Scale-Free Phenomena in Communication Networks: A cross-Atlantic Comparison

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

[*] The School of Public Policy, George Mason University, USA; email: lschintl@gmu.edu; sgorman1@gmu.edu; [**] Department of Economics, University of Bologna, Italy; email: reggiani@economia.unibo.it; patuelli@tin.it; [***] Department of Spatial Economics, Free University of Amsterdam, The Netherlands; email: pnijkamp@feweb.vu.nl

Abstract

One of the key features of our modern world is its gradual transition to a network society. Many networks like the Internet have been found to possess scale-free and small-world network properties exhibited by power law distributions.

Scale-free properties evolve in large complex networks through self-organizing processes and more specifically, preferential attachment. New nodes tend to attach themselves to other vertices that are already well-connected. Because traffic is routed mainly through a few highly connected 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 puts scale-free networks at risk for becoming disconnected or significantly disrupted, when super-connected nodes are removed either intentionally or through a targeted attack.

This paper will examine and compare three communication networks: bandwidth capacity between major metropolitan areas within the United States (97-01), inter-district phone traffic in Italy (1989-1993) and a particular peer-to-peer data exchange network. Each network will be examined in terms of its network topology and specifically whether or not they are evolving into scale-free networks.

Finally, the paper will conclude with some preliminary thoughts and reflections aiming at a consolidation and exploration of further comparative studies on communication networks in Europe and North America, in the light of the STELLA objectives.

1 Introduction

One of the key features of our modern world is its gradual transition to a network society. This development has prompted many intriguing research questions, not only on the structural properties of networks (e.g., connectivity properties), but also on its evolutionary properties. The present paper will address the latter type of questions.

Many networks like the Internet have been found to be small-world networks possessing so-called scale-free properties exhibited by power law distributions reflecting their non-linear dynamic features. Scale-free properties evolve in large complex networks through self-organizing processes and more specifically, preferential attachment. New nodes tend to attach themselves to other vertices that are already well-connected. Systems with this topology are generally viewed as falling into a larger class of networks that exhibit a small-world phenomenon. A small-world network is characterized by a high degree of local clustering and a short average minimum path or diameter through the network. Because traffic is routed mainly through a few highly connected vertices, the diameter of the network is small in comparison to other network structures, and movement through the network is therefore efficient. In this context, Watts and Strogatz (1998), two pioneers of "small-world" network analysis, argue that "models of dynamical systems with small-world coupling [for example] display enhanced signal propagation speed, computational power, and synchronizability." While an efficient network topology is good in certain respects, it also presents some problems emerging from its high connectivity. Bad elements such as contagious diseases, forest fires and Internet viruses tend to spread more freely in "small-world networks." Also, a network with "small-world" properties is more vulnerable to major disruption or a shutdown, when super-connected nodes are removed either intentionally or through a targeted attack.

This paper examines small-world phenomena in communications systems focusing specifically on three networks each operating in different geographical spheres. The first one is the logical IP (Internet Protocol) fiber optic infrastructure that connects major metropolitan areas in the United States (for the years 1997 through 2000), the second one a portion of the Italian telephone network using outgoing landline calls by district to capture network traffic dynamics, while the third one is a peer-to-peer data network for the international exchange of music for a particular group of independent people. Power law distributions are generated for each network to look for scale-free properties. The implications of the results of these experiments for transportation policy and planning, and the way in which they may vary depending on the geography at hand – for example, whether or not a network operates in Europe versus the United States, or whether it is one with no geographical boundaries and rather has an international dimension – are extensively discussed. The paper offers also some thoughts about the analytical methodologies, visualization techniques and data that are needed to facilitate a valid and informative cross-Atlantic comparison of communication networks in this context.

The paper contains three sections in addition to the present introduction. In Section 2, the small-world network concept as it has evolved over the past few decades and the way in which it has been applied to various systems in a variety of fields such as transportation and communications is described. Section 3 presents the results of the empirical experiments involving the US backbone network, telephone traffic in Italy and the peer-to-peer musical data exchange system. Finally, Section 4 presents the implications of the results and directions for cross-Atlantic research in this area.

2 Small-world Network Analysis

2.1 Prefatory remarks

The concept of "small-world" networks has recently received much attention, although its origins stem from early work done some forty years ago on large, complex systems. Erdös and Renyi (ER) (1960) were pioneers in this area applying "probabilistic methods" to solve problems in graph theory, where a large number of nodes were involved (Albert and Barabasi 2002, p. 54). Under this assumption, they modeled large graphs utilizing algorithms where N nodes were randomly connected according to probability p, and found that when vertices were connected in this fashion they followed a Poisson distribution (Albert and Barabasi 2002, p. 49). A more thorough review of random graphs can be found in the survey work of Bollobás (1985). Following ER's findings, their random models of network formation were widely used in several disciplines examining networks, the most topical to this research being Internet topology generators (Radoslavov *et al.*, 2000).

The absence of detailed topological data for complex networks left random network models 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, p. 48-49):

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 density, called degrees, often follow power laws.

Watts and Strogatz (WS) (1998) formalized this concept of clustering for large, complex networks, although others had introduced several years before some of the features of "small-world" phenomena using smaller graphs (e.g., see Zipf, 1949). Using several large data sets, WS found that the real-world networks studied were not entirely random, but instead displayed significant clustering at the local level. According to WS, "small-world" networks are characterized by their average path length L(p) and the degree to which there is local connectivity in the network, measured by a clustering coefficient C(p). The variable L(p) measures the average minimum path in the network and C(p) the connectivity of an average neighborhood in the network. More specifically, L(p) is the smallest number of links it takes to connect one node to another, averaged over the entire network, and clustering is the fraction of adjacent nodes connected to one another. One may view L(p) as a global property of the network and C(p) a local property.

WS (1998) showed that a "small-world" network falls somewhere in between a regular lattice and a random network. To demonstrate this, they began with a regular lattice with n vertices and k edges, and rewired it in such a way that it approached a random network. Specifically, beginning with a vertex, the edge connected to its nearest neighbor was reconnected with probability p to another vertex chosen randomly from the rest of the lattice. No rewiring occurred, if there already exists a connection to that vertex. They continued the process by moving clockwise around the lattice, and randomly rewiring each edge with probability p, until the lap was completed. Next, the same process was repeated for vertices and their second nearest neighbors. Because they considered a network with only first-order and second-order connections in each direction of the vertex, they terminated the rewiring process after two laps.

In general, for a network with k nearest neighbors, WS found that rewiring would stop after k/2 laps. As the network is rewired, shortcuts through the network are created, resulting in an immediate drop in L(p). Local clustering, or C(p), remains relatively high up to a point after which it begins to drop rapidly. The results of this process suggest that the global connectivity of a regular network can significantly improve with the addition of just a few shortcuts; in essence, a "small world" network is one with high degree of local clustering and a short average minimum path.

The short cuts across the graph to different clusters of vertices introduced a level of efficiency[1] not predicted in the ER model. The distribution was not Poisson as with the ER model, but was bounded and decayed exponentially for large sets of vertices (Watts and Strogatz, 1999). Watts (2003) has extended this work recently to cover topics ranging from: "epidemics of disease to outbreaks of market madness, from people searching for information to firms surviving crisis and change, from the structure of personal relationships to the technological and social choices of entire societies (p.1)." The work by WS was not the first though, to investigate the effects of rewiring, as is witnessed by the following quotation (Hayes, 2000, p. 106): "*...Fan R. K. Chung, in collaborations with Michael R. Garey of AT&T Laboratories and Béla Bolobás of the University of Memphis, studied various ways of adding edges to cyclic graphs. They found cases where the diameter is proportional to log n.*"

The finding of WS spurred a flurry of work into understanding the attributes of complex networks, while new findings and discoveries quickly followed. Two parallel studies by Albert, Jeong and Barabasi (1999) of Notre Dame and Adamic and Huberman (1999) at Xerox Parc found that when one looks at the World Wide Web as a graph (web pages are vertices and hyperlinks connecting them are edges), it follows not a Poisson or exponential distribution, but a power law distribution.

2.2 Methodological approach: power law 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 both the ER and WS model; the probability of finding a highly connected vertex in the ER and WS model decreases exponentially, so that "vertices with high connectivity are practically absent"[2] (Barabasi and Albert 1999, p.510). The reason, according to Barabasi and Albert (1999), 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 (see Figure 1 for a graphic representation of the function). 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.
- 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.

The difference between the random model of Erdös and Renyi and the model described by Barabasi and Albert becomes clearer when seen in a visual representation. In their node diagram, Barabasi (2001, p. 1) show that more than 60% of nodes (green) can be reached from the five most connected nodes (red) compared with only 27% in the random network. This demonstrates the key role that hubs play in the scale-free network. Both networks contain the 130 nodes and 430 links.

This leaves the rather fuzzy question of what is a small world and what is a scale-free network. As stated earlier, Albert and Barabasi (2002) see small worlds and scale-free networks as explanations for two different phenomena occurring in complex networks. The WS small world model explains clustering and the scale-free model explains power law degree distributions (Albert and Barabasi 2002, p.49). There have though, been other opinions on how small world and scale-free networks should be classified; Amaral *et al.* (2000) argue that scale-free networks are a sub- class of small world networks. Further, they argue that there are three classes of small world networks (Amaral *et al.*, 2000, p.11149):

a) Scale-free networks, characterized by a vertex connectivity distribution that decays as a power law;
b) Broad-scale networks, characterized by a connectivity distribution that has a power law regime followed by a sharp cutoff;
c) Single-scale networks, characterized by a connectivity distribution with a fast decaying tail.

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.

2.3 Small-world network applications

"Small-world" network phenomena have been explored in the context of many large, complex networks. Not only has the World Wide Web found to fall into a scale-free organization, but so has the Internet. The Faloutsos brothers (1999) found that the Internet followed power laws at both the router level and autonomous system (AS) level. The router level entails the fiber optic lines (edges) and the routers (vertices) that direct traffic on the Internet, and the AS level entail networks (AT&T, UUNet, C&W etc.) as vertices and their interconnection as edges. This means that the physical fabric of the Internet and the business interconnections of the networks that comprise the Internet both qualify as scale-free networks. Before these discoveries, the Internet had been modeled as a distinct hierarchy or random network, and the new finding had many implications throughout the field of computer science.

Scale-free theory and BA model have not been without debate. Several arguments have been made stating that the BA model is too simplistic for the Internet and additional corollaries need to be made (Chen *et al.*, 2001). The re-wiring principle was one of Albert and Barabasi's (2000) responses to these criticisms, but overall the model has held. Tests of network generators based on power laws have been found to produce better models, and many efforts are being made to base new Internet protocols on these discoveries (Tangmunarunkit *et al.*, 2001, Radoslavov *et al.*, 2001). While these discoveries have paved the way for advances in several fields, the question of the geography and location of these networks remains to be addressed.

Small-world properties have also been found in transportation networks. Amaral *et al.* (2000) found that the airline network was a small world because of its small average path length, and other transportation networks such as the Boston subway has also been found to be small worlds (Latora and Marchiori 2001). Schintler and Kulkarni (2001) discovered the emergence of "small-world" phenomena in a congested road network. One may argue though, that transportation networks are less prone to evolve into a scale-free structure over time given the fact that they tend to be planar. The number of edges that can be connected to a single node is limited by the physical space available to connect them and it is this fact that makes the large number of connections needed for a power law distribution quite difficult to obtain. But even in some non-planar networks this may be a problem as well. Airline networks, for example, have similar properties. The number of connections is limited by the space available at the airport, and "such constraints may be the controlling factor for the emergence of scale-free networks" (Amaral *et al.*, 2000, p. 11149).

Finally, there is an interesting parallel between the study of scale-free networks from a mathematical-statistical perspective and network externalities from an industrial organization perspective. Network externalities refer to unpaid benefits for users or subscribers of a network facility as a result of additional entry of new members. Given the direct and indirect connectivity increase of one additional member, a non-linear evolutionary growth is obtained. This drives the network to a more than proportional performance and explains the rapid introduction rate of new forms of network technologies (e.g., mobile phone).

More applications and experiments aiming to verify the power law behavior or scale-free assumptions can be found in different fields, ranging from biology, e.g., with the study of the metabolic network of the E. Coli bacterium (Jeong, 2003), to networks in linguistics (Albert and Barabasi, 2002). Annex A presents a brief review of these applications.

3 Empirical Experiments

3.1 Introduction

In this section, we look more closely at three communications networks and use the power law distribution methodology to assess whether or not they possess scale-free properties. The networks examined include the US IP (Internet protocol) fiber optic infrastructure, the landline telephone network in Italy and a peer-to-peer musical data exchange network. The results are mixed, most likely owing perhaps to the diversity of the networks examined and differences in the social, economic and political factors underpinning these networks.

3.2 Bandwidth network in the United States

The US IP (Internet protocol) infrastructure is an interesting case to study features of small-world networks. The logical network itself is planar, while the underlying physical infrastructure – i.e., the fiber that is positioned in the ground, is planar, meaning there is a spatial aspect attached to it that may hinder the development of scale-free attributes in the network. This study will examine whether the vertex connectivity of these networks exhibits a power law distribution over time and whether other scale-free properties such as preferential attachment and rewiring have also taken place.

Data on the US IP fiber optic infrastructure data was collected for the years 1997-2000. While 1997 and 1999 data was obtained from New York University's Information Technology and the Future of Environment project (SBR-9817778) (Moss and Townsend, 2000), the 1998 data was compiled from CAIDA's MapNet application, and the 2000 data was obtained from the University of Florida's The Infrastructure of the

Internet: Telecommunications Facilities and Uneven Access project (BCS-9911222) (Malecki, 2002). All four data sets cover the backbone layer-three transit providers of the USA Internet and are very similar in composition. It should be noted that data in all three data sources is not always 100% accurate, since carriers often advertise more bandwidth and lines than are actually in service and topological errors have been found in the past. These have been corrected for to the maximum extent. However, for the gross level of aggregate analysis in this paper these data sets are a viable and useful information source.

For all four data sets the total bandwidth connecting to a consolidated metropolitan area (CMSA) was tabulated. For the 1998 and 2000 data sets this was done through the construction of a matrix and the calculation of an accessibility index based on the bandwidth capacity of the links for each CMSA. For 1997 and 1999, the data was provided with total bandwidth connected to the CMSA already tabulated. Capacity was totaled for each CMSA as the total number of mega bits per second (Mbps) of fiber optic connections to the CMSA, running IP. Since binary connectivity data was not available for 1997 and 1999, total capacity was utilized for comparison across the four years of data. Other researchers, including Amaral *et al.* (2000) in their analysis of airline networks, utilize the weight of a link in their methodology to determine if a network is scale-free. This approach is commonly used when structural network data is not available or the number of nodes is too small for a log-log plot. Utilizing capacity as a measure of connectivity also makes sense, since it takes into account the large number of lines connected to any one CMSA and the common practice of partitioning these lines. The vast majority of fiber optic partitions are as T-1's that carry 1.544 Mbps, thus the Mbps total for each city, can very roughly approximate the number of T-1 lines possibly available.

The data for 1997-2000 was individually plotted as rank order distributions with log-log plots and fitted with a power law (see Figures 1-4). For each graph the x-axis is the total bandwidth connected to a CMSA and the y-axis is the CMSA ranked in descending order. Prior research by Moss and Townsend (2000) found a high level of similarity in the exponential curves for the rank-size distribution plots by number of edges connected to a metropolitan area. The power law exponent used provides a rate of increase indicator, an exponent of 2 would indicate an increasing sequence of 1,4,9,16,25 or an exponent of 3 would indicate an increasing sequence of 1,8,27,64,125. The USA's backbone network has been incrementally increasing its power law exponent for each year, except for a small decrease from 1997 to 1998, but has also been increasingly moving away from a power law distribution. By 2000, the network's power law exponent is 1.82, approaching the range found in other real world scale-free networks but the distribution is far from a power law. In fact the 2000 and 1999 data appears to be two different trends lines occurring. A closer examination reveals that there is a break between the top 110 CMSA's and the bottom 37 each with a distinctly different slope. Interestingly, of the bottom 37 CMSA's, 33 do not have any of the high-speed 2.5, 5, or 10 Gbps connections.

The first distribution runs from the minimally connected locations at 45 Mbps and follows the power law until a tail forms, starting with Fort Pierce, FL with 4,976 Mbps and ending with Syracuse at 5,624 Mbps, consisting of 18 city vertices. The distribution then resumes a normal power law trend to the top connected locations. The jump from Laredo, TX with 2,488 Mbps to Brownsville, TX at 4,976 appears at first glance to indicate a critical mass at which cities gain a level of preferential attachment into the network. Theoretically, as the USA Internet continues to evolve, these kinks in the distribution will work out as connectivity spreads to more nodes, erasing clustered hierarchies. A closer examination of the data reveals that the reason for this clustering is a technology shock in the network. Beginning, for the most part, in 1999 several networks began provisioning dense wave dimension multiplexing (DWDM) lines with capacities of 2448 Mbps in their networks, a large increase in capacity from the more common 45 and 155 Mbps lines. A connection to two cities provides 4,976 Mbps and caused a whole cluster of cities to be bumped up into the 4,976-5,624 Mbps noted in the distribution. Massive investment in Internet backbone capacity has occurred between 1998 and 2000 in the US. In early 1998, only two of 38 national backbones offered bandwidth at OC-48 (2488 Mbps or 2.488 Gbps). By mid-2000, fully 17 of 41 backbone networks (41%) had installed capacity at bandwidths of 2488 Mbps or faster, as opposed to just 5% in 1998 (Gorman and Malecki, 2002). Such bandwidths easily overwhelm networks with a slower capacity: a single OC-48 cable has the same bandwidth as 55 of the older DS-3 (45 Mbps) capacity. The current standard is OC-192, which moves data at speeds of nearly 10 gigabits per second, and work is underway to implement OC-768 (40 Gbps) in the near future.

The existence of a break and multiple slopes in the 2000 data could indicate that the diffusion of new high-speed technologies is not even across space and does not follow a power law. In order to test this assumption, a binary connectivity distribution was built for 1998 and 2000 data, but unfortunately this was not possible for the 1997 and 1999 data. The binary connectivity distributions can be found in figures 5 (1998) and 6 (2000). When bandwidth capacity is stripped from the network, the trend reverses, and from 1998 to 2000 the connectivity distribution is getting closer to a power law fit. The exponent is also increasing, but the number is considerably lower than the bandwidth plots. This is most likely explainable, since the binary connectivity distribution does not take into account all connections in the network, but only whether there is – or is not – a connection between cities. The reality of the network topology is most likely somewhere between the two, following a power law with an exponent higher than binary but lower than bandwidth. It would seem that when weighted links are used to examine scale-free networks, there is the possibility of shocks in the network, in this case a technology shock. Interestingly while the IP networks studied do generally follow a power law, the diffusion of new technologies across the network do not follow a power law geographically or topologically. This also leaves the question whether applying curves, like power laws, is too simple an approach for weighted networks, but this issue falls out of the scope of this paper to answer.

The next condition for the BA model is preferential attachment. There is a higher probability for a new or existing node to connect or reconnect to a vertex that already has a large number of links than there is to (re)connect to a low degree vertex (Barabasi and Albert, 1999). As the network grows incrementally, it expands following preferential attachment. The probability (Π) that a new vertex will connect with another vertex (i) depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = k_i / \Sigma_j k_j$ (Barabasi and Albert, 1999). Because of 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. This characteristic can be seen in the urban hierarchy of backbone connections. The Internet largely evolved out of Washington DC, through NSFNET and one of the original network access points, MAE East. Washington, DC has leveraged this historical preferential attachment to average the highest ranking over the four years of backbone connectivity data in the time series. While the rank order of the top ten cities has shifted, they have consistently benefited from preferential connectivity to maintain the majority of connections in the network. Although it should be noted, that early first mover advantage for preferential attachment has succumbed to market size in many cases; the most obvious in the data being New York's move from a sixth to a first position. Over the four time series, the top ten cities have on average accounted for 57.4% of total bandwidth. The 1997-2000 time series appears to establish evidence of preferential attachment as one of many factors in the growth of the network. The actual testing of the BA equation to the time series was not possible, since matrix connectivity data was not available for all four years. This is a future research avenue that could yield interesting results, especially in regard to predicting the future connectivity and growth of the network.

The last condition established by the BA model is re-wiring within the network. While this is not a feature of the network that can be tested explicitly, it can be addressed anecdotally outside what has been cited in the literature. Re-wiring of the Internet occurs at many levels but at dramatically different rates. The backbone network, in general, operates at layer 3 of the OSI (open system interconnect) networking model. This is the layer where routing between networks occurs and rewiring within this virtual network occurs on a very frequent basis. Topologies and routes change frequently as new peering arrangements occur on one hand and the actual path of traffic changes constantly as congestion and traffic fluctuate on the other. The physical fiber that is installed in the ground is re-wired at a much slower pace, but re-wiring does occur. Fiber into a city is typically leased from a carrier's carrier, like Enron, Williams, or Qwest. The long haul transit fiber into a city most often surfaces at a co-location facility, network access point, or a metropolitan area network interchange. At these junctures, the individual conduits leased by multiple different backbone carriers are split off and run by various networks to their customer's locations. This allows for a considerable amount of fluidity in re-wiring topologies within backbone networks without actually digging up, turning off, or laying new fiber. The most dramatic example of this type of re-wiring was the change in Cable and Wireless's network when they acquired MCI's network. The network was significantly re-wired from a star topology focusing on connectivity to coastal cities to a partial mesh topology concentrating connectivity to interior vertices (Gorman and Malecki, 2000). While the re-wiring principle occurs at various levels of the data examined and at different rates, it is very much a factor affecting the distribution and connectivity of the network.

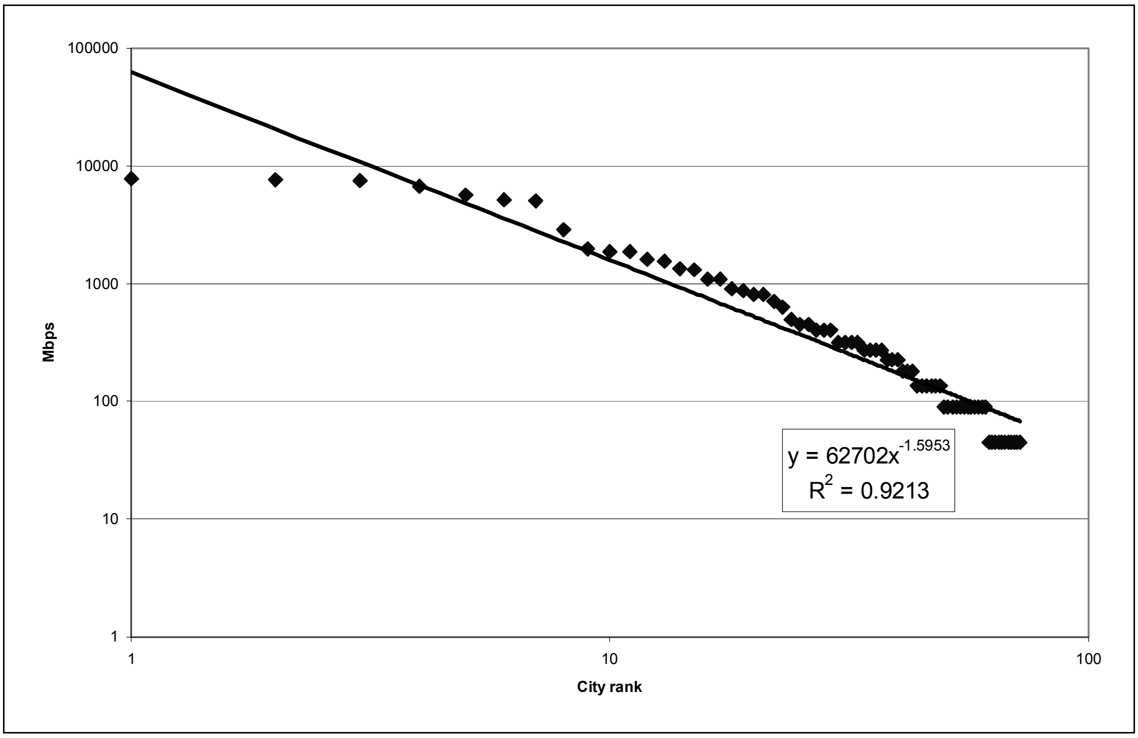

Figure 1. Power Law Distribution of the USA Internet 1997

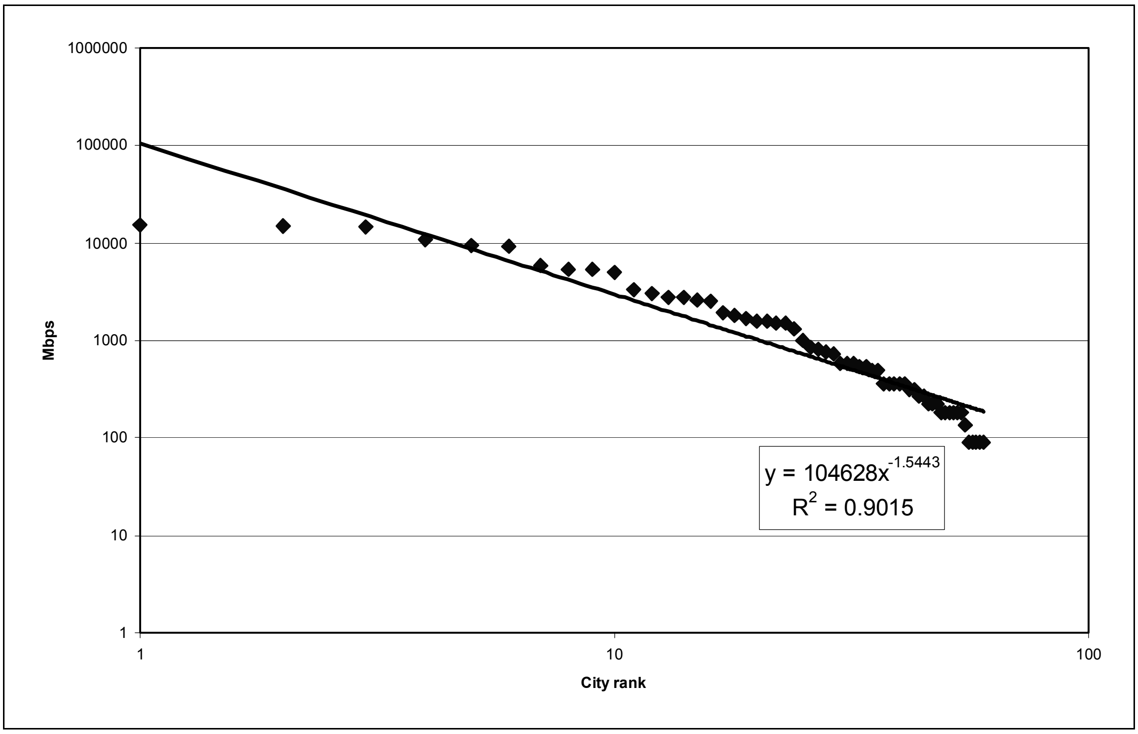

Figure 2. Power Law Distribution of the USA Internet 1998

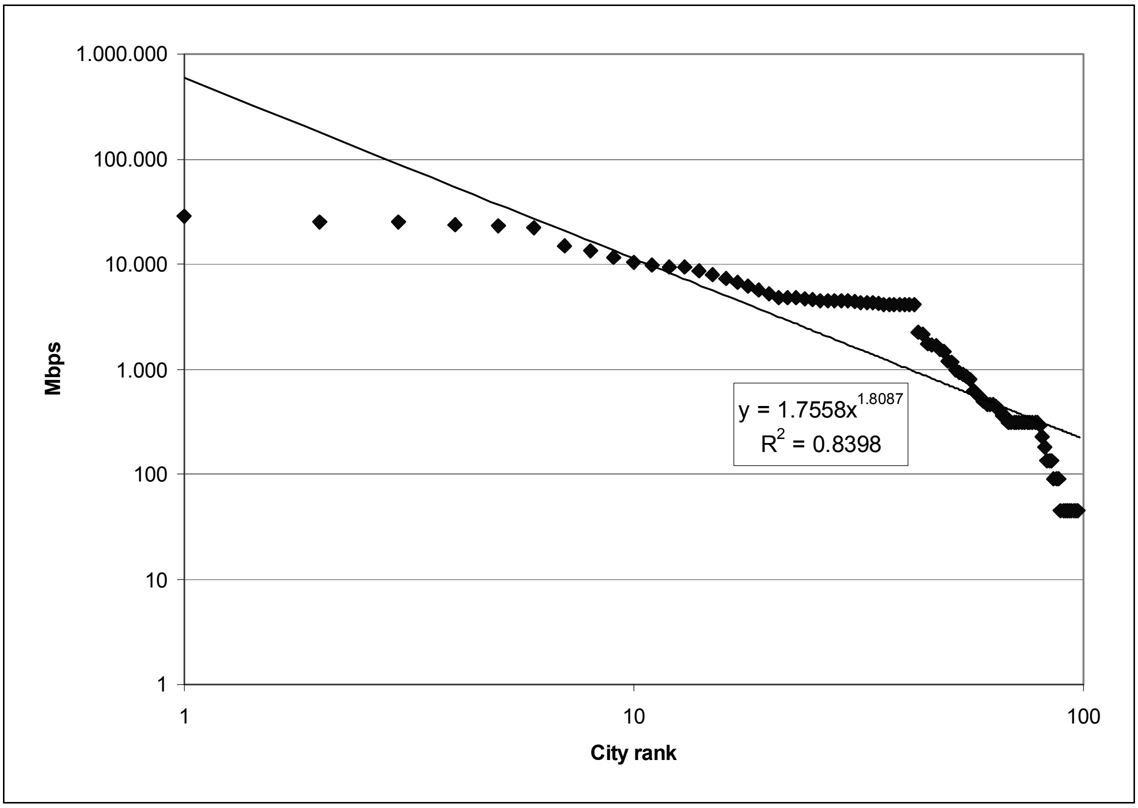

Figure 3. Power Law Distribution of the USA Internet 1999

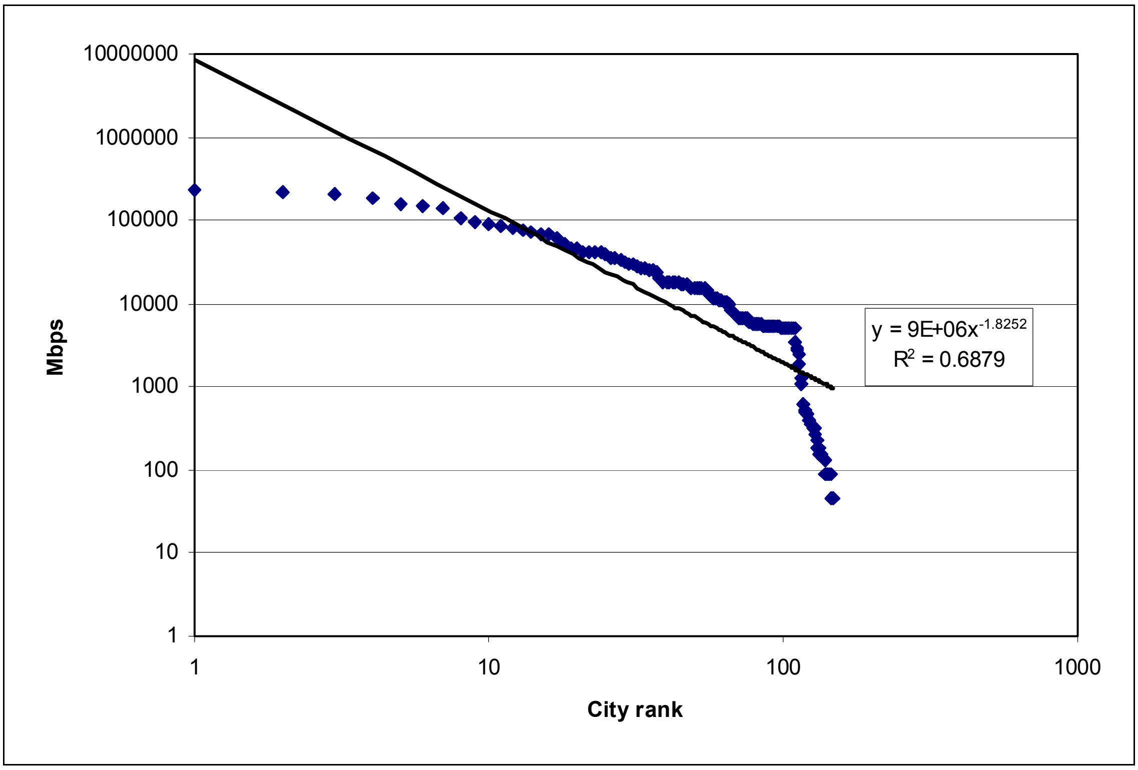

Figure 4. Power law distribution of the USA Internet 2000

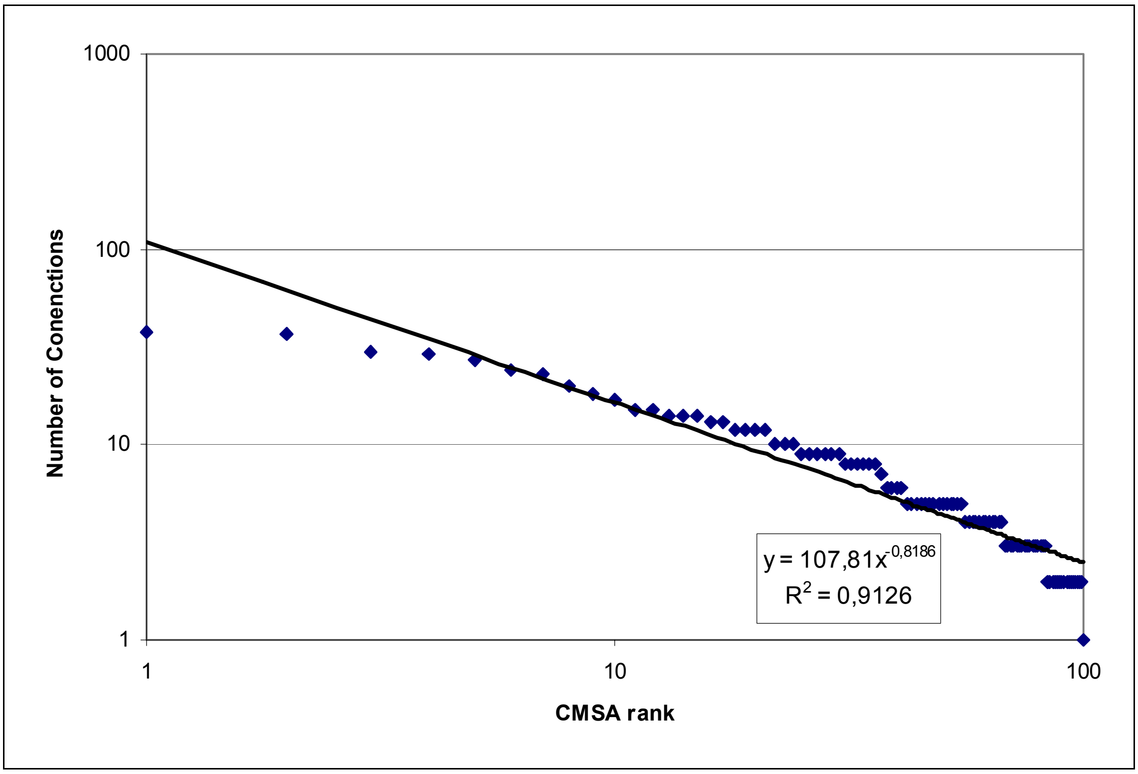

Figure 5. Binary connectivity distribution of the USA Internet 1998

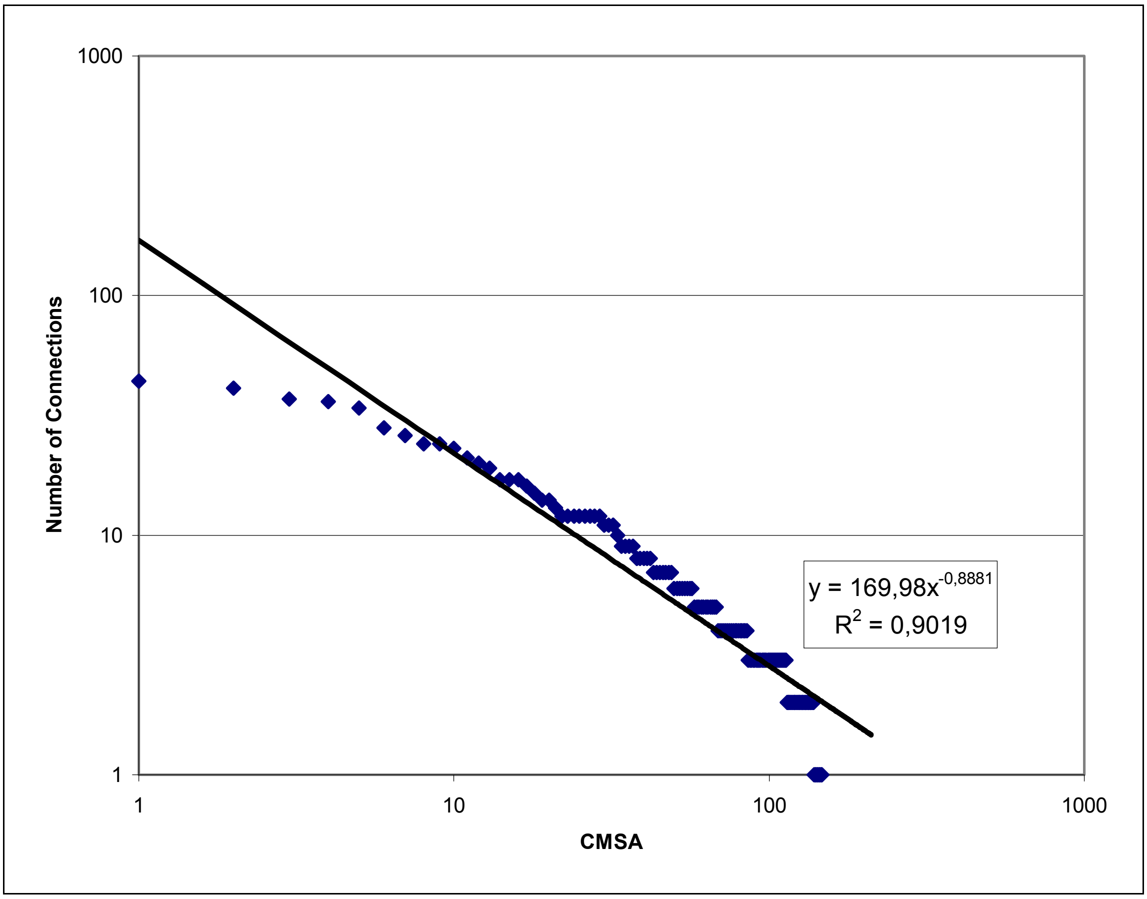

Figure 6. Binary connectivity distribution of the USA Internet 2000

3.3 Italian telephone network

The telecommunication system in Italy has tended to be under-developed. The SIP-Telecom Italia monopoly on landline telecommunications, the delay in the use of the Internet and the slow decay of the price for long-distance calls prevented a development similar to that in the United States or other European countries. In particular, the lack on competition on the market did not bring about descending prices for long-distance – and international – calls.

Only recently, Italy has started a path leading to descending prices and a liberalization of the market. This process is still on its way, and is also strongly supported by the European Union, in an effort to homogenize European telecommunication prices.

The database used in this study includes landline phone calls in Italy, for 101 districts.[3] The data cover the volume of outgoing phone calls by district and their cumulative length. Calls are divided into 4 types: urban, inter-district, international and intercontinental.

In order to verify whether the Italian telecommunication network, in our case represented by the volume of landline phone inter-district calls, qualifies as a free-scale network, we have to fit the data to a power law distribution. The expected exponent of the power law distribution for scale-free networks falls in between 2.1 and 3.

By creating an ascending rank of the telephone use data and plotting it on a log-log scale, we can compare the obtained curve to a theoretic power law distribution (best-fitting the data curve). The exponent of the power law function and the R^2 value will give an indication of the adaptation of the data to a scale-free network.

Figure 7 shows the plotting of the phone data on a log-log graph. The data are here interpolated with two functions: a power law function and an exponential function.

The graph shows that the data can be interpolated by a power law function having a power equal to about 0.65, while the R^2 value for the data adaptation is about 0.795. The exponent of the power law function is much lower than the value expected for scale-free networks, which is usually found between 2.1 and 3.

The data fits best to an exponential function.[4] The R^2 value of 0.986 shows a very high degree of correlation of the phone data to the curve. This result suggests that long-distance calls are distributed in Italy more as an exponential function than as a power law.

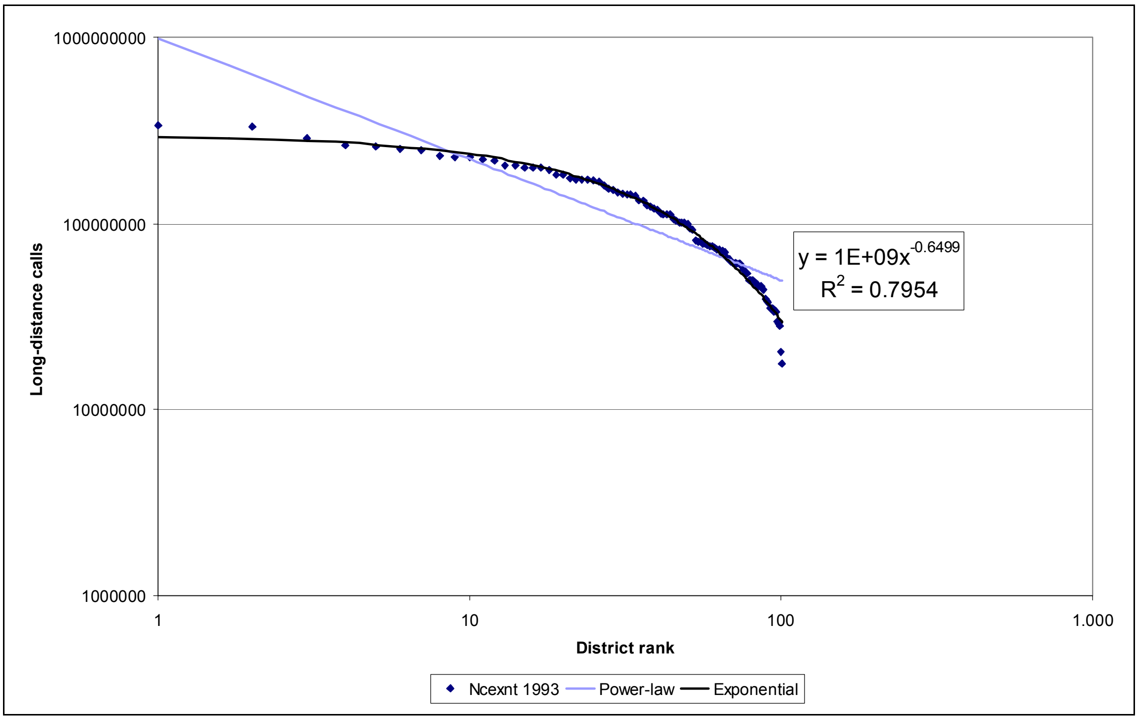

Figure 7. Log-log plotting of Italian landline phone calls for 1993

The exponent of the interpolating power law function does apparently not change over time (see Figure 8). Although the level of calls is increasing, log-log power laws from different years run indeed parallel. This result is due to the strong relationship between number of phone calls and the population size of the cities.[5] In fact, similar exponents for the power law function can be found plotting city population in the same way as for telephone traffic. Similar results have been obtained using data for cumulative phone time (instead of volume of calls) and different types of calls, like urban, international and intercontinental calls. Splitting the database in order to verify whether the North of Italy had a different behavior did not generate significant results, as a similar exponent for the power law function was found.

Concluding, we can say that inter-district phone calls in Italy do not show any appearance of scale-free behavior. Indeed, the distribution of phone traffic shows a distinct exponential shape, imitating the distribution characteristic of many phenomena, as for instance, city size.

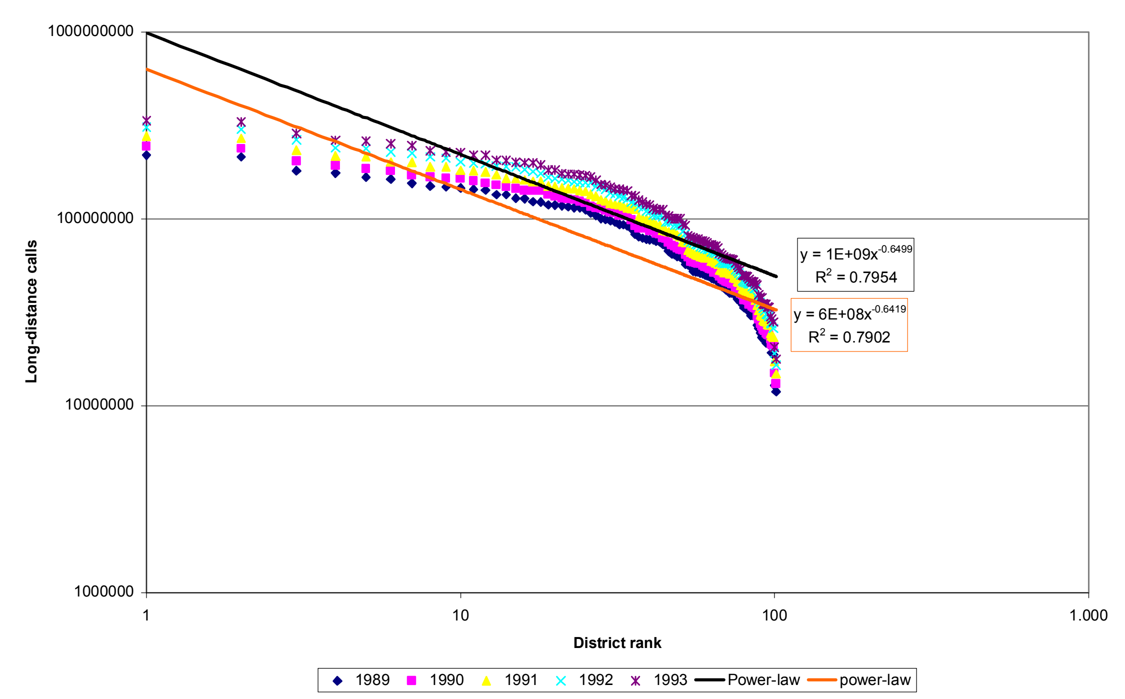

Figure 8. Italian phone calls over time (1989-1993) and interpolating power laws

3.4 Peer-to-peer sharing networks: a sample case

Peer-to-Peer networks (P2P) have been on a rising edge in the past years. Although it has been around for a while, only recently this networking model has been pushed to new heights. P2P networks can be seen as a model of a more efficient way for data exchange. In addition, applications of P2P networks in virtual enterprises and Business-to-Business operations are estimated to be potentially successful (Gorman and Malecki, 2002; Singh, 2001).

Since the inception of Napster, P2P has become the easiest way to share, exchange, find documents and files on the Internet. Endless software had seen birth in the last years, expanding the heritage of the famous P2P software and joining millions of users worldwide in large communities of peers.

More importantly, second-generation P2P software generally shares a new characteristic, which is the following. Information on available files on the network is not treated on centralized servers, as it was the case for Napster. It is instead available through the service given by private users who dispose of broadband connections and are willing to give up a small portion of bandwidth in order to serve as a node for the network.

Each computer serving as a main node to the network receives and passes on file requests – and availability information – to the computers that make a connection to the network through it. In some cases, each computer connected to the network contributes to the passage of the information. This flow expands at any passage, as each node of the network is transmitting file requests on the attached nodes. The process goes on for a predefined amount of levels.[6]

This kind of network seems to be structured in a small-world fashion. In order to properly analyze this issue, a snapshot of online users for a P2P network would be required. This kind of data would let us determine how the files usually available on P2P networks, amounting to thousands of Gigabytes, are distributed on the network's nodes, that is, the users.

In recent literature, Jovanovic *et al.* (2001) stress that P2P networks (the Gnutella network was used in their case study) have "strong small-world properties", displaying higher clustering coefficients and shorter characteristic path lengths compared to random networks and 2D meshes.[7]

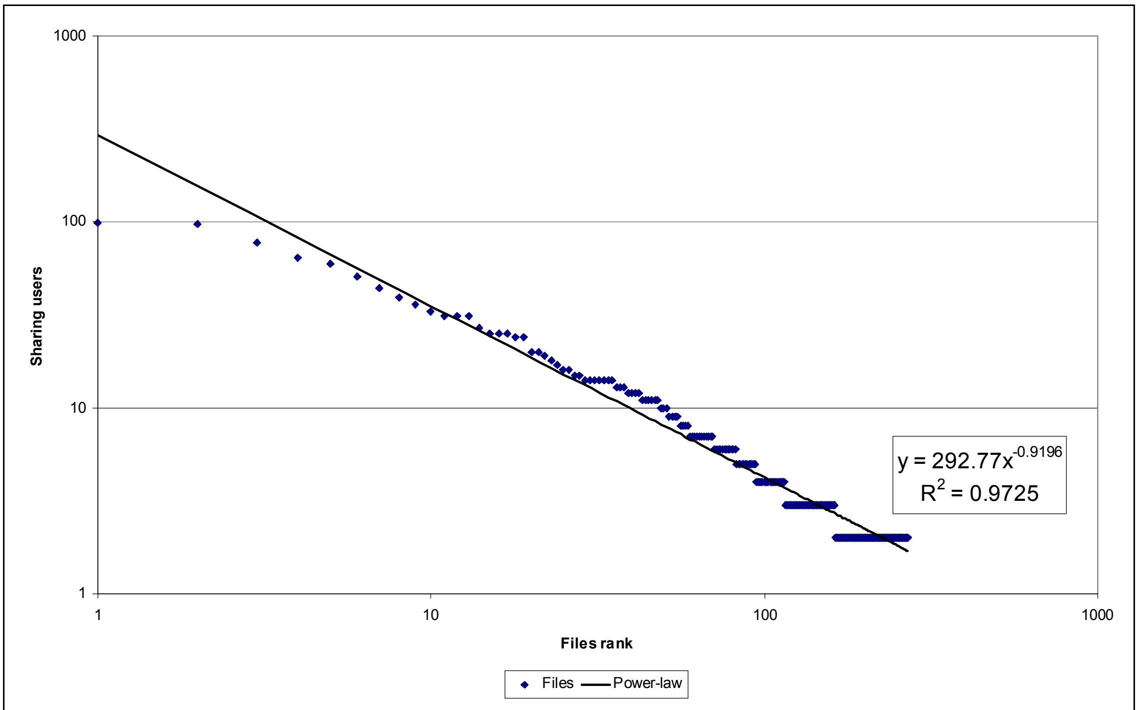

Figure 9. Distribution of shared files over the network users (2003)

The aspect of P2P networks that we will now investigate concerns the results of a file search on the network. For our experiment, we used a well-known P2P software and generated a returning list of files by typing a simple keyword and interrogating the network. The list of results obtained (in the month of April 2003) comprises 2801 file matches. More importantly, a portion of the matches – amounting to 270 files – is shared by multiple users.

The data set comprising the files shared by at least two users will now be analyzed. Results range from the most common file, which is shared by 99 users, to 101 files shared by just two users. By creating a descending rank size rule of the amounts of users sharing the matching files, we can plot the data on a two-axis graph (see Figure 9).

The log-log plot of our data shows that the files' sharing seems to follow a power law distribution.[8] The interpolating function's R^2 value equals to 0.9725, showing good approximation to the data, while the function's equation shows an exponent value of -0.9196. The good fitting of the model is confirmed by analysis of residuals, which show normal distribution.[9]

As we explained above, the data set generating the graph in Figure 9 only comprised files shared by at least two users. Allowing in the remaining cases (single-user shared files), which anyway are the majority of the search results (2531 results on 2801), brings to severe change in the function shape (see Figure 10). The interpolating function's fitting (R^2) decreases down to 0.6441 and the power exponent is only –0.4101.[10] This change in the results can be explained by a simple consideration. The files that are exchanged on the network – especially since we used a keyword for our search – are of a limited number. The remaining files can be considered peripheral in the perspective of the network.[11] If each user's available files are mostly a result of past exchanges on the network, unique files show scarce interest by the network. Furthermore, the choice of the exchange source, for an incoming file request, is not casual. This is usually guided by "preferential connectivity", explainable here as a "the more, the better" behavior. The user searching for a document or file will likely choose the file that is shared by as many users as possible – and obviously still respecting the search criteria.

Although the power law's exponent in this experiment is low in comparison to scale-free networks (see Section 3.2), it should be noted that the search results over the peer-to-peer network do not represent a 'topological' network, but are just an indirect result of it. The present experiment only stands as a first explorative step towards a more in-depth analysis of P2P communication networks and the data exchange within.

A second – and better – step might be the analyses of the users connected to the network. A snapshot of the network's online users and their file availability would permit to identify the hypothetical presence of super-connectors – users disposing of large bandwidth and archives – who serve a large part of the network demand or equity in file sharing.

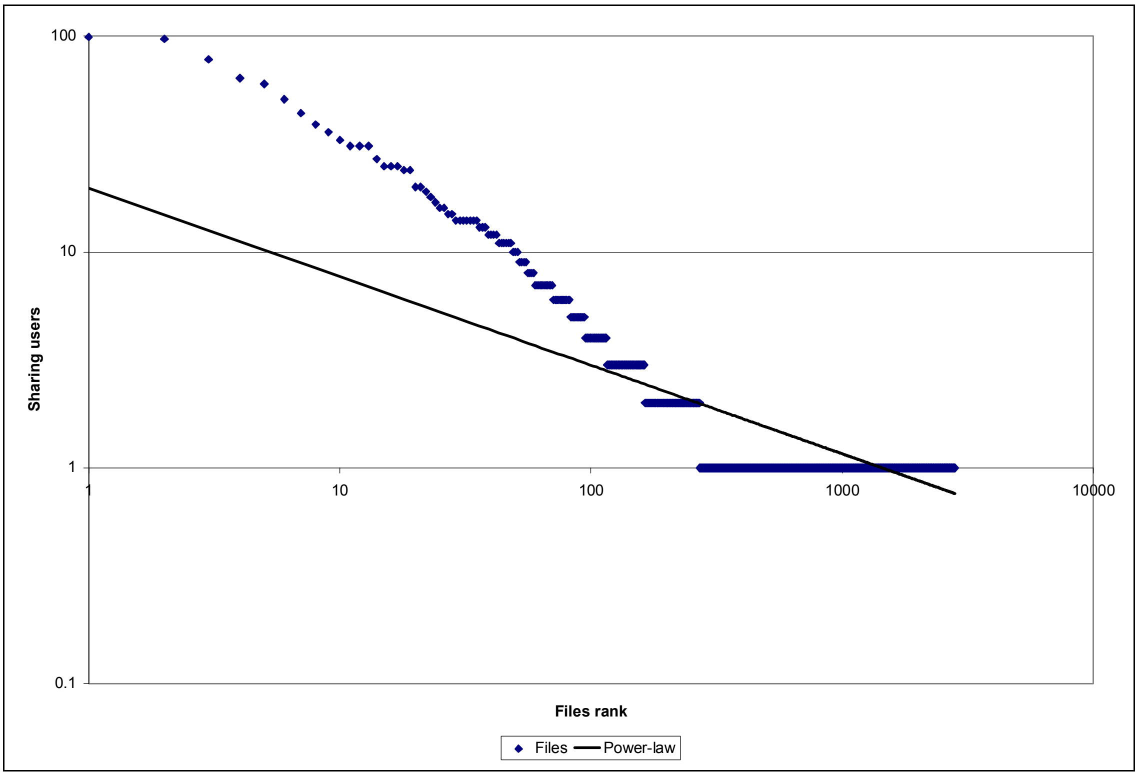

Figure 10. Files' sharing comprising single-user shared files (2003)

3.5 Concluding remarks

The experiments presented in this section demonstrate how "small-world" network properties are not universal across all networks. While a power law distribution was found to exist under certain conditions in a particular peer-to-peer data exchange network and was generally discovered in the US IP backbone network, small-world network properties were not found in the Italian phone traffic network and they appear to have diminished over time in the US IP backbone network. In fact, the Italian phone network, at least with the network indicators used, appears to more closely match an exponential function. These conclusions raise some interesting empirical, theoretical and methodological issues, and demonstrate the need to conduct a more extensive cross-Atlantic comparison of networks in terms of small-world network properties.

4 Conclusions: Towards a Cross-Atlantic Agenda

The results of the experiments undertaken in the previous sections are preliminary and they highlight the need for new data and methodologies to conduct cross-Atlantic comparisons of communications networks. One question that surfaces from the analysis presented here is why in certain cases scale-free properties exist and why in other communications networks they don't, and whether or not the influence of these factors may vary geographically. Would one see differences in the findings when comparing similar networks between the United States and Europe? Thus there is a need for research into this conditioning factor of scale-free features of complex systems.

Furthermore, there are several factors that may contribute to disparities across networks and that may require the use of different data sets on such complex phenomena. First, the underlying spatial structure of the infrastructure – in particular, the question whether it does have aspects to it that are planar – will likely play a prominent role in the ability of a network to evolve gradually into a small-world network. While the logical IP network for the United States may look similar to that in Europe, there may be differences in the spatial layout of the physical fiber in each region. Data on the European IP backbone network and the location of physical telecommunications infrastructure would help facilitate a more thorough study of this issue. Second, there may be social, economic or cultural factors that may contribute to differences in the findings across geographical regions. For example, the macroeconomic structure of Italian telecommunications – in particular, the question whether there is a monopoly, is much different from that in other countries in Europe or the United States. Do these and other factors significantly contribute to differences across regions? And third, there is the question of the quality and appropriateness of data that is currently available to undertake a small-world network analysis. In the case of the telephone network in Italy, outgoing landline calls were used to capture network dynamics indirectly, but in order to measure this more directly inter-district or point-to-point flows would be more accurate and appropriate. Or in the case of peer-to-peer data exchange networks, perhaps a breakdown of the networks by region would allow for a more fruitful comparison of these types of networks by region.

Clearly, there are also methodological issues to be resolved. The case study of the US IP backbone network highlighted the fact that the power law distribution methodology may be too simple for weighted networks. In fact, to date only a few studies have addressed the challenge of developing a robust and statistically sound technique for examining scale-free properties in weighted networks. This should also be explored in future studies.

Footnotes

[1] Efficiency in this case refers to the network characteristic of a large number of nodes having a low diameter.

[2] 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".

[3] Main cities (Milan, Turin, Rome and Naples) are here divided in several districts. In addition, a database using aggregated metropolis has been used, providing similar results.

[4] The equation of the exponential function interpolating the data (for 1993) is $y = 3 * 10^8 e^{-0.0229x}$, while the power-law equation is $y = 10^9 x^{-0.6499}$.

[5] A second factor of importance may be economic activity creating more than proportional phone traffic, because of intensive office activity.

[6] The Time-To-Live (TTL) parameter defines how deep (how many hops) the information and requests of the user will go through the network. For instance, in the case of Gnutella, probably the most known P2P software by now, peers usually keep a default value of 7 hops, which generally provide link up to ten thousand peers and a million files (Gorman and Malecki, 2001).

[7] In order to retrieve the Gnutella network's topology, Jovanovic et al. (2001) employed a distributed network crawler based on the software protocol. The Gnutella network's parameters for clustering and characteristic path length are compared to the ones generated by both a random graph and a 2D mesh of the same size.

[8] The equation for the interpolating power-law function (Figure 1) is as follows: $y = 292.77x^{-0.9196}$.

[9] Komogorov-Smirnov normality test carried out on the residuals confirms the goodness of the model.

[10] The resulting equation (Figure 2) is now: $y = 19.77x^{-0.4101}$.

[11] A similar aspect was investigated by Xerox Palo Alto Research Center (PARC). They discovered that up to 50% of the search results on a Gnutella network is actually provided by the top 1% hosting peers, somehow centralizing the sharing process and probably letting peripheral peers serve for less mainstream file requests (Gorman and Malecki, 2001; Adar and Huberman 2000).

Acknowledgements

The European authors wish to thank their STELLA colleague Carlo Tesauro (IPIGET, Naples, Italy) for kindly providing the Italian data set on phone network.

References

Abello J., P.M. Pardalos and M.G.C. Resende (1999), On Maximum Cliques Problems in Very Large Graphs. In: Abello J. and J. Vitter (Eds.), *External memory algorithms*. DIMACS Series on Discrete Mathematics and Theoretical Computer Science, vol. 50, American Mathematical Society, pp. 119-130.

Adamic L.A. and B.A. Huberman (1999), Growth Dynamics of the World Wide Web. *Nature* 401, p. 131.

Aiello W., F. Chung and L. Lu (2000), A Random Graph Model for Massive Graphs. *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pp. 171-180.

Albert R. and Barabasi A.-L. (2000), *Topology of Evolving Networks: Local Events and Universality*. Physical Review Letters 85, p. 5234.

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

Albert R., H. Jeong and A.-L. Barabasi (1999), The Diameter of the World Wide Web. *Nature* 401, pp. 130-131.

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.

Barabasi A.-L. (2001), The Physics of the Web. *Physics World*, July 2001. (http://www.physicsweb.org/article/world/14/7/09).

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

Barabasi A.-L. and Albert R. (1999), Emergence of scaling in random networks. *Science* 286, pp. 509-512.

Bollobás B. (1985), *Random Graphs*. Academic Press, New York.

Broder A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins and J. Wiener (2000), Graph Structure in the Web. *Comput. Networks* 33, pp.309-320.

Chen Q., C. Hyunseok, R. Govindan, J. Sugih, S. Schenker and W. Willinger (2001), The Origin of Power Laws in Internet Topologies Revisited. Submitted for publication.

Erdos P. and Renyi A. (1960), *On the Evolution of Random Graphs*. Publication of the Mathematical Institute of the Hungarian Academy of Science, Vol. 5, pp. 17-67.

Faloutsos C., P. Faloutsos and M. Faloutsos (1999), On Power-Law Relationships of the Internet Topology. *Proceedings of the ACM SIGCOMM Sept. 1999.*

Faloutsos M. and C. Faloutsos, Analysis of the Internet Topology. DARPA NMS, no. 00-1-8936 (http://www.cs.ucr.edu/~michalis/PROJECTS/NMS/TALKS/kickOff.ppt).

Ferrer i Cancho R., R.V. Solé (2001), Santa Fe working paper 01-03-016.

Gorman S.P. and E.J. Malecki (2000), The Networks of the Internet: an Analysis of Provider Networks. *Telecommunications Policy* 24, pp. 113-134.

Gorman S.P. and E.J. Malecki (2002), Fixed and Fluid: Stability and Change in the Geography of the Internet. *Telecommunications Policy* 26, pp. 389-413.

Govindan R. and H. Tangmunarunkit (2000), Heuristics for Internet Map Discovery. *Proc. IEEE Infocom 2000*, Tel Aviv.

Hayes B. (2000), Graph Theory in Practice: Part II. *American Scientist* 88 (02), pp. 104-109.

Jeong H. (2003), Complex Scale-Free Networks. *Physica A* 321, pp. 226-237.

Jeong H., B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabasi (2000), The Large-Scale Organization of Metabolic Networks. *Nature* 407, pp. 651.654.

Jeong H., S.P. Mason, Z.N. Oltvai and A.-L. Barabasi (2000), Lethality and Centrality in Protein Networks. *Nature* 411, pp. 41-42.

Jovanovic M.A. and F.S. Annexstein, Kenneth A. Berman (2001), Scalability Issues in Large Peer-to-Peer Networks - A Case Study of Gnutella. University of Cincinnati Technical Report.

Kumar R., P. Raghavan, S. Rajalopagan and A. Tomkins (1999), Applications of linear algebra in information retrieval and hypertext analysis. *Proceedings of the 9th ACM Symposium on Principles of Database Systems*, p.1.

Latora V. and M. Marchiori (2002), Is the Boston Subway a Small-World Network? *Physica A* 314, p. 109.

Liljeros F., C.R. Edling, L.A.N. Amaral, H.E. Stanley and Y. Aberg (2001), The Web of Human Sexual Contacts. *Nature* 411 (21), pp. 908-909.

Malecki E.J. (2002), The Economic Geography of the Internet's Infrastructure. *Economic Geography* 78.

Montoya J.M. and R.V. Solé (2000). cond-mat/0011195.

Moss M.L. and A. Townsend (2000), The Internet Backbone and the American Metropolis. *The Information Society* 16, pp. 35-47.

Newman M.E.J. (2001), *Proc. Nat. Acad. Sci. USA* 98, p. 404.

Radoslavov P., H. Tangmunarunkit, H. Yu, R. Govindan, S. Schenker and D. Estrin (2000), On Characterizing Network Topologies and Analyzing their Impact on Protocol Design. Tech Report 00-731, University of Southern California, Dept. of CS.

Redner S. (1998), How Popular Is Your Paper? An Empirical Study of the Citation Distribution. *Euro. Phys. Journ. B* 4, p. 131.

Schintler L.A. and R.G. Kulkarni (2000), The Emergence of Small World Phenomenon in Urban Transportation Networks: an Exploratory Analysis. In Reggiani A. (Ed.), *Spatial Economic Science: New Frontiers in Theory and Methodology*, Springer-Verlag, Heidelberg, Germany, pp. 419-434.

Singh, M.P. (2001), Peering at Peer-to-Peer Computing. *IEEE Internet Computing* 5, 1: pp. 4-5.

Tangmunarunkit H., R. Govindan, S. Jamin, S. Schenker and W. Willinger (2001), Network Topologies, Power Laws, and Hierarchy. Submitted for publication.

Watts D.J. (2003), *Six Degrees: the Science of a Connected Age*. W.W. Norton, New York.

Watts D.J. and S.H. Strogatz (1998), Collective Dynamics of Small-World Networks. *Nature* 363, pp. 202-204.

Zipf P.K. (1949), *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.

Annex A

Table A.1. Power law applications in recent literature. Table adapted from Albert and Barabasi (2002).

Authors	Year	Exp.	Network
Faloutsos M. and C. Faloutsos	1995	2.83	WWW (pair of nodes within h hops)
Faloutsos M. and C. Faloutsos	1997	2.15	WWW (frequency of outdegree)
Faloutsos M. and C. Faloutsos	1998	0.74	WWW (outdegree of Internet nodes)
Faloutsos M. and C. Faloutsos	1998	0.48	WWW (eigenvalues of adjacency matrix)
Redner S.	1998	3	Papers' citations
Kumar R. *et al.*	1999	2.1	WWW
Adamic L.A. and B.A. Huberman	1999	1.8	WWW (distribution of documents on domains)
Barabasi A.-L. and R. Albert; Amaral L.A.N.	1999-2000	2.3	Network of movie actors
Abello J. *et al.*; Aiello W. *et al.*	1999-2000	2.1	Telephone-call network
Montoya J.M. and R.V. Solé	2000	1.05; 1.13	Ythan estuary; Silwood park
Jeong H. *et al.*	2000	2.4	Protein (S. Cerev.)
Jeong H. *et al.*	2000	2.2	Metabolic (E.Coli)
Govindan R. and H. Tangmunarunkit	2000	2.4	WWW (router)
Broder A. *et al.*	2000	2.1	WWW
Jeong H. *et al.*	2000	2.4	Cellular networks
Ferrer i Cancho R. and R.V. Solé	2001	2.7	Words (occurrence)
Liljeros F. *et al.*	2001	3.4	Sexual contacts
Newman M.E.J.; Barabasi A.-L. *et al.*	2001	1.2; 2.1; 2.5	Science collaboration graph