
The structure of interurban traffic: a weighted network analysis

Andrea De Montis

Dipartimento di Ingegneria del Territorio, Sezione Construzioni e Infrastrutture, Università degli Studi di Sassari, Via De Nicola, 07100 Sassari, Italy; e-mail: andreadm@uniss.it

Marc Barthélemy

School of Informatics, Center for Biocomplexity and Department of Physics, Indiana University, Bloomington, USA; e-mail: marc.barthelemy@cea.fr

CEA-DIF, Département de Physique Théorique et Appliquée, 91680 Bruyères-le-Châtel, France; e-mail: barthelemy@cea.fr

Alessandro Chessa

Dipartimento di Fisica, Università degli Studi di Cagliari, Cittadella Universitaria di Monserrato, 09042 Monserrato, Italy; e-mail: alessandro.chessa@dsf.unica.it

Alessandro Vespignani

School of Informatics, Center for Biocomplexity and Department of Physics, Indiana University, Bloomington, IN 47408, USA; e-mail: alexv@indiana.edu

Received 14 July 2005; in revised form 9 May 2006; published online 9 March 2007

Abstract. We study the structure of the network representing the interurban commuting traffic of the Sardinia region, Italy, which amounts to 375 municipalities and 1 600 000 inhabitants. We use a weighted network representation in which vertices correspond to towns and the edges correspond to the actual commuting flows among those towns. We characterize quantitatively both the topological and weighted properties of the resulting network. Interestingly, the statistical properties of the commuting traffic exhibit complex features and nontrivial relations with the underlying topology. We characterize quantitatively the traffic backbone among large cities and we give evidence for a very high heterogeneity of the commuter flows around large cities. We also discuss the interplay between the topological and dynamical properties of the network as well as their relation with sociodemographic variables such as population and monthly income. This analysis may be useful at various stages in environmental planning and provides analytical tools for a wide spectrum of applications ranging from impact evaluation to decision making and planning support.

1 Introduction

The informatics revolution has enabled the systematic gathering and handling of a large array of large-scale network datasets, and has allowed the analysis of their detailed structural features. In particular, mapping projects of the World Wide Web and the physical Internet offered the first chance to study the topology and traffic of large-scale networks. Gradually other studies followed describing networks of practical interest in social science, critical infrastructures, and epidemiology (Albert and Barabási, 2002; Dorogovtsev and Mendes, 2003; Pastor-Satorras and Vespignani, 2004). The possibility of accessing and mining large-scale datasets allows a more detailed statistical analysis and theoretical characterization of correlation patterns, hierarchies, and community structure of these networks, along with the setup of new modeling frameworks. In particular, the systematic study of these systems has shown the ubiquitous presence of *complex* features, mathematically encoded in statistical distributions with heavy tails, diverging fluctuations, self-organization, and emerging phenomena and patterns.

The statistical properties induced by these complex features indicate the presence of topological and structural properties that do not find an explanation in the paradigm

put forward by Erdős and Rényi (1960) with the random graph model. Indeed, even if the Erdős–Rényi model rationalizes the small-world property of networks (ie short distances compared with the number of links among nodes), it fails in reproducing the high level of local cohesiveness observed in many networks (Watts and Strogatz, 1998). Additionally, several of these networks are characterized by the statistical abundance of nodes with a very large degree k ; that is, the number of connections to other nodes. For these ‘scale-free’ networks, the degree probability distribution $P(k)$ spans over a wide range of k values, which signals the appreciable occurrence of large degree nodes, the ‘hubs’ of the system (O’Kelly, 1998; Taaffe et al, 1996). Finally, it is important to remark that the topological features of networks turn out to be extremely relevant as they have a strong impact in assessing the networks’ physical properties, such as robustness, vulnerability, or the ability to spread a disease (Barabási and Albert, 1999; Dorogovtsev and Mendes, 2003; Pastor-Satorras and Vespignani, 2004).

More recently, the activity on complex networks has been extended to the characterization of weighted networks. This representation allows the consideration of features pertaining to the dynamics and traffic flows occurring on networks, adding another dimension in the description of these systems. Also, in this case, the analysis and the characterization of weighted quantities have pointed out the presence of large-scale heterogeneity and nontrivial correlations (Barrat et al, 2004). The weighted graph representation provides valuable elements that open the path to a series of questions of fundamental importance to the understanding of networks. Among these are the issues of how dynamics and structure affect each other and the impact of traffic flows on the basic properties of spreading and congestion phenomena. Finally, this representation spurs the more theoretical questions on how all these properties may be considered in generative-network models.

In the realm of urban and environmental planning, spatial analysis, and regional science, the interest for complex networks has noticeably increased during recent years. In his investigation of the structure of the modern urban fabric, Salingaros (2001; 2005, page 143) invokes the concepts of a small-world network and scale-free properties, and suggests a parallelism between the city and an ecosystem. Starting from the principle that connective webs are the main source of urban life, Salingaros invites planners to apply complex-network analysis to understand properties such as resilience and self-organization. While these studies are mostly based on conceptual arguments, other authors (Batty, 2001; 2003; Shiode and Batty, 2000) promote the development of quantitative analyses with a twofold perspective: (i) explaining the behavior of complex networks such as the World Wide Web, in terms of geography, demography, and economics; and (ii) improving the efficiency of advanced spatial analysis GIS-based tools, by integrating the foundation of complex-network analysis into geographic information science. Batty (2001), along with Salingaros, acknowledges the existence of a link between efficient urban connectivity and self-preserving behaviors, and fosters the development of complex networks to study the functional mechanisms of cities. It is interesting to note that a similar approach has been extremely fruitful in the study of the Internet in relation to the geographical and social environment (Barthélemy et al, 2003; Gorman and Kulkarni, 2004; Pastor-Satorras and Vespignani, 2004; Schintler et al, 2005; Yook et al, 2002).

In the domain of geographic information science, Jiang and Claramunt (2004) proposed the use of complex-network tools to investigate quantitatively the urban space syntax. Aiming at the improvement of the effectiveness of network GIS advanced analysis, they used topological measures of connectivity, average path length, and generalized clustering in order to compare three large street networks of the cities of Gävle, Munich, and San Francisco. Further work on street network analysis can be

found in Crucitti et al (2006). Fertile research directions are also currently explored in the domain of transportation systems and human traffic analysis. For example, the small-world properties of the Boston subway system have been characterized by Latora and Marchiori (2000; 2001; 2002; 2003), and Sen et al (2003) studied the topology of the Indian railway network. At the urban level, a network study has been carried out by Chowell et al (2003) using large-scale simulations (with census and demographic data integration) in order to describe urban movements of individuals in Portland, Oregon. On a larger scale, recent studies have focused on the network properties of major transportation infrastructure, such as the US Interstate highway network and the airport network (Gastner and Newman, 2004; Guimera et al, 2003). Finally, Barrat et al (2004) have provided the first study of the worldwide airports network including the traffic flows and their correlation with the topological structure. In this research area, the network approach provides relevant information on issues such as traffic analysis and risk assessment in case of damages and attacks.

In the present work we use a network approach to study the Sardinian inter-municipal commuting network (SMCN), which describes the everyday work and study-led movements among 375 municipalities in the Italian region of Sardinia. We obtain a weighted network representation in which the vertices correspond to the Sardinian municipalities and the valued edges correspond to the amount of commuting traffic among them. We provide a detailed quantitative study of the resulting network, aiming at a first characterization of the structure of human traffic at the intercity level and its relation to the topological structure defined by the connectivity pattern among cities. Very interestingly, though the network topology appears to fit within the standard random graph paradigm, the weighted description offers a very different perspective with complex statistical properties of the commuting flows, and will enable us to display the highly nontrivial relations between weighted and topological properties. We discuss these properties in relation with the geographical, social, and demographical aspects of the system and show the potentiality of the network approach as a valuable tool at various stages of the policy making and environmental planning process.

The paper is organized as follows. In the next section we describe the dataset. In section 3 we carry out the analysis of the topology of the network. In section 4 we analyze the traffic and weight properties of this network, and in section 5 we discuss the implications of our network analysis results in terms of urban and environmental planning. In section 6 we compare topological and weighted network analysis results with socioeconomic characteristics of the Sardinian municipalities, and we discuss their relations. We finally summarize our results in section 7, and in section 8 we discuss the different research directions that this work suggests.

2 Setting the case study: data and geographical features

Sardinia is the second largest Mediterranean island with an area of approximately 24 000 km² and 1 600 000 inhabitants. Its geographical location and morphological characters have determined an important history of commercial and cultural relations with transborder external communities, and have also favored the development of a strong social identity and of important political movements toward self-reliance and autonomy. As of 1991 the island was partitioned into 375 municipalities—the municipality being the second simplest body in the Italian public administration—each one of those generally corresponding to a major urban centre [in figure 1(a) we report the geographical distribution of the municipalities]. For the whole set of municipalities the Italian National Institute of Statistics has issued the origin–destination table (ODT) corresponding to the commuting traffic at the intercity level (Istat, 1991a). The ODT was constructed from

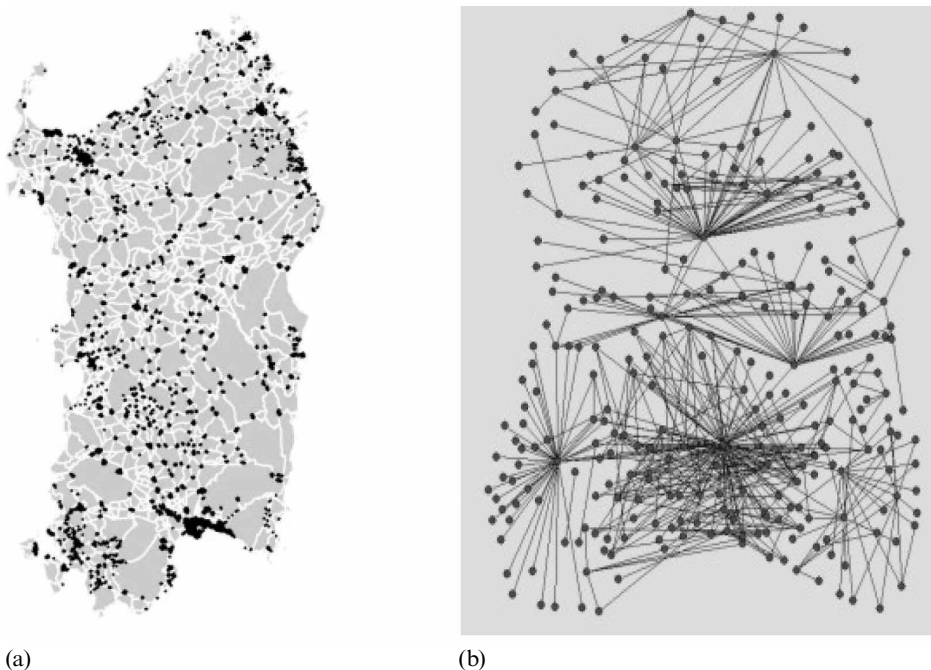


Figure 1. Geographical versus topological representation of the Sardinian intermunicipal community network: the nodes (points) correspond to the towns, while the links correspond to a flow value larger than fifty commuters between two towns.

the output of a survey about the commuting behaviors of Sardinian citizens. This survey refers to the daily movement from the habitual residence (the origin) to the most frequent place for work or study (the destination): the data comprise both the transportation means used and the time usually spent for displacement. Hence, ODT data give access to the flows of people regularly commuting among the Sardinian municipalities. In particular we have considered the *external flows* $i \rightarrow j$ which measures the movements from any municipality i to the municipality j and we will focus on the flows of individuals (workers and students) commuting throughout the set of Sardinian municipalities by all means of transportation.

This data source allows the construction of the SMCN, in which each node corresponds to a given municipality and the links represent the presence of a nonzero flow of commuters among the corresponding municipalities. The standard mathematical representation of the resulting network is provided by the adjacency matrix \mathbf{A} of elements (a_{ij}) . The elements on the principal diagonal (a_{ii}) are set equal to 0, since intramunicipal commuting movements are not considered here. Off-diagonal terms a_{ij} are equal to 1 in the presence of any nonzero flow between i and j ($i \rightarrow j$ or $j \rightarrow i$) and are equal to 0 otherwise. The adjacency matrix is then symmetric, $a_{ij} = a_{ji}$, and describes regular bidirectional displacements among the municipalities.

The adjacency matrix contains all the topological information about the network but the dataset also provides the number of commuters attached to each link. It is therefore possible to go beyond the mere topological representation and to construct a weighted graph, where the nodes still represent the municipal centers but where the links are valued according to the actual number of commuters. Analogously to the adjacency matrix \mathbf{A} , we thus construct the symmetric weighted adjacency matrix \mathbf{W} , in which the elements w_{ij} are computed as the sum of the $i \rightarrow j$ and $j \rightarrow i$ flows between

the corresponding municipalities (per day). The elements w_{ij} are null in the case of municipalities i and j which do not exchange commuting traffic and by definition the diagonal elements are set to zero $w_{ii} = 0$. According to the assumption of regular bidirectional movements along the links, the weight matrix is symmetric and the network is described as an undirected weighted graph. The weighted graph provides a richer description since it considers the topology along with the quantitative information on the dynamics occurring in the whole network. It is however important to stress that, while the nodes correspond to municipalities located in the physical space, the graph representation does not contain any information explicitly related to geographical distances and other spatial characteristics of the network. The definition of a network representation that correlates topological and traffic characteristics with the spatial properties of the SMCN requires a more refined investigation, which is in progress.

3 The analysis of the topological properties of the SMCN

In this section we analyze the topology of the SMCN and we propose possible territorial explanations for our findings. This network is relatively small, being characterized by $N = 375$ vertices and $E = 8124$ edges. The degree k of nodes, which measures the number of links of each node, ranges between 8 and 279 and exhibits a high average value $\langle k \rangle = 43.33$. The degree can be considered as a first indication of the topological centrality of a vertex: the importance of the corresponding municipal center as an attraction point for commuters in the network. A measure of the topological distance within the network is given by the shortest path length defining the distance between nodes in terms of links to be traversed by using the shortest possible path between the nodes. On average, the shortest path length is $l = 2.0$, and the maximum path length is 3.0. These values are small compared to the number of nodes, in agreement with a small-world behavior for which l typically scales as the logarithm of N .

Further information on the network's topology is provided by the degree distribution $P(k)$, defined as the probability that any given node has degree k . In this case it is essential to distinguish between an 'exponential' network similar to the usual random graph for which $P(k)$ decreases exponentially fast, and scale-free networks indicative of hub-like hierarchies for which $P(k)$ decreases typically as a power law (Albert and Barabási, 2002; Amaral et al, 2000). In the present case the degree probability distribution is skewed and is relatively peaked around a mean value of order 40 (figure 2).

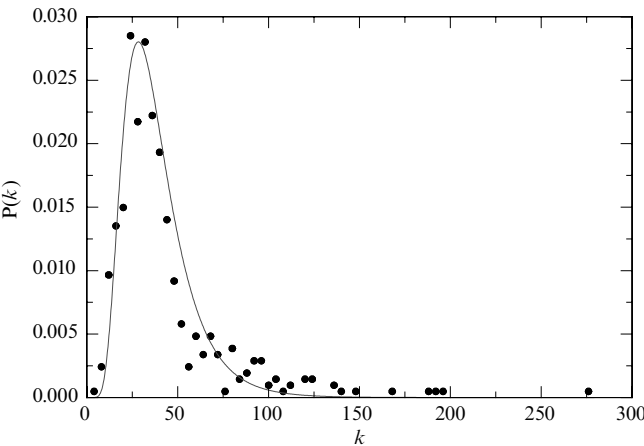


Figure 2. Plot of the probability distribution of the degree. The line is a lognormal fit.

The tail of the distribution contains outliers with respect to a regular exponential decay, signaling the presence of municipal hubs, which exchange commuters with many other municipalities. However, a closer inspection of the distribution tail shows that the decay is not compatible with a heavy-tailed distribution. The tail has indeed a power-law behavior, but with an exponent that ensures the convergence of the relevant moments of the distribution and which does not imply a scale-free heterogeneity. In particular, the variance converges to a finite limit for infinitely large systems in contrast with scale-free networks, which have an infinite variance for infinite size networks. This result implies that the topology of the Sardinian network is close to a random graph with a quasi-absence of dominant hubs.

A further relevant topological quantity is the clustering coefficient, defining a measure of the level of cohesiveness around any given node. It is expressed as the fraction of connected neighbors:

$$C(i) = \frac{2E(i)}{k_i(k_i - 1)}, \quad (1)$$

where $E(i)$ is the number of links between the k_i neighbors of the node i and $k_i(k_i - 1)/2$ is the maximum number of possible interconnections among the neighbors of the node. This quantity is defined in the interval $[0, 1]$ and measures the level of local interconnectedness of the network. If $C(i) = 0$, the neighbors of the node i are not interconnected at all, while $C(i) = 1$ corresponds to the case where all the neighbors are interconnected. A large clustering thus indicates the existence of a locally well-connected neighborhood of nodes. It is often convenient to average $C(i)$ over all nodes with a given degree k , giving the clustering spectrum:

$$C(k) = \frac{1}{NP(k)} \sum_{i|k_i=k} C(i), \quad (2)$$

where $NP(k)$ is the total number of nodes of degree k , and the summation is performed over the nodes i such that their degree k_i is equal to k . In figure 3(a) we plot the clustering spectrum of the SMCN: we observe a decaying behavior with $C(k)$ that varies from 0.8 to values lower than 0.2.

This indication of a decreasing clustering suggests that the SMCN behaves like other studied real networks, in that low degree municipalities belong to well interconnected working or studying communities, while the high degree vertices, the urban hubs, connect otherwise disconnected regions. It is worth noting that this result is similar to the one obtained by Chowell et al (2003) for urban movement.

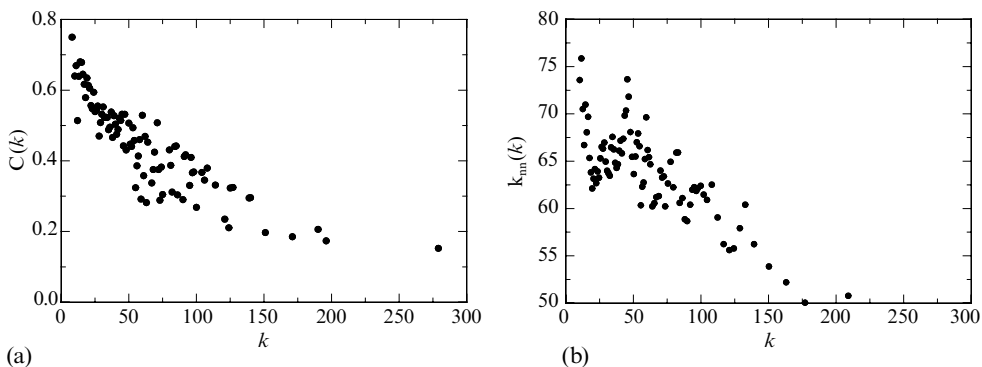


Figure 3. (a) Scatterplot of the clustering coefficient versus degree; (b) assortativity of the Sardinian intermunicipal commuting network, showing a slight disassortative behavior.

There is a striking difference however, as the SMCN is topologically very close to a random graph whereas the network studied in Chowell et al (2003) is a scale-free network. This similarity with a random graph is also confirmed by the average value of the clustering coefficient $\langle C(k) \rangle = 0.26$, which is of the same order as the clustering coefficient theoretically computed for the case of a generalized random graph, which is

$$\langle C(k) \rangle_{\text{RG}} = \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{N \langle k \rangle^3} = 0.24 . \quad (3)$$

Another important property is the degree similarity of the neighbors of a node. This is measured by the average degree of the nearest neighbors of a given node i :

$$k_{\text{nn}}(i) = \frac{1}{k_i} \sum_{j \in V(i)} k_j , \quad (4)$$

where $V(i)$ denotes the set of neighbors of i . Analogously to the clustering coefficient, we can average the assortativity over nodes with a given degree leading to

$$k_{\text{nn}}(k) = \frac{1}{NP(k)} \sum_{i/k_i=k} k_{\text{nn}}(i) . \quad (5)$$

This spectrum measures the tendency of vertices to be connected with vertices which have the same degree properties. If $k_{\text{nn}}(k)$ increases with k , nodes have a tendency to connect to nodes with a similar or larger degree and the network is said to be assortative. In contrast, if $k_{\text{nn}}(k)$ is decreasing with k , small degree nodes connect preferentially to hubs and the network is defined as disassortative. The result obtained for the SMCN is shown in figure 3(b), and displays a $k_{\text{nn}}(k)$ with a decreasing trend. This result seems to indicate that the SMCN belongs to the class of disassortative mixed networks. The disassortative behavior is typical of a technological and transportation network and signals the presence of a global hierarchy in which the hubs tend to provide the connectivity to the small degree nodes at the periphery of the network.

Finally, we measure an important topological quantity, the betweenness centrality (BC) (Freeman, 1977), which characterizes the centrality of nodes (or edges) in large networks. This quantity is defined as the total fraction of shortest path going through a given node. More precisely the BC is given by,

$$g(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} , \quad (6)$$

where $\sigma_{st}(i)$ is the number of shortest paths going from node s to node t and passing through node i , and σ_{st} is the number of shortest paths from s to t . Central nodes are therefore part of more shortest paths than less-important nodes, and act as ‘bridges’ between different parts of the network. Moreover, the BC is often used in transport networks to provide an estimate of the traffic handled by the vertices, assuming that the number of shortest paths is a first-order approximation to the frequency of use of a given node. Interestingly, it has been observed that for the airport network (Guimera et al, 2003), a large BC does not necessarily correspond to a large degree. In this case a characterization using the BC is then very important, since it conveys more information than a simple local quantity such as the degree.

In the case of the SMCN, we see in figure 4 that, in contrast with the airport network (Guimera et al, 2003), the fluctuations are small and the centrality of a node increases with its degree. This implies that the BC does not provide extra information with respect to centrality measures based on degree and other local quantities.

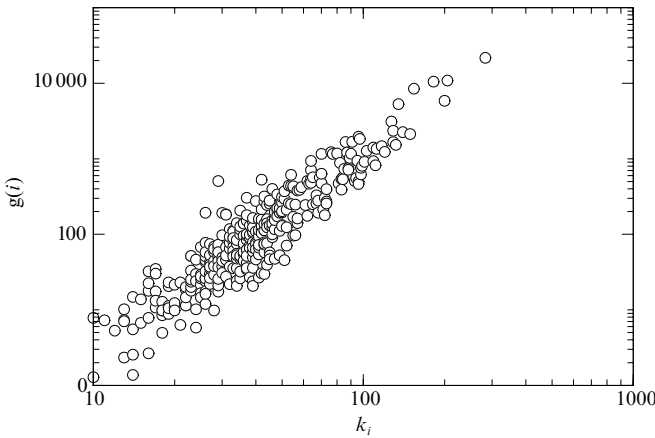


Figure 4. Scatterplot of the betweenness centrality (BC) versus the degree. Fluctuations are rather small and to a good approximation a large BC corresponds to a large k .

4 The SMCN weighted network analysis

The ODT dataset provides the number of commuters among municipalities, allowing the construction of the weighted graph representation of the SMCN. The analysis of these commuting flows informs us of the dynamical behavior of the system and of the needs of the commuters in terms of regional transportation services and infrastructure. This information might provide relevant indication to researchers and decision makers for addressing environmental policies and planning.

4.1 Weight analysis

The values of the weights between pairs of vertices range between 1 and $w_{\max} = 13\,953$ with an average value $w \cong 23$, much lower than w_{\max} . This fact is a signature of a high level of heterogeneity in the SMCN, since a peaked distribution of weights would be very unlikely to give a maximum value that is three orders of magnitude larger than the average. Indeed, the probability distribution $P(w)$ that any edge has a weight w displays a power-law decay— $P(w) \sim w^{-\gamma_w}$ with an exponent $\gamma_w \approx 1.8$ [figure 5(a)].

Another relevant quantity that characterizes the traffic is given by the strength of a node, defined as

$$s(i) = \sum_{j \in V(i)} w_{ij} \quad . \tag{7}$$

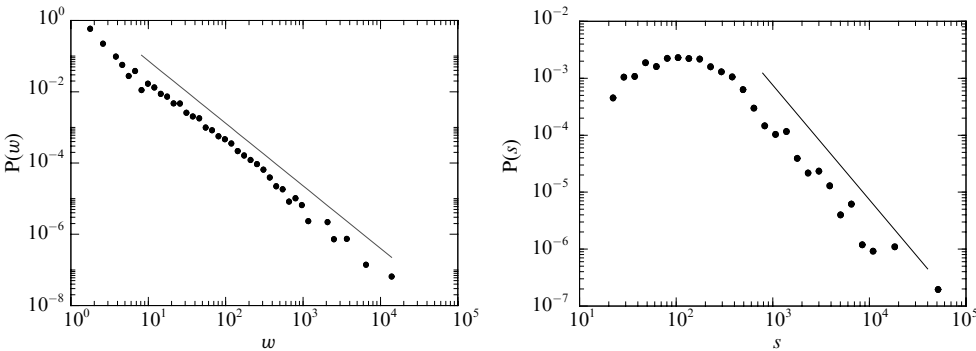


Figure 5. (a) Log–log plot of the probability distribution of the weights (the straight line is a power-law fit with exponent 1.8); (b) log–log plot of the probability distribution of the strength (as a guide to the eye we also plot a power-law fit with exponent 2).

This quantity corresponds to the total traffic commuting in the town i and can be considered as a measure of the traffic centrality of a municipality. For the SMCN the strength varies between 1 and 64 834, with an average value equal to 1010. The probability distribution of the strength $P(s)$, that any given node has a strength s , is fitted by a power-law regime $P(s) \sim s^{-\gamma_s}$ with exponent $\gamma_s \approx 2$ [figure 5(b)]. The strength provides an intuitive parameter for the description of the centrality of Sardinian municipalities according to the actual traffic handled, generated, and received by the municipality.

In the first instance the variation in the weight and strength values implies that commuters' flows are distributed in a broad spectrum with a large heterogeneity. It is worth commenting on the implications of a power-law behavior of the statistical distributions $P(s)$ and $P(w)$. This behavior defines the so-called heavy-tailed distributions which have a virtually unbounded variance. A remarkable fact about a distribution with a heavy tail is that there is a finite probability of finding vertices with weight or strength much larger than the average values. In other words, the consequence of heavy tails is that the average behavior of the system is not typical. The characteristic weight (or strength) is the one that, picking up a vertex at random, should be encountered most of the time. This is evident in the bell-shaped distribution observed for the degree of the SMCN, in which the average value is very close to the maximum of the distribution and represents the most probable value in the system. On the contrary, the distributions shown in figure 5 are highly skewed. Vertices have an appreciable probability of having large strength values, yet all intermediate values are probable and their mean does not represent any special value for the distribution. The power-law behavior and the relative exponent thus represent a quantitative measure of the level of heterogeneity of the network's traffic.

Another important question concerns the distribution of the traffic among the different roads. All the connections could carry a similar flow or on the contrary one connection could dominate. A convenient measure of this is given by the disparity (Barthélemy et al, 2005), defined as

$$Y_2(i) = \sum_j \left(\frac{w_{ij}}{s(i)} \right)^2. \quad (8)$$

For hubs ($k \gg 1$), this quantity enables us to distinguish situations [figure 6(b)] for which all weights are of the same order [$Y_2(i) \approx 1/k \ll 1$] from situations where only a small number of connections dominate [$Y_2(i)$ is of order $1/n$, where $n \ll k$].

We computed the value of $Y_2(i)$ for each node and we average these quantities for each degree in order to obtain $Y_2(k)$. The result is shown in figure 6(a). We obtain here an average behavior of the form $kY_2(k) = k^{1-\theta}$ with $\theta \approx 0.4$. This result shows that the weights on the links attached to a given node are very heterogeneous. In other words, there are only a few dominant connections, while the traffic on all the other roads is small.

4.2 The analysis of traffic – topology correlations

In order to inspect in more detail the relation between the degree and the traffic, we show in figure 7 the strength of nodes as a function of their degree. Despite the existence of inevitable fluctuations, we observe over a wide range of degrees a power-law behavior of the form $s(k) \sim k^\beta$ with an exponent $\beta \approx 1.90$. Independent variables would give a value $\beta = 1.0$ (Barrat et al, 2004) and the result obtained here reveals that there is a strong correlation between the traffic and the topology. The strength of the nodes grows faster than their degree: the more a municipality is connected with other

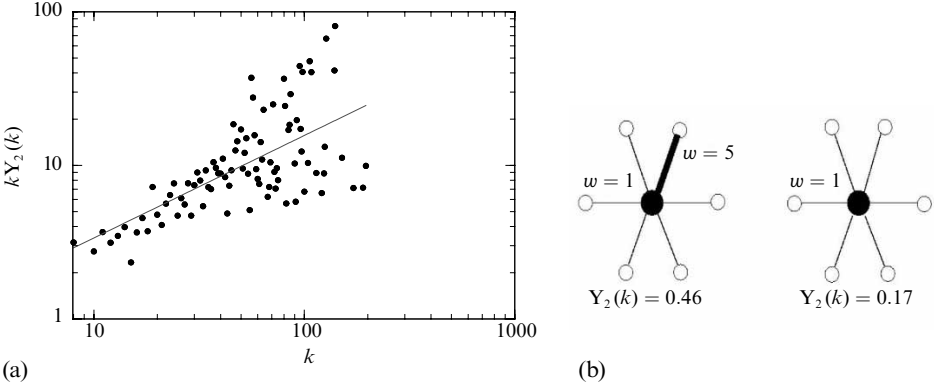


Figure 6. (a) Log–log plot of the disparity versus degree for the Sardinian intermunicipal community network; (b) illustration of the disparity for two very different cases. When a few connections dominate, $Y_2(k)$ is of order 1; in contrast, if all connections have the same weight, $Y_2(k)$ is of order $1/k$.

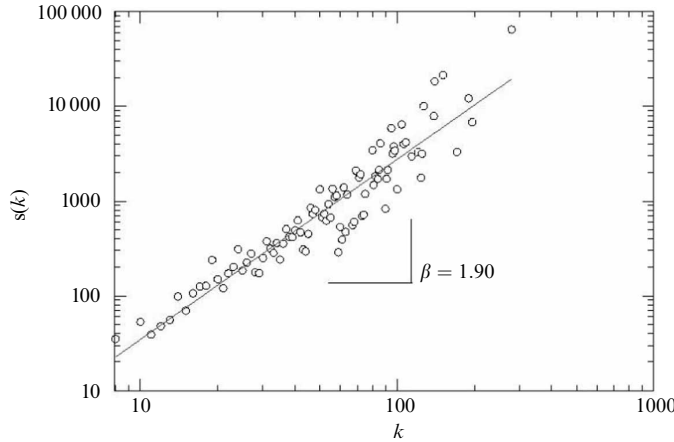


Figure 7. Average strength of the municipalities as a function of the degree.

centers the much more it is able to exchange commuters’ flows or, in other terms, the traffic per connection is not constant and increases with the number of connections.

The presence of weights also alters the definition of other structural measures. For instance, the standard topological clustering coefficient used in the previous section needs to be augmented by the inclusion of the actual traffic. In order to understand the weight structure and its relation with topology, the weighted clustering coefficient can be defined as follows (Barrat et al, 2004):

$$C^w(i) = \frac{1}{s(i)(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{ih} a_{jh} \quad , \quad (9)$$

where a_{lm} is an element of the adjacency matrix and where $s(i)(k_i - 1)$ is a normalization factor which ensures that $C^w(i)$ belongs to $[0, 1]$. As for the topological case, it is possible to average over all nodes of the same degree k obtaining the clustering spectrum $C^w(k)$. The weighted clustering coefficient counts the weight of the two participating edges starting from i , for each triple formed in the neighborhood of the vertex i . In this way, we are not just considering the number of closed triangles but also their total relative weight with respect to the vertex strength. In the case of random

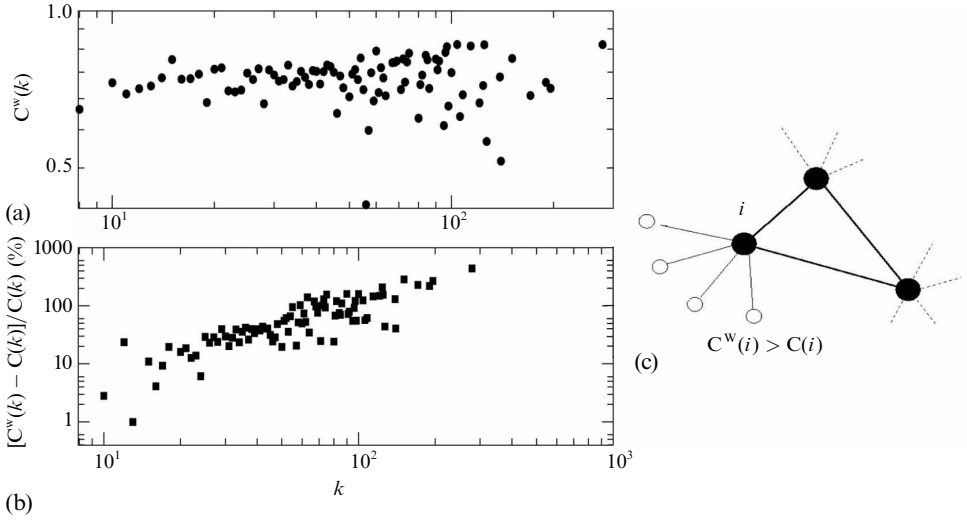


Figure 8. (a) Weighted clustering coefficient versus degree; (b) relative difference between the weighted and the topological clustering coefficients versus degree; (c) large cities are well interconnected by large flow links (bold edges must be considered with a larger weight).

networks $C^w(i) = C(i)$, but, in real weighted networks, we can face two different situations. If $C^w(i) < C(i)$, the topological clustering is generated by edges with low weight and therefore the cohesiveness is less important in terms of traffic properties. On the contrary, if $C^w(i) > C(i)$, we are in the presence of a network in which the interconnected triples are more likely formed by edges with larger weight. We show in figure 8(c) an example of such a situation. It is clear that the interconnected triple has a major role in the network dynamics and organization, and that the clustering properties are not completely exploited by a simple topological analysis. The weighted clustering coefficient takes into account the weighted elements by an appropriate mathematical form.

In the SMCN the values of the weighted clustering coefficient are larger than the corresponding topological values $C^w(k) > C(k)$ [figure 8(a) and 8(b)] over the entire range of degree values. In contrast with the topological clustering, the weighted clustering coefficient is approximately constant over the whole range of degree.

This result reveals the existence of a rich-club phenomenon in the sense that important cities form a group of well interconnected nodes and that weight heterogeneity is enough in order to balance the lack of topological clustering. The rich-club phenomenon implies that, in the SMCN, the interconnected triplets are more frequently built by edges with a higher weight [figure 8(c)]: in terms of the local cohesiveness of the workers' and students' commuting system, two different destinations available from a given city are more likely to be connected if the traffic flow leading to them is large. A typology of various arising substructures and hub-and-spokes networks was proposed by O'Kelly and Miller (1994) and the connected structure of hubs shown in figure 8(c) corresponds to a protocol-A network in this classification, which satisfies three conditions: (i) all hubs are interconnected, (ii) all nodes (nonhubs) are connected to one hub only, and (iii) there are no nonhub to nonhub direct connections. We note that from a purely topological point of view this structure is absent (the network is close to a random graph with a large clustering), but that, if we include the traffic in a weighted network analysis, we indeed observe such a substructure.

Similarly to the weighted clustering, we need to integrate the information on weights in a sensible definition of assortativity and disassortativity. This can be easily done by introducing the weighted average degree of the nearest neighbors of a given node i :

$$k_{nn}^w(i) = \frac{1}{s(i)} \sum_{j \in V(i)} w_{ij} k_j . \quad (10)$$

In this case we perform a local weighted average of the nearest-neighbor degree according to the normalized weight of the connecting edges. This definition implies that $k_{nn}^w(k) > k_{nn}(k)$ if the edges with the larger weights are pointing to the neighbors with larger degree and $k_{nn}^w(k) < k_{nn}(k)$ in the opposite case. The function $k_{nn}^w(k)$ thus measures the effective affinity to connect with high-degree or low-degree neighbors according to the magnitude of the actual interactions. In the SMCN the weighted average degree of the nearest neighbors displays a slightly increasing pattern, which is a signature of assortativity (figure 9). This result is in sharp contrast with the behavior of the topological disassortativity and indicates an affinity of high-degree municipalities with other large centers with which they exchange a high number of commuters: hubs have more frequently large traffic flows to and from municipalities with the same or higher-order degree. This phenomenon is illustrated in figure 9(c): large weights connect hubs, while the connection with nodes of smaller degree has smaller weights.

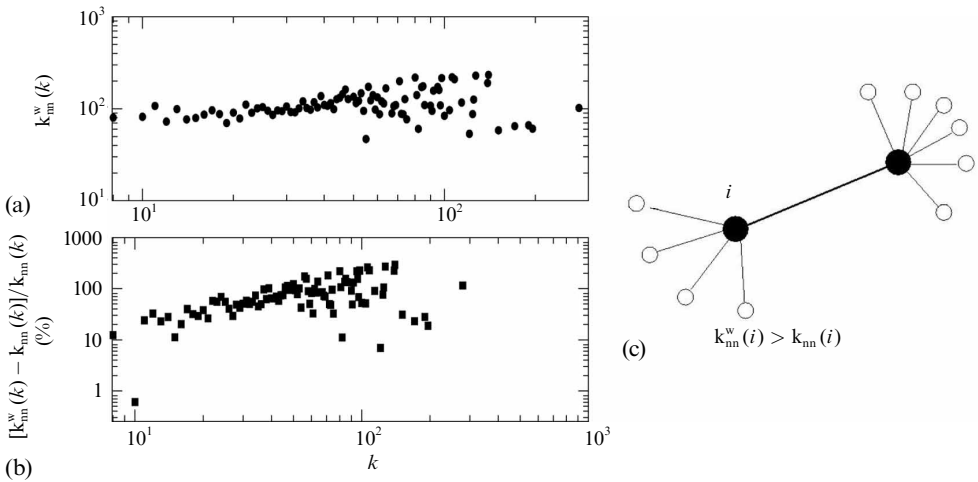


Figure 9. (a) Weighted assortativity; (b) relative difference between the weighted and the topological assortativity; (c) a difference between the weighted and the topological assortativity indicates that large cities are connected by large-flow links.

5 Discussion: the SMCN and its relation with real environmental systems

In this section we propose a discussion of the underlying environmental characteristics of Sardinia in the light of the results obtained about the topological and weighted properties of the SMCN.

5.1 Topology and environmental properties

The ranking obtained with the degree (table 1), an index of topological centrality, confirms the existence of administrative and socioeconomic hierarchies: among the hubs, Cagliari is the most populated and regional capital town, Nuoro, Oristano, and Sassari are province capital towns, Macomer and Quartu Sant'Elena are emerging productive and residential centers.

Table 1. Ranking of Sardinian municipalities by their degree k .

Rank	Sardinian municipal center	k
1	Cagliari	279
2	Nuoro	196
3	Oristano	190
4	Macomer	171
5	Sassari	151
6	Quartu Sant' Elena	140
7	Assemini	139
8	Selargius	127
9	Villacidro	125
10	Ottana	124

While the SMCN network is not scale free, the analysis of the degree probability distribution still reveals the presence of a relatively large number of well-connected towns that attract commuters from a fairly high number of satellite centers. The behavior of the clustering coefficient with respect to the degree indicates that, for $k < 50$, the clustering is relatively high, and is a signature of the fact that small towns (small k) form very well connected clusters of municipalities exchanging students and workers. On the contrary, hub towns ($k > 50$) have less cohesive neighborhoods since they have commuting flows with several different local clusters of municipalities not interconnected among them. This kind of star-like topology, with large municipalities acting as hubs for several small groups of municipalities, is confirmed by the analysis of the average nearest-neighbor degree spectrum, which suggests that the SMCN belongs to the class of disassortative mixed networks. In this case, high-degree hubs link preferentially with low-degree municipalities. This topological analysis would suggest the existence of intermunicipal commuting districts of small towns, which pivot around a few urban poles, and could be interpreted as the signature of a phenomenon often observed in urban and transportation systems when top functional rank towns attract commuters from small centers, behaving as urban poles. In such an environment, citizens of the satellite centers—being interested in regional-level services such as public administration, health, and finance—prefer commuting to the pole rather than to a peer-level town. The heterogeneity detected in the pattern of municipal centers could be seen as functional to an efficient behavior of the whole network, since each urban hub exchanges commuters, and thus goods, services, and revenues, with several small towns. Turning to the geographical displacement of the hub centers, we can note that, as expected, they are located far away from each other. The distance between Cagliari and Sassari is 210 km, and it is 160 km from Cagliari to Nuoro, with respect to the total length of the island, 230 km.

5.2 Traffic and environment in the weighted characterization of the SMCN

The heterogeneity of the weights, as we saw in the previous sections, is very high. This implies that the links are not all equivalent, and thus warns against any interpretation based on a sole topology. Indeed, the analysis of the weighted network introduces new insights into the interpretation, elaborated on the result of the topological properties under a variety of perspectives.

Since the weights are distributed in a very heterogeneous pattern, the corresponding commuters' flows display a wide range of values, which reflect the existence of a territorial hierarchy embedded and supported by the underlined transportation network. Precise indications emerge about dominant relations between some of the major poles of the island (table 2): Cagliari, Sassari, Olbia, Assemini, Porto Torres, and Capoterra.

Table 2. Ranking of pairs of Sardinian municipal centers by the weight of their connections.

Rank	Pairs of Sardinian connected municipal centers	w
1	Cagliari – Sassari	13 953
2	Sassari – Olbia	7 246
3	Cagliari – Assemini	4 226
4	Porto Torres – Sassari	3 993
5	Cagliari – Capoterra	3 731

Table 3. Ranking of Sardinian municipalities by their strength.

Rank	Sardinian municipality	s
1	Cagliari	64 834
2	Sassari	21 437
3	Quartu Sant’ Elena	18 431
4	Oristano	12 130
5	Selargius	10 084
6	Assemini	7 915
7	Porto Torres	6 886
8	Nuoro	6 834
9	Carbonia	6 616
10	Iglesias	6 479

In particular, the analysis of the strength probability distribution shows that municipalities identified as topological poles behave also as attractors of traffic flows. The strength is broadly distributed along the nodes, revealing a strong hierarchical dominance of the dynamic hubs over other smaller level centres.

The analysis of the relation between strength and degree shows that traffic grows superlinearly with the degree, thereby confirming the fact that, the more the topological hub towns are connected, the larger their traffic centralities (ie their capacity to attract Sardinian commuting students and workers). The ranking of Sardinian municipalities according to their strength is reported in table 3.

From the behavior obtained for the disparity (figure 5), we see that only a few connections emerging from the hub towns carry a large amount of traffic, indicating a strong hierarchy and heterogeneity also in the local structure of the flows connecting large municipalities of the SMCN. This mirrors the underlying road network structure, which comprises a very small set of main highways (state roads), a large set of medium-sized roads (provincial roads), and a very large set of local roads (municipal roads). In particular, the link Cagliari – Sassari corresponds to state road number 131, named ‘Carlo Felice’, which is the highway crossing the island longitudinally and can be considered as the backbone of the SMCN structure to which the lower hierarchical level transportation structures connect.

The use of the traffic information provides further insight into the structure and hierarchy of the network. In particular, the difference detected between weighted and topological clustering coefficients and weighted versus topological assortativity indicates the presence of a distribution of traffic flows that might be indicative of what is sometimes referred to as the rich-club phenomenon. Indeed, it is possible to observe that large traffic flows characterize the commuter traffic among municipalities with high centrality ranks as measured in terms of degree and strength. The accumulation of weight on these centers attenuates the decaying of the clustering and the average nearest-neighbor spectrum, which appears to invert their behavior by using the weighted definitions. This is the sign of the existence of a backbone of commuting

patterns among large municipalities which are all interconnected, while the flows toward peripheral municipalities are indeed of small size. According to this picture, small municipalities are mostly connected with each other through small traffic links, which correspond to the underlying second-order or third-order roads. This completes the topological information and offers a clear picture of the ordering principles of the SMCN, in which a large number of small cities tend to become satellites of higher rank municipalities, thereby shaping an overall network structure widely punctuated with star-like subsystems pivoting around important urban poles connected through dominant roads and the corresponding traffic flows. In this respect it appears that the network analysis presented here is able to provide an extremely detailed and quantitative characterization of the structure of the interurban traffic system.

6 Relating topology and traffic to socioeconomic phenomena

In this section we provide a discussion of the emergent topological and traffic properties in relation to the demographic and economic properties of the towns of the SMCN.

In the previous sections, we observed that the values of degree k and of the strength s , as indexes of topological and weighted centrality, quantitatively confirm historical structural hierarchies among Sardinian towns. The heterogeneity of the strength and the degree can be visualized as shown in the thematic map representation of figure 10.

We explore now in quantitative terms the relation between the network properties and the environmental and economic indicators. One goal is to inspect whether the commuting network hubs correspond to economically relevant towns. As measures of the social and economic centrality of each town, we consider two variables: the total resident population (p) in 1991 (Istat, 1991b) and the average monthly income (m). The average monthly income, an index of urban aggregated wealth, is defined as the product of the average monthly income per worker times the number of workers (Carcangiu et al, 1999). A map representation using these indicators is shown in figure 11.

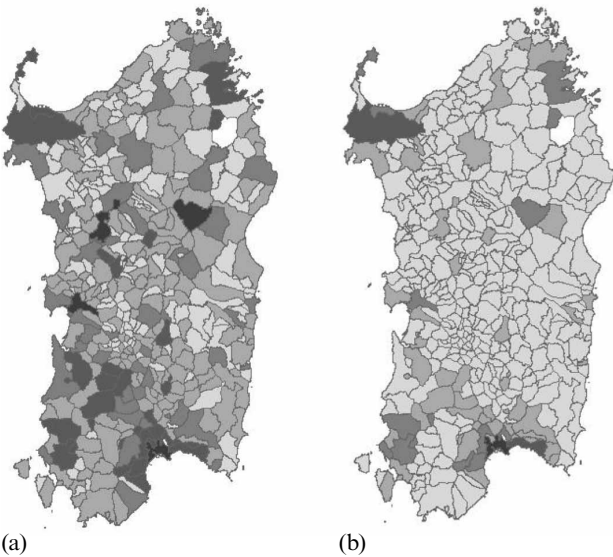


Figure 10. (a) Map of the degree k ; (b) map of the strength s for each Sardinian municipality (scale: the darker the shade, the higher the value).

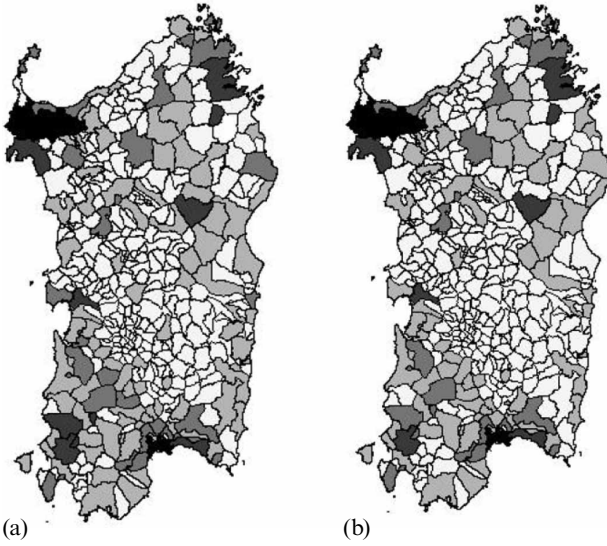


Figure 11. (a) Map of the resident population p ; (b) map of the municipal monthly income m for each Sardinian municipality in 1991 (scale: the darker the shade, the higher the value).

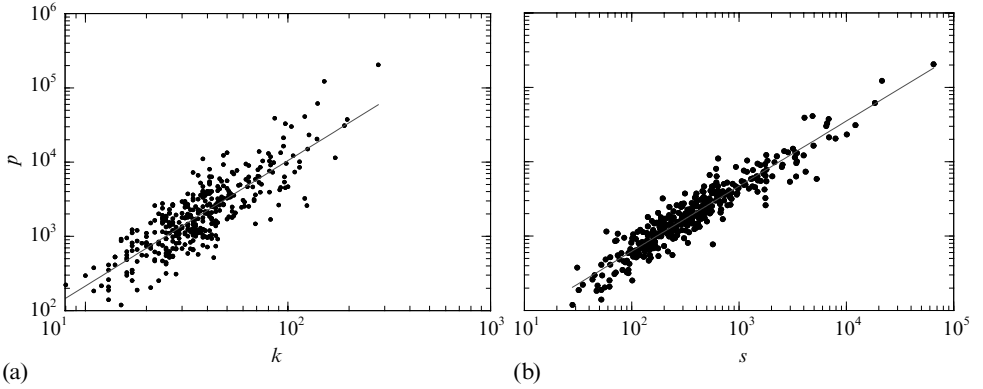


Figure 12. (a) Log–log plot of population p versus the degree k (slope coefficient equal to 1.70); (b) log–log plot of population p versus the strength s (slope coefficient equal to 0.90).

In figure 12, we plot the variable p as a function of the degree k and the strength s and we observe a clear positive correlation. The association between the population and these variables has a power-law behavior with two different exponents related by a factor 1.9. This is expected since we have already observed that traffic and degree are related by an empirical law $s \sim k^{1.9}$. This implies that, if the population has a scaling $p \sim s^a$, the simple substitution of variable yields $p \sim k^{1.9a}$. This is confirmed in the analysis, where the slopes of the two curves are related by a factor close to two. It is interesting to note that the population scales almost linearly with the traffic, thereby suggesting that a constant fraction of the population of every city commutes to another town, irrespectively of the size of the city.

In figure 13 the monthly municipal income (m) is plotted versus the degree k and the strength s . A positive correlation is signaled by a linear behavior on a log–log scale, which indicates the presence of a power-law association. Also in this case the exponents of the two relations confirm the general result of the (almost) quadratic relation between strength and degree.

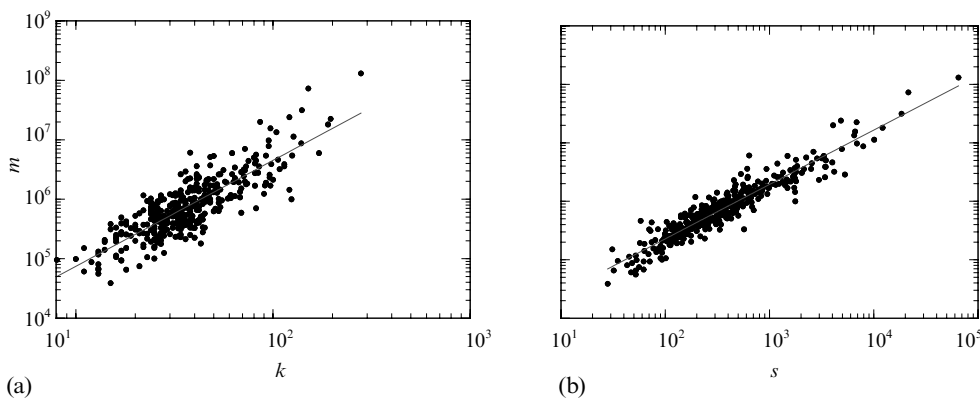


Figure 13. (a) Log–log plot of m versus the degree k (slope coefficient equal to 1.80); (b) log–log plot of m versus the strength s (slope coefficient equal to 0.90).

The results of this section confirm that the centrality measures (strength and degree) display trends consistent with the behavior of social and economic indicators such as population and income. In this sense, the higher the network centrality of a node, the higher is its demographic and economic size. The aforementioned territorial size measures, p and m grow much faster with the degree k than with the strength s . We can therefore identify a group of variables related to population (s , p , m) that are correlated by association laws with a behavior very close to the linear one. Their relation with the degree is, however, highly nonlinear, thereby indicating that these variables grow much faster than the connectivity of municipalities. Indeed, inverting the relation between degree and traffic (or population) we have $k \sim s^{1/2}$. The degree of a municipality is an indicator of the level of economic and social exchange both with other municipalities and with the transport infrastructure. The above relation implies that the economic and transportation infrastructures seem to grow at a pace slower than population indicators. It is not clear if this is a peculiar characteristic of the Sardinia region, due to political and urban planning decisions, or a general property of population flows, indicating a kind of adaptive equilibrium of a societal system. It is interesting to note that the worldwide airport network exhibits association trends among the network infrastructure and the population size very similar to those observed in this study (Barrat et al, 2004), thereby suggesting a possible level of universality. The detailed study of other commuting and transportation networks is however needed in order to gain further insight on this issue.

7 Conclusions

In this paper the interurban commuting system of the island of Sardinia has been represented as a weighted network. This representation allows the quantitative characterization of the capacity of attraction of each urban center on workers and students. By comparing the different levels of hierarchies emerging in the network with the pattern of demographic and economic poles, we shed light on several structural and organizational properties of the intermunicipal traffic system.

At first sight the SMCN is similar to a small-world random graph (Watts and Strogatz, 1998). The clustering organization of the SMCN, however, diverges from the usual random paradigm and, similarly to other technological networks, such as the airline network, it suggests a hierarchy of nodes in which small municipalities are densely interconnected locally, while large municipalities (regional hubs) provide the long-range connectivity for different areas of the islands which are otherwise disconnected.

This picture is reinforced by the evidence of a disassortative mixed network in which the hub municipalities are found to act as starlike vertices connecting, on average, municipalities with a low degree and a low centrality ranking.

The weighted representation of the network allowed by the inclusion of traffic data complements the resulting picture with important information relating the commuter traffic to the available topological connectivity. We find that the weights and the strengths are very broadly distributed, confirming the necessity of including those measures in a realistic description of the SMCN. In particular, the probability distributions of weights and strengths (commuter traffic handled by the municipality) display a power-law regime over a wide range of degree values: the SMCN behaves as a scale-free weighted network. In addition, the traffic is strongly correlated to the connectivity of the municipality and moreover, as shown by the disparity, it is largely concentrated on a very few links. This indicates that, the more a town is connected, the much more it attracts commuters along a small number of dominant roads. Our analysis of weighted clustering suggests that important towns form a group of well-interconnected centers. It also shows that in the SMCN the main fraction of traffic exchanges occurs among the hubs of the network: hub centers accumulate their commuter traffic to same-order or higher-order municipalities. In terms of environmental system analysis, a large number of small-size centers are commuting satellites of higher rank municipalities. The SMCN appears to be widely punctuated with starlike subsystems pivoting around important urban poles connected preferentially through high-traffic roads. In addition, the fact that the weighted clustering coefficient is large for the whole range of connectivity suggests that on the local scale there are microcommuting basins, rationalized as intermunicipal districts of highly clustered small towns, which pivot around a few urban poles belonging to macrocommuting basins.

Interestingly enough, the weight distribution and the clustering obtained in the present study show striking similarities with those obtained for the networks of population movements at the urban level (Chowell et al, 2003): this evidence suggests the existence of a set of basic mechanisms leading to common features in transportation system at various granularity levels. In table 4 the main characteristics of different transportations networks at different scales are shown. In particular, it seems that, while the topological properties can vary, the broadness of the traffic distribution is a common feature to all these networks. Another common feature is the existence of correlation between the traffic and the topology.

Finally, the existence of the municipality hierarchy that we have uncovered with the help of a weighted network analysis is confirmed by comparing demographic and economic urban polarization in Sardinia. Population and wealth, as proxies of the endowment of local resources, display a positive correlation with topological and dynamic centrality, strengthening the values of the network approach as a general tool for the analysis of interurban transportation systems.

Table 4. Comparison of properties of transportation networks. SNCM denotes the Sardinian intermunicipal commuting network.

Network	$P(k)$	$P(s)$	s	$Y_2(k)$
Global: WAN	heavy tail	broad	$\beta = 1.5$	$\theta = 1.0$
Intercities: SNCM	light tail	broad	$\beta = 1.9$	$\theta = 0.4$
Intraurban (Chowell et al, 2003)	heavy tail	broad	$\beta = 1.0$	–

8 Outlook

These different results show that a complex network approach could be relevant in uncovering nontrivial correlations between the flow of commuters and the territorial organization. The network approach allows one to uncover a rich structure in the flow pattern that could otherwise be very difficult to extract and to analyze. The results presented here provide a solid basis for further investigations using this approach. The indication of network centrality provides a powerful measure of the strength of municipalities in attracting workers and students daily. The present study is also the first step in a complex network description that takes into account simultaneously both the complexity of the topological feature and the dynamics occurring in the system. This approach might be useful in developing modeling approaches and in practical applications based on community detection and network optimization. In particular, the weighted network approach allows the consideration of clustering algorithms based on the strength and topology of the commuting patterns. These approaches could provide automated compartmentalization and territorial studies, and work is in progress in this direction. These studies may therefore provide insights for bottom-up processes of decision making and planning based on the definition of the existing emerging behaviors of the system. Finally, the SMCN is obviously not independent from geography and space. It refers to a physical network, mostly based on road systems, which display precise geographical characteristics. Further studies which include spatial aspects of the network are currently under progress.

Acknowledgements. A V is partially funded by the European commission under the project contract 001907 (DELIS). M B is on leave of absence from CEA-Centre d'Etudes de Bruyères-le-Châtel, Département de Physique Théorique et Appliquée, BP12, 91680 Bruyères-le-Châtel, France. A C gratefully acknowledges financial support by Fondazione Banco di Sardegna and by the project MIUR-PON CyberSar. The authors wish to thank Michele Campagna for his useful comments and discussions. We thank one of the referees for suggesting useful references.

References

- Albert R, Barabási A L, 2002, "Statistical mechanics of complex networks" *Review of Modern Physics* **74** 47–97
- Amaral L A N, Scala A, Barthélemy M, Stanley H E, 2000, "Classes of small-world networks" *Proceedings of the National Academy of Sciences* **87** 11149–11152
- Barabási A L, Albert R, 1999, "Emergence of scaling in random networks" *Science* **286** 509–512
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A, 2004, "The architecture of complex weighted networks" *Proceedings of the National Academy of Sciences* **11** 3747–3752
- Barthélemy M, Gondran B, Guichard E, 2003, "Spatial structure of the Internet traffic" *Physica A* **319** 633–642
- Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani V, 2005, "Characterization and modeling of weighted networks" *Physica A* **346** 34–43
- Batty M, 2001, "Cities as small worlds" *Environment and Planning B: Planning and Design* **28** 637–638
- Batty M, 2003, "Network geography", paper 63, Centre of Advanced Spatial Analysis (CASA), University College London
- Carcangiu R, Sistu G, Usai S, 1999, "Struttura socio-economica dei comuni della Sardegna. Suggestimenti da un'analisi cluster", WP3, CRENOS, Cagliari, <http://www.crenos.it/working/singwp.php?idwpp=68>
- Chowell G, Hyman J M, Eubank S, Castillo-Chavez C, 2003, "Scaling laws for the movement of people between locations in a large city" *Physical Review E* **68** 066102
- Crucitti P, Latora V, Porta S, 2006, "Centrality measures in spatial networks of urban streets" *Physical Review E* **73** 036125
- Dorogovtsev S N, Mendes J F F, 2003 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford)
- Erdős P, Rényi P, 1960, "On the evolution of random graphs" *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** 17–60
- Freeman L C, 1977, "A set of measures of centrality based upon betweenness" *Sociometry* **40** 35–41

- Gastner M T, Newman M E J, 2006, "The spatial structure of networks" *European Physical Journal B: Condensed Matter and Complex Systems* **49** 247–252
- Gorman S P, Kulkarni R, 2004, "Spatial small worlds: new geographic patterns for an information economy" *Environment and Planning B: Planning and Design* **31** 273–296
- Guimera R, Mossa S, Turtshi A, Amaral L A N, 2003, "Structure and efficiency of the world-wide airport network" *Proceedings of the National Academy of Science* **102** 7794–7799
- Istat, 1991a, "13° Censimento generale della popolazione e delle abitazioni, Matrice origine destinazione degli spostamenti pendolari della Sardegna", Italian National Institute of Statistics, 16 Via C Balbo, Rome 00184
- Istat, 1991b, "13° Censimento generale della popolazione e delle abitazioni", Italian National Institute of Statistics, 16 Via C Balbo, Rome 00184
- Jiang B, Claramount C, 2004, "Topological analysis of urban street networks" *Environment and Planning B: Planning and Design* **31** 151–162
- Latora V, Marchiori M, 2000, "Harmony in the small-world" *Physica A* **285** 539–546
- Latora V, Marchiori M, 2001, "Efficient behavior of small-world networks" *Physical Review Letters* **87** 198701
- Latora V, Marchiori M, 2002, "Is the Boston subway a small-world network?" *Physica A* **314** 109–113
- Latora V, Marchiori M, 2003, "Economic small-world behavior in weighted networks" *The European Physical Journal B* **32** 249–263
- O'Kelly M E, 1998, "A geographer's analysis of hubs-and-spoke networks" *Journal of Transport Geography* **6** 171–186
- O'Kelly M E, Miller H J, 1994, "The hub network design problem" *Journal of Transport Geography* **2** 31–40
- Pastor-Satorras R, Vespignani A, 2004 *Evolution and Structure of the Internet* (Cambridge University Press, Cambridge)
- Salingaros N A, 2001, "Remarks on a city's composition" *Resource for Urban Design Information* <http://www.cartage.org/16/en/themes/arts/civicsarts/AreaPlanning/urbanstructure/chapt8/remarks.htm>
- Salingaros N A, 2005 *Principles of Urban Structure* (Techné Press, Amsterdam)
- Schintler L A, Gorman S P, Reggiani A, Patuelli R, Gillespie A, Nijkamp P, Rutherford J, 2005, "Complex network phenomena in telecommunication systems" *Networks and Spatial Economics* **4** 351–370
- Sen P, Dasgupta S, Chatterjee A, Sreeram P A, Mukherjee G, Manna S S, 2003, "Small world properties of the Indian railway network" *Physical Review E* **67** 036106
- Shiode N, Batty M, 2000, "Power law distribution in real and virtual worlds", paper 19, Centre of Advanced Spatial Analysis (CASA), University College London
- Taaffe E J, Gauthier H L, O'Kelly M E, 1996 *Geography of Transportation* 2nd edition (Prentice-Hall, Englewood Cliffs, NJ)
- Watts D J, Strogatz S H, 1998, "Collective dynamics of 'small-world' networks" *Nature* **393** 440–442
- Yook S H, Jeong H, Barabási A L, 2002, "Modeling the Internet's large-scale topology" *Proceedings of the National Academy of Science* **99** 13382–13386

Conditions of use. This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.