Papers in Evolutionary Economic Geography

10.06

The Geography of Internet Infrastructure: An evolutionary simulation approach based on preferential attachment

Sandra Vinciguerra & Koen Frenken & Marco Valente

Utrecht University
Urban & Regional research centre Utrecht

http://econ.geo.uu.nl/peeg/peeg.html

The Geography of Internet Infrastructure:

An evolutionary simulation approach based on preferential attachment

Sandra Vinciguerra [1,2]

Koen Frenken [1,3]

Marco Valente [4]

[1] Section of Economic Geography
Urban and Regional research centre Utrecht (URU)
Utrecht University

[2] Department of Geography
Loughborough University

[3] Eindhoven Centre for Innovation Studies (ECIS)
School of Innovation Sciences
Eindhoven University of Technology

[4] Department of Economics, University of L'Aquila

31 March 2010

Abstract: We model the evolution of infrastructure networks as a preferential attachment process. We assume that geographical distance and country borders provide barriers to link formation in infrastructure networks. The model is validated against empirical data on the European Internet infrastructure network covering 209 cities. We successfully simulate the average path length and average clustering coefficient of the observed network. Furthermore, the simulated network shows a significant correlation with the observed European Internet infrastructure network. We end with a discussion on the future uses of preferential attachment models in the light of the literature on world cities and global cities.

The Geography of Internet Infrastructure:

An evolutionary simulation approach based on preferential attachment

1. Introduction

The marked differences in cities' connectivity through global transportation infrastructure networks have been considered as one of the main determinant of increasing income inequalities across cities and regions in the past two or three decades (Castells, 1996). Generally speaking, places that are well connected in transportation networks provide transportation services at lower costs due to scale economies at the hub and along the connections. At the hub, fixed costs can be recovered more easily from the large amount of traffic. Along the connections, traffic can be organised using larger transportation equipment. This is why the geography of network industries is generally characterised by a hub-and-spokes logic. Most places are connected to nearby hubs, which function as long-distance transfer points. In this manner, demand for transportation from multiple origins and to multiple destinations can be pooled and transmitted cheaply over large distance.

Connectivity in global transportation infrastructure networks can also be associated with 'world cities' (Friedmann, 1986) hosting multinational corporations that strategically locate in cities with high levels of connectivity as well as 'global cities' (Sassen, 1991) hosting globally operating producer service firms that support the organisation of global production systems. Multinational companies in general, and multinational producer service firms in particular, are dependent on high-quality access to transportation and information networks, while new investments in such networks are, in turn, often motivated by the presence of such companies or the wish to attract them (Keeling, 1995).

In this light, it may not come as a surprise that infrastructure networks, in particular airline and telecommunication networks, are increasingly being exploited to map the structure of the system of world cities and of global cities as well as to analyse changes herein over time (Gorman and Malecki, 2000; Gorman and Malecki, 2002; Moss and Townsend, 2000; Smith and Timberlake, 2001; Rutherford et al., 2004; Derudder and Witlox, 2005; Choi et al., 2006; Zook and Brunn, 2006; Derudder et al., 2007; Devriendt et al., 2008). Yet, the interpretation of the analyses of infrastructure networks in the light of the concept of world cities and the concept of global cities is far from straightforward (Taylor 1999; Derudder, 2006; Taylor et al., 2007). Data on transportation flows generally do not include information on origin and destination, which implies that the importance of cities with airports with many transfer passengers is over-estimated. Another problem concerns the uneven distribution of tourist passengers, which lead one to over-estimate the importance of cities that are popular tourist destinations (though some argue that tourist passengers should be taken into account, see Pirie, this issue). Furthermore, most airports serve a large region, or even a whole country, which makes airline passenger data less suitable to compare the connectivity of cities. This is especially problematic in the context of global cities, which mainly refer to the central business districts, while it is less of a problem for the world city concept, which refers to large metropolitan regions.

Despite these disadvantages of the use of transportation data, it is precisely the symbiotic relationship between the location of multinational firms and investments in infrastructure networks that legitimates the use of transportation data as a “proxy” of world cities and of global cities (Keeling, 1995; Smith and Timberlake, 2001). What is more, the data are generally more comprehensive and easier to obtain than data on multinational enterprises or globally operating producer service firms. The latter data are generally used to analyse in great detail the corporate networks among the largest centres worldwide, yet infrastructure have the advantage of covering cities of any size as long as such cities are part of the infrastructure network in question.

One of the central research questions in the literature on world cities and global cities is what determines the position of a city in the global urban network. One way of indicating such positions is by ranking the connectivity of cities, where connectivity can be based on inter-city corporate network data (Beaverstock et al., 2000; Alderson and Beckfield, 2004) or, as will be done below, on the basis of transportation network data (Cattan, 1995; Smith and Timberlake, 2001). One of the core questions is how one can explain the differences in connectivity between cities. Obviously, the connectivity of a city in infrastructure networks is generally closely related to city size. This can be explained by the fact that the connectivity is to be supported by local demand. However, this line of reasoning cannot explain the fact that important deviations exist in that some relatively small cities can obtain high connectivity and relatively large cities can remain poorly connected, nor can it deal with changes in connectivity of cities over short periods of time (Smith and Timberlake, 2001; Taylor and Aranya, 2008).

In an evolutionary perspective, the connectivity of cities can be said to be contingent upon small historical events early on in the emergence of a network. A city that gained a high connectivity early on is more likely to become a hub than a city that has gained little connectivity early on, *ceteris paribus*. This can be understood from preferential attachment: new cities entering an infrastructure network will prefer to create links with nodes that are already well connected as to profit from transfer opportunities. Below, we develop a simulation model of infrastructure evolution in which we explain the structure of an infrastructure network from an evolutionary process in which cities sequentially enter the network. The model is based on two elementary ideas. First, we assume that cities with high connectivity at one moment in time have a higher probability to increase their connectivity at the next moment in time, because new cities that enter the network, prefer to attach themselves to well-connected nodes. Well-connected nodes provide reliable services and many transfer opportunities at low costs due to scale economies. This principle is known as preferential attachment and implies that locations that enter the network early in time have a higher probability of becoming well connected than nodes entering later in time (Barabasi and Albert, 1999). Second, we assume that 'geography matters': as the costs of a network connection increases with geographical distance, the probability that a new node attaches itself to an existing node decreases with increasing geographical distance. We further assume that a network connection that crosses a national boundary is also more costly, as the planning and construction of a network connection is expected to be more costly if not one, but two different institutional regimes are involved. The barriers due to geographical distance and country borders will thus counteract the tendency towards an extreme hub-and-spokes structure, as these barriers will render the distribution of network connectivity more equal.

The remainder of this paper is organized as follows. In section 2 we introduce the simulation model. In section 3, the model is validated against empirical data on the Internet backbone network in Europe as observed in 2001 (Rutherford et al., 2004). The model successfully reproduces both the average path length and average clustering coefficient for a single set of parameters. We end in section 4 with a discussion on the future uses of preferential attachment models in the light of the literature on world cities and global cities.

2. A spatial model of network evolution

2.1 The Barabasi-Albert model

We start from the Barabasi-Albert model of evolving networks (Barabasi and Albert, 1999). In this model, a network grows through the addition of nodes. At every time step, a new node is created, which attaches itself stochastically to another node with a bias towards better connected nodes. More precisely, the probability that the new node *i* attaches itself to an existing node *j* is proportional to the latter's connectivity, a mechanism referred to as 'preferential attachment'. We get:

$$\prod_{i \to j} = \frac{k_j}{\sum_j k_j} \qquad (1)$$

where k_j is the degree of node *j* (where degree means the number of links attached to a node).

What characterizes the preferential attachment model is that each time a node is added to the network, the network's topology changes and, as a consequence, the probabilities for attachment of future nodes change: a case of path dependence (Arthur, 1989; Page, 2006). This means that, on average, nodes entering early in time with end up having a higher connectivity than nodes entering late in time. Furthermore, among the nodes entering early in time large differences in connectivity occur. Nodes that are 'lucky' and obtain many links early on will continue to increase their connectivity and end up being 'hubs' with high connectivity, while nodes that are 'unlucky' and obtain few or no links will end up with low connectivity later in time.

This mechanism of preferential attachment is sufficient for the generation of power law degree distributions with exponents between 2 and 3. Such degree distributions are characterized by the presence of few nodes with a large number of links (generally called hubs), while other nodes will have just one or few links. It has been found that many real-world networks exhibit degree distributions that follow power laws, including hyper-link networks, citation networks, co-authorship networks, movie actor networks, telephone call networks, and many others (Albert and Barabasi, 2002; Newman et al., 2006).

The logic of preferential attachment also holds for the evolution of infrastructure networks. A city that gains a high connectivity early on in the growth of a network is more likely to become a hub than a city that has gained little connectivity early on in this process, *ceteris paribus*. This can be understood from preferential attachment, since cities that enter an infrastructure network will prefer to create links with nodes that are already well connected, as these nodes provide the maximum number of transfer opportunities and have the infrastructure in place at the node to handle many of such transfers efficiently. Indeed, the model of preferential attachment has been used to understand the network topology of infrastructure networks like airline networks (Barthelemy, 2003), road networks (De Montis et al., 2007) as well as urban growth stemming from trade networks (Andersson et al., 2003).

The Barabasi-Albert model, however, can be criticised on empirical grounds as many networks do not follow the predicted power law distribution (Newman et al., 2006). Most notably, the degree distribution of networks in which the cost of a link increases with some variable (such as geographical distance) tends to be much less skewed than networks in which the cost of a link is close to zero. This observation is highly relevant to transportation networks where the costs of creating a link indeed increase with geographical distance (Barthelemy, 2003).[1]

Another important difference between the original Barabasi-Albert model and transportation networks holds that, in the latter network, existing connections between nodes can be extended with additional capacity. This feature is common to transportation networks where investments can be made to increase the link capacity or link *weight* (as it is generally called). This is the case, for example, when a new railway track is constructed alongside an existing track between two train stations, or when a second airline is entering service between two airports. In the case of Internet infrastructure, such increases in the strength of an existing connection between two cities are frequent. In fact, it is generally much cheaper to increase the capacity of an existing connection by adding cables than digging into the ground to create a completely new connection between two cities. In the Barabasi-Albert model of preferential attachment, this means that at each time t, one node, whether an already existing node in the network or a new node entering the network, is selected randomly with equal probability. Given the node selected, a second node is determined following formula (1) to which the selected node will connect. In this setup, it becomes possible that two cities with one link among them, obtain an additional link. The resulting network will no longer be *unweighted* network as in the original Barabasi-Albert model (where links between nodes are either

[1] In airline networks, however, the cost of creating a new link is largely insensitive to geographical distance. Yet, the cost of operating an airline connection increases with distance as gasoline and personnel costs are roughly proportional to distance.

present or absent), but a *weighted* network where the number of links existing between two cities is expressed by the link weight.[2]

Other criticisms have been voiced against this model, especially for the simplicity of it. For example, the model does not allow for a redistribution of links (Albert and Barabasi, 2000). And, the increasing number of nodes that will connect to the same hub may well cause congestion (Holme and Kim, 2002). Yet, for our case of fibre-optic cables making up the Internet infrastructure network, these two criticisms do not hold; once a cable is installed is very unlikely that the same cable will be redistributed at a later moment in time. And, congestion at hubs plays no role in fibre-optic networks as it can be dealt with by increasing the computing power at the node.

2.2 Geography matters

Due to the cost structure of the Internet infrastructure as explained before, the costs of connecting two cities increases with geographical distance. It is therefore important to account for the effect of geographical distance on the probability two places connect (Andersson et al., 2003; Barthelemy, 2003; Ozik et al. 2004). We therefore assume the probability of city i to connect to city j to be inversely dependent on their geographical distance d_{ij} such that:

$$\prod_{i \rightarrow j} = \frac{k_j}{\sum_j k_j} \frac{1}{(d_{ij})^{\alpha}} \qquad (2)$$

[2] In weighted networks, the preferential attachment model can also be modified in such a way that the probability a node to connect to another node is not dependent on the latter's degree, but on the latter's strength. The strength of a node stands for the sum of weights of links coming from a node (Yook et al., 2001; Barrat et al., 2004). We also implemented this procedure, but the results were very close to the results discussed below. The reason why results are so similar stems from the fact that degree and strength are highly correlated.

with $\alpha \geq 0$. The probabilities are to be normalised such that the sum of all probabilities add up to one. Note that for $\alpha = 0$, we obtain again the original Barabasi-Albert model of preferential attachment as in equation (1).

National borders are also expected to constitute a barrier to connectivity between cities. The weight of a link between two cities at some geographical distance in two different countries is expected to be smaller than the weight of a link between two cities at the same geographical distance in the same country. One reason for this national bias holds that most demand for Internet traffic is domestic (Zook, 2005). A second reason holds that constructing cross-border infrastructure projects involve two different institutional systems, which render the planning and construction of such projects more costly. One can therefore speak of institutional distance between cities located in different countries, as a second form of distance in addition to geographical distance (Boschma, 2005).

To account for domestic biases in the Internet infrastructure network, we introduce a parameter γ, which takes on, in all simulations, the value of 1 if cities *i* and *j* are located in the same country, and which takes on the value of γ ($\gamma \geq 1$) for cities *i* and *j* being located in two different countries. The preferential attachment formula is then given by:

$$\prod_{i \to j} = \frac{k_j}{\sum_j k_j} \frac{1}{\left(d_{ij}\right)^{\alpha}} \frac{1}{\gamma} \qquad (3)$$

with $\alpha \geq 0$, $\gamma = 1$ for domestic connections, and $\gamma \geq 1$ for cross-border connections. The probabilities are normalised such that the sum of all probabilities add up to one. Note that for

$\alpha = 0$ and $\gamma = 1$, we obtain again the original Barabasi-Albert model of preferential attachment as in equation (1).[3]

In this way, we can tune the parameter γ for cross-border connections only as to evaluate the barriers encountered by cross-border operators. For example, if $\gamma = 2$, it means that the probability of a link added at some geographical distance is twice as low for cross border relations compared to domestic relations for which holds that $\gamma = 1$.

The final restriction of the Barabasi-Albert model we want to lift is the restriction that a new node cannot attach itself to a node that has zero connectivity. For this reason, the network structure that is simulated by the Barabasi-Albert model always consists of a single network component, in which all nodes are all linked, directly or indirectly. By contrast, in the early evolution of transportation networks there are often multiple network components, i.e. subsets of cities that are connected, but not inter-connected. These components only become integrated in a single network later in time. If one allows nodes to attach themselves to nodes without any connectivity, this restriction of the Barabasi-Albert model can be avoided, and subsets of interconnected nodes can emerge. We therefore add a positive constant *c* to the connectivity to a node, such that there is some positive probability that an isolated node with zero connectivity is chosen. The final equation becomes:

$$\prod_{i \to j} = \frac{(c + k_j)}{\sum_j (c + k_j)} \frac{1}{\left(d_{ij}\right)^{\alpha}} \frac{1}{\gamma} \qquad (4)$$

[3] The model is implemented in LSD (Laboratory of Simulation Development), an open simulation platform developed by Marco Valente (Valente, 2008).

Throughout the analysis, we keep c at 0.1 but any other small value would not fundamentally alter our results.

Finally, it should be noted that we purposefully ignore the effect of city size on the probability of a links being added. City size undoubtedly affects the connectivity of a node as repeatedly shown in spatial interaction models of the gravity type and related studies (cf. Alderson et al., this issue; Orozco-Pereira and Derudder, this issue). However, we aim to explore to what extent preferential attachment, geographical distance and country borders alone can explain the network structure, not only the interaction strength between each pair of cities as in gravity models, but also global network properties such as the average path length and the average clustering coefficient (as further explained below). In this manner, we can examine to what extent preferential attachment, geographical distance and country borders, as factors independent of city size, are driving the network structure. In future extensions of the model, city size differentials can be added, which will then further improve the model fit.

3. Model results

Figure 1 shows the Internet infrastructure network in 2001 in Europe based on the data made available by Rutherford (Rutherford et al., 2004), which in turn were based on data provided by www.telegeography.com. The number of nodes in the network adds up to 209 cities. The weight between two cities is expressed by the number of infrastructure providers active between two cities. This number will generally correlate closely with other network data such as the amount of traffic between two cities or the capacity between two cities. Though different companies operate on different routes, one can consider the Internet infrastructure as a single network as Internet exchange points allows different Internet service providers (ISPs) to exchange traffic between their networks through mutual peering agreements (without cost).

The network is weighted in the sense that more than one provider can be active between two cities. The network is expressed by a symmetric matrix with off-diagonal values being integer values ranging between 0 and the maximum number of providers (which is 14 for the Hamburg-Bremen connection). The average distance between cities that are connected by either one or more providers is only 196 kilometers. This already suggests a strong hampering effect of geographical distance on connectivity.

The sum of weights in the network adds up to 1276. This means that there are 1276 links, rounded 1300 links, to be allocated among cities. We do so by iterating the simulation model 1300 times. With each iteration of the model, we randomly select one of the 209 cities in the network with equal probability. We then determine probabilistically to which other city the selected city will connect following equation (4) with parameters α for all city pairs, and γ for cross-border city pairs.

We first simulated a network with the baseline parameters corresponding to the original Barabasi-Albert model of preferential attachment ($\alpha = 0$, $\gamma = 1$). This means that in this simulation geographical proximity and country borders have no effect whatsoever on the probability that the strength of a link between city *i* and city *j* increases with one link. Figure 2 shows the outcome of one simulation run. Clearly, the resulting network structure is very different from the real network structure as shown in Figure 1. A correlation analysis using the quadratic assignment procedure (QAP), available in UCINET software, indeed shows that the correlation value of 0.004 is insignificant (p-value = 0.3).[4]

We then simulated the model for different parameter values of α and γ. We first validate the model result against the observed data by looking at the average path length and the average clustering coefficient as an overall characterization of the main properties of a network (Watts and Strogatz, 1998). The average path length is given by the average number of steps along the shortest paths for all possible city pairs in the (unweighted) network. The average path length in the observed Internet infrastructure network in 2001 is 4.383, which means that, on average, each city is 4.383 steps away from any other city. Table 1 and Figure 3 provide the average path length of the simulated networks for different parameter settings. The values for each pair of values for α and γ are averaged over 10 simulations. Judging from the average path length of 4.383 for the observed network, the parameter couple ($\alpha = 3$, $\gamma = 4$) yields by far the best results with a value of 4.2839 with the second closest result being 4.0220 for parameter couple ($\alpha = 3$, $\gamma = 3$).

Table 2 and Figure 4 provide the average clustering coefficient for the simulated networks for different parameter settings, each averaged over 10 simulations. The clustering coefficient of a node is given by the number of links between its neighbouring nodes divided by the total number of links that could possibly exist between them. In other words, the clustering coefficient of a node is the fraction of possible triangles that are triangles, which yields a value between 0 and 1 (Watts and Strogatz, 1998). The average clustering coefficient is then given by the average of all 209 clustering coefficients. In the observed Internet infrastructure network, this value is 0.347 meaning that, on average, about 35 percent of all city pairs that

[4] The quadratic assignment procedure (QAP) is a test based on random permutation to compare the similarity of two networks. Contrary to standard correlation measures, it accounts for the inherent dependencies between social network data (Krackhardt, 1988).

connected to a particular city, are also connected among them. Again, we find that the closest value is found for the parameter couple ($\alpha = 3$, $\gamma = 4$). Since we can reproduce both network properties (average path length and average clustering coefficient) for the *same* set of parameters ($\alpha = 3$, $\gamma = 4$), the model is likely to capture the fundamental mechanism underlying the evolution of this transportation network (Windrum et al. 2007).

Looking at the QAP correlation between the simulated network obtained for $\alpha = 3$ and $\gamma = 4$ and the observed network data from 2001, we observe a highly significant correlation. For 10 simulations, we find that the QAP correlation is on average 0.332 with a p-value of 0.002, with a minimum value of 0.294 and a maximum value of 0.342. This means that the correlation that is found is robust despite the fact the evolution of the network in terms of the order of entry of cities in the network, is completely random. The robust results show that the structure of the network is determined for an important part by the location of cities and their relative distance, as well as by country borders between cities.

To further illustrate this result on the QAP regressions, Figure 5 shows out of the ten simulations the simulated network with the highest correlation (0.342, p-value = 0.002) and the simulated network with the lowest correlation (0.294, p-value = 0.002). For the simulated network with the highest correlation in Figure 5, we also provide in Figure 6 three snapshots of the state of the network at different time steps (after 50 periods, 650 periods, and 1300 periods). These snapshots give an impression how the network evolves over time during a simulation run.

Looking at the two final simulated networks in figure 5 and comparing these with the observed Internet infrastructure network from 2001, one can observe parts of the networks that are fairly well reproduced and other parts where the simulation does a poorer job. In particular, the simulations are relatively poor in predicting the connectivity among Eastern European cities. The simulations generate a dense network of connections between these cities, while in reality the connections are quite sparse. The over-estimation of the connectivity is understandable given the lower levels of income and productivity – and the resulting lower rates of usage of Internet – in Eastern European cities compared to other European cities. As productivity and income differentials are not taken into account in the simulation model, the model tends to over-estimate the connectivity of Eastern European cities. Another aspect of the network that is not particularly well reproduced are the network positions of London and Paris. These cities have in reality a high degree with high weights, while in the simulation results these cities tend to obtain degrees and weights that are only slightly higher than the average. This deviation follows from the fact that the simulation model is driven by preferential attachment only, ignoring the effect of city size differentials. Still, the simulations reproduce – apart from the general network properties (average path length and average clustering coefficient) – a large part of the inter-city patterns reading from the QAP correlations.

What the simulations show is that the preferential attachment logic of network growth, combined with barriers to connect due to geographical distance and county borders, is successfully reproducing a significant part of an infrastructure network. This result can be considered to be quite remarkable since we did not rely in any way on differences between city size or income, which are the main determinants in gravity models used to estimate flows between territories (Tinbergen, 1962). What the results indicate is that a large part of network structure is caused by the location of cities and country borders alone. The high density of connections in Western Europe, as found in the simulated networks and the real network, can thus be understood from the high density of cities in a relatively small region in Europe.

4. Discussion

The modelling framework introduced in this paper is meant as a first attempt to approach spatial infrastructures as an outcome of an evolutionary process in which each new link changes the probabilities of subsequent link to occur. As such, the approach can be considered as complementing the evolutionary programme in economic geography (Boschma and Frenken, 2006). Admittedly, the model of preferential attachment is a rather abstract way to understand the evolution of networks. It implies that the links of new nodes entering a network are driven by a preference to connect to nodes that are already well connected as to profit from its transfer opportunities to other nodes. Though simple, the principle of preferential attachment is theoretically well founded in the context of transportation networks, because nodes is such networks can only develop a high connectivity by virtue of the transfer opportunities they provide.

A particular feature of the Barabasi-Albert model that has not been explored here nor has it elsewhere (as far as we know), holds that one can simulate the effect of specific “historical events” (Arthur, 1989). For example, one can ask the question whether the entry of particular cities early in the network has had a long-lasting effect on the resulting network structure that emerged later on. Such an effect can occur as an early entrant has much more time to collect links from other cities. What is more, as its connectivity is growing over time, it becomes more attractive to link to for other cities.[5] Two types of analyses can then be carried out. First, one can assess ex post whether the early entrance of particular cities, as can be retrieved from historical records, have had a lasting effect on the structure of the network (cf. Polese, this issue). Such an analysis is useful to understand exceptional cases of cities that have a much higher (lower) connectivity than their location and relative distance to other cities would predict. Second, in a counterfactual exercise, one can analyse whether the network structure would have been significantly different if the order of entry that occurred historically, would have been different.[6]

In a more general sense, the logic of preferential attachment and geographical barriers can be viewed as a model to analyse urban networks as the outcome of an evolutionary process, in which urban economic growth opportunities arise from connectivity to other cities. This idea is rather novel to urban economists who hitherto concentrated on attributes of cities as determinants of growth, rather than on the specific network structure in which they are positioned. Notable exceptions are studies that model inter-regional knowledge spillovers as stemming from citation networks (Peri, 2005), innovation networks (Maggioni et al., 2007) and university-industry collaborations (Ponds et al., 2009) and simulation studies of urban growth based on the preferential attachment concept (Andersson et al., 2003; Andersson et al., 2006).

The evolutionary approach based on preferential attachment can also be applied to the world city concept and the global city concept using data on inter-city corporate networks that are

[5] In the particular case of the European Internet infrastructure network, such early entrants could be port cities connecting Europe to the United States or research cities setting up Internet connections early on.

[6] In both cases, one is in need of historical data on the entry time of each city, which is lacking in the particular case of the European Internet infrastructure network, which is why we did not engage in these exercises.

commonly used in this context. The preferential attachment modelling framework is versatile with respect to the inclusion of any form of distance that provide barriers for linkages between cities or, inversely, any form of proximity that provide particular opportunities for linkages between cities (cognitive, language, cultural, *et cetera*). The form of proximity that is central to world city and global city research concerns organizational proximity. Following Boschma (2005), organizational proximity, as the inverse of organizational distance, can be understood as the extent to which any two entities are under a common hierarchical control. Where the entities refer to cities, organizational proximity between two cities can then be measured as the number of firms with subsidiaries in both two cities, be it multinational firms in general (Friedmann, 1986) or multinational producer service firms in particular (Sassen, 1991).[7] One can expect that the more firms are present in both two cities, the more demand there exist for direct connections within a particular transportation infrastructure network. Or, equivalently, the fewer firms that are present in two cities, the larger the organizational distance between two cities, the less likely a new link will be added between the two cities. In this way, one can use the inter-city corporate network data to extend the simulation model of preferential attachment with an organizational distance variable.

[7] Indeed, this methodology has been underlying much of the empirical literature since Beaverstock et al. (2000) proposed this methodology.

Bibliography

Albert, R. Barabasi, A. L. (2000) Topology of evolving networks: Local events and universality, *Physical Review Letters*, 85(24), pp. 5234-5237.

Albert, R. Barabasi, A. L. (2002) Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74(1), pp. 47-97.

Alderson, A. S., Beckfield, J. (2004) Power and position in the world city system, *American Journal of Sociolog*y, 109, pp. 811-851.

Alderson, A. S., et al. this issue.

Andersson, C., Frenken, K., Hellervik, A. (2006). A complex network approach to urban growth, *Environment and Planning A* 38(10): 1941–1964.

Andersson, C., Hellervik, A., Lindgren, K., Hagsson, A., Tornberg, J. (2003) Urban economy as a scale-free network, *Physical Review E*, 68, 036124 Part 2

Arthur, W. B. (1989) Competing technologies, increasing returns, and lock-in by historical events, *The Economic Journal*, 99, pp. 116–131.

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

Barrat, A., Barthelemy, M., Vespignani, A. (2004) Weighted evolving networks: coupling topology and weights dynamics, *Physical Review Letters*, 92: 228701.

Barthelemy, M. (2003) Crossover from scale-free to spatial networks, *Europhysics Letters*, 63(6), pp. 915-921.

Beaverstock, J. V., Smith, R. G., Taylor, P. J. (2000) World city network: a new metageography? *Annals of the Association of American Geographers*, 90(1), pp. 123-134.

Boschma, R. A. (2005) Proximity and innovation. A critical assessment, *Regional Studies*, 39(1), pp. 61–74.

Boschma, R. A., Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography, *Journal of Economic Geography*, 6(3), pp. 273-302.

Castells, M. (1996) *The Rise of the Network Society*. Oxford: Blackwell.

Cattan, N. (1995) Attractivity and internationalisation of major European cities: The example of air traffic, *Urban Studies* 32, pp. 303-312.

Choi, J., Barnett, G., Chon, G. (2006) Comparing world city networks: A network analysis of Internet backbone and air transport intercity linkages, *Global Networks*, 6, pp. 81–99.

Derudder, B. (2006) On conceptual confusion in empirical analyses of a transnational urban network, *Urban Studies* 43(11), pp. 2027-2046.

Derudder, B., Devriendt, L., Witlox, F. (2007) Flying where you don't want to go: an empirical analysis of hubs in the global airline network, *Tijdschrift voor Economische en Sociale Geografie* 98(3), pp. 307-324.

Derudder, B., Witlox F. (2005) An appraisal of the use of airline data in assessing the world city network: A research note on data, *Urban Studies* 42(13), pp. 2371-2388.

De Montis, A., Barthélemy, M., Chessa, A., Vespignani, A. (2007) The structure of interurban traffic: a weighted network analysis, *Environment and Planning B: Planning and Design*, 34(5), pp. 905-924.

Devriendt, L., Derudder, B., Witlox, F. (2008) Cyberplace and Cyberspace: two approaches to analyzing digital intercity linkages, *Journal of Urban Technology* 15(2), pp. 5-32.

Friedmann, J. (1986) The world city hypothesis, *Development and Change* 17, pp. 69-83.

Gorman, S. P., Malecki, E. J. (2000) The networks of the Internet: an analysis of provider networks in the USA, *Telecommunications Policy*, 24(2), pp. 113-134.

Gorman, S. P., Malecki, E. J. (2002) Fixed and fluid: stability and change in the geography of the Internet, *Telecommunications Policy*, 26(7-8), pp. 389-413.

Holme, P., Kim, B. J. (2002) Vertex overload breakdown in evolving networks, *Physical Review E* 65(6), 066109 Part 2

Keeling, D. J. (1995) Transport and the world city paradigm, in: P. L. Knox and P. J. Taylor (eds.) *World Cities in a World-System*. Cambridge: Cambridge University Press, pp. 115-131.

Krackhardt, D. (1988) Predicting with networks: A multiple regression approach to analyzing dyadic data, *Social Networks*, 10, pp. 359-381

Maggioni, M. A., Nosvelli, M., Uberti, T. E. (2007) Space versus networks in the geography of innovation. A European analysis, *Papers in Regional Science*, 86(3), pp. 471-494.

Mahutga, M. C., et al., this issue.

Moss, M. L., Townsend, A. M. (2000) The Internet backbone and the American metropolis, *Information Society*, 16 (1) pp. 35-48.

Newman, M. E. J., Barabási, A.-L., Watts, D. J. (2006) *The Structure and Dynamics of Networks*, Princeton: Princeton University Press.

Orozco-Pereira, R. A., Derudder, B., this issue.

Ozik, J., Hunt, B. R., Ott, E. (2004) Growing networks with geographical attachment preference: Emergence of small worlds, *Physical Review E*, 69, 026108

Page, S. E. (2006) Path dependence, *Quarterly Journal of Political Science*, 1(1), pp. 87-115.

Peri, G. (2005) Determinants of knowledge flows and their effect on innovation, *The Review of Economics and Statistics*, 87(2), pp. 308-322.

Pirie, G., this issue.

Polese, M., this issue.

Ponds, R., Van Oort, F., Frenken, K. (2009) Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach, *Journal of Economic Geography*, in press, doi:10.1093/jeg/lbp036

Rutherford, J., Gillespie, A., Richardson, R. (2004) The territoriality of pan-European telecommunications backbone networks, *Journal of Urban Technology* 11(3), pp. 1-34.

Sassen, S. (1991) *The Global City*, Princeton: Princeton University Press.

Smith, D., Timberlake, M. (2001) World city networks and hierarchies, 1977-1997: an empirical analysis of global air travel links, *American Behavioral Scientist* 44 (10), pp. 1656-1678.

Taylor, P. J. (1999) So-called 'world cities': the evidential structure within a literature, *Environment and Planning A*, 31(11), pp. 1901-1904.

Taylor, P. J., Aranya, R. (2008) Global ‘urban roller coaster’? Connectivity changes in the world city network, 2000–2004, *Regional Studies* 42(1), pp. 1-16.

Taylor, P. J., Derudder, B., Witlox, F. (2007) Comparing airline passenger destinations with global service connectivities: a worldwide empirical study of 214 cities. *Urban Geography*, 28(3), pp. 232-248.

Tinbergen, J. (1962) *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: The 20th Century Fund.

Valente, M. (2008) *Laboratory for Simulation Development - LSD*, LEM Papers Series 2008/12, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy

Watts, D. J., Strogatz, S. H. (1998) Collective dynamics of ’small-world’ networks, *Nature*, 393, pp. 440-442.

Windrum, P., Fagiolo, G., Moneta, A. (2007) Empirical validation of agent-based models: Alternatives and prospects, *Journal of Artificial Societies and Social Simulation* 10(2), 8, <http://jasss.soc.surrey.ac.uk/10/2/8.html>.

Yook, S. H., Jeong, H., Barabási, A.-L., Tu, Y. (2001) Weighted evolving networks, *Physical Review Letters*, 86, pp. 5835-5838.

Zook, M. A. (2005) *The Geography of the Internet Industry: Venture Capital, Dot-coms and Local Knowledge*. Oxford: Blackwell Publishers.

Zook, M.A., Brunn, S.D. (2006) From podes to antipodes: Positionalities and global airline geographies. *Annals of the Association of American Geographers*, 96, pp. 471-490.

	γ=1	γ=2	γ=3	γ=4
α=0	2.4105	2.4159	2.4150	2.4312
α=1	2.4415	2.4779	2.5167	2.5481
α=2	2.7422	2.8769	3.0130	3.0984
α=3	3.5509	3.8423	4.0220	4.2839
α=4	5.1471	5.6037	6.1265	6.5047

Table 1. Average path length of simulated networks for different values of α and γ (average of 10 simulations per parameter setting)

	γ=1	γ=2	γ=3	γ=4
α=0	0.0880	0.0914	0.0940	0.0925
α=1	0.1083	0.1112	0.1263	0.1408
α=2	0.1707	0.1906	0.2234	0.2348
α=3	0.2682	0.2983	0.3288	0.3412
α=4	0.3406	0.3568	0.3706	0.3748

Table 2. Average clustering coefficient of simulated networks for different values of α and γ (average of 10 simulations per parameter setting)

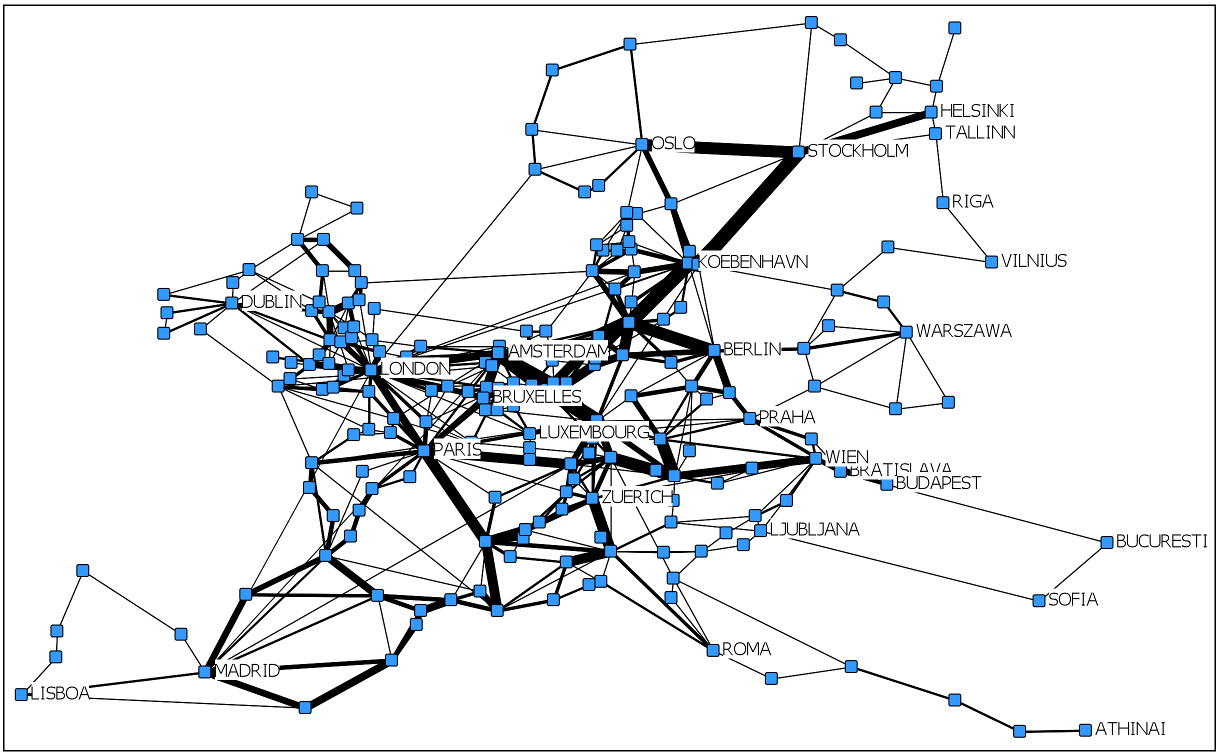

Figure 1. Graph representing the Internet Infrastructure network in 2001. The thickness of the lines (from 1 to 14) is directly proportional to the number of providers serving that connection, where 14 is the maximum number of providers present on the connection between Hamburg and Bremen.

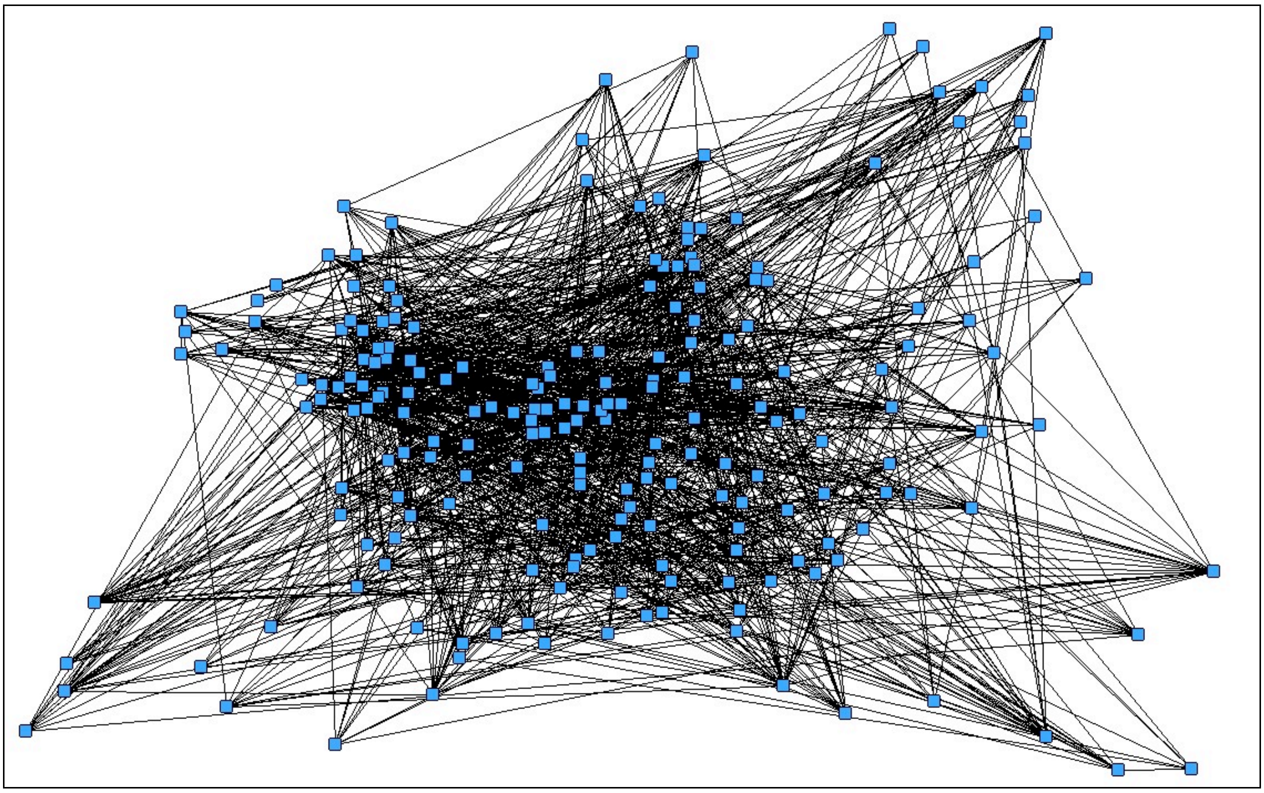

Figure 2. Simulated network for $\alpha = 0$ and $\gamma = 1$

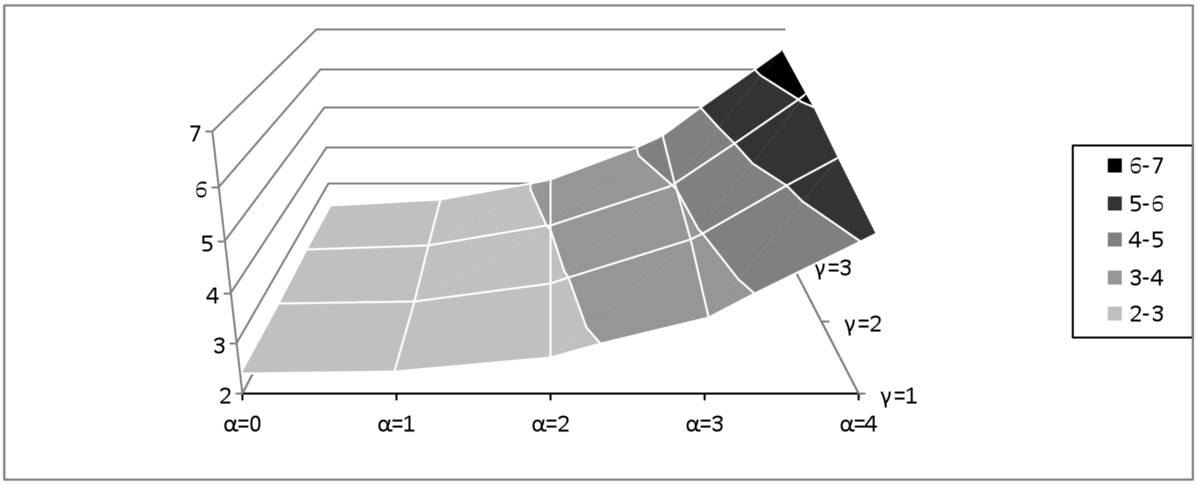

Figure 3. Average path length for simulated networks for different values of α and γ (values as in Table 1)

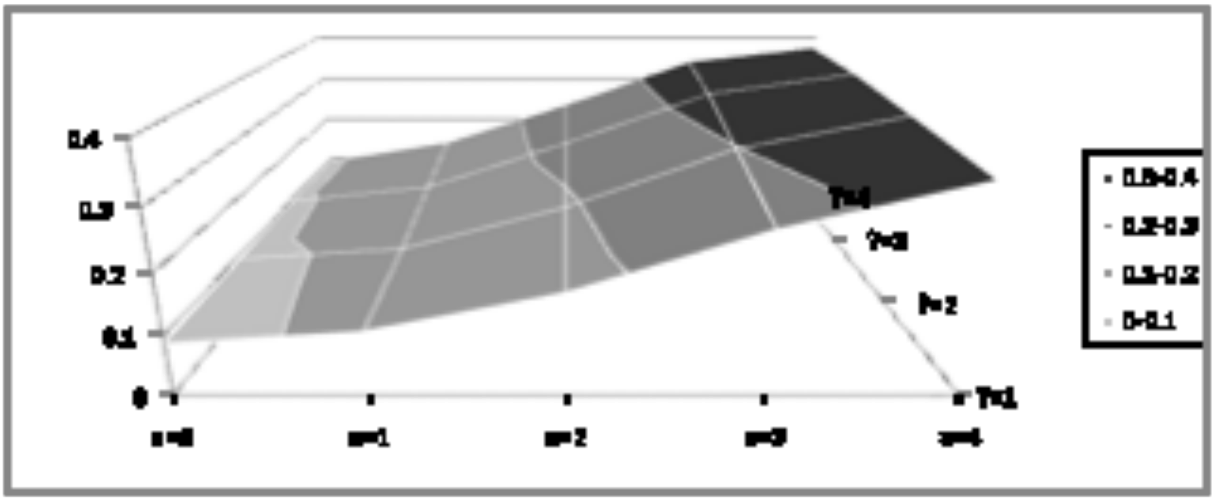

Figure 4. Average clustering coefficient for simulated networks for different values of α and γ (values as in Table 2)

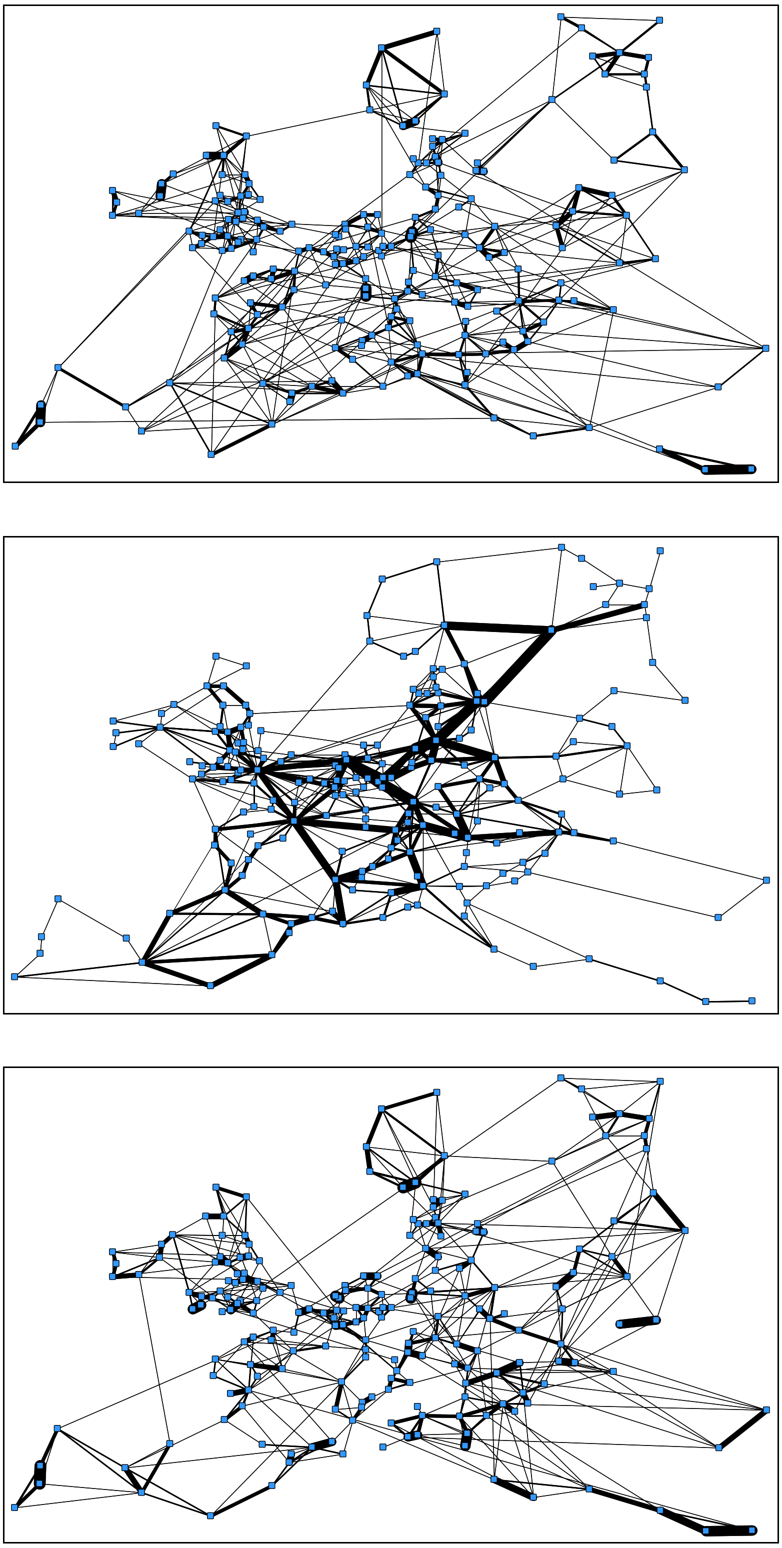

Figure 5: Comparison between simulated and real networks. In all three networks the thickness of lines is directly proportional to the number of providers serving that connection. All networks have the same scale from 1 to 16, where 16 is the maximum number of providers present on a connection in one of the simulated networks. Upper: α = 3, γ = 4, best simulation (QAP correlation = 0.342, p-value = 0.002), middle: observed Internet infrastructure network 2001, lower: α = 3, γ = 4, worst simulation (QAP correlation = 0.294, p-value = 0.002)

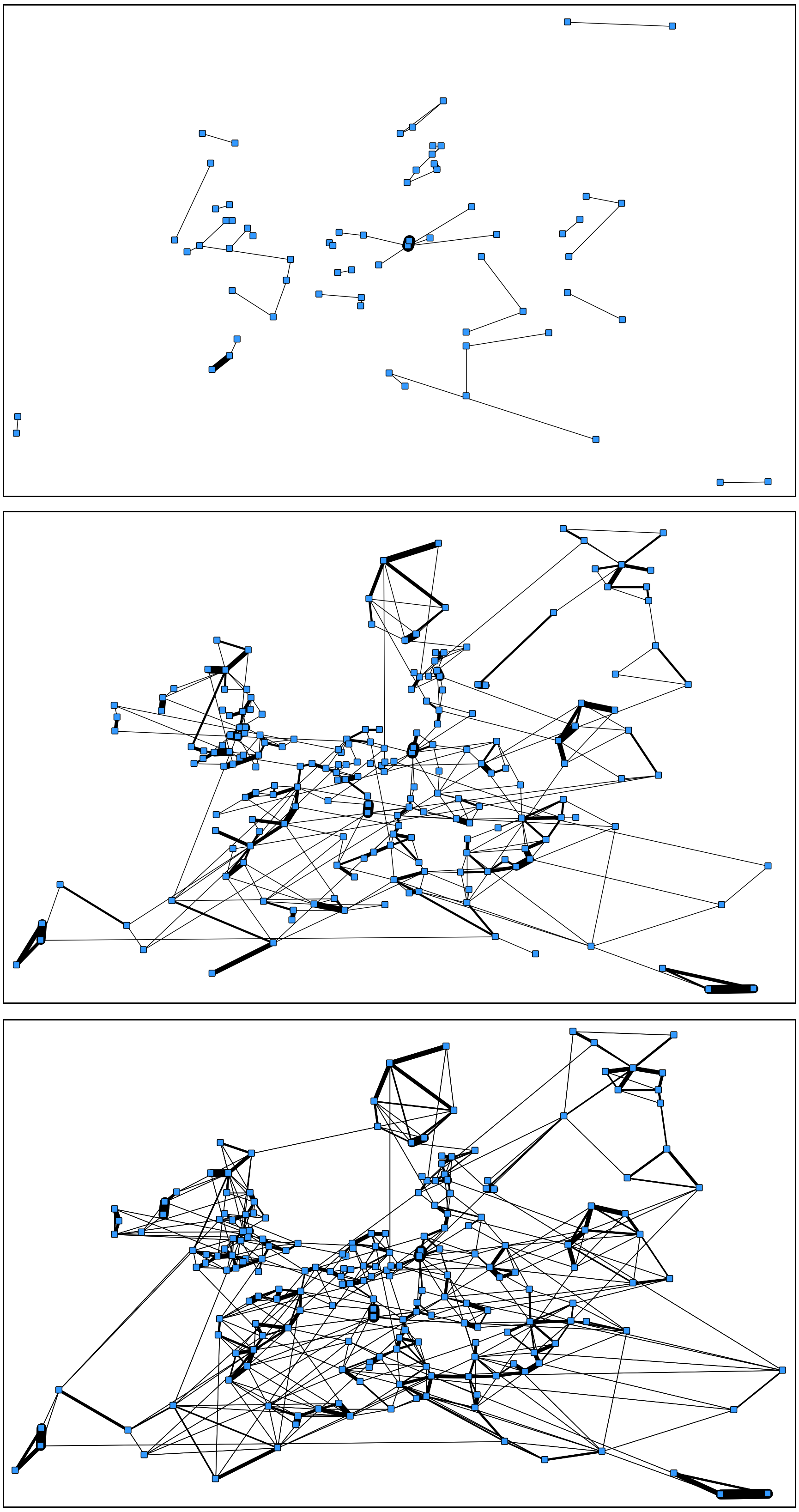

Figure 6: Example of a simulation run which represents the evolution of the network over time. Parameters are $\alpha = 3$ and $\gamma = 4$. Respectively, from top to bottom, networks show iteration at time step 50, 650 and 1300.