Network Analysis of Commuting Flows: A Comparative Static Approach to German Data

**Roberto Patuelli · Aura Reggiani · Sean P. Gorman ·
Peter Nijkamp · Franz-Josef Bade**

Abstract The analysis of complex networks has recently received considerable attention. The work by Albert and Barabási presented a research challenge to network analysis, that is, growth of the network. The present paper offers a network analysis of the spatial commuting network in Germany. First, we study the spatial evolution of the commuting network over time. Secondly, we compare two spatial interaction model (SIM) specifications, in order to replicate the actual network structure. Our findings suggest that the commuting network appeared to become more dense and clustered, while the SIMs seem to require more sophisticated specifications, in order to replicate such a connectivity structure.

Keywords Complex networks · Commuting · Infrastructure · Dynamics ·
German districts

1 Introduction

Network theory has received increasing attention over recent years, because of the implications that the structure and functioning of networks have for a wide set of

R. Patuelli (✉) · P. Nijkamp
Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105,
1081HV Amsterdam, The Netherlands
e-mail: rpatuelli@feweb.vu.nl

P. Nijkamp
e-mail: pnijkamp@feweb.vu.nl

A. Reggiani
Department of Economics, Faculty of Statistics, University of Bologna, 40126 Bologna, Italy
e-mail: aura.reggiani@unibo.it

S. P. Gorman
School of Public Policy, George Mason University, Fairfax, VA 22030-4444, USA
e-mail: sgorman1@gmu.edu

F.-J. Bade
Faculty for Spatial Planning, University of Dortmund, 44221 Dortmund, Germany
e-mail: franz-josef.bade@uni-dortmund.de

societal phenomena. Transportation is one of these. In particular, commuting networks, because of their association with residential choices and job location, are a fruitful research domain in this context.

Commuting behaviour has become a relevant research issue in many fields (Rouwendal and Nijkamp 2004). While it has long been studied in terms of forecasting and flows approximation (see, for example, White 1977, 1986; Fotheringham 1983), less efforts have been made in studying the structure and connectivity properties resulting from commuting. However, a growing literature is available that studies commuting in a spatial framework. Rouwendal (2004) introduces search theory and spatial behaviour in commuting choice modelling, while Ma and Banister (2007) analyse the relationship between urban spatial structure decentralization and average commuting distance. Cörvers and Hensen (2003) use regional modelling in order to study functional relationships between regions that maximize internal commuting. Spatial behaviour had also been introduced in more traditional approaches, as in the works by Fotheringham (see, mainly, Fotheringham 1983), who introduced the competing destinations models. This approach made it possible to introduce in traditional spatial interaction models (SIMs) an element representing the effects of the clustering of destinations, by means of accessibility measures. Network approaches to commuting have also been proposed, both at an urban level (Sheffi 1985; Sohn 2005) and a zonal level (Thorsen et al. 1999). A graph theory approach has instead been proposed by Binder et al. (2003).

The main objective of this paper is to investigate the changes that occur over time in commuting networks, as well as the associated relationships with the underlying network configuration. This is done—in a first stage—by considering commuting networks as graphs, where flows of commuters between two locations (seen as nodes in the network) represent a logical link between them, according to the recently developed theories by Albert and Barabási (2000, 2002). Our purpose is to compare the network connectivity models of Albert and Barabási with the conventional SIMs. As a case study, we consider the network of the commuting flows in Germany, for 1995 and 2004. In this framework, the connectivity properties of the network concerned will be first analysed in a graph theory perspective, since they influence the functioning, reliability and efficiency of the network itself. For example, a network moving toward a high centralization or denser clustering implies changes in the opportunity costs of commuters, and, consequently, in the future economic development of the areas of interest. Secondly, we will observe the network structure of commuting flows in Germany from the viewpoint of the network spatial interaction, that is, the spatial mobility. In this context, we will try to identify the appropriate deterrence form inherent to spatial interaction modelling. We will then carry out two specifications of the impedance function concerning an unconstrained SIM: (a) a power-law function; (b) an exponential function, in order to evaluate how well they simulate the underlying network structure observed for the real data.

The paper unfolds as follows: Section 2 reviews the main aspects of complex network theory, in the light of the Albert and Barabási developments. In Section 3, we discuss two SIM formulations employed in the paper, as well as their interpretation in terms of preferential attachment. Section 4 describes the empirical application that was carried out. First, a discussion of the data is presented (Section 4.1), followed by an exploration of the data (Section 4.2). Section 4.3

presents the network analysis of the commuting network, for 1995 and 2004. Subsequently, Section 4.4 compares the two specifications of unconstrained SIMs described earlier in terms of their connectivity properties. Finally, in Section 5, we draw some conclusions and outline future research directions.

2 Complex network theory

This section briefly reviews the main issues related to complex network theories, and in particular their implications for transportation networks. While complex networks have had considerable attention in recent years, the study of such networks is not particularly new. Before Albert and Barabási's discoveries, original research had in fact already been carried out some 40 years ago by Erdös and Renyi (1960), whose major assumption was an underlying random network structure. However, because of insufficient computational power and suitable data, for the majority of the twentieth century, these random theories formed the basis for the most common methods of network simulation, although they were not adequately challenged (Barabási 2001). Recently, Albert and Barabási (2002) found that (large) complex networks were actually behaving according to three main characteristics:

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

In detail, short average path length indicates that any two nodes on the network can be reached with a limited number of hops. High clustering, on the other hand, occurs because of nodes locating topologically close to each other in cliques that are well connected to each other. This property had been formalized by Watts and Strogatz (1998). Finally, the frequency distributions of node density (or, more generally, number of connections) are called degrees and can follow power-law and exponential distributions. This third property implies connections that cut across the graph, directly linking different clusters of vertices. These direct links between clusters bring an increased level of efficiency—in terms of number of hops—to the network. This result shows the limits of the Erdös and Renyi models, in which the exponential decay of the degree distribution did not imply a higher number of connections available to the most important nodes. The novelty in the Albert and Barabási approach was incorporating an additional component: network growth. Consequently, not only can the number of nodes in the network increase, but new nodes are found to have a higher probability of connecting to other nodes that are already well-connected (preferential attachment).

A certain amount of literature is now available on the analysis of transportation networks in terms of complex theory (Reggiani and Schintler 2005). Because of their short average path length, airline networks have been considered by Amaral et al. (2000) as a *small-world* network, referring to the model presented by Watts and Strogatz (1998). On the other hand, the same authors note that the structural limitation of airline networks, such as the limited space available in the airports, may hinder the emergence of scale-free properties. Other authors have found similar results. Latora and Marchiori (2002) analysed the Boston subway network, while

Schintler and Kulkarni (2000) observed congested road networks. Both articles found small-world network properties in the analysed networks.

Generally, it might be argued that transportation networks are less prone to evolve into a scale-free structure over time, given the fact that they tend to be planar. In fact, in planar networks, the maximum number of connections for a single node can be limited by the physical space available to connect it to other nodes, and it is this fact that makes the large number of connections needed for finding a power-law distribution more difficult to obtain. Further, it may be observed that highly centralized transportation networks can be subject to threats to their viability, in the case of the destruction of large hubs (Kwan et al. 2003). Scale-free networks have many implications, but a far-reaching consequence of their unique hub structure is that they are very fault tolerant, while also susceptible to attack (Albert et al. 2000). Specifically, a scale-free network model remains connected when up to the 80% of nodes are randomly removed from the network, but when the most connected nodes are removed, the average path length of the network increases rapidly, doubling its original value when the top 5% of nodes are removed (Albert et al. 2000). In short, targeting the most-connected nodes can cause significant damage to a scale-free network, making it highly susceptible to a coordinated and targeted attack. Spatial analysis of network failure has also been done for airline networks, finding similar results for the Indian airline network (Cliff et al. 1979).

Starting from these considerations, the next section will present the SIMs that were modelled as approximations of preferential attachment, in order to compare them—in the framework of spatial mobility—with a scale-free model inspired by the theories described above.

3 The spatial interaction model as an approximation of preferential attachment

3.1 Spatial interaction models for identifying commuter flows in the German labour market network

Spatial interaction models are arguably one of the most common methods employed and studied for estimating commuting flows (see, recently, Thorsen and Gitlesen 1998; Johansson et al. 2003; Jörnsten et al. 2004). Generally, SIMs have long been a popular technique for describing and explaining behavioural, demographic and economic phenomena in space (for an extensive presentation of the family of methods, see Sen and Smith 1995).[1] The main reason for the widespread utilization

---

[1] The most common specification of SIM has its origins in analogy with Isaac Newton's law of universal gravitation. The idea of utilizing models derived from this theory had already been introduced, in the nineteenth century, in the field of social sciences by Carey (1858) and Ravenstein (1885), and subsequently mathematically formalized by Stewart (1941). Remarkably, SIMs have been shown to have theoretical justification in entropy theory and in utility maximization/cost minimization (see, for example, Nijkamp 1975; Nijkamp and Reggiani 1992). While Isard (1960) first suggested the use of SIMs in regional science, the entropy root of SIMs introduced by Wilson (1967; 1970) and, subsequently, the micro-economic derivation introduced by McFadden (1974, 1979) contributed to make SIMs more suitable to interpret spatial-economic phenomena.

of SIMs is their simple mathematical form, in addition to the intuitive assumptions underlying the approach.

The common form of a SIM (here presented as double-constrained) is as follows:

$$T_{ij} = A_i B_j O_i D_j f(\beta, c_{ij}), \quad \text{for } i = 1, ..., I; \ j = 1, ..., J, \quad (1)$$

where:

$$A_i = 1 \bigg/ \sum_j B_j D_j f(\beta, c_{ij}); \quad (2)$$

$$B_j = 1 \bigg/ \sum_i A_i O_i f(\beta, c_{ij}). \quad (3)$$

T_{ij} measures the flow of interaction between the origin i and the destination j, depending on the stock variables O_i and D_j, as well as on the deterrence function f (β, c_{ij}), and on the balancing factors A_i and B_j (see Reggiani 2004).

The deterrence function in Eq. (1) depends on the deterrence factor β and the interaction costs c_{ij}. The variable c_{ij} might also be considered as generalized costs. In our experiment, distances were used as a proxy of the interaction costs (such as congestion), since the analysis was carried out at the German district level (*kreise*). The functional form of the deterrence function is also a relevant issue. While in its first formulations the distance deterrence function was shaped as a power-law function—as used in the Newtonian formula—Kulldorf (1955) showed that an exponential deterrence function seemed to better fit migration phenomena. Subsequently, the exponential deterrence form emerged mathematically from the entropy maximization approach developed by Wilson (1967).

In our analysis, both the power-law and exponential specifications were used. The power-law form shows a larger amount of flows—compared with the exponential form—in the presence of long distances or travel times. In addition to the shape of the deterrence function, the value of the β deterrence factor was researched for both specifications (see Section 4.4). In detail, we adopted two unconstrained[2] SIM forms, specified as follows:

$$T_{ij} = K E_i E_j d_{ij}^{\beta}; \quad (4)$$

$$T_{ij} = K E_i E_j e^{\beta d_{ij}}. \quad (5)$$

In Eqs. (4) and (5), the flows T_{ij} are the employees commuting from the origin district i to the destination district j. They are a function of the number of persons E_i and E_j employed[3] in the two districts, as well as of the distance d_{ij} between the two,

---

[2] We are aware that a doubly constrained model is a better specification of Eqs. (4) and (5). However, we used it as a first approximation in our experiments. The next phase of the research will certainly contemplate the use of the doubly constrained SIM (see also Section 5).

[3] It should be noted that the use of E_j is formally correct according to spatial interaction theory, since it is proportional to the inflows D_j. Concerning the outflows O_i, the use of the variable E_i is a necessary approximation due to data availability.

in addition to a scaling factor K. The models that we propose are, of course, overly simple. However, what is most relevant for our experiments is not the exact estimation of the German commuting flows, but the connectivity structure of the network that underlies the numerical data (see Section 4.3).

When employing a SIM for estimating inter-urban commuting flows, additional issues should be cited. One of them is the treatment of internal commuting. In particular, the distance between the working and living areas is, by definition, null (although travel time or costs would not necessarily be). This issue is at times solved by assigning an arbitrary value to the distance for internal commuting. Alternatively, the flows assigned to internal commuting can be omitted in the analyses. A number of additional ways to treat internal commuting are available in the literature. The method suggested by Thorsen and Gitlesen (1998) starts from the consideration that intra-commuting might imply different transportation means, such as biking or walking. Thorsen and Gitlesen suggest an additional component to be added to the deterrence function exponent. This component would represent—depending on the case—either a start-up (generalized) cost for commuting between different zones, or a premium, expressing the benefit of intra-commuting. An example model with these characteristics, reminiscent of the Champernowne deterrence function (see, for example, Sen and Smith 1995), is presented by Thorsen and Gitlesen (1998, p. 279) for a double-constrained specification. Alternatively, the authors suggest that labour market characteristics might be used to influence the elements on the diagonal of the O/D matrix.

In our case, the elements of the diagonal are omitted from the analysis. This choice was taken mainly due to the network approach to commuting identified in the paper. As we analyse the connectivity and structural properties of the German commuting network, the measure of the number of commuters within a certain district would not add additional information about the network, apart from the 'socio/economic weight' of a certain node. On the other hand, the total number of employees in each district already embraces this aspect.

3.2 Interpretation of spatial interaction behaviour as preferential attachment

The usual practice in the use of SIMs, when dealing with commuting flows, is to employ the models in forecasting future flows, given certain conditions. In our experiments, we propose the utilization of the simple power-law-specified SIM shown in Eq. (4) as a tool for approximating the connectivity and structural properties of a commuting network, as opposed to the more mainstream and studied exponential specification [Eq. (5)]. In particular, we want to verify if a SIM can allow for preferential attachment behaviour. In the models introduced by Barabási and Albert, nodes have a higher probability of connecting to other nodes that are already well-connected. The hypothesis that we will test in the next section is that commuting networks follow a similar preferential attachment-based behaviour in terms of connectivity and structure. They would not be the first transportation network to be referred to in these terms. In fact, hub-n-spoke networks operated by airlines are a well-known example of preferential attachment behaviour (see, for example, Bowen 2002, and, most importantly, Wojahn 2001).

An additional reason for the consideration of commuting networks in such a framework can be found if we think of preferential attachment as a maximization of

utility levels. The idea is that utility is maximized by connecting to the most-connected nodes of the network, as they give access to other points in the network by a minimal number of hops (therefore minimizing generalized costs). If so, this hypothesis would be consistent with the theoretical basis of utility maximization that justifies the use of SIMs. In particular, the hub-n-spoke network might—conceptually—be interpreted as a network tree consistent with a nested logit/hierarchical SIM structure (for the compatibility between the nested-logit and double-constrained SIM, see Nijkamp and Reggiani 1992).

Given these premises, the next section presents the empirical application carried out in this paper. We first describe the data available for the experiment (see Section 4.1), and subsequently carry out a statistical exploration of the data (Section 4.2). In Section 4.3, a network analysis of the commuting data and the SIM results follows. Finally, an analysis of the results of the SIMs is presented in Section 4.4).

4 Empirical analysis: commuting networks over time

4.1 The data and the experiments on SIMs

The primary data employed in the analyses presented in this paper consist of information on German commuting flows. In detail, for each origin–destination (O–D) pair (i, j), the number of commuters is represented by the number of employees living in district i, and working in district j. These are therefore home-to-work data, which are available for 1995 and 2004. The level of disaggregation, at district (*kreise*) level, allows for the analysis of 441 districts, and corresponds to a NUTS-III classification. It should be noted that, in this data set, Berlin is still classified as two separate districts (coded 11100 and 11200), while the latest classification includes them in one larger district (11000). A similar *caveat* should be made about the Hannover area, as the main city district and the its surrounding region are kept separate in the data set, although they belonged to the same joint district in recent years.

Further, the SIMs estimated for our experiment employ, as shown in Section 3.1, two types of data:

a. The number of employees working in each German district (both part- and full-time). Data were collected—as part of a yearly survey—in the year 2004.
b. The distance between each pair of origins and destinations. This is expressed in kilometres, and acts as a proxy for more effective measurements, such as cost or travel time.

The next section will therefore analyse the connectivity properties of the observed commuting network, and of those developed by the models presented in our exercise.

4.2 A statistical comparison of commuting flows for 1995 and 2004

A first step in our network analysis is to statistically explore the commuting data introduced in the previous section. In particular, we focus our attention on the

statistical comparison of the observations collected for the first and the last year of the data set.

Concerning the first point, the four graphs in Fig. 1 show the distribution of the commuter flows, for both 1995 and 2004, at different scales. While the top-left graph shows the entire range of the flows, the remaining graphs reduce the visualization to flows lesser than 70,000, 10,000 and 1,000 commuters (bottom-right), respectively. A high correlation between the observations for the 2 years can be observed. This result might imply a certain stability in the relationships between centres, or the absence of dramatic changes in transport infrastructure over the period considered, which are among the prime determinants of variations in commuting patterns. As we observe the smallest ranges of commuters, more spreading starts to be seen. The R^2 obtained regressing the data for the year 2004 on those for 1995 decreases from 0.975 to 0.898, when considering the whole range of flows or only those with fewer than 1,000 commuters.

In the top-left graph of Fig. 1, the most visible outlier, which has the highest number of commuters, represents the commuting flows between the two formerly separate East and West Berlin. Workers who live and work on opposite 'sides' of Berlin seem to have increased in the 9-year period, which seems to be a reasonable and somehow expected finding, considering that the data for the first year conceived, 1995, were collected only 5 years after reunification. Therefore, it makes sense to expect a gradual redistribution of residential and business location
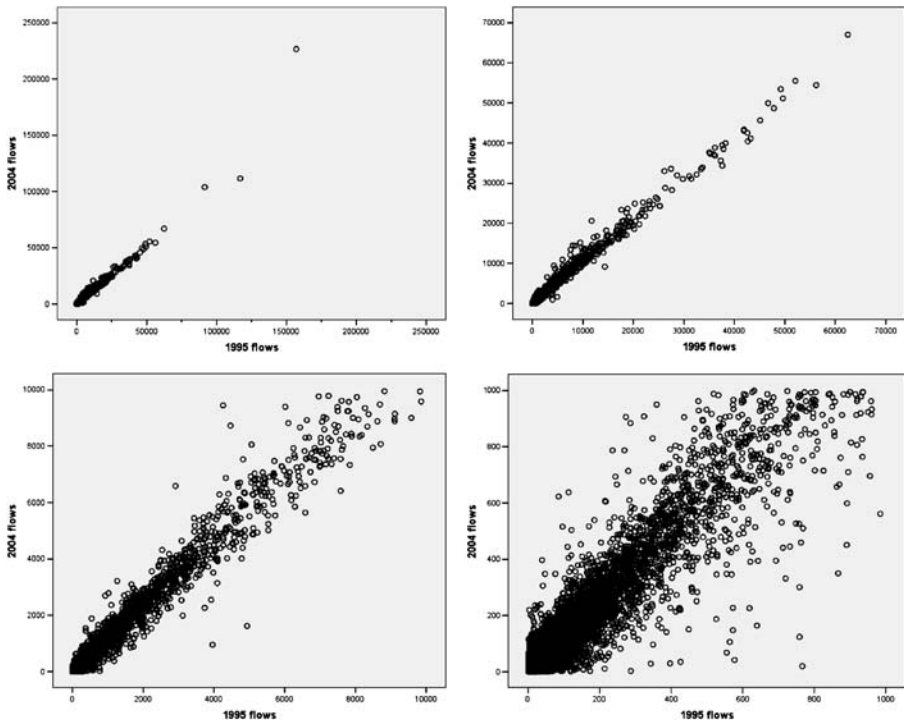

Fig. 1 Scatter plot of commuting flows in 1995 and 2004, at different scales

Table 1 Statistical exploration of the commuting flows data sets, for 1995 and 2004

Year	Maximum value	Sum	Mean	Mean 2004/1995 ratio	Median 2004/1995 ratio
1995 flows	156,999	8,616,362	107.59	1.670	1.214
2004 flows	226,700	9,944,326	118.53		

choices on both sides of Berlin, This might be particularly true for relocations into the formerly Soviet side of the city, where rents are, or were, supposedly cheaper (see, for example, Kemper 1998). As explained in the previous section, the two Berlin districts are kept separate in this data set. Although it could be considered non-realistic to separately analyse areas that, as a matter of fact, belong to the same city, this allows us to observe the huge amount of mobility that has been generated, within the city, by the reunification.

In addition to the above considerations, a more in-depth exploration of the data is necessary. Table 1 summarizes the statistical results obtained for the two data sets. The summary statistics show the change in the commuting flows over the years. Consequently, the total number of commuters also increased, by 15.41%. The average number of commuters per O–D pair increases from about 108 to about 119. However, the large increase in the commuting flows within Berlin suggests that the 15.41% increase may not be distributed uniformly over the network. The difference between the mean and the median of the ratio statistics for the two data sets seems to suggest an increased disparity in the distribution of flows, as they stand at 1.67 (mean) and 1.21 (median).

Having statistically explored the data in this section, the next section will examine, from a network perspective, the implications and properties of the logical links derived from the data set analysed.

4.3 Network analysis: the results

We want to examine the network structure underlying the commuting flows data. In order to do so, we may consider each O–D pair (i, j) as a link between nodes i and j, within our commuting network. The nodes of the network are, therefore, the districts the commuters travel to and from on the network.

The commuting flows are translated in a network structure by means of a simple procedure. Each O–D pair that has at least a given number of commuters (for example, (1) contributes to generate a vertex index, which is a counter of the number of links that attach to one or another node of the network. The final product of this operation is a ranked list of the nodes (districts) in the network, ordered according to the number of connections they enjoy. For example, the presence of commuters on the link between Munich and Rostock increases by one the number of connections of both districts. The threshold for the minimum amount of flows to be observed on each link in order to be valid is, of course, subjective. For computational reasons, we set this threshold at ten commuters for the analysis that follows. Once the vertex list is complete, a publicly available graph software such as Pajek[4] can visualize the

---

[4] Available at http://vlado.fmf.uni-lj.si/pub/networks/pajek.

resulting network and compute a set of parameters that characterize it. Table 2 presents the results obtained by computing this set of network parameters for the data collected in 1995 and 2004.

The network parameters computed for the two data sets provide rather interesting results. In both years, the networks appear to be dense, with an exceedingly large number of connections for the rather small number of nodes (districts). In the roughly 10 years between the two samples, the commuting network became more clustered, more efficient, and less centralized. The growth observed is fascinating, since the network did not only become more efficient, but it also grew more resilient and diverse. The lower diameter of the network implies that commuters can cross the network in a more efficient manner (lesser hops through the nodes), while the lower betweenness means that the load across any one link in the network has decreased, from both a maximal and average perspective. The result for the diffusion of link frequency (betweenness), which decreased in 2004, might be an indicator of congestion avoidance, where job seekers tend to seek employment in locations that have less difficult commutes. Generally, the results would seem to indicate that the structure is becoming more hub-like, despite the increasing density of the network. The average degree and minimum degree of the network both increased, and the standard deviation in connectivity rose as well. However, there is no dramatic shift, as these are all small changes.

The above considerations should be *caveated* by the fact that the network we are examining is purely *logical*, and it does not represent the physical links on which the commuters are travelling. A one-hop trip on the network (for example, from Munich to Rostock) could hardly be a single link in reality, but would instead be a trip crossing many physical links and nodes between the two locations.

The changes outlined above can also be seen in a ranked plot of the number of connections by district, shown in Fig. 2. In addition, the decay of the number of connections seems to better fit an exponential distribution, rather than a power-law distribution, which is an indicator of a tendency toward scale-free, preferential attachment behaviour. When counting connections for the entire data set, a large number of nodes seem to enjoy connections to all other nodes in the network. If we consider a list of the most- and the least-connected districts for the 2 years (see

Table 2 Network parameters computed for the commuting networks, for 1995 and 2004

Index	1995	2004
Clustering coefficient	367.448	378.468
Weighted clustering coefficient	223.696	232.359
Diameter	3	2
Average degree	113.832	133.909
Standard deviation Degree	72.053	79.219
Max degree	430	437
Min degree	25	28
Centralization	3.76%	2.65%
Betweenness mean	163.098	153.045
Betweenness std. dev.	457.403	368.555
Max betweenness	3787.278	2705.86

For the definitions of the main network parameters, see Appendix A, Table 4

Fig. 2 Log–log plot of the connections of German districts, for observed (1995 and 2004) and estimated data (2004)
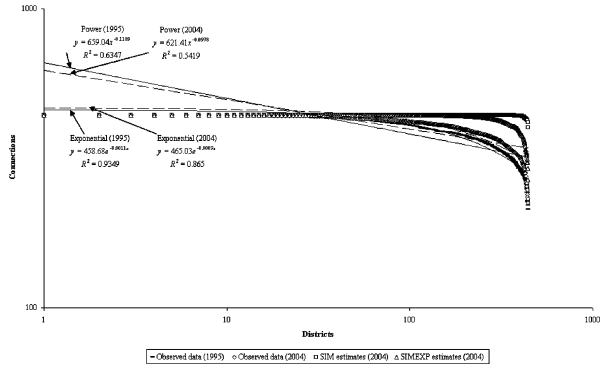
Appendix B, Table 5), and we follow the district classification by the BfLR/BBR[5] (Böltgen and Irmen 1997), we can see that most of the districts with the highest number of connections belong to type 1, that is, 'central cities in regions with urban agglomerations'. Conversely, the least connected districts are mostly classified as types 8 or 9, which refer to regions with rural features. However, the connections are still quite high in the rural regions as well, showing a rather good connectivity structure over the entire network. The high value of the connectivity shows indeed that we are dealing with more general connectivity rather than daily commuting. The findings presented here—a high number of fully connected districts and a slow decay of the number of connections—can be explained by the limited number of nodes (districts) in the network, and by one of the conditions considered in the Albert and Barabási framework: network growth. In our case, no new node can be added to the network over time, unless new districts are introduced.

The network analyses presented in this and the preceding section showed the distribution of the commuting data from two points of view: numerical and structure-wise. It is also interesting, at this point, to explore the network's characteristics (that is, homogeneity of the network) from the perspective of the deterrence function in the commuting flows. In other words, it is worth examining, by means of appropriate models, like SIMs, whether the network under analysis shows—in its deterrence function—an exponential function, reflecting a homogeneous network, or alternatively a power function—reflecting a hub structure. The next section investigates how the SIMs introduced earlier fit the data, and, most importantly, which specification (power-law or exponential) is more suitable to approximate the deterrence form of the commuting network.

4.4 Spatial interaction models: the results

The last steps in our analysis are: (a) the calibration of the SIMs for 2004; and (b) the test of the distribution properties of the commuting network in Germany, in the light

---

[5] The BfLR/BBR (BBR is the Bundesanstalt für Bauwesen und Raumordnung, Bonn, which former name was Bundesforschugsanstalt für Raumordnung und Landeskunde (BfLR) district typologies are (1) regions with urban agglomeration: (a) central cities, (b) highly urbanized districts, (c) Urbanized districts, (d) Rural districts; (2) regions with tendencies towards agglomeration: (a) central cities (b) highly urbanized districts, (c) rural districts; (3) regions with rural features: (a) urbanized districts (b) rural districts.

of its impedance function. Concerning the calibration phase, it should be noted that the SIMs were calibrated in the unconstrained specification. A value of −1.658 was estimated for the β deterrence factor of the power-law specification [Eq. (4)], while a value of −0.006 was computed for the exponential specification [Eq. (5)]. The results of the power-law SIM calibration are consistent with the findings of Olsson (1980) for Sweden. From this first analysis, the power-law coefficient seems more appropriate than the exponential coefficient, since the latter suggests a rather low propensity towards mobility in the German commuting network.

Concerning the test phase, we are then interested in comparing the connectivity structure emerging from the real data with the results of the SIMs from the O–D perspective. This is achieved by first ranking the observed and estimated data in decreasing order of commuters per O–D pair. We can now analyse the decay of the flows and fit a curve to the data. Table 3 shows the results obtained when fitting both a power-law and an exponential distribution to the data, while Fig. 3 provides a 'partial' visualization of the distribution of the two data sets and models, on a log–log scale.

According to the results shown in Table 3, the distribution of the observed flows fits a power-law distribution better than an exponential distribution, from both a statistical and spatial-economic viewpoint. However, from a purely statistical viewpoint, both functions could be suitable. In this context, refinements of these two functions might be adopted, for example, by means of a Box–Cox transformation (Box and Cox 1964). The two SIMs specified earlier [Eqs. (4) and (5)] seem to better fit the respective functions (power-law and exponential) at the basis of their computation, although with lower R^2 values. Two considerations may be made regarding the SIMs. On the one hand, the modelling results tend to be smoothed out in comparison with the observed data. The model data show a lower R^2, and a lower exponent for the power-law function, although its implications are not straightforward (for a discussion of power-law exponent values, see, for example, Albert and Barabási 2002). On the other hand, fitting a power-law function implies aiming at a more than proportional concentration of commuting flows over a few routes, with the number of commuters decreasing rapidly afterwards. This does not seem to happen with the real data.[6]

Next, by ranking the number of connections per district emerging from the SIMs (Fig. 2), it can be seen that the exponential-specified SIM better approximates the commuting network's connectivity structure, as it shows a cut-off, for the less-connected districts, that is more similar to the one of the observed data. This can be explained by the fact that the data themselves fit an exponential distribution better (see also Russo et al. 2007). Further, our finding of a slow decay of connections can also be considered to be a consequence of trends, more or less recent, due to the

---

[6] We can visualize the distribution of the flows estimated by the power-law SIM by plotting them against the data observed for the year 2004 (see Appendix B, Fig. 4, the scales are the same as Fig. 1). It is evident that, although the data employed refer to the same year, the clear correlation patterns found in Fig. 1 are not matched in these new plots. It can be noted, in the top-left graph, that three observations in particular are wrongly estimated by the SIM. The model underestimates the commuting flows between Hannover and its surrounding region, while it again overestimates the flows between the cities of Munich and Bamberg and their surrounding districts, respectively. Overall, the R^2 obtained by regressing the observed data on the SIM results is 0.415. Similarly to the discussion above of the years 1995 and 2004 (see Section 4.1), the R^2 decreases (to 0.405) when only observations with less than 1,000 commuters are considered. The bottom-right plot of Fig. 4 confirms the wide spread of the data. Generally, a more marked tendency to underestimation can be seen for mid-range flows (bottom-left graph).

Table 3 Fitting exponential and power-law distributions to commuting flows observed and estimated for 2004

| Distributions	R^2	F	Degrees of freedom	Parameter estimates	
				Constant	b_1
Observed flows					
Power	0.972	2917991.957	83895	97636479.740***	−1.589***
Exponential	0.839	438229.205	83895	92.631***	−0.00006***
Estimated flows (power law-specified SIM)					
Power	0.934	1377917.160	96840	28033260.494***	−1.405***
Exponential	0.908	956887.156	96840	123.979***	−0.00005***
Estimated flows (exponential-specified SIM)					
Power	0.901	847,235.717	93,295	1,290,073.530***	−1.176***
Exponential	0.941	1,489,485.743	93,295	47.716***	−0.00004***

All parameters are significant at the 99% level (***)

overcrowding of the main cities, such as the tendency to suburbanization, which causes an increase in commuting.

5 Conclusions

The present paper has provided an overview of the network properties found for home-to-work commuting patterns in Germany. First, an exploration of the commuting data was carried out, showing a significant increase of flows on the network, with a tendency to a more pronounced skewness and a power-law distribution. Second, a network analysis was considered in order to investigate the connectivity properties of the network. The analysis highlighted that the commuting network in Germany has become more concentrated and more clustered over the period considered (1995–2004). Also, the number of average connections per district increased, showing a denser net of reciprocal connections between cities. To some degree, the results also suggest a tendency toward a hub-based structure, although this trend is not clearly picked up by the analysis. This hub-process is somehow inhibited by the no-growth condition of the network (the number of nodes in the network is fixed), which hinders significant

Fig. 3 A partial visualization of the observed and the estimated commuting flows for the year 2004
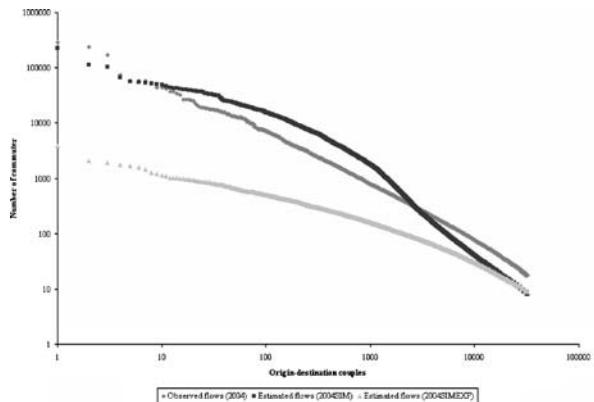
modifications of the network. This happens because the number of districts does not grow. In fact, it actually decreases, as a few districts were merged over time (such as in the case of Hannover and Berlin). At this stage, however, these mergers are not considered in our analysis. Generally, the German transportation network (commuting means are not specified) seems to make travel possible to approximately any point of the network. This is also evident from the distribution of the number of connections per district (Fig. 2), which shows a slow decay of the connectivity degree of the districts. In other words, the least-connected districts also still enjoy connections to the majority of the nodes in the network (see Appendix B, Table 5). Consequently, the increase in commuting over the years can be attributed to a better efficiency of the transportation network already in place.

In addition to the network analysis, two SIMs (alternatively using a power-law and an exponential deterrence function) have been utilized in order to detect the network structure underlying the flows. While the SIM modelling results for the flow variables concerned were quite ambiguous (see Footnote 3), most probably because of the simplicity of the models employed (unconstrained SIMs), the network connectivity structures generated by the two SIMs seem to favour the use of the exponential specification [see Eq. (5)], highlighting the homogeneity of the observed data.

Future research should further investigate the network properties of commuting, by employing better specified, doubly constrained SIMs that fully account for the total flows on the network, using a more suitable proxy of the travel opportunity cost than distance. In addition, it is desirable to go beyond the purely logical analysis of connections carried out here, by considering the real routes that, in the case of surface transportation, commuters have to follow. This could be done by means of a shortest-path analysis of the road network (or data on train routes), in which the commuting flows are utilized to weigh the network. A careful consideration of the level of spatial aggregation used should also be pursued. In fact, a recombination of smaller districts in 'macro-districts' might significantly influence the network structure found in this paper. A final interesting analysis might be the investigation of pre- and post-reunification mobility patterns, as well as the effects of the occasional merging of districts.

Appendix A

Table 4 Definition of the main network parameters

	Equation	Parameters
Clustering	$C(G) = \frac{1}{N}\sum_i C_i$	C_i=ratio of existing links and max. possible links G_I=subgraph of neighbourhood of vertex i N=nodes in the network
Diameter	$L(G) = \frac{1}{N(N-1)}\sum_{i \neq j} d_{ij}$	d_{ij}=shortest path links from i to j N=nodes in the network
Average degree	$k = 2L/N$	L=links in the network N=nodes in the network
Centralization	$C = \dfrac{\sum_{i=1}^{N}[C(p*)-C(p_i)]}{\max\sum_{i=1}^{N}[C(p*)-C(p_i)]}$	$C(p_i)$=centrality of node i $C(p*)$=max centrality of nodes in the network
Betweenness	$C_b(v) = \sum_{s \neq v \neq t} \sigma_{st}(v)/\sigma_{st}$	$\sigma_{st}(v)$=geodesic paths between nodes s and t that pass through node v σ_{st}=total geodesic paths between s and t

Appendix B

Table 5 Classification of most- and least-connected districts, years 1995 and 2004

| 1995			2004		
District	Type of district	# of connections	District	Type of district	# of connections
Most connected districts					
Hamburg, Freie und Hansestadt (W)	1	440	Hamburg, Freie und Hansestadt (W)	1	440
Hanover, Stadt (W)	1	440	Hanover, Stadt (W)	1	440
Cologne, Stadt (W)	1	440	Düsseldorf, Stadt (W)	1	440
Frankfurt am Main, Stadt (W)	1	440	Bonn, Stadt (W)	1	440
Stuttgart (W)	1	440	Cologne, Stadt (W)	1	440
Munich, Landeshauptstadt (W)	1	440	Frankfurt am Main, Stadt (W)	1	440
West Berlin, Stadt (E)	1	440	Offenbach (W)	2	440
East Berlin, Stadt (E)	1	440	Stuttgart (W)	1	440
Dresden, Stadt (E)	1	440	Esslingen (W)	2	440
Düsseldorf, Stadt (W)	1	439	Karlsruhe (W)	1	440
Offenbach (W)	2	439	Mannheim (W)	1	440
Esslingen (W)	2	439	Munich, Landeshauptstadt (W)	1	440
Bremen, Stadt (W)	1	438	Munich (W)	2	440
Munich (W)	2	438	Nuremberg, Stadt (W)	1	440
Nuremberg, Stadt (W)	1	438	West Berlin, Stadt (E)	1	440
Main-Kinzig-Kreis (W)	3	437	East Berlin, Stadt (E)	1	440
Leipzig, Stadt (E)	1	437	Dresden, Stadt (E)	1	440
Least connected districts					
Sonneberg (E)	8	259	Regen (W)	9	263
Straubing, Stadt (W)	9	257	Stralsund (E)	9	261
Kaufbeuren, Stadt (W)	9	252	Kaufbeuren, Stadt (W)	9	254
Regen (W)	9	251	Emden, Stadt (W)	8	249
Emden, Stadt (W)	8	243	Kusel (W)	7	249
Pirmasens, Stadt (W)	6	243	Pirmasens, Stadt (W)	6	248
Lüchow-Dannenberg (W)	9	228	Freyung-Grafenau (W)	9	240
Freyung-Grafenau (W)	9	227	Zweibrücken, Stadt (W)	6	235
Zweibrücken, Stadt (W)	6	220	Wismar (E)	8	231
Wismar (E)	8	214	Lüchow-Dannenberg (W)	9	225
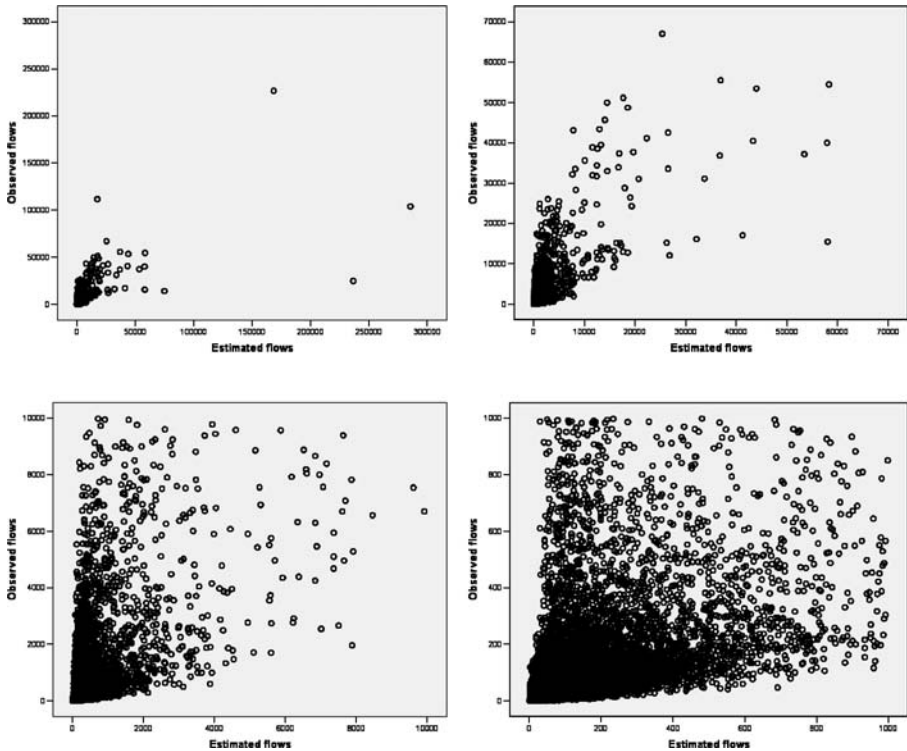
Fig. 4 Scatter plot of observed and estimated (power-law SIM) commuting flows, year 2004, at different scales

References

Albert R, Barabási A-L (2000) Topology of evolving networks. Phys Rev Lett 85:5234–5237

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47

Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance in complex networks. Nature 406:378–382

Amaral LAN, Scala A, Barthélémy M, Stanley HE (2000) Classes of small-world networks. Proc Nat Acad Sci USA 97:11149–11152

Barabási A-L (2001). The physics of the web. Physics World July 2001

Binder J, Haag G, Rabino G (2003) Analysis and modelling of commuter flows: application to the regions of Stuttgart and Turin. Jahrbuch der Regionalwissenschaften 23:117–139

Böltgen F, Irmen E (1997) Neue Siedlungsstrukturelle Regions- und Kreistypen. Mitt Inf BfLR H 1, S. 4–5

Bowen J (2002) Network change, deregulation, and access in the global airline industry. Econ Geogr 78:425–439

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B 26:211–246

Carey HC (1858) Principles of social science. Lippincott, Philadelphia

Cliff A, Haggett P, Ord H (1979). Graph theory and geography. In: Wilson R, Beineke L (eds) Applications of graph theory. Academic, London, pp 293–326

Cörvers F, Hensen M (2003) The regionalization of labour markets by modelling commuting behaviour. Paper presented at the 43rd ERSA, Jyvaeskylae

Erdös P, Renyi A (1960) On the evolution of random graphs, vol 5. Publication of the Mathematical Institute of the Hungarian Academy of Science

Fotheringham AS (1983) A new set of spatial-interaction models: the theory of competing destinations. Environ Plan A 15:15–36

Isard W (1960) Methods of regional analysis. MIT Press, Cambridge, MA

Johansson B, Klaesson J, Olsson M (2003). Commuters' non-linear response to time distances. J Geogr Syst 5:315–329

Jörnsten K, Thorsen I, Ubøe J (2004) Replication/prediction problems in the journey to work. Environ Plan A 36:347–364

Kemper F-J (1998). Restructuring of housing and ethnic segregation: recent developments in Berlin. Urban Stud 35:1765–1789

Kulldorf G (1955). Migration probabilities, Lund studies in geography, series B, No. 14. Lund, Sweden: Department of Geography, Lund University

Kwan M-P, Murray AT, O'Kelly ME, Tiefelsdorf M (2003) Recent advances in accessibility research: representation, methodology and applications. J Geogr Syst 5:129–138

Latora V, Marchiori M (2002) Is the Boston Subway a small-world network? Physica A 314:109–113

Ma K-R, Banister D (2007) Urban spatial change and excess commuting. Environ Plann A 39(3):630–646

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic, New York, pp 105–142

McFadden D (1979) Qualitative methods for analysing travel behaviour of individuals: some recent developments. In: Hensher DA, Stopher PR (eds) Behaviour travel modelling. Croom Helm, London, pp 279–318

Nijkamp P (1975) Reflections on gravity and entropy models. Reg Sci Urban Econ 5:203–225

Nijkamp P, Reggiani A (1992) Interaction, evolution and chaos in space. Springer, Berlin

Olsson G (1980) Birds in egg. Pion, London

Ravenstein EG (1885). The laws of migration. JR Stat Soc 48:167–235

Reggiani A (2004) Evolutionary approaches to transport and spatial systems. In: Stopher PR, Button K, Haynes K, Hensher DA (eds) Handbook of transport geography and spatial systems. Elsevier, Amsterdam, pp 237–253

Reggiani A, Schintler LA (eds) (2005) Methods and models in transport and communications. Cross Atlantic Perspectives. Springer, Berlin

Rouwendal J (2004) Search theory and commuting behavior. Growth Change 35:391–418

Rouwendal J, Nijkamp P (2004). Living in two worlds: a review of home-to-work decisions. Growth Change 35:287–303

Russo G, Reggiani A, Nijkamp P (2007). Spatial activity and labour market patterns: a connectivity analysis of the German space-economy. Annals of Regional Science (forthcoming)

Schintler LA, Kulkarni RG (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, Berlin, pp 419–434

Sen A, Smith TE (1995) Gravity models of spatial interaction behavior. Springer, Heidelberg

Sheffi Y (1985) Urban transportation networks. Prentice-Hall, Englewood Cliffs, NJ

Sohn J (2005) Are commuting patterns a good indicator of urban spatial pattern? J Transp Geogr 13:306–317

Stewart JQ (1941) In inverse distance variation for certain social influences. Science 93:89–90

Thorsen I, Gitlesen JP (1998) Empirical evaluation of alternative model specifications to predict commuting flows. J Reg Sci 38:273–292

Thorsen I, Ubøe J, Nævdal G (1999) A network approach to commuting. J Reg Sci 39(1):73–101

Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 363:202–204

White MJ (1977) A model of residential location choice and commuting by men and women workers. J Reg Sci 17(1):41–52

White MJ (1986) Sex differences in urban commuting patterns. Am Econ Rev 76:368–372

Wilson AG (1967). A statistical theory of spatial distribution models. Transp Res 1:253–269

Wilson AG (1970) Entropy in urban and regional modelling. Pion, London

Wojahn OW (2001) Airline network structure and the gravity model. Transp Res E 37(4):267–279