	*17*	9	*9*	3	*22*	14	*8*
Rome	2	*38*	9	*23*	1	*43*	21	*10*	1	*28*	6	*15*	0	*40*	14	*9*
Barcelona	6	*27*	17	*12*	7	*22*	20	*12*	2	*24*	9	*10*	2	*24*	11	*10*
Prague	13	*13*	18	*11*	9	*19*	30	*8*	5	*16*	5	*18*	5	*18*	9	*11*
Stockholm	23	*7*	21	*8*	19	*7*	17	*13*	25	*6*	11	*7*	20	*5*	9	*12*
Lisbon	18	*10*	10	*21*	16	*10*	17	*13*	1	*25*	7	*14*	2	*25*	9	*13*
Athens	12	*15*	14	*14*	16	*10*	15	*16*	1	*33*	7	*11*	3	*23*	8	*14*
Manchester	1	*44*	21	*8*	1	*43*	21	*10*	0	*34*	7	*12*	0	*40*	7	*15*
Oslo	22	*8*	10	*18*	7	*22*	12	*20*	9	*10*	7	*13*	5	*17*	6	*16*
Helsinki	11	*18*	12	*15*	5	*29*	12	*19*	3	*18*	6	*16*	5	*19*	5	*17*
Warsaw	10	*22*	17	*12*	10	*16*	17	*13*	1	*31*	3	*23*	6	*15*	5	*18*
Berlin	7	*24*	5	*36*	1	*43*	12	*20*	3	*21*	2	*25*	0	*43*	4	*19*
Düsseldorf	4	*33*	10	*18*	6	*25*	13	*18*	5	*15*	4	*20*	8	*12*	4	*20*
Geneva	11	*18*	6	*29*	10	*16*	6	*33*	7	*14*	3	*24*	8	*13*	4	*21*
Brussels	22	*8*	12	*15*	12	*14*	8	*28*	32	*5*	4	*21*	12	*11*	3	*22*
Cologne	2	*38*	8	*24*			14	*17*	0	*45*	1	*32*			3	*23*
Bucharest	7	*24*	10	*21*	7	*22*	12	*20*	1	*32*	2	*29*	1	*28*	3	*24*
Dublin	5	*30*	11	*17*	5	*29*	6	*31*	2	*23*	4	*19*	4	*20*	3	*25*
Vienna	24	*5*	5	*34*	28	*5*	6	*31*	7	*13*	2	*27*	14	*8*	3	*26*
Stuttgart	1	*44*	7	*27*	9	*19*	10	*24*	0	*36*	2	*31*	0	*36*	3	*27*
Hamburg	12	*15*	6	*29*	12	*14*	11	*23*	8	*12*	2	*26*	13	*10*	3	*28*
Zürich	13	*13*	41	*5*	14	*13*	9	*25*	8	*11*	10	*8*	7	*14*	2	*29*
Lyon	2	*38*	7	*27*			8	*28*	2	*22*	2	*28*			2	*30*

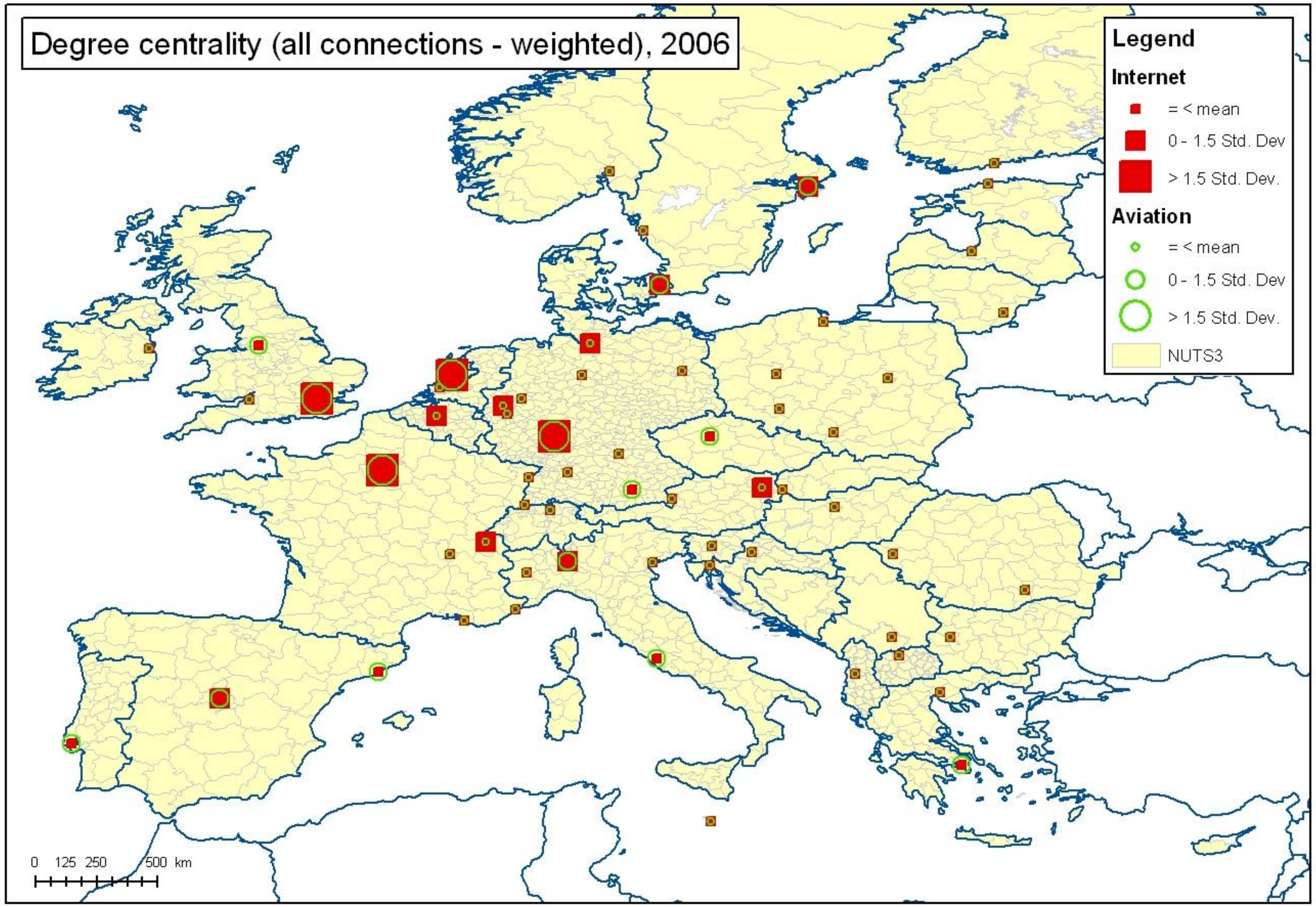
Data sources: Telegeography 2007, ICAO 2008

However, the main differences with the Internet network are observed not in the top four cities, but in the next tier of cities, as can also be seen by the maps in Figures 5.3 and 5.4. Madrid is the 5th city regarding the passenger flows and is followed by Milan and Copenhagen. Madrid is higher in the aviation hierarchy than in the Internet backbone one. However, lately its position in the informational network has been improved probably because of the upgrade of the link with London, as can be seen by Figure 4.1 and 4.2. The Danish capital manages to maintain almost the same position in both infrastructural networks both for the valued and the binary versions. Milan on the other hand, appears to be more central for the binary version of the Internet backbone network, but when the analysis moves to the capacity and the flows, its position is more central for the aviation network. Rome is another city which seems to be preferred by the international aviation carriers. Contrary to the fact that Rome was not directly connected with any international backbone Internet network in 2006 as Milan was the country's joint point with the global network, it is the 9th most central city in 2006 concerning the weighted aviation network. In addition, cities such as Munich, which is part of Europe's pentagon, as well as some famous tourist destinations such as Barcelona, Prague and Lisbon gain positions compared to the Internet bandwidth centrality ranking. On the other hand, cities out of Europe's pentagon such as Stockholm, Vienna and Bratislava and Budapest as well as the more core ones such as Hamburg, Brussels, Düsseldorf, Geneva and Zurich (only for 2006) seem to be more central in the international Internet backbone network than in the international aviation one. In addition, it is worth mentioning that a few of the 'sub-dominant' cities within national spaces (e.g. Manchester with respect to London, but also Barcelona, Munich, Rome) seem to have lower Internet rankings than aviation, presumably on the basis that much of the international Internet traffic passes through the major city.

In general, it could be said that regarding the cities' ranking based on the binary connections, it follows more or less the weighted one with very few exceptions such as Geneva. This case is fairly interesting as Geneva is less central in terms of binary connections than it is when the passenger flows are taken into consideration. Geneva gains 12 places in the relevant ranking when the passenger flows are taken into consideration, reflecting fewer but more crowded air routes. The fact that there are less discontinuities between the degree centrality ranking for the weighted and the binary aviation network, like these occurred for the Internet network with cities with high centralities for the weighted and low centralities for the binary networks (or the reverse), is not surprising. Airline hubs with a large number of air connections and small number of passengers would just not be feasible for the private air carriers because of the high maintenance costs and the actual low payoff of the low passenger flows. On the other hand, this discontinuity reflects the ease of Internet backbone providers to upgrade their networks and to adjust their capacity provision to the demand.



Figures 5.3: Degree centralities based on the binary links and all the cities, 2006



Figures 5.4: Degree centralities based on the weighted links and all the cities, 2006

London’s global character is reflected in the aviation network as well. As is stated in Table 5.5, in 2006 almost half of the passengers travelling through, from or to London (55% of all connections) had a non European origin or destination. Table 5.6 gives a more detailed view for the most extroversive cities. The most popular origins (or destinations since the data is symmetrised) for London are North American cities (24% of the passengers and 17% of connection) followed by Asian and Pacific cities (17% and 25% respectively). In addition, its dominance for almost all the destinations is obvious. For instance, 38% of all the passengers travelling between Europe and US and Canada, travel through, from or to London. The same almost applies for the binary links. Nonetheless, just like the Internet backbone networks, its dominance is smaller in comparison to the actual flows. In other words, the binary links are more widely distributed in Europe than the passenger flows. However, this inconsistency between binary and weighted connections is not that evident for the aviation network for reasons explained above.

Yet, the most extroversive city is Frankfurt, since 54% of all its passengers and 64% of all its links are extra-European. Frankfurt seems to have a different role as a hub city for Asia since 34% of all its binary connections and 23% of all its passenger flows had an origin or a destination in Asia and Pacific region. In terms of dominance, it is the second most important hub for this continent after London. Surprisingly enough, the financial capital of Germany seems to have a distinctive connection with Asia and Pacific region in general: for both infrastructural networks Frankfurt is the second most important European hub for this region regarding the binary network and also Asia and Pacific is the region outside Europe which shares the most links.

In general, apart from the four most central cities, the rest of the cities appear to be more introversive. For all of them, the percentage of extra-European travel flows is less than 30%. However, Lisbon and Madrid are again interesting cases since both of them have important links with the Latin America and Caribbean region. In 2006, 25% of Lisbon’s binary connections and 13% of the passenger loads were with Latin America and 26% of Madrid’s links and 22% of its passenger flows were with this region as well. In terms of dominance, Madrid is the main gateway city for this region both in terms of binary and weighted links. In comparison with the Internet backbone network, the importance of the ties between Madrid and Lisbon and Latin America and Caribbean remains or even increases both in terms of dominance as European hubs but also as a proportion of these cities' total links and flows. However, what is decreased in comparison to the Internet backbone network is the share of these cities' binary and weighted links with Africa. In the Internet backbone network London is the dominant city for communications with Africa. However, the share of Lisbon's and Madrid's binary links with this continent was high (56% and 20% respectively). Regarding the aviation network though, London and Paris are both the main hubs and also have the highest proportions of links with this continent, decreasing this way the role of Lisbon and Madrid in communications with Africa. All in all, Lisbon and Madrid have distinctive roles as gateway cities with Latin America in both infrastructural networks, but they only have such roles with Africa in the Internet backbone.

Table 5.5: Percentage of extra-European connections

City	**Internet 2006**		**Aviation 2006**	
	Weighted	Binary	Weighted	Binary
Frankfurt	11%	49%	54%	64%
London	36%	67%	47%	55%
Paris	19%	57%	40%	48%
Manchester	0%	0%	34%	29%
Amsterdam	17%	31%	31%	43%
Madrid	4%	47%	29%	39%
Nice	0%	0%	26%	50%
Brussels	0%	33%	25%	25%
Munich	0%	17%	24%	33%
Milan	6%	32%	23%	36%
Zürich	3%	14%	20%	31%
Copenhagen	15%	35%	18%	24%
Lisbon	4%	69%	18%	38%
Rome	0%	0%	18%	35%
Marseille	0%	0%	13%	25%
Athens	21%	25%	13%	32%
Dublin	4%	20%	13%	11%
Düsseldorf	8%	17%	9%	16%
Berlin	0%	0%	9%	18%
Ljubljana	0%	0%	9%	17%
Oslo	9%	29%	8%	18%
Cologne	100%	100%	8%	20%
Helsinki	0%	0%	7%	22%
Hamburg	3%	25%	6%	19%
Bucharest	0%	14%	6%	12%
Stockholm	8%	32%	6%	17%
Barcelona	0%	0%	5%	24%
Riga	2%	20%	5%	9%
Warsaw	1%	20%	5%	13%
Prague	0%	0%	5%	9%

Data sources: Telegeography 2007, ICAO 2008, Author's calculations
The rest of the 62 common cities are presented in Annex of this chapter

It should also be highlighted that the centrality of some cities such as Vienna, Geneva and Budapest for the aviation network has been underestimated due to the exclusion of some airlines from the ICAO dataset, such as Austrian Airlines. However, the spatial pattern of aviation centrality as it is presented here remains a good approximation of the reality.

Table 5.6: Geographic allocation of the aviation edges of the most extroversive cities (%)

	Regions	London		Paris		Amsterdam		Frankfurt		Brussels		Madrid		Munich		Lisbon	
weighted	Africa	4	*39*	4	*25*	1	*3*	4	*16*			1	*2*			1	*1*
	Asia & Pacific	17	*36*	14	*17*	14	*11*	23	*21*			0	*0*	13	*4*		
	Europe	53	*16*	60	*10*	69	*8*	46	*6*	75	*1*	71	*5*	76	*3*	82	*2*
	Latin America & Caribbean	1	*7*	3	*15*	1	*3*	2	*9*			22	*50*			13	*10*
	Rest of Europe	1	*17*	1	*10*	2	*9*	2	*11*			0	*2*	1	*2*	1	*1*
	U.S. & Canada	24	*38*	18	*16*	14	*8*	23	*16*	25	*1*	5	*2*	10	*2*	4	*1*
	Total	100		100		100		100		100		100		100		100	
binary	Africa	6	*22*	7	*17*	3	*5*	5	*12*			4	*5*			4	*2*
	Asia & Pacific	25	*24*	19	*13*	22	*11*	34	*23*			4	*1*	18	*5*		
	Europe	45	*7*	52	*6*	57	*5*	36	*4*	75	*1*	61	*4*	67	*3*	63	*2*
	Latin America & Caribbean	5	*13*	4	*10*	1	*3*	6	*15*			26	*35*			25	*15*
	Rest of Europe	2	*7*	3	*7*	5	*9*	2	*4*			2	*2*	4	*4*	4	*2*
	U.S. & Canada	17	*20*	15	*13*	13	*8*	17	*15*	25	*2*	4	*2*	11	*4*	4	*1*
	Total	99		100		100		100		100		100		100		100	

5.5.2 Betweenness and eigenvector centrality

The next centrality indicators that are presented here are the betweenness and the eigenvector centrality. As mentioned above, these indicators are more suitable for the Internet backbone network because they take into consideration the indirect linkages, something which is fundamental for the Internet function. Table 5.7 presents these indicators.

Table 5.7: Betweenness and eigenvector centrality

	Betweenness								**Eigenvector**							
	Internet 2001		Aviation 2001		Internet 2006		Aviation 2006		Internet 2001		Aviation 2001		Internet 2006		Aviation 2006	
London	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*
Paris	49	*2*	33	*2*	35	*3*	50	*2*	77	*2*	63	*2*	70	*2*	67	*2*
Amsterdam	17	*4*	23	*4*	13	*5*	33	*4*	66	*3*	59	*3*	52	*4*	54	*3*
Frankfurt	22	*3*	27	*3*	64	*2*	39	*3*	59	*4*	46	*4*	55	*3*	47	*4*
Madrid	2	*20*	5	*8*	4	*13*	22	*5*	8	*9*	36	*5*	22	*5*	41	*5*
Milan	12	*7*	0	*22*	12	*6*	9	*8*	9	*8*	17	*12*	9	*8*	34	*6*
Rome	0	*39*	0	*26*	0	*39*	1	*20*	0	*37*	18	*10*	0	*52*	31	*7*
Barcelona	1	*29*	0	*21*	1	*22*	4	*11*	1	*18*	25	*6*	3	*17*	30	*8*
Copenhagen	5	*12*	8	*6*	8	*8*	3	*14*	11	*6*	23	*7*	10	*7*	25	*9*
Munich	3	*17*	2	*14*	0	*36*	14	*6*	1	*21*	20	*9*	1	*28*	24	*10*
Lisbon	15	*5*	7	*7*	9	*7*	11	*7*	1	*22*	17	*11*	3	*19*	21	*11*
Athens	1	*28*	4	*11*	5	*12*	5	*10*	1	*19*	16	*13*	2	*21*	19	*12*
Stockholm	9	*9*	4	*9*	2	*19*	0	*23*	10	*7*	20	*8*	5	*13*	19	*13*
Prague	2	*21*	4	*10*	3	*15*	9	*9*	1	*23*	10	*21*	3	*18*	18	*14*
Geneva	5	*13*	0	*31*	0	*28*	0	*39*	6	*11*	10	*22*	7	*12*	15	*15*
Oslo	8	*10*	2	*16*	0	*33*	0	*27*	3	*14*	15	*15*	3	*20*	13	*16*
Berlin	0	*31*	0	*39*	0	*39*	0	*24*	0	*26*	9	*24*	0	*44*	11	*17*
Manchester	0	*40*	1	*18*	0	*39*	3	*13*	0	*42*	12	*19*	0	*43*	11	*18*
Helsinki	2	*26*	0	*23*	0	*39*	0	*28*	1	*20*	12	*17*	1	*27*	11	*19*
Warsaw	3	*18*	2	*15*	6	*11*	1	*21*	0	*25*	7	*26*	5	*15*	11	*20*
Dublin	2	*23*	3	*12*	2	*17*	0	*29*	2	*15*	14	*16*	8	*10*	10	*21*
Brussels	13	*6*	0	*25*	0	*25*	0	*34*	41	*5*	9	*23*	17	*6*	10	*22*
Düsseldorf	0	*40*	1	*19*	0	*37*	1	*22*	6	*10*	11	*20*	9	*9*	10	*23*
Hamburg	4	*16*	0	*41*	3	*14*	0	*25*	4	*13*	8	*25*	8	*11*	8	*24*
Venice		*40*	0	*34*	0	*39*	0	*42*			5	*30*	0	*46*	7	*25*
Stuttgart	0	*40*	0	*30*	1	*21*	3	*12*	0	*41*	6	*28*	0	*39*	7	*26*
Vienna	12	*8*	0	*37*	14	*4*	0	*36*	1	*17*	6	*27*	4	*16*	7	*27*
Cologne	0	*40*	0	*27*		*39*	2	*19*	0	*38*	4	*32*			6	*28*
Bucharest	0	*37*	2	*13*	0	*24*	2	*18*	0	*31*	4	*33*	1	*24*	6	*29*
Lyon	2	*23*	0	*29*		*39*	0	*32*	0	*29*	6	*29*			6	*30*

Data sources: Telegeography 2007, ICAO 2008

These measures' value added in the comparative study of the two infrastructural networks is the clearer illustration of cities which only perform an important role in one of the two networks. Just like with the degree centrality, the cluster of the four main cities remains as it is. However, although Amsterdam appears to be only one position higher in the aviation hierarchy, its betweenness centrality is 20 units (out of 100) higher for this network. However, in regards with the eigenvector centrality the difference between the two networks for this city are minor.

By observing Table 5.7 it seems that there is a group of cities which appear to be more central in the aviation network than in the Internet backbone one: Rome, Barcelona, Munich, Manchester and even Athens and Prague are more central for the network of the airlines. They appear to gain up to 30 (out of the 62 common cities) positions in the hierarchy when the focus changes from the Internet backbone to the aviation network. Not surprisingly, some of them are unique touristic destinations and their superiority in this network could be explained by that. On the other hand, cities such as Geneva but mainly Brussels, Hamburg and most importantly Vienna are clearly less central in the aviation than in the Internet backbone network.

All in all, a few solid conclusions can be drawn from the comparative analysis of the cities' centralities between the two infrastructural networks. No matter that the four main hubs in the Internet backbone network retain their central roles in the aviation network as well, it cannot be said that this is the case for all the 62 common cities. On the contrary, the distinctive roles of some cities in the two infrastructural networks emerged from the above analysis of the different centrality indicators. Moreover, the discontinuities that appeared in the Internet backbone network between the binary and the weighted network's centralities do not characterize the aviation network. This is probably one of the reasons why cities' extroversive character is more intensive for the Internet backbone network than for the aviation one. Going back to Chapter 4, cities appear to be more extroversive mainly for the binary network, while in terms of capacity the main extra-European links are concentrated in a smaller number of cities. However, because of cities' similar performance in the binary and the weighted centrality for the aviation network for reasons explained above, the extroversive character is less intensive in total and is more agglomerated across a handful of cities. As a result these cities serve the continent as hubs with the rest of the world, indicating again the hub and spoke structure of the aviation network.

5.6 City centrality correlations and Quadratic Assignment Procedure

This section will evaluate the similarity between the two infrastructural networks. This comparison can be done at two levels: first at the level of nodes, focusing on the comparison of the network attributes of the nodes of the two infrastructural networks; and second at the level of the overall structure of the network. For both of them though, networks with the same size (same number of nodes) are needed; this is the reason why the networks of the 62 common cities were extracted. In order to do this two-level comparison two different statistical techniques are employed. For the first level, the focus is on the different centrality measures for the two different networks. In order to evaluate the association of the two networks' nodes' centrality, Pearson's correlation coefficient is used. For the second level of the comparative analysis Quadratic Assignment Procedure (QAP) correlation analysis is applied. This is a non-parametric test of correlation, which does not refer to each actor individually, but to the whole network by correlating each pair of the networks via permutations (Choi et al 2006, Chon 2004).

Table 5.8 presents Pearson's correlation coefficient for the Internet backbone and the aviation network based on all the centrality measures which were discussed before. The centrality measures used here are based on the global extents of the two networks but only the centralities for the 62 common cities are included in the correlation analysis. In general it could be said that the correlation coefficients for all the different centralities are very high indicating high associations between the two networks nodes' centralities. Comparing the correlations of the different centralities there are some patterns that can be indicated. First, weighted degree centralities' correlation coefficients are higher than the ones for the binary networks, and eigenvector centralities' correlation coefficients are higher than the ones for the betweenness centralities. In other words the capacity and the flows that the cities create or attract are more associated between the aviation and the Internet backbone network than the number of the links that the cities are served by. This can be linked with the previous finding that the weights of the links are more related with economic geography than the structure of the binary links.

Table 5.8: Centrality correlations

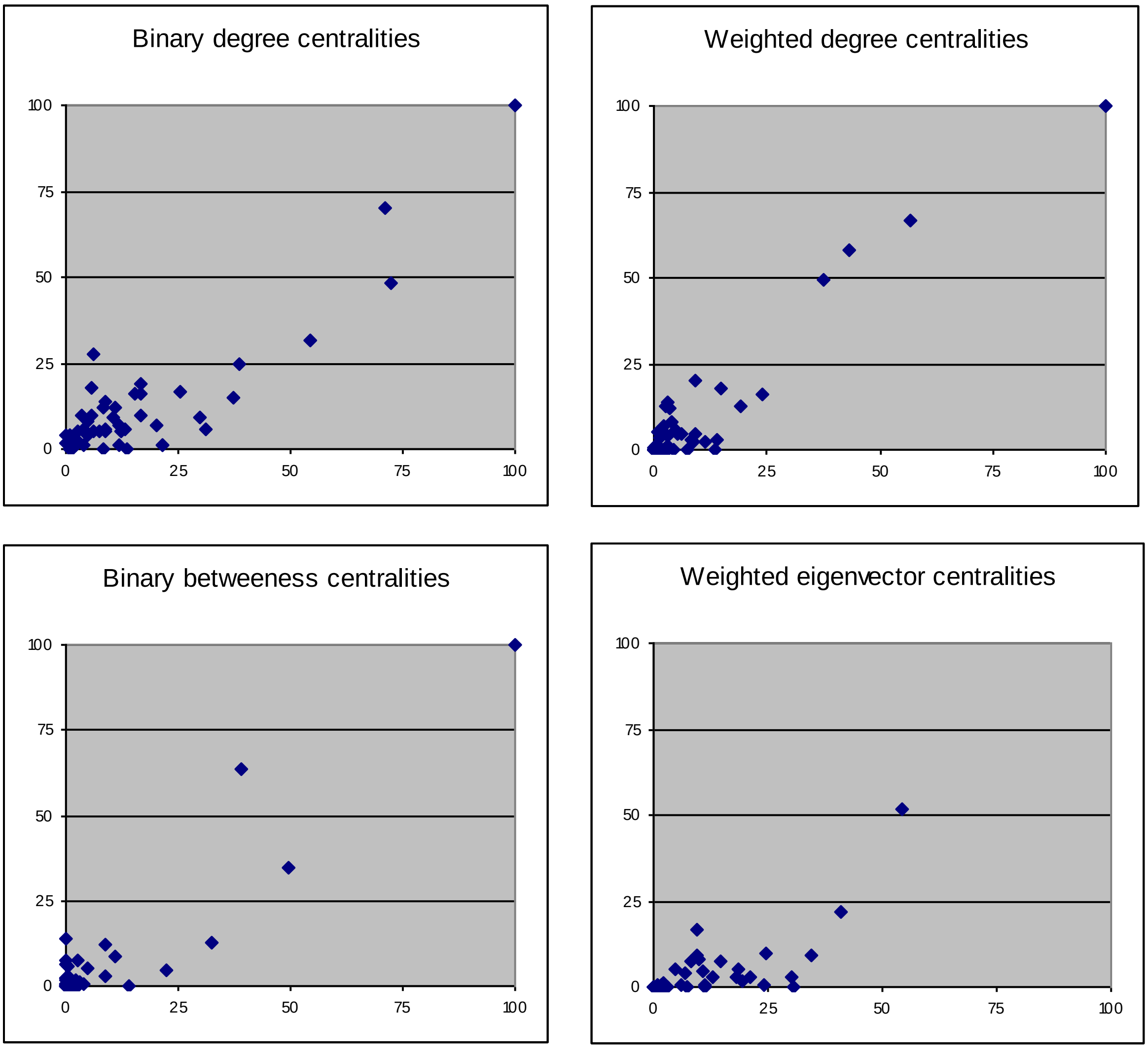
Degree (w)	Degree (b)	Betweeness (b)	Eigenvector (w)
0.925	0.906	0.960	0.896
0.948	0.911	0.962	0.919
0.964	0.910	0.962	0.936
0.955	0.910	0.956	0.938
0.967	0.890	0.950	0.931
0.957	0.889	0.923	0.902

All coefficients are significant at 0.01 level
Data sources: Telegeography 2007, ICAO 2008

The second point refers to the longitudinal comparison of the different coefficients. Through time, correlation coefficients for the centrality measures based on the weighted network (degree and eigenvector) are increased contrary to the decrease that is observed for the correlation coefficients for the centrality indicators based on the binary networks. In simple terms, through time the high capacity and high passenger volume links tend to concentrate more around the same cities, while links in general and more specifically links of lower capacity and passenger volumes do not follow the same pattern

However, more interesting patterns emerge after focusing not on the entire 62 cities, but on the different clusters that are shaped. Going back to the results of the cluster analysis of Chapter 4, the four most central cities emerged as a clear cluster highly differentiated from the remaining cities. This can also be seen in Figure 5.5, where the scatter plots of the four different centralities measures for the two different networks are presented. In order to have a more accurate reflection of how associated the centrality measures for the two infrastructural networks are, Table 5.9 presents the correlation coefficients separately for the four most central cities and for the remaining 58. The first observation from Table 5.9 is that at any case the correlation coefficient for the city centralities between the two networks is much higher for the four most central cities than for the rest. This is not unexpected though. What is interesting is the fact that the centralities for the remaining 58 cities are very low in comparison with the four most central. So, it could be said that the central roles of the four main European hubs (London, Paris, Amsterdam and Frankfurt) are consistent both in the Internet backbone and the aviation network. What is not consistent though are the centralities of the remaining cities. The correlation coefficients for the 58 cities vary from 0.39 to 0.603 indicating a mediocre

Figure 5.5: Scatter plots of the centrality measures for the two networks



association between the centralities of the two networks. In other words, apart from the four central cities, the centralities of the rest of the 62 common cities vary significantly between the two networks.

Another interesting point is that the correlation coefficients for the centralities of the 58 least central cities for the two networks are higher for the two degree centrality indicators than for betweenness and eigenvector. The correlation coefficients for the betweenness centrality are, for all six years but one, lower than the coefficients for the binary degree centrality and accordingly the correlation coefficients for the eigenvector centrality are always lower than the coefficients for the weighted degree centrality. This means that when only the direct links are taken into consideration (i.e. degree centrality) the centrality measures of the 62 common cities are more associated between the two infrastructural networks. On the contrary, when the indirect links are taken into consideration (i.e. betweenness and eigenvector centrality) the centrality measures of the nodes of the two networks appear to be less associated. No matter that the differences are rather small, especially between the degree and the eigenvector centralities for the weighted network, the consistency of these differences indicates the dissimilar philosophies which lie behind these network structures. The Internet backbone network structure enables and promotes communications through indirect links, while for the aviation network indirect links usually result in more inconvenience and diminish the network's efficiency.

Table 5.9: Centrality correlations for sub-groups

Degree (W)		**Degree (B)**		**Betweeness (B)**		**Eigenvector (W)**	
58 cities	4 cities	58 cities	4 cities	58 cities	4 cities	58 cities	4 cities
0.535**	0.995**	0.527**	0.980*	0.323*	0.963*		0.983*
0.601**	0.974*	0.572**	0.992**	0.458**	0.966*	0.496**	0.911**
0.628**	0.990**	0.567**	0.993**	0.317*	0.992**	0.593**	0.968*
0.592**	0.973*	0.534**	0.995**		0.970*	0.579**	0.999**
0.604**	0.995**	0.507**	0.956*	0.390**	0.975*	0.598**	0.998**
0.606**	0.997**	0.524**		0.603**		0.587**	0.981*

coefficients significance: ** at 0.01 level, * at 0.05 level, blank for insignificant coefficient

Data sources: Telegeography 2007, ICAO 2008

However, the above are only focused on the attributes of the cities themselves and the conclusion that have been drawn do not say anything about the network structure similarities. In order to focus on the latter, QAP is applied for the two networks of the 62 common cities. The results are presented in Table 5.10. QAP based correlations can be interpreted as usual correlations. For the case of the binary networks and for the links between the 62 common cities, the structural similarities between the two infrastructural networks are low, but significant for the whole 6 year period. However, there is a trend for higher QAP correlation coefficient over time, which means that through the six year time period, the strength of the correlation between the two infrastructural networks increased. Comparing the above analysis with other similar studies, the QAP correlation coefficient for the two infrastructural networks for the 62 common cities is much lower than the coefficient for the same networks at a global scale. Choi et al (2006) compared the same datasets for 82 world cities for 2002 and they found a QAP correlation coefficient of 0.46, almost 4 times higher than ours. So, it could be said that at a global scale there is some similarity in the way that these infrastructural networks are rolled out, but on a lower scale the similarity is very weak. This is not surprising because of the networks' character. By definition, they are global networks and their objective is to serve the global centres of production and consumption. At such a scale, it is expected that they would share such structural similarities since they serve the same cities. However, at a regional scale and when less important nodes of the urban network are included in the analysis, the location decisions of the infrastructural networks' designers are less obvious and the networks interconnect more cities than the well-known global ones, reflecting different scalar geographies.

Table 5.10: QAP correlations between the Internet and aviation networks

Year	QAP
2001	0.126
2002	0.145
2003	0.145
2004	0.146
2005	0.153
2006	0.160

All coefficients are significant at 0.01 level
Data sources: Telegeography 2007, ICAO 2008

To sum up, although that the structure of the two infrastructural networks does not appear to be very similar at least at this scale, the centralities of the 62 common cities seem to be highly correlated. However, the high correlation is mainly due to the 4 most central cities, which retain their central roles in both networks. Nonetheless, the differences in correlation coefficients of the 58 least central cities can lead us to some conclusions regarding the structure of these networks. These are the Internet's inclination for indirect communications, the economic explanation of the allocation of the networks' weights, as well as the higher concentration over time of the high capacity and passenger volumes links around the same cities.

5.7 Complex networks analysis

In this section the comparison between the two infrastructural networks is extended by taking into consideration complex network theory. As was mentioned in the previous chapter, complex network theory is used in order to better explain complicated network structures. In this section, the two infrastructural networks are compared with theoretical models in order to identify similarities and differences. Just like in Chapter 4, the main network elements that are under scrutiny are the average distances and the cluster coefficients, the vertex degree distribution and the s metric. In order to better compare the two infrastructural networks, in addition to the global and the European extent, the networks of the 62 common cities are also studied.

Table 5.11 presents some network statistics for the two infrastructural networks taking into consideration only the binary connections. The network statistics for the global and the European Internet backbone network have been presented in the previous chapter (Table 4.12) but are repeated here for comparison reasons. In general, average distances as well as the diameter are always higher for the case of the Internet backbone network. This applies to the whole six year period and for all the different extents, including the same size networks of the 62 common cities. As mentioned in the previous chapter, short average distances and diameters are related with the networks' greater efficiency. For the case of airline networks short distances indicate fewer connection flights and obviously less inconvenience for the travellers while for the case of the Internet backbone links, fewer hops and latency. The shorter network distances for the aviation network can be justified by the fact that airlines network's efficiency are more sensitive to short network distances compared to the backbone networks, since the inconvenience is greater for the former when the geodesics are longer. What is also important is that through time average distances are decreased, indicating a network efficient improvement for both of the infrastructural networks.

The same applies to the clustering coefficient as well. The aviation network appears to be more clustered than the Internet backbone for all the years and for all the different extents. The differences in the clustering coefficient reflect the fact that direct links are more important for the aviation network than for the Internet backbone. The clustering coefficient is the ratio between the existing numbers of edges among a node's nearest neighbours and the maximum number of these edges (Albert and Barabási 2002). So, for the case of the aviation network, more links exist among a node's nearest neighbours than for the Internet backbone. Obviously, for the latter the indirect links serve all the nodes in the neighbourhood while for the aviation network the need for direct links forces the creation of new edges in the neighbourhood which result in higher clustering coefficients.

Table 5.11: Network statistics

Networks		Average Distance (1)	Average Distance RN (2)	Diameter (3)	Diameter RN (4)	CC (5)	CC RN (6)
Internet backbone networks	2001 a.c.	2.861	3.600	6	*7*	0.478	0.025
	2002 a.c.	2.789	3.565	6	*8*	0.457	0.042
	2003 a.c.	2.716	3.538	6	*7*	0.522	0.022
	2004 a.c.	2.669	3.378	6	*7*	0.597	0.024
	2005 a.c.	2.611	3.413	5	*7*	0.581	0.039
	2006 a.c.	2.725	3.457	7	7	0.574	0.033
	2001 E.c.	2.762	2.579	7	5	0.424	0.082
	2002 E.c.	2.641	2.719	5	6	0.524	0.114
	2003 E.c.	2.555	2.653	5	5	0.539	0.075
	2004 E.c.	2.495	2.549	6	5	0.562	0.063
	2005 E.c.	2.477	2.570	5	5	0.563	0.086
	2006 E.c.	2.549	2.631	6	5	0.571	0.112
	2001 c.c.	2.337	2.670	5	6	0.598	0.107
	2002 c.c.	2.195	2.756	4	6	0.669	0.071
	2003 c.c.	2.191	2.673	4	6	0.634	0.087
	2004 c.c.	2.087	2.546	4	5	0.594	0.109
	2005 c.c.	2.206	2.533	5	5	0.601	0.072
	2006 c.c.	2.250	2.508	5	5	0.61	0.092
Aviation networks	2001 a.c.	2.348	3.002	5	6	0.742	0.027
	2002 a.c.	2.367	2.983	5	5	0.748	0.032
	2003 a.c.	2.346	2.950	5	5	0.75	0.029
	2004 a.c.	2.335	2.921	4	5	0.731	0.038
	2005 a.c.	2.351	2.910	5	5	0.742	0.036
	2006 a.c.	2.344	2.891	5	5	0.75	0.036
	2001 E.c.	2.309	2.451	5	5	0.634	0.078
	2002 E.c.	2.278	2.442	4	4	0.682	0.078
	2003 E.c.	2.271	2.435	4	5	0.671	0.072
	2004 E.c.	2.218	2.362	4	4	0.67	0.084
	2005 E.c.	2.275	2.442	5	5	0.672	0.091
	2006 E.c.	2.267	2.445	5	4	0.679	0.063
	2001 c.c.	1.992	1.993	4	3	0.657	0.157
	2002 c.c.	1.927	1.975	4	3	0.676	0.171
	2003 c.c.	1.928	1.946	4	3	0.652	0.195
	2004 c.c.	1.912	1.925	4	3	0.64	0.184
	2005 c.c.	1.976	1.976	4	3	0.668	0.176
	2006 c.c.	2.004	1.939	4	3	0.677	0.177

a.c. = all cities; E.c. = European cities; c.c. = common cities

Data sources: Telegeography 2007, ICAO 2008

In addition, for both the Internet backbone network and the aviation network average network distances and diameters for almost all the cases are shorter than the same measures for the same size random networks (RN – columns 2, 4 and 6). Moreover, clustering coefficients are always higher than the same coefficients for the same size RN networks. So, just like the conclusion we drew in the previous chapter, both networks appear to have SW characteristics.

Furthermore, the vertex degree distribution of the two networks are also analysed here. Figures 5.6 and 5.7 illustrate the cumulative distribution function (CDF) of the degree centrality for the two networks for all the years, both for the binary and weighted versions and for all the different extents. In addition, Table 5.12

Figure 5.6: Internet backbone network’s degree distribution

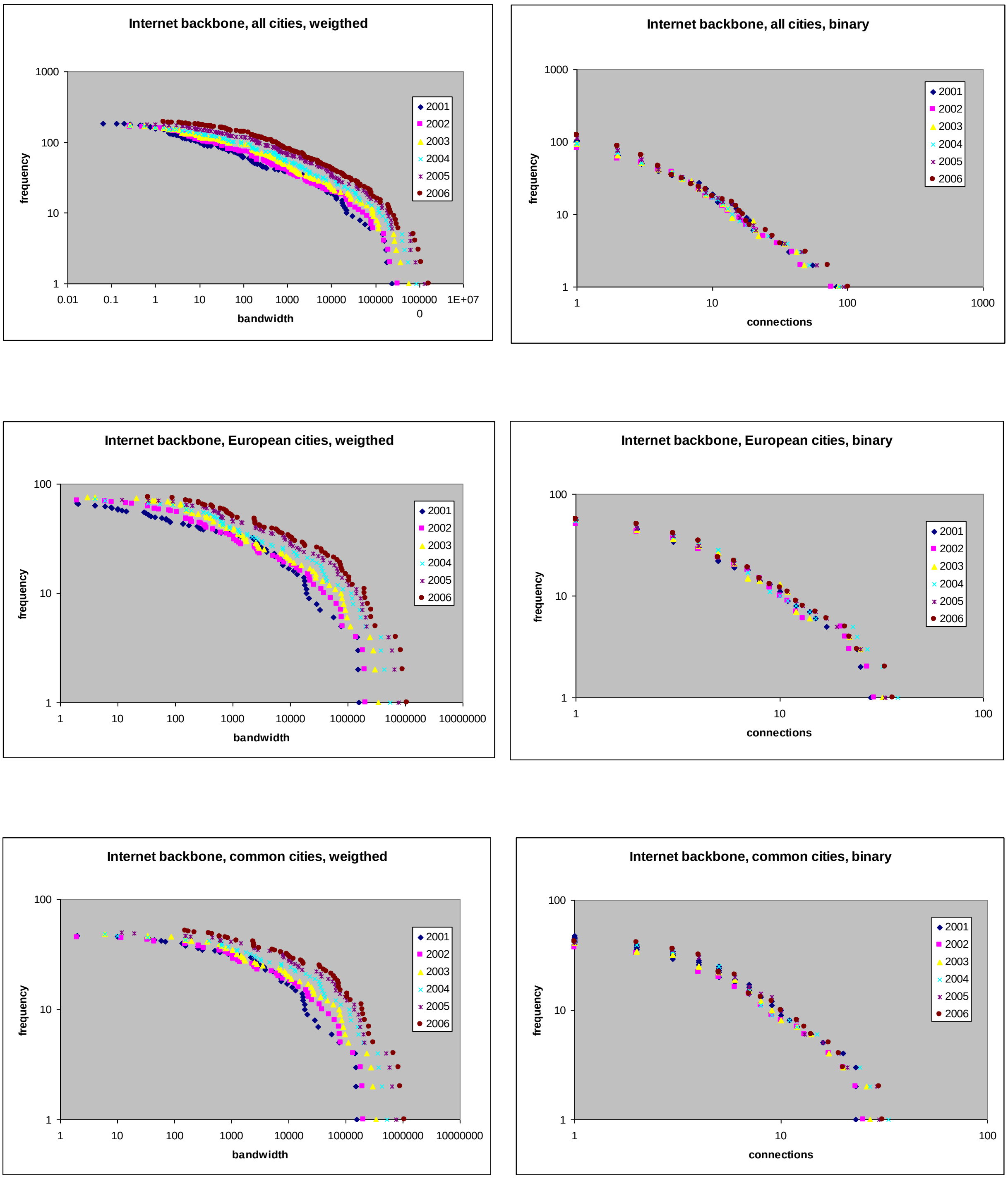
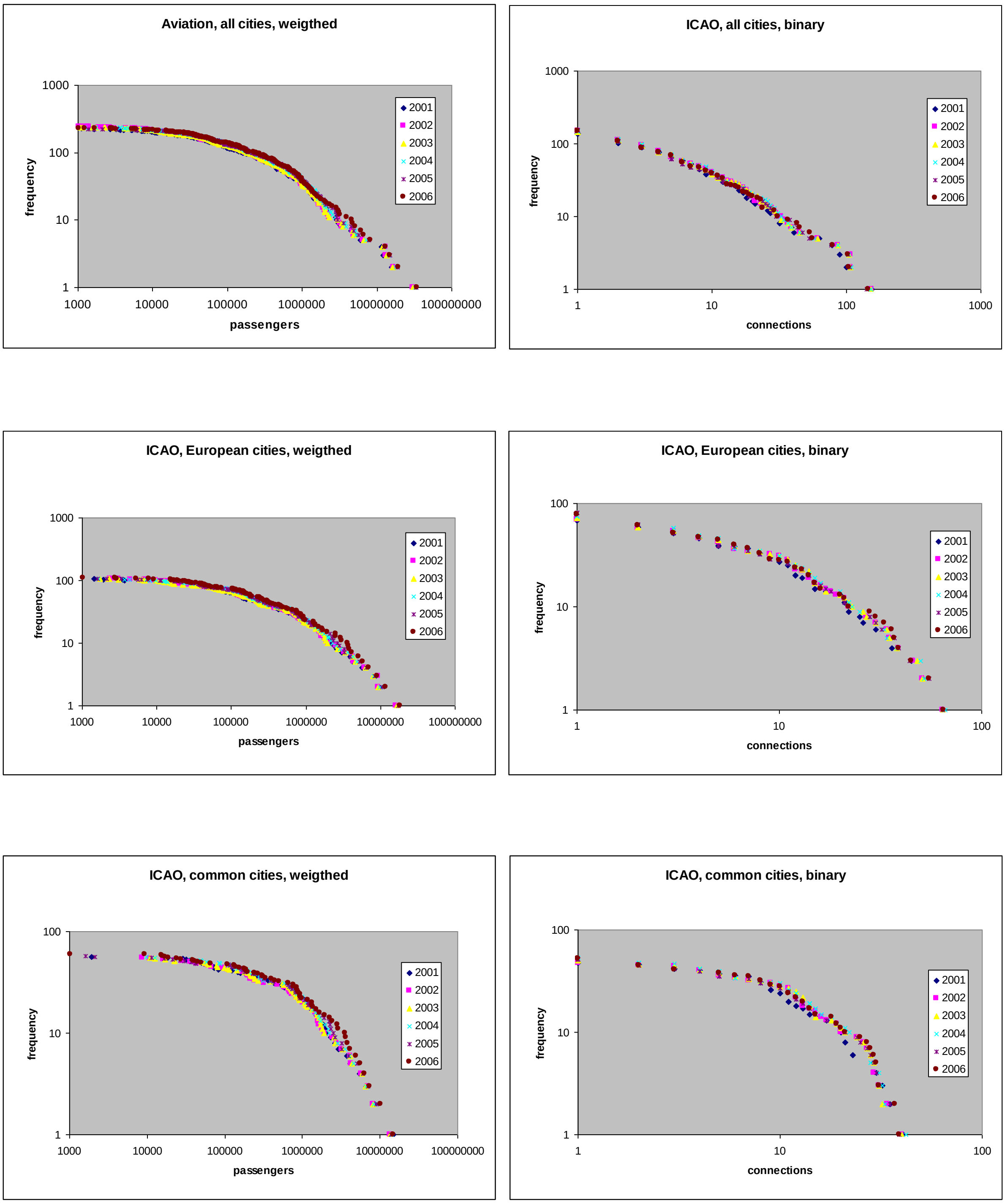


Figure 5.7: Aviation network’s degree distribution



presents the results of the curve estimation using OLS and Table 5.13 the s metric for both networks. Just like before, the figures for the Internet backbone network are repeated for comparison reasons.

Table 5.12: Power and exponential law fit (OLS)

Network		**Weighted**			**Binary**		
		Exp R^2	Power R^2		Exp R^2	Power R^2	
Internet backbone	2001 a.c.	0.525	**0.921**	*SF*	0.762	**0.966**	*SF*
	2002 a.c.	0.578	**0.919**		0.810	**0.934**	
	2003 a.c.	0.557	**0.892**		0.760	**0.949**	
	2004 a.c.	0.571	**0.886**		0.750	**0.947**	
	2005 a.c.	0.546	**0.874**		0.703	**0.968**	
	2006 a.c.	0.543	**0.876**		0.671	**0.964**	
	2001 E.c.	0.727	**0.790**	*SF*	**0.962**	0.894	*SW*
	2002 E.c.	0.785	**0.855**		**0.965**	0.859	
	2003 E.c.	0.755	**0.869**		**0.950**	0.881	
	2004 E.c.	0.782	**0.839**		**0.924**	0.876	
	2005 E.c.	0.776	**0.837**		**0.967**	0.873	
	2006 E.c.	0.758	**0.844**		**0.941**	0.873	
	2001 c.c.	**0.800**	0.736	*SW*	**0.967**	0.830	*SW*
	2002 c.c.	**0.878**	0.744		**0.985**	0.834	
	2003 c.c.	**0.854**	0.790		**0.977**	0.841	
	2004 c.c.	**0.870**	0.734		**0.946**	0.828	
	2005 c.c.	**0.849**	0.726		**0.968**	0.801	
	2006 c.c.	**0.830**	0.792		**0.953**	0.816	
Aviation	2001 a.c.	0.492	**0.836**	*SF*	0.738	**0.928**	*SF*
	2002 a.c.	0.494	**0.831**		0.737	**0.936**	
	2003 a.c.	0.498	**0.832**		0.747	**0.935**	
	2004 a.c.	0.503	**0.830**		0.747	**0.936**	
	2005 a.c.	0.520	**0.839**		0.747	**0.953**	
	2006 a.c.	0.536	**0.831**		0.749	**0.947**	
	2001 E.c.	0.692	**0.830**	*SF*	**0.960**	0.850	*SW*
	2002 E.c.	0.709	**0.843**		**0.972**	0.849	
	2003 E.c.	0.712	**0.821**		**0.972**	0.843	
	2004 E.c.	0.716	**0.826**		**0.966**	0.850	
	2005 E.c.	0.725	**0.826**		**0.954**	0.887	
	2006 E.c.	0.748	**0.821**		**0.959**	0.873	
	2001 c.c.	**0.833**	0.770	*SW*	**0.990**	0.710	*SW*
	2002 c.c.	**0.864**	0.793		**0.977**	0.690	
	2003 c.c.	**0.867**	0.783		**0.969**	0.691	
	2004 c.c.	**0.866**	0.789		**0.981**	0.724	
	2005 c.c.	**0.878**	0.701		**0.978**	0.729	
	2006 c.c.	**0.899**	0.730		**0.973**	0.718	

a.c. = all cities; E.c. = European cities; c.c. = common cities

Data sources: Telegeography 2007, ICAO 2008

In general, an interesting attribute is that the plots for weighted Internet networks seem to be more differentiated through time in comparison to the aviation network. This happens because of the technological improvements which take place, such as the diachronic bandwidth upgrade in the backbone links. On the contrary, weighted aviation networks do not demonstrate such differences because the passenger volumes remain relatively stable over time, indicating the maturity of the aviation networks contrary to the early stages of the Internet backbone network.

In more detail, the Internet backbone and aviation networks for all the cities appear to have SF properties both for weighted and binary versions according to their degree distribution. R square for power law fit is above 0.90 for binary versions and around 0.90 for the weighted Internet backbone network and slightly smaller for the aviation one. What is interesting is that R square for the power law fit decreased through time for the weighted Internet network for all the cities while the R square for the exponential law increased. The latter reflects the tendency towards a more homogenous bandwidth distribution across cites. In addition, after a closer observation, it seems that CDF plots for the weighted aviation networks consist of two different slopes, one almost parallel to the x axis for the least connected cities and one steeper for the most connected ones, forming this way a power law tail. Indeed, if only the 100 most connected cities are plotted, R square for power law fit is around 0.97 for the whole time period. The latter indicates that a SF structure, which is related with the power law degree distribution and with the existence of a small but considerable number of highly connected nodes and a bulk of poorly connected nodes, only takes place on the first tier of the most well connected cities, while the second tier seems to be more homogenous.

Table 5.13: *s* metric for both networks

	Internet a.c.	Internet E.c.	Internet c.c.	Aviation a.c.	Aviation E.c.	Aviation c.c.
2001	0.233	0.606	0.651	0.189	0.454	0.687
2002	0.242	0.571	0.619	0.191	0.472	0.712
2003	0.216	0.513	0.596	0.197	0.478	0.726
2004	0.216	0.482	0.545	0.199	0.471	0.719
2005	0.196	0.528	0.471	0.210	0.463	0.710
2006	0.178	0.511	0.565	0.209	0.463	0.704

a.c. = all cities; E.c. = European cities; c.c. = common cities
Data sources: Telegeography 2007, ICAO 2008

It is also of interest to explore the structural importance of those highly connected nodes using the s metric. As is presented in Table 5.13, the s metric, which takes values between 0 and 1 is very low for both Internet backbone and aviation networks for the whole time period. The indicator's change through time is also interesting. The s metric for the Internet backbone steadily decreased while for the aviation network it almost steadily increased. In other words, there is an increasing tendency of the Internet backbone hubs to be connected to less connected vertices, while connectivity among highly connected aviation hubs is being increased through the 6 year period. Nonetheless, for all the cases, the s metric is very low and prevents us from identifying these networks as SF, despite their power law fit. However, it should be highlighted here that distributions close to power law and hubs with neighbours of low connectivity were expected for the networks of all the cities. As it was noted before, for the non-European cities, only their edges with the European ones are included in the analysis while their links with the non-European cities are missed. As a result, these cities appeared to be poorly connected.

In order to overcome the above, the analysis now focuses on the networks consisting only by European cities. From a first look at R square, it seems that the weighted versions of both networks follow power laws, while the binary ones are more homogenous since they better fit with exponential laws for the whole time period. The fit for the binary networks over the 6 year period is higher than 0.94 and also higher than the ones for the weighted networks. Again, the weighted aviation network seems to consist of two different slopes, one almost parallel to the x axis for the least connected cities and another steeper for the most connected ones. The slopes for the 50 most connected cities for the whole time period fit better with power laws and R square is above 0.94, indicating a power law tail for the passengers' distribution. However, this is the case either for the binary aviation networks or for the Internet backbone ones. Even for this small sample of the 50 most connected cities, CDF slopes fit better with exponential laws for the binary networks. This can also be graphically proven since no different slopes can be observed for any case apart from the weighted aviation networks. All in all, both aviation and Internet backbone binary networks seem to be more homogenous and more compared to the previous set of networks. However, even for this scale, the bandwidth's and passengers' distributions better fit with power laws indicating SF properties, which are related with the existence of some very-well connected (in terms of bandwidth and passenger loads) hubs and a bulk of less connected cites. However, it should be highlighted here the existence of a sharp cut-off which appears for the four most connected cities in the weighted Internet backbone network. In order to better fit with such a network structure and a SF distribution, these cities should have been characterized by greater bandwidth.

Table 5.13 also presents the s metric for the intra-European network. No matter that this indicator is slightly higher for the case for the Internet network, it is still low for both infrastructural networks, indicating the lack of hubs which hold the network together. So, it could be said that there is quite an uneven bandwidth and passenger distribution across the European cities, which is a characteristic of SF networks, but it cannot be said that those network follow SF models.

Going further, we analyze the Internet and aviation networks for the case of 62 common cities in Europe. For the overall networks and for the whole time period exponential laws seem to fit better. For the weighted networks the difference between R square for power and exponential law are lower than those for the binary networks. Again, the tails of the distribution of the weighted aviation network seem to fit better with a power law and R square for the 20 most connected cities is higher than 0.97. The common cities’ networks are characterized by higher s metric values comparing to the previous cases. However, they are still low and only for some years for the aviation network exceed 0.7. But even this value of s metric is not enough for the emergence of SF hubs. If the five most connected cities for 2006 of both networks were removed (which is 8% of all networks’ nodes), then the average distance among all reachable cities for the Internet backbone would be increased only by 10% while for the aviation networks by 24% indicating this way the fact that hubs have a more dominant role in the aviation network than in the Internet backbone. For a SF network the increase would be more than 100% according to Albert et al (2000).

To sum up, the Internet backbone and aviation networks do not fit with the SF model. Both Internet and aviation networks have some highly connected hubs, but they cannot hold the network together. The distribution of the technology (bandwidth) and the passenger loads is less homogenous than the distribution of the actual connections, indicating some SF properties for the weighted networks but clear SW characteristics for the binary networks. In addition, because of the aviation networks' truncated degree distribution, they could also be characterized as broad-scale networks, following Amaral et al's (2000) definition. Overall, it could be said that between the two infrastructural networks, the aviation one seems to resemble more the structure of SF networks with more important hubs, for which for some years and only for the extracted network of the 62 common cities the s metric was above 0.7. However, in general the low s metric values for almost all the cases indicate network structures which do not fit with the existence of highly connected super-hubs, which hold the networks together.

5.8 Economic realization of the two infrastructural networks

In order to better understand the results of the above topological and structural analysis of the two infrastructural networks and explain the reasons why these structures appeared, there is a need to use a wide range of argumentation, from economic theory to network engineering, and to apply it not only on the aggregated level but also on the disaggregated – micro level of the individual carriers.

As mentioned above, the two infrastructural networks appear to have different structures at the micro level of individual carriers. Airlines tend to roll out their networks using a hub and spoke topology, while Internet Backbone Providers usually do not clearly follow such a topology [21,22]. At this level aviation networks are more centralized and the hub-cities have a very important role for the network function. Before going further to the macro level of the aggregated networks of different carriers which is the main focus of this analysis, it is worth analyzing why the developers of the two different infrastructural networks choose different topologies for their networks. This assignment will assist in better explaining the structures at the macro level.

The hub and spoke networks in the aviation industry appeared as a result of the deregulation process in the US and from 1977 onwards more and more carriers adopted this network structure (Button and Stough 2000). The increased competition

---

[21] Examples are provided in the Annex of this chapter.

[22] However, this is not the case for the low cost airlines, which base their function on direct links and prefer structures different than the hub and spoke, such as the fully connected network (Gillena and Morrison 2003).

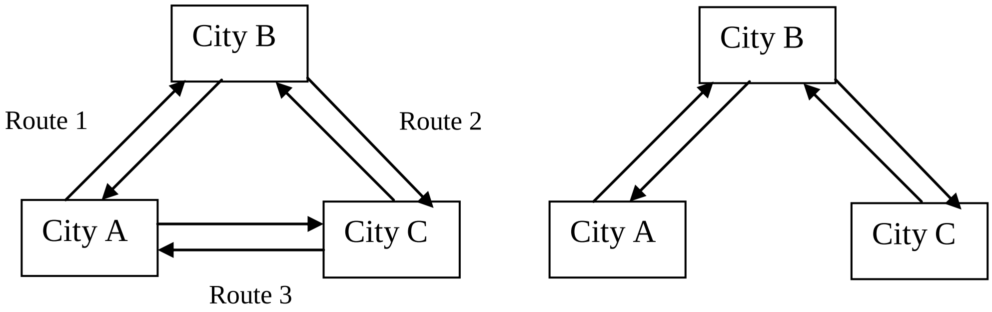
forced air carriers to reduce the direct flights and re-route them through specific hub cities because of the economic externalities that the hub and spoke network structure produces. Holloway (2003) indicates three types of externalities because of the hub and network structure:

- *Economies of scope*. These externalities appear because of the reduced cost after channelling passengers with different origins and destinations through a specific hub instead of operating all the different origin-destination routes. Passenger flows between different origins and destinations are combined together at least for a leg of their total route, reducing in this way the total cost of operating the full network with lower frequency flights and smaller capacity aircraft.
- *Economies of density*. These externalities appear when the cost is reduced because the increased passenger flows on a specific spoke due to the hub and spoke structure, enable the carrier to use larger aircrafts, which are characterised by lower set-mile cost. Alternatively, the size of the aircraft can be a traded by smaller aircraft operating in higher frequencies. In general economies of density can be referred also as ‘aircraft economies of scale’.
- *Marketing economies of scale*. Carriers can take advantage of the information economies which appear because by choosing a hub airport they can be identified as the dominant players in this hub and consequently in its hinterland, as it is signified by the hub’s spokes. KLM’s presence in Schiphol airport gives the impression to the potential customers that it is the dominant air-carrier in Amsterdam’s hinterland or in other words to KLM’s spokes, preventing customers from searching for alternatives.

However, as mentioned above, the hub and spoke does not seem to be the preferred network structure by the Internet Backbone Providers. This business choice can be rationalised after taking into consideration the network economics using the example illustrated by Shy (2001, 216-7). Figure 5.8 presents a three city urban network served by a fully connected network in the left part and by a hub and spoke network in the right part. In the left part passengers can fly to all destinations using the direct links while in the right part city B acts as the hub, through which all the flows from A to C and from C to A are channelled. In order to simplify the example, only one-way travel can be assumed with passenger volume n_1 from city A to city B, n_2 the volume from B to C and n_3 the passenger volume from city A to C. Then, the total cost of a carrier can be denoted as function of n that is $TC(n_1, n_2, n_3)$. According to Shy (2001, 217) an airline is favoured by economies of network when

$$TC(n_1, n_2, n_3) < TC(n_1, 0, 0) + TC(0, n_2, 0) + TC(0, 0, n_3) \tag{5.1}$$

Figure 5.8: Fully connected (left) and hub and spoke (right) networks



Source: Shy 2001, 216

This means that the above economies of scope, density and marketing because of the network structure decrease the operational cost of the three different routes when they are conducted by the same carrier at a level lower than the sum of operational costs of three different carriers, when each one of them only operates one of the three routes.

For the needs of the example we can assume that there is only one carrier operating across the three routes. In this case the total cost will be

$$TC(n_1, n_2, n_3) = c(n_1) + c(n_2) + c(n_3) \tag{5.2}$$

$$c(n_i) = \phi + n_i^{\,2} \tag{5.3}$$

So, in this example the operational cost of each route is the sum of a fixed cost ϕ and a variable cost linked to the number of passengers. The former is due to costs such as gate renting, hiring local staff, landing fees etc., while the latter rises quadratically with passenger volumes for reasons such as aircraft capacity limits.

Going back to Figure 5.8 and from the equations 5.2 and 5.3 the total cost for a carrier to operate the fully connected network is $TC^{FC} = 3\phi + {n_1}^2 + {n_2}^2 + {n_3}^2$ while the total cost for the hub and spoke network is $TC^{HS} = 2\phi + (n_1 + n_3)^2 + (n_2 + n_3)^2$. If we assume equal passenger volumes among the three cities $(n_1 = n_2 = n_3 = n)$, then

$TC^{HS} < TC^{FC}$ only if $\phi > 5n^2$.

So, in simple terms, when the fixed cost is much higher than the passenger volume, then the hub and spoke structure is the most cost efficient network structure. On the contrary, when the fixed cost of operating a route is small, then the full network appears to be the network structure with the least cost. And because usually the fixed cost is high for most airlines, they tend to roll out their network at a hub and spoke structure. On the contrary, low cost airlines, because of the lower fixed cost tend to use direct links and fully or partially connected networks (Gillena and Morrison 2003, Shy 2001).

No matter how simple the above example is, it illustrates the economic interpretation of the adoption of the hub and spoke network structure by the airlines. However, the economic explanation of the Internet backbone networks structure is not that straightforward. Contrary to what applies to the aviation networks, the main cost for the roll out of such a network is not the fixed cost of switching which takes place at the network nodes, and was denoted by ϕ above, but the high sunk cost of the fibre optic cables' installation. Kharif (2001) noted that in extreme cases the cost of fibre installation can reach even \$1 million per mile because of the excavation cost but also because of the trench and pipes property rights. However, such high prices usually only occur for highly urbanized areas, where the above costs rise dramatically. So, the main economic constraint in the Internet backbone network roll out is the total length of the edges of the network as well as the number of the edges, but not the switching process as it is in the aviation industry, the cost of which is relatively small in comparison to the whole network's cost. What is for sure is that a fully connected network, where the physical length of the edges as well as their total number is maximized, is not the optimal choice for the Internet backbone networks designers.

Parenthetically, it should also be noted here that the use of indirect links in the Internet data packets transportation might also create negative externalities because of latency. As explained in Chapter 2 the extensive use of indirect links and routes with many hops is related with data transmission delays, known as latency, which result in network's lower efficiency.

At the other side of the spectrum, hub and spoke networks are suitable for minimising the number of the edges. In addition, by channelling data packets through specific hubs and achieving higher utilisation of the existing links, economies of scope and density can be exploited in a way similar to the aviation networks. Although, in regards to the total physical length of the network edges, it cannot be claimed that hub and spoke networks always result in shorter total installed fibre optic length than ring networks. This depends on the actual geography of the nodes that need to be connected.

The complexity of the interpretation of the structure of the Internet backbone networks is increased when other non-economic factors are taken into consideration, such as the networks' resilience. *Survivability* is among the main restrictions in the process of the network's design. According to the US Department of Commerce (1996, S-34) *survivability* is defined as a

> "property of a system, subsystem, equipment, process, or procedure that provides a defined degree of assurance that the named entity will continue to function during and after a natural or man-made disturbance; *e.g.*, nuclear burst. *Note:* For a given application, survivability must be qualified by specifying the range of conditions over which the entity will survive, the minimum acceptable level or post-disturbance functionality, and the maximum acceptable outage duration".

Yet, the hub and spoke networks or otherwise the star like networks are the most vulnerable when the network is under attack. No matter the fact that the extreme and oversimplified case of the hub and spoke or star like network does not fit with the complex structure of a SF network, both of them share the similarity that they are vulnerable in the case of an attack from an informed agent. Such an agent will target the hub node(s), the break down of which will result in the collapse or significant malfunction of the whole network (Albert et al 2000). On the contrary, such networks face less danger from an accidental (i.e. random) node failure. The likelihood of having the main hub down because of a failure is much smaller than having one of the spokes down, simply because there is only one hub contrary to the bulk of spokes. Even at the disaggregated level of the individual Internet Backbone Providers, their commercial success depends to a degree on their ability to provide security guarantees. This is why the extreme case of a clear hub and spoke or star like network is not the preferable choice for Internet backbone network designers contrary to the aviation ones, where in the extreme scenario of an informed attack the survivability of the network is not the main priority.

The above discussion is not something new in the Internet history. On the contrary survivability was the main determinant for the initial Internet's design. This is the reason why Paul Baran in 1964 on behalf of RAND, the US defence think tank, highlighted the vulnerability of centralized networks. Instead, the structure of the distributed or mesh network was suggested by him for "a future all-digital-data […] network which provides common user service for a wide range of users having different requirements" (Baran 1964, v). This future network was the ARPANET, the Advanced Research Projects Agency Network developed by the US Department of Defence, the ancestor of today's Internet[23]. The main objective for such a network was to become the main communication grid, able to survive even after a nuclear attack. However, as Moss and Townsend (2000) note, through time and because of the gradual privatization of the Internet (from US Department of Defence to the National Science Foundation – NSF – and then to private ISPs) the Internet backbone network lost its decentralized attribute and moved towards a more centralized structure in order to meet the demand for such services. And because the Internet is an urban phenomenon (Rutherford et al 2004) the demand for such localised services is mostly allocated in metropolitan areas.

In order to sum up the above, it could be said that the Internet Backbone Providers in order to roll out their long-haul networks need to consider the following constraints:

- Minimise the length of the installed fibre optics and the number of network edges because of the fibre installation cost;

---

[23] For a detailed presentation of the Internet history, the reader can refer among others to Castells 2001, Townsend 2001b, and Rutherford et al 2004.

- Create hubs in order to achieve economies of scope and density;
- Avoid absolute hub and spoke structure because of the increased vulnerability;
- Decrease the number of switching points in order to decrease latency, as explained in Chapter 2;

Obviously, there is no 'one size fits all' solution for an optimal backbone network structure[24]. It depends on the specific conditions of each different case since all the above need to be taken into consideration. As a result Internet backbone networks combine many different structures. Highly connected networks such as the fully connected mesh ones are avoided just like the minimally connected ring networks. On the contrary, usually the latter are enriched with redundant links and the result is partially connected mesh structures, which also include rings and hubs in order to meet the above constraints. Cities with high demand for Internet connectivity appear as hubs in the Internet Backbone Providers topology maps because of their redundant connections. Cities also appear as hubs because of their role in the network to act as gateway cities for their hinterland, as is defined by the spoke cities.

All the above factors are endogenous to the (network) business models adopted by Internet Backbone Providers. Nevertheless, there are also exogenous factors which need to be considered by the individual Internet Backbone Providers such as *path dependency*. As described in Chapter 2, ISPs in order to achieve universal connectivity need to interconnect with other ISPs in order to exchange data. Peering takes place either in private POPs under private peering agreements or in public or commercial IXPs, with the number of the latter being steadily increased in Europe as was illustrated in Chapter 2. Consequently, Internet Backbone networks need to be present in these peering points in order to gain the valuable universal connectivity. It is not surprising that the three IXPs with the highest Internet traffic in 2007 were some of the first IXPs established in the early 1990s in Europe and they are located in Amsterdam (AMS-IX), London (LINX) and Frankfurt (DE-CIX) (Euro-IX 2008)[25]. As a result, no matter the geographical extent that the Internet Backbone

---

[24] The annex of this chapter presents some examples of different Internet backbone networks.

[25] The first IXP in Europe is CIXP, which is the IXP of the well known research centre of CERN in Geneva (CIXP 2008). However, it did not grow like other IXPs simply because its main goal was to facilitate the research centre and not the private ISPs.

Providers want to cover, it makes sense for them to link their networks with these cities in order to enjoy the emerging economies of scope because of the peering process. Consequently, the central role of the above cities in the Internet backbone network is somehow related to the early establishment and the success of the IXPs in these cities.

The result of the above process is that more and more Internet backbone networks are connected to these hub cities and at the aggregated level of the sum of the Internet backbone networks, which is the focus of this study, these cities appear to be highly centralized and at any case much more centralized than they appear at the disaggregated level of the individual networks. Or in other words, the centralized pattern of the overall Internet backbone network is the result of the aggregation process since the individual Internet backbone networks are not that centralized, but most of them include some or all of the hub cities. On the contrary, the hubs in the overall aviation network are not the result of the overlay of different mesh networks, but rather the sum of the different hub and spoke networks of different airlines, which are spread around different cities for a number of reasons including the marketing economies of scale and the airports’ physical limitations to accommodate hub functions for several carriers.

5.9. Conclusions

The main objective of this chapter was to shed some light on the way the two infrastructural networks interconnect the European cities. Because of the network nature of the Internet backbone and the airlines infrastructure, topology is very important in order to better understand how these infrastructural networks facilitate the global city process. Through this comparative network analysis, the attributes of both networks but mainly the Internet backbone network, which is the main focus of this study, can better emerge.

The first conclusion is that the European parts of these two infrastructural global networks are structured in a way to assure low average distances and diameters, enabling an efficient interaction between European cities. In addition, the high clustering coefficient supports the existence of highly connected clusters among the European cities for both networks. No matter the relative low links density, which is justified by the high cost of links establishment, it can be claimed that European cities are interconnected in a fairly efficient way in terms of network distance. However, the lower densities, the longer average distances and the lower clustering coefficients for the Internet backbone network highlight the importance of direct intercity connections for the aviation network. On the contrary, the Internet backbone network can still function efficiently even with lower density and higher average distances, because it is based on indirect communications.

Secondly, both of the aggregated networks are characterized by the existence of some very well connected cities which play the role of hubs. However, while the aviation network is the result of the aggregation of different hub and spoke networks, the Internet backbone network is the result of the overlay of networks with diverse but at any case not clear hub and spoke structures. The results of the complex network analysis, which did not recognize any clear SF network but indicated structures closer to SF for the aviation network, which according to the theory are related with a very few super connected nodes and a bulk of low connectivity nodes, are justified by the nature of these networks. All in all, maybe the hubs play more important roles in the aviation network, but in both networks they are not important enough to keep the network together.

Another distinction is on the way the two networks perform in regards with the edges' weights. When the capacity and the flows are taken into consideration, both networks appear to be less homogenous and more centralized since the importance of the hub cities is greater, justifying the conclusion of the previous chapter that the geography of the weighted links is more related to economic geography.

Moreover, the distinctive character of a telecommunications network such as the Internet is reflected in its network structure as well. This is why the aviation network appears to be more centralised when only the direct links are included in the analysis while the Internet backbone network is more centralised than the aviation one when the indirect connections are taken into consideration. However, for all the different indicators the aviation network decreases its centralization over time while the Internet backbone appears to be more centralized at the end of the six year period, indicating this way that the change which was observed for the US backbone network by Moss and Townsend (2000) towards a more centralized network at the aggregated level has also taken place in Europe. It should be noted though that the aviation network, which is the most expanded one, has a very different level of maturity. Part of the explanation for the above different trends might be due to the still evolving character of the Internet backbone network.

All in all, the Internet backbone network appears to be more homogenous than the aviation one with an emerging and rather differentiated geography. London, Paris, Amsterdam and Frankfurt are the main the hubs in Europe for both global infrastructural networks. However, apart from these cities, the rest of the interconnected European cities have rather different centralities and distinctive roles with respect to the two networks. This indicates an emerging contemporary geography of Internet connectivity in Europe, since it enables cities, which are not part of the traditional core, to become part of the first tier of the most connected cities. This allow cities to play a more important role in Europe than the one they usually perform, which is represented by the more traditional geography of the aviation network.

Chapter 5

Annex

Table A5.1: Centralization indicators

	Binary networks								**Weighted networks**							
	Internet 2001		Aviation 2001		Internet 2006		Aviation 2006		Internet 2001		Aviation 2001		Internet 2006		Aviation 2006	
London	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*
Paris	66	*2*	65	*2*	49	*3*	72	*2*	76	*2*	50	*2*	66	*2*	57	*2*
Frankfurt	45	*3*	58	*3*	70	*2*	71	*3*	68	*4*	37	*4*	58	*3*	43	*3*
Amsterdam	40	*4*	51	*4*	32	*4*	54	*4*	72	*3*	38	*3*	50	*4*	38	*4*
Madrid	11	*18*	25	*7*	15	*12*	37	*6*	9	*9*	19	*5*	16	*7*	24	*5*
Milan	24	*5*	10	*18*	25	*6*	39	*5*	13	*8*	5	*17*	13	*9*	19	*6*
Copenhagen	17	*12*	27	*6*	17	*9*	26	*9*	18	*7*	14	*6*	18	*6*	15	*7*
Munich	18	*10*	21	*8*	6	*25*	31	*7*	4	*17*	9	*9*	3	*22*	14	*8*
Rome	2	*38*	9	*23*	1	*43*	21	*10*	1	*28*	6	*15*	0	*40*	14	*9*
Barcelona	6	*27*	17	*12*	7	*22*	20	*12*	2	*24*	9	*10*	2	*24*	11	*10*
Prague	13	*13*	18	*11*	9	*19*	30	*8*	5	*16*	5	*18*	5	*18*	9	*11*
Stockholm	23	*7*	21	*8*	19	*7*	17	*13*	25	*6*	11	*7*	20	*5*	9	*12*
Lisbon	18	*10*	10	*21*	16	*10*	17	*13*	1	*25*	7	*14*	2	*25*	9	*13*
Athens	12	*15*	14	*14*	16	*10*	15	*16*	1	*33*	7	*11*	3	*23*	8	*14*
Manchester	1	*44*	21	*8*	1	*43*	21	*10*	0	*34*	7	*12*	0	*40*	7	*15*
Oslo	22	*8*	10	*18*	7	*22*	12	*20*	9	*10*	7	*13*	5	*17*	6	*16*
Helsinki	11	*18*	12	*15*	5	*29*	12	*19*	3	*18*	6	*16*	5	*19*	5	*17*
Warsaw	10	*22*	17	*12*	10	*16*	17	*13*	1	*31*	3	*23*	6	*15*	5	*18*
Berlin	7	*24*	5	*36*	1	*43*	12	*20*	3	*21*	2	*25*	0	*43*	4	*19*
Düsseldorf	4	*33*	10	*18*	6	*25*	13	*18*	5	*15*	4	*20*	8	*12*	4	*20*
Geneva	11	*18*	6	*29*	10	*16*	6	*33*	7	*14*	3	*24*	8	*13*	**4**	*21*
Brussels	22	*8*	12	*15*	12	*14*	8	*28*	32	*5*	4	*21*	12	*11*	3	*22*
Cologne	2	*38*	8	*24*			14	*17*	0	*45*	1	*32*			3	*23*
Bucharest	7	*24*	10	*21*	7	*22*	12	*20*	1	*32*	2	*29*	1	*28*	3	*24*
Dublin	5	*30*	11	*17*	5	*29*	6	*31*	2	*23*	4	*19*	4	*20*	3	*25*
Wien	24	*5*	5	*34*	28	*5*	6	*31*	7	*13*	2	*27*	14	*8*	3	*26*
Stuttgart	1	*44*	7	*27*	9	*19*	10	*24*	0	*36*	2	*31*	0	*36*	3	*27*
Hamburg	12	*15*	6	*29*	12	*14*	11	*23*	8	*12*	2	*26*	13	*10*	3	*28*
Zürich	13	*13*	41	*5*	14	*13*	9	*25*	8	*11*	10	*8*	7	*14*	2	*29*
Lyon	2	*38*	7	*27*			8	*28*	2	*22*	2	*28*			2	*30*
Venice			4	*37*	1	*43*	4	*37*			1	*33*	0	*39*	2	*31*
Budapest	11	*18*	3	*40*	18	*8*	6	*33*	1	*26*	0	*43*	4	*21*	1	*32*
Sofia	12	*15*	3	*40*	5	*29*	9	*25*	0	*40*	0	*46*	0	*38*	1	*33*
Vilnius	7	*24*	6	*29*	6	*25*	9	*25*	0	*42*	1	*39*	1	*31*	1	*34*
Riga	6	*27*	6	*29*	5	*29*	8	*30*	0	*38*	1	*37*	1	*33*	1	*35*
Tallinn	4	*33*	3	*40*	8	*21*	5	*35*	0	*35*	1	*40*	1	*27*	1	*36*
Gothenburg			8	*25*	1	*43*	2	*42*			2	*30*	0	*48*	1	*37*
Marseille			4	*37*	5	*29*	3	*41*			1	*36*	1	*30*	1	*38*
Zagreb	5	*30*	5	*34*	10	*16*	3	*40*	0	*41*	1	*35*	0	*35*	1	*39*
Thessaloniki			3	*43*	4	*35*	5	*35*			0	*41*	0	*50*	1	*40*
Bratislava	8	*23*	1	*47*	6	*25*	4	*37*	3	*20*	0	*51*	5	*16*	1	*41*
Ljubljana	6	*27*	4	*37*	5	*29*	4	*37*	0	*37*	0	*44*	1	*29*	1	*42*
Bristol			1	*50*	1	*43*	1	*45*			0	*54*	1	*32*	1	*43*
Turin	1	*44*	2	*45*	1	*43*	1	*45*	1	*29*	1	*38*	0	*54*	0	*44*
Nuremberg			2	*45*	1	*43*	2	*42*			0	*42*	0	*47*	0	*45*
Hannover			6	*29*	1	*43*	1	*45*			1	*34*	1	*26*	0	*46*
Basel	1	*44*	1	*50*	3	*39*	2	*42*	1	*29*	0	*48*	0	*34*	0	*47*
Timisoara					3	*39*	1	*45*					0	*42*	0	*48*
Msida	2	*38*	1	*50*	3	*39*	1	*45*	0	*44*	0	*50*	0	*46*	0	*49*

Wroclaw			1 *43*	1 *53*			0 *48*	0 *50*
Dortmund	1 *44*	1 *50*		1 *53*	0 *46*	0 *55*		0 *51*
Poznan		1 *47*	1 *43*	1 *53*		0 *53*	0 *43*	0 *52*
Nice	2 *38*	8 *25*	1 *43*	1 *45*	0 *39*	4 *22*	0 *53*	0 *53*
Gdansk	2 *38*	1 *50*		1 *53*	0 *43*	0 *49*		0 *54*
Skopje	4 *33*	3 *43*	4 *35*	1 *53*	0 *47*	0 *45*	0 *45*	0 *55*
Strasbourg	4 *33*			1 *45*	3 *19*			0 *56*
Katowice		1 *50*		1 *53*		0 *52*		0 *57*
Salzburg				1 *45*				0 *58*
Rijeka				1 *53*				0 *59*
Tirane	4 *33*		4 *35*	1 *53*	0 *48*		0 *51*	0 *60*
Rotterdam	5 *30*	1 *50*	2 *42*		1 *27*	0 *47*	0 *37*	
Pristina		1 *47*	4 *35*			0 *56*	0 *52*	

Data sources: Telegeography 2007, ICAO 2008

Table A5.2: Percentage of extra-European connections

City	**Internet 2006**		**Aviation 2006**	
	Weighted	Binary	Weighted	Binary
Frankfurt	11%	49%	54%	64%
London	36%	67%	47%	55%
Paris	19%	57%	40%	48%
Manchester	0%	0%	34%	29%
Amsterdam	17%	31%	31%	43%
Madrid	4%	47%	29%	39%
Nice	0%	0%	26%	50%
Brussels	0%	33%	25%	25%
Munich	0%	17%	24%	33%
Milan	6%	32%	23%	36%
Zürich	3%	14%	20%	31%
Copenhagen	15%	35%	18%	24%
Lisbon	4%	69%	18%	38%
Rome	0%	0%	18%	35%
Marseille	0%	0%	13%	25%
Athens	21%	25%	13%	32%
Dublin	4%	20%	13%	11%
Düsseldorf	8%	17%	9%	16%
Berlin	0%	0%	9%	18%
Ljubljana	0%	0%	9%	17%
Oslo	9%	29%	8%	18%
Cologne	100%	100%	8%	20%
Helsinki	0%	0%	7%	22%
Hamburg	3%	25%	6%	19%
Bucharest	0%	14%	6%	12%
Stockholm	8%	32%	6%	17%
Barcelona	0%	0%	5%	24%
Riga	2%	20%	5%	9%
Warsaw	1%	20%	5%	13%
Prague	0%	0%	5%	9%
Sofia	100%	100%	96%	92%
Lyon			96%	92%
Stuttgart	100%	89%	98%	80%
Vilnius	100%	83%	100%	92%
Wien	100%	86%	100%	100%
Geneva	83%	70%	100%	100%
Bratislava	100%	100%	100%	100%
Budapest	99%	83%	100%	100%
Hannover	0%	0%	100%	100%
Tallinn	98%	88%	100%	100%
Bristol	100%	100%	100%	100%
Basel	100%	100%	100%	100%
Zagreb	100%	100%	100%	100%
Venice	100%	100%	100%	100%
Timisoara	100%	100%	100%	100%
Poznan	100%	100%	100%	100%
Skopje	100%	100%	100%	100%
Msida	100%	100%	100%	100%

Nuremberg	100%	100%	100%	100%
Gothenburg	100%	100%	100%	100%
Wroclaw	100%	100%	100%	100%
Thessaloniki	93%	75%	100%	100%
Tirane	100%	100%	100%	100%
Turin	100%	100%	100%	100%
Strasbourg			100%	100%
Gdansk			100%	100%
Dortmund			100%	100%
Salzburg			100%	100%
Katowice			100%	100%
Rijeka			100%	100%
Rotterdam	50%	50%		
Pristina	100%	100%		

Data sources: Telegeography 2007, ICAO 2008

Table A5.3: Betweenness and eigenvector centrality

	Betweenness								**Eigenvector**							
	Internet 2001		Aviation 2001		Internet 2006		Aviation 2006		Internet 2001		Aviation 2001		Internet 2006		Aviation 2006	
London	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*	100	*1*
Paris	49	*2*	33	*2*	35	*3*	50	*2*	77	*2*	63	*2*	70	*2*	67	*2*
Amsterdam	17	*4*	23	*4*	13	*5*	33	*4*	66	*3*	59	*3*	52	*4*	54	*3*
Frankfurt	22	*3*	27	*3*	64	*2*	39	*3*	59	*4*	46	*4*	55	*3*	47	*4*
Madrid	2	*20*	5	*8*	4	*13*	22	*5*	8	*9*	36	*5*	22	*5*	41	*5*
Milan	12	*7*	0	*22*	12	*6*	9	*8*	9	*8*	17	*12*	9	*8*	34	*6*
Rome	0	*39*	0	*26*	0	*39*	1	*20*	0	*37*	18	*10*	0	*52*	31	*7*
Barcelona	1	*29*	0	*21*	1	*22*	4	*11*	1	*18*	25	*6*	3	*17*	30	*8*
Copenhagen	5	*12*	8	*6*	8	*8*	3	*14*	11	*6*	23	*7*	10	*7*	25	*9*
Munich	3	*17*	2	*14*	0	*36*	14	*6*	1	*21*	20	*9*	1	*28*	24	*10*
Lisbon	15	*5*	7	*7*	9	*7*	11	*7*	1	*22*	17	*11*	3	*19*	21	*11*
Athens	1	*28*	4	*11*	5	*12*	5	*10*	1	*19*	16	*13*	2	*21*	19	*12*
Stockholm	9	*9*	4	*9*	2	*19*	0	*23*	10	*7*	20	*8*	5	*13*	19	*13*
Prague	2	*21*	4	*10*	3	*15*	9	*9*	1	*23*	10	*21*	3	*18*	18	*14*
Geneva	5	*13*	0	*31*	0	*28*	0	*39*	6	*11*	10	*22*	7	*12*	15	*15*
Oslo	8	*10*	2	*16*	0	*33*	0	*27*	3	*14*	15	*15*	3	*20*	13	*16*
Berlin	0	*31*	0	*39*	0	*39*	0	*24*	0	*26*	9	*24*	0	*44*	11	*17*
Manchester	0	*40*	1	*18*	0	*39*	3	*13*	0	*42*	12	*19*	0	*43*	11	*18*
Helsinki	2	*26*	0	*23*	0	*39*	0	*28*	1	*20*	12	*17*	1	*27*	11	*19*
Warsaw	3	*18*	2	*15*	6	*11*	1	*21*	0	*25*	7	*26*	5	*15*	11	*20*
Dublin	2	*23*	3	*12*	2	*17*	0	*29*	2	*15*	14	*16*	8	*10*	10	*21*
Brussels	13	*6*	0	*25*	0	*25*	0	*34*	41	*5*	9	*23*	17	*6*	10	*22*
Düsseldorf	0	*40*	1	*19*	0	*37*	1	*22*	6	*10*	11	*20*	9	*9*	10	*23*
Hamburg	4	*16*	0	*41*	3	*14*	0	*25*	4	*13*	8	*25*	8	*11*	8	*24*
Venice		*40*	0	*34*	0	*39*	0	*42*			5	*30*	0	*46*	7	*25*
Stuttgart	0	*40*	0	*30*	1	*21*	3	*12*	0	*41*	6	*28*	0	*39*	7	*26*
Wien	12	*8*	0	*37*	14	*4*	0	*36*	1	*17*	6	*27*	4	*16*	7	*27*
Cologne	0	*40*	0	*27*		*39*	2	*19*	0	*38*	4	*32*			6	*28*
Bucharest	0	*37*	2	*13*	0	*24*	2	*18*	0	*31*	4	*33*	1	*24*	6	*29*
Lyon	2	*23*	0	*29*		*39*	0	*32*	0	*29*	6	*29*			6	*30*
Zürich	5	*14*	9	*5*	2	*18*	0	*30*	5	*12*	16	*14*	5	*14*	5	*31*
Sofia	6	*11*	0	*45*	0	*38*	2	*15*	0	*33*	1	*45*	0	*35*	3	*32*
Marseille		*40*	0	*33*	2	*16*	0	*41*			2	*35*	0	*36*	3	*33*
Vilnius	2	*19*	0	*35*	0	*30*	2	*16*	0	*44*	1	*41*	0	*33*	2	*34*
Budapest	1	*30*	0	*44*	6	*10*	0	*33*	0	*24*	0	*51*	1	*22*	2	*35*
Tallinn	0	*40*	0	*45*	0	*32*	0	*35*	0	*28*	1	*43*	1	*29*	2	*36*
Riga	2	*22*	0	*32*	0	*35*	0	*31*	0	*40*	1	*40*	0	*34*	2	*37*
Ljubljana	2	*27*	2	*17*	2	*20*	2	*17*	0	*36*	1	*44*	0	*32*	2	*38*
Zagreb	5	*15*	0	*24*	8	*9*	0	*37*	0	*32*	2	*36*	0	*38*	2	*39*
Bristol		*40*	0	*45*	0	*39*	0	*44*			0	*54*	1	*23*	2	*40*
Hannover		*40*	0	*40*	0	*39*	0	*44*			4	*34*	1	*25*	2	*41*
Bratislava	0	*33*	0	*45*	0	*29*	0	*26*	0	*27*	0	*55*	0	*30*	2	*42*
Gothenburg		*40*	1	*20*	0	*39*	0	*43*			4	*31*	0	*41*	1	*43*
Turin	0	*40*	0	*45*	0	*39*	0	*44*	0	*45*	2	*37*	0	*48*	1	*44*
Thessaloniki		*40*	0	*38*	1	*23*	0	*38*			0	*47*	0	*42*	1	*45*
Nuremberg		*40*	0	*45*	0	*39*	0	*44*			1	*39*	0	*50*	1	*46*
Basel	0	*40*	0	*45*	0	*39*	0	*44*	0	*35*	0	*46*	0	*31*	1	*47*
Msida	0	*35*	0	*45*	0	*27*	0	*40*	0	*47*	1	*42*	0	*45*	1	*48*
Timisoara		*40*		*45*	0	*39*	0	*44*					0	*37*	0	*49*

Wroclaw		*40*		*45*	0	*39*	0	*44*					0	*40*	0	*50*
Katowice		*40*	0	*45*		*39*	0	*44*		*##*	0	*48*			0	*51*
Nice	0	*38*	0	*28*	0	*39*	0	*44*	0	*43*	12	*18*	0	*54*	0	*52*
Poznan		*40*	0	*42*	0	*39*	0	*44*			0	*53*	0	*47*	0	*53*
Rijeka		*40*		*45*		*39*	0	*44*							0	*54*
Dortmund	0	*40*	0	*45*		*39*	0	*44*	0	*39*	0	*50*			0	*55*
Gdansk	0	*40*	0	*45*		*39*	0	*44*	0	*34*	0	*49*			0	*56*
Strasbourg	2	*23*		*45*		*39*	0	*44*	2	*16*					0	*57*
Salzburg		*40*		*45*		*39*	0	*44*							0	*58*
Skopje	0	*32*	0	*36*	0	*34*	0	*44*	0	*47*	0	*52*	0	*51*	0	*59*
Tirane	0	*34*		*45*	0	*26*	0	*44*	0	*46*			0	*49*	0	*60*
Rotterdam	0	*36*	0	*45*	0	*39*		*44*	0	*30*	1	*38*	1	*26*		
Pristina		*40*	0	*43*	0	*31*		*44*			0	*56*	0	*52*		

Data sources: Telegeography 2007, ICAO 2008

Figure A5.1: GTS European backbone network in Europe
Source: GTS 2009

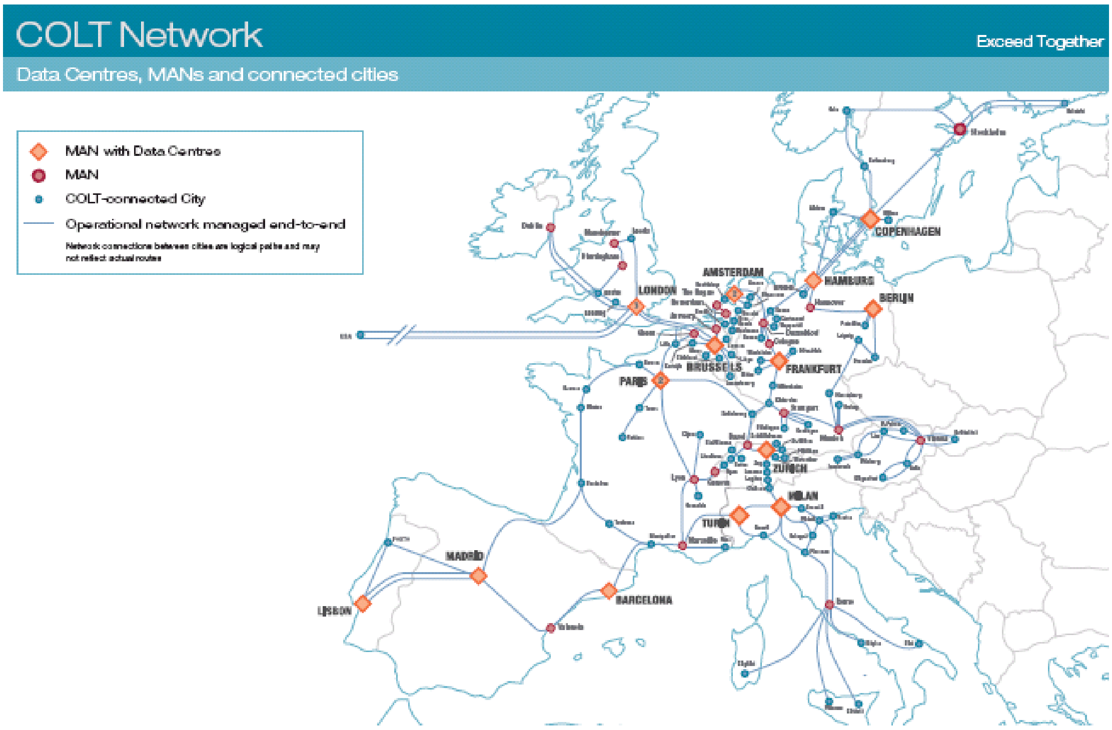


Figure A.5.2: COLT backbone network in Europe
Source: Colt 2009

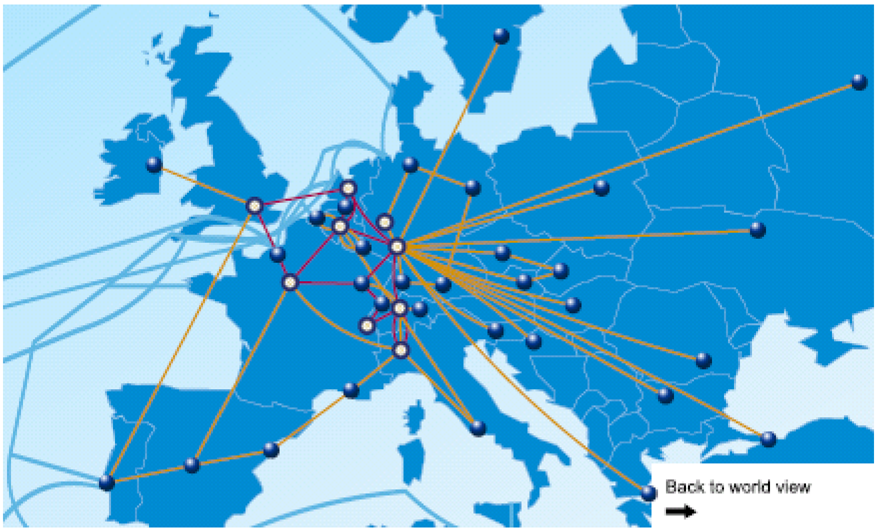


Figure A5.3: Belgacom backbone network in Europe
Source: Belgacom 2009

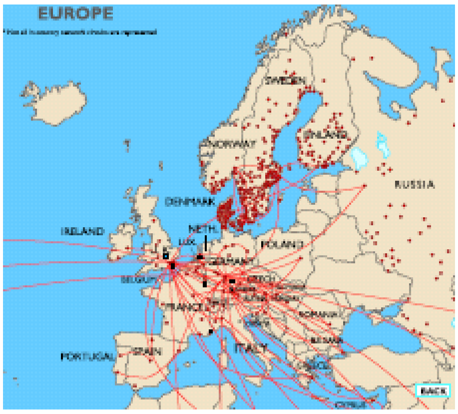


Figure A5.4: AT&T backbone network in Europe
Source: AT&T 2009

Chapter 6

Explanatory analysis of the spatial distribution of the Internet backbone provision in Europe

6.1 Introduction

This chapter attempts to explain the factors that determine the spatial distribution of Europe's Internet backbone networks[26]. As analyzed before in Chapter 2, these backbone networks can be regarded as the infrastructural underpinning that enables the Internet to function, seamlessly and apparently place-lessly from the viewpoint of the user.

The aim of this chapter is then to identify, through the use of statistical methods, the factors that influence the likelihood of European cities being connected to the Internet's backbone networks. In fact three measures of connectivity were used (described in more detail below in section 6.3): firstly, whether a city-region is *connected or not* to one or more backbone networks; secondly, the *level of connectivity* of those city-regions that are connected to at least one Internet backbone; and thirdly, the *number* of different backbone networks that a city is connected to.

The next part outlines the data used for the analysis and section 6.3 describes the quantitative methodology used in this chapter in order to explain the spatial distribution of backbone networks in Europe, the basic elements of which are principal components analysis and regression analysis. The results of the analysis are presented in 6.4, and the chapter finishes by presenting the main conclusions of this explanatory analysis.

6.2. Internet backbone data

The data for the Internet backbone networks used in this chapter is not based on the Telegeography (2007) dataset, but on a similar dataset provided by KMI Research Group (2001). This dataset was initially provided as a map of the different pan-European Internet backbone networks built or planned in 2001 by the KMI Research Group to CURDS for the needs of ESPON 1.2.2 Project (ESPON 2005b). According to KMI, the pan-European networks included in this dataset refer to those backbone providers that installed their own fibre optic cable in more than one European country (ibid). This is the first difference with the Telegeography (2007)

---

[26] This chapter extensively draws from Tranos and Gillespie (2009).

data used in the previous chapters: while the Telegeography dataset includes only the aggregated (it is not possible to distinguish the networks of different backbone providers) international links, the KMI dataset also includes the intra-country links as long as they are part of an international backbone network. The initial map of KMI (see extract from this below in Figure 6.1) presented all the different Pan-European backbone networks contrary to the aggregated approach of Telegeography's data. The data extracted from this map was used for the needs of the ESPON project (ESPON 2005b) but also for other publications such as Rutherford et al. 2004 and 2005, Rutherford forthcoming and Schintler et al. 2004. For the needs of this doctoral thesis, only the extracted data was available.

The second difference between the two data sources lies in the main measure of the data extracted from the KMI map that is *connectivity*. The latter is defined here as the number of international backbone connections that each city is served by (no bandwidth is included in this dataset). However, the definition of connectivity in the extracted data from KMI map is quite different from the degree centrality which was used in Chapters 4 and 5. According to the KMI dataset the connectivity of each city is measured as the number of cities which can be reached directly or indirectly from the origin one without using more than one backbone network for each route. For example, although Luxemburg according to our binary degree centrality measure developed in the previous chapters would have centrality equal to 6 (i.e. 6 direct connections to Frankfurt, Metz, Folkstone, Brussels, Namur and Liege as is highlighted in Figure 6.1), according to the data extracted from the KMI map data, Luxemburg is connected to 125 city-regions in Europe using 233 redundant[27]direct and non-direct links, owned by 4 different backbone providers. According to the measure of connectivity adopted here, Luxemburg's connectivity is equal to 233 and underlies the hypothesis that as long as two cities are connected with the same backbone network, their network distance is equal to 1 regardless of the intermediate hops. Additionally, this measure of connectivity incorporates the notion of redundancy. The above two differences resulted in a rather low correlation coefficient (0.327 and significant at 0.05 level) between the binary degree centrality based on the European links and the connectivity based on KMI's data. This mostly

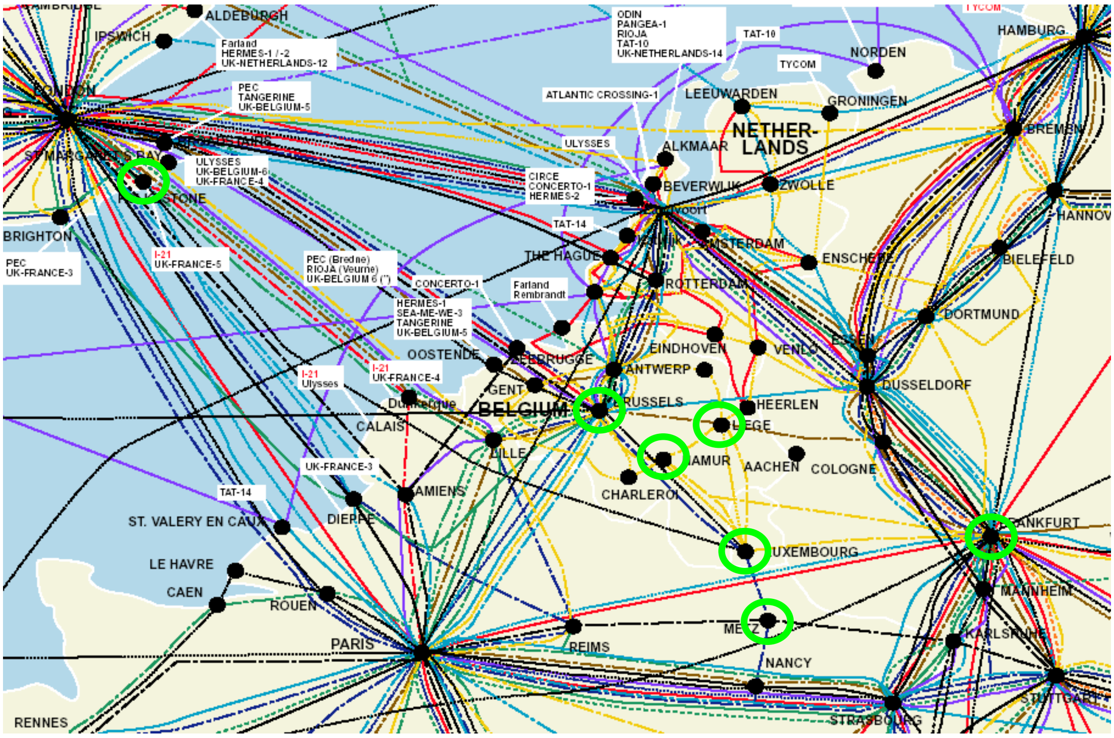
---

[27] The redundant links between any two cites refer to the overlapping links connecting those two cities and can guarantee Internet connectivity even if one of them is down (Gorman and Malecki, 2002).

reflects the different connectivity/centrality measures but also the difference between the two sources (Telegeography and KMI) in mapping the international and domestic Internet backbone networks as well as the gaps between the two data sources[28]. However, both of them can be approached as different measures and representations of the aggregated Internet backbone networks in Europe. While degree centrality reflects the topology of the Internet backbone network, the connectivity measure based on KMI's data is sensitive to the network extent of each different backbone provider.

The discontinuity because of the use of a different data source for the analysis presented in this chapter can be justified by the fact that KMI does not provide such data any more. The analysis presented in this chapter which is based on the KMI data took place in the initial stages of this doctoral research, when the Telegeography data, which is used in the rest of the doctoral thesis, was not available. Despite the above differences, the analysis presented fits well with the needs of this doctoral thesis as it provides an explanatory analysis of the geography of the Internet

Figure 6.1: Extract from KMI' map



Source: Rutherford et al. 2005

---

[28] 69 city-regions were connected with a least one backbone network in 2001 according to Telegeography (2007) and 65 according to KMI Group (2001). However, only 48 city-regions were present in both databases.

infrastructure. Such an analysis is the intermediate stage between the descriptive analysis of the geography of the Internet backbone network in Europe and the final part of this research which concerns the developmental impacts of such infrastructure.

6.3 Methodology

The first step for the explanatory analysis was to construct a database of socio-economic variables that were hypothesised as being likely to exert an influence on the geography of the Internet's backbone across Europe's urban system. A data set of 37 socio-economic variables for EU25 NUTS3 and 27 variables for NUTS2 regions was established (the choice of NUTS2 and NUTS3 regions is explained below). A description of these variables, the data sources and the time reference can be found in the Annex of this chapter. What should be noted here is that the main modelling strategy was to collect as many socio-economic variables as possible, for which there are reasons derived from theory that they will help explain the Internet's geography. The selected variables can be grouped into the following thematic areas:

- *Development level.* A number of variables were selected to test the proposition that backbone networks will be located in cities with advanced levels of development. The indicators selected include whether or not the city is part of an Objective 1 region in EU policy terms (i.e. classified as under-developed); its level of GDP and change in GDP; its population level and change in population; and whether or not the city is located in the core of Europe, the 'pentagon', which contains 14% of the EU27 area, accommodates 32% of Europe's population and produces 43% of its GDP (ESPON, 2005a).

- *Services and the Knowledge economy.* A number of variables were selected to test the proposition that the level of development and sophistication of the service sector and the knowledge economy is one of the most significant factors in attracting backbone networks to a city (Malecki 2004). For the NUTS3 spatial level the only available relevant indicators were the percentage of total employment in the service sector; service sector GVA; and service GVA per employee. However, more indicators related to the knowledge economy, such as employment in various knowledge-intensive sectors as well as education levels, were available for larger NUTS2 regions, so a database was also constructed at this spatial level.

- *Spatial structure***.** The third group of variables were selected to test the influence of spatial structure – including levels of urbanisation, population density, and levels of artificial land surfaces, as well as locations on coasts and near national borders, on the distribution of Internet backbone networks.
- *Physical transport and accessibility.* The final group of variables were selected to test the extent to which Internet infrastructures are co-located with transport infrastructures, and/or are located in cities/regions with high levels of physical accessibility. The variables selected include the number of commercial airports, seaports, the length of the road network, the length of the rail networks, levels of population accessibility (by car, air and rail), and accessibility times to market.

The analysis was conducted for the EU25, rather than for the whole of Europe, as regional data-sets for non-EU countries usually have many gaps. Although the previous discussion would imply that data at the level of urban areas would be most appropriate, given that the concern is with infrastructures connecting urban centres, the lack of comparable socio-economic data for European urban areas led to the NUTS3 and NUTS2 regional levels being chosen instead. Comparability requirements also resulted in effort being expended to select variables relating to the year 2001, or as near to that year as possible[29].

Turning now to the methods for analysing the data, the explanatory analysis undertaken was based on statistical modelling techniques and more specifically on different regression analysis methods. Instead of using the whole set of the independent variables collected as the regressors in the modelling procedure, an alternative method was selected in order to avoid multicollinearity problems, which would have occurred if all the explanatory variables were included in the regression models. Principal Components Regression (PCR) is a combination of two different methods, Principal Components Analysis (PCA) and Regression Analysis and it is

---

[29] According to ESPON (2005b) this is the year that the roll out of those networks stopped. This timing will be related with the dot-com bubble burst and the subsequent lack of willingness by telecommunications companies to invest in new technologies after the 2001 crash.

known in the relevant literature as one of the few methods in order to surpass multicollinearity problems (Massy 1965, Mardia et al 1979, Afifi and Clark 1996, Liu et al 2003, Fekedulegn et al 2002, Filzmoser and Croux 2002, Basilevsky 1994, Abdul-Wahab et al 2005). In more details, PCR is a two step method. First a set of principal components is calculated using ordinary PCA. These components are linear combinations of the original independent variables. In addition, because of the orthogonal transformation that takes place during the PCA process, the components are uncorrelated and consequently no multicollinearity problems appear in the subsequent regression analysis. Then a selected number of the principal components replace the original independent variables as the new regressors (Filzmoser and Croux 2002, Fekedulegn et al 2002). The main objective of the components' selection process is to eliminate the non-significant principal components and it is based on stepwise regression procedure (Fekedulegn et al 2002, Abdul-Wahab et al 2005, Filzmoser and Croux 2002, Liu et al 2003). At the end, regression analysis is carried out and the regression coefficients for the reduced set of orthogonal components are calculated. The latter can be mathematically transformed using the linear relations between the initial independent variables and the orthogonal components, resulting in this way to a final equation with the original independent variables. However, as Mardia et al (1979, 244) highlighted "[i]f the principal components have a natural intuitive meaning, it is perhaps best to leave the regression equation expressed in terms of the components". Although PCR was initially used in science, recent applications of the approach can also be found in social science (Sufian 2005) as well as in the field of urban and regional studies (Rodríguez-Pose and Crescenzi 2008, Blume and Sack 2008).

Three explanatory statistical models were produced, using the above method. The first is based on logistic regression and tries to explain the likelihood of a NUTS3 region being connected with at least one backbone network. The dependent variable for this model, which is based on KMI Research Group Maps (2001), is a binary variable indicating whether a region is part of at least one backbone network or not. The data for this variable was initially provided for cities and was then converted into a NUTS3 regional level measure across the whole of the EU25. The independent variables for this logistic regression model are the principal components produced from the PCA, which was applied to the socio-economic variables data set of the 1206 NUTS3 EU25 regions.

The second model is also at NUTS3 level, but it is focused only on the interconnected regions (i.e. those with at least one backbone network node within them) and tries to explain the socio-economic factors that affect a connected region’s level of connectivity, using a multiple linear regression model. The dependent variable is the number of redundant connections that a region shares with the rest of Europe. For example, Hamburg region shares 894 redundant links with 200 cities, which are located in 175 NUTS3 regions, while Naples is connected with 5 cities, which are located in 5 regions. As before, this variable is based on KMI Research Group Maps (2001) and was originally provided at the urban level. So, a summarization took place at NUTS3 level, excluding the intra-region connections. The independent variables are the result of the PCA which was applied to the socio-economic data set for the 184 interconnected NUTS3 regions.

The third model tries to explain the number of different Internet backbone providers present in each region (only including those that are connected), which can be regarded as another expression of regional connectivity, using again multiple linear regression, but this time at the NUTS2 level. The shift from NUTS3 to the larger NUTS2 regions took place in order to use variables related to the knowledge economy, which are not available at NUTS3 level. As before, the independent variables are the components that resulted from PCA applied to the socio-economic dataset for the 139 interconnected NUTS2 regions. The dependent variable is the number of different ISPs with at least one network node in each region, which again is based on KMI Research Group Maps (KMI Research Group, 2001).

6.4 Results

The results of the models described above are presented here. Model I is the logistic regression model. The first step was the exclusion of highly correlated variables (Pearson > 0.9 or <-0.9). After some additional tests, some more variables were excluded because of multicollinearity problems, resulting in a final set of 27 variables. PCA was carried out on these variables for the 1206 NUTS3 regions, resulting in six principal components being identified, which together explained

67.5% of the total variance and fulfilled all the proposed tests for the validity of this type of analysis (Field 2005).

A rotation of the six principal components then took place in order to distribute better the initial variables across the components. The method used for the latter was varimax, an orthogonal rotation method and the most widely used. This procedure resulted in the components loadings presented in Table 6.1, which help us interpret the principal components, since they represent the correlation between the initial variables and the components.

The *first component* is associated with *urban regions with high levels of market accessibility*. It is most highly correlated with regions with above levels of population density, with regions classified as urban, with regions with high levels of potential accessibility by air and by rail, and, to a lesser extent, regions located in the European ‘pentagon’.

The *second component* is labelled as *Europe’s prosperous core,* as it identifies prosperous, developed regions in the territorial heartlands of northern Europe. It is positively associated with non-objective 1 and non-lagging regions and with regions located in Europe’s pentagon, GDP per capita, and with population growth in the 1995-2003 period and negatively associated with high development growth rates.

The *third component* identifies the *major urban centres and transport hubs*, which are characterized by high levels of total population, by high levels of endowment in transport infrastructure (railways, roads, large airports), and by high levels of total service sector GVA.

The *fourth component* identifies the *most urbanised regions*, which are characterised by the highest population densities and by the largest percentage of artificial surface, as well as by high levels of GDP per capita and high levels of total service sector GVA. The *fifth component* identifies mainly *inland regions*, since it is negatively correlated with the number of seaports and the number of airports (which in abundance usually characterise island regions), and positively correlated with non-coastal regions and with potential accessibility by rail. The *sixth and final component* identifies *service-dominated regions*, which are characterized by a high employment share in the service sector.

Table 6.1: Model I, PCA

Principal Components **Variables**	**PC 1:** Urban regions with high levels of market accessibility	**PC 2:** Europe's prosperous core	**PC 3:** Major urban centres and transport hubs	**PC 4:** Most urbanised regions	**PC 5:** Inland regions	**PC 6:** Service-dominated regions
human_intervention (binary)	0.841					
urban_influence (b)	0.835					
Settlement (b)	0.771					
P_access_air	0.707	0.438				
P_access_rail	0.663	0.442			0.466	
D_market_access_car	0.608	0.454				
acc_typo (b)	0.594					
obj1 (b)		0.819				
srvc_productivity		0.799				
Lagging (b)		0.760				
gdp_ppp02_cap		0.596		0.571		
Pentagon (b)	0.486	0.560				
pop9503		0.501				-0.409
gpd9802euro_cap		-0.500				
pop			0.809			
railway			0.670			
road			0.637			
gva_srvc			0.629	0.566		
traffic_airports			0.580			
pop_density				0.851		
artificial_srfc	0.424			0.741		
Coast (b)					0.793	
seaports					-0.718	
airports			0.484		-0.506	
Border (b)						
empl_ndstr						-0.753
empl_srvc						0.680

b = binary variable

The next step was to feed a logistic regression model, based on the Backward Stepwise method, with the above principal components. After three steps, the model selected the following four components as contributing most to predicting the likelihood of a region being connected to a backbone network; the *major urban centres and transport hubs* (Component 3) and the *most urbanised regions* (Component 4) were the most important, followed at some distance in importance by the *inland regions* (Component 5), though this was *negatively* associated with being connected to an Internet backbone, and the *service-dominated regions* (Component 6). Because there is no R^2 for the Logistic Regression, the Nagelkerke R^2 is used in order to test the model's goodness of fit. For this case it is 0.364, a value that could be regarded as acceptable for logistic regression models. Table 6.2 presents the components that were finally entered in the logistic regression, and their main statistics. The residuals of the regression model do not cause any concern, since the only outliers (studentized residuals greater than 3) are the regions of West Inner London in UK and Hauts de Seine in France. The odds in both cases are overestimated because neither of these regions have an Internet backbone node within them but they both share many socio-economic characteristics with the highly interconnected neighbouring metropolitan regions of London and Paris (and from which they are likely to derive access to the Internet backbone though Metropolitan Area Networks).

Table 6.2: Model I, logistic regression model for NUTS3 regions

Variables	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
PC 3	0.982	0.102	93.405	1	0.000	2.671	2.188	3.259
PC 4	1.027	0.109	87.995	1	0.000	2.793	2.253	3.461
PC 5	-0.571	0.085	45.560	1	0.000	0.565	0.479	0.667
PC 6	0.458	0.101	20.591	1	0.000	1.581	1.297	1.927
Constant	-2.163	0.110	384.959	1	0.000	0.115		

So, the likelihood of an EU NUTS3 region being interconnected with at least one backbone network is greater if it is a major urban centre and transport hub; if it is a highly urbanised region with high per capita GDP and a high level of service sector GVA, and if its employment is service-dominated. This confirms of course the expectation that being connected to an Internet backbone is primarily a metropolitan phenomenon; a region's degree of metropolitan-ness is a more powerful predictor of whether it will be connected to a backbone network than is its location with respect to the geographical core of Europe or its level of wealth per se.

In addition, and less expectedly, it seems that the location of a region on Europe's coast increases that region's likelihood of being connected to backbone network. This phenomenon emerges both in the well developed countries in terms of ICTs, which also happen to be primarily coastal, such as Denmark, Finland and the UK, but also in the "gateway cities for high-bandwidth backbone connections" (Rutherford et al 2004, 19), whose connectivity may take place because they act as gateways for the backbone networks' onward connections. The latter refers to cities such as Bari in Italy, which is the gateway city for the Greek submarine broadband connection, or the French west coast, which connects UK with continental Europe. In addition, coastal regions' connectivity in terms of Internet infrastructures probably also mirrors their transport connectivity, in that the roll-out of backbone networks follows the previous layers of network infrastructure (which tend to run along low-lying coasts rather than inland, across mountains) simply because it is easier to install fibre cables next to or underneath an existing road or rail network rather than building a new network from scratch (see for example Rutherford 2005).

The second of our three models attempts to explain the factors that determine not the presence or absence of a backbone connection amongst all of Europe's 1206 NUTS 3 regions, but rather the degree of inter-connectedness of the (considerably fewer) regions with at least one backbone connection. The measure of inter-connectedness used is the number of redundant backbone connections that interconnected NUTS3 regions share with all the other interconnected ones. It refers thus to the 184 interconnected NUTS3 regions, and is based on linear regression, rather than the logistical regression of the first model.

Following the same methodology as the previous model, after the correlation tests, the 27 non-correlated variables were entered in a PCA model, which resulted in 7 principal components, which together explain 74% of the total variance (Table 6.3). It should be highlighted here that although the set of independent variables for this model is the same as for the previous one, the PCA results are different because (a) the number of regions included in this model is much smaller compared with model I (i.e. 1206 NUTS3 in model I and only 184 in model II), and (b) because the dependent variable is different (that is, the level of inter-connection of connected regions, rather than the presence or absence of a connection which constituted the dependent variable in Model I).

The *first component* identifies *urbanized and accessible regions*, with the highest correlations being with the binary variables representing above average shares of artificial surfaces, above average population densities and regions typologised as densely populated city core regions. Relatively high correlations are also found for potential accessibility by air and by rail, and for daily market accessibility by car. The *second component* identifies *Europe's largest metropolitan regions*, with the highest correlations being with measures of metropolitan scale; the size of total service sector GVA, population size and with the amount of traffic in commercial airports (measured in millions of passengers per year). Relatively high correlations are also found with potential accessibility by air, a typology identifying central or very central regions in multimodal accessibility terms, and with population density. The *third component* identifies Europe's *small urbanized tertiary centres*. Positive correlations are found with population density, the share of employment in the service sector and the percentage of artificial surfaces, while negative correlations are found with population size and the length of road and rail networks (probably acting here as surrogates for the geographical scale of the region). The *fourth component* identifies Europe's *inland accessible regions*. It is positively correlated with non-coastal regions and with potential accessibility by rail and daily market accessibility by car, but it is negatively correlated with the number of seaports and airports. The *fifth component* identifies Europe's *dynamic, prosperous regions*, since it refers to regions which are non-lagging regions and non-objective 1, with high levels of per capita GDP, high levels of GVA per employee in services, and experiencing population growth. The *sixth component* identifies *established tertiary centres*, which are characterized by high levels of GVA per employee in services and a high percentage accounted for by service activities, as well as a relatively high potential accessibility by rail, but which are negatively correlated with GDP growth and with the share of employment in secondary sector industries. The *seventh* and final component identifies *border regions*, which are characterised by border locations within countries.

The components were entered into a stepwise linear regression model, which achieved an R^2 of 55%. Regression's coefficients and the main statistics for this model can be found in Table 6.4. Regarding the regression's residuals, they do not create any concern. Only the region of Roma has a residual greater than 3 standard deviations, indicating an overestimation of the region's connectivity.

According to the results of the model, the regional characteristic that most positively influences Internet backbone connectivity is *metropolitan scale* (Component 2), followed by *established tertiary centres* (Component 6), *inland accessible regions* (Component 4), *urbanised and accessible regions* (Component 1), and *dynamic prosperous regions* (Component 5). Given the explanatory dominance of Component 2 in the overall model, it can be confirmed that the importance of metropolitan scale (as expressed in total service sector GVA, population size and the volume of commercial air traffic) to explaining the likelihood of a connected region having a high degree of backbone Internet connectivity with other connected regions.

Table 6.3: Model II, PCA

Principal Components	**PC 1:** Urbanized and accessible regions	**PC 2:** Europe's largest metropolitan regions	**PC 3:** Small urbanized tertiary centres	**PC 4:** Inland accessible regions	**PC 5:** Dynamic prosperous regions	**PC 6:** Established tertiary centres	**PC 7:** Border regions
human_intervention (b)	0.843						
urban_influence (b)	0.837						
Settlement (b)	0.816						
P_access_air	0.549	0.532					
gva_srvc		0.882					
pop		0.780	-0.410				
traffic_airports		0.644					
pop_density		0.486	0.472				
acc_typo	0.402	0.441					
road			-0.837				
railway			-0.755				
artificial_srfc			0.597				
seaports				-0.812			
Coast (b)				0.807			
P_access_rail	0.479			0.571		0.480	
airports				-0.502			0.454
D_market_access_car	0.441			0.483			
Lagging (b)					0.815		
obj1 (b)					0.773		
srvc_productivity					0.639	0.599	
gdp_ppp02					0.594		
pop9503					0.496		
gpd9802euro						-0.744	
empl_ndstr			-0.454			-0.634	
empl_srvc			0.450			0.590	
Pentagon (b)							
Boarder (b)							-0.663

b = binary variable

Interestingly, although the first model suggested that the likelihood of a region being connected with at least one backbone network is increased if it is located on the coast of Europe, when the focus is on the levels of connectivity of the (many fewer) inter-connected regions, as it is in this model, it is found that higher connectivity is associated with inland regions. Perhaps unsurprisingly given the scale of the European landmass and the complex history of its settlement and development, the most inter-connected urban regions are not, primarily, located around its extensive coastal periphery.

Table 6.4: Model II, linear regression model for NUTS3 regions

Variables	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95% Confidence Interval for B**	
	B	Std. Error	Beta			Lower Bound	Upper Bound
Constant	270.211	11.827		22.847	0.000	246.868	293.554
PC 2	136.939	11.860	0.585	11.546	0.000	113.531	160.347
PC 6	60.516	11.860	0.258	5.103	0.000	37.109	83.924
PC 4	53.022	11.860	0.226	4.471	0.000	29.614	76.430
PC 1	52.343	11.860	0.224	4.413	0.000	28.935	75.751
PC 5	48.669	11.860	0.208	4.104	0.000	25.261	72.076

The third and final model is another linear regression model which has as its dependent variable a different measure of connectivity, that of the *number of different backbone network providers* with at least one node in the region concerned. This is then a measure of how attractive regions are to the suppliers of backbone networks, which it is anticipated being in turn related to their expectations of levels of demand in particular regions, within the context of their commercial network deployment strategies. The spatial scale of the analysis also shifts, from NUTS3 to NUTS2 regions, with the analysis concerning the 139 interconnected NUTS2 regions. Although the shift from a larger number of NUTS3 regions to a smaller number of NUTS2 regions means that some geographical detail has been lost, this is compensated for by being able to include a larger number of independent variables relating to the knowledge economy, which are only available at the NUTS2 level (see table A2). While both this model and the previously described Model II are attempting to explain the level of backbone connectivity of connected regions, different results are anticipated because (a) a different measure of inter-connectivity is being used as the dependent variable; (b) because the analysis in being conducted for the 139 NUTS2 inter-connected regions rather than the 184 inter-connected NUTS3 regions, and (c) because the shift to the larger NUTS2 regions has enabled us to include a much wider array of independent variables relating to the knowledge economy intensity and characteristics of different areas, which are only available for these larger areas.

As with the previous cases, for the 22 non-correlated variables PCA was applied. This resulted in 5 principal components being identified (Table 6.5), which explain 73% of the total variance. The *first component* identifies *knowledge-intensive service regions*, being strongly positively related to the percentage of employment in knowledge intensive services (and specifically in market services, financial services and high-tech services), the share of employment in service industries, performance with respect to the Lisbon Agenda, the level of human resources in science and technology, GDP per capita, and, albeit much less strongly, accessibility. The *second component* identifies *major corporate and service hubs*, being positively related to the scale of service sector GVA, the level of population in total and the highly educated population, the number of headquarters from the top 1500 companies that can be found in the region, and the volume of traffic to the region's airports. The *third component* identifies Europe's *inland core regions*, as it is negatively correlated with the number of seaports and airports and positively correlated with inland regions, a general accessibility classification and with location in Europe's pentagon. The *fourth component* identifies *large transport-rich regions* (in terms of the length of road and railway networks within them) while the fifth and final component is slightly correlated with regions with *high employment rates*.

After entering the above components in a stepwise linear regression model, three components were identified that determine the number of Internet backbone providers present in a region, with a goodness of fit of 57%. Regression coefficients and the regression's main statistics are illustrated in Table 6.6. The results demonstrate that the number of Internet backbone providers operating within a region is positively associated with *knowledge-intensive service regions* (Component 1), with the *major corporate and service* hubs (Component 2), and with locations in *Europe's inland core* (Component 3). Regarding the regression's residuals, no concerns emerge since the only outliers are the regions of Hamburg and Île de France. The number of Internet backbone providers for the former is underestimated, since this region is the most well-connected one. On the contrary, Paris' connectivity seems to be overestimated by the model; despite its importance in the European urban hierarchy, it is not the most interconnected region in Europe in terms of the number of Internet backbone providers.

Table 6.5: Model III, PCA

Principal Components	**PC 1:** knowledge-intensive service regions	**PC 2:** Major corporate and service hubs	**PC 3:** Inland core regions	**PC 4:** Large transport-rich regions	**PC 5:** High employment rates
se_kis_tot	0.912				
empl_srvc	0.828				
Spatial classification_lisbon (o)	0.823				
se_kis_ms	0.812				
hrst	0.792				
Se_kis_ht	0.761				
gdp02ppp_cap	0.729				
Se_kis_fs	0.529				
gva_gp		0.891			
pop		0.888			
edu		0.864			
Top_1500_companies		0.827			
airport_trffc		0.701			
seaport			-0.754		
airport			-0.730		
Spatial classification_access (o)	0.436		0.684		
Coast (b)			0.666		
Pentagon (b)			0.648		
railways				0.835	
road				0.817	
empl_T	0.502				0.598
Spatial classification_tech (o)	0.513				-0.591

b = binary, o = ordinal

The results from the third model confirm the importance of the knowledge economy in shaping the Internet's geography. The regions in Europe with the highest number of Internet backbone providers are those with the highest incidence of knowledge-intensive services (particularly of market services, financial services and high-tech services), with the highest level of human resources in science and technology, with the highest levels of service sector GVA and with the highest number of corporate headquarters.

Table 6.6: Model III, linear regression model for NUTS2 regions

Variables	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95% Confidence Interval for B**	
	B	Std. Error	Beta			Lower Bound	Upper Bound
Constant	5.691	0.269		21.134	0.000	5.158	6.223
PC 1	2.598	0.270	0.543	9.613	0.000	2.063	3.132
PC 2	2.256	0.270	0.472	8.349	0.000	1.722	2.791
PC 3	1.094	0.270	0.229	4.049	0.000	0.560	1.629

6.5 Conclusions

This chapter has attempted to shed some light on the factors that determine the spatial allocation of the Internet's backbone networks in Europe. Such backbone networks form the infrastructural underpinning of the modern knowledge economy, and it is expected that the geography of the knowledge economy in Europe both shapes, and to an extent is shaped by, the spatial allocation of backbone networks. In a Europe in which the Internet can be accessed over a variety of widely deployed network technologies, including DSL over copper telephone lines, cable TV networks, Wi-Fi networks and 3G mobile phones, the Internet appears to be 'everywhere', to be ubiquitously available. However, from the examination of the places at which the Internet's usage is aggregated into nodal points and funnelled through fibre-optic cables, it becomes clear that the Internet has a distinctive geography, that it is not thinly spread and ubiquitous, but rather highly aggregated and geographically-differentiated. It is this aggregation and differentiation that is revealed by the Internet's backbone infrastructure; thus of the 1206 NUTS3 regions in our analysis, less than one-in-seven have an Internet backbone node within them, and of the relatively small proportion that do have a node within them, their level of connectivity with other regions and the number of providers operating with them are further highly differentiated. Using the results of the analyses presented above, the rest of the section attempts to explain the distinctive geography of Internet backbone network provision in Europe.

Table 6.7 summarizes the principal components which proved to be significant regressors for our three measures of backbone connectivity. For the first model, it was the major urban centres and transport hubs that emerged as having the highest likelihood of being connected to an Internet backbone, while for those NUTS3 regions that were connected (model 2), the highest levels of connectivity were associated with Europe's largest metropolitan regions. In the third model, in which the spatial scale changed to NUTS2 regions in order to open up a wider range of knowledge economy variables, it was clearly the most knowledge-intensive regions which emerged as having the highest number of backbone network providers.

In section 6.3 above, a number of types of socio-economic variables were advanced as influences upon the location of backbone networks; these were grouped into levels of development; services and the knowledge economy; spatial structure; and physical transport and accessibility. In each of our three models, all of these types of socio-economic variable emerged as having some explanatory power. Or, to put it another way, there are no mono-causal explanations for the Internet's backbone geography.

The *level of development* has some purchase in all three models, though it tended to be measures of the scale of development – such as the absolute size of population or GDP – that were more significant than relative measures of wealth or prosperity in influencing the Internet's backbone geography. The variables relating to *services and the knowledge economy* were of considerable explanatory importance, particularly in models 2 and 3 which concerned levels of connectivity, rather than connectivity as opposed to non-connectivity. Thus in model 2, measures of the service economy such as the share of employment in services and service sector productivity had explanatory power, while in model 3, in which a wider variety of knowledge economy measures could be included (because of being at NUTS2 level), the knowledge-intensity of employment, the incidence of higher level skills and the presence of corporate headquarters emerged as overwhelmingly important predictors of the number of Internet backbone providers.

Spatial structure also emerged as a prime predictor of the Internet's geography, particularly in the sense of levels of urbanisation and population densities. Thus being connected to the Internet's backbone is an overwhelmingly urban region phenomenon, while the levels of connectivity of regions which are connected is clearly related to their degree of 'metropolitan-ness'. Another, less expected, spatial variable which emerged as having relevance to understanding the Internet backbone's geography was location with respect to Europe's coastline; being connected to an Internet backbone was positively associated with coastal locations, whereas high degrees of connectivity in the relatively small number of connected regions was, on the contrary associated with inland, more centrally located regions.

Table 6.7: Principal components, which were included in the three regression models

Principal Components	**Model I (NUTS 3)**	**Model II (NUTS 3)**	**Model III (NUTS 2)**
Dependent Variable	Likelihood of being connected	Level of connectivity of interconnected regions	Number of backbone providers in connected regions
Most important Component	Major urban centres and transport hubs (+)	Europe’s largest metropolitan regions (+)	Knowledge-intensive service regions (+)
2nd Most important Component	Most urbanised regions (+)	Established tertiary centres (+)	Major corporate and service hubs (+)
3rd Most important Component	Inland regions (+)	Inland accessible regions (+)	Europe’s inland core regions (+)
4th Most important Component	Service-dominated regions (-)	Urbanized and accessible regions (+)	
5th Most important Component		Dynamic, prosperous regions (+)	

Finally, *physical transport and accessibility* also play a role in influencing the geography of the Internet; generally speaking, the backbone networks of the Internet tend to locate in regions which are already well provided with transport infrastructure and which have airports with substantial volumes of passenger traffic.

While it is clear that all four of these groups of variables have explanatory power in helping us to understand the geography of Internet backbone provision, it can be concluded that the factors with the greatest explanatory purchase are urban size, metropolitan status and knowledge-intensity. Although there is no simple or single explanation of the backbone geography in EU25, it is concluded that it is, nevertheless, both familiar and predictable, since it is reflects largely the existing spatial, development and knowledge economy structures of metropolitan Europe. The Internet and its backbone networks seem not challenge existing paths but rather to bolster the present metropolitan core areas of Europe.

Chapter 6

Annex

Table A6.1: Model 1 and 2 variables (NUTS3 Regions)

Thematic area		Variables	Description	source	time
	Model 1 dependent	network_0/I *(binary)*	1: existence of one or more backbone networks in the region; 0: no backbone network	ESPON 2005b (own transformation to binary variable)	2001
	Model 2 dependent	cities_cnnctnsSUM	Number of total connections between two regions	ESPON 2005b	2001
1	development level	obj1 *(binary)*	Eligible areas typology; 1: not objective 1 region; 0: objective 1 region	ESPON 2006	2000
2		pop	Annual average population	Eurostat 2006	2001
3		pop9503	Change in average population (%)	ESPON 2006	1995-2003
4		empl_ndstr	Employment in secondary sector (%)	Eurostat 2006	2001
5		gdp_ppp02_cap	GDP (PPP per capita)	ESPON 2006	2002
6		gpd9801euro_cap	Change of GDP (euro per capita)	ESPON 2006	1998-2002
7		gdp01_ppp	GDP at current market prices (mil. PPP)	Eurostat 2006	2001
8		pentagon *(binary)*	Pentagon typology; 1: region in pentagon; 0 region not in pentagon	ESPON 2006	2003
9		lagging *(binary)*	Lagging regions typology; 1: non lagging regions; 0: lagging and potential lagging regions according to GDP per inhabitant and unemployment rate	ESPON 2006 (own transformation to binary data)	2000

10	Services and knowledge economy	gva_srvc	Service sector GVA at basic prices (mil. Euros)	Eurostat 2006	2001
11		empl_srvc	Employment in service sector (%)	Eurostat 2006	2001
12		srvc_productivity	Productivity of service industries (GVA per employee in service industries)	Eurostat 2006 (own calculation)	2001
13	spatial structure	airports	Number of commercial airports	ESPON 2006	2001
14		seaports	Number of commercial seaports	ESPON 2006	2001
15		road	Length of road network	ESPON 2006	2001
16		railway	Length of rail network	ESPON 2006	2001
17		traffic_airports	Traffic in commercial airports (in million passengers/year)	ESPON 2006	2000
18		Connectivity_airports_car	Connectivity to commercial airports by car of the capital or centroid representative of the NUTS3 (in hours)	ESPON 2006	2001
19		Connectivity_seaports_car	Connectivity to commercial airports by car of the capital or centroid representative of the NUTS3 (in hours)	ESPON 2006	2001
20		Time_motorway	Time to the nearest motorway access, by car of the capital or centroid representative of the NUTS3 (in hours)	ESPON 2006	2001
21		D_pop_access_car	Daily population accessible by car (in clear	ESPON 2006	1999

		accessibility units)		
22	D_market_access_car	Daily market accessible by car in terms of GDP (mil. euros / capita * 1.000.000)	ESPON 2006	2000
23	P_access_air	Potential accessibility air, ESPON space = 100	ESPON 2006	2001
24	P_access_rail	Potential accessibility rail, ESPON space = 100	ESPON 2006	2001
25	P_access_road	Potential accessibility road, ESPON space = 100	ESPON 2006	2001
26	P_access_multimodal	Potential accessibility multimodal, ESPON space = 100	ESPON 2006	2001
27	Access_time_market_road	Accessibility time to market by road half-life mesoscale (25), weighted by Population	ESPON 2006	1997
28	Access_time_market_rail	Accessibility time to market by rail half-life mesoscale (25), weighted by Population	ESPON 2006	1997
29	Access_time_market_rail_road	Accessibility time to market by rail and road half-life mesoscale (25), weighted by Population	ESPON 2006	1997
30	urban_influence *(binary)*	Urban influence typology1: population density above average (107 inh./km2 in ESPON space) and/or at least European level FUA; 0: population density below average and no European level FUA.	ESPON 2006 (own transformation to binary data)	1996/1999

31	human_intervention *(binary)*	Human intervention typology; 1: share of artificial surfaces (and possibly some other land use) above average (3,48%); 0: share of agricultural (and possibly residual) land use above average (50,36%); Low: only the share of residual land use above average (46,16%)	ESPON 2006 (own transformation to binary data)	1996/1999
32	settlement *(binary)*	Urban - rural typology; 1: city core region, very densely populated, densely populated, city core region, densely populated region; 0: rural region, more densely populated rural region, less densely populated rural region.	ESPON 2006 (own transformation to binary data)	1999
33	pop_density	Population density	Eurostat 2006	2001
34	acc_typo *(binary)*	Multimodal potential accessibility typology; 1: very central, central; 0: intermediate, peripheral, very peripheral	ESPON 2006 (own transformation to binary data)	2001
35	coast *(binary)*	Coast region typology; 1: no coast, 0: coast	ESPON 2006	2003
36	border *(binary)*	National border region typology; 1: no border, 0: border	ESPON 2006	2003
37	artificial_srfc	Share of artificial surfaces (%)	ESPON 2006	1986-1996

For all the binary variables value 1 dedicates centrality and value 0 peripherality

Table A6.2: Model 3 variables (NUTS2 Regions)

Thematic area		Variables	Description	source	time
Model 3				ESPON 2005b	2001
dependent		D_Ntwrks	Number of different ISPs present in NUTS2 region		
1	Development level	empl_T	Total employment (percentage of active population)	Eurostat 2006	2001
2		gdp02ppp_cap	GDP (PPP per hab.)	ESPON 2006	2002
3		gdp01	GDP at current market prices (mil. PPP)	Eurostat 2006	2001
4		productivity	productivity (gdp per employer)	Eurostat 2006 (own calculation)	2001
5		pop	Annual average population	Eurostat 2006	2001
6		pentagon *(binary)*	Pentagon typology; 1: region in pentagon; 0 region not in pentagon	ESPON 2006	2003
7		top_1500_companies	Top-1500 companies headquarters location	ESPON 2005b	2003
8	Services and Knowledge economy	se_kis_tot	Total knowledge-intensive services (percentage of total employment)	Eurostat 2006	2001
9		se_kis_ht	Knowledge-intensive high-technology services (percentage of total employment)	Eurostat 2006	2001
10		se_kis_ms	Knowledge-intensive market services (excluding financial intermediation and high-tech services - percentage of total employment)	Eurostat 2006	2001

11		se_kis_fs	Knowledge-intensive financial services (percentage of total employment)	Eurostat 2006	2001
12		se_kis_ot	Other knowledge-intensive services (percentage of total employment)	Eurostat 2006	2001
13		g_h_p	Wholesale and retail trade, hotels and restaurants, private households (percentage of total employment)	Eurostat 2006	2001
14		frb	Financial intermediation, real estate, renting and business activities (without computers and R&D - percentage of total employment)	Eurostat 2006	2001
15		empl_srvc	Employment in service sectors (percentage of total employment)	Eurostat 2006	2001
16		hrst	Human Resources in Science and Technology (percentage of active population)	Eurostat 2006	2001
17		gva_gp	Service sector GVA at basic prices (mil. Euros)	Eurostat 2006	2001
18		edu	Population aged 15 at highest level of education attained	Eurostat 2006	2001
19		airport	Number of commercial airports	ESPON 2006	2001
20	Spatial structure	seaport	Number of commercial seaports	ESPON 2006	2001
21		road	Length of road network	ESPON 2006	2001
22		railways	Length of rail network	ESPON 2006	2001
23		Airport_trffc	Traffic in commercial airports (in million passengers/year)	ESPON 2006	2000

24	coast *(binary)*	Coast region typology; 1: no coast, 0: coast	ESPON 2006	2003
2525	Spatial classification_lisbon *(ordinal)*	Classified Lisbon performance; 1=highly below average; 2=below average; 3=average; 4=above average; 5=highly above average.	ESPON 2006	2001-2003
2626	Spatial classification_tech *(ordinal)*	Classified technological hazards; 1=highly below average; 2=below average; 3=average; 4=above average; 5=highly above average	ESPON 2006	2003
2727	Spatial classification_access *(ordinal)*	Classified accessibility; 1=highly below average; 2=below average; 3=average; 4=above average; 5=highly above average	ESPON 2006	2003

For all the binary variables value 1 dedicates centrality and value 0 peripherality

Chapter 7

Internet infrastructure and regional economic development

7.1 Introduction

This chapter focuses on testing the impact of the Internet infrastructure on regional economic development. Drawing on the conceptual framework presented in section 2.5.5, effort is spent here to empirically examine whether the Internet infrastructure, as it is reflected in the Internet backbone network, affects the level of economic development of Europe’s city-regions. As discussed in Chapter 3 and illustrated in figure 2.4, it is established nowadays in the relevant literature that ICTs have a direct impact on productivity at a macro level. And because of the Internet’s GPT characteristics, it is known that in order for these productivity gains to be achieved and diffused in the economy, there is a need for an infrastructural layer. Additionally, according to our analysis presented in Chapter 4 (and partially in Chapter 5), but also according to previous studies, which were extensively discussed in section 2.3, this Internet infrastructure which is responsible for the utilization of the productivity gains due to ICTs and the Internet, is unequally distributed across the nodes of the urban network. Consequently, the following question emerges: does the unequally distributed Internet infrastructure impact on the economic development level of Europe’s city-regions?

This chapter aims to address the above research question by using econometric analysis. Additionally, this chapter aims to shed some light on the well known problem of defining the direction of causality between infrastructure provision and economic development, which was discussed in section 2.2.5. The structure of this chapter is as follows: the next section briefly presents the method used and the data; section 7.3 presents the panel data analysis and section 7.4 the causality analysis; this chapter ends with discussion of the results.

7.2 Data and methods

In order to answer the above research question an empirical quantitative approach and modelling techniques have been adopted. The results of the network analysis of the Internet backbone network presented in Chapter 4 feed the econometric models, which assess the regional economic development impacts of the (unequally distributed) Internet backbone network. More specifically, the key point of this chapter is to appraise the explanatory value of the different centrality indicators produced in Chapter 4 in explaining the economic development level of city-regions in Europe over time.

In order to achieve this, a *panel data* approach is used. As was briefly described in Chapter 3, panel data refers to a two-dimension database where observations exist over time for a number of cross-section units. It is differentiated by simple cross-section data because it includes observations over time – contrary to a single point in time character of the cross-section. It is also differentiated by time series because it has observations for multiple cross-section units over time contrary to the one-unit approach of the time-series (Maddala 2001). Additionally, it is differentiated by the *repeated cross section* or *trend* data, which contains data for the same variables but for different cross-section units over time (Finkel 1995).

The panel data approach is preferred to a simple cross-sectional approach for various reasons: (a) panel datasets provide a large number of data points, increasing the degrees of freedom and reducing the collinearity among explanatory variables and as a result improve the efficiency of econometric estimates; (b) the longitudinal dimension of the panel data allow the analysis of a number of important economic questions which need sequential observations in order to be answered; and most importantly (c) panel data improves the researchers ability to control for missing or unobserved variables (Hsiao 2003). Such omitted-variable bias as a result of unobserved heterogeneity is common in cross-section models (Rodríguez-Pose and Tselios 2009a). The above advantages according to the theory but also the structure of the existing data, which is derived from the network analysis led in adopting the panel data approach.

In our case, the panel dataset consists of the different centrality indicators for the period 2001-2006 which derived from the network analysis of the Internet backbone. These observations are included for all the city-regions connected with at least one backbone and at least for one year during the period 2001-2006. The spatial unit used here is again NUTS3 regions. These variables will be used as the main independent variables in the regression models. In the first phase of our analysis, the dependent variable will be GDP per capita. In more detail, the regression analysis will model the impact of the independent variables (various centrality measures) on the dependent variable (GDP per capita). Additionally, the panel dataset also includes control variables such as employment in the service sector, the unemployment rate and information about whether the region was Objective 1 or not, whether it is located in Europe's pentagon and whether it is a coastal region. The last three are dummy and time invariant variables. While employment in the service sector is more linked with the knowledge economy, unemployment and Objective 1 regions are more linked with the level of regional economic development. The coastal location and the location in Europe's pentagon as was stated in Chapter 6 are linked with the backbone distribution. All these additional variables will perform the role of control variables and increase the robustness of our models. It should also be noted here again the difficulty in identifying relevant socio-economic variables for a scale as detailed as NUTS3 level for the whole of Europe. This explains the small number of control variables. Table 7.1 presents the variables used in this analysis as well as their main characteristics.

Another important point is the dynamic character of the models. This term refers to the time lag between the dependent and independent variables (Maddala 2001). This modelling strategy is introduced in order to address the *endogeneity* problem (Banerjee and Duflo 2003). The latter is defined as the "simultaneous determination of response variables and regressors" (Baum 2006, 185). In our case, the endogeneity problem can be interpreted as follows: the economic development of a region in a year t might not be affected by the provision of the Internet infrastructure in year t but rather in year t-1. Such complications should be expected because in order for an infrastructure or a new technology to be utilized there is always a need for an adoption period (see for example the discussion about the productivity paradox in section 2.5.2). Because of the rather narrow time dimension of the panel dataset, only one year time lag is used for this analysis.

Based on the above the panel data models will have the following form:

$$gdp\_pc_{i,t} = \beta_1 central\_\ln_{i,t-1} + \beta_2 x_{i,t} + u_{i,t} \qquad (7.1)$$

with *i* denoting city-regions (i = 1, 2, …, N) and *t* denoting time (2001, 2002, …, 2006); *gdp_pc* denotes the GDP per capita in euro and 2000 constant prices; *central_ln* denotes the natural logarithm of one of the centrality measures presented in Table 7.1; *x* is a vector of the control variables, employment in service sector, unemployment, location in objective 1 region, location in pentagon or in a coastal

Table 7.1: Variables

Variables	Variables	Years	Number of observations (total: 6x102=612)	Mean	Std. Dev.	Min	Max
D_intra_w_ln	Nat. logarithm of the weighted degree centrality of the intra-European links	2001-2006	437	7.569	2.986	0.693	13.870
D_intra_b_ln	Nat. logarithm of the binary degree centrality of the intra-European links	2001-2006	437	1.259	0.980	0.000	3.638
D_all_w_ln	Nat. logarithm of the weighted degree centrality of all the links	2001-2006	449	7.640	2.970	0.693	14.321
D_all_b_ln	Nat. logarithm of the binary degree centrality of all the links	2001-2006	449	1.437	1.114	0.000	4.615
btwnss_all~n	Nat. logarithm of the betweeness centrality of all the links	2001-2006	449	-1.622	7.963	-11.513	9.055
eigen_all_ln	Nat. logarithm of the eigenvector centrality of all the links	2001-2006	449	-7.432	3.465	-13.816	-0.472
gdp_pc	GDP per capita in euro, constant prices 2000	2001-2006	465	26,636	20,650	1,894	153,212
emp_gp	Percentage in employment in services	2001-2006	384	0.729	0.113	0.313	0.946
un	Unemployment rate	2001-2006	451	8.096	4.460	1.600	26.100
obj1	Objective 1 regions or not	time inviriant	528	0.250	0.433	0.000	1.000
pentagon	Location in Europe's pentagon	time inviriant	612	0.275	0.447	0.000	1.000
coast	Costal location	time inviriant	612	0.353	0.478	0.000	1.000

region; β are the coefficients and $u_{i,t}$ is the composite error, which can be further analysed $u_{i,t} = \upsilon_i + \varepsilon_{i,t}$ with υ_i denoting the time-invariant unobserved heterogeneity and ε_{it} the error term.

According to the relevant literature, there are three main modelling approaches in estimating a panel data model (Johnston and DiNardo 1997; Maddala 2001): the *pooled OLS*, the *fixed effects* model (FE), and the *random effects* model (RE). The former is the simplest one and its main characteristic is that it ignores the panel structure of the data and uses simple OLS to estimate the model. In reality, pooled OLS only differentiates from simple OLS as the latter has information about *N* observations while the former includes information about *N* observations for *T* time periods resulting in *NxT* total number of observations (Johnston and DiNardo 1997). The assumptions that underlie this method reflect the assumptions of the classic linear model namely that the composite error is uncorrelated with the dependent variables. Although pooled OLS is the simplest method, usually it is not appropriate for estimating panel data models (Johnston and DiNardo 1997).

A derivative of the pooled OLS is the FE model also known as least-squares dummy-variable approach (LSDV). In this estimation dummy variables are introduced to “account for the effects of those omitted variables that are specific to individual cross-sectional units but stay constant over time, and the effects that are specific to each time period but are the same for all cross-section units” (Hsiao 1986). More simply put, the FE estimation is not based on the variation *between* the different cross-section units but rather on the variation *within* the cross-section units, removing the bias of the unobserved heterogeneity occurred by omitted variables. This is achieved by using the main attribute of panel datasets, the cross-sectional observations over time. Based on this, instead of using the (7.1) the first difference of this equation is used for the estimation:

$$gdp\_pc_{i,t} - gdp\_pc_{i,t-1} = \beta(central\_\ln_{i,t-1} - central\_\ln_{i,t-2}) + \beta(x_{i,t} - x_{i,t-1}) + (u_{i,t} - u_{i,t-1}) \quad (7.2)$$

Because of the above subtraction all the time-invariant x control variables are dropped from the estimation. Additionally, the error term is only based on the ε_{it} as the u_i time invariant factor is also dropped because of the subtraction. This process results in unbiased coefficients estimation using OLS (Johnston and DiNardo 1997).

However, the main drawback of this estimation is that the cross-section variation is vastly downgraded (Rodríguez-Pose and Tselios 2009a).

The third suggested model for estimating panel data is the RE. This approach focuses on the serial correlation in the composite error and the model is estimated using a Generalised Least Squares framework (GLS) (Wooldridge 2003). Contrary to the FE approach where the effects of the omitted cross-sectional variables are considered as fixed over time, in this case the cross-sectional specific effects are considered as random variables (Hsiao 1986). The υ_i are assumed to be independent of the u_{it} as well as mutually independent (Maddala 2001). The main attribute of the RE is that the cross-sectional differences are retained similarly to the pooled OLS and contrary to the FE coefficients (Mairesse 1990, Rodríguez-Pose and Tselios 2009a).

The selection of the most appropriate estimation is based on how the time invariant and individual-specific effect υ_i is treated (Rodríguez-Pose and Tselios 2009a). In order to select which of the above models is the most appropriate for our panel data, two tests are suggested by the literature (Johnston and DiNardo 1997): the rejection of the Breusch and Pagan (1980) Lagrangian multiplier test leads in rejecting the validity of the pooled OLS and the adoption of the FE model as the appropriate one; the rejection of the Hausman (1978) test leads in choosing the FE; respectively, failure in rejecting the Hausman test, enables RE to be used as an alternative to the FE model.

Additionally, effort is spent in this chapter to further investigate the direction of causality between infrastructure provision and the economic development level. The first step is to use the above specifications for the panel data regression model, but interchange the dependent with the main independent variable. So, the *reverse* models can be represented as follows:

$$central\_\ln_{i,t} = \beta_1 gdp\_pc_{i,t-1} + \beta_2 x_{i,t} + u_{i,t} \tag{7.3}$$

Such models can be the first step for investigating the impact of the economic development level in attracting Internet infrastructure. If using the same specifications, GDP per capita proves to be a significant predictor of the allocation of the different centrality measures across the European cities, this could be a first indicator for the existence of a non uni-direct causal relationship between Internet infrastructure and the economic development level.

However, in order to further investigate the direction of causality between these two variables there is a need for the use of a specialised econometric method. If the causality exists simultaneously in both directions (i.e. Internet infrastructure both generates and is attracted by GDP per capita at the same time) then OLS estimation will produce biased and inconsistent estimates because of the endogenous relationship of the two variables. In order to overcome this limitation Granger causality is suggested as the most widely used method (Erdil and Yetkiner 2009). The Granger causality test (Granger 1969) was initially introduced for time series. However, recent developments in panel data analysis enable the use of such a test with panel data (Hoffmann at al 2005). Hood III et al (2008) highlight three reasons why the Granger causality test works better with panel data: (a) panel data provides more flexibility in modelling the cross-section units than time-series analysis separately for each cross-section; (b) panel data allows more observations to be included in the analysis and consequently more degrees of freedom than time series data; (c) finally and also because of the above, the Granger test is more efficient with panel than with time series data (Hurlin and Venet 2003). Indeed, apart from the usual gains because of panel data usage, such as the ability to control omitted variable bias, the Granger causality test with panel data enables the researcher to take account of heterogeneity of the cross section units and even of their subgroups (Shiu and Lam 2008).

There are two main strands of methods for the Granger causality test in the relevant literature (Erdil and Yetkiner 2009). The first one is based on estimation and testing of vector autoregressive coefficients (VAR) in panel data. Autoregressive coefficients and regression slope coefficients are considered and included as variables in the model (Holtz-Eakin et al. 1985, Hsiao 1986). However, this approach does not take heterogeneity into consideration as the variation of causality among the cross-section units is not addressed (Hood III et al 2008). And this is the significance of the second strand of methods as it is mostly represented by the work of Hurlin and Venet (2003), which addresses the heterogeneity problem by treating autoregressive coefficients and regression coefficients as constants (Erdil and Yetkiner 2009). Because of the last attribute, the last method is preferred for this study.

This application of the Granger causality test in panel data was first introduced by Hurlin and Venet (2003) and was applied later by Hood III et al (2008)[30]. This method is based on the following model:

$$y_{i,t} = a_i + \sum_{k=1}^{p} \gamma^{(k)} y_{i,t-k} + \sum_{k=0}^{p} b_i^{(k)} x_{i,t-k} + \varepsilon_{i,t} \tag{7.4}$$

For each cross-section unit *i* and for all *t* [1,T] the regressors are lagged values of the dependent variable ($y_{i,t-k}$) and the lagged values of the independent one ($x_{i,t-k}$), both of them subset by the cross-section unit. a_i represents the fixed effects, $e_{i,t}$ the error term, *k* the lags and ρ the time periods (Hood III et al 2008). Following Hurlin and Venet (2003) and in order to maintain enough degrees of freedom, it is assumed that γ^k is constant and identical for all cross-section units and β_i^k is constant for all $k \in [1, p]$. While the former prevents variation in the autoregressive coefficient among cross-section units, the latter prevents variation in the regression coefficients from time period to time period. However, it should also be noted that coefficients are allowed to vary across lag lengths. Based on the above specifications and after testing specific hypotheses with the use of constraints, three possible causal scenarios can result (Hood III et al 2008):

1. A homogenous causal relationship between x and y for all cross-section units.
2. No causal relationship between x and y for any cross-section units.
3. A causal relationship between n (n<N) cross-section units without a constant causality character.

7.3 Panel data regressions

This section starts with the relevant tests for choosing the most appropriate panel data model. Table 7.2 reports the above two tests. For most of the cases both of the tests are significant (i.e. rejection of the null hypothesis) so the FE appears to be the preferred model. However, there are cases where RE appears to be an alternative to the FE model. Additionally, when the regression is bivariate, the Hausman test

---

[30] The .do file from Hood III et al (2008) for applying the Granger causality test in STATA was provided after request from Trey Hood III. Based on this, a modified version was created for the needs of this study.

cannot be calculated because the chi-square appears to be negative. Nonetheless, this inconsistency does not appear in the multivariate models. Based on the results of these tests, it can be said that overall FE appears to be the preferred model. However, in this section apart from the FE, the pooled OLS will also be presented as they reflect different approaches. The results of the RE are illustrated in the Annex of this chapter.

Table 7.2: Breusch and Pagan and Hausman tests

regressor	Test		(1)	(2)	(3)	(4)	(5)	(6)
d_intra_w_ln	Breusch and Pagan	chi2	491.43	197.79	175.17	163.92	113.95	113.9
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-4.02	21.32	19.61	10.45	5.04	4.08
		Prob > chi2	-	0.000	0.0002	0.0151	0.169	0.2529
d_intra_b_ln	Breusch and Pagan	chi2	523.55	208.09	183.13	168.28	114.86	114.85
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-9.31	15.65	14.49	7.5	1.45	0.46
		Prob > chi2	-	0.0004	0.0023	0.0576	0.6929	0.9286
d_all_w_ln	Breusch and Pagan	chi2	498.85	203.9	179.92	167.36	120.42	120.42
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-3.80	20.17	18.83	9.99	4.53	3.58
		Prob > chi2	-	0.000	0.0003	0.0187	0.2095	0.3108
d_all_b_ln	Breusch and Pagan	chi2	524.05	214.25	188.83	175	124.39	124.37
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-10.89	14.97	14.07	6.67	0.32	-0.62
		Prob > chi2	-	0.0006	0.0028	0.0832	0.9557	0.000
btwnss_all_ln	Breusch and Pagan	chi2	552.15	209.62	182.52	171.06	118.33	118.27
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-1.9	15.12	14.42	8.26	4.08	3.33
		Prob > chi2	-	0.0005	0.0024	0.0409	0.2529	0.3429
eigen_all_ln	Breusch and Pagan	chi2	519.59	197.12	175.47	164.85	118.65	118.68
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	-11.42	15.62	14.37	6.52	0.21	-0.77
		Prob > chi2	-	0.000	0.002	0.089	0.976	-

First, the results of the pooled OLS estimation testing the affect of the weighted degree centrality based on all the links (intra- and extra-European) on the GDP per capita are presented in Table 7.3. In order to test the robustness of this analysis, we gradually insert in the model the control variables described above. When all the control variables are used, the model manages to estimate 66% of the variance in the GDP per capita. As was expected, employment in the service sector has the most important impact on the economic development level of those city-regions which are included in our dataset. Just to clarify again, the NUTS3 city-regions included in the analysis here are city-regions which are linked with at least one Internet backbone network. Additionally, the location of a city-region in Europe's pentagon has also a positive impact, contrary to the negative effect of unemployment and coastal location. The important observation though is that the lagged value (lag = 1 year) of the weighted degree centrality of all the links remains a significant predictor of the GDP per capital for all the regressions. It should also noted that its contribution in explaining the GDP per capita appears to be lower in comparison to the other explanatory variables, as can be seen from the beta coefficients. However, the fact that the Internet infrastructure has higher explanatory value in explaining GDP per capita than objective 1 regions – although the latter has a negative effect –is at least an interesting finding. Additionally, almost the exact same results come out when the weighted degree centrality of only the intra-European links is used as the main regressor. The estimation of this model is presented in the Annex of this chapter[31].

The next centrality indicator tested here is the binary degree centrality for all the links. Table 7.4 presents the results of the model. Again the results are similar with Table 7.3. All the control variables have the same effect. However, according to the beta coefficients the binary degree centrality appears to have slightly lower explanatory value especially when it is compared with the impact of the weighted degree centrality presented in the previous table. Again, the results are the same when only the intra-European links are taken into consideration for the calculation of the degree centrality. The latter as well as the robust OLS are presented in the Annex of this chapter.

---

[31] The robust OLS has also been estimated. The results are the same and they can be found in the appendix. The robust OLS refers to this estimation method where special treatment takes place for the outliers and for the cases with high leverage. The existence of the same results can be used an indication of high robustness of the model.

Table 7.3: Pooled OLS on GDP per capita; lagged degree centrality based on all the weighted links

	(1)	(2)	(3)	(4)	(5)	(6)
L1.d_all_w_ln	4,256.30	1,742.37	1,623.36	1,451.92	1,427.05	1,425.95
	(418.680)***	(420.729)***	(461.096)***	(433.069)***	(370.955)***	(374.903)***
emp_gp		91,487.02	106,855.94	88,490.94	76,373.48	76,441.13
		(11,938.260)***	(13,430.082)***	(13,056.464)***	(11,277.851)***	(11,669.145)***
Un			-777.493	-216.47	-242.458	-243.239
			(259.766)***	(265.777)	(227.671)	(230.692)
Obj1				-13,416.85	-6,477.29	-6,472.41
				(2,559.256)***	(2,345.385)***	(2,360.836)***
pentagon					15,140.32	15,126.44
					(1,819.435)***	(1,917.577)***
coast						-41.672
						(1,774.21)
Constant	-3,972.53	-54,330.11	-59,130.37	-45,337.48	-43,170.63	-43,187.02
	(3,537.65)	(7,788.251)***	(8,677.346)***	(8,541.913)***	(7,321.163)***	(7,373.890)***
Observations	274	208	193	193	193	193
R-squared	0.275	0.438	0.468	0.536	0.661	0.661
Significant beta coefficients						
L1.d_all_w_ln	0.525	0.260	0.229	0.205	0.201	0.201
emp_gp		0.482	0.521	0.431	0.372	0.373
un			-0.162			
obj1				-0.301	-0.145	-0.145
pentagon					0.398	0.398
coast						

Standard errors in parentheses,
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.4: Pooled OLS on GDP per capita; lagged degree centrality based on all the binary links

	(1)	(2)	(3)	(4)	(5)	(6)
L1.d_all_b_ln	9,221.61	2,405.21	2,046.43	2,846.23	3,070.42	3,057.38
	(998.466)***	(979.184)**	(1,004.467)**	(935.878)***	(798.037)***	(802.348)***
emp_gp		104,715.32	120,801.19	91,514.14	77,157.34	77,708.58
		(11,711.305)***	(12,971.122)***	(12,956.354)***	(11,171.170)***	(11,486.348)***
Un			-896.4	-231.185	-248.77	-255.604
			(262.145)***	(266.909)	(227.482)	(230.245)
Obj1				-15,212.98	-8,220.30	-8,166.49
				(2,592.435)***	(2,358.407)***	(2,377.504)***
pentagon					15,428.85	15,300.42
					(1,820.359)***	(1,919.272)***
Coast						-381.562
						(1,764.99)
Constant	16,338.10	-54,432.31	-59,169.36	-39,997.26	-36,837.96	-36,999.57
	(1,899.176)***	(8,314.188)***	(9,271.019)***	(9,149.084)***	(7,806.176)***	(7,861.773)***
Observations	274	208	193	193	193	193
R-squared	0.239	0.409	0.445	0.531	0.661	0.661
			Significant beta coefficients			
L1.d_all_b_ln	0.489	0.151	0.128	0.178	0.192	0.191
emp_gp		0.551	0.589	0.446	0.376	0.379
un			-0.187			
obj1				-0.341	-0.184	-0.183
pentagon					0.406	0.402
coast						

Standard errors in parentheses,
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.5: Pooled OLS on GDP per capita; lagged betweenness centrality of all the links

	(1)	(2)	(3)	(4)	(5)	(6)
L1.btwnss_all_ln	730.169	26.74	-24.891	90.126	233.117	230.05
	(161.720)***	(129.890)	(135.299)	(127.995)	(110.815)**	(112.293)**
emp_gp		118,022.62	135,003.37	108,098.82	89,094.82	89,640.31
		(11,064.772)***	(12,331.923)***	(12,535.133)***	(10,960.809)***	(11,365.812)***
un			-926.454	-304.074	-302.515	-308.859
			(264.644)***	(272.421)	(233.031)	(236.059)
obj1				-14,376.36	-7,612.42	-7,561.42
				(2,660.728)***	(2,415.458)***	(2,436.855)***
pentagon					15,810.44	15,687.67
					(1,890.667)***	(2,004.889)***
coast						-343.375
						(1,826.51)
Constant	31,422.91	-60,939.69	-66,809.44	-47,870.19	-40,779.49	-40,982.71
	(1,355.618)***	(8,527.536)***	(9,582.193)***	(9,601.347)***	(8,256.706)***	(8,348.363)***
Observations	274	208	193	193	193	193
R-squared	0.07	0.391	0.433	0.509	0.643	0.643
			Significant beta coefficients			
L1.btwnss_all_ln	0.264				0.103	0.101
emp_gp		0.621	0.658	0.527	0.434	0.437
un			-0.193			
obj1				-0.322	-0.171	-0.169
pentagon					0.416	0.413
coast						

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.6: Pooled OLS on GDP per capita; lagged eigenvector centrality of all the links

	(1)	(2)	(3)	(4)	(5)	(6)
L1.eigen_all_ln	3,709.97	1,729.47	1,507.68	1,276.73	1,112.58	1,108.33
	(330.696)***	(342.994)***	(368.137)***	(349.740)***	(302.791)***	(303.726)***
emp_gp		84,277.91	101,706.95	86,351.87	77,438.45	78,285.44
		(11,930.434)***	(13,434.298)***	(13,017.458)***	(11,298.725)***	(11,566.046)***
un			-688.077	-172.897	-215.501	-226.725
			(260.142)***	(265.686)	(229.561)	(232.191)
obj1				-12,832.72	-6,222.15	-6,143.63
				(2,561.029)***	(2,359.366)***	(2,374.873)**
pentagon					14,744.57	14,533.49
					(1,829.398)***	(1,924.809)***
coast						-637.133
						(1,766.63)
Constant	55,754.43	-23,479.51	-32,987.49	-24,162.19	-25,391.45	-25,666.55
	(2,559.366)***	(10,614.533)**	(11,626.314)***	(11,089.726)**	(9,580.512)***	(9,633.123)***
Observations	274	208	193	193	193	193
R-squared	0.316	0.458	0.479	0.54	0.659	0.659
			Significant beta coefficients			
L1.eigen_all_ln	0.562	0.317	0.268	0.227	0.198	0.197
emp_gp		0.444	0.496	0.421	0.377	0.382
un			-0.144			
obj1				-0.287	-0.139	-0.138
pentagon					0.388	0.382
coast						

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

However, as highlighted in Chapters 4 and 5 the above centrality indicators can be approached as the infrastructural capital since they represent the accumulated bandwidth or the direct backbone links with the rest of the world. It is worth testing the impact of the other two centrality measures which are more linked with the network function of the Internet and were also estimated in Chapter 4: betweenness and eigenvector centrality (Table 7.5 and 7.6). In regards to the betweenness centrality, the effect is again positive but the significance is not stable across the control variables and the impact is small as is reflected on the low beta values. The significance for the eigenvector centrality measure appears to be more stable and its explanatory value, as is represented by the beta values, is still low but higher than for the betweenness centrality. The latter can be justified because of the inclusion of the bandwidth in the eigenvector centrality measure. Again, the impact of the control variables is the same. The results of robust OLS are presented in the Annex of this chapter.

To sum up the above, the two centrality indicators, which are more linked with the infrastructural capital as well as the eigenvector centrality, appear to be rather good and robust predictors of GDP per capita in Europe. However, the betweenness centrality which reflects more the network function does not seem to have the same explanatory value.

Nevertheless, as was illustrated in table 7.2, the FE is the preferred model for estimating our panel dataset. Table 7.7 presents the FE model when the weighted degree centrality for all the links is used as the main predictor of regional GDP per capita. Because of the specification of the FE model, the time invariant variables are dropped here. The unemployment rate appears to be a non significant predictor, but again the most important observation is that the past level of the Internet infrastructure, as reflected in the weighted degree centrality, appears to have a robust significant impact on the economic development level of the city-regions which are favoured by at least one international Internet backbone link. The robustness of the impact of the weighted degree centrality is justified by the fact that the same results occur when the RE models are used, but also when only the intra-European links are included in the analysis. The RE as well as the robust FE and RE can be found in the Annex of this chapter.

The next centrality measure examined here in a FE framework is the binary degree centrality when all the links are included in the analysis (Table 7.8). Interestingly enough, these models do not seem to work properly. Only employment in service sectors is a significant predictor of the GDP per capita. But most importantly, the past Internet infrastructure capacity as measured by the binary degree centrality does not have a significant impact on the economic development level of city-regions. And this seems to be a robust conclusion because the same results occur from analysis of the intra-European links, the robust FE and the RE models both for the intra-European and all the links. All these models are presented in the Annex of this chapter. These results are in accordance with the previous findings of the descriptive analysis presented in Chapters 4 and 5. In these chapters, when the geography of the Internet backbone networks in Europe was the key question, it was stated that in broad terms while the weighted degree centrality seems to reflect more the economic geography of European city-regions, the binary degree centrality is influenced by other factors as well. More specifically, when the capacity is excluded from the analysis the hub and spoke structure of the Internet backbone network and the distinctive role of some cities as gateways for their hinterland emerge. Among other reasons, the importance of physical location is highlighted in order to explain such connectivities. Notable also is the political geography which is reflected on the binary extra-European links, since some post-colonial relations seem

Table 7.7: FE on GDP per capita; lagged degree centrality based on the all the weighted links

	(1)	(2)	(3)
L1.d_all_w_ln	479.605	256.82	295.901
	(80.338)***	(68.803)***	(78.027)***
emp_gp		40,950.43	43,450.92
		(9,378.832)***	(10,068.873)***
Un			-31.647
			(67.983)
obj1			
pentagon			
coast			
Constant	26,164.13	-4,644.05	-6,353.13
	(646.305)***	(6,970.75)	(7,512.55)
Observations	274	208	193
R-squared	0.153	0.25	0.262

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

to affect the connectivity patterns with out of Europe regions. Based on the above, it seems rational that the accumulation of backbone links regardless of their capacity is not a significant regressor of GDP per capita. On the contrary, the distribution of (low capacity) backbone links seems to affect other factors such as network efficiency and reflect political and physical geography, but such themes are out of the scope of this chapter.

Lastly, the two centrality measures which reflect more the Internet network function are presented in Tables 7.9 and 7.10. Again, these models do not appear to work properly: the one-year lag of the betweenness centrality is only significant at 0.1 level while eigenvector centrality is not significant even at this level. Nevertheless, these results should have been expected because as was highlighted in Chapters 4 and 5, betweenness and eigenvector centrality mostly reflect the network function rather than infrastructural accumulation. It can be said that just like the binary degree centrality, past year betweenness and eigenvector centrality are not significant predictors of the economic development level of European city-regions which are favoured by the existence of at least one international Internet backbone link. The results of the robust FE and RE are the same as the one presented here and can be found in the Annex of this chapter.

Table 7.8: FE on GDP per capita; lagged degree centrality based on all the binary links

	(1)	(2)	(3)
L1.d_all_b_ln	101.519	8.408	62.216
	(296.328)	(217.229)	(233.909)
emp_gp		50,583.35	53,017.17
		(9,468.717)***	(10,295.510)***
un			-16.814
			(72.185)
obj1			
pentagon			
coast			
Constant	29,840.90	-9,968.61	-11,571.08
	(447.704)***	(7,166.60)	(7,813.09)
Observations	274	208	193
R-squared	0.001	0.174	0.177

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.9: FE on GDP per capita; lagged betweenness centrality of all the links

	(1)	(2)	(3)
L1.btwnss_all	-37.197	-27.813	-28.646
	(22.578)	(15.162)*	(16.691)*
emp_gp		49,209.45	51,496.52
		(9,382.081)***	(10,213.618)***
un			-16.017
			(70.803)
obj1			
pentagon			
coast			
Constant	29,918.27	-8,972.72	-10,382.62
	(99.052)***	(7,098.11)	(7,739.60)
Observations	274	208	193
R-squared	0.014	0.194	0.196

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.10: FE on GDP per capita; lagged eigenvector centrality of all the links

	(1)	(2)	(3)
L1.eigen_all_ln	122.892	38.709	33.805
	(104.846)	(76.141)	(81.16)
emp_gp		49,308.29	51,892.93
		(9,788.494)***	(10,608.552)***
un			-16.166
			(71.704)
obj1			
pentagon			
coast			
Constant	30,844.61	-8,723.32	-10,394.56
	(733.495)***	(7,561.14)	(8,208.78)
Observations	274	208	193
R-squared	0.007	0.175	0.178

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

7.4 Causality analysis

As described above, the first indication for the existence of a more complicated causal relationship is the existence of a significant model when the dependent variable is interchanged with the independent one. In our case, such a model would have the structure of 7.3. Following the same process as before, Breusch and Pagan and Hausman tests are presented in Table 7.11 for the case of the degree centrality for all the weighted links. As can be seen, the FE is suggested by the Hausman test as the appropriate model.

Table 7.11: Breusch and Pagan and Hausman tests for the reverse models

regressor	Test		(1)	(2)	(3)	(4)	(5)	(6)
d_all_w_ln	Breusch and Pagan	chi2	289.80	148.32	158.11	154.82	154.5	148.87
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	chi2	27.6	12.31	18.72	19.04	19.36	19.02
		Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000

Table 7.12 presents the FE model with the natural logarithm of weighted degree centrality for all the links as the dependent variable. In addition and for comparison reasons, Table 7.13 presents the RE model with the natural logarithm of the binary degree centrality as the dependent variable. In both cases, the lagged value of GDP per capita appears to be a significant predictor of the different centrality measures with a very small coefficient[32].

---

[32] Similar results occurred for the other centrality indicators, but they are not presented here as the focus of this section is the degree centrality for all the weighted links, which is the strongest predictor of the GDP per capita according to the section 7.3.

Table 7.12: FE on the natural logarithm of the degree centrality based on all the weighted links; lagged gdp per capita

	(1)	(2)	(3)
L1.gdp_pc	0.00037	0.00037	0.00028
	(0.000)***	(0.000)***	(0.000)***
emp_gp		44.228	53.694
		(10.152)***	(10.106)***
un			-0.068
			-0.065
obj1			
pentagon			
Coast			
Constant	-2.476	-35.482	-39.89
	(1.676)	(7.226)***	(7.147)***
Observations	279	210	196
R-squared	0.17	0.254	0.285

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7.13: RE on the natural logarithm of the degree centrality based on all the binary links; lagged gdp per capita

	(1)	(2)	(3)	(4)	(5)	(6)
L1.gdp_pc	0.00003	0.00002	0.00001	0.00002	0.00002	0.00002
	(0.000)***	(0.000)**	(0.000)*	(0.000)**	(0.000)***	(0.000)**
emp_gp		3.975	5.042	5.035	4.801	5.009
		(1.296)***	(1.391)***	(1.376)***	(1.372)***	(1.421)***
Un			0.002	-0.004	-0.003	-0.004
			(0.018)	(0.019)	(0.019)	(0.019)
obj1				0.478	0.35	0.349
				(0.338)	(0.346)	(0.348)
pentagon					-0.469	-0.512
					(0.320)	(0.331)
Coast						-0.158
						(0.275)
Constant	0.461	-2.106	-2.864	-3.044	-2.824	-2.895
	(0.195)**	(0.861)**	(0.910)***	(0.911)***	(0.914)***	(0.925)***
Observations	279	210	196	196	196	196

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

However, as was stated in section 7.2, the existence of such *reverse* relationships does not provide enough evidence for concluding about the actual direction of causality between the Internet infrastructure – as it is reflected in the

Internet backbone centrality measures – and GDP per capita. In order to examine the existence of a causal relation between the two variables, but also the direction of this relationship, the Granger causality test for panel data is used and presented below. For this analysis, the centrality measure used is the degree centrality when all the weighted links are taken into consideration, which according to the above panel data regressions is the best linked with the economic development level centrality measure.

Following the method introduced by Hurlin and Venet (2003) and the application of this method by Hood III et al (2008), the first step is to examine whether both time series (weighted degree centrality and GDP per capita) are *stationary*. A time series appears to be stationary if its expected value and its population variance are independent of time (Douhgerty 2002). The stationarity condition is crucial for the consistency of the OLS coefficients and therefore before moving on to the Granger causality test it is worth testing this. In order to do so, two different tests are utilised (Hood III et al 2008): Levin, Lin and Chu (Levin et al 2002) and Im, Pesaran and Shin (Im et al 2003). Table 7.14 presents the results of these tests. Both of the tests are significant which confirms that weighted degree centrality and GDP per capita in constant prices are stationary.

Table 7.14: Stationary tests

	degree	gpd pc
Ipshin	-3.774***	-4.295***
Levinlin	-58.110***	-19.610***

*** sig. at 1%

However, it should be noted here that the number of cross-section units included in the causality analysis is smaller than the initial panel data used for the panel regression models in section 7.3. There was a need for removing these cross section units, the variation of which over time was low. More specifically, because the Granger causality model includes fixed effects for the cross-section units (a_i in 7.4) and consequently the model is focused on the variation across time instead of the variation across the cross-section units and the total time periods are only six, there was a need to 'clean' the panel data from these cross-section units which had not enough observations. For example, some of the cities in the initial panel data were missing values for two or more of the six time periods because they were not connected with any backbone network. However, this created a problem in running the model which will be described below. In order to avoid this, cities with more than two missing values for the weighted degree centrality were removed. This resulted in a balanced panel data with the degree centrality and the GDP per capita in constant prices for 48 city-regions over the six year time period.

As indicated in section 7.3, the next step is to investigate the three scenarios by testing the relevant hypotheses. Figure 7.1 below illustrates this process. The first scenario refers to the existence of a homogenous causal relationship between *x* and *y* for all the cross-section units. In order to research the first scenario the following hypothesis is first tested:

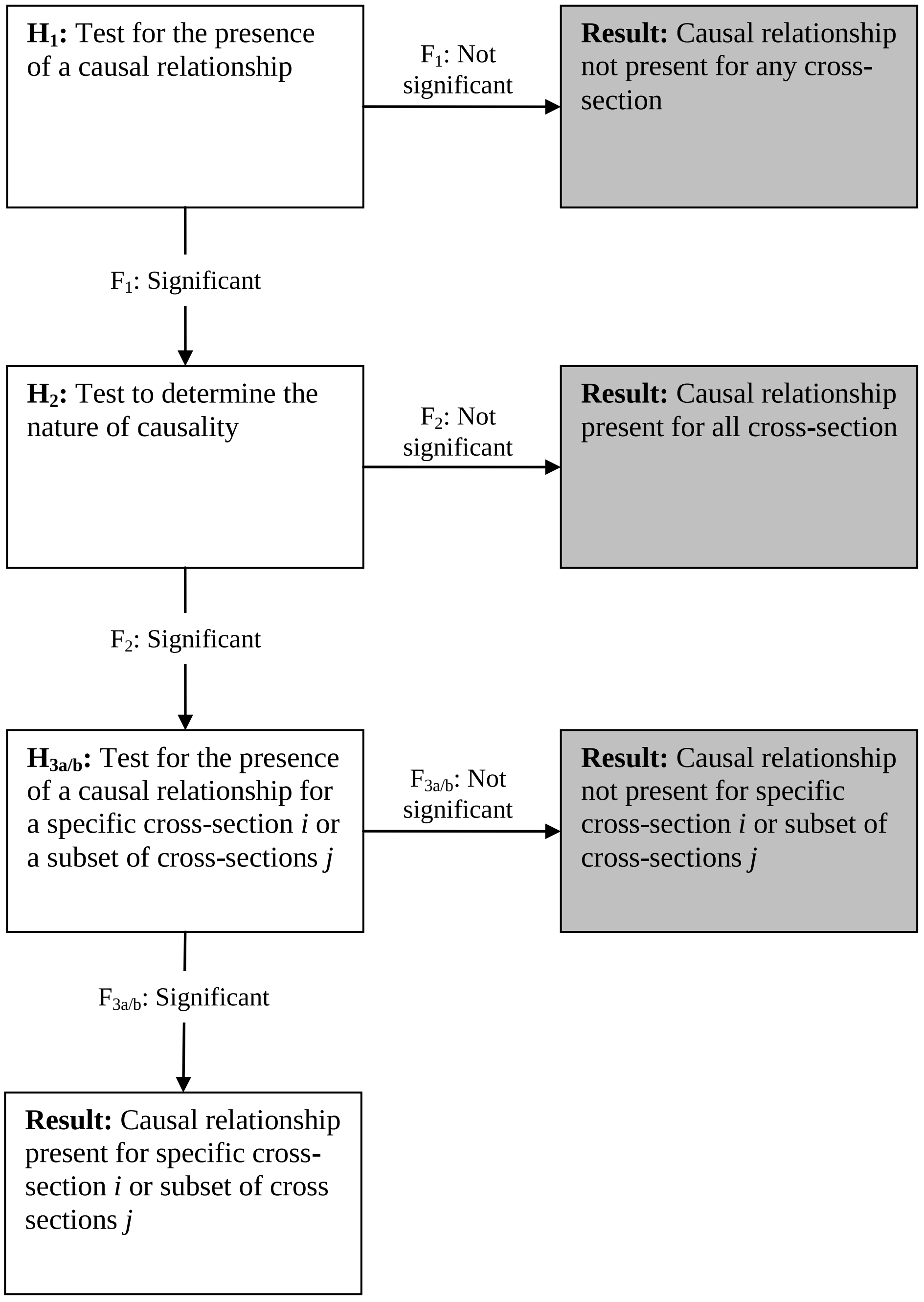
H_1: For all *i*, *x* does not cause *y* (Hood III et al 2008).

In order to test this hypothesis the following statistic test is calculated:

$$F_1 = \frac{(RSS_2 - RSS_1)/Np}{RSS_1/[NT - N(1+p) - p]}, \text{(Hurlin and Venet 2003)} \qquad (7.5)$$

This test aims to examine whether the inclusion of the lagged independent variable increases or not the value of the model 7.4 in predicting the dependent variable. In order to do so, the model described in 7.4 with *N* the number of the cross-section units, *p* the number of lags and *T* number of the time periods is run twice; the first time no restrictions are induced (unrestricted) while for the second time some restrictions are introduced. Then the sum of the squared residuals for the unrestricted (RSS_1) and the restricted model are calculated (RSS_2). The restriction introduced for the restricted model refers to the nullity of the regression coefficients for all the lags. This restriction leads the prediction of the dependent variable to be dependent only on the fixed effects and on the lagged version of the dependent variable (Hood III et al 2008). After calculating the RSS_1 and RSS_2 the F_1 test is calculated using 7.5. The significance of the test is calculated using the *F* distribution with Np and $NT - N(1+p) - p$ degrees of freedom for the nominator and the denominator respectively.

Figure 7.1: Granger causality test for panel data



Source: Hood III et al 2008, 311

This hypothesis is tested for both directions, i.e. the lagged weighted degree centrality impacts on GDP per capita and vice versa for one and two year lags. The results of these tests are shown in Table 7.15. Based on these results, the first hypothesis (H_1) can be rejected only when the one year lag of the Internet weighted degree centrality is used as the explanatory variable for GDP per capita. In simpler terms, this means that for this case the inclusion of the independent variable (weighted degree centrality) increases the explanatory value of the model in predicting the dependent variable (GDP per capita). This is the only case that a causal relation exists. For all the other cases, the H_1 cannot be rejected (i.e. not a significant F test) so no causal relation exists as the inclusion of the independent variable (two year lagged centrality, one and two year lagged GDP per capita) does not increase the explanatory value in predicting the dependent variable (GDP per capita, weighted degree centrality respectively).

Table 7.15: F_1 tests

	centrality → GDP pc	**GDP pc → centrality**
t-1	3.561***	0.458
t-2	0.902	0.213

***p < 0.01

Following figure 7.1 the next step is to examine the nature of the one causal relationship which resulted from the implementation of the F_1. This test proved that there is a causal relationship from the one year lagged Internet degree centrality to the GDP per capita. However, F_1 cannot conclude whether this causal relationship exists for one or for all the cross-section units. In order to examine this, a second hypothesis is tested:

H_2: *x* causes *y* for all *i* (Hood III et al 2008).

This hypothesis is tested using the following F test:

$$F_2 = \frac{(RSS_3 - RSS_1)/p(N-1)}{RSS_1/[NT - N(1+p) - p]}, \text{ (Hurlin and Venet 2003)} \qquad (7.6)$$

Again RSS_1 refers to the sum of the square residuals of the unrestricted model while RSS_3 is the sum of the square residuals of the new restricted model based on 7.4. The restriction here is that the regression coefficients are equal for each cross-section unit ($\beta_{t-1} = \beta_{t-k}$) (Hood III et all 2008). This restriction will enable us to examine the homogeneity of the causal relationship. The F_2 test for the causal relationship from the one year lagged weighted degree centrality to GDP per capita is 2.540, which is significant at p<0.01. The latter enables the rejection of the H_2

hypothesis which means that the above causal relation is not homogenous across the cross-section units, and it only exists for a subset of the 48 city-regions.

In order to investigate for which city-regions this causal relationship is true, a third hypothesis is tested:

H$_3$: For *i, x* does not cause *y* (Hood III et al 2008)

Just as before, in order to investigate this hypothesis, a test is calculated for each one of the 48 city-regions included in the analysis.

$$F_3 = \frac{(RSS_2, i - RSS_1)/p(N-1)}{RSS_1/[NT - N(1+p) - p]}, \text{ (Hurlin and Venet 2003)} \qquad (7.7)$$

For the restricted model the nullity of the coefficient of the lagged explanatory variable for each cross-section unit is imposed (i.e. $\beta_i^k = 0$). In order to calculate F_3, model 7.4 is calculated N=48 times separately for each cross-section unit. Then the significance of these 48 F_3 tests is examined and according to this it can be concluded for which cross-section units the Internet weighted degree centrality affects GDP per capita. Table 7.16 presents the results of the F_3 test for all the city-regions.

The first conclusion that can be drawn from this final test is that for 28 out of the 48 city-regions included in the analysis, there is a causal relation running from the weighted degree centrality of all the backbone links to GDP per capita. This finding is important as it is proves that for most of the city-regions included in the analysis, the Internet infrastructure provision, as reflected in the weighted degree centrality of all the backbone links, impacts on the regional GDP per capita rather than the GDP per capita being a pull factor for this infrastructure allocation.

Also interesting is the geographical representation of the results, as they are presented in Figure 7.2. Maybe the pattern is not clear-cut, but still there is a visible higher concentration of city-regions with a significant causal relationship from the Internet infrastructure to the economic development level in the northern part of Europe. Conversely, for most of the city-regions located in the southern part of Europe (i.e. Iberian peninsula and the Mediterranean arc with the exception of Athens), but also for some of the Eastern and Central European city-regions, there is no statistically significant causal relation between Internet infrastructure and the economic development level. From the above it can be said that there is a north-south divide for the role of the Internet infrastructure as a significant – causal – predictor of the regional economic development.

Table 7.16: F_3 for all the city-regions

City-region	F_3	Causality	City-region	F_3	Causality
London	58.236***	degree ➔ GDP pc	Tallinn	5.278***	degree ➔ GDP pc
Paris	3.822***	degree ➔ GDP pc	Bucharest	1.092	no
Frankfurt	2.73***	degree ➔ GDP pc	Ljubljana	3.458***	degree ➔ GDP pc
Amsterdam	1.274	no	Vilnius	2.366***	degree ➔ GDP pc
Stockholm	8.917***	degree ➔ GDP pc	Riga	2.548***	degree ➔ GDP pc
Madrid	0.91	no	Luxembourg	17.471***	degree ➔ GDP pc
Voesendorf	0.182	no	Stuttgart	0.728	no
Milan	0.182	no	Rotterdam	2.548***	degree ➔ GDP pc
Hamburg	4.186***	degree ➔ GDP pc	Hilden	2.366***	degree ➔ GDP pc
Brussels	1.82***	degree ➔ GDP pc	Rome	0	no
Düsseldorf	6.188***	degree ➔ GDP pc	Berlin	0	no
Warsaw	6.006***	degree ➔ GDP pc	Bielsko-Biala	0.182	no
Bratislava	1.456	degree ➔ GDP pc	Malmö	3.64***	degree ➔ GDP pc
Prague	4.186***	degree ➔ GDP pc	Msida	0.182	no
Helsinki	9.281***	degree ➔ GDP pc	Nuremberg	6.188***	degree ➔ GDP pc
Dublin	14.013***	degree ➔ GDP pc	Gothenburg	1.638	degree ➔ GDP pc
Budapest	4.368***	degree ➔ GDP pc	Graz	2.548***	degree ➔ GDP pc
Munich	1.456**	degree ➔ GDP pc	Nice	1.274	no
Athens	2.548***	degree ➔GDP pc	Turin	0	no
Barcelona	0.728	no	Ehingen	0.91	no
Lisbon	0	no	Maribor	0.91	no
Brno	0.546	no	Cologne	1.638**	degree ➔ GDP pc
Palermo	0.182	no	Portsmouth	0.364	no
Hannover	8.189***	degree ➔ GDP pc	Nicosia	0.182	no

***p < 0.01, **p<0.05

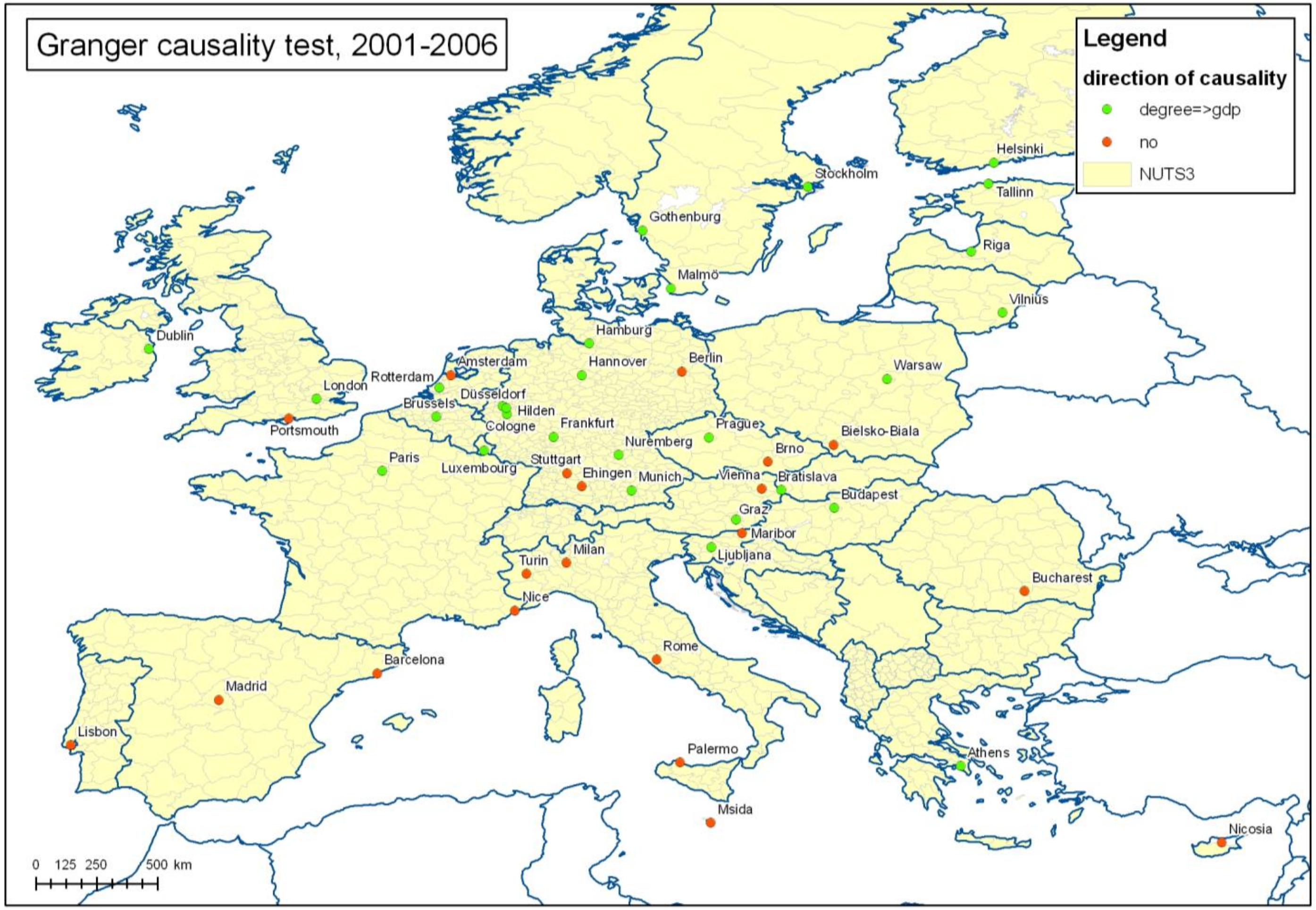


Figure 7.2: Granger causality test, 2001-2006

7.5 Discussion

The main finding of the above analysis is that the Internet infrastructure is a significant predictor for the level of regional economic development. Using panel data analysis we empirically verify that the past level of Internet infrastructural capital, as it is reflected in the accumulation of high capacity backbone networks, is a significant predictor of GDP per capita. But more importantly, the above analysis has shed light on the direction of causality of the relationship between Internet infrastructure and the economic development level: the Granger causality tests verified that in most of the interconnected city-regions in Europe where there is a significant causal relationship, the causality runs from the Internet infrastructure to the economic development level.

The geographic analysis of causality also highlighted something equally important: this causal relationship is not homogenous for all the European interconnected city-regions. On the contrary, an interesting almost north-south divide emerged. Overall, the Internet infrastructure appears to cause economic development mostly in the northern European city-regions, while in the south of Europe no significant causal relationship has been identified. Of course, there are exceptions to this pattern such as the significant causal effect of the Internet infrastructure in Athens and the lack of such a significant relationship in Berlin and in Amsterdam. Although the value of the F-test for the latter was very close to the cut-off value for 90% significance, it is still a surprise and difficult to explain why the analysis did not identify a significant causal relation for Amsterdam – one of the most well-connected cities in Europe. However, apart from this, a general pattern of north-south divide in the existence of a causal relationship emerges.

Despite the fact that to my knowledge no research has taken place about the regional development impacts of Internet infrastructure, the results of the above analysis are in accordance with previous studies dealing with ICTs and (regional) economic development. Indeed, almost half of the studies presented in Table 2.4 resulted in a unidirectional causal relationship running from ICTs to economic development. Additionally, the heterogeneous nature of the causal relationship between Internet infrastructure and regional economic development, in addition to the empirical confirmation that the Internet infrastructure is a significant predictor for regional economic development, concurs with Capello and Nijkamp's empirical results (1996b, 26) that "mere accessibility to advanced telecommunications infrastructure and services does not necessarily lead to a better corporate and regional performance". Certainly, as happens with infrastructure in general (Banister and Berechman 2003; Huddlestone and Pangotra 1990), ICTs infrastructure is a necessary, but not a sufficient condition for economic development (Gillespie and Robins 1989; Graham 1999; Gibbs and Tanner 1997; Hackler 2003). The lack of causal relationship in 20 out of 48 city-regions might have occurred due to the lack of the other necessary but also sufficient factors for economic development. These critical factors can be recognised as the *capacity* of a city or a region to exploit the Internet infrastructure, which can support the development process in the frame of the digital economy (Antonelli 2003). In the emerging stage of the digital economy, Capello and Nijkamp (1996a, 226) identified this phenomenon:

> "The exploitation of advanced computer networks requires organizational, managerial, technical and strategic knowledge, which is not present everywhere, and is not at all a 'public good'. For this reason it would be misleading to think that the impacts of these telecommunications technologies on the performance of firms and regions are similar everywhere".

Lastly, from the geography point of view, it is interesting to analyse how this heterogeneity in the causal relationship between Internet infrastructure and regional economic development is projected on space. If the above argument about the regional capacity in exploiting this infrastructure is valid, then the spatial heterogeneity in the causal relationship should be explained by the spatial differentiation of the regional capacity. Indeed, this north-south divide in the significance of the causal relationship corresponds to the well-established socio-economic north-south divide in Europe, which is related among others with differences in economic development level and prosperity, technology adoption, innovation, and human capital level (e.g. Cutrini 2009; Rodríguez-Pose and Tselios 2009b; Paci and Usai 2000; EC1999). In short, it seems that it is more difficult for the south of Europe to take advantage and use as a development tool the Internet infrastructure contrary to the higher efficiency in exploiting this infrastructure which was observed in the northern part of Europe.

Chapter 7

Annex

Table A7.1: Robust pooled OLS on GDP per capita; lagged degree centrality based on all the weighted links

	1	2	3	4	5	6
L1.d_all_w_ln	4,256.30	1,742.37	1,623.36	1,451.92	1,427.05	1,425.95
	(617.481)***	(353.553)***	(392.721)***	(378.869)***	(326.054)***	(344.838)***
emp_gp		91,487.02	106,855.94	88,490.94	76,373.48	76,441.13
		(10,990.192)***	(13,523.631)***	(12,653.269)***	(11,079.207)***	(10,210.662)***
un			-777.493	-216.47	-242.458	-243.239
			(200.537)***	-216.207	-174.166	-171.901
obj1				-13,416.85	-6,477.29	-6,472.41
				(1,678.123)***	(1,747.013)***	(1,705.141)***
pentagon					15,140.32	15,126.44
					(1,813.196)***	(2,197.669)***
coast						-41.672
						(2,070.80)
Constant	-3,972.53	-54,330.11	-59,130.37	-45,337.48	-43,170.63	-43,187.02
	-4,420.60	(8,208.308)***	(9,702.325)***	(9,130.891)***	(8,326.078)***	(7,983.608)***
Observations	274	208	193	193	193	193
R-squared	0.275	0.438	0.468	0.536	0.661	0.661

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.2: Robust pooled OLS on GDP per capita; lagged degree centrality based on all the binary links

	1	2	3	4	5	6
L1.d_all_b_ln	9,221.61	2,405.21	2,046.43	2,846.23	3,070.42	3,057.38
	(1,624.686)***	(1,092.254)**	(1,094.424)*	(1,069.371)***	(959.720)***	(994.726)***
emp_gp		104,715.32	120,801.19	91,514.14	77,157.34	77,708.58
		(10,269.471)***	(12,728.278)***	(11,121.669)***	(9,790.463)***	(9,448.779)***
un			-896.4	-231.185	-248.77	-255.604
			(209.606)***	(221.771)	(179.431)	(178.239)
obj1				-15,212.98	-8,220.30	-8,166.49
				(1,722.739)***	(1,727.422)***	(1,728.812)***
pentagon					15,428.85	15,300.42
					(1,835.755)***	(2,206.725)***
coast						-381.562
						(2,063.82)
Constant	16,338.10	-54,432.31	-59,169.36	-39,997.26	-36,837.96	-36,999.57
	(1,979.174)***	(7,794.284)***	(9,609.316)***	(8,184.270)***	(7,330.728)***	(7,088.788)***
Observations	274	208	193	193	193	193
R-squared	0.239	0.409	0.445	0.531	0.661	0.661

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.3: Pooled OLS on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3	4	5	6
L1.d_intra_w_ln	4,371.09	1,858.70	1,696.60	1,479.39	1,504.22	1,517.75
	(425.895)***	(415.542)***	(462.687)***	(435.820)***	(368.843)***	(375.520)***
emp_gp		90,636.51	106,081.59	88,711.58	75,493.59	74,838.37
		(11,814.976)***	(13,442.265)***	(13,044.502)***	(11,145.459)***	(11,612.387)***
un			-681.408	-133.983	-128.834	-120.98
			(266.560)**	(271.502)	(229.771)	(233.461)
obj1				(13,341.61)	(6,237.61)	(6,274.68)
				(2,586.113)***	(2,338.634)***	(2,351.537)***
pentagon					15,726.47	15,851.56
					(1,824.427)***	(1,926.021)***
coast						369.932
						(1,783.34)
Constant	-4,602.03	-54,385.75	-59,719.10	-46,212.80	-44,073.20	-43,925.18
	-3,565.25	(7,768.700)***	(8,684.310)***	(8,552.343)***	(7,242.042)***	(7,295.929)***
Observations	269	205	190	190	190	190
R-squared	0.283	0.447	0.471	0.538	0.671	0.671

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.4: Robust pooled OLS on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3	4	5	6
L1.d_intra_w_ln	4,371.09	1,858.70	1,696.60	1,479.39	1,504.22	1,517.75
	(598.720)***	(348.786)***	(384.966)***	(367.455)***	(313.813)***	(341.479)***
emp_gp		90,636.51	106,081.59	88,711.58	75,493.59	74,838.37
		(11,261.034)***	(13,833.541)***	(12,998.877)***	(11,348.124)***	(10,355.035)***
un			-681.408	-133.983	-128.834	-120.98
			(198.749)***	(221.183)	(173.068)	(171.581)
obj1				-13,341.61	-6,237.61	-6,274.68
				(1,708.407)***	(1,765.160)***	(1,715.762)***
pentagon					15,726.47	15,851.56
					(1,833.152)***	(2,245.112)***
coast						369.932
						(2,113.60)
Constant	-4,602.03	-54,385.75	-59,719.10	-46,212.80	-44,073.20	-43,925.18
	-4,181.75	(8,309.741)***	(9,811.870)***	(9,274.952)***	(8,455.315)***	(8,080.207)***
Observations	269	205	190	190	190	190
R-squared	0.283	0.447	0.471	0.538	0.671	0.671

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.5: Pooled OLS on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3	4	5	6
L1.d_intra_b_ln	10,053.88	2,495.13	1,924.01	2,805.15	3,213.55	3,228.06
	(1,193.747)***	(1,129.475)**	(1,176.61)	(1,097.554)**	(929.038)***	(945.281)***
emp_gp		106,373.07	123,306.69	94,515.35	78,934.22	78,647.98
		(11,813.014)***	(13,166.768)***	(13,153.430)***	(11,264.362)***	(11,730.171)***
un			-848.458	-161.586	-138.598	-135.086
			(269.287)***	(275.964)	(233.307)	(237.142)
obj1				-15,129.97	-8,056.87	-8,080.93
				(2,633.164)***	(2,371.344)***	(2,392.610)***
pentagon					15,993.20	16,049.57
					(1,848.355)***	(1,955.410)***
coast						163.043
						(1,803.35)
Constant	16,803.33	-55,387.00	-60,970.14	-42,298.24	-38,939.73	-38,848.64
	(1,989.973)***	(8,380.979)***	(9,313.515)***	(9,195.833)***	(7,783.553)***	(7,869.382)***
Observations	269	205	190	190	190	190
R-squared	0.21	0.406	0.441	0.526	0.663	0.663

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.6: Robust pooled OLS on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3	4	5	6
L1.d_intra_b_ln	10,053.88	2,495.13	1,924.01	2,805.15	3,213.55	3,228.06
	(1,753.335)***	(1,082.689)**	(1,110.214)*	(1,061.825)***	(943.123)***	(1,027.276)***
emp_gp		106,373.07	123,306.69	94,515.35	78,934.22	78,647.98
		(11,143.276)***	(13,716.114)***	(12,303.155)***	(10,738.633)***	(9,928.093)***
un			-848.458	-161.586	-138.598	-135.086
			(205.000)***	(225.626)	(178.711)	(178.266)
obj1				-15,129.97	-8,056.87	-8,080.93
				(1,749.517)***	(1,747.195)***	(1,716.894)***
pentagon					15,993.20	16,049.57
					(1,874.490)***	(2,296.048)***
coast						163.043
						(2,161.25)
Constant	16,803.33	-55,387.00	-60,970.14	-42,298.24	-38,939.73	-38,848.64
	(1,818.282)***	(8,355.121)***	(10,266.572)***	(8,955.499)***	(8,001.344)***	(7,562.007)***
Observations	269	205	190	190	190	190
R-squared	0.21	0.406	0.441	0.526	0.663	0.663

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.7: Robust pooled OLS on GDP per capita; lagged betweenness centrality of all the links

	1	2	3	4	5	6
L1.btwnss_all_ln	730.169	26.74	-24.891	90.126	233.117	230.05
	(184.633)***	(118.785)	(121.954)	(117.66)	(97.701)**	(103.557)**
emp_gp		118,022.62	135,003.37	108,098.82	89,094.82	89,640.31
		(12,718.478)***	(15,401.886)***	(14,207.827)***	(11,958.854)***	(11,063.384)***
un			-926.454	-304.074	-302.515	-308.859
			(205.249)***	(231.77)	(184.847)	(181.470)*
obj1				-14,376.36	-7,612.42	-7,561.42
				(1,716.478)***	(1,768.196)***	(1,753.080)***
pentagon					15,810.44	15,687.67
					(1,882.725)***	(2,275.761)***
coast						-343.375
						(2,113.43)
Constant	31,422.91	-60,939.69	-66,809.44	-47,870.19	-40,779.49	-40,982.71
	(1,563.936)***	(9,398.522)***	(11,412.634)***	(10,214.876)***	(8,824.845)***	(8,436.152)***
Observations	274	208	193	193	193	193
R-squared	0.07	0.391	0.433	0.509	0.643	0.643

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.8: Robust pooled OLS on GDP per capita; lagged eigenvector centrality of all the links

	1	2	3	4	5	6
L1.eigen_all_ln	3,709.97	1,729.47	1,507.68	1,276.73	1,112.58	1,108.33
	(465.725)***	(282.509)***	(295.012)***	(296.666)***	(266.530)***	(271.042)***
emp_gp		84,277.91	101,706.95	86,351.87	77,438.45	78,285.44
		(10,193.708)***	(12,595.302)***	(11,950.292)***	(10,802.017)***	(10,185.558)***
un			-688.077	-172.897	-215.501	-226.725
			(196.192)***	(218.813)	(176.933)	(175.957)
obj1				-12,832.72	-6,222.15	-6,143.63
				(1,747.978)***	(1,765.850)***	(1,731.353)***
pentagon					14,744.57	14,533.49
					(1,798.276)***	(2,166.151)***
coast						-637.133
						(2,026.03)
Constant	55,754.43	-23,479.51	-32,987.49	-24,162.19	-25,391.45	-25,666.55
	(4,016.016)***	(7,964.408)***	(9,582.788)***	(8,762.855)***	(8,123.549)***	(7,939.180)***
Observations	274	208	193	193	193	193
R-squared	0.316	0.458	0.479	0.54	0.659	0.659

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.9: Robust FE on GDP per capita; lagged degree centrality based on the all the weighted links

	1	2	3
L1.d_all_w_ln	479.605	256.82	295.901
	(126.333)***	(87.438)***	(101.394)***
emp_gp		40,950.43	43,450.92
		(11,928.550)***	(12,990.748)***
un			-31.647
			(85.773)
obj1			
pentagon			
coast			
Constant	26,164.13	-4,644.05	-6,353.13
	(1,008.095)***	(9,017.26)	(9,980.90)
Observations	274	208	193
R-squared	0.153	0.25	0.262

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.10: RE on GDP per capita; lagged degree centrality based on the all the weighted links

	1	2	3	4	5	6
L1.d_all_w_ln	520.196	247.244	291.929	291.352	299.596	300.907
	(82.994)***	(71.305)***	(81.264)***	(79.506)***	(79.219)***	(79.108)***
emp_gp		58,109.50	61,674.34	58,163.00	55,881.92	54,801.94
		(8,576.846)***	(9,209.647)***	(9,020.525)***	(8,845.512)***	(8,918.526)***
un			-65.611	-47.055	-48.993	-47.43
			(70.555)	(69.226)	(68.936)	(68.844)
obj1				-17,504.83	-9,431.14	-9,686.97
				(4,646.905)***	(4,633.098)**	(4,667.392)**
pentagon					16,404.27	16,977.54
					(4,024.401)***	(4,101.322)***
coast						3,184.46
						(3,783.50)
Constant	23,449.48	-17,349.72	-19,746.23	-13,238.34	-19,755.15	-20,261.12
	(2,204.842)***	(6,544.742)***	(6,995.752)***	(7,041.360)*	(6,900.867)***	(6,944.673)***
Observations	274	208	193	193	193	193

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.11: Robust RE on GDP per capita; lagged degree centrality based on the all the weighted links

	1	2	3	4	5	6
L1.d_all_w_ln	520.196	247.244	291.929	291.352	299.596	300.907
	(104.172)***	(68.563)***	(78.450)***	(77.850)***	(79.791)***	(79.933)***
emp_gp		58,109.50	61,674.34	58,163.00	55,881.92	54,801.94
		(9,550.946)***	(10,364.653)***	(9,976.928)***	(9,499.511)***	(8,791.935)***
un			-65.611	-47.055	-48.993	-47.43
			(68.103)	(66.269)	(63.258)	(63.481)
obj1				-17,504.83	-9,431.14	-9,686.97
				(3,432.792)***	(3,307.356)***	(3,384.038)***
pentagon					16,404.27	16,977.54
					(4,933.267)***	(5,376.252)***
coast						3,184.46
						(5,419.40)
Constant	23,449.48	-17,349.72	-19,746.23	-13,238.34	-19,755.15	-20,261.12
	(1,880.626)***	(6,189.393)***	(6,585.655)***	(6,580.391)**	(6,503.129)***	(6,889.329)***
Observations	274	208	193	193	193	193

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.12: Robust FE on GDP per capita; lagged degree centrality based on the all the binary links

	1	2	3
L1.d_all_b_ln	101.519	8.408	62.216
	(203.627)	(244.087)	(266)
emp_gp		50,583.35	53,017.17
		(13,795.621)***	(14,921.961)***
un			-16.814
			(87.864)
obj1			
pentagon			
coast			
Constant	29,840.90	-9,968.61	-11,571.08
	(301.482)***	(10,498.78)	(11,536.09)
Observations	274	208	193
R-squared	0.001	0.174	0.177

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.13: RE on GDP per capita; lagged degree centrality based on the all the binary links

	1	2	3	4	5	6
L1.d_all_b_ln	326.82	101.434	184.932	180.76	196.339	190.816
	(304.092)	(223.385)	(241.635)	(237.865)	(237.27)	(236.963)
emp_gp		66,581.82	70,827.54	67,330.46	64,841.46	63,842.74
		(8,555.925)***	(9,271.595)***	(9,107.550)***	(8,934.510)***	(9,017.398)***
un			-57.879	-38.044	-40.199	-38.36
			-74.354	(73.426)	(73.242)	(73.162)
obj1				-17,389.59	-9,573.76	-9,802.27
				(4,612.617)***	(4,604.220)**	(4,643.925)**
pentagon					15,935.18	16,437.16
					(3,998.893)***	(4,080.664)***
coast						2,760.35
						(3,764.82)
Constant	26,870.32	-22,011.55	-24,773.37	-18,312.20	-24,392.74	-24,774.23
	(2,206.879)***	(6,635.047)***	(7,144.068)***	(7,202.507)**	(7,048.466)***	(7,090.023)***
Observations	274	208	193	193	193	193

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.14: Robust RE on GDP per capita; lagged degree centrality based on the all the binary links

	1	2	3	4	5	6
L1.d_all_b_ln	326.82	101.434	184.932	180.76	196.339	190.816
	(252.616)	(215.153)	(220.302)	(218.63)	(224.999)	(223.619)
emp_gp		66,581.82	70,827.54	67,330.46	64,841.46	63,842.74
		(10,104.682)***	(11,116.087)***	(10,679.287)***	(10,026.170)***	(9,314.356)***
un			-57.879	-38.044	-40.199	-38.36
			-69.85	(68.22)	(66.008)	(65.979)
obj1				-17,389.59	-9,573.76	-9,802.27
				(3,615.018)***	(3,461.697)***	(3,530.122)***
pentagon					15,935.18	16,437.16
					(4,843.561)***	(5,297.219)***
coast						2,760.35
						(5,392.81)
Constant	26,870.32	-22,011.55	-24,773.37	-18,312.20	-24,392.74	-24,774.23
	(1,988.158)***	(6,663.220)***	(7,279.945)***	(7,128.522)**	(6,969.525)***	(7,352.860)***
Observations	274	208	193	193	193	193

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.15: FE on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3
L1.d_intra_w_ln	519.821	239.156	271.7
	(83.171)***	(70.148)***	(79.802)***
emp_gp		40,339.92	42,952.45
		(9,368.231)***	(10,083.657)***
un			-31.555
			(68.466)
obj1			
pentagon			
coast			
Constant	25,826.24	-4,038.94	-5,769.74
	(662.279)***	(6,963.80)	(7,522.13)
Observations	269	205	190
R-squared	0.169	0.235	0.244

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.16: Robust FE on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3
L1.d_intra_w_ln	519.821	239.156	271.7
	(125.831)***	(91.555)**	(109.011)**
emp_gp		40,339.92	42,952.45
		(12,141.657)***	(13,146.894)***
un			-31.555
			(88.749)
obj1			
pentagon			
coast			
Constant	25,826.24	-4,038.94	-5,769.74
	(994.170)***	(9,155.85)	(10,111.81)
Observations	269	205	190
R-squared	0.169	0.235	0.244

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.17: RE on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3	4	5	6
L1.d_intra_w_ln	562.789	231.473	268.341	267.726	276.854	278.417
	(85.891)***	(72.918)***	(83.243)***	(81.409)***	(81.187)***	(81.087)***
emp_gp		57,769.23	61,417.24	57,916.44	55,687.65	54,579.67
		(8,585.483)***	(9,234.109)***	(9,041.392)***	(8,867.403)***	(8,941.683)***
un			-64.814	-46.108	-47.598	-46.032
			(71.172)	(69.803)	(69.57)	(69.484)
obj1				-17,502.77	-9,396.81	-9,658.15
				(4,656.269)***	(4,633.494)**	(4,665.014)**
pentagon					16,459.06	17,047.23
					(4,025.094)***	(4,099.812)***
coast						3,266.21
						(3,782.05)
Constant	23,183.24	-16,967.43	-19,371.75	-12,873.03	-19,466.17	-19,986.46
	(2,206.608)***	(6,553.943)***	(7,014.312)***	(7,057.331)*	(6,914.185)***	(6,955.867)***
Observations	269	205	190	190	190	190

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.18: Robust RE on GDP per capita; lagged degree centrality based on the intra-European weighted links

	1	2	3	4	5	6
L1.d_intra_w_ln	562.789	231.473	268.341	267.726	276.854	278.417
	(105.953)***	(69.848)***	(81.566)***	(80.781)***	(81.989)***	(82.441)***
emp_gp		57,769.23	61,417.24	57,916.44	55,687.65	54,579.67
		(9,642.479)***	(10,469.701)***	(10,072.561)***	(9,558.414)***	(8,843.007)***
un			-64.814	-46.108	-47.598	-46.032
			(69.928)	(68.219)	(65.265)	(65.553)
obj1				-17,502.77	-9,396.81	-9,658.15
				(3,431.795)***	(3,308.924)***	(3,386.643)***
pentagon					16,459.06	17,047.23
					(4,921.595)***	(5,379.055)***
coast						3,266.21
						(5,455.25)
Constant	23,183.24	-16,967.43	-19,371.75	-12,873.03	-19,466.17	-19,986.46
	(1,833.762)***	(6,231.848)***	(6,651.487)***	(6,631.652)*	(6,554.736)***	(6,938.715)***
Observations	269	205	190	190	190	190

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.19: FE on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3
L1.d_intra_b_ln	213.306	77.345	116.86
	(320.083)	(222.571)	(236.761)
emp_gp		48,524.92	51,058.63
		(9,450.563)***	(10,238.071)***
un			-34.1
			(71.934)
obj1			
pentagon			
coast			
Constant	29,654.70	-8,512.80	-10,004.90
	(427.670)***	(7,129.27)	(7,752.71)
Observations	269	205	190
R-squared	0.002	0.169	0.173

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.20: Robust FE on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3
L1.d_intra_b_ln	213.306	77.345	116.86
	(300.396)	(300.464)	(321.192)
emp_gp		48,524.92	51,058.63
		(13,621.904)***	(14,793.413)***
un			-34.1
			(88.303)
obj1			
pentagon			
coast			
Constant	29,654.70	-8,512.80	-10,004.90
	(392.305)***	-10,321.52	(11,405.14)
Observations	269	205	190
R-squared	0.002	0.169	0.173

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.21: RE on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3	4	5	6
L1.d_intra_b_ln	423.722	127.896	188.602	184.342	203.315	202.056
	(327.429)	(229.005)	(244.778)	(240.661)	(240.059)	(239.754)
emp_gp		64,514.67	68,612.79	65,202.59	62,816.52	61,811.97
		(8,561.477)***	(9,254.640)***	(9,086.537)***	(8,918.432)***	(8,997.746)***
un			-70.368	-51.043	-52.976	-51.421
			(73.958)	(72.919)	(72.697)	(72.629)
obj1				-17,444.19	-9,547.38	-9,785.17
				(4,674.083)***	(4,669.884)**	(4,705.838)**
pentagon					16,080.39	16,608.02
					(4,055.752)***	(4,134.602)***
coast						2,923.59
						(3,814.29)
Constant	26,811.28	-20,500.43	-23,009.46	-16,596.61	-22,831.26	-23,277.15
	(2,225.335)***	(6,631.290)***	(7,131.432)***	(7,188.088)**	(7,046.644)***	(7,090.396)***
Observations	269	205	190	190	190	190

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.22: Robust RE on GDP per capita; lagged degree centrality based on the intra-European binary links

	1	2	3	4	5	6
L1.d_intra_b_ln	423.722	127.896	188.602	184.342	203.315	202.056
	(278.638)	(230.643)	(234.583)	(233.919)	(240.254)	(242.703)
emp_gp		64,514.67	68,612.79	65,202.59	62,816.52	61,811.97
		(10,021.977)***	(11,029.752)***	(10,607.267)***	(9,974.254)***	(9,298.485)***
un			-70.368	-51.043	-52.976	-51.421
			(69.033)	(67.479)	(65.199)	(65.336)
obj1				-17,444.19	-9,547.38	-9,785.17
				(3,663.239)***	(3,521.266)***	(3,595.211)***
pentagon					16,080.39	16,608.02
					(4,898.546)***	(5,351.859)***
coast						2,923.59
						(5,477.33)
Constant	26,811.28	-20,500.43	-23,009.46	-16,596.61	-22,831.26	-23,277.15
	(2,016.046)***	(6,588.072)***	(7,211.468)***	(7,091.780)**	(6,913.443)***	(7,308.431)***
Observations	269	205	190	190	190	190

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.23: Robust FE on GDP per capita; lagged betweenness centrality of all the links

	1	2	3
L1.btwnss_all_ln	-37.197	-27.813	-28.646
	(20.112)*	(11.433)**	(12.305)**
emp_gp		49,209.45	51,496.52
		(13,624.381)***	(14,750.686)***
un			-16.017
			(81.203)
obj1			
pentagon			
coast			
Constant	29,918.27	-8,972.72	-10,382.62
	(39.434)***	(10,308.61)	(11,318.32)
Observations	274	208	193
R-squared	0.014	0.194	0.196

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.24: RE on GDP per capita; lagged betweenness centrality of all the links

	1	2	3	4	5	6
L1.btwnss_all_ln	-31.965	-23.856	-24.144	-24.53	-24.008	-24.176
	(22.895)	(15.522)	(17.209)	(16.896)	(16.88)	(16.861)
emp_gp		64,932.92	69,095.63	65,610.05	63,463.93	62,489.31
		(8,467.161)***	(9,203.615)***	(9,035.142)***	(8,876.475)***	(8,953.930)***
un			-50.328	-31.772	-33.258	-31.723
			(72.488)	(71.35)	(71.211)	(71.153)
obj1				-17,600.22	-9,721.28	-9,957.82
				(4,750.008)***	(4,749.536)**	(4,786.590)**
pentagon					16,027.35	16,552.18
					(4,124.592)***	(4,204.799)***
coast						2,919.35
						(3,879.61)
Constant	27,130.04	-20,749.93	-23,409.00	-16,903.83	-23,274.21	-23,741.75
	(2,309.928)***	(6,590.488)***	(7,108.151)***	(7,170.368)**	(7,047.050)***	(7,096.046)***
Observations	274	208	193	193	193	193

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.25: Robust RE on GDP per capita; lagged betweenness centrality of all the links

	1	2	3	4	5	6
L1.btwnss_all_ln	-31.965	-23.856	-24.144	-24.53	-24.008	-24.176
	(20.575)	(12.690)*	(13.506)*	(13.113)*	(13.086)*	(13.060)*
emp_gp		64,932.92	69,095.63	65,610.05	63,463.93	62,489.31
		(10,065.636)***	(11,094.770)***	(10,713.717)***	(10,113.009)***	(9,452.875)***
un			-50.328	-31.772	-33.258	-31.723
			(64.101)	(63.1)	(60.953)	(61.172)
obj1				-17,600.22	-9,721.28	-9,957.82
				(3,802.757)***	(3,643.109)***	(3,718.763)***
pentagon					16,027.35	16,552.18
					(4,976.347)***	(5,413.577)***
coast						2,919.35
						(5,539.67)
Constant	27,130.04	-20,749.93	-23,409.00	-16,903.83	-23,274.21	-23,741.75
	(2,143.367)***	(6,617.287)***	(7,269.478)***	(7,199.828)**	(7,013.096)***	(7,424.666)***
Observations	274	208	193	193	193	193

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.26: Robust FE on GDP per capita; lagged eigenvector centrality of all the links

	1	2	3
L1.eigen_all_ln	122.892	38.709	33.805
	(90.262)	(56.539)	(61.976)
emp_gp		49,308.29	51,892.93
		(13,902.018)***	(15,012.931)***
un			-16.166
			(88.262)
obj1			
pentagon			
coast			
Constant	30,844.61	-8,723.32	-10,394.56
	(626.808)***	(10579.53)	(11,649.75)
Observations	274	208	193
R-squared	0.007	0.175	0.178

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.27: RE on GDP per capita; lagged eigenvector centrality of all the links

	1	2	3	4	5	6
L1.eigen_all_ln	226.307	47.034	44.979	41.635	47.815	47.902
	(109.222)**	(79.054)	(84.591)	(82.65)	(82.113)	(81.98)
emp_gp		66,101.36	70,174.68	66,241.56	63,446.23	62,424.75
		(8,871.081)***	(9,592.875)***	(9,390.681)***	(9,203.237)***	(9,283.604)***
un			-54.584	-33.513	-35.074	-33.43
			(74.471)	(72.998)	(72.509)	(72.415)
obj1				-17,422.51	-9,583.37	-9,817.10
				(4,628.236)***	(4,659.936)**	(4,700.497)**
pentagon					16,004.97	16,521.84
					(4,046.882)***	(4,129.778)***
coast						2,851.80
						(3,809.53)
Constant	28,994.59	-21,175.16	-23,748.11	-16,994.94	-22,818.54	-23,224.45
	(2,249.129)***	(7,028.947)***	(7,561.197)***	(7,579.359)**	(7,419.520)***	(7,460.101)***
Observations	274	208	193	193	193	193
Number of code	76	70	65	65	65	65

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A7.28: Robust RE on GDP per capita; lagged eigenvector centrality of all the links

	1	2	3	4	5	6
L1.eigen_all_ln	226.307	47.034	44.979	41.635	47.815	47.902
	(89.944)**	(64.939)	(67.707)	(63.637)	(66.128)	(66.558)
emp_gp		66,101.36	70,174.68	66,241.56	63,446.23	62,424.75
		(10,283.454)***	(11,306.529)***	(10,831.525)***	(10,198.838)***	(9,494.925)***
un			-54.584	-33.513	-35.074	-33.43
			(68.815)	(67.066)	(64.295)	(64.406)
obj1				-17,422.51	-9,583.37	-9,817.10
				(3,652.285)***	(3,529.888)***	(3,602.296)***
pentagon					16,004.97	16,521.84
					(4,903.891)***	(5,351.900)***
coast						2,851.80
						(5,455.88)
Constant	28,994.59	-21,175.16	-23,748.11	-16,994.94	-22,818.54	-23,224.45
	(2,138.464)***	(6,746.281)***	(7,404.617)***	(7,255.616)**	(7,080.377)***	(7,448.122)***
Observations	274	208	193	193	193	193
Number of code	76	70	65	65	65	65

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Chapter 8

Conclusions

The aim of this chapter is to summarise the key findings of this doctoral thesis and further discuss its results. Firstly, based on the empirical results, the three research questions stated in the first chapter are briefly addressed here. Additionally, the contribution to the relevant literature is also highlighted. The next section focuses on policy recommendations based on the research's findings. This chapter ends with identifying the limitations of this study and recommendations for future research.

8.1 Research questions and further contribution to the literature

As analysed in the first chapter, there are three main research questions which initiated this doctoral research:

RQ1: How is the Internet infrastructure allocated across the European city-regions?

> As explained in Chapter 3, in order to approach the Internet infrastructure, the international Internet backbone network was studied. The geographical analysis firstly highlighted this network's trend to expand over time, including more links and also interconnecting more cities. Interestingly enough, the European part of the network appears to grow faster than the network of the extra-European links, a relative faster expansion of this infrastructure inside the European borders. Additionally, if the capacity of the backbone links is not taken into consideration, the Internet backbone network appears to be moderately centralised. However, when the capacity is included in the analysis, the network appears more centralised, indicating a higher concentration of the high capacity links. In more geographic terms, different roles were recognised for different cities, but one thing is for sure: over the six year period, London is the main hub of the international Internet backbone network in Europe. Its importance is not limited only to the European part of the global network, but London as well as New York, is one of the cities with the highest accumulation of bandwidth. Apart from London, Paris, Amsterdam and Frankfurt appear to be key locations for the topology of this network. Indeed, more than half of the total bandwidth accumulated in European city-regions is allocated across these four city-regions. Regardless of the importance of these hubs though, analysis

showed that they are not important enough to hold the network together, as no clear evidence of SF attributes were found. However, the existence of SW properties, such as low average distance and high clustering coefficient, highlight the efficiency of this network. The latter though appears to be quite different when technology (i.e. bandwidth) is excluded from the analysis as the actual topology of the Internet backbone links emerges. Based on this analysis, the clear role of some city-regions as gateway locations for their hinterlands emerged. The important thing is that these hinterlands are not only independent of country borders, but sometimes overcome Europe's borders, as they reflect post-colonial relations. Additionally the binary links also reflect location and physical geography, contrary to the bandwidth distribution which seems to be more related with economic geography.

RQ2: Which are the socio-economic factors that shape the distribution of the Internet infrastructure across the European city-regions?

The analysis identified a set of components which appear to affect the connective-ness and the connectivity of the city-regions with the Internet backbone network in Europe. In general, it can be said that the level of development, services and the knowledge economy, spatial structure and the physical transport and accessibility level are significant predictors of the likelihood of a city-region to be connected to a backbone network but also the level of the connectivity. Regardless of the lack of a unique explanation of the geography of the Internet backbone network in Europe, it can be concluded that the explanatory analysis concluded in both familiar and rather predictable results, since it is reflects largely the existing spatial, development and knowledge economy structures of metropolitan Europe.

RQ3: What are the impacts that the Internet infrastructure can generate on the development of city-regions in Europe?

The econometric modelling concluded that the Internet infrastructure, as is reflected in the weighted degree centrality, is a significant predictor for the level of regional economic development. Additionally,

> the analysis proved that wherever a significant causal relationship between the Internet infrastructure and regional economic development exists, then the causality runs from the Internet infrastructure to regional economic development and not the other way round. From the geography point of view, an interesting conclusion was the emergence of an almost north-south pattern of the causal relationship, with the northern city-regions appearing to be more efficient in exploiting the installed Internet infrastructure.

Apart from the above empirical results, which directly correspond to the three research questions, this doctoral research further contributes to the relevant literature. Firstly, as the Internet geography of Europe has not been very well examined, this study is valuable in understanding how the European city-regions participate in this global infrastructural network and how they are benefited by this. Indeed, as highlighted in section 2.3.2, the majority of the empirical studies in the emerging field of the Internet geography are concerned with the US Internet backbone network, but also the quantitative data used in these papers only refers to the early 2000s. This doctoral study comes to fill in the gap of empirical research about the European Internet geography, but also to present and analyse recent data after the telecoms crash of the early 2000s.

Furthermore, the explanatory analysis for the connectivity and the connectiveness of the European city-regions with the backbone network is on its own a contribution to the literature. As highlighted in section 2.3.2 apart from a very few and fragmented exceptions, no study has attempted in the past to analyse the socio-economic reasons affecting the distribution of the Internet backbone links across the urban network neither in the US nor in Europe.

Additionally, this doctoral research contributes to the field of world cities research. The adoption of a relational approach – whenever this is possible – is a contribution on its own due to the lack of such studies in this field (Taylor 2004; Short et al. 1996). In order to do so, this research contributes to the empirical justification of Castells' (1996) space of flows, bringing into light the inter-urban relations to the degree they are reflected in this infrastructural network. In short, this doctoral research contributes in bridging the gap between the theoretical work of

Castells (1996), Sassen (2000) and others and the lack of empirical research about the emerging space of flows, as identified by Taylor (2004).

However, the Internet is not the only facilitator of the world city process and as Taylor (2004) suggested the first layer of the space of flows consists both of the Internet infrastructure and the aviation network: while the latter transports the main actors of the global city process – the managerial elites as they are identified by Castells (1996), the former transports the ideas and the products of the digital economy. Additionally, a great discussion has taken place in the communications geography literature about the link between telecommunications and transportation. The burgeoning question is whether and to what degree there are complementarities between the two networks. This doctoral thesis draws upon these points by comparing the topology and the geography of these two networks as well as explaining the differences and the similarities in the way these two networks facilitate the world city process. The empirical results of this analysis, such as the illustration of the different roles that different cities perform in these two networks, contribute both to the world city but also to the communications geography literature.

In addition, this doctoral thesis contributes to the field of regional science as it not only confirms that the Internet infrastructure is a significant predictor of the level of economic development of the city-regions in Europe, but it also verifies the existence of a significant causal relationship running from the Internet infrastructure to regional economic development. These empirical results are a contribution to the relevant literature for two reasons. Firstly, it is the first time, to my knowledge at least, that empirical research has confirmed such a causal relationship. Regional scientists, economic geographers and urban planners have avoided including in their research such networks, because of their complex technical nature and the lack of data. However, the radical expansion of the Internet and the implementation of the digital economy raise the need for including this digital infrastructure in the research agenda. The robust results of this study can support this process. Secondly, there was always a debate about the direction of causality between infrastructural capital in general and more specifically within telecommunications – as was highlighted in section 2.5.6 – and (regional) economic development. This study contributes to the debate by the empirical results of the causality analysis. This doctoral thesis is the first study performing such a causal analysis for the Internet infrastructure.

To sum up, this doctoral research not only answers precisely the three research questions by conducting advanced empirical quantitative research tailored to the special needs of the study, but it also contributes to relevant but at the same time diverse fields of literature. Moreover, as it will be illustrated in the next section, the empirical analysis of this doctoral thesis can also result in policy implications.

8.2 Policy insights

Apart from the rather limited interest in relative terms that the research community has demonstrated for the link between ICTs and (regional) economic development, the same – and maybe at a higher degree – applies to policy makers (Cohen-Blankshtain et al 2004; Graham and Marvin 1996). Hence, issues such as ICTs, telecommunications, the Internet and the Internet infrastructure were rather neglected or in best cases misinterpreted by policy makers. However, recently a change has been noticed as policy makers have started showing interest in ICTs. A quite straightforward explanation for this relative lack of interest is the *invisible* and complex nature of this infrastructure (Batty 1990; Graham and Marvin 1996; Hackler 2003), contrary to the visible nature of other more traditional infrastructural capital such as the various transportation networks. As explained in Chapter 2, this *digital* infrastructure only becomes visible when it stops working (Star 1999). And apart from the lack of institutional data and knowledge about ICTs and the fact that policy-makers are more familiar with the characteristics of the transport infrastructure (Cohen et al 2002), it could be also assumed that the policy makers are reluctant to invest in an infrastructural element which is not directly visible to their voters.

Apart from the academic impact that this doctoral research attempts to generate in supporting the inclusion of ICTs and the Internet infrastructure in the field of regional science, this study also aspires to justify and promote the inclusion of these elements in the local and regional development agenda. Indeed, this study proved econometrically that for most of the interconnected (with international backbone networks) European city-regions, the accumulated bandwidth is a significant cause of economic development. The geographic analysis of the outcomes of the econometric modelling for the causal relation and their interpretation with the use of relevant theoretical approaches and results of prior empirical studies pointed out that the resulting north-south pattern in the significance of a causal relation is likely to be due to the spatial differentiation of the regional capacity for exploiting this infrastructure.

The above findings can justify the inclusion of elements such as the Internet infrastructure in the local and regional development agenda. As literature suggests (Cieslik and Kaniewska 2004; Wolde-Rufael 2007), the existence of a causal relationship between ICTs and economic development where the causality runs from the ICTs to economic development, can justify the inclusion of ICTs infrastructure in a policy framework for stimulating the economy. However, as the econometric analysis showed, the causal effect is not homogenous in space and a north-south pattern emerges, which can be explained by the regional capacity as outlined above. This non-homogenous character of the impacts of ICTs on (regional) economic development is the main reason for policy makers' misspecification about the investment on such infrastructure:

> “The impact of telecommunications technologies on regional development is not a straightforward mechanism. One of the greatest mistakes would be to expect a direct linkage between the supply of new technologies and economic and regional development. The link between these two elements, technology on one side and economic and regional development on the other, is a rather complex phenomenon. Its successful results stem mainly from a collection of essential elements which have to be present and have to be exploited in the right way” (Capello and Nijkamp 1996a, 235).

ICT related policies can be divided into three groups regarding their strategic approach (Cohen et al 2002):

- *Direct policies*. Such policies aim at promoting the availability and the use of ICTs. They target both the supply and demand of ICT infrastructure and services. They include a variety of policy tools from strategic city plans for ICTs implementation to policies for bridging the digital gap.
- *Indirect policies*. This group of policies intends to achieve non-ICT goals with the use of ICTs. For instance, the use of ICTs as a tool for stimulating the (regional) economy is such an example. Again, both supply and demand oriented policies are included here such as the provision of ICT infrastructure – supply side – or e-governance related services – demand side.
- *By the way policies*. These are policies of which the outcome only accidentally affects – directly or indirectly – the field of ICTs. The best example in the

literature is the US defence industry, which pushed ICT developments such as the Internet, but the policy goal was far from ICT related (Markusen 1988 cited in Cohen et al 2002).

However, the implementation of an indirect ICT policy for enhancing regional economic activity which only focuses on the supply side might not be enough. The investment in infrastructure appears to be a precondition for economic development, but it cannot automatically result in enhancing economic activity via the micro and the territorial effects described above (Gibbs and Tanner 1997; Gillespie 1991). Hence, as highlighted above, the supply side policies should be accompanied by an existing regional capacity for exploiting this production input or/and by direct demand side policies in order to advance this regional capacity.

However, unlike transport infrastructure, this precondition for development is mostly in the hands of the private sector (Priemus 2007). Telcos decide where and how they roll out their networks based on market assessments. The outcome of this process is usually the cumulative strengthening of the core regions and the resulting widening of the quality gap between core and peripheral regions in terms of connectivity, which in the long term might turn out to be a substantial burden for development (Camagni and Capello 2005). The question is whether and how can policy makers react on this process? Literature suggests the adoption of a moderate interventional approach in order to *correct* the outcome of the market forces (Cohen-Blankshtain et al 2004) and the inclusion of ICTs in urban policies (Horan and Jordan 1998; Couclelis 2000; and Cohen et al 2002). Indeed, recently cities have started being proactive in improving the level of ICT infrastructure in their territories (Hackler 2003). Examples include the extension of the municipal Internet broadband networks, the partnership with private telecom firms to extend their fibre networks, the building of networks and the afterwards opening to ISPs (Cohen-Blankshtain and Nijkamp 2004; van Winder and Woets 2004) as well as a number of EU initiatives (RACE, ESPRIT, BRITE, STAR, DRIVE – Capello and Spairani 2008, Camagni and Capello 2005).

To sum up, this doctoral thesis suggests that the level of the Internet infrastructure – at least as this is reflected in the accumulated international Internet backbone capacity – is a significant causal factor for economic development and because of this attribute, policy makers can use it as a means for stimulating the regional economy. However, two difficulties arise. Firstly, the effect of such infrastructure is not homogenous. Regardless of the need for such connection in order to achieve global Internet connectivity, not all places can be benefitted by this precondition for development. Regional capacity and related policies as well as other demand side policies are necessary in order for a city-region to exploit such endowment. Secondly, as this infrastructure is mostly a responsibility of the private sector and policy makers are not fully aware of its special (technical and non-technical) attributes, it is difficult to integrate such policy goals in urban and regional policies. Indeed, “ICT is a young concept, and ICT policy-making is still in its infancy” (Cohen et al 2002, 34) but still, according to Martin (2003 cited in Cieslik and Kaniewska 2004), investments in ICTs are preferred to financing highways as they promote technological convergence among regions with the use of public programmes for telecommunications, the Internet, and training of human capital.

8.3 Limitations and further research

This section is focused on presenting the main limitations of this study and also in raising some issues which would be worth being researched in the future. The limitations that this doctoral thesis faces are mostly situated at two levels: data and methodological limitations. Firstly, the data limitations are discussed.

As analysed extensively in this study, one of the main reasons why ICTs and more specifically the Internet are not included among the favourite research subjects of geographers, planners and regional scientists is the scarce and the inherent technical (and not only) complexity of relevant data. As a result, for the needs of this study and for reasons explained in Chapter 3, secondary data purchased by Telegeography (2007) is used. This strategic choice is accompanied by a set of potential limitations. Firstly, the data is collected for uses other than this doctoral thesis (White 2003). Indeed, Telegeography’s main customers are telecommunications companies and therefore the data collection process is designed to fit the needs of the industry. Secondly, in general the use of secondary data raises issues of trust about the accuracy of the observations and the consistency of the data collection process (ibid). Thirdly, it is common for secondary data to face comparability problems (Clark 2005). Hence, as stated in Chapter 6, the Telegeography data (2007) is not directly comparable with the KMI Research Group data (2001), which used to be the alternative provider for Internet infrastructure data.

And lastly, secondary data is rather inflexible as it cannot be modified once it is gained (Clark 2005).

The above general limitations for secondary data use can be addressed here by two arguments. Firstly, nowadays Telegeography is the only available source for such data. Regardless of the fact that such data is gathered for reasons different than this research, it fits with the needs of this doctoral thesis, with the lack of intra-country inter-city links being the main deficiency. However, this happens due to the complexity of the actual infrastructure, which makes it difficult for Telegeography to draw the distinction between inter-city and intra-city (MAN) networks in the same country (Telegeography 2007). Secondly, the long history of Telegeography in the field (Telegeography 2009) and its wide acceptance by researchers but also by the industry is a proof of its trustworthiness. However, in regards to the inflexibility and the lack of comparability, this is a common problem for studies based on secondary data which was taken into consideration during the research design. Apparently, similar limitations apply for the other secondary data sources used for this study.

Apart from the general limitations of the use of secondary data, another limitation of the existing dataset is the short time period of observations. This is also linked with the methodological limitations discussed below. Indeed, data was available only for six years. Regardless of the fact that this is the first study using panel data for modelling the impacts of Internet infrastructure, it would have been very useful especially for addressing the third research question about the regional economic impacts and for the application of the Granger causality tests, if more observations over time had been available. This would increase the robustness of the results. However, it should be underlined here that even with this rather narrow time dimension the results of the econometric analysis and more specifically the Granger causality tests are significant and robust.

Furthermore, another technical limitation which is also related with the Granger causality test is the actual test itself. Indeed, the Granger causality test is based on a bivariate model and does not allow controlling for the combined effect of other variables on the direction of causality (Hood III et al 2008). In spite of the need for further development of this method, it is still one of the most well accepted methods for testing causal relationships.

Regarding the methods used to address the first research question and mostly the networks analysis, it could be flagged as a possible limitation to the rather descriptive character of this method, at least on the way it was applied here. Contrary to what happens with statistical and econometric analysis, where usually the descriptive analysis is mostly the preparatory stage for the modelling applications, in network analysis the descriptive part is more important. Indeed, it is common to find papers in academic journals using the methods and techniques applied here. The modelling part of the network analysis, where the objective is to model the structure and the topology of the network, was out of the scope of this study. As the initial point of this study is the geography and not the network science, the topology and analytical tools related with this are used here as a means to explore and explain the geography of these networks. Therefore, in spite of being rather descriptive, the network analysis methodological tools utilised here fit with the needs and the objectives of this study.

Apart from the above limitations of this doctoral study, the novel character of the research field and the still emerging digital economy present an opportunity for further extensive research. Firstly, more *geography* can be incorporated in the research by including in the analysis the physical distance between cities. This element can show whether Tobler's (1970; 236) first law of geography[33] is valid for the infrastructural layer of the digital economy. Additionally, further normalization can also be introduced. The infrastructural capital could be weighted by the population or by the labour force in knowledge and technology intensive sectors.

Furthermore, the weighted or not variables of the infrastructural capital can be the input in an economic model such as a *production function*. Hence, apart from the research on the impact of the Internet infrastructure on the economic development level of city-regions with the use of econometric methods, it would be more robust from the economic theory point of view to study the economic impact of this infrastructure in a production function framework.

Moreover, another empirical but also novel research question that is worthy of studying is the use of the *network dependence* for exploring the relationship between the Internet infrastructure and regional economic development. The main hypothesis

---

[33] "Everything is related to everything else, but near things are more related than distant things."

for such research is that the economic development level of a city-region, apart from other factors, also depends on the economic development level of the city-regions that it interacts with. According to the widely used concept of *spatial dependence*, these city-regions are usually the direct neighbours or city-regions which are located under a specific distance threshold. In the proposed research question, instead of using spatial continuity or spatial distance, the network distance or even better the network capacity – bandwidth – between directly connected city-regions can be used.

Another field of further expansion of this doctoral study is the identification of power law distribution. As explained in section 4.6 OLS is a valid method in distinguishing power from exponential law distribution. However, it would be worth trying other methods suggested by the literature such as the maximum likelihood (Newman 2005), the Kolmogorov – Smirnov statistic (Clauset et al 2008), and the J test, KS test and the encompassing test (Russo et al 2007).

An additional analysis that needs to be undertaken as future research is an explanatory analysis of the Internet backbone distribution based on Telegeography data. As mentioned in Chapter 7, the explanatory analysis on KMI research took place in the very first stages of this doctoral thesis, when the Telegeography data was not available. Regardless of the comparability problem, the results of the explanatory analysis on the KMI data fit well with the overall results based on the Telegeography data. However, it would be worth exploring the factors that shape the distribution of the international Internet backbone network as it is represented by the Telegeography data. Apart from the methods used in Chapter 7, an alternative approach would be to use a panel data specification and decrease the explanatory variables using a method such as the *Theil's sequential elimination procedure*. Such a method can enhance the robustness of such an analysis and the definition of the pre-existing model could be based on the results of the current study.

Additionally, other elements of the long-haul Internet infrastructure such as the IXPs could be utilised. This element could shed light on the actual nodes of the Internet. In this doctoral study, the aggregation process took place at the city level. So for instance, London appears in the network as one node. In reality though, London is served by six IXPs (Telegeography 2009), which means that London has six nodes in the Internet backbone network. Moreover, such an analysis can shed some light into the actual flows transported over the backbone networks as statistics about the IP data flows pass through (or end at) the IXPs are available.

This last point leads to the main area suggested for further research: the geography of the inter-city IP data flows. From the geography standpoint, such a research project would enhance the field of the Internet geography as it would enable the research community to study the actual inter-city interactions in the digital economy. While this doctoral research is concerned with the supply side of the intercity interactions in the framework of the digital economy, such proposed research would shed light on the demand side of these interactions. Moreover, if the present doctoral thesis is related with the first – infrastructural – layer of the space of flows, this proposed study would focus on the second layer of the space of flows: the actual flows of the network society. The main difficulty for proceeding with such research is the lack of data, as the backbone providers, which collect data for IP origin-destination data-packets for their own usage (network maintenance), are not keen in disclosing this data for competition reasons. Additionally, there is no central authority with responsibility of collecting data for traffic flows (Kende 2000). Nonetheless, recent examples of partnerships between telcos and researchers from the fields of geography, planning and regional science can be considered as indications of progress and raise confidence for the feasibility of such a research project. These examples include the MIT SENSEable City Laboratory *NYTE* project, which focused on mapping the IP flows between New York City and the rest of the world cities (Ratti et al 2008), and the *currentcity.org* (2009) project which uses mobile phone calls data for urban scale modelling. From the complexity point of view, the availability and the analysis of such data which captures everything from our communications to our whereabouts may lead to the establishment of a wider foundation of a theory of complexity; and interestingly enough, “the complex system that we are most likely to tackle first in a truly quantitative fashion may not be the cell or the Internet but rather society itself” (Barabási 2009).

The above examples suggest fruitful avenues for further exploring the geography of the Internet at a variety of geographical scales.

Reference List

Abdul-Wahaba, S. A., Bakheitb, C. S. and Al-Alawi, S. M. (2005) 'Principal component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentrations', *Environmental Modelling & Software,* 20, pp. 1263–1271.

Ache, P. (2000) 'Cities in old industrial regions between local innovative milieu and urban governance - Reflections on city region governance', *European Planning Studies,* 8, (6), pp. 693-709.

Adamic, L. A. (2000) 'Zipf, Power-laws, and Pareto - a ranking tutorial', *Information Dynamics Lab, HP Labs,* [Online]. Available at: http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html (Accessed: 27/8/2009).

Adamic, L. A. and Huberman, B. A. (2002) 'Zipf's law and the Internet', *Glottometrics,* 3, pp. 143-150.

Afifi, A. A. and Clark, V. (1996) *Computer-aided multivariate analysis.* London: Chapman & Hall.

Aghion, P. and Howitt, P. (1998) *Endogenous growth theory.* Cambridge, MA: The MIT Press.

Albert, R. and Barabási, A.-L. (2002) 'Statistical mechanics of complex networks', *Reviews of modern physics,* 74, pp. 47-97.

Albert, R., Jeong, H. and Barabási, A.-L. (2000) 'Error and attack tolerance of complex networks', *NATURE,* 406, pp. 378-381.

Amaral, L. A. N., Scala, A., Barthelemy, M. and Stanley, H. E. (2000) 'Classes of small-world networks', *PNAS,* 97, pp. 11149–11152.

Amin, A. and Thrift, N. (1992) 'Neo-Marshallian nodes in global networks', *International Journal of Urban and Regional Research,* 16, (4), pp. 571-87.

Antonelli, C. (2003) 'The digital divide: understanding the economics of new information and communication technology in the global economy', *Information Economics and Policy,* 15, (2), pp. 173-99.

Armstrong, H. and Taylor, J. (2000) *Regional economics and policy.* Massachusetts: Blackwell Publishers.

Arnaud, B. S. (2009) *Green IT/Broadband and Cyber-Infrastructure*. Available at: http://green-broadband.blogspot.com/ (Accessed: 21-8-2009).

AT&T (2009) *AT&T backbone network in Europe*. Available at: http://www.corp.att.com/globalnetworking/media/network_map.swf

(Accessed: 7/8/2009).

Atkinson, R. D. and Mckay, A. (2007) *Digital Prosperity.* Washington, DC: ITIF

Audestad, J. A. (2007) 'Internet as a multiple graph structure: the role of the Transport Layer', *Information Security Technical Report,* 12, (1), pp. 16-23.

Bakis, H. (1981) 'Elements for a geography of the telecommunication', *Geographic Research Forum,* 4, pp. 31-45.

Banerjee, A. V. and Duflo, E. (2003) 'Inequality and growth: what can the data say?' *Journal of Economic Growth*, 8, (3), pp. 267-99.

Banister, D. and Berechman, J. (2003) *Transport investment and economic development.* London: Routledge.

Banister, D. and Stead, D. (2004) 'Impact of Information and Communications Technology on Transport', *Transport Reviews,* 24, (5), pp. 611-32.

Barabási, A.-L. (2009) 'Scale-Free networks: A decade and beyond', *SCIENCE*, 325, (5939), pp. 412-3.

Barabási, A.-L. and Albert, R. (1999) 'Emergence of Scaling in Random Networks', *SCIENCE,* 286, (15 October), pp. 509-512.

Baran, P. (1964) *On distributed communications: I Introduction to distributed communications networks.*

Basilevsky, A. (1994) *Statistical factor analysis and related methods.* New York: John Willey & Sons, INC.

Batty, M. (1990) 'Invisible cities', *Environment and Planning B,* 17, pp. 127-30.

Batty, M. (1991) 'Urban Information Network: The Evolution and Planning of Computer-Communications Infrastructure', in Brotchies, J., Batty, M., Hall, P. and Newton, P.(eds) *Cities of the 21st century.*New York: Wiley and Sons, Inc.

Batty, M. (1997) 'Virtual Geography', *Futures,* 29, (4/5), pp. 337-352.

Batty, M. (2001) 'Cities as small worlds', *Environment and Planning B,* 28, pp. 637 - 6 38.

Baum, C. F. (2006) *An Introduction to Modern Econometrics Using Stata,* Boston, MA: STATA Press.

Beil, R., Ford, G. and Jackson, J. (2005) 'On the relationship between telecommunications investment and economic growth in the United States', *International Economic Journal,* 19, (1), pp. 3-9.

Belgacom (2009) *Belgacom backbone network in Europe*. Available at:

http://www.belgacom-ics.com/Home/Network/TransportNetwork/tabid/116/Default.aspx (Accessed: 7/8/2009).

Biehl, D. (1991) 'The role of infrastructure in regional development', in Vickerman, R. W.(ed), *Infrastructure and Regional Development.* London: Pion Limited.

Black, W. R. and Nijkamp, P. (2006) 'Transportation, Communication and Sustainability: In Search of a Pathway to comparative Research ', in Reggiani, A. and Schintler, L.(eds) *Methods and Models in Transport and Telecommunications Cross Atlantic Perspectives* Berlin: Springer.

Blum, B. S. and Goldfarb, A. (2006) 'Does the internet defy the law of gravity?' *Journal of International Economics,* 70, pp. 384-405.

Blume, L. and Sack, D. (2008) 'Patterns of Social Capital in West German Regions', *European Urban and Regional Studies,* 15, (3), pp. 229-48.

Bonacich, P. (2007) 'Some unique properties of eigenvector centrality', *Social Networks,* 29, pp. 555–564.

Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002) 'UCINET for Windows: Software for Social Network Analysis', *Harvard, MA: Analytic Technologies*.

Borland, J. and Hu, J. (2004) *Why policies must change.* CNET News.com Digital Agenta Broadband: Breaking the digital gridlock

Braudel, F. (1984) *The Perspective of the World.* London: Collins.

Brenner, N. (1998a) 'Global cities, glocal states: global city formation and state territorial restructuring in contemporary Europe', *Review of International Political Economy,* 5, (1), pp. 1-37.

Brenner, N. (1998b) 'Between fixity and motion: accumulation, territorial organization and the historical geography of spatial scales', *Environment and Planning D,* 16, pp. 459-81.

Bresnahan, T. F. and Trajtenberg, M. (1995) 'General purpose technologies ‘Engines of growth’?' *Journal of Econometrics,* 65, (1), pp. 83-108.

Breusch, T. S. and Pagan, A. R. (1980) 'The Lagrange multiplier test and its applications to model specification in econometrics', *Review of Economic Studies,* 47, (1), pp. 239-53.

Bronzini, R. and Piselli, P. (2009) 'Determinants of long-run regional productivity with geographical spillovers: The role of R&D, human capital and public infrastructure', *Regional Science and Urban Economics,* 39, pp. 187-99.

Burton-Jones, A. (1999) *Knowledge capitalism.* Oxford: Oxford Press.

Button, K. (2000) 'New approacjes to spatial economics', *Growth and Change*, 31, pp. 480-500.

Button, K. and Stough, R. (2000) *Air transportation networks. Theory and policy implications.* Cheltenham, UK - Northampton, USA: Edward Elgar.

CAIDA (2009) *The Cooperative Association for Internet Data Analysis*. Available at: http://www.caida.org (Accessed: 28/5/2009).

Cairncross, F. (2001) *The death of distance 2.0.* London: Texere Publishing Limited.

Camagni, R. and Capello, R. (2005) 'ICTs and territorial competitiveness in the era of internet', *Annals of Regional Science*, 39, (3), pp. 421-38.

Capello, R. and Nijkamp, P. (1996a) 'Regional variations in production network externalities', *Regional Studies*, 30, (3), pp. 225 - 237.

Capello, R. and Nijkamp, P. (1996b) 'Telecommunications technologies and regional development: theoretical considerations and empirical evidence', *Annals of Regional Science*, 30, (1), pp. 7-30.

Capello, R. and Spairani, A. (2008) 'Ex-ante evaluation of European ICT policies: efficiency vs. cohesion scenarios', *International Journal of Public Policy* 3, (3-4), pp. 261-80.

Castells, M. (1996) *The Rise of the Network Society.* Oxford: Blackwell.

Castells, M. (2001) *The internet galaxy.* Oxford: Oxford University Press.

Chakraborty, C. and Nandi, B. (2003) 'Privatization, telecommunications and growth in selected Asian countries: an econometric analysis', *Communications and Strategies*, 52, (3), pp. 31-47.

Chamberlain, G. (1982) 'Multivariate regression models for panel data', *Journal of Econometrics*, 18, (1), pp. 5-45.

Charles, D., Bradley, D., Chatterton, P., Coombes, M. and Gillespie, A. (1999) *Core Cities: Key Centres for Regeneration, Synthesis Report.* Newcastle upon Tyne: CURDS

Cheshire, P. (1990) 'Explaining the Recent Performance of the European Community's Major Urban Region', *Urban Studies*, 27, (3), pp. 311-33.

Cheshire, P., Carbonaro, G. and Hay, D. (1986) 'Problems of Urban Decline and Growth in EEC Countries: Or Measuring Degrees of Elephantness', *Urban Studies*, 23, (2), pp. 131-49.

Cho, S. and Mokhtarian, P. L. (2007) 'Telecommunications and travel demand and supply: Aggregate structural equation models for the US', *Transportation Research Part A,* 41, (1), pp. 4-18.

Choi, J. H., Barnett, G. A. and Chon, B.-S. (2006) 'Comparing world city networks: a network analysis of Internet backbone and air transport intercity linkages', *Global Networks,* 6, (1), pp. 81–99.

Chon, B.-S. (2004) 'The dual structure of global networks in the entertainment industry: interorganizational linkage and geographical dispersion', *The International Journal on Media Management,,* 6, (3), pp. 194 - 206.

Chu, N., Oxley, L. and Carlaw, K. (2005) 'ICT and causality in the New Zealand economy', *Proccedings of the 2005 International Conference on Simulation and Modelling*.

Cieslik, A. and Kaniewska, M. (2004) 'Telecommunications Infrastructure and Regional Economic Development: The Case of Poland', *Regional Studies,* 38, (6), pp. 713-25.

CIXP (2008) *CERN Internet eXchange Point*. Available at: http://cixp.web.cern.ch/cixp/ (Accessed: 7/8/2009).

Clark, F. (2005) 'Secondary data', in Flowerdew, R. and Martin, D.(eds) *Methods in human geography.* Essex: Pearson.

Clauset, A., Shalizi, C. R. and Newman4, M. E. J. (2008) 'Power-law distributions in empirical data', *physics.data-an*.

Cohen, G., Salomon, I. and Nijkamp, P. (2002) 'Information-communications technologies (ICT) and transport: does knowledge underpin policy?' *Telecommunication Policy,* 26, (1-2), pp. 31-52.

Cohen, S. S., DeLong, J. B. and Zysman, J. (2000) 'Tools for thought: what is new and important about the ‘e-conomy’', *Berkeley Roundtable on International Economics, Berkeley CA,* Working paper no. 138.

Cohen, S. S., DeLong, J. B., Weber, S. and Zysman, J. (2001) *Tracking a transformation: e-commerce and the terms of competition in industries.* BRIE-IGCC E-conomy Project Task Force, Washington, DC: Brookings Press.

Cohen-Blankshtain, G. and Nijkamp, P. (2004) 'The appreciative system of urban ICT policies: an Analysis of perceptions of urban policy makers', *Growth and Change,* 35, (2), pp. 166-97.

Cohendet, P. and Steinmueller, W. E. (2000) 'The codification of knowledge. A conceptual and empirical exploration in industrial and corporate change', *Industrial and Corporate Change,* 9, (2), pp. 195-209.

COLT (2009) *COLT backbone network in Europe*. Available at: http://www.colt.net/oracleUCM/groups/public/documents/digitalasset/colt_011195.pdf (Accessed: 7/8/2009).

Corey, K. E. (1982) 'Transactional forces and the metropolis', *ekistics,* 297, pp. 416-23.

Cornford, G. and Gillespie, A. (1993) 'Cable systems, telephony and local economic development in the UK', *Telecommunication Policy,* 17, (8), pp. 589-603.

Couclelis, H. (2000) 'From sustainable transportation to sustainable accessibility: Can we avoid a new tragedy of the commons?' in Janelle, D. and Hodge, D.(eds) *Information, place and cyberspace: Issues in accessibility.* Berlin: Springer.

Cronin, F. J., Parker, E. B., Colleran, E. K. and Gold, M. A. (1991) 'Telecommunications infrastructure and economic growth: an analysis of causality', *Telecommunications Policy,* 15, (6), pp. 529-35.

Cronin, F. J., Parker, E. B., Colleran, E. K. and Gold, M. A. (1993a) 'Telecommunications infrastructure investment and economic development', *Telecommunications Policy,* 17, (6), pp. 415-30.

Cronin, F. J., Colleran, E. K., Herbert, P. L. and Lewitzky, S. (1993b) 'Telecommunications and growth: the contribution of telecommunications infrastructure investment to aggregate and sectoral productivity', *Telecommunications Policy,* 17, (9), pp. 677-90.

Crucitti, P., Latora, V., Marchiori, M. and Rapisarda, A. (2003) *Energy and information transfer in biological systems: how physics could enrich biological understanding.* Italy,

Crucitti, P., Latorab, V., Marchioric, M. and Rapisarda, A. (2004) 'Error and attacktolerance of complex networks', *Physica A,* 340, pp. 388 – 394.

Cukier, K. N. (1998) *Telegeography 1999.* Washington, DC: Telegeography

currentcity.org (2009) *currentcity.org home page*. Available at: http://currentcity.org (Accessed: 15/7/2009).

Cushman & Wakefield. (2008) *European cities monitor.* Cushman & Wakefield: Global real estate solutions

Cutrini, E. (2009) 'Specialization and concentration from a twofold geographical perspective: evidence from Europe', *Regional Studies,* forthcoming.

Dasgupta, P. and David, P. A. (1994) 'Toward a new economics of science', *Policy Research,* 23, pp. 487-521.

Davoudi, S. (2003) 'Polycentricity in Euroepan spatial planning: from an analytical tool to a normative agenda', *European Planning Studies,* 11, (8), pp. 979-99.

Derudder, B. and Witlox, F. (2005) 'An appraisal of the use of airline data in assessing the world city network: a research note on data', *Urban Studies,* 42, (13), pp. 2371–2388.

Derudder, B. and Witlox, F. (2008) 'Mapping world city networks through airline flows: context, relevance, and problems', *Journal of Transport Geography,* 16, (5), pp. 305-312.

Devriendt, L., Derudder, B. and Witlox, F. (2008) 'Cyberplace and cyberspace: two approaches to analyzing digital intercity linkages', *Journal of Urban Technology,* 15, (2), pp. 5-32.

Dickinson, R. E. (1947) *City, region and regionalism.* London: Routledge.

D'Ignazio, A. and Giovannetti, E. (2007) 'Spatial dispersion of interconnection clusters in the European Internet ', *Spatial Economic Analysis,* 2, (3), pp. 219-36.

Dixon, R. J. and Thirlwall, A. P. (1975) 'A model of regional growth rate differentials along Kaldorian lines', *Oxford Economic Press,* 27, pp. 201-14.

Dodge, M. (2008) *Understanding Cyberspace Cartographies: A Critical Analysis of Internet Infrastructure Mapping.* Thesis. UCL.

Dodge, M. and Kitchin, R. (2000) *Mapping Cyberspace.* London and New York: Routledge.

Dodge, M. and Shiode, N. (2000) 'Where on Earth is the Internet? An empirical investigation of the geography of Internet real estate', in Wheeler, J. O., Aoyama, Y. and Warf, B.(eds) *Cities in the Telecommunications Age: The Fracturing of Geographies.*New York: Routledge, pp. 42-53.

Dodge, M. and Zook, M. (2009) 'Internet Based Measurement', in Kitchin, R. and Thrift, N.(eds) *The International Encyclopedia of Human Geography.* Elsevier.

Dougherty, C. (2002) *Introduction to econometrics.* Oxford: Oxford University Press.

Drucker, P. F. (1998) 'From Capitalism to Knowledge Society', in Neef, D.(ed), *The* *knowledge economy.* Woburn, MA: Butterworth-Heinemann.

Dutta, A. (2001) 'Telecommunications and economic activity: an analysis of granger causality', *Journal of Management Information Systems,* 17, (4), pp. 71-95.

EC (2009) *NUTS Regions*. Available at: http://ec.europa.eu/eurostat/ramon/nuts/splash_regions.html (Accessed: 6/8/2009).

EC. (1999) *ESDP - European Spatial Development Perspective.* Luxembourg: Office for Official Publications of the European Communities

EC. (2004) *Strengthening competitiveness through co-operation. European research in information and communication technologies.* Brussels:

Economist. (2002) 'The great telecoms crash', 18 July 2002.

Erdil, E. and Yetkiner, I. H. (2009) 'The Granger-causality between health care expenditure and output: a panel data approach', *Applied Economics,* 41, pp. 511-18.

Erdös, P. and Rényi, A. (1959) 'On random graphs', *Publicationes Mathematicae Debrecen,* 6, pp. 290–97.

ESPON (2006) *Database Public Files*. Available at: http://www.espon.eu/ (Accessed: 28/5/2009).

ESPON. (2005a) *Potentials for polycentric development in Europe.* Luxemburg:

ESPON. (2005b) *Project 1.2.2. Telecommunication Services and Networks: Territorial Trends and Basic Supply of Infrastructure for Territorial Cohesion.* Luxemburg:

Euro-IX (2008) *The European IXP Scene*. Available at: http://www.euro-ix.net/news/meetevent/gpf3_euro-ix.pdf (Accessed: 12/10/2008).

Euro-IX. (2006) *2006 Report on European IXPs.* European Internet Exchange Association

Eurostat (2009) *Regional Data*. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/ (Accessed: 28/5/2009).

Evans-Cowley, J., Malecki, E. J. and McIntee, A. (2002) 'Planning responses to telecom hotels: What accounts for increased regulation of co-location facilities?' *Journal of Urban Technology,* 9, (3), pp. 1-18.

Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999) 'On power-law relationships of the Internet topology', *Computer Communication Review,* 29, pp. 251-262.

Fekedulegn, B. D., Colbert, J. J., R.R. Hicks, J. and Schuckers, M. E. (2002) 'Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology', *United States Department of Agriculture, Forest Service, Northeastern Research Station*, pp. Research Paper NE-721.

Field, A. (2005) *Discovering Statistics using SPSS for Windows.* London, Thousand Oaks, New Delhi: SAGE Publications.

Filzmoser, P. and Croux, C. (2002) 'A projection algorithm for regression with collinearity', in Jajuga, K., Sokolowski, A. and Bock, H.(eds) *Classification, Clustering and Data Analysis.* Vol. 227 – 234 Berlin: Springer.

Finkel, S. E. (1995) *Causal analysis with panel data.* Thousan Oaks, California: SAGE Publications, Inc.

Florens, J.-P. and Mouchart, M. (1982) 'Note on noncausality', *Econometrica,* 50, (3), pp. 583-91.

Fortune. (1999) 'The need for speed', *Fortune Technology Buyer's Guide*, pp. 122-6.

Fosco, C. (2004) *Local preferential attachment*. Available at: http://merlin.fae.ua.es/constanza/ (Accessed: 16/05/2008).

Freeman, C. (1987) *Technoloy policy and economic performance: lessons from Japan.* London: Pinter Publishes Limited.

Freeman, C. and Perez, C. (1988) 'Structural crisies of adjustment: business cycles and investment behaviour', in Dosi, G., Freeman, C., Nelson, R., Silverberg, G. and Soete, L.(eds) *Technical Change and Economic Theory.* London: Pinter Publishers.

Freeman, C. and Soete, L. (1997) *The economics of industrial innovation.* London, New York: CONTINUM.

Friedmann, J. (1986) 'The world city hypothesis', *Development and Change,* 17, (1), pp. 69-84.

Friedmann, J. (1995) 'Where we stand: a decade of world city research', in Knox, P. L. and Taylor, P. J.(eds) *World cities in a world-system.* Cambridge: Cambridge University Press, pp. 21-47.

Fujita, M. and Krugman, P. (2004) 'The new economic geography: Past, present and the future', *Papers in Regional Science,* 83, (1), pp. 139-164.

Garcia, B. E. (2000) *Global crossing to build facility*. Available at: http://www.internetcoast.org/index.php?src=news&refno=370 (Accessed: 3/12/2008).

Gastner, M. T. and Newman, M. E. J. (2005) 'The spatial structure of networks', *The European Physical Journal B - Condensed Matter and Complex Systems,* 49, (2), pp. 247-252.

Geddes, P. (1915) *Cities In Evolution.* London: Williams.

Geels, F. W. and Smit, W. A. (2000) 'Failed technology futures: pitfalls and lessons from a historical survey', *Futures,* 32, (9/10), pp. 867–85.

Gibbs, D. and Tanner, K. (1997) 'Information and Communication Technologies and local economic development policies: The British Case', *Regional Studies,* 31, (8), pp. 765-74.

Gibson, W. (1984) *Neuromancer.* New York: Ace books.

Gilder, G. (1995) *Forbes ASAP,* February 27, p.56.

Gilder, G. (2000) *Telecosm.* New York: Free Press.

Gillena, D. and Morrison, W. G. (2003) 'Regulation, competition and network evolution in aviation', *Journal of Air Transport Management,* 11, (3), pp. 161-174.

Gillespie, A. (1991) 'Advanced communications networks, territorial integration and local development', in Camagni, R.(ed), *Innovation networks: spatial perspectives.* London: Belhaven.

Gillespie, A. and Robins, K. (1989) 'Geographical Inequalities: the spatial bias of the new communications technology', *Journal of Communication,* 39, (3), pp. 7-18.

Gillespie, A., Richardson, R. and Cornford, J. (2001) 'Regional development and the new economy', *European Investment Bank Papers,* 6, (1), pp. 109-31.

Goldberger, A. S. (1964) *Econometric Theory.* New York: John Wiley & Sons.

Gore, A. (1993) *Remarks by Vice President Al Gore at National Press Club, December 21.* Available at: http://www.ibiblio.org/nii/goremarks.html (Accessed: 31/1/2009).

Gorman, S. P. and Kulkarni, R. (2004) 'Spatial small worlds: new geographic patterns for an information economy', *Environment and Planning B,* 31, (2), pp. 273-296.

Gorman, S. P. and Malecki, E. J. (2000) 'The networks of the Internet: an analysis of provider networks in the USA', *Telecommunications Policy,* 24, (2), pp. 113-134.

Gorman, S. P. and Malecki, E. J. (2002) 'Fixed and fluid: stability and change in the geography of the Internet', *Telecommunications Policy,* 26, (7-8), pp. 389-413.

Gorman, S. P. and McIntee, A. (2003) 'Tethered connectivity? The spatial distribution of wireless infrastructure', *Environment and Planning A,* 35, (7), pp. 1157-1171.

Gottmann, J. (1961) *Megalopolis: The urbanized Northeastern Seaboard of the United States.* New York, NY:

Gottmann, J. (1983) *The coming of the Transactional City.* College Park, MD: University of Maryland, Institute for Urban Studies.

Graham, S. (1999) 'Global grids of glass: on global cities, telecommunications and planetary urban networks', *Urban Studies,* 36, (5-6), pp. 929-49.

Graham, S. (2004) 'Excavating the material geographies of cybercities', in Graham, S.(ed), *The cybercities reader.* London: Routledge.

Graham, S. and Marvin, S. (1996) *Telecommunications and the city.* London and New York: Routledge.

Graham, S. and Marvin, S. (2001) *Splintering urbanism.* London, New York: Routledge.

Granger, C. W. J. (1969) 'Investigating causal relations by econometric and cross-spectral methods', *Econometrica,* 37, (3), pp. 424-38.

Greenstein, S. M. (2004) 'The Economic Geography of Internet Infrastructure in the United States', in Cave, M., Majumdar, S. and Vogelsang, I.(eds) *Handbook of Telecommunications Economics, Volume II.* Elsevier.

Grubesic, T. H. and O'Kelly, M. E. (2002) 'Using points of presence to measure accessibility to the commercial Internet', *Professional Geographer,* 54, (2), pp. 259-278.

GTS (2009) *GTS European backbone network in Europe*. Available at: http://www.gtsce.com/maps/index.php (Accessed: 7/8/2009).

Guimera, R., Mossa, S., Turtschi, A. and Amaral, L. A. N. (2005) 'The worldwide air transportation network : Anomalous centrality, community structure and cities' global roles', *Proceedings of the National Academy of Sciences of the United States of America,* 102, (22), pp. 7794-7799.

Hackler, D. (2003) 'Invisible Infrastructure and the City: The Role of Telecommunications in Economic Development', *American Behavioral Scientist,* 46, (8), pp. 1034-55.

Hall, P. (1966) *The World Cities.* London: Weidenfeld and Nicolson.

Hall, P. (1998) 'Globalization and the world cities', in Lo, F.-C. and Yeung, Y.-m.(eds) *Globalization and the world of large cities.* Tokyo: The United Nations University Press, pp. 17-36.

Hall, P. (2009) 'Looking backward, looking forward: the city region of the mid-21st century', *Regional Studies,* 43, (6), pp. 803-17.

Hall, P. and Pain, K. (2006) *The polycentric metropolis: learning from mega-city regions in Europe.* London: Earthscan.

Hanneman, R. A. and Riddle, M. (2005) 'Introduction to social network methods', *Riverside, CA: University of California*, pp. Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/).

Hardy, A. (1980) 'The role of the telephone in economic development', *Telecommunication Policy,* 4, (4), pp. 278-86.

Harris, G. R. (1998) 'The Internet as a GTP: Factor market implications', in Helpman, E.(ed), *General Purpose Technologies and Economic Growth.* Cambridge: MIT Press, pp. 145-66.

Harvey, D. (1982) *The limits to capital.* Chicago: University of Chicago Press.

Hausman, J. A. (1978) 'Specification tests in econometrics', *Econometrica,* 46, (6), pp. 1251-71.

Hawkinson, J. and Bates, T. (1996) *Guidelines for creation, selection, and registration of an Autonomous System (AS).* IETF

Helpman, E. (1998) 'General purpose technologies and economic growth: Introduction', in Helpman, E.(ed), *General Purpose Technologies and Economic Growth.* Cambridge: MIT Press, pp. 1-13.

Hepworth, M. (1989) *Geography of the Information Economy.* London: Belhaven Press.

Hirschman, A. (1958) *The strategy of economic development.* New York: Yale University Press.

Hoffmann, R., Lee, C.-G., Ramasamy, B. and Yeung, M. (2005) 'FDI and pollution: a granger causality test using panel data', *Journal of International Development,* 17, (3), pp. 311-17.

Holloway, S. (2003) *Traight and level: Practical airline economics.* Hampshire: Ashgate.

Holtz-Eakin, D., Newey, W. and Rosen, H. (1985) 'Implementing causality tests with

Panel data, with an example from local public finance', *NBER Technical Paper Series No. 48*.

Hood III, M. V., Kidd, Q. and Morris, I. L. (2008) 'Two Sides of the Same Coin? Employing Granger Causality Tests in a Time Series Cross-Section Framework', *Political Analysis,* 16, (3), pp. 324-44.

Hoover, K. D. (2001) *Causality in macroeconomics.* Cambridge: Cambridge University Press.

Horan, T. A. and Jordan, D. R. (1998) 'Integrating transportation and telecommunications planning in Santa Monica', *Journal of Urban Technology,* 5, (2), pp. 1-20.

Hsiao, C. (1986) *Analysis of panel data.* Cambridge: Cambridge University Press.

Hsiao, C. (2003) *Analysis of panel data.* Cambridge: Cambridge University Press.

Huddleston, J. R. and Pangotra, P. P. (1990) 'Regional and local economic impacts of transportation investment', *Transportation Quartely,* 44, (4), pp. 579-94.

Hurlin, C. and Venet, B. (2003) 'Granger causality test in panel data models with fixed coefficients', *Working Paper, University of Oleans*.

IANA (2007) *Internet Assigned Number Authority website*. Available at: (Accessed: 12/10/2007).

ICAO (2008) *Data from the International Civil Aviation Organization*. Available at: http://www.icaodata.com/ (Accessed: 03/02/2009).

Im, K. S., Pesaran, M. H. and Shin, Y. (2003) 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics*, 115, (1), pp. 53-74.

ITU. (2004) *Via Africa. Creating local and regional IXPs to save money and bandwidth.* Geneva:

Jochimsen, R. (1966) *Theories der Infrastructur, Grundlagen der Marktwirtschaftlichen Entwicklung.* Tubingen: J C B Mohr.

Johnston, J. and Dinardo, J. (1997) *Econometric methods.* New York: McGraw-Hill.

Kaldor, N. (1970) 'The case for regional policies', *Scottish Journal for Political Economy,* 18, pp. 337-48.

Kam, J. (2006) 'No pain, No Gain: Rethinking the Telecoms Crash', *Technology Analysis & Strategic Management,* 18, (5), pp. 497-514.

Kay, J. (1993) 'Efficiency and private capital in the provision of infrastructure', in OECD(ed), *Infrastructure Policies for the 1990s.*Paris: OECD.

Keeling, D. J. (1995) 'Transport and the world city paradigm', in Knox, P. L. and

Taylor, P. J.(eds) *World cities in a world system.* Cambridge: Cambridge University Press.

Kellerman, A. (1993) *Telecommunications Geography.* London: Belhaven Press.

Kellerman, A. (2002) *The Internet on Earth.* West Sussex: Wiley.

Kende, M. (2000) 'The Digital Handshake: Connecting Internet Backbones', *OPP Working Paper No. 32*, [Online]. Available at: http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp32.pdf (Accessed: 27/8/2009).

Kharif, O. (2001) 'The Fiber-Optic "Glut" - in a New Light', *Business Week*, 31/8/2001.

Kitchin, R. (1998a) *Cyberspace.* New York: Wiley.

Kitchin, R. (1998b) 'Towards geographies of cyberspace', *Progress in Human Geography,* 22, (3), pp. 385-406.

KMI Research Group (2001) *PAN European fiber optic network routes planned or in place*. Available at: http://www.kmiresearch.com/ (Accessed: not available anymore).

Kolko, J. (1999) 'The death of cities? The death of distance? Evidence from the geography of commercial Internet Usage', *Selected papers from the telecommunications policy research conference 1999.* Newcastle.

Kondratieff, N. (1926) 'The Long Waves in Economic Life', *Archiv für Sozialwissenschaft und Sozialpolitik*, 56, (3), pp. 573-609, *reproduced and translated in Review of Economic Statistics Vol. XVII, No. 6, November 1935, pp. 105-115*.

Krugman, P. (1991a) *Geography and Trade.* Cambridge: MIT Press.

Krugman, P. (1991b) 'Increasing Returns and Economic-Geography', *Journal of Political Economy,* 99, (3), pp. 483-499.

Krugman, P. (1998) 'What's new about the new economic geography?' *Oxford Review of Economic Policy,* 14, (2), pp. 7-17.

Kunzmann, K. R. (1998) 'World city regions in Europe: Structural change and future challenges', in Lo, F.-C. and Yeung, Y.-m.(eds) *Globalization and the world of large cities.*Tokyo: The United Nations University Press, pp. 37-75.

Lambooy, J. G. (2002) 'Knowledge and urban economic development: an evolutionary perspective', *Urban Studies,* 39, (5-6), pp. 1019-35.

Lambooy, J., Nagengast, E., Raat, N. and Veldkamp, L. (2000) *De ruimtelijke* *effecten van ICT in Nederland. Een essay [The Spatial Effects of ICT in the Netherlands. An Essay]*. Amsterdam: Regioplan Stad en Land BV.

Landes, D. (1998) *The Wealth and Poverty of Nations.* New York: W.W. Norton.

Latora, V. and Marchiori, M. (2001) 'Efficient behavior of small-world networks', *Physical Review Letters,* 87.

Latora, V. and Marchiori, M. (2002) 'Is the Boston subway a small-world network?' *Physica A,* 314, pp. 109-113.

Lee, H.-S. (2009) 'The networkability of cities in the international air passenger flows 1992–2004', *Journal of Transport Geography,* 17, (3), pp. 166-75.

Levin, A., Lin, C.-F. and Chu, C.-S. J. (2002) 'Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties', *Journal of Econometrics,* 108, (1), pp. 1-24.

Leydesdorff, L. (2006) The knowledge-based economy: modeled, measured, simulated. Boca Raton, FL: Universal Publishers.

Li, L., Alderson, D., Tanaka, R., Doyle, J. C. and Willinger, W. (2005) 'Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications (Extended Version)', *Internet Mathematics,* 2, (4), pp. 431-523.

Liebenau, J., Atkinson, R., Karrberg, P., Catro, D. and Ezell, S. (2009) The UK's Digital Road to Recovery. LSE Enterprise LTD. and the Information Technology and Innovation Foundation.

Lipsey, R. G., Carlaw, K. I. and Bekar, C. (2005) *Economic Transformations: General Purpose Technologies, and Long Term Economic Growth.* Oxford: Oxford University Press.

Liu, R. X., Kuang, J., Gong, Q. and Hou, X. L. (2003) 'Principal component regression analysis with SPSS', *Computer Methods and Programs in Biomedicine,* 71, pp. 141-47.

Louter, P. (2001) *Ruimte voor de digitale economie. Verkenning van de relaties tussen ICT en uimtelijkeconomische ontwikkeling [Space for the Digital Economy. Survey of the Relationships between ICT and Spatial-economic Development]*. Delft: TNO Inro.

Lundvall, B.-A. (ed.) (1992) *National innovation systems: towards a theory of innovation and interactive learning.* London: Pinter.

Maddala, G. S. (2001) *Introduction to econometrics.* Baffins Lane, Chichester: John Wiley and Sons Ltd.

Maignan, C., Pinelli, D. and Ottaviano, G. I. P. (2003) 'ICT, clusters and regional cohesion: A summary of theoretical and empirical research', *NOTA DI LAVORO - working paper* 58, pp. 1-17.

Mairesse, J. (1990) 'Time-series and cross-sectional estimates on panel data: why are they different and why they should be equal', in Hartog, J., Ridder, G. and Theeuwes, J.(eds) *Panel data and labor market studies.*New York: North Holland.

Malecki, E. J. (2002a) 'The economic geography of the Internet's infrastructure', *Economic Geography,* 78, (4), pp. 399-424.

Malecki, E. J. (2002b) 'Hard and soft networks for urban competitiveness', *Urban Studies,* 39, (5-6), pp. 929-45.

Malecki, E. J. (2004) 'Fibre tracks: explaining investment in fibre optic backbones', *Entrepreneurship and Regional Development,* 16, (1), pp. 21-39.

Malecki, E. J. and Gorman, S. P. (2001) 'Maybe the death of distance, but not the end of geography: the Internet as a network', in Leinbach, T. R. and Brunn, S. D.(eds) *Worlds of E-Commerce: Economic, Geographical and Social Dimensions.*West Suusex: Wiley.

Malecki, E. J. and Moriset, B. (2008) The digital economy. New York: Routledge.

Malecki, E. J. and Wei, H. (2009) 'A wired world: the evolving geography of submarine cables and the shift to Asia', *Annals of the Association of American Geograghers,* 99, (2), pp. 360-82.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate analysis.* London: Academic Press Inc.

Markusen, A. (1988) *Regions: The economics and politics of territory.* Totawa, NJ: Rowman & Littlefield.

Martin, P. (2003) 'Public policies and economic geography', in Funck, B. and Pizzatti, L.(eds) *European Integration, Regional Policy and Growth.* Washington, DC: World Bank, pp. 19-32.

Martin, R. (1999) 'Critical survey. The new 'geographical turn' in economics: some critical reflections', *Cambridge Journal of Economics,* 23, (1), pp. 65-91.

Massy, W. F. (1965) 'Principal components regression in exploratory statistical research', *Journal of the American Statistical Association,* 60, (309), pp. 234-256.

Matsumoto, H. (2007) ' International air network structures and air traffic density of world cities ', *Transportation Research Part E,* 43, pp. 269–282.

McCann, P. and Acs, Z. J. (forthcoming) 'Globalisation: countries, cities and multinationals'.

Miles, I. and Matthews, M. (1992) 'Information technology and the information economy', in Robins, K.(ed), *Understanding information.* London: Belhaven Press.

Milgram, S. (1967) 'The Small World Problem', *Psychology Today,* 2, pp. 60-67.

Mokhtarian, P. L. (1990) 'A typology of relationships between telecommunications and transportation', *Transportation Research,* 24A, (3), pp. 231-42.

Mokhtarian, P. L. (2002) 'Telecommunications and travel: The case for complementarity', *Journal of Industrial Ecology,* 6, (2), pp. 43–57.

Moriset, B. (2003) 'The new economy in the city: emergence and location factors of Internet-based companies in the metropolitan area of Lyon, France', *Urban Studies,* 40, (11), pp. 2165–86.

Moss, M. L. (1987) 'Telecommunications. world cities and urban policy', *Urban Studies,* 24, pp. 534-46.

Moss, M. L. and Mitra, S. (1998) ''Net equity: A report on income and Internet access', *Journal of Urban Technology,* 5, (3), pp. 23-32.

Moss, M. L. and Townsend, A. (1997) 'Tracking the net: Using domain names to measure the growth of the Internet in US cities', *Journal of Urban Technology,* 4, (3), pp. 47-60.

Moss, M. L. and Townsend, A. M. (2000) 'The Internet backbone and the American metropolis', *Information Society,* 16, (1), pp. 35-47.

Mumtaz, K. (2003) *Schupeter innovation and growth.* Kathmandu, Nepal: Ashgate Publishing Group.

Musgrave, R. A. and Musgrave, P. B. (1984) *Public finance in theory and practise.* New York: McGraw-Hill.

Mydral, G. (1957) *Economic theory and underdeveloped regions.* London: Duckworth.

Neef, D. (1998) 'The knowledge economy: an introduction', in Neef, D. (ed), *The knowledge economy.* Woburn, MA: Butterworth-Heinemann.

Negroponte, N. (1995) *being digital.* London: Mackays of Chatham.

Newman, M. E. J. (2005) 'Power laws, Pareto distributions and Zipf's law', *arXiv:cond-mat/0412004v3.*

Newman, M. E. J. (2008) 'Mathematics of networks', in Blume, L. E. and Durlauf, S. N.(eds) *The New Palgrave Encyclopedia of Economics.*Basingstoke: Palgrave Macmillan.

Nijkamp, P. and Jonkhoff, W. (2001) 'The city in the Information and Communication Technology age: a comparative study on path dependency', *International Journal of Technology, Policy and Management,* 1, (1), pp. 78-99.

Nijkamp, P., Rietveld, P. and Spierdijk, L. (2000) 'Classification Techniques in Quantitative Comparative Research: A Meta-Comparison', *Spatial Economic Science, Dept. of Economics, Free University, Amsterdam,* 102-124 Research Memorandum, 1999-56.

Nooy, W. d., Mrvar, A. and Batagelj, V. (2005) *Exploratory Social Network Analysis with Pajek.* Cambridge University Press.

Obraczka, K. and Silva, F. (2000) 'Network latency metrics for server proximity', *IEEE,* pp. 421-427.

OECD. (1996) *The knowledge-based economy.* Paris: OECD

OECD. (2003) *The sources of economic growth in OECD countries.* Paris: OECD

O'Kelly, M. E. and Grubesic, T. H. (2002) 'Backbone topology, access, and the commercial Internet, 1997-2000', *Environment and Planning B,* 29, (4), pp. 533-552.

Paci, R. and Usai, S. (2000) 'Technological enclaves and industrial districts: An analysis of the regional distribution of innovative activity in Europe', *Regional Studies,* 34, (2), pp. 97-114.

Paltridge, S. (2002) *Internet traffic exchange and the development of end-to-end international competition.* Paris: OECD

Parr, J. B. (2005) 'Pespectives on the city-region', *Regional Studies,* 39, (5), pp. 555-66.

Patuelli, R., Reggiani, A., Gorman, S. P., Nijkamp, P. and Bade, F.-J. (2007) 'Network analysis of commuting flows: a comparative static approach to German data', *Networks and Spatial Economics,* 7, pp. 315-331.

Pelletiere, D. and Rodrigo, C. G. (2001) 'Economic geography and policy in the network age', in Feldman, M. P. and Link, A. N.(eds) *Innovation policy in the knowledge-based economy.* Boston: Kluwer Academic Publishers.

Perez, C. (1983) 'Structural change and the assimilation of new technologies in the economic system', *Futures,* 15, (4), pp. 357-75.

Peterson, L. L. and Davie, B. S. (2003) *Computer Networks.* San Francisco: Morgan Kaufmann Publishers.

Pike, A., Rodríguez-Pose, A. and Tomaney, J. (2006) *Local and regional development.* London and New York: Routledge.

Porat, M. (1977) *The information Economy: Definition and Management.* Washington, DC: Special Publication 77-12, Office of Telecommunications, US Department of Commerce.

Priemus, H. (2007) 'The network approach: Dutch spatial planning between substratum and infrastructure networks', *European Planning Studies,* 15, (5), pp. 667-86.

Prufer, J. and Jahn, E. (2007) 'Dark clouds over the Internet?' *Telecommunications Policy,* 31, (3-4), pp. 144-154.

Quah, D. T. (1996) 'The invisible hand and the weightless economy', *Centre for Economic Performance, London School of Economics and Political Science,* Occational Paper 12.

Quah, D. T. (1998) 'A weightless economy', *The UNESCO Courier*, December 1998.

Ratti, C., Sassen, S., Mitchell, W. J., Townsend, A., Moss, M., Research, A. T. L. and Calabrese, F. (2008) *NYTE.* Cambridge, Massachusetts: SA+P Press.

Reggiani, A. and Vinciguerra, S. (2007) 'Network Connectivity Models: An Overview and Applications to the Space-Economy', in Friesz, T.(ed), *Network Science, Nonlinear Science and Infrastructure Systems.* New York: Springer-Verlag.

Richardson, R. and Gillespie, A. (2000) 'The economic development of peripheral rural areas in the information age', in Wilson, M. I. and Corey, K. E.(eds) *Information Tectonics.*Wiley.

Rimmer, P. J. (1998) 'Transport and telecommunications among world cities', in Lo, F.-C. and Yeung, Y.-m.(eds) *Globalization and the world of large cities.* Tokyo: United Nations University Press, pp. 433-70.

Roberts, D. (2006) 'Broadband gluttons', *Financial Times*, 15 April 2006.

Rodríguez-Pose, A. (2008) 'The rise of the "city-region concept and its development policy implications', *European Planning Studies,* 16, (8), pp. 1025-46.

Rodríguez-Pose, A. and Crescenzi, R. (2008) 'Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe', *Regional*

Studies, 42, (1), pp. 51-67.

Rodríguez-Pose, A. and Tselios, V. (2009a) 'Inequalities in income and education and regional economic growth in Western Europe', *Annals of Regional Science,* forthcoming.

Rodríguez-Pose, A. and Tselios, V. (2009b) 'Mapping regional personal income distribution in Western Europe: income per capita and inequality', *Czech Journal of Economics and Finance,* 5, (1), pp. 41-70.

Rogerson, P. A. (2006) *Statistical methods for geography.* London: SAGE.

Rommer, P. M. (1986) 'Increasing returns and long-run growth', *Journal of Political Economy,* 94, pp. 1002-37.

Rommer, P. M. (1990) 'Endogenous technological change', *Journal of Political Economy,* 98, (5), pp. 71-102.

Russo, G., Reggiani, A. and Nijkamp, P. (2007) 'Spatial activity and labour market patterns: A connectivity analysis of commuting flows in Germany', *Annals of Regional Science,* 41, pp.:789–811.

Rutherford, J. (2004) *A tale of two global cities.* Hants: Ashgate.

Rutherford, J. (2005) 'Networks in cities, cities in networks: Territory and globalisation intertwined in telecommunications infrastructure development in Europe', *Urban Studies,* 42, pp. 2389-2406.

Rutherford, J. (forthcoming) 'A technological urban hierarchy of Europe? Analysing inter-city telecommunications connections as measurements of world city network formation', *Urban Studies*.

Rutherford, J., Gillespie, A. and Richardson, R. (2004) 'The territoriality of Pan-European telecommunications backbone networks', *Journal of Urban Technology,* 11, (3), pp. 1-34.

Rutherford, J., Gillespie, A. and Richardson, R. (2005) *Technological connectivities of European cities? The potentials and pitfalls of the use of telecommunications data in measurements of world city network formation*. Available at: http://www.lboro.ac.uk/gawc/rb/rb181.html (Accessed: 8/2/2009).

Salomon, I. (1986) 'Telecommunications and travel relationships: A review', *Transportation Research,* 20A, (3), pp. 223–238.

Sassen, S. (1991) *The global city. New York, London, Tokyo.* Princeton, New Jersey: Princeton University Press.

Sassen, S. (2000) *Cities in a world economy.* California: Pine Forge Press.

Sassen, S. (2004) *The global city. New York, London, Tokyo.* Princeton, New Jersey: Princeton University Press.

Sassen, S. (2008) 'New York City's Two Global Geographies of Talk', in Ratti, C.(ed), *NYTE.*New York: MIT, *senseable* lab.

Schintler, L., Gorman, S. P., Reggiani, A., Patuelli, R., Gillespie, A., Nijkamp, P. and Rutherford, J. (2004) 'Complex network phenomena in telecommunication Systems', *Networks and Spatial Economics,* 5, pp. 351-370.

Scott, A. J. (1998) *Regions and the world economy: the coming shape of global production, competition and political order.* Oxford: Oxford University Press.

Scott, A. J. (2001) 'Globalization and the rise of city-regions', *European Planning Studies*, 9, (7), pp. 813-26.

Scott, A. J. and Storper, M. (2003) 'Regions, globalization, development', *Regional Studies*, 37, (6-7), pp. 579-93.

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G. and Manna, S. S. (2003) 'Small-world properties of the Indian railway network', *Physical Review E*, 67.

Shiller, D. (1999) *Digital Capitalism: networking the global market system.* Cambridge, MA: MIT Press.

Shinjo, K. and Zhang, X. (2004) 'ICT capital investment and productivity growth: Granger causality in Japanese and the USA industries', *Working Paper*.

Shiu, A. and Lam, P.-L. (2008) 'Causal relationship between telecommunications and economic growth in China and its Regions', *Regional Studies*, 42, (5), pp. 705-18.

Short, J., Kim, Y., Kuus, M. and Wells, H. (1996) 'The dirty little secret of world cities research: Data problems in comparative analysis', *International Journal of Urban and Regional Research,* 20, (4), pp. 697-719.

Shy, O. (2001) *The economics of network industries.* Cambridge: Cambridge University Press.

Smith, D. and Timberlake, M. (2002) 'Hierarchies of dominance among world cities: a network approach', in Sassen, S.(ed), *Global networks; linked cities.* New York. London: Routledge.

Solow, R. M. (1956) 'A contribution to the theory of economic growth', *Quarterly*

Journal of Economics, 70, pp. 65-94.

Star, S. L. (1999) 'The ethnography of infrastructure', *American Behavioral Scientists*, 43, (3), pp. 377-91.

Stephenson, N. (1996) 'Mother earth mother board', *Wired Magazine*, 4.12, [Online]. Available at: http://www.wired.com/wired/archive/4.12/ffglass.html (Accessed: 27/8/2009).

Stevens, S. (1998) 'The knowledge-driven economy', in Neef, D.(ed), *The knowledge economy.* Woburn, MA: Butterworth-Heinemann.

Sufian, A. J. M. (2005) 'Analyzing collinear data by principal component regression approach — An Example from Developing Countries', *Journal of Data Science*, 3, pp. 221-32.

Taaffe, E. J., Gauthier, E. L. and O'Kelly, M. E. (1996) *Geography of Transportation.* New Jersey: Prentice-Hall, Inc.

Tanenbaum, A. S. (2003) *Computer Networks.* Prentice Hall: Pearson Education.

Tassey, G. (1992) *Technology infrastructure and competitive position.* Boston, MA: Kluwer Academic Press.

Tassey, G. (2008) 'Modeling and measuring the economic roles of the technology infrastructure', *Economics of Innovation and New Technology,* 17, (7&8), pp. 615-29.

Taylor, P. (1997) 'Hierarchical tendencies amongst world cities: a global research proposal', *Cities*, 14, (6), pp. 323-32.

Taylor, P. (1999) 'So-called "World Cities": The Evidential Structure within a Literature', *Environment and Planning A*, 31, (11), pp. 1901-4.

Taylor, P. (2004) *World city network.* London and New York: Routledge.

Telegeography (2009) *Telegeography's home page*. Available at: http://www.telegeography.com (Accessed: 28/5/2009).

Telegeography. (2007) *Global Internet Geography.* Washington, D.C.: PriMetrica, Inc.

Tobbler, W. (1970) 'A computer movie simulating urban growth in the Detroit region', *Journal of Economic Geography,* 46, (2), pp. 234-40.

Toffler, A. (1980) *Third way.* New York: William Morrow.

Townsend, A. (2003) *Wired / Unwired: The Urban Geography of Digital Networks.* Thesis. MIT.

Townsend, A. M. (2001a) 'Network cities and the global structure of the Internet', *American Behavioral Scientist*, 44, (10), pp. 1697-1716.

Townsend, A. M. (2001b) 'The Internet and the rise of the new network cities, 1969-1999', *Environment and Planning B*, 28, pp. 39-58.

Tranos, E. and Gillespie, A. (2009) 'The spatial distribution of Internet backbone networks in Europe: a metropolitan knowledge economy perspective', *Journal of European Urban and Regional Studies*, 16, (4), 423-37.

Triplett, J. E. (1998) *The Solow productivity paradox: what computers do to productivity?* Washington, DC: Brookings Institution

UN (2006) *Information Economy Report.* New York and Geneva: UN

US Department of Commerce. (1996) *National Telecommunications and Information Administration, Institute for Telecommunications Services*. Federal Standard 1037C.

van Oort, F., Raspe, O. and Snellen, D. (2003) *De ruimtelijke effecten van ICT [The spatial impacts of ICT]*. Rotterdam: NAi Uitgevers.

van Winden, W. and Woets, P. (2004) 'Urban broadband Internet policies in Europe: A critical review', *Urban Studies*, 41, (10), pp. 2043-59.

Venables, A. J. (2001) 'Geography and international inequalities: the impact of new technologies', *Journal of Industry, Competition and Trade*, 1, (2), pp. 1566-79.

Wallerstein, I. (2004) *World-Systems Analysis.* Durham, NC: Duke University Press.

Walsh, J. A. (2007) *People and place. A census of the Republic of Ireland.* Maynooth: National Institute for Regional and Spatial Analysis.

Warf, B. (2001) 'Segueways into cyberspace: multiple geographies of the digital divide', *Environment and Planning B*, 28, (1), pp. 1-11.

Watts, D. J. and Strogatz, S. H. (1998) 'Collective dynamics of small-world networks', *Nature*, 393, pp. 440-442.

Wheeler, D. C. and O'Kelly, M. E. (1999) 'Network topology and city accessibility of the commercial internet', *Professional Geographer*, 51, (3), pp. 327-339.

White, P. (2003) 'Making use of secondary data', in Cliffor, N. J. and Valentine, G.(eds) *Key methods in geography*. London: SAGE.

Wolde-Rufael, Y. (2007) 'Another look at the relationship between telecommunications investment and economic activity in the United States', *International Economic Journal*, 21, (2), pp. 199-205.

Wooldridge, J. M. (2003) *introductory econometrics: a modern approach.* Mason, Ohio: South-Western.

Yoo, S. H. and Kwak, S. J. (2004) 'Information technology and economic development in Korea: a causality study', *International Journal of Technology Management*, 27, (1), pp. 57-67.

Youtie, J. (2000) 'Field of dreams revisited: economic development and telecommunications in LaGrange, Georgia', *Economic Development Quarterly,* 14, pp. 146–153.

Zook, M. A. (2000) 'The Web of Production: The Economic Geography of Commercial Internet Content Production in the United States.' *Environment and Planning A*, 32, pp. 411-26.

Zook, M. A. (2001) 'Old hierarchies or new networks of centrality? – The Global Geography of the Internet Content Market', *American Behavioral Scientist,* 44, (10), pp. 1679-96.

Zook, M. A. (2006) 'The geographies of the Internet', *Annual Review of Information Science and Technology,* 40, pp. 53-78.