Centrality measures in spatial networks of urban streets

Paolo Crucitti,[1] Vito Latora,[2] and Sergio Porta[3]

[1]*Scuola Superiore di Catania, Italy*

[2]*Dipartimento di Fisica e Astronomia, Università di Catania, and INFN Sezione di Catania, Italy*

[3]*Dipartimento di Progettazione dell'Architettura, Politecnico di Milano, Italy*

(Received 20 October 2005; published 24 March 2006)

We study centrality in urban street patterns of different world cities represented as networks in geographical space. The results indicate that a spatial analysis based on a set of four centrality indices allows an extended visualization and characterization of the city structure. A hierarchical clustering analysis based on the distributions of centrality has a certain capacity to distinguish different classes of cities. In particular, self-organized cities exhibit scale-free properties similar to those found in nonspatial networks, while planned cities do not.

A particular class of complex networks [1–3] is composed of those embedded in real space—i.e., networks whose nodes occupy a precise position in two- or three-dimensional Euclidean space and whose edges are real physical connections. With a few exceptions [4–6], most of the works in the literature have focused on the characterization of the topological (relational) properties of *spatial networks*, while the spatial aspect has received less attention, when not neglected at all. However, it is not surprising that the topology of spatial networks is strongly constrained by their geographical embedding. For instance, the number of long-range connections [4–6] and the number of edges that can be connected to a single node [5] are limited by the spatial embedding. This is particularly evident in planar networks (e.g., those networks forming vertices whenever two edges cross) [7], as urban streets or ant networks of galleries [5], and has important consequences on the possibility to observe a small-world behavior or scale-free degree distributions [1–3]. Consequently, spatial networks are different from relational networks, and as such they need to be treated.

Centrality has remained a fundamental concept in network analysis since its introduction in structural sociology [8,9]. The network approach has also a long tradition in economic geography and city planning, where it has been used to investigate the territorial relationships among communication flows, population, wealth, and land uses [10]. However, when dealing with *urban street patterns*, centrality has been studied in relational networks only [11–13], neglecting a fundamental aspect such as the geography. In such an approach, known as the *dual representation* [12,13] or *information city network* [14], a city is transformed into a relational (topological) graph by mapping the streets onto the graph nodes and the intersections between streets onto edges between the nodes. In the present paper, we study centrality in urban street patterns of different world cities represented as spatial networks. In our approach, which is opposite to the dual one, we work within a fully metric framework in which the distance has to be measured not just in topological terms (steps), like in the dual representation of a city [12–14] or in social [9] and other complex systems [1], but rather in properly spatial terms (meters, miles). The results indicate that a spatial analysis based on a set of different centrality measures (properly extended for spatial graphs) allows (1) a vi-

sual characterization of the structural properties of a city, (2) the evidence that planned and self-organized cities belong to two different classes, and (3) to find scale-free properties similar to those found in the degree distributions of relational (nonspatial) networks.

We have selected 18 one-square-mile samples of different world cities from Ref. [15], imported them in a GIS (Geographic Information System) environment and constructed spatial graphs of street networks. In our approach, each urban street sample is turned into an undirected, valued graph G, where intersections are nodes and streets are edges. We denote by N the number of nodes and by K the number of edges. The nodes are characterized by their position $\{x_i, y_i\}_{i=1,\ldots,N}$ in the unit square. The obtained graphs can be described by the adjacency matrix A, whose entry a_{ij} is equal to 1 when there is an edge between i and j and 0 otherwise, and by a matrix L, whose entry l_{ij} is the value associated to the edge, in our case the length of the street connecting i and j. The considered cities exhibit striking differences in terms of cultural, social, economic, religious, and geographic context and can be roughly divided into two large classes: (1) patterns, such as Ahmedabad, Cairo, and Venice, grown throughout a largely self-organized, fine-grained historical process, out of the control of any central agency, and (2) patterns, such as Los Angeles, Richmond, and San Francisco, realized over a short period of time as the result of a single

TABLE I. Basic properties of the spatial graphs obtained from 18 one-square-mile samples of different world cities.

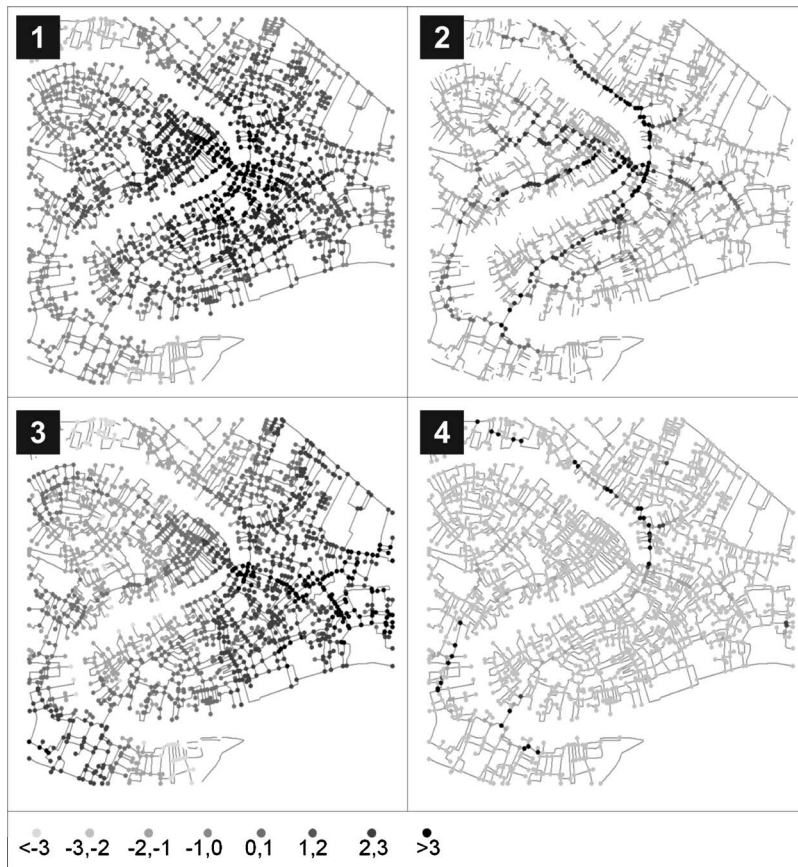|	Case	N	K		Case	N	K
1	Ahmedabad	2870	4387	10	Paris	335	494
2	Barcelona	210	323	11	Richmond	697	1086
3	Bologna	541	773	12	Savannah	584	958
4	Brasilia	179	230	13	Seoul	869	1307
5	Cairo	1496	2255	14	San Francisco	169	271
6	Los Angeles	240	340	15	Venice	1840	2407
7	London	488	730	16	Vienna	467	692
8	New Delhi	252	334	17	Washington	192	303
9	New York	248	419	18	Walnut Creek	169	197

FIG. 1. Color-coded maps representing the spatial distributions of node centrality in Venice. The four indices (1) closeness C^C, (2) betweenness C^B, (3) straightness C^S, and (4) information C^I are visually compared over the spatial graph. Different colors represent classes of nodes with different values of centrality (the classes are defined in terms of multiples of standard deviations from the average, as reported in the color legend).

plan, usually exhibiting a regular grid like structure. The basic characteristics of the derived graphs are reported in Table I. N and K assume widely different values, notwithstanding the fact we have considered the same amount of land. The edge value (i.e., the street length) distribution $P(l)$ is single peaked in self-organized cities, while it shows many peaks in single planned cities, due to their grid patterns [16]. Other characteristics of the same graphs, such as the fractal space distribution of the nodes, meshedness coefficients, and cycles, efficiency, and costs, have been investigated in Ref. [17].

For each of the 18 cities we have evaluated the 4 following node centrality indices.

(1) *Closeness centrality* C^C measures to which extent a node i is near to all the other nodes along the shortest paths and is defined as [9]

$$C_i^C = \frac{N-1}{\sum_{j \in G; j \neq i} d_{ij}}, \tag{1}$$

where d_{ij} is the shortest path length between i and j, defined, in a valued graph, as the smallest sum of the edge length l throughout all possible paths in the graph between i and j.

(2) *Betweenness centrality* C^B is based on the idea that a node is central if it lies between many other nodes, in the sense that it is traversed by many of the shortest paths connecting couples of nodes. The betweenness centrality of node i is [9,18]

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

where n_{jk} is the number of shortest paths between j and k, and $n_{jk}(i)$ is the number of shortest paths between j and k that contain node i.

(3) *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 lenght d_{ij} [19,20]. The straightness centrality of node i is defined as

$$C_i^S = \frac{1}{N-1} \sum_{j \in G, j \neq i} d_{ij}^{Eucl}/d_{ij}, \tag{3}$$

where d_{ij}^{Eucl} is the Euclidean distance between nodes i and j along a straight line, and we have adopted a normalization recently proposed for geographic networks [21]. This measure captures to which extent the connecting route between nodes i and j deviates from the virtual straight route.

(4) *Information centrality* C^I relates the node centrality to the ability of the network to respond to the deactivation of the node [22,23]. The information centrality of node i is defined as the relative drop in the network efficiency $E[G]$ caused by the removal from G of the edges incident in i:
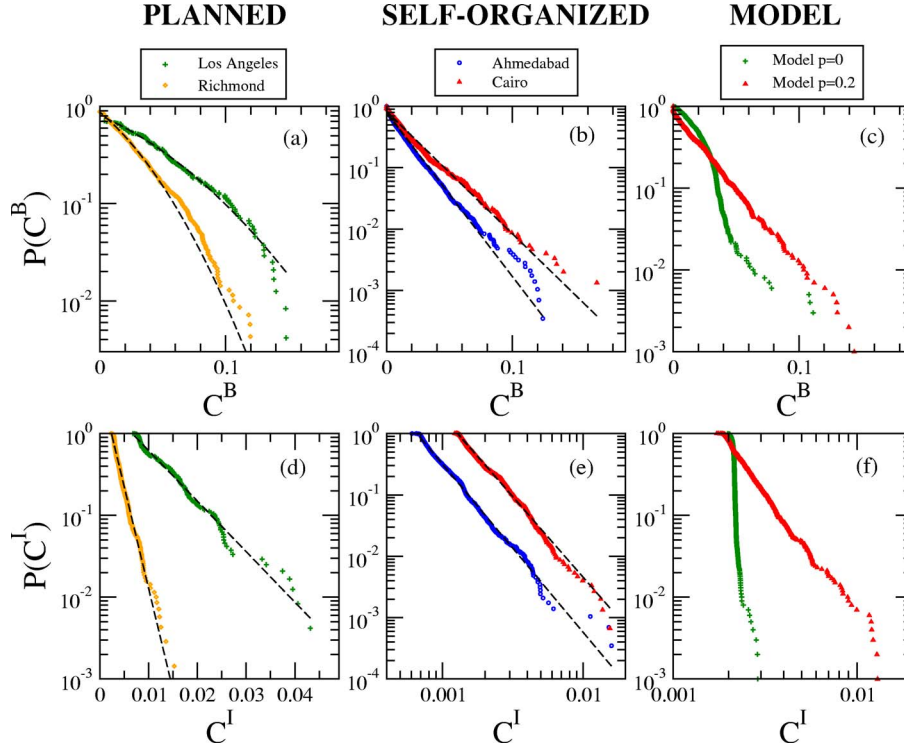
FIG. 2. (Color online) Cumulative distributions of betweenness C^B (a), (b), (c) and information C^I (d), (e), (f) for two planned cities (Los Angeles and Richmond), two self-organized cities (Ahmedabad and Cairo), and the model discussed in the text. The cumulative distributions $P(C)$ are defined by $P(C) = \int_C^{+\infty} \frac{N(C')}{N} dC'$, where $N(C)$ is the number of nodes having centrality equal to C. $P(C^B)$ are single scale in all cases: the dashed lines in panels (a) and (b) are, respectively, exponential, $P(C) \sim \exp(-C/s)$ ($s_{Ahm} = 0.016$, $s_{Cai} = 0.022$), and Gaussian, $P(C) \sim \exp(-x^2/2\sigma^2)$ ($\sigma_{LA} = 0.078$, $\sigma_{Rich} = 0.049$), fits to the empirical distributions. Conversely, $P(C^I)$ differentiates self-organized cities from planned ones: the dashed lines in the log-log plot of panel (d) indicate that the information centrality follows a power law $P(C) \sim C^{-\gamma}$ for the two self-organized cities ($\gamma_{Ahm} = 2.74$, $\gamma_{Cai} = 2.63$), whereas the dashed lines in panel (e) indicate an exponential distribution $P(C) \sim \exp(-C/s)$ for the two planned cities ($s_{LA} = 0.007$, $s_{Rich} = 0.002$). In panel (f), $P(C^I)$ is exponential in the model with $p=0$ and power law for $p=0.2$.

$$C_i^I = \frac{\Delta E}{E} = \frac{E[G] - E[G']}{E[G]}, \qquad (4)$$

where the efficiency of a graph G is defined as [19,20]

$$E[G] = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} d_{ij}^{Eucl}/d_{ij} \qquad (5)$$

and where G' is the graph with N nodes and $K - k_i$ edges obtained by removing from the original graph G the k_i edges incident in node i. An advantage of using the efficiency to measure the performance of a graph is that $E[G]$ is finite even for disconnected graphs.

Spatial distribution of centralities. The spatial distribution of node centralities can be visualized by means of color-coded maps such as the one of Venice reported in Fig. 1. The figures for the other cities can be downloaded from our website [24]. A comparison with the spatial distribution of the same measures in a pure connectivity network can be found in Ref. [25]. As shown in Fig. 1, C^C exhibits a strong trend to group higher scores at the center of the image. This is both due to the nature of the index and to the artificial boundaries imposed by the one-square-mile map representation. The spatial distribution of C^B nicely captures the continuity of prominent urban routes across a number of intersections,

changes in direction, and focal urban spots. In the case of Venice the most popular walking paths and squares ("campi") and the Rialto bridge over the Canal Grande emerge along the red nodes routes. In most of the cities considered, C^B is also able to identify the primary structure of movement channels as different from that of secondary, local routes [24]. The spatial distribution of C^S depicts both linear routes and focal areas in the urban system: C^S takes high values along the main axes, even higher at their intersections. Finally C^I exhibits a spatial distribution that is in many cases similar to that of C^B.

Statistical distribution of centralities. Notwithstanding the similarities in the color maps, C^I and C^B exhibit different statistical distributions. This is illustrated in Fig. 2, where we report an example of the cumulative distributions for the two categories of cities. Closeness, straightness (not shown in figure), and betweenness distributions are quite similar in both self-organized and planned cities, despite the fact that the diversity of the two cases in sociocultural and economic terms could not be deeper. In particular, C^B exhibits a single-scale distribution [26] in self-organized and in planned cities, the former having an exponential distribution, the latter having a Gaussian distribution, as, respectively, shown in Figs. 2(a) and 2(b). Conversely, the distribution of C^I is single scale for planned cities and broad scale for self-organized
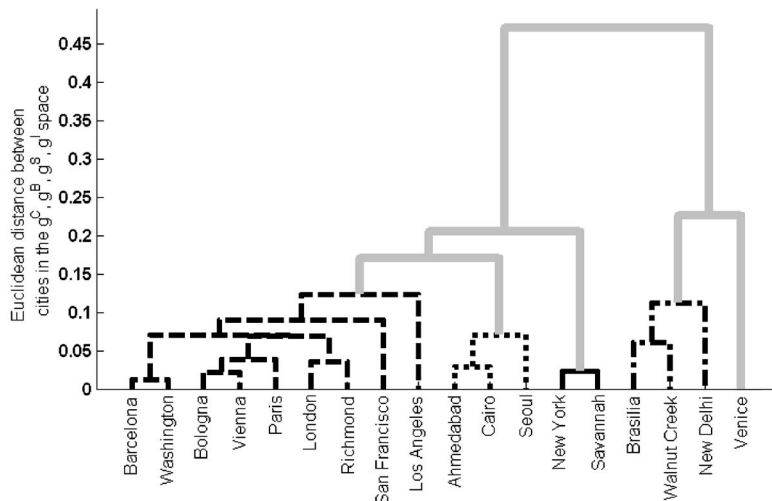
FIG. 3. Hierarchical tree based on g^C, g^B, g^S, and g^I. The complete linkage method, based on the largest distance between objects in different clusters, has been applied. By choosing a maximum distance equal to 0.15 for two cities to belong to the same cluster, we find five clusters: a first cluster from Barcelona to Los Angeles, a second cluster from Ahmedabad to Seoul including self-organized cities, a third cluster made up by New York and Savannah, a fourth cluster from Brasilia to New Delhi, and a fifth cluster with only Venice.

cities: e.g., Los Angeles and Richmond are well fitted by exponential curves [Fig. 2(d)], while Ahmedabad and Cairo are fitted by power laws $P(C) \sim C^{-\gamma}$ with exponents $\gamma_{Ahm} = 2.74$ and $\gamma_{Cai} = 2.63$ [Fig. 2(e)]. Among the self-organized cities, Venice is the one with the smallest value of the exponent: namely, $\gamma_{Ven} = 1.49$. This is due to the particular environmental constraints that have shaped the historical structure of the city. The identified power laws indicate a highly uneven distribution of C^I over networks of self-organized cities. This can be considered as the analog, for spatial networks of urban streets, of the power laws observed in the degree and in the betweenness distributions of many nonspatial complex networks from biology and technology [1–3].

To reproduce the empirical distributions we have considered the following model: N nodes are initially placed on a rectangular grid; with a probability p each node is moved to a random position in the unit square; for each node i, two new edges (i,j) and (i,k) are added, where j and k are the two nearest nodes among those not yet connected to i. The model interpolates from a regular grid, for $p=0$, to a graph with randomly placed nodes, for $p=1$. The distribution of C^B in the model is single scale for any value of p. In particular, for values of p in the range 0.1–0.3, $P(C^B)$ is exponential as in self-organized cities. Conversely, $P(C^B)$ is single scale for low values of p and follows a power law for intermediate values of p. We have found that the centrality distribution in planned cities is well reproduced by the model with $p \sim 0$ (or by triangular and square grids), while self-organized cities by the model with $p \sim 0.1$–0.3. The distributions obtained for $N=900$, $p=0$, and $p=0.2$ are reported in Figs. 2(c) and 2(f).

Inequalities in the distribution of centrality among the nodes of the network can be quantified by evaluating the Gini coefficient of the distribution. The Gini coefficient is a measure used in economics and ecology to describe inequalities in the distribution of a given resource among the individuals of a population. To compute the Gini coefficient g of a given empirical distribution, one needs first to define the so-called Lorenz curve of the ranked empirical distribution, which is a curve that shows, for the bottom $x\%$ of individuals in the population, the percentage $y\%$ of the

total size (resource) which they have. In the case of a perfectly uniform distribution (in which every person has the same size), the bottom $x\%$ of individuals would always have $y = x\%$ of the total size. Thus the Lorenz curve for a uniform distribution is the straight line $y = x$: this line is usually called the line of perfect equality. On the other hand, a distribution in which one single individual has all the resources and everyone else has none produces a Lorenz curve that is at $y=0$ for all $x < 100$ and $y=100$ when $x=100$: this is called the line of perfect inequality. Finally, the Gini coeffiecient g is defined as the area between the Lorenz curve of the empirical distribution and the line of perfect equality, divided by the area between the line of perfect equality and the line of perfect inequality [27]. The coefficient g ranges from a minimum value of zero, when all individuals are equal, to a maximum value of 1, in a population in which every individual, except one, has a size of zero. For each of the cities we have evaluated four Gini coefficients g^C, g^B, g^S, and g^I, one for each of the centrality measures. E.g., the Gini coefficient g^I is 0.12 for New York, 0.19 for Richmond, and 0.23 for Cairo, thus indicating that Cairo has a more heterogeneous information centrality distribution than that of Richmond and New York. In Fig. 3 we show the results of a hierarchical clustering analysis based on the homogeneity and heterogeneity of the networks, as measured by the four Gini coefficients. The iterative pairing of cities obtained captures some basic classes of urban patterns: it is the case of the early association of Barcelona and Washington or New York and Savannah, all grid-iron planned cities as well as that of Bologna, Wien, and Paris, all mostly medieval organic patterns, or that of Ahmedabad and Cairo. Brasilia, Walnut Creek, and New Delhi, in this respect, share a planned, large-scale modernist formation. Venice is the last association, which tells of the unique mix of fine-grained pattern and natural constrains that have shaped the historical structure of the city.

We have proposed a comparative analysis of different centrality measures in spatial networks of urban streets. Each centrality captures a different aspect of one place's "being central" in geographic space, and by the use of many centrality measures it is possible to capture structural similarities and dissimilarites across cities. Our work opens up to an

in-depth investigation of the correlation between the structural properties of the system and the relevant dynamics on the system, like pedestrian and vehicular flows and retail commerce vitality, all information traditionally associated with spatial graphs. We expect that some of these factors are more strictly correlated to some centrality indices than to others, thus giving informed indications for strategies of urban planning and design.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] S.-H. Yook, H. Jeong, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **99**, 13382 (2002).

[5] J. Buhl, J. Gautrais, R. V. Solé, P. Kuntz, S. Valverde, J. L. Deneubourg, and G. Theraulaz, Eur. Phys. J. B **42**, 123 (2004).

[6] M. T. Gastner and M. E. J. Newman, e-print cond-mat/0407680.

[7] D. B. West, *Introduction to Graph Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

[8] A. Bavelas, Hum. Organ **7**, 16 (1948).

[9] S. Wasserman and K. Faust, *Social Networks Analysis* (Cambridge University Press, Cambridge, England, 1994).

[10] G. A. Wilson, *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis* (Prentice-Hall, Upper Saddle River, NJ, 2000).

[11] B. Hillier and J. Hanson, *The Social Logic of Space* (Cambridge University Press, Cambridge, England, 1984).

[12] B. Jiang and C. Claramunt, Environ. Plan. B: Plan. Des. **31**, 151 (2004).

[13] S. Porta, P. Crucitti, and V. Latora, e-print cond-mat/0411241; Physica A (to be published).

[14] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).

[15] A. Jacobs, *Great Streets* (MIT Press, Boston, 1993).

[16] P. Crucitti, V. Latora, and S. Porta, Chaos (to be published).

[17] A. Cardillo, S. Scellato, V. Latora, and S. Porta, e-print physics/0510162.

[18] L. C. Freeman, Soc. Networks **1**, 215 (1979).

[19] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[20] V. Latora and M. Marchiori, Eur. Phys. J. B **32**, 249 (2003).

[21] I. Vragovìc, E. Louis, and A. Dìaz-Guilera, Phys. Rev. E **71**, 036122 (2005).

[22] V. Latora and M. Marchiori, e-print cond-mat/0402050.

[23] V. Latora and M. Marchiori, Phys. Rev. E **71**, 015103(R) (2005).

[24] Website at http://www.ct.infn.it/˜latora/1sqml.html.

[25] S. Porta, P. Crucitti, and V. Latora, e-print physics/0506009, Environ. Plann. B (to be published).

[26] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[27] C. Dagum, Econ. Appl. **33**, 327 (1980); P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, Ecology **68**, 1548 (1987).