TI 2004-118/3
Tinbergen Institute Discussion Paper

Complex Network Phenomena in Telecommunication Systems

*Laurie A. Schintler**
*Sean P. Gorman**
*Aura Reggiani***
*Roberto Patuelli**
*Andy Gillespie****
*Peter Nijkamp*****
*Jonathan Rutherford****

** The School of Public Policy, George Mason University, US,*
*** Department of Economics, University of Bologna, Italy,*
**** Centre for Urban and Regional Development Studies, University of Newcastle, Newcastle upon Tyne, UK,*
*****Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam, and Tinbergen Institute.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Amsterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

COMPLEX NETWORK PHENOMENA IN TELECOMMUNICATION SYSTEMS

Laurie A. Schintler[*], Sean P. Gorman[*], Aura Reggiani[**], Roberto Patuelli[*], Andy Gillespie[***], Peter Nijkamp[****] and Jonathan Rutherford[***]

PN170LSSGARRPAGJR

[*] The School of Public Policy, George Mason University, US; email: lschintl@gmu.edu; sgorman1@gmu.edu, rpatuel1@gmu.edu

[**] Department of Economics, University of Bologna, Italy; email: reggiani@economia.unibo.it

[***] Centre for Urban and Regional Development Studies, University of Newcastle, Newcastle upon Tyne NE1 7RU, United Kingdom; email: Andy.Gillespie@newcastle.ac.uk; J.A.Rutherford@newcastle.ac.uk

[***] Department of Spatial Economic, Free University of Amsterdam, The Netherlands; email: pnijkamp@feweb.vu.nl

ABSTRACT

Many networks such as the Internet have been found to possess scale-free and small-world network properties reflected by so-called power law distributions. Scale-free properties evolve in large complex networks through self-organizing processes and more specifically, preferential attachment. New nodes in a network tend to attach themselves to other vertices that are already well-connected. Because traffic is routed mainly through a few highly connected and concentrated vertices, the diameter of the network is small in comparison to other network structures, and movement through the network is therefore efficient. At the same time, this efficiency feature puts scale-free networks at risk for becoming disconnected or significantly disrupted when super-connected nodes are removed either unintentionally or through a targeted attack or external force.

The present paper will examine and compare properties of telecommunications networks for both the United States and Europe. Both types of networks will be examined in terms of their network topology and specifically whether or not they are scale-free networks to be further explored by identifying and plotting power law distributions. Next, economic, political and cultural factors may be used to explain differences in network structures between the United States and Europe. In addition, the paper will identify data and modeling tools that are needed to facilitate further cross-Atlantic comparative studies of communications networks.

1. Introduction

Network connectivity has increasingly become a major research issue. With the advent of the World Wide Web and the Internet, complex network formations have emerged that call for a thorough functional and geographic analysis. In recent years several contributions have emerged at the intersection of geography and network analysis (see inter alia Barabasi and Albert, 1999; Albert and Barabasi, 2000; Amaral et al., 2000; Gorman and Kulkarni, 2000; Malecki and Gorman, 2001; Rinaldi et al., 2001; Malecki, 2002; Barthelemy, 2003; Gorman and Kulkarni, 2004). Such studies address the position of nodes (e.g., cities, hubs) as well as the diffusion patterns of flows in complex networks, with due emphasis on route length, nodal clustering, and power law and exponential connectivity distributions. Most complex networks appeared not to have a random formations, but a locally organized structure of nodes or clusters, leading to 'small worlds' (cf. Watts and Strogatz, 1998). Such 'small worlds' in complex networks tend to have a heavy tailed connectivity distribution (with an inverse relationship between the number of nodes and the number of connecting links), in contrast to a random distributions (with an exponentially declining probability to find highly connected nodes). In the case of incremental growth in a complex network, new nodes tend to be more likely to connect with existing well-linked nodes. Consequently, hubs tend to reinforce themselves. Such networks are often coined 'scale-free networks'.

This paper compares the distribution and connectivity properties of two telecommunications networks, the physical fiber network of North America and the pan-European fiber optic network. There are five sections to the paper following the introduction. The first provides a historical perspective of telecommunications deployment in each region along with a description of the regulatory and institutional frameworks that have contributed to these trends. Section three discusses complex network theory and the methodology used in this study to compare the United States and Europe in terms of network connectivity. The results of the empirical investigations are presented in Section four and concluding remarks and directions for future research in Section five.

2. Europe and the United States: Comparing Development and Deregulation Policies

2.1. Preface

The nations of Europe, on average, lag the United States in terms of telecommunications deployment and market development, although this is changing, as Europe transitions to a more competitive

regime. This section provides a historical perspective of these trends focusing on how the institutional and regulatory frameworks have evolved in the two regions, establishing a base from which the findings of the empirical analysis can be interpreted.

2.2. Telecommunications Networks in the United States and the Role of Deregulation

The telecommunications revolution was initiated by deregulation of the telecommunications industry, which turned state-subsidized monopolies into private market competitors. Private firms, even before regulation, had supplied advanced telecommunication services, but deregulation has moved the vast majority of the world's telecommunication networks into private control (OECD, 2002). Privately operated networks in a competitive environment will respond mainly to market pressures of supply and demand (Gillespie and Robins, 1989). The largest network of interconnected privately operated networks is the Internet, and, as a result, it is a largely unregulated system into which corporate networks have connected (Schiller, 1999). Crandall (1997) argues that deregulation is a key ingredient for the successful growth of advanced telecommunication services like the Internet, "*the case for government support and direction of telecommunication infrastructure investment remains very weak*" (Crandall, 1997, p.161). In fact, the Federal Communication Commission's (FCC) official status for the Internet is as an unregulated network (Oxman, 1999).

The exception to Internet's lack of regulation is "last mile" access that provides services like broadband to homes. "Last mile" infrastructure is predominantly owned by regional monopolies, either the regional Bell telephone companies or regional cable television provider. In some areas of the United States the regional monopolies have been forced to share or unbundle their network infrastructure to allow competitors to offer broadband service to the home. The current policy debate over whether to completely open these markets to unfettered competition (i.e. open access) (Bittlingmayer and Hazlett, 2002), or limit broadband provision to only the regional monopolies, focusing regulation on encouraging monopolies to provide access to underrepresented areas (Kalakota et al., 2002), continues. The debate has taken the form of legislative actions, like the proposed Tauzin-Dingall Bill, which would limit broadband provision to only the regional monopolies. Some have prognosticated that wireless local loop solution could be the solution to the "last mile" dilemma, but this too depends on public policy actions[1] that make electro-magnetic spectrum available for new technologies to broadcast over (Economist, 2003).

[1] More radical policy strategies call for the deregulation of spectrum allocation where spectrum is treated as property, allowing the owner to use their frequency for whatever application they see best fit, or most profitable (Economist, 2003).

In the current deregulated market, where demand is greatest, telecommunications will be supplied, resulting in geographic biases. The principal bias is locations with agglomerations of population and economic activity will be disproportionately supplied telecommunications services and infrastructure (Gasper and Glaeser, 1998; Moss, 1998; Malecki, 2002). This suggests that communication innovations propagate a core-periphery effect of IT, where a more refined spatial division of "*information labour*" has developed (Hepworth, 1987, p. 157). The agglomeration of demand and skilled labor is the basis of the opposing argument: that we will not see communication technologies causing the "end of geography" or the decentralization of economic activity. The fundamental question separating the two arguments is whether the core (large urban areas) or the periphery (rural and small urban areas) will grow and develop as a result of the telecommunications and IT revolution, and at what rate each will grow.

2.3. Telecommunications Networks in Europe and the Liberalization

The picture of telecommunications networks in Europe is rather different from the one depicted for the United States. Although now developing at a higher speed, Europe is still several steps behind the US in the development of advanced systems of telecommunications infrastructures and services. The deployment of physical infrastructure in Europe has been slowed down by several factors, such as institutional frameworks determined by intra- and inter-country regulations, in contrast to the United States, where an earlier deregulation in 1996 allowed for faster development, which was also pushed by rapidly rising demand.

Europe is now beginning to catch up. Just between July 2003 and January 2004, broadband users grew by 32%, reaching a total of 23.2 million users. The overall growth since the month of July 2002 was 152% (Balbi, 2004). By 2007, 62 million broadband users are expected to be located in Europe and 40% of households in Europe – in particular Western Europe – are estimated to be equipped with high-speed connections (INA, 2004). Such impressive figures will require extensive investments for the deployment of capable, sophisticated and integrated networks all over Europe. Recently, the development of these infrastructures advanced at different paces across countries in Europe. Table 1 shows the recent development of Internet hosts and users in European countries between 2000 and 2001, as long as the values of several economic, social and technological variables. Over the short period of a year, Internet users and hosts increased in most of the listed countries, often at rates between +50 and +100%. While the values of the Education Index (see UNDP, 2002 for details) seem

to suggest that the most educated countries enjoy higher levels of Internet penetration, variables such as per capita GDP do not show a clear common pattern if compared with the Internet-related variables.

If one considers the European network in terms of Internet infrastructure, one can observe that there are clear differences between Europe and the US in terms of Internet network infrastructure. Several studies have been carried out on this phenomenon. For example, Gorman (2002) has calculated the alpha and gamma indices, among others, for Europe and the United States, to measure the network's redundancy and interconnectivity, respectively[2]. Gorman's findings show a 51% redundancy for the United States in 2000, while Europe only had a 28% redundancy rate. It is important to underline, though, how Europe has progressed, by improving its index from 3% in 1998 to 28% in 2000. Europe shows a lower value for the interconnectivity gamma index compared to the United States. The gamma index value for Europe is 26.2%, while the US score a 55%. Since the gamma index has previously been shown to be correlated to economic development,[3] the low value observed for Europe has to be explained by complementary or different factors. A major factor in this setting is regulation, or, perhaps more precisely, lack of deregulation (Gorman, 2002).

In the framework of the 1998 liberalization of the telecommunications industry, the European Commission laid down a set of key requirements for the operators to meet and the National Regulatory Authorities (NRAs) to enforce. These requirements are related in particular to:

- control of retail prices;
- control of access prices;
- Universal Service Obligation (USO).

[2] The alpha index is calculated as the ratio between the number of loops observed over a network and the maximum possible number of loops in the given network. The gamma index is instead computed as the ratio between observed and theoretical maximum number of nodes belonging to the given network. See, between others, Haggett and Chorley (1969) for more details on network indices.

[3] Among others, Kansky (1963) and Garrison (1968) demonstrated the correlation between economic development and the gamma index.

Table 1 – Social, economic and technological variables for selected European Countries.

List of Countries	*Total Population (M) 2001*	*Online Population 1999–2002*	*Internet Hosts per 10k inhab. 2001*	*Internet Hosts per 10k inhab. 2000*	*Internet Users per 10k inhab. 2001*	*Internet Users per 10k inhab. 2000*
Austria	-	3550000	400.51	595.01	3194.1	2585.89
Denmark	5.37	3230000	1045.38	626.6	4471.77	3658.52
Estonia	1.43	429700	356.92	284.25	3004.59	2720.96
Finland	5.20	2690000	1707.25	1022.53	4302.83	3722.95
Germany	82.36	30200000	294.58	248.05	3642.54	2917.6
Italy	58.02	19250000	117.28	177.97	2757.76	2303.75
Netherlands	16.1	9280000	1634.77	1015.55	3291.72	2439.47
Norway	4.53	2460000	673.82	1009.31	5962.9	4905.24
Portugal	10.3	3600000	239.28	62	3494.13	2494.11
Sweden	8.91	5740000	825.14	670.79	5162.74	4558.29
Switzerland	7.22	3410000	730.74	364.39	4040.17	2962.22
United Kingdoms	60.08	34000000	371.37	280.75	3995.01	3011.75
Average	23.59	9819975	699.75	529.77	3943.36	3190.06
Median	8.91	3575000	537.17	479.70	3818.78	2939.91
United States	285.93	165750000	3714.01	2928.32	4995.10	4506.96

	Total Internet Bandwidth (Mbps) 2002	*Technology Achievement Index 1999*	*GDP per capita (PPPUS$) 2000*	*R&D Expenditure as % of GNP 1990–2000*	*Human Development Index (HDI) Value 2000*	*Education Index 2000*
Austria	17729	0.544	26765	1.6	0.926	0.96
Denmark	43456	-	27627	1.9	0.926	0.98
Estonia	517	-	10066	0.8	0.826	0.95
Finland	7820	0.744	24996	-	0.930	0.99
Germany	207669	0.583	25103	2.3	0.925	0.97
Italy	35771	0.471	23626	1.0	0.913	0.94
Netherlands	173154	0.630	25657	2.0	0.935	0.99
Norway	21638	0.579	29918	1.7	0.942	0.98
Portugal	3522	0.419	17290	0.6	0.880	0.94
Sweden	60349	0.703	24277	3.8	0.941	0.99
Switzerland	40057	-	28769	2.6	0.928	0.94
United Kingdoms	238074	0.606	23509	1.8	0.928	0.99
Average	70813	0.587	23967	1.8	0.917	0.97
Median	37914	0.583	25050	1.8	0.927	0.98
United States	273770	0.733	34142	2.5	0.939	0.98

Sources: ITU, NUA, Packet Geography, UNDP.

The second and third requirements are of particular importance in the context of our research. Control over access prices is, in fact, to a certain degree, a regulation of the interconnection market. The prices for leased lines contribute to determining the willingness, by an incoming operator, to invest capital in infrastructure or rent lines. Leasing lines in Europe is still much more expensive than in the United States, although prices are slowly moving in agreement with international rates.[4] This effect may be due to the emerging competition in recent years. Perhaps the 1998 liberalization contributed to this trend, which was the intention of the European policy-makers. In addition to interconnection and interoperability conditions, the development of telecommunications networks is influenced by factors such as the possibility of building up alternative technology platforms and acquiring infrastructures and essential facilities (Cave and Crandall, 2001).

In addition, the requirement regarding USO consists of the obligation to serve every customer at the same price.[5] This requirement influences the ability of operators to offer lower prices to preferred/more convenient areas or customers.

Some have suggested that price regulation should be assigned to a centralized European telecommunications authority. This issue has been discussed since the early 90's, but the creation of such an institution would require a loss of powers by the NRAs, which are, in many countries, under control of the government. Most of all, a change in the European Treaty would be needed, which would then require ratification, often via a referendum, by all the member countries (Bergman et al., 1998; Cave and Crandall, 2001). On the other hand, a central authority would be able to deal with conflicts between NRAs and European regulators, and would complement EU policies.[6] This trend goes in the opposite direction of the United States, where some discussions have been going on about a possible elimination of the United States' FCC.

Despite the issues discussed above, the liberalization process was, until now, surprisingly successful, as progress has been made in several fields. A number of telecommunications operators have, in fact, been privatized, and access barriers for the telecom market have been eliminated, by liberalizing entry. Prices have also dropped, in particular, beginning in 1997. On the other hand, the deployment of infrastructure is still limited and tends to agglomerate near important business centers,

[4] Cave and Crandall (2001) hypothesize the existence of a community of interest may have delayed the decrease in line leasing prices. Such a behavior might have been possible by the lack of deregulation in the matter and the financial barriers for the entry in the market.

[5] In detail, the USO requirement consists of 3 dimensions: a) universal geographical coverage; b) geographical averaging of tariffs; c) basic telecom service provided at affordable price (Cave and Crandall, 2001).

e.g., Frankfurt or Brussels. An important policy concern are long-term commitments have to be favored, in order to facilitate the deployment of new broadband infrastructures. Moreover, Bergman et al. (1998) stress how the promotion of innovation is tied to such long-term commitments. The financial needs associated with investments in infrastructures are one of the reasons for the trend directed towards the agglomeration of firms in strategic or financial alliances (see, e.g., Allen, 1996). In fact, because of the large volume of customers required to make the connection to a new network profitable, large and global operators "*are the main target group for direct fibre optic connections*" (Pehrsson, 2001, p. 203). In an analysis of the United Kingdom's and Sweden's telecommunications markets, Pehrsson notes, in fact, the perception of deregulation and the capital needed for investments in infrastructure are important barriers to the entrance of new operators.

Finally, in summary the above discussion recalls that the development of a competitive telecommunications market is a strategic landmark in the growth of the networks. In this context, the concept of scale-free network came recently to the fore, by showing the potential of long-term correlations, expressed by power law distributions (Reggiani and Schintler, 2005). The emergence of scale-free phenomena in telecommunication networks might be, therefore, tied to the efficiency of the liberalization process and its enforcement.

The next section will briefly describe the scale-free properties, which essentially evolve in large complex networks through self-organized processes and, more specifically, preferential attachment. The subsequent Section 4 will then illustrate the empirical findings emerging from the scale-free approach, with reference to the analysis of US and European telecommunications data. The potential factors that might have caused networks to grow following different patterns and trends will be outlined in this framework.

3. Scale-Free Networks and Power Law Distributions

For the majority of the 20th century the absence of detailed topological data for complex networks left random network models (Erdos and Renyi) as the most widely used method of network simulation (Barabasi, 2001). As computing power increased and real world network data began to become available, several empirical findings emerged. Three network characteristics frequently resulted from the analysis of complex networks (Albert and Barabasi, 2002, pp. 48–49):

[6] Alternative approaches have been proposed, such as a self-regulation regime to be set between the NRAs, or, as a middle-of-the-way approach, a two-tier regulation, in which the European Commission, the NRAs, telecommunications operators

1. Short average path length
2. High level of clustering
3. Power law and exponential degree distributions.

Short average path length indicates that the distance between any two nodes on the network is short; they can be reached in a few number of hops along edges. Clustering occurs when nodes locate topologically close to each other in cliques that are well connected to each other. Lastly, the frequency distributions of node connectivity, called degrees, often follow power laws or exponential distributions.

In a power law distribution there is an abundance of nodes with only a few links, and a small but significant minority that have a very large number of links (Barabasi, 2002). It should be noted that this is distinctly different from the models developed by Erdos and Renyi (1960) (ER) or Watts and Strogatz (1998) (WS); the probability of finding a highly connected vertex in the ER and WS models decreases exponentially, so that "*vertices with high connectivity are practically absent*"[7] (Barabasi and Albert, 1999, p. 510). The reason, according to Barabasi and Albert, was that their model added another perspective to complex networks, incorporating network growth; the number of nodes does not stay constant as in the WS and ER model. The Barabasi-Albert (BA) models added growth over time and the idea that new vertices attach preferentially to already well-connected vertices in the network.

Barabasi and Albert (1999) formalized this idea in "Emergence of Scaling in Random Networks". They stated that in a complex network like the World Wide Web the probability P(k) that a vertex in the network interacts with k other vertices decays as a power law following $P(k) \sim k^{-\gamma}$ where the power law exponent is equal to three. When studying real world scale-free networks, empirical results have ranged from 2.1 to 4 (Barabasi and Albert, 1999). While the model set up by Barabasi and Albert produces an exponent of three, they demonstrate how the model can be altered to produce results other than three for different network conditions. The BA model is based on three mechanisms that drive the evolution of graph structures over time to produce power law relationships (Chen et al., 2001, p. 5):

- Incremental growth – Incremental growth follows from the observation that most networks develop over time by adding new nodes and new links to an existing graph structure.

---

and manufacturing firms, consumer unions and other relevant agents are involved (Bergman et al., 1998).

[7] Barabasi and Albert's definition of high connectivity is relative to the number of nodes in the network, and in this context, it simply means a large proportion on the total connections in the network. The odds of a node having a large proportion on connections in a network are small enough that they are likely to be "practically absent".

- Preferential connectivity – Preferential connectivity expresses the frequently encountered phenomenon that there is a higher probability for a new or existing node to connect or reconnect to a node that already has a large number of links (i.e. high vertex degree) than there is to (re)connect to a low degree vertex.
- Re-wiring – Re-wiring allows for some additional flexibility in the formation of networks by removing links connected to certain nodes and replacing them with new links in a way that effectively amounts to a local type of re-shuffling connection based on preferential attachment.

In many ways, hub-and-spoke[8] airline networks serve as a good practical example of a scale free network. Hub-and-spoke networks show preferential connectivity to the major airports (hubs) from the minor airports, while the major hubs are fully interconnected. The major hub airports are small in number, yet they have the vast majority of the connections in the global airline network. From a technical perspective, though, airline network connectivity distributions do not quite fit a power law; they have an exponential cut off, because there is a physical constraint of the number of planes an airport can handle, preventing a complete power law fit. (Amaral et al., 2000). A large literature is now available on hub-spoke network. See, for example, Oum et al., 1995; Wojahn, 2001; Bowen, 2002.

There have, though, been other opinions on how complex networks should be classified; Amaral et al.(2000) argue that scale-free networks are a sub-class of small world networks. An exact delineation of where small world and scale-free networks diverge is still somewhat fuzzy in the literature, but the area of study is still evolving. It can be safely said that the two are inter-related and that generally speaking scale-free networks exhibit the clustering and short average path length of small world networks, but not all small world networks exhibit the power law distribution of scale-free networks.

The next section investigates the emergence of scale-free characteristics in the US and European telecommunications networks, by looking at the distribution of fiber optic connections between cities of the data set.

[8] In hub-and-spoke networks, traffic is directed towards few major hubs or transfer facilities, which serve as interconnections between more peripheral points of the network. Therefore, traffic does not move anymore directly from origin to destination. This type of network aims to "minimize the transportation cost plus the cost of hub operation." (Ballou, 2004, p. 595).

4. The Empirical Applications to US and European Communication Networks

4.1. Preface

The United States and Europe differ quite considerably in terms of the connectivity of their telecommunications networks – i.e., as seen in previous studies of their comparative graph theoretical indices of connectivity (Gorman 2002). This section presents the results of the analysis conducted to examine each regions network properties, including a cluster analysis to group cities within each region in terms of connectivity properties. Section 4.4 briefly discusses some of the cultural, economic, political and technological factors that have likely contributed to these differences, although a more rigorous examine of this is warranted.

4.2. Empirical Findings for the European Networks

This section describes the results of our analysis of European telecommunications networks. It will show that the European data mirror two different types of statistical distributions, viz. an exponential and a power law distribution, and verify how the statistical fit changes when splitting our data in several clusters.

The data set used in the analysis is fiber optic cable deployed and owned by pan-European networks in 209 cities, distributed over 25 countries. A summary of the data set can be found in Annex A, Table A1. Only networks operating in more than one country are considered here, therefore excluding domestic networks. Only international links are considered and no redundancy within networks is assumed in this case.[9] The data set dates back to the third quarter of 2001, thus it is fairly current. But unfortunately, it was not possible to capture the latest European trends that we outlined above. In particular, the significant increases in the demand of broadband connections (a 152% increase observed between July 2002 and January 2004) may have prompted new investments and new characteristics of the European telecommunications networks. On the other hand, considerable over-capacity emerged in 2000/2001, as a result of over-speculation in network-building. Consequently, much of the recent broadband-related growth in demand will have been using fibre that had been installed up to 2001, but that had not necessarily been lit, due to the overcapacity problem. It is, therefore, clear that more recent data are a main need in the analysis of telecommunications networks.

[9] No information on interconnection is available from this data set. A ranking of the most connected cities in terms of redundant links shows a mostly similar set of cities, if compared to the top ranks of physical connections (see Rutherford et al., 2004, Table 3).

It is noteworthy that two variables are available in our data set (represented in Figure 1). The *networks* variable refers to the number of pan-European networks present in each city of the data set, while the *connections* variable is the number of other European cities linked to each city in the data set via the network. For example, the city of Hamburg has links (*connections*) to 200 different European cities. These links are provided by 21 *networks* that are in place.

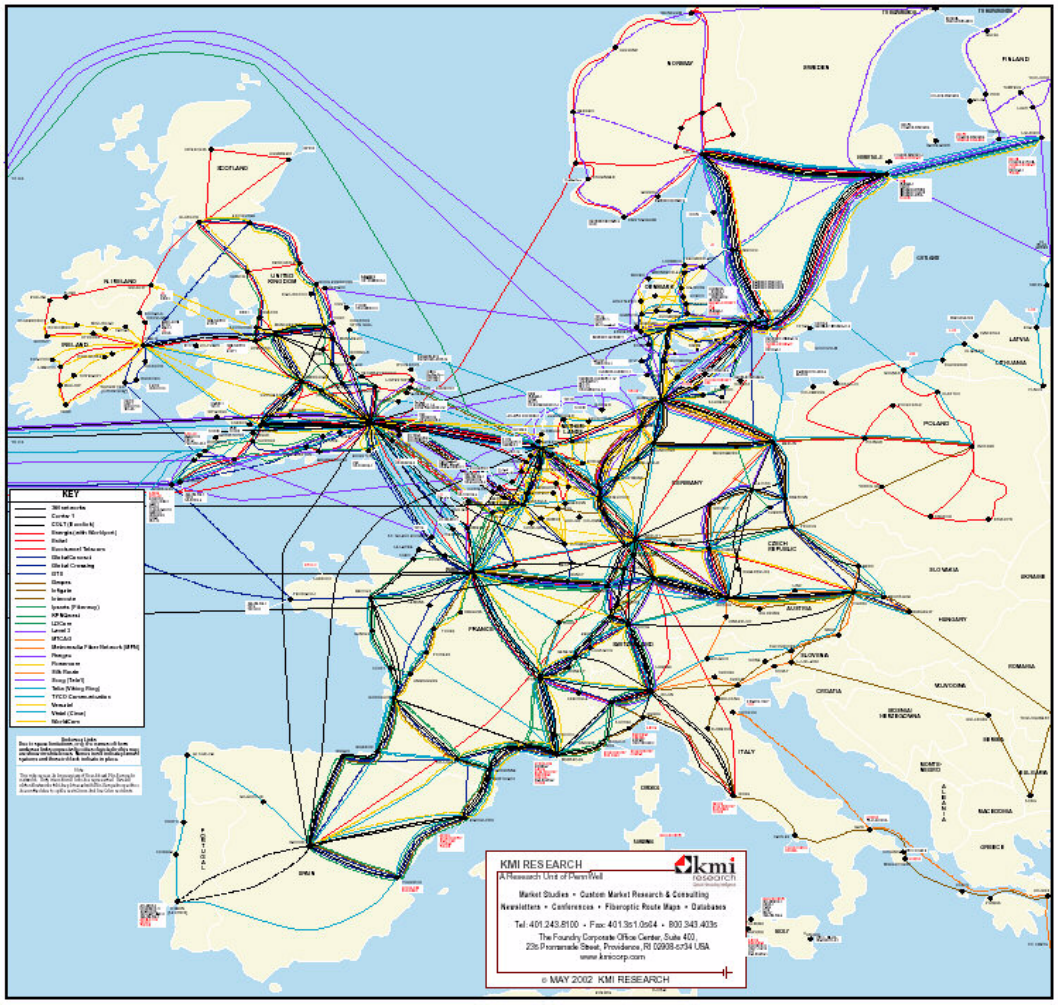

Figure 1 – Pan-European fiber optic networks (planned or in place)

Source: Rutherford et al. (2004, Figure 1). Copyright of map is with KMI Research, used by permission.

After having ranked the data set by connection, we can observe that the decay in the number of connections is gradual, in particular after the first major cities of the list. The most connected cities in the data set are Hamburg and, second, London, while the least connected cities are five cities served by only five connections each. Moreover, in a rank order of the data set of cities computed by number of networks, Hamburg and London are still at the top of the list. If we compute instead subtotals for each

of the 25 countries, the United Kingdom is at the top, followed by France and Germany. These three nations have an overall number of connections that is much higher than every other country in Europe. Considering the number of connections per capita, we can note instead that relatively small countries, such as Luxembourg and Denmark, are at the top of the list, either because of their relevant geographical position or because of higher diffusion of telecommunication technologies.

The presence in the top ranks of cities that are not considered to be between the most important in Europe – see for example Hamburg, which is topping the list – is due to the role that these cities acquired. They are, in fact, important gateways for providing large amounts of bandwidth in Europe. Some cities, like Prague or Copenhagen, also act as the main "door" through which more peripheral regions are connected and provided with bandwidth (Rutherford et al., 2004).

Once the data set has been ranked according to one variable, the aim is to analyze how the variable is distributed over the data set. Figure 2 shows, on a log-log scale, the distribution of the number of connections over the European telecommunications network. The graph shows that the data mostly fit an exponential distribution, resulting in a R^2 value of 0.73. If we investigate the existence, in our data set, of a power law distribution, then, again on a log-log scale, there is a straight line. The R^2 referring to a power law distribution for the connection variable is equal to 0.43, while the exponent for the resulting function is 0.54. In addition, the distribution of the *networks* variable shows an exponential distribution as well, reaching a R^2 value of 0.97.

The results illustrated above show a network that seems to be missing scale-free properties, which are associated with a power law distribution (see, e.g., Albert and Barabasi, 2002). A slightly different twist is given to these results by splitting our dataset in separate clusters. We have, therefore, computed, according to the values of the two available variables, five clusters out of the complete data set. A hierarchical-associative method was used in order to compute the clusters, namely the one of the most remote neighbor (or complete linkage) (see, e.g., Anderberg, 1973; Hatigan, 1975). The furthest neighbor computations result in more homogeneous clusters, which are computed by means of aggregation. The algorithm iteratively computes the Euclidean distance between the furthest members of two clusters, which are initially composed by single cities, and aggregates the two clusters minimizing this measure.[10] The iterations stop when the ideal number of clusters is found. This

[10] The Euclidean distance is computed as:

$$d_{ij} = \left\{ \sum_{k=1}^{2} (x_{ik} - x_{jk})^2 \right\}^{1/2},$$

where i,j identify two clusters and $i \neq j$, and k is the number of variables, which, in our case, are networks and connections.

corresponds to the maximization of the objective function, which is the incremental distance computed between the clusters after each aggregation.

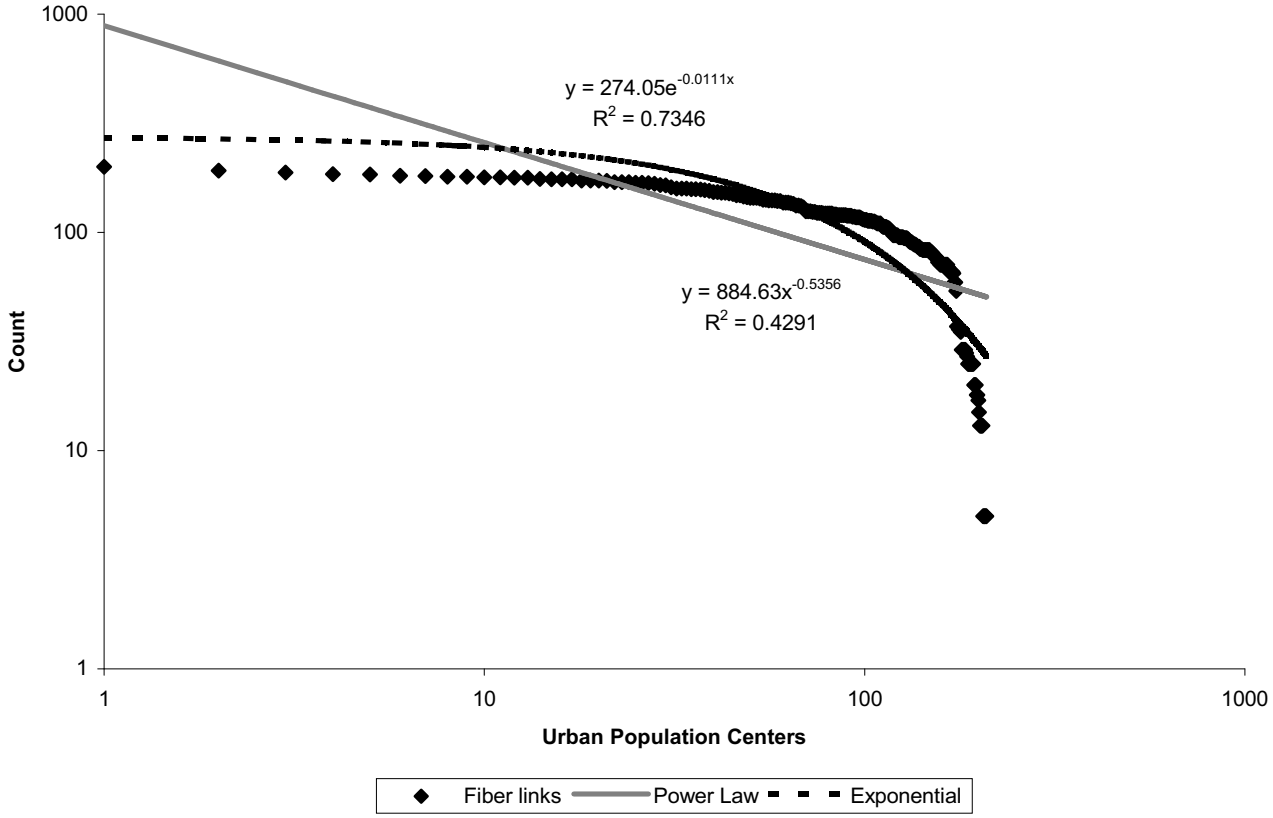

Figure 2 – The distribution of connections in European cities.

Table 2 presents the results obtained for the five clusters and shows the values of the coefficients for both an exponential distribution and a power law distribution fitting the data. The first cluster observed, containing the 21 best-connected cities, better fits a power law distribution than an exponential distribution. The value of the power law exponent is equal to 0.99 and therefore much higher than the one for the complete data set (0.43). A map of the cities belonging to the first cluster is presented in Annex A, Figure A1. It is interesting to note that the 21 cities belonging to the first cluster also correspond, aside for only 4, to the top European cities for estimated total bandwidth of backbone connectors (Rutherford et al., 2004, reported in Annex A). The cities excluded by the first clusters are, anyway, between the first cities of the second cluster. The two lists seem, therefore, to be consistent to a certain degree.

In contrast to the first cluster, the remaining four clusters confirm a better fit to exponential distributions. One objective here is to understand why the most connected cities show results that are different from the overall ones. Because of a reduced number of cases analyzed, it is clear that any judgment on this cluster should be given with caution, but one is still able to give an interpretation of this result. As previously explained in the description of the European telecommunications networks, infrastructure development often concentrated in the proximity of important business centers, which would allow for larger economies of scale for entering or reinvesting operators. The results observed for the first cluster of data seem to suggest that new adoption patterns and development are taking place – and are starting, as expected – from the bigger metropolitan centers, which are the nearest group to scale-free behavior. The evolution of this cluster over time would, with the presence of updated data, allow for a confirmation of this centrifugality hypothesis.

Table 2 – The exponent and coefficient values for the exponential and power law functions fitting the number of connections in European telecommunications networks.

Number of Clusters	*Number of Nodes*	*Exponential Coeff.*	*Exponential R^2*	*PowerLaw Coeff.*	*Power Law R^2*
1	21	0.0056	0.8458	0.0459	0.9866
2	48	0.0057	0.9924	0.8405	0.8405
3	70	0.0057	0.9439	0.1042	0.6390
4	35	0.0106	0.9080	0.1072	0.6581
5	35	0.0552	0.8164	0.5368	0.5473
Totals	209	0.0111	0.7346	0.5356	0.4291

4.3. Empirical Findings for the United States Communication Networks

Data on the North American long haul fiber optic infrastructure was collected for 2003, from Platts Inc. TelcoMap GIS product. The data set covers over forty different physical fiber providers in North America, and contains a route confidence level for the geographic accuracy of right of way routings. Figure 3 shows the backbone network by fiber density for the United States. The diagram illustrates that there are areas, within the United States, that are very well-connected, and most of these points are located within or near major population centers. Also there are large densities of physical fiber connecting the major metro demand centers.

In order to get a sense of the connectivity distribution of the US long haul fiber network, a novel GIS-based technique was implemented. One of the difficulties in examining physical fiber networks is determining what is a link and what is a node in the network. When very dense networks, like the

North American one, are examined, it can be exceedingly difficult. While the physical location data can be very specific, it is impossible to know what parts of the fiber are nodes for interconnection or routing or even who is leasing the fiber and how it is being utilized. Just because two fiber lines cross in physical space, it does not imply that there is an operational routing node located there. Further, there can be hundreds of operational network utilizing any one fiber right of way. To circumvent this problem, an operation was executed in GIS, to calculate the number of fiber lines intersecting with metropolitan statistical area, providing a proxy of the number of fiber lines originating from the area. The fiber connectivity for every metropolitan area above 50,000 people was then rank ordered and plotted to determine the distribution. The results of the analysis demonstrate that the US fiber link network follows an exponential distribution, similarly to the European equivalent. Figure 4 shows the exponential distribution for the United States. It can be seen that the model has a very good fit, as indicated by the R^2 of 0.98. In the case of power law distribution fitting, the R^2 value is 0.81, while the exponent is 1.33, which would suggest a steep slope and significant divide between the most and least connected nodes in a network (from an economic perspective this would indicate a high Gini coefficient[11]), according to the literature on complex networks. Regarding the distribution of fiber links, it can be noted that the most connected cities of the data set show values lying *above* the estimated curve for an exponential distribution. This evidence might imply that bigger – or more connected – cities are now developing faster than the others. A similar phenomenon has already been observed for IP backbone infrastructure (Gorman and Kulkarni, 2000; Moss and Townsend, 2000).

Clusters for the US data on the number of links, consisting of 326 cities/metropolitan areas, turned out to be rather different from the ones developed for Europe. Whether by using population (in addition to the number of links) in order to aggregate the cities, or by not using it, only two clusters are proposed. The difference in the resulting clusters stands in the number of cities belonging to the first cluster. When population is also considered, the first clusters enlarges, from comprising only New York, to embracing Los Angeles and Chicago. Additional analyses, utilizing Ward's hierarchical clustering method[12], showed that two different clusters can be obtained. The first cluster comprises the top 13 cities in terms of links. Consequently, the second cluster contains the remaining 313 cities. The small size of the first cluster does not allow us to thoroughly examine its connectivity distribution, as R^2 values for both exponential and power law distributions show high values.

[11] The Gini index is a measure of inequality in distributions. The index measures the extent to which the distribution of a given variable differs from a perfectly equal distribution. The index can assume values ranging from 0 (perfect equality) to 100 (perfect inequality). The most common use of the Gini index is for measuring income inequality.

[12] For details on Ward's clustering method see Ward (1963) or more recently, El-Hamdouchi and Willett (1986).

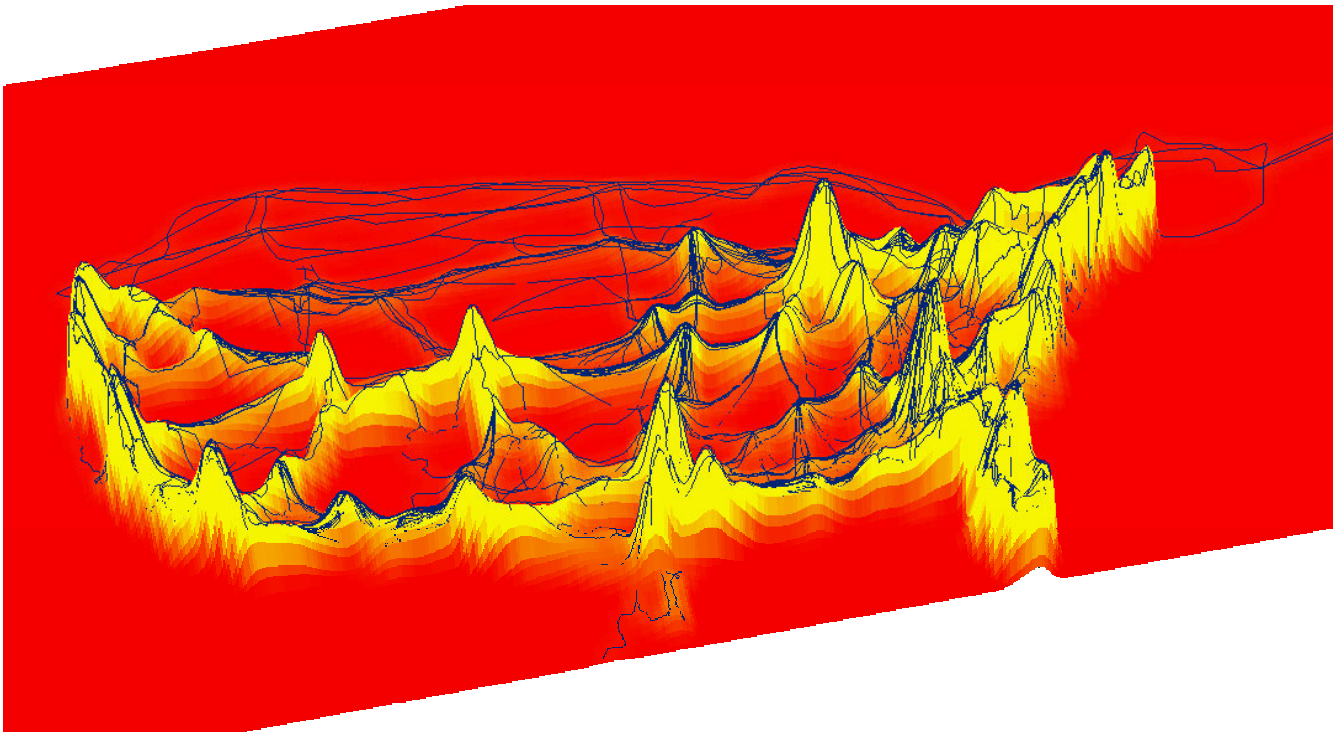

Figure 3 – The backbone network by bandwidth density for the United States

Singularly, the cities comprised in the first cluster do not correspond to the top 13 cities in terms of population. While several major cities and metropolitan areas belong to the cluster, we should note the presence of relatively minor areas, such as Fort Lauderdale-West Palm Beach, El Paso, or Bridgeport. This occurrence can be explained by the geographical characteristics of such cities. Due to the planar nature of fiber networks, these cities are gateways on the road to major end-of-the-line locations, which, for this reason, enjoy a larger number of links. For example, El Paso is an important routing hub between the large Los Angeles and Dallas markets. While El Paso does not have a large demand itself it is an important geographic location routing wise. When examining the security of an infrastructure most often prominent impetus is placed on the largest cities, but as this analysis points out smaller but geographically strategic cities are also important to consider. This also illustrates a critical difference between the analysis of logical networks like the Internet and physical networks like fiber optic cable. In an Internet network Dallas and Los Angeles would have a direct connection to each other and El Paso would be a minor location. Based on the number of logical Internet connections El Paso is 47th, but ranks 7th in physical fiber. Bridgeport is even more dramatic as a routing point between New York City and Boston – ranking 214th in logical Internet connections and 11th in physical fiber connections. The difference between planar and non-planar networks is not only seen in connectivity distributions,

but also which places in those distributions are important. Also it reinforces the fact that any one connectivity distribution can represent a wide variety of real world topologies.

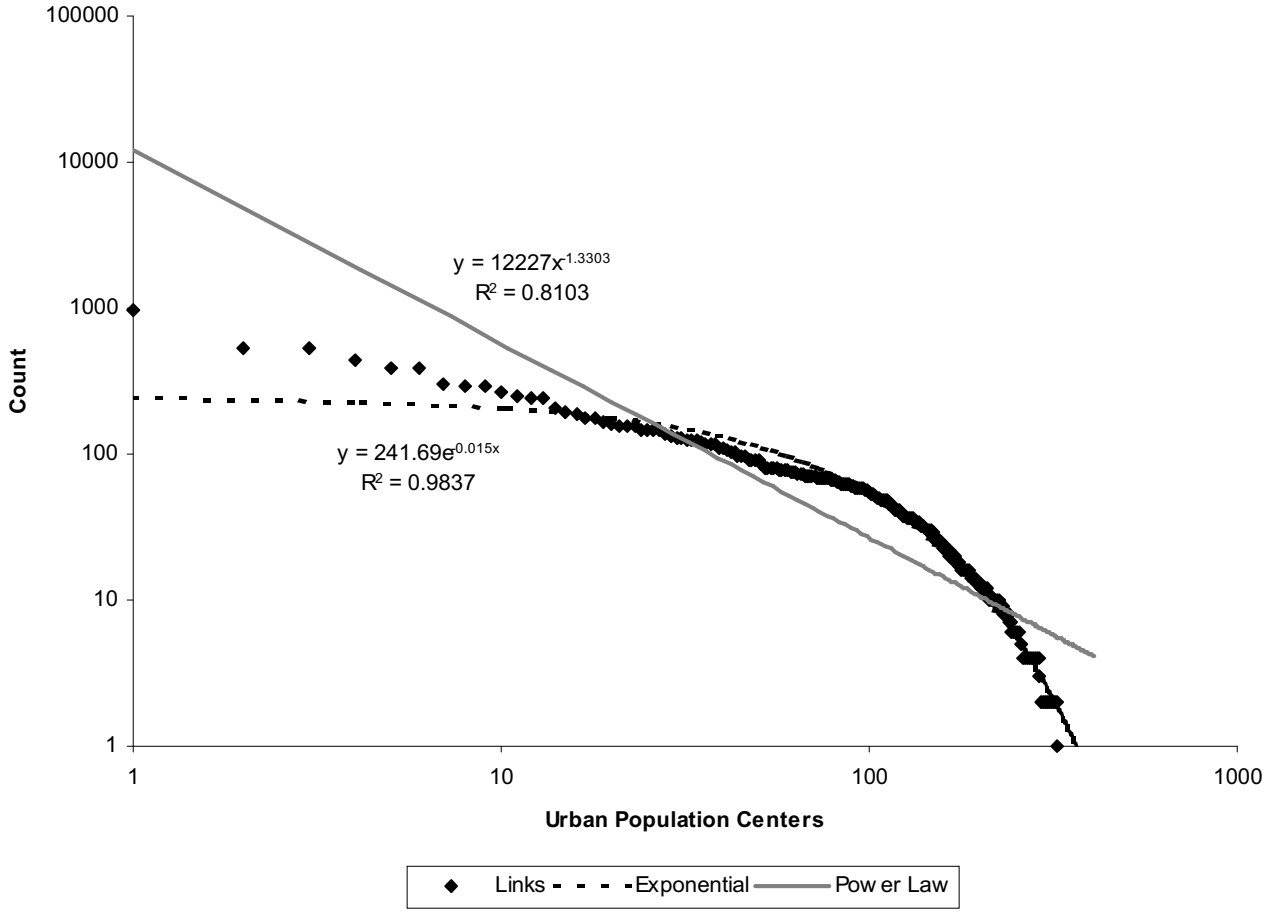

Figure 4 – The distribution of fiber links in US cities

4.3. A Comparison of the Results Obtained for the US and Europe

Differences between the United States and Europe in terms of the distribution properties of their telecommunications networks can be explained by technological, economic, cultural, and policy factors, all of which are very much intertwined. As noted in Table 1, the U.S. leads Europe in terms of some of the variables that reflect these factors – i.e., GDP per capita, Human Development Index, Education Index and R&D expenditures. Further, there is a greater spatial disparity between these factors in Europe, whereas in the United States they are relatively uniform across metropolitan areas. The exception is rural areas, although they are not considered in this analysis.

Europe has also historically had an institutional framework that one may argue has contributed to vast spatial disparities in the deployment of telecommunications infrastructure. Access costs, along

with a slow approval process, hinder free access into the market. These same forces have prevented the telecommunications network in Europe from evolving according to the factors that are necessary ingredients for the emergence of power law distributions: rewiring, incremental growth and preferential attachment.

While these institutional and economic differences have a clear effect on the development of IP networks seen in previous studies (Schintler et al. 2003, Gorman and Kulkarni 2004) they do not seem to influence the distribution of fiber cable links, which we observed in Sections 4.1–2. The reason for the uniform results found in this paper might lie in the planar nature of fiber connection networks. The deployment of such networks requires large fixed costs, which are an obstacle to the development of competition as fast and dynamic as the one observed for IP networks.

5. Conclusions

The analysis evokes some interesting questions that might be the focus of further investigations. A more rigorous analysis of some of the factors that have led to differences between the US and Europe in terms of their telecommunications networks is recommended. In particular, further study is needed to explore how cultural, political, economic and technological factors hinder the processes necessary for the development of a network that has power law distribution properties. This could either be done mathematically, using statistical models or simulation, or through case studies, perhaps examining some of the nations in Europe that represent outliers in the analysis. It may also be interesting to explore how the clusters identified in this paper have evolved over time, as the market expanded in the US and Europe, and to utilize alternative clustering methods to verify the consistency of the results presented in this paper.

Acknowledgements

The authors wish to thank Sandra Vinciguerra (University of Bologna, Italy) for the cluster analysis elaboration.

References

Albert, R. and A.-L. Barabasi (2000). Topology of Evolving Networks. *Physical Review Letters* 85, pp. 5234-5237.

Albert, R. and A.-L. Barabasi (2002). Statistical Mechanics of Complex Networks. *Review of Modern Physics* 74, pp. 47.

Allen, D. (1996). Dynamic Industry Structure as Policy? Three Alliances and World Telecommunications. In: *Telecommunication: New Dynamics and Driving Forces*, A. Bianchi and G. Richeri, Eds. Amsterdam, IOS Press.

Amaral, L.A.N., A. Scala, M. Barthélémy and H.E. Stanley (2000). Classes of Small-World Networks. *Proc. Nat. Acad. Sci. USA* 97, pp. 11149-11152.

Anderberg, M.R. (1973). *Cluster Analysis for Applications*. New York, Academic Press.

Balbi, A. (2004). Banda larga: è boom in Italia e in Europa. www.repubblica.it. http://www.repubblica.it/2004/c/sezioni/scienza_e_tecnologia/bandalarga/bandalarga/bandalarga.html.

Ballou, R.H. (2004). *Business Logistics / Supply Chain Management: Planning, Organizing, and Controlling the Supply Chain*. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Barabasi, A.-L. (2001). The Physics of the Web. *Physics World* July 2001.

Barabasi, A.-L. (2002). *Linked: the New Science of Network*. New York, Perseus Publishing.

Barabasi, A.-L. and R. Albert (1999). Emergence of Scaling in Random Networks. *Science* 286, pp. 509-512.

Barthelemy, M. (2003). Crossover from Scale-Free to Spatial Networks. *Europhysics Letters* 63 (6), pp. 915-921.

Bergman, L., C. Doyle, J. Gual, L. Hultkrantz, D. Neven, L.-H. Röller and L. Waverman (1998). *Europe's Network Industries: Conflicting Priorities*. London, Centre for Economic Policy Research.

Bittlingmayer, G. and T.W. Hazlett (2002). The Political Economy of Cable 'Open Access'. AEI-Brookings Joint Center for Regulatory Studies, Working Paper 01-06.

Bowen, J. (2002). Network Change, Deregulation, and Access in the Global Airline Industry. *Economic Geography* 78 (4), pp. 425-439.

Cave, M. and R.W. Crandall, Eds. (2001). *Telecommunications Liberalization on Two Sides of the Atlantic*. Washington, D.C., AEI-Brookings Joint Center for Regulatory Studies.

Chen, Q., C. Hyunseok, R. Govindan, J. Sugih, S. Schenker and W. Willinger (2001). *The Origin of Power Laws in Internet Topologies Revisited*. IEEE Infocom, 2002.

Crandall, R.W. (1997). Are Telecommunications Facilities "Infrastructure"? If They Are, So What? *Regional Science and Urban Economics* 27 (161-179).

Economist (2003). Unfixing Fixed Wireless. http://www.economist.com/science/tq/displayStory.cfm?story_id=1620752.

El-Hamdouchi, A. and P. Willett (1986). *Hierarchic document classification using Ward's clustering method*. Proceedings of The 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, ACM Press, New York, USA.

Erdos, P. and A. Renyi (1960). *On the Evolution of Random Graphs*, Publication of the Mathematical Institute of the Hungarian Academy of Science.

Garrison, W. (1968). Connectivity of the Interstate Highway System. In: *Spatial analysis; a reader in statistical geography*, B.J.L. Berry and D.F. Marble, Eds. Englewood Cliffs, N.J.,, Prentice-Hall, pp. 239-249.

Gasper, J. and E.L. Glaeser (1998). Information Technology and the Future of Cities. *Journal of Urban Economics* 43, pp. 136-156.

Gillespie, A. and K. Robins (1989). Geographical Inequalities: The Spatial Bias of the New Communications Technologies. *Journal of Communication* 39.

Gorman, S.P. (2002). The Network Advantage of Regions: The Case of the USA, Europe and China. School of Public Policy, George Mason University, Fairfax, VA.

Gorman, S.P. and R. Kulkarni (2000). The Networks of the Internet. *Telecommunications Policy* 26, pp. 113-134.
Gorman, S.P. and R. Kulkarni (2004). Spatial Small Worlds. *Environment & Planning B* 31, pp. 273-296.
Haggett, P. and R.J. Chorley (1969). *Network analysis in geography*. New York, St. Martin's.
Hatigan, J.A. (1975). *Clustering Algorithms*. New York, John Wiley and Sons.
Hepworth, M. (1987). *Geography of the Information Economy*. New York, Guilford Press.
INA (2004). Market Overview. Southeast Europe ICT Weekly Update, 23 January 2004. Thessaloniki.
Kalakota, R., P. Gundepudi and J. Wareham (2002). The Economics of DSL Regulation. *Computer* 35 (10).
Kansky, K. (1963). Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics. University of Chicago, Department of Geography, Research papers.
Malecki, E.J. (2002). The Economic Geography of the Internet's Infrastructure. *Economic Geography* 78, pp. 399-424.
Malecki, E.J. and S.P. Gorman (2001). Maybe the Death of Distance but not the End of Geography. In: *The Worlds of Electronic Commerce*, S.D. Brunn and T.R. Leinbach, Eds. New York, John Wiley, pp. 87-105.
Moss, M. (1998). Technology and Cities. *Cityscape* 3, pp. 107-127.
Moss, M.L. and A. Townsend (2000). The Internet Backbone and the American Metropolis. *The Information Society* 16, pp. 35-47.
OECD (2002). Information Technology Outlook 2002. http://www.oecd.org/dataoecd/63/60/1933354.pdf.
Oum, T.H., A. Zhang and Y. Zhang (1995). Airline Network Rivalry. *Canadian Journal of Economics* 28 (4a), pp. 836-857.
Oxman, J. (1999). The FCC and the unregulation of the Internet. OPP Working Paper No. 31, Federal Communications Commission, Washington, D.C.
Pehrsson, A. (2001). *Strategy in Emerging Markets : Telecommunications Establishments in Europe*. London ; New York, Routledge.
Reggiani, A. and L.A. Schintler, Eds. (2005). *Methods and Models in Transport and Communications. Cross Atlantic Perpectives*. Berlin, Springer Verlag (Forthcoming).
Rinaldi, S.M., J.P. Peerenboom and T.K. Kelley (2001). Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. *Control Systems Magazine* 21 (6), pp. 11-25.
Rutherford, J., A. Gillespie and R. Richardson (2004). The Territoriality of pan-European Telecommunications Backbone Networks. *Journal of Urban Technology* 11 (3).
Schiller, D. (1999). *Digital Capitalism: Networking the Global Market System*. Cambridge, MA, MIT Press.
UNDP (2002). *Human Development Report 2002*. New York, Oxford, Oxford University Press.
Vinciguerra, S. (2004). Modelli di rete: sviluppi metodologici ed applicazioni empiriche con riferimento al settore delle telecomunicazioni, Degree Thesis, Faculty of Statistics, University of Bologna (in Italian, forthcoming).
Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, pp. 236-244.
Watts, D.J. and S.H. Strogatz (1998). Collective Dynamics of Small-World Networks. *Nature* 363, pp. 202-204.
Wojahn, O.W. (2001). Airline Network Structure and the Gravity Model. *Transportation Research: Part E: Logistics and Transportation Review* 37 (4), pp. 267-279.

Annex A

Table A1 – A summary of the data set for European cities, showing the subdivision in clusters.

Clusters	*City*	*Country*	*Networks*	*Connections*	*Fitting Function*
1° cluster	Hamburg	Germany	21	200	Power Law
	…	…	…	…	
	Bristol	United Kingdom	12	172	
2° cluster	Koebenhavn	Denmark	12	170	Exponential
	...	...	...	...	
	Bilbao	Spain	7	130	
3° cluster	Luxembourg	Luxembourg	4	125	Exponential
	…	…	…	…	
	Basingstoke	United Kingdom	3	86	
4° cluster	Herning	Denmark	3	84	Exponential
	…	…	…	…	
	Regensburg	Germany	1	54	
5° cluster	Frederikshavn	Denmark	3	37	Exponential
	…	…	…	…	
	Napoli	Italy	1	5	

Table A2 – Top European cities in estimated total bandwidth of Internet backbone links

City	*Estimated Total Bandwidth (Gbps)*
London	31.00
Paris	30.25
Frankfurt	30.25
Dusseldorf	29.75
Hamburg	26.50
Amsterdam	23.25
Zurich	18.50
Berlin	17.50
Munich	17.50
Lyon	17.25
Strasbourg	16.50
Brussels	16.00
Madrid	15.75
Copenhagen	15.50
Marseille	15.00
Stuttgart	13.75
Cologne	13.50
Hannover	13.25
Milan	12.25
Stockholm	11.75
Leipzig	11.50
Geneva	11.50

Source: Rutherford et al. (2004, Table 4).

Figure A.1 – The map of the 21 cities belonging the first cluster (see Table A1). Source: Vinciguerra (2004).