Environment and Planning B: Planning and Design 2013, volume 40, pages 1071 – 1086

doi:10.1068/b38216

Urban street networks, a comparative analysis of ten European cities

Emanuele Strano
Laboratory of Geographic Information Systems, School of Architecture, Civil and Environmental Engineering, Ecole Polytechique Fédérale de Lausanne (EPFL), and Urban Design Studies Unit, Department of Architecture, University of Strathclyde, Glasgow, Scotland; e-mail: emanuele.strano@gmail.com

Matheus Viana, Luciano da Fontoura Costa
Institute of Physics of São Carlos, University of São Paulo, São Paulo, Brazil; e-mail: vianam@gmail.com, luciano@ifsc.usp.br

Alessio Cardillo
Departamento de Física de Materia Condensada, Universidad de Zaragoza, E-50009 Zaragoza, Spain, and Institute for Biocomputation and Physics of Complex Systems, Universidad de Zaragoza, E-50018 Zaragoza, Spain, and Dipartimento di Fisica e Astronomia, Università de Catania and INFN, Via S Sofia, 64, 95123 Catania, Italy; e-mail: alessio.cardillo@unizar.es

Sergio Porta
Urban Design Studies Unit, Department of Architecture, University of Strathclyde, Glasgow, Scotland; e-mail: sergio.porta@strath.ac.uk

Vito Latora
School of Mathematical Sciences, Queen Mary, University of London, London, England, and Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via S Sofia, 64, 95123 Catania, Italy, and Laboratorio sui Sistemi Complessi, Scuola Superiore de Catania, Via San Nullo 5/I, 95123 Catania, Italy; e-mail: latora@qmul.ac.uk

Received 1 December 2011; in revised form 27 July 2012

Abstract. We compare the structural properties of the street networks of ten different European cities using their primal representation. We investigate the properties of the geometry of the networks and a set of centrality measures highlighting differences and similarities between cases. In particular, we found that cities share structural similarities due to their quasiplanarity but that there are also several distinctive geometrical properties. A principal component analysis is performed on the distributions of centralities and their respective moments, which is used to find distinctive characteristics by which we can classify cities into families. We believe that, beyond the improvement of the empirical knowledge on streets' network properties, our findings can open new perspectives into the scientific relationship between city planning and complex networks, stimulating the debate on the effectiveness of the set of knowledge that statistical physics can contribute for city planning and urban-morphology studies.

Keywords: complex street networks, urban form, city classification, centrality

1 Introduction

Defining urban form is certainly an important and difficult issue, especially if one wants to supply useful knowledge to urban planners and urban designers or new knowledge for city scientists. In this paper we address this question, and we try to improve the empirical-based knowledge on the structure of a city by studying the urban street networks of ten

European cities. The form of cities is the subject of an area of urban studies named urban morphology. Urban morphology in its current form emerged between the 1940s and the 1960s from the work of two scholars as prominent as different: the German-born and then British urban geographer Conzen (1960), and the Italian architect and historian Muratori (1960). In this area the main subject of investigation is the urban fabric of the city at the scale of the neighborhood, street, plot, and building.

A different branch of urban morphology has stemmed from the sciences of complex systems building on a long-standing tradition in regional analysis, economic geography, and modeling (Anas et al, 1998). Complexity in the built environment is investigated here with the same instruments used for other classes of self-organized phenomena in nature, technology, and society (Batty, 2005). These works are now flanked by a growing interest in complex spatial networks within the community of physicists (Boccaletti et al, 2006).

Spatial networks, and in particular planar graphs, are suited to modeling a number of real phenomena (Barthélemy, 2011). Here we are interested in the study of a particular class of spatial networks that describe the street patterns of cities. The beginning of these studies can be traced back to the classical works on regional transportation networks based on graph theory (Garrison and Marble, 1962; Kansky and Danscoine, 1989). The advent of complex system science and its paradigm (Albert and Barabasi, 1999; 2002), jointly with the increasing availability of spatial and time georeferenced data, has given a new boost to these studies, and several important contributions have appeared recently (Barthélemy, 2011; Strano et al, 2012).

Masucci et al (2009) studied the structural property of the London street network in its dual and primal representation. In Jiang (2007), the authors, by using forty urban networks in a dual representation, found a small-world structure and a scale-free property for both street length and connectivity degree, and used various centrality indices as indicators of the importance of streets. Lämmer et al (2006) developed a comparative analysis of the betweenness distribution for twenty cities in Germany, suggesting a relationship with vehicular traffic. Others have focused on centralities in primal and dual representations of street networks (Crucitti et al, 2006; Hillier, 1999; Hillier and Hanson, 1984; Porta et al, 2006a; 2006b), and on other structural features, such as the number of cycles of a given length (Cardillo et al, 2006).

However, an important and still open question in urban morphology has to do with the characterization of classes of cities based on their form. This study is a preliminary step towards a comparative analysis aimed at the classification of cities. In this paper we propose a classification based on the distribution of street centrality by cross-comparing real cases and, therefore, with no use of null models. We limit our study to the characterization of city form without exploring its impact on collective behaviors, an area of research which, at the scale of entire cities, is now finding new opportunities through the exploitation of massive datasets from online geosocial networks and mobile georeferenced systems (Expert et al, 2011; Ratti et al, 2010).

In our study, we observe first the geometry of the networks following the approach recently suggested by Chan et al (2011); we consider the distribution of basic indices of the primal street networks such as street angles at intersections and street lengths. Secondly, we investigate four different centrality indices computed over the entire urban street network, highlighting their distribution and their mutual correlation.

We considered ten European cities: Edinburgh, Leicester, Sheffield, Oxford, Worcester, and Lancaster (in the UK), Catania and Bologna (in Italy), Barcelona (in Spain), and Geneva (in Switzerland). We show that these cities share some structural geometric properties, which are the same as those found in other planar spatial systems such as in leaves (Bohn et al, 2002; Couder et al, 2002; Perna et al, 2011), which suggests that planarity in itself is a driver across spatial systems in various domains.

However, we also show that cities are different. We witness that some cities, like Catania, stand alone in terms of basic geometric properties and, more importantly, that the various distributions of centrality are reconcilable to one common pattern (a power law) only if a largely minoritarian subset of streets is taken into consideration. These results have to do with the extreme heterogeneity of the cities' visible traits as resulting from the interplay of entirely different phenomenon in time, such as historical accidents (including planning), physical constraints, or just random events (Batty, 2005). In particular, we show how cases tend to cluster in groups after the whole set of centrality-measure distributions are considered in a way that suggests major planning interventions and physical geographic constraints are key.

The paper is organized as follows. In section 2 the case-study cities are presented and their geometric properties are introduced and analyzed. In section 3 the study of the four centrality indices is illustrated along with the clusterization of cases according to their combined behavior as a result of the application of a principal component analysis (PCA) of the distribution of those four centralities. Finally in section 4 we offer a discussion of the results and our conclusions.

2 Basic proprieties of urban networks

We address the analysis of the ten European cities (figure 1). These cities are variously located and present remarkably different economic, cultural, and climatic conditions along with a great variety of characteristics, such as population and area (table 1). We represent the street network of these cities such that intersections are nodes and streets are links between nodes, that is, a primal representation (Porta et al, 2006a). We analyse such networks in terms

Figure 1. Geographical location of the cities studied.

Table 1. Basic proprieties of the primal networks.

City	Population	Area (km^2)	N	E	$\langle k \rangle$	ρ (km^{-1})	L (km)	$\langle \ell \rangle$ (m)	σ_ℓ (m)	f (%)
Barcelona	1615908	82.0	6452	11071	3.43	15.15	1242	110.7	105.1	0.1
Bologna	380878	88.6	5200	7359	2.84	9.19	814	119.1	158.0	23.0
Catania	293811	34.0	11099	14039	2.52	23.91	813	55.9	57.3	17.0
Edinburgh	477660	195.6	5021	13063	2.43	8.96	1752	110.0	147.0	24.0
Geneva	191237	95.0	6183	8681	2.80	11.18	1062	122.4	119.7	0.9
Lancaster	45952	77.7	5913	15567	2.51	9.28	721	96.8	153.6	18.0
Leicester	288000	122.2	7186	8896	2.47	7.23	883	98.5	94.5	18.0
Oxford	149800	51.1	4372	11071	2.32	10.27	525	103.0	133.7	28.0
Sheffield	520700	187.5	14583	17674	2.42	10.58	1983	111.0	129.8	21.0
Worcester	94300	45.1	4685	5538	2.36	11.75	530	94.8	126.6	21.0

Notes. N and E are the number of nodes and number of edges, respectively, while $\langle k \rangle$ is the average degree. The density ρ is given by the ratio between the total length (L) and the area. $\langle \ell \rangle$ indicates the average length of the edges, while σ_ℓ corresponds to its standard deviation. f indicates the percentage of tree-like appendixes.

of their basic properties and several geometric indices such as street length and the angle streets form at intersections.

Before getting deeper into the analysis, we introduce some basic concepts of graph theory. A graph (or network) is a mathematical object which consists of two sets: $\mathcal{N}$ and $\mathcal{L}$. The N elements of the former are called *nodes*, while the E elements of the latter (unordered pair of nodes) are called *links*.

There are many ways to represent a graph, but the most common one is the *adjacency matrix* $\mathbf{A}$, an $N \times N$ square matrix whose entry a_{ij} $(i, j = 1, \ldots, N)$ is equal to 1 if a link between nodes i and j exists and 0 otherwise. The *degree* k_i of a node i is the number of links incident with it. The *average degree* $\langle k \rangle = 2E/N$, is the average of the degrees over all the nodes in the network.

Networks of street patterns belong to a particular class of graphs called *planar graphs*, that is, graphs whose links cross only at nodes. In our case the nodes represent street intersections, while the links are the centerlines of the streets and a network made using this convention is called a *primal network* (figure 2). Our networks are also weighted and each link (i, j) carries a numerical value w_{ij} expressing the intensity of the connection. The natural choice, in our case, for the functional form expressing the weight of a link connecting nodes i and j is to put w_{ij} equal to the length l_{ij} of the connecting street.

Our analysis starts by importing the street system into a geographical information system (GIS) environment. Data of the street systems have been retrieved from different sources: for example, in all UK cases we have used the Ordnance Survey maps, while in Italian cases we have used data from the planning offices of the city councils, and in Barcelona we worked with a dataset provided by the Agència d'Ecologia Urbana de Barcelona. Given that these geographical street networks had been built mainly for the sake of traffic navigation or planning, they presented characteristics that do not always fit the purposes of a centrality analysis; for example, multilane streets were usually represented with one link per lane rather than one link per street. As a result we have prepared our databases by first cleaning the networks accurately to remove link redundancies, fix short missing links, and, when needed, collapse unconnected links on the same node, while continuously confronting the networks with aerial images of the real cities drawn from remote sensing sources such as Google Earth. Such procedures were undertaken both manually and through ad hoc tools in a GIS

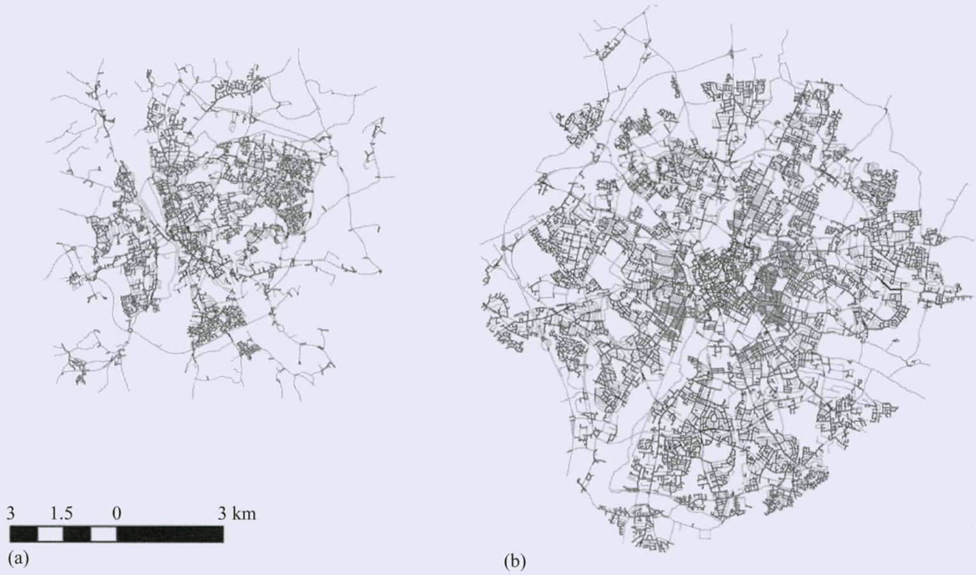

Figure 2. Example of street networks, (a) Leicester and (b) Worcester. The streets with lengths that are not following a power-law distribution are shown in black, while streets with lengths that follow a power-law distribution are shown in grey.

environment. For the definition of the boundary of the urban systems, we followed the border of the built-up area extending it outwards by roughly 1 km.

Considering the entries in table 1, we can see how the various selected cities compare, for example, in terms of size, from small cases like Lancaster to large ones like Barcelona, or in terms of street intersection density, from very dense cities like Catania to more sprawling ones like Edinburgh. We have selected cities with different levels of geographic constraints, from those such as Geneva and Oxford that are traversed or limited by large natural water features to those like Catania and Bologna that sit on uninterrupted plains, and cities with a different prominence of planning history, from those self-organized or only fragmentarily planned like Leicester or Bologna to one like Barcelona whose street layout had been heavily determined by one single planning vision (the 1859 Cerdá Plan). We see that the particular planning history of Barcelona is reflected in the low values of both the standard deviation of the street length and the percentage of dead-end streets to the total number of streets, both resulting from the extensive adoption of a rigid homogeneous grid layout. The extreme diversity of the selected cases has been pursued in order to make the comparative analysis of similarities and differences more profound.

The study of the geometric properties of the networks has focused on the distribution of three indices with the aim of finding common patterns: street length, angles formed between street intersections, and the relationship between dead-end link length and the area of the *cycle* they belong to. Following the definition given by Chan et al (2011), we consider cycles as polygons formed by closed loops of links. We build our approach on previous findings that have identified universal geometric patterns under seemingly diverse street networks in cities (Barthélemy and Flammini, 2008; Bohn et al, 2002; Couder et al, 2002; Låmmer et al, 2006; Masucci et al, 2009; Perna et al, 2011) and extend our exploration to focus on local patterns that actually make the uniqueness of each case or clusters of cases.

In a city long and short streets both play an important role. The former allow the connecting of distant locations, while the latter act as shortcuts between longer streets reducing the average path length in the navigation of the system. Short and long streets have a different historical meaning in the evolution of cities, as street length tends to diminish with increasing density of the urban pattern, following a rule in which the more 'urban' the area, the shorter the street.

The overall average street length $\langle \ell \rangle$ of selected networks represents a simple and good indicator of the diversity of cities, and looking at table 1 we find a considerable variation so that even putting aside the special case of Catania with $\langle \ell \rangle = 56$ m, which is due to the extreme density of the historical core, the average street length lies between 94.8 m (Worcester) and 122.4 m (Geneva).

We observed the distribution of street length in the selected cases. Since we are comparing cities of different sizes, we considered the normalized length ℓ, that is, the street length divided by the diameter of the network, which is defined as the maximum Euclidean distance between any couple of nodes belonging to the network.

In figure 3 we confirm the findings of Barthélemy and Flammini (2008), that the relationship between the total length L and the number of nodes N scale as $N^{1/2}$. However, we notice that our cases are not closely distributed along a straight line, indicating a significant variance that can be explained by the different natures of our datasets: we are, in fact, comparing a small number of cases; moreover, our cases are large networks representing entire cities, which means that we are here handling nonhomogeneous and invariant street networks made of parts derived from different historical formations and shapes. It is exactly this variability that we want to capture with a closer look at the differences emerging from the data.

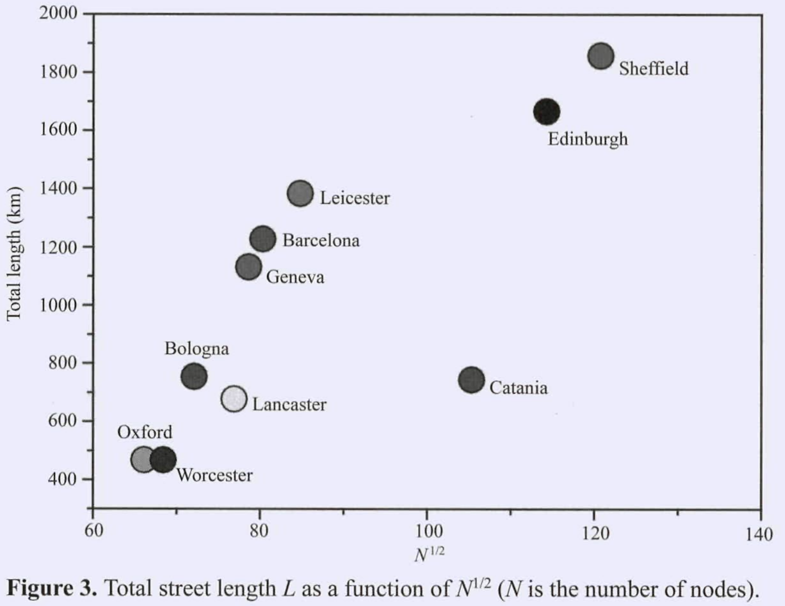

Figure 3. Total street length L as a function of $N^{1/2}$ (N is the number of nodes).

The distribution of street lengths is the simplest and first indicator that we use for examining such differences. Previous findings proposed a power-law distribution with a cut-off for the longest street segments (Masucci et al, 2009) and a bimodal shape distribution with a plato region above 30 m (Chan et al, 2011). We found slightly different results. If we look at figure 4(a), we see that a power law emerges in the distribution tail though, of course, the accordance between the distribution and the fitting becomes worse with decreasing of the street lengths. However, since we want to look at local patterns as well as global patterns, we must investigate what happens in the region that is not well fitted by a power law.

In figure 4(b), we plot also the same distribution in a semilog scale. Here it is possible to see that most of the distributions exhibit a peak in the region between 8×10^{-4} and 3×10^{-3}, which means that the majority of streets have a normalized length around these values. However Catania, Barcelona, and Worcester have different behaviors whose causes might be traced back to historical accidents. Catania falls out of these boundaries because of its abundance of very short streets in the historical urban center, which are possibly a consequence of the complete reconstruction of the city center after the disastrous volcanic eruption of Etna in 1669. Barcelona presents an anomalous peak clearly related to the Plan Cerdá mentioned above, with a massive grid-iron plan covering the central part of the current urban area. Worcester exhibits a double peak in the street length distribution, which is a consequence of the post XXII planning process that clearly shapes most of the periphery. From this we can appreciate the impact of specific historical occurrences that mostly impacted the urban form in terms of an interplay between planned and nonplanned urban forms.

This variety of street patterns gets entirely hidden in the conventional representation of data through log–log charts and if we look at which part of the street layout falls within the region that is correctly fitted by a power-law function (table 2), we see that it actually represents a minority of the entire network in all our cases.

Table 2. The total number of streets following a power-law distribution is very low and includes only the longer streets in the city. In terms of total street length the percentage of streets not following the power-law distribution are always the majority with the exception of Lancaster.

City	Threshold (m)	% of streets in a power-law distribution	% of street length not in the power-law distribution
Barcelona	347	4	86
Bologna	122	28	37
Catania	143	6	77
Edinburgh	194	13	59
Geneva	218	13	63
Lancaster	93	29	32
Leicester	211	9	74
Milano	233	8	71
Oxford	184	13	58
Sheffield	245	8	70
Worcester	166	14	55

In terms of sheer numbers of streets, using the method proposed by Clauset et al (2009) we observed that the percentage of streets falling inside the power-law region ranges from 4% in Barcelona to the 29% in Lancaster as shown in table 2. Of course, streets falling in the power-law region are the longest, so they cover a larger share of the system in terms of street length; however, the percentage of the total street length belonging to streets falling outside the power-law region in most of the cities is up to the 60%, as shown in the last column of table 2. We can appreciate visually the geographic consistency and character of the portion of the street network in Leicester and Worcester that is outside the power-law region: clearly, this portion represents not only the majority of the street network, but also the part that is historically more important, the denser and the more central, which is no surprise if we think that it is comprised of the shortest streets. However, generalizations and conclusions may be improved by further and deeper tests on larger areas. For the time being, on the basis of these results, we may argue that cities are composed of streets following two distributions that may reflect different dynamics of urban evolution.

Streets are not always straight lines. In order to study the distribution of angles formed by streets at intersections we must use an equivalent network in which all the streets are

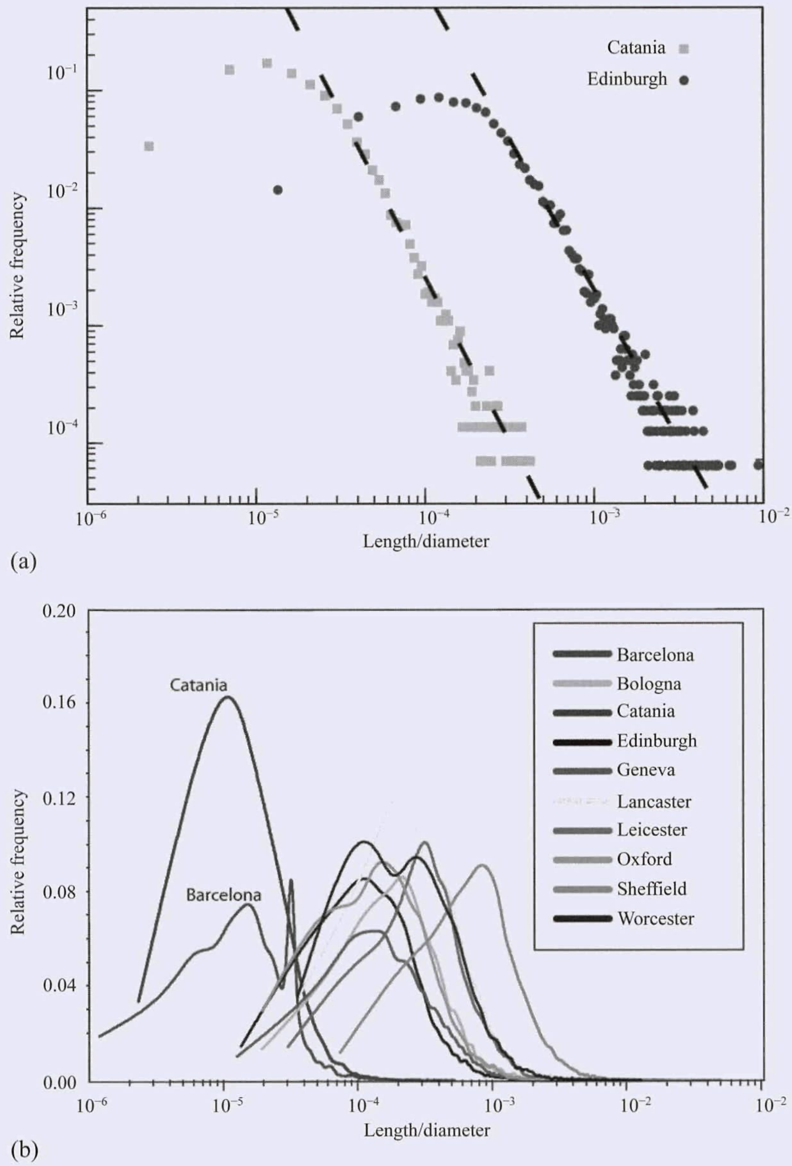

Figure 4. Distribution of normalized length ℓ for (a) two representative cities, Catania and Edinburgh, in the log–log scale and (b) the ten cities in semilog scale. It is possible to observe how different visualization give different results. In the semilog plots the distribution of the majority of streets seem different, while in the log–log plots they have the same trend.

represented by straight lines (ie, substituting straight streets for curved streets) and where the link weight is equal to the Euclidean distance between the end nodes. We name this a *Euclidean network*. As for all the generalization models, results should be interpreted with caution due to the effect that the approximation may produce on the real structure of the network. Such caution suggests a preliminary test. The inset in figure 5 shows small divergences between the distribution of street length in the original network and in the Euclidean network for the city of Leicester. Such a simple test confirms that the Euclidean generalization does not lose relevant information, that is, that streets in cities are not always straight but are predominantly straight, confirming the finding of Chan et al (2011). Therefore

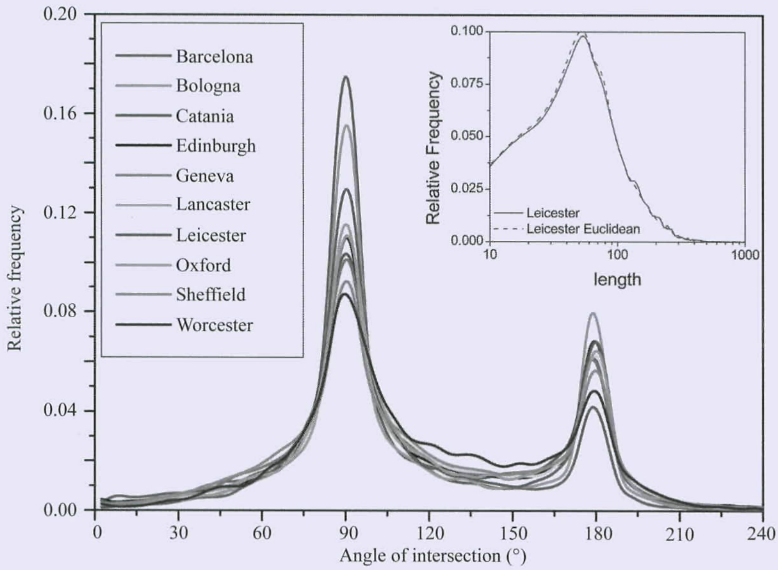

Figure 5. Distributions of the angles formed by street intersections of the cities' Euclidean networks. All the distributions show the same shape. The inset shows the street length distribution of the original network and the Euclidean network for the city of Leicester.

we are confident that using the distribution of the angles formed by street intersections in the Euclidean network is a reliable approximation of the original. The distributions of the angles formed by streets at intersections are shown in figure 5. At a glance, we note that all the cities share the same behavior, exhibiting a double peak-shaped distribution around the characteristic values of 90° and 180°. This finding confirms the analogous results that have been found in other forms of spatial planar networks, such as those of leaf venations (Bohn et al, 2002; Couder et al, 2002), as a result of tensorial stress fields or simple force models.

The incidence of certain motifs in complex networks is a long-standing object of investigation (Milo et al, 2002). Here we want to focus not on the cycle's shape and quantity, but on the relationship between the total length of dead links within a given cycle and its area. Even if we are measuring static systems, that is, systems that do not change over time, we should remember that a cycle is the result of an evolutionary process that starts with short dead-end streets sprouting from the longest edges of the cycles and then extending towards the opposite edge until splitting the original cycle in two smaller subcycles. Dead-end streets can be interpreted as sprouts of new cycles in parcels still subjected to evolution or as crystallized fractures that do not undergo further development. Their quantity is given by the index f shown in table 1 and it can be thought of as an estimator of the abundance of cycles in the intermediate evolutionary stage of their lifetime as suggested by Barthélemy and Flammini (2008). Such assumptions are supported by the result shown in figure 6, where we report the sum of the length of dead links inside each cycle versus the area of the cycle itself for each city. The distribution shows a clear power-law behavior with a common exponent close to 0.8. It is worth noting that the power-law behavior is not affected by any factor such as the fraction of dead ends f, or the average degree $\langle k \rangle$. Similar results have appeared in Lämmer et al (2006) and Perna et al (2011). Of course, our findings can be truly confirmed only by investigating the evolution of urban streets over time with the support of empirical data.

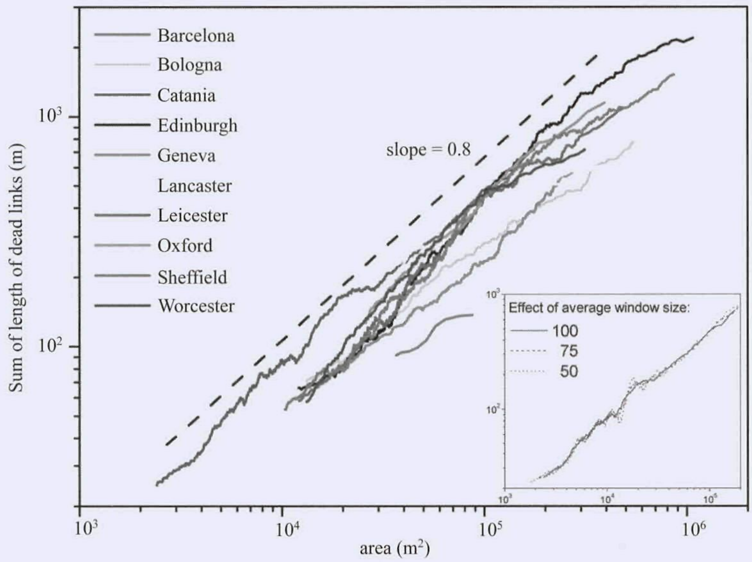

Figure 6. Total length of tree-like (dead ends) links as a function of the cycle area for the considered cities. The distributions show a power-law behavior with an exponent around 0.8.

3 Centralities and city classification

The concept of centrality has been used for many years in network and social science and, starting from the seminal work by Wasserman and Faust (1994), there has been a growth of the literature regarding centrality in social networks as well as other kinds of networks. Depending on the definition, centrality can be understood as meaning proximity between nodes, accessibility from other nodes, or being in a strategic position for connecting a couple of nodes. It is clear that from different definitions of centrality, a node actor can be placed at different centrality ranks and that the same node can have a high value for one form of centrality while yielding weak values for others. Therefore, for different cities, we can reasonably expect slightly different distributions of centralities. Moreover, we can identify how centralities are mutually dependent and correlated. What makes centralities particularly suited for geographical studies is that they can be visualized and mapped. We are interested in understanding if and how these ranges of correlations and fluctuations can help to classify cities that share the main network morphology. For example, a grid-like network can be different from a radial one, but only looking at the statistical distribution of centralities might not be enough for a proper classification. This problem has been already analysed by Porta et al (2006a; 2006b) where the statistical distributions of closeness, betweenness, straightness, and information centralities have been analysed for a sample of twenty one square mile city parcels. Crucitti et al (2006) found significant differences between cities and through a cluster analysis they proposed a classification of different urban patterns. The approach proved to be effective in capturing essential features of urban form as emerging in limited samples selected for their inner morphological consistency. However, dealing with entire cities poses the problem of the classification of internally complex objects predominantly composed of different parts, each of which may exhibit different properties. So, the question about the validity of such a procedure for a whole city still needs a response. In order to validate if centrality indices can be used in the classification of entire cities, we propose a clustering method based on a PCA made on the distribution of centralities and on their moments. Before discussing this part of the research, we must introduce briefly the adopted centrality indices.

Betweenness centrality, C^{B}, is based on the idea that a node is more central when it is traversed by a large number of the shortest paths connecting all pairs of nodes in the network. More precisely, the betweenness of a node i is defined in Wasserman and Faust (1994) and Freeman (1977; 1979) as:

$$C_i^{\mathrm{B}} = \frac{1}{(N-1)(N-2)} \sum_{\substack{j,k \in N \\ i \neq k,\, j \neq k}} \frac{n_{jk}(i)}{n_{jk}}\,, \tag{1}$$

where n_{jk} is the number of shortest paths connecting j and k, while $n_{jk}(i)$ is the number of shortest paths connecting j and k and passing through i.

Straightness centrality, C^{S}, originates from the idea that the efficiency in the communication between two nodes i and j is equal to the inverse of the shortest path length, or geodesic, d_{ij} (Latora and Marchioni, 2001). In the case of a spatial network embedded into a Euclidean space, the straightness centrality of node i is defined as:

$$C_i^{\mathrm{S}} = \frac{1}{N-1} \sum_{j \in N,\, i \neq i} \frac{d_{ij}^{\mathrm{Eucl}}}{d_{ij}}\,, \tag{2}$$

where d^{Eucl} is the Euclidean distance between nodes i and j along a straight line. This measurement captures the extent to which the connecting route between nodes i and j, such as between each street junction, deviates from a virtual straight route.

The *closeness centrality*, C^{C}, of a node i is based on the concept of minimum distance, in the topological sense, that is, the minimum number of edges traversed to get from i to j (Boccaletti et al, 2006) and is defined in Scott (2000) and Sabidussi (1966) as:

$$C_i^{\mathrm{C}} = \frac{1}{L_i} = \frac{N-1}{\sum_{j \in G} d_{ij}}\,, \tag{3}$$

where L_i is the average distance from i to all the other nodes. Closeness centrality is a classical centrality index that has been widely used in urban geography and econometrics as well as in regional planning, where it gives an idea of the cost that spatial distance loads on many different kinds of relationships that take place between places, people, activities, and markets.

The *accessibility*, C^{A}, is a measure recently introduced by Travençolo and da F Costa (2008a; 2008b). It has been used for studying the property of very different spatial networks. In the case of urban networks, it has been used to investigate the relationship between subway systems and road systems (da F Costa et al, 2011). In addition, the accessibility has been found to be closely related to the borders of networks (Travençolo et al, 2009), in the sense that nodes with low accessibility tend to belong to these borders. The C^{A} of node i measures the ratio of neighboring nodes that are effectively reached by an agent randomly navigating the network against the actual number of nodes that belong to the neighborhood. More precisely, C^{A} takes into account the number of nodes effectively accessed by each node of the network, as well as the probabilities of such accesses. First, we evaluate the transition probability $P_{i,j}(h)$ which describes the probability of an agent leaving from node i to reach node j after h steps along a given type of walk. At each step, the agent located at node q, chooses a random neighbor of q and jumps to it. These rules define a random walk over the network. When the transition probabilities are very heterogeneous, we have low values of accessibility, meaning that the random walks are biased towards a certain number of nodes, which is less than the number of nodes that can be reached after h steps. On the other hand, when the transition probabilities are homogeneous, all nodes that can be reached after h steps are accessed, on average, the same number of times. This case corresponds to the highest values of accessibility. The heterogeneity of the transition probabilities is quantified in terms of the classical concept of entropy, so that the mathematical definition of accessibility of a node i with respect to the number of steps h is given as

$$C_i^{\mathrm{A}}(h) = \exp\left[-\sum_{j=1}^{N} P_{i,j}(h) \ln P_{i,j}(h)\right]. \qquad (4)$$

Also, we have considered the transition probability for unitary step ($h = 1$), as:

$$P_{i,j}(1) = P_{i,j} = \frac{w_{i,j}}{\sum_{j=1}^{N} w_{i,j}}, \qquad (5)$$

where $w_{i,j}$ is the weight of the edge (i, j). In order to take the geography into account, we considered $w_{i,j} = 1/d_{ij}^{\mathrm{Eucl}}$. For disconnected nodes, we assume $d_{ij}^{\mathrm{Eucl}} = \infty$ such that $w_{i,j} = 0$.

In order to investigate if the distribution of centralities can describe the main geographical differences within cities, we use the PCA approach. This well-established method of multivariate statistics implements a projection of the distribution of objects (in our case, cities) from a higher into a lower dimensional space such that the maximum dispersion of the data is observed at the first new axis (or principal variable), and so forth. This projection is optimal in the sense of fully decorrelating the original data, therefore removing all correlations between the original measurements describing the objects. So, since the data dispersion is better described by the first principal axes, the remaining axes can be discarded. PCA is therefore particularly relevant to the present studies because: (i) it decorrelates the original measurements; (ii) it provides a projection of the data that maximizes their dispersion (ie, the differences between the cities); and (iii) it allows the visualization of the distribution of objects (when projected onto 2D or 3D spaces).

The PCA consists of obtaining the covariance matrix of the original data and then extracting its eigenvalues and respective eigenvectors. The eigenvalues can be shown to correspond to the variances along each new axis, and then each respective eigenvector component provides the coefficient of the linear combination of the original measurements used to project the original data onto the respective axis. Therefore, the effectiveness of PCA in projecting the data can be inferred by inspecting the eigenvalues in descending order. For instance, if the two largest eigenvalues account for 75% of the overall variance, it can be understood that these two axes are describing the original distribution of points in an effective way, and that the other axes can be overlooked without losing much information.

First, we evaluate the histogram of each centrality measurement considering 20 bins. In this way, each dimension corresponds to the relative frequency of centrality values in a small range. The histogram of each one of the four centralities were merged in order to create an eighty-dimension feature vector for each city. For instance, figure 7(a) shows the feature vector obtained for the city of Leicester. In the second approach, the feature vectors were created by considering the 20th first moments of the centrality distributions merged together, also generating an eighty-dimension vector. It is important to note that, both first and second approaches provided similar results. Finally, the dimension of the feature vectors were reduced from eighty to two, in which each original dimension has a contribution according the values shown in figure 7(b), which shows the explanation of each dimension. By plotting the first two dimensions, it is possible to account for almost 50% of the dispersion, while a clear differentiation between cities can be seen in figure 8. Since it is impossible to visualize the high-dimensional data, in order to try to recognize the clusters we used an agglomerative clustering method, the so-called complete linkage method, to perform this task. In this method, the Euclidean distance between two clusters is given by the value of the shortest distance between any object belonging to these clusters. The final result is shown

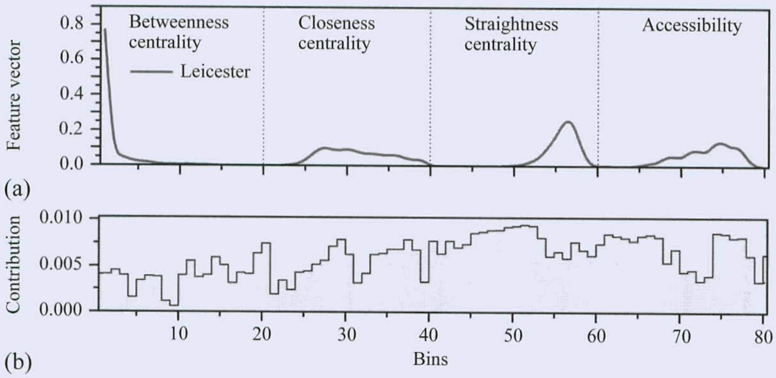

Figure 7. The complete vector centralities and the contribution of each bin to the explanation illustrated in figure 8.

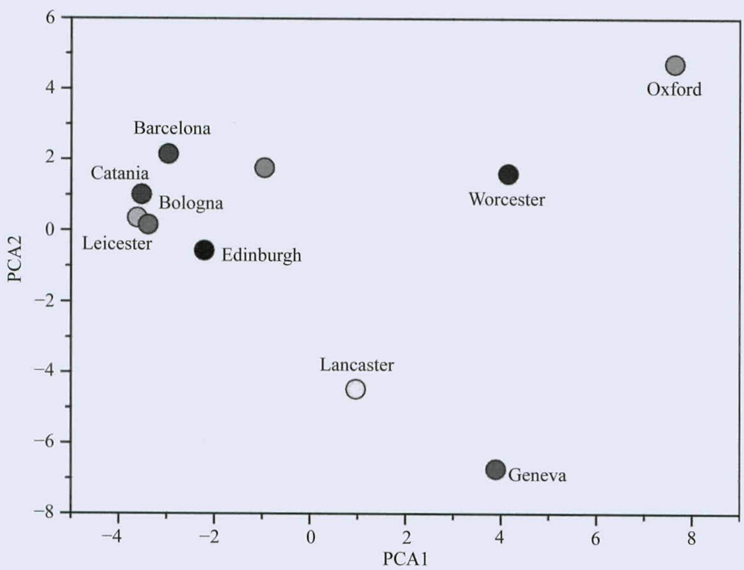

Figure 8. Principal component analysis (PCA) using only straightness and accessibility.

in figure 9 as a dendrogram, where the colors (black and grey) correspond to the two clusters identified by using a threshold parameter 0.7 of the maximum distance between two cities.

This figure clearly shows the emergence of a cluster of cities (grey) and a group of cities that are very different from each other but belong to a single family (black), so we may interpret this agglomeration as the existence of two main different classes or typologies of cities. In this classification a clear group composed of Catania, Bologna, Sheffield, Edinburgh, and Leicester is separated from Geneva, Lancaster, Oxford, Worcester, and Barcelona. Though the reasons behind this separation may be manifold and needs further experimental tests, we notice that by comparing the plans of those cities and the topography of their terrain it is possible to conclude that physical constraints may be a key factor in classifying cities. We have found that cities traversed by rivers or bordered by a lake were separated from those where the growth has never been bounded or divided by physical constraints. The exception of Barcelona, a city with a major planning event (the Plan Cerdà), confirms this hypothesis placing the great urban planning operations in the list of the geographical constraints.

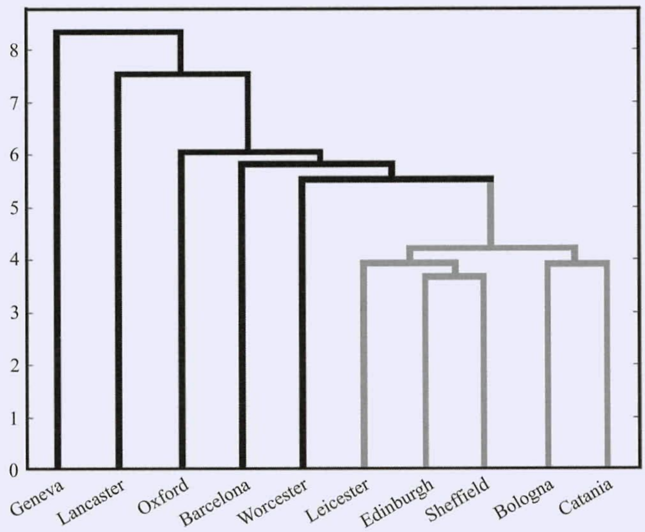

Figure 9. The dendrogram is based on the agglomerative complete linkage method of the twelve descriptive components illustrated in figure 8. The dendrogram suggests the existence of two main families of cities that are, basically, with and without big geographical constraints. The only exception is Barcelona, which even without a natural constraint was subjected to a massive and strong planning operation.

4 Discussion

In this research we analyse the street network of ten European cities represented in a primal way, where intersections are translated into the nodes of a network and the connecting streets into the links. We study first the geometric properties of the networks and then the way centrality is distributed over them according to four different definitions of centrality. We show that the selected cities share several universal geometric patterns, such as the average street length (which is limited within a remarkably narrow range of values), the distribution of angles between streets at intersections, and the distribution of the total length of dead-end streets as a function of the area of the cycle they belong to. In addition, we confirm that the distribution of street lengths follows universal power-law behavior in the long tail of the dataset, that is, for high values of street length. However, we highlight that the conventional way to represent the distribution of street lengths in search of the power-law behavior leaves out of the picture the vast majority of the data, that is, the number of streets whose length distribution cannot be accurately fitted by a power-law function is of the order of 90% of the whole dataset. We therefore investigated the actual behavior of the street-length distribution in a semilog scale finding the emergence of remarkably different behaviors that seemingly reflect the diversity of local history and conditions. Finally, we found that the distribution of the four centrality indices over the street networks allows us to characterize clearly two different clusters of cities that appear to be predominantly informed by major topographic and geographic local features such as the presence of rivers or lakes. We also found that planning events of extreme magnitude such as the occurrence of the Plan Cerdà in Barcelona may account for the uniqueness of this city as measured through the distribution of street centrality.

We show how the analysis of simple geometric properties of the street network, as well as a more complex evaluation of the street centrality, can highlight differences between cities and how these differences can be used for classifying cities in categories. It is clear that global geometric characteristics of the street network exhibit 'universal' behaviors but it is reasonable to argue that, as these behaviors also emerge in a vast range of transportation networks in nature and technology that, like the street networks, are mainly planar, the universality of those features may be related to the planarity of the systems rather than a particular 'nature' of cities as such.

Regarding the analysis of the morphology of cities, we believe our study highlights that if the universal rules governing the evolution of planar hierarchical networks are not discriminated, the extreme differences of their inner urban structure may be easily underestimated and, from a qualitative point of view, it may leave local patterns out of the picture. We focused on this discrimination process, showing how to operate at a global level in order to highlight local patterns where the extreme diversity of our cities emerges and should be accounted for, especially when we move on to the problem of classification, that is, the problem of finding a taxonomy for urban types.

Acknowledgments. The authors thank Shibu Raman, Tim Jones, Mike Jenks, and Colin Pooley for providing some of the data. Emanuele Strano is grateful to Marc Barthélemy, Salvatore Scellato, and Andrea Perna for their suggestions and help and to the Faculty of Engineering at the University of Strathclyde for financial support. Luciano Da Fontoura Costa is grateful to FAPESP (05/00587-5) and CNPq (301303/06-1 and 573583/2008-0) and Matheus Viana to FAPESP (2010/16310-0) for financial support.

References

Albert R, Barabási A-L, 1999, "Emergence of scaling in random networks" *Science* **286** 509–512

Albert R, Barabási A-L, 2002, "Statistical mechanics of complex networks" *Review of Modern Physics* **74** 47–97

Anas A, Arnott R, Small K A, 1998, "Urban spatial structure" *Journal of Economic Literature* **36** 1426–1464

Barthélemy M, 2011, "Spatial networks" *Physics Reports* **499** 1–101

Barthélemy M, Flammini A, 2008, "Modeling urban street patterns" *Physical Review Letters* **100** 138702

Batty M, 2005 *Cities and Complexity. Understanding Cities with Cellular Automata, Agent-based Models, and Fractals* (MIT Press, Cambridge, MA)

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U, 2006, "Complex networks: structure and dynamics" *Physics Reports* **424** 175–308

Bohn S, Andreotti B, Douady S, Munzinger J, Couder Y, 2002, "A constitutive property of the local organization of leaf venation networks" *Physical Review E* **65** 061914

Cardillo A, Scellato S, Latora V, Porta S, 2006, "Structural properties of planar graphs of urban street patterns" *Physical Review E* **73** 066107

Chan S H Y, Donner R V, Lämmer S, 2011, "Urban road networks - spatial networks with universal geometric features?" *European Physical Journal B* **84** 563–578

Clauset A, Shalizi C, Newman M E J, 2009, "Power-law distributions in empirical data" *SIAM Review* **51** 661–703

Couder Y, Pauchard L, Allain C, Adda-Bedia M, Douady S, 2002, "The leaf venation as formed in a tensorial field" *European Physical Journal B* **28** 135–138

Crucitti P, Latora V, Porta S, 2006, "Centrality measures in spatial networks of urban streets" *Physical Review E* **73** 036125

da F Costa L, Rodrigues F A, Travieso G, Villas Boas P R, 2007, "Characterization of complex networks: a survey of measurements" *Advances In Physics* **56** 167–242

da F Costa L, Travençolo B A N, Viana M P, Strano E, 2011, "On the efficiency of transportation systems in large cities" *Europhysics Letters* **91**(1) 18003

Expert P, Evans T S, Blondel V D, Lambiotte R, 2011, "Uncovering space-independent communities in spatial networks" *Proceedings of the National Academy of Sciences* **108** 7663–7668

Freeman C L, 1977, "A set of measures of centrality based on betweenness" *Sociometry* **40** 35–41

Freeman C L, 1979, "Centrality in social networks conceptual clarification" *Social Networks* **1** 215–239

Garrison W L, Marble D F, 1962, "The structure of transportation networks", TR, Northwest University, Evanston, IL

Hillier B, 1999 *Space is the Machine: A Configurational Theory of Architecture* (Cambridge University Press, Cambridge)
Hillier B, Hanson J, 1984 *The Social Logic of Space* (Cambridge University Press, Cambridge)
Jiang B, 2007, "A topological pattern of urban street networks: universality and peculiarity" *Physica A Statistical Mechanics and its Applications* **384** 647–655
Kansky K, Danscoine P, 1989, "Measures of network structure" *Flux* **5** 89–121
Lämmer S, Gehlsen B, Helbing D, 2006, "Scaling laws in the spatial structure of urban road networks" *Physica A Statistical Mechanics and its Applications* **363** 89–95
Latora V, Marchiori M, 2001, "Efficient behavior of small-world networks" *Physical Review Letters* **87**(19) 198701
Masucci A P, Smith D, Crooks A, Batty M, 2009, "Random planar graphs and the London street network" *European Physical Journal B* **71** 259–271
Milo R, Shen-Orr S, Itzkovitz S, Kashan N, Chklovskii D, Alon U, 2002, "Network motifs: simple building blocks of complex networks" *Science* **298** 824–827
Perna A, Kuntz P, Douady S, 2011, "Characterization of spatial networklike patterns from junction geometry" *Physical Review E* **83** 066106
Porta S, Crucitti P, Latora V, 2006a, "The network analysis of urban streets: a primal approach" *Environment and Planning B: Planning and Design* **33** 705–725
Porta S, Crucitti P, Latora V, 2006b, "The network analysis of urban streets: a dual approach" *Physica A Statistical Mechanics and its Applications* **369** 853–866
Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, Strogatz S H, 2010, "Redrawing the map of Great Britain from a network of human interactions" *PLoS ONE* **5** e14248
Sabidussi G, 1966, "The centrality index of a graph" *Psychometrika* **31** 581–603
Scott J, 2000 *Social Network Analysis: A Handbook* 2nd edition (Sage, London)
Strano E, Nicosia V, Latora V, Porta S, Barthélemy M, 2012, "Elementary processes governing the evolution of road networks" *Scientific Reports* **2** 296
Travençolo B A N, da F Costa L, 2008a, "Accessibility in complex networks" *Physics Letters A* **373** 89–95
Travençolo B A N, da F Costa L, 2008b, "Hierarchical spatial organization of geographical networks" *Journal of Physics A: Mathematical and Theoretical* **41** 224004
Travençolo B A N, Viana M P, da F. Costa L, 2009, "Border detection in complex networks" *New Journal of Physics* **11** 063019
Wasserman S, Faust K, 1994 *Social Network Analysis* (Cambridge University Press, Cambridge)