The spatial structure of networks

M.T. Gastner[1,a] and M.E.J. Newman[2]

[1] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[2] Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109–1040, USA

Abstract. We study networks that connect points in geographic space, such as transportation networks and the Internet. We find that there are strong signatures in these networks of topography and use patterns, giving the networks shapes that are quite distinct from one another and from non-geographic networks. We offer an explanation of these differences in terms of the costs and benefits of transportation and communication, and give a simple model based on the Monte Carlo optimization of these costs and benefits that reproduces well the qualitative features of the networks studied.

PACS. 89.75.Hc Networks and genealogical trees – 87.23.Ge Dynamics of social systems – 05.90.+m Other topics in statistical physics, thermodynamics, and nonlinear dynamical systems – 64.60.Ak Renormalization-group, fractal, and percolation studies of phase transitions

1 Introduction

There has in the last few years been considerable interest within the physics community in the analysis and modeling of networked systems including the world wide web, the Internet, and biological, social, and infrastructure networks [1–3]. Some of these networks, such as biochemical networks and citation networks, exist only in an abstract "network space" where the precise positions of the network nodes have no particular meaning. But many others, such as the Internet, live in the real space of everyday experience, with nodes (e.g., computers in the case of the Internet) having well-defined positions. Most previous studies of real-world networks have ignored geography, concentrating instead on other issues. Here we argue that geography matters greatly, and to ignore it is to miss some of these systems' most interesting features.

A network in its simplest form is a set of nodes or *vertices* joined together in pairs by lines or *edges*. We consider networks in which the vertices occupy particular positions in space. The edges in these networks are often real physical constructs, such as roads or railway lines in transportation networks [4], optical fiber or other connections in the Internet [5,6], cables in a power grid [7], or oil pipelines [8]. In other cases the edges may be more ephemeral, such as flights between airports [9], business relationships between companies [10], or wireless communications [11].

Interest in the spatial structure of networks dates back to the economic geography movement of the 1960s [12,13] and particularly the work of Kansky [14]. Early work was

hampered however by limited data and computing resources, and geographers' attention moved on after a while to other topics. Networks have come back into the limelight in recent years, particularly as a result of interest among physicists. However, empirical studies of networks, even networks in which geography plays a pivotal role, have, with some exceptions [6,9,15], focused mostly on topological features [4,17,18]. Similarly, the best-known theoretical models of networks either make no reference to space at all [19,20], or they place vertices on simple regular lattices whose structure is quite different from that of real systems [7,21]. Recently several models have been proposed that incorporate more sophisticated geographical considerations [22–26], although the empirical data against which one might verify such models is still lacking.

Here we analyze empirical data for a number of spatial networks, finding some clear patterns that imply a connection between network structure and geography. We also propose a simple model of a spatial network that offers an explanation for some of these patterns.

2 Empirical data

In this paper we look in detail at three specific networks, particularly concentrating on their spatial form. The three networks are the Internet, a road network, and a network of passenger flights operated by a major airline. To make comparison between the networks easier we limit our studies to the United States, and we exclude Alaska and Hawaii to avoid problems of disjoint maps.

[a] e-mail: mgastner@santafe.edu

The first of our three networks is the Internet. We examine the network in which the vertices are autonomous systems (ASes) and the edges are data connections between them (technically, direct-peering relationships). The topology of the connections between ASes can be inferred from routing tables. In our studies we have made use of the collection of routing tables compiled by the University of Oregon's Route Views project (http://www.routeviews.org). To determine the geographical parameters of the network we use NetGeo (http://www.caida.org/tools/utilities/netgeo), a software tool that can return approximate latitude and longitude for a specified AS. Combining these two resources a geographic map of the Internet was created, from which were then deleted all nodes falling outside the lower 48 states. This leaves a network of 7049 nodes and 13 831 edges for data from March 2003.

Our second network is the US interstate highway network in which the vertices represent intersections, termination points of highways, and country borders, and the edges represent highways. Vertex positions and edges were extracted from GIS databases. For data from the year 2000 the network has 935 vertices and 1337 edges. Our third network, the airline network, is similarly straightforward. In this network the vertices represent airports and there is an edge between every pair of airports connected by a scheduled flight. The particular case we study is the published schedule of flights for Delta Airlines for February 2003, for which there are 187 vertices and 825 edges. Geographic locations of airports were found from standard directories.

We focus initially in our analysis of these networks on three fundamental properties: edge lengths, network diameter, and vertex degrees. In Figure 1 we show the distribution of the lengths in kilometers of edges in each of our networks. Common to all three networks is a clear bias towards shorter edges, which is unsurprising since long edges are presumably more expensive to create and maintain than short ones. When we look more closely, however, the networks show some striking differences. The road network has only very short edges, on the order of 10 km to 100 km, while the Internet and airline network have much longer ones. The latter two networks also both have bimodal distributions, with a large fraction of edges of length 2000 km or less, and then a smaller but distinct peak of longer edges around 4000 km[1]. (These are continent-spanning edges, like coast-to-coast flights in the airline network.)

Simple Euclidean distance between vertices is not the only measure of distance in a network however. Another commonly used measure is the so-called graph distance,

---

[1] Although we do not dwell on it in this paper, the Internet and the airline network do differ at the shortest edge lengths, the Internet having a strong peak for edges of 100 km or less, while the airline network seems deliberately to avoid such short edges, having a dip in the distribution at the shortest length scales. This is presumably an effect of economic pressures: very short airline flights are uneconomical because passengers can conveniently drive the same distance for less money.
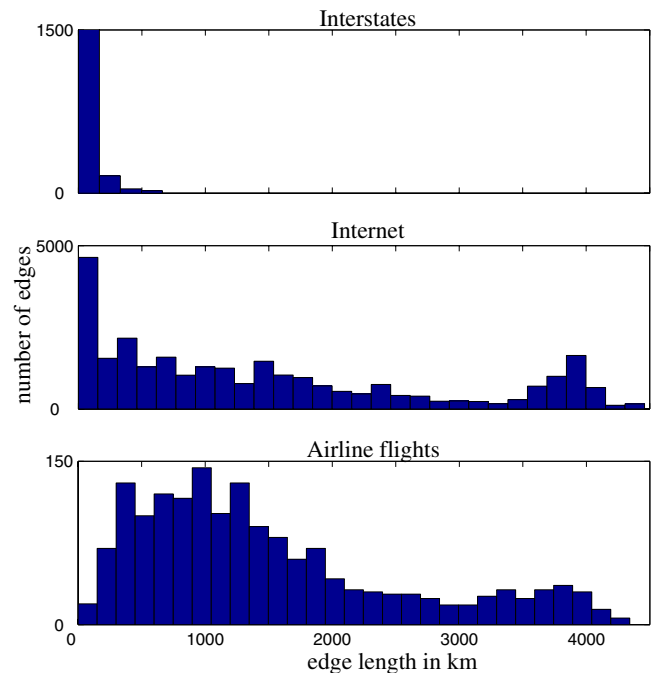
Fig. 1. Histograms of the lengths of edges in the three networks studied here.

which measures the number of edges traversed along the shortest path from one vertex to another – the number of "legs" of air travel, for instance, or the number of "hops" an Internet data packet would make. The largest graph distance between any two points in a network is called the graph diameter, and it varies widely between our networks. For the highway network for example the diameter is 61, but it is just 8 for the Internet, even though the latter network has far more vertices. And for the airline network the diameter is only 3. In the jargon of the networks literature, the Internet and the airline network form "small worlds," while the interstate network does not.

Euclidean edge lengths and graph distances are not unrelated: in a graph like the road network, which is composed mainly of short edges, one will need to traverse a lot of such edges to make a long journey, so we would expect the diameter to be large. Conversely, the presence of even just a few long edges makes for much smaller diameters, as demonstrated recently by Watts and Strogatz [7]. Thus there seems to be a pay-off between Euclidean distance and number of legs in a journey, an idea that we exploit below to help explain the observed structure of our networks.

Another way in which our networks differ is in the degrees of their vertices. (The degree of a vertex is the number of edges connected to it.) The highest degree of any vertex in the highway network is 4, which means that the best connected vertex links directly to only 0.4% of other vertices. In the airline network by contrast, the maximum degree is 141 or 76% of the network, while for the Internet it is 2139 or 30%. High-degree vertices that connect to a significant fraction of the rest of the network are commonly called "hubs"; the airline network and Internet

thus both contain at least one hub (in fact each contains several), whereas the road network contains none[2].

We would like to understand how the observed structure of our networks is related to their geographical nature, and the origin of the marked differences between the networks. We present two approaches that shed light on these questions. The first is empirical in nature, the second theoretical.

3 Network dimension

At the empirical level, many of the features we observe in these networks can be explained in terms of spatial dimension. Each of our networks is of course two-dimensional in a geographic sense, since it lives on the two-dimensional surface of the Earth. However, one can also ask about the effective dimension of the network itself [27]. We find that, in a sense we will shortly define, the Internet and airline networks are not really two-dimensional at all, but the road network is.

The road network is, in fact, almost planar. That is, it can be drawn on a map without any edges crossing. This automatically gives it a two-dimensional form and helps us to understand why its edges are so short: if edges are not allowed to cross then they cannot travel far before they run into one another. It also goes some way towards explaining the network's low vertex degrees: it can be proved that the mean degree $\bar{k}$ of a planar graph is strictly less than 6 [28] and indeed we find that the mean degree of the road network is $\bar{k} = 2.86$. For the airline network on the other hand $\bar{k} = 8.82$, so this network cannot be planar. This is not an entirely persuasive argument however. The Internet has mean degree $\bar{k} = 3.93$, which is not large enough to rule out planarity, and the highway network is actually not perfectly planar, having a small number of road crossings so that rigorous demonstrations of planarity such as Kuratowski's theorem [28] or the Hopcroft–Tarjan planarity algorithm [29] fail. We would like, therefore, some other more flexible way of probing the dimension of our networks. We propose the following.

On an infinite regular d-dimensional lattice, such as a square or cubic lattice, the dimension d can be calculated from $d = \lim_{r\to\infty} d \log N_v(r)/d \log r$ where $N_v(r)$ is the number of vertices r steps or less from a given vertex v [27]. On finite lattices one cannot take the limit $r \to \infty$, but good results for d can be achieved by plotting $\log N_v$ against $\log r$ for some central vertex v and measuring the slope of the initial part of the resulting line. Applying the same method to other networks allows us to define a dimension in more general cases as well. It's important to realize, as pointed out in references [27,30], that this dimension need not coincide with the dimension of the underlying space (if any) on which our networks are built. If it does coincide, however, it provides a strong

indication that the network structure has been influenced by this space.

We show results for $N_v(r)$ for the interstate network and the Internet in Figure 2, panels (a) and (b). (To reduce statistical errors, N_v is averaged over all vertices v.) As the figure shows, the slope of the plot is close to 2 for the interstates, indicating that this network is essentially two-dimensional. For the Internet on the other hand, the plot grows much faster with r, indicating that the network has high dimension, or perhaps no well-defined dimension at all (similar results are seen for the airline network).

If a network is fundamentally two-dimensional, then we would expect it to have a diameter that, like any two-dimensional system, varies as the square root of the network size. Networks without a well-defined low dimension, by contrast, have diameters varying much more slowly, usually logarithmically with network size. Thus, we propose a tentative explanation of the structure of our geographic networks as follows. All our networks appear to show a preference for short edges over long ones, which is a natural effect of geography. However, the road network has much shorter edges, lower degrees, and larger diameter than the other two. These are all expected consequences of a two-dimensional or near-planar form, and when we measure dimension we do indeed find that the road network is fundamentally two-dimensional, while the other networks are not.

4 A model of a spatial network

This is a satisfying finding, certainly, but to some extent it just passes the intellectual buck: our measurements can be explained in terms of network dimensionality, but why do the networks have different dimension in the first place? As we now show, it is possible to construct a simple model that explains the basic features of geographic networks, including their dimension, in terms of competing preferences for either short Euclidean distances between vertices or short graph distances.

First, let us assume that the cost of building and maintaining a network is proportional to the total length of all its edges:

$$\text{cost} = \sum_{\text{edges } (i,j)} d_{ij}, \qquad (1)$$

where d_{ij} is the Euclidean length of the edge between vertices i and j. This result is only approximately true in most cases, but it is a plausible starting point.

From a user's perspective, a network will usually be better if the paths between points are shorter. As we have seen, however, the way we measure path length can vary. In a road network most travelers look for routes that are short in terms of miles, while for airline travelers the number of legs is often considered more important. To account for these differences, we assign to each edge an effective length thus:

$$\text{effective length of edge } (i,j) = \lambda\sqrt{n}\, d_{ij} + (1 - \lambda), \quad (2)$$

---

[2] The existence of hubs in the airline network is of course well known to travelers, and their existence in the Internet has also been known for some years [17].
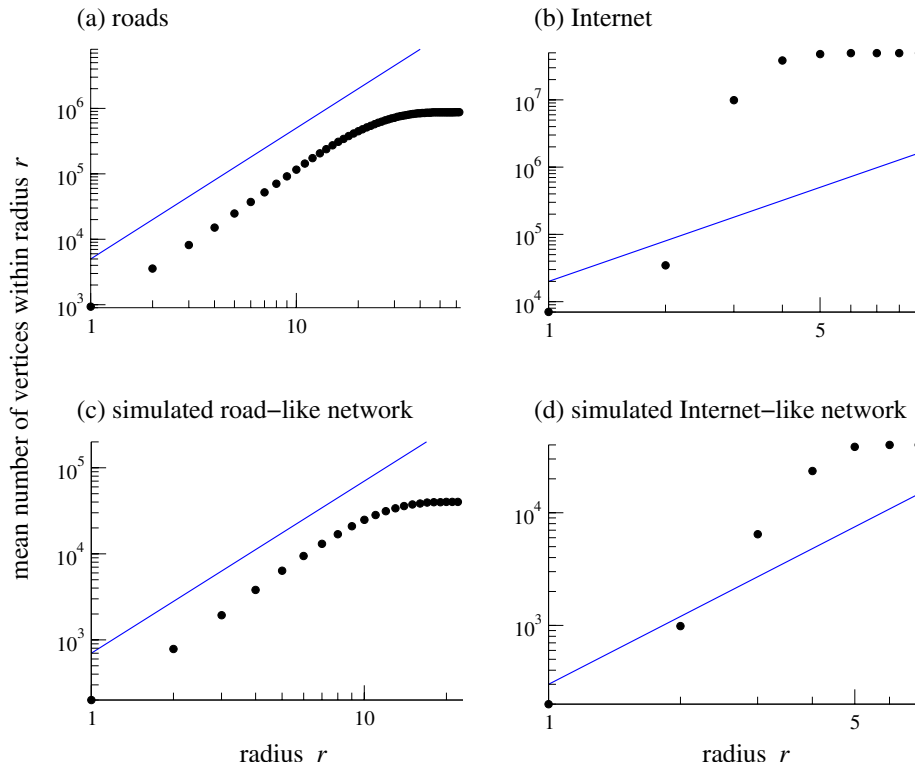
Fig. 2. The size of neighborhoods vs. their radius on doubly-logarithmic plots (a) for interstate highways, (b) for the Internet, (c) and (d) for simulations based on the optimization model described in the text ($\lambda = 1.0$ and $\lambda = 0.1$ in Eq. 2). The straight lines have slope 2 and indicate the expected growth for two-dimensional networks.

where $0 \leq \lambda \leq 1$ and n is the number of vertices. The parameter λ determines the user's preference for measuring distance in terms of miles or legs. (The factor of $\sqrt{n}$ is not strictly necessary but it is convenient; it compensates for the scaling of nearest-neighbor distances $d_{ij} \sim n^{-1/2}$ with system size.) If there is no path between i and j we formally set the effective length to infinity. Now we define the total distance between two (not necessarily adjacent) vertices to be the sum of the effective lengths of all the edges along a path between them, minimized over all paths.
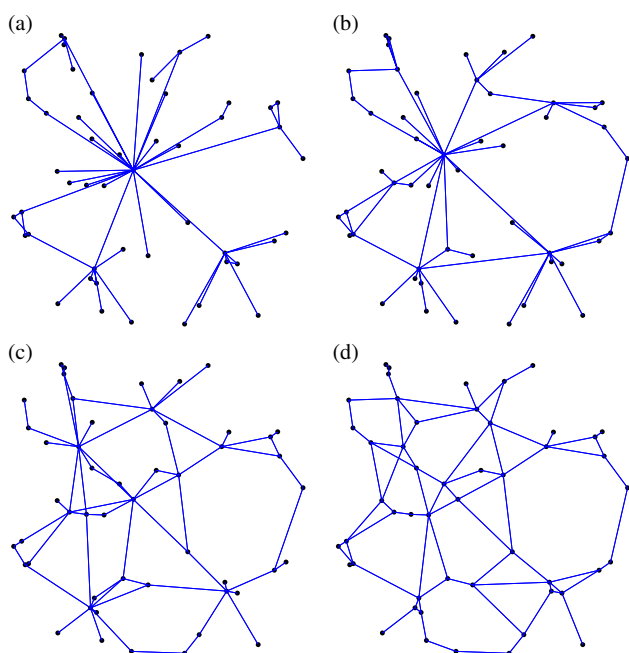
We now construct a model network as follows. We suppose we are given the positions of n vertices that we are to connect[3], we are given a budget, equation (1), for building the network, and we are given the preference of the users, meaning we are given a value of λ. We then search for network structures that connect all the vertices, can be built within budget, and minimize the mean vertex–vertex distance between all vertex pairs, for edge lengths defined as above. If the budget is smaller than the cost of building the minimum spanning tree, no connected network can be built and the mean distance is infinity, whereas an unlimited budget would allow us to construct a fully connected network with $n(n-1)/2$ edges. For intermedi-

ate budgets we have to solve a combinatorial optimization problem [32], for which we can derive good (though usually not perfect) solutions using simulated annealing. Here the mean effective vertex–vertex distance is treated as an energy whose minimum we wish to calculate in the spirit of [33].

Figure 3 shows four networks generated in this fashion for $n = 50$ vertices placed at random within a square. For $\lambda = 0$ and $\lambda = 1$ we find networks strongly reminiscent of airlines and roads respectively – tree-like structures with long edges and hubs in the first case and structures with many loops but neither long edges nor hubs in the second. For intermediate values of λ the model finds a compromise between hub formation and local links. Note that we see hub formation for small λ even though there is no "preferential attachment" mechanism in our model, unlike some previous models [23–25,34].

To make the comparison with the real-world networks more concrete, we have generated networks using our model with the same mean degrees as the real-world networks. For $n = 200$ nodes, we find that the maximum degree of the model networks varies between 7 (3.5% of the network) and 143 (71.5%) as we vary λ from 0 to 1. At the same time, the diameter decreases from a sizable 21 to a small-world-like 4. In Figures 2c and 2d we show the mean size of the neighborhood $N_v(r)$ of a vertex as a function of distance r, as we did for our empirically observed networks. As the figure shows, the results indicate a network with a roughly two-dimensional form for large λ

---

[3] The formation of cities usually predates the existence of highway or airline networks, so in the context of this paper it is reasonable to assume that the vertices are given from the start. The alternative possibility, that the vertices are added as they connect to the network has been investigated by, for example, Fabrikant et al. [31].

Fig. 3. Optimized networks created using the model described in the text starting with 50 points placed uniformly at random in the unit square, a total budget of 10.0, and (a) $\lambda = 0$, (b) $\lambda = \frac{1}{3}$, (c) $\lambda = \frac{2}{3}$, (d) $\lambda = 1$. Networks (a) and (d) resemble airline and road networks respectively, while (b) and (c) show structure intermediate between the two extremes.
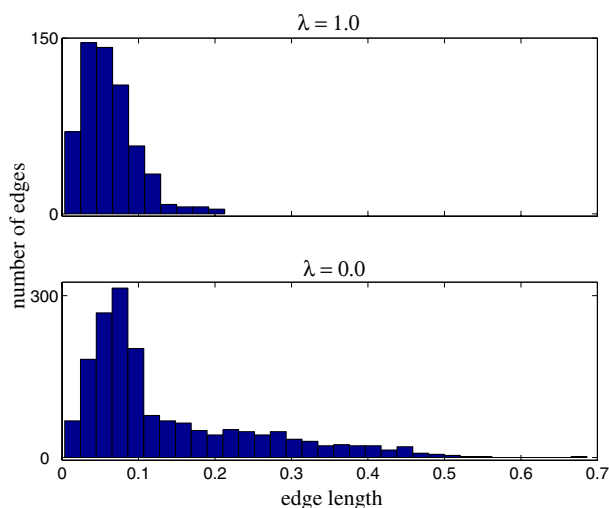
Fig. 4. Histograms of the Euclidean lengths of edges in simulated networks for $\lambda = 1.0$ (top) and $\lambda = 0.0$ (bottom). The vertices were placed uniformly at random in the unit square.

(Fig. 2c) and a strongly super-quadratic form for small λ (Fig. 2d).

In Figure 4 we show a histogram of the distribution of the lengths of edges in the model for $\lambda = 0$ and 1. Again the model shows behavior qualitatively similar to the real-world networks (Fig. 1). For large λ most edges are short, but for smaller λ a small fraction of edges span

distances on the order of the system size. These "short-cut" edges substantially reduce the mean graph distance between vertices, as in the model of reference [7], but cost relatively little because they are few in number.

5 Conclusions

In this paper we have studied empirically a number of networks whose vertices exist in the two-dimensional geographic space defined by the surface of the Earth. Our analysis indicates that there are strong connections between the geographical and topological features of these networks, as well as substantial qualitative differences between the various networks. In particular, we find that a simple measure of effective dimension distinguishes strongly between networks such as the road network, which is fundamentally two-dimensional, and the airline network, which has a much higher dimension.

We have also proposed a simple model of a geographic network that balances the extent to which users of the network care about geographic distance between vertices, as in the road network, against their concern for graph distance – the number of legs or hops between vertices – as in an airline network or the Internet. We find that this model reproduces well the qualitative features of our real-world networks, generating two-dimensional road-like networks in one limit and high-dimensional hub-and-spoke networks in another.

The results presented here are inevitably only the beginnings of a detailed study of spatial networks. Many other features of these networks deserve scrutiny, such as, for instance, the effects of population distribution. We hope that others will also investigate this interesting class of systems and look forward eagerly to their results.

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
3. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
4. P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, S.S. Manna, Phys. Rev. E **67**, 036106 (2003)
5. B.M. Waxman, IEEE J. Selected Areas Comm. **6**, 1617 (1988)
6. S.H. Yook, H. Jeong, A.-L. Barabási, Proc. Natl. Acad. Sci. USA **99**, 13382 (2001)
7. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
8. J. Brimberg, P. Hansen, K.-W. Lih, N. Mladenovic, M. Breton, Oper. Res. **51**, 0228 (2003)
9. R. Guimerà, S. Mossa, A. Turtschi, L.A.N. Amaral, Proc. Natl. Acad. Sci. USA **102**, 7794 (2005)

10. M.S. Mizruchi, *The American Corporate Network, 1904–1974* (Sage, Beverley Hills, 1982)
11. L.E. Miller, J. Res Natl. Inst. Stand. Technol. **106**, 401 (2001)
12. W.L. Garrison, Papers and Proceedings of the Regional Science Association **6**, 121 (1960)
13. P. Haggett, R.J. Chorley, *Network Analysis in Geography* (St. Martin's Press, New York, NY, 1969)
14. K.J. Kansky, *Structure of Transportation Networks: Relationships Between Network Geometry and Regional Characteristics* (University of Chicago, Chicago, 1963)
15. S.P. Gorman, R. Kulkarni, Environment and Planning B **31**, 273 (2003)
16. A. Barrat, M. Barthélemy, A. Vespignani, J. Statist. Mech., P05003 (2005)
17. M. Faloutsos, P. Faloutsos, C. Faloutsos, Comp. Commun. Rev. **29**, 251 (1999)
18. L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000)
19. M. Molloy, B. Reed, Random Structures and Algorithms **6**, 161 (1995)
20. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
21. J.M. Kleinberg, Nature **406**, 845 (2000)
22. J. Dall, M. Christensen, Phys. Rev. E **66**, 016121 (2002)
23. S.S. Manna, P. Sen, Phys. Rev. E **66**, 066114 (2002)
24. M. Barthélémy, Europhys. Lett. **63**, 915 (2003)
25. P. Sen, S.S. Manna, Phys. Rev. E **68**, 026104 (2003)
26. T. Petermann, P. de los Rios, preprint `cond-mat/0501420` (2005)
27. M.E.J. Newman, D.J. Watts, Phys. Rev. E **60**, 7332 (1999)
28. D.B. West, *Introduction to Graph Theory* (Prentice Hall, Upper Saddle River, NJ, 1996)
29. J.E. Hopcroft, R.E. Tarjan, J. ACM **21**, 549 (1974)
30. G. Csányi, B. Szendrői, Phys. Rev. E **70**, 016122 (2004)
31. A. Fabrikant, E. Koutsoupias, C.H. Papadimitriou, in *ICALP* (Springer, 2002), Lect. Notes Comput. Sci., Vol. 2380, pp. 110–112
32. A.J. Scott, Transportation Research **3**, 201 (1969)
33. J. Berg, M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002)
34. R. Xulvi-Brunet, I.M. Sokolov, Phys. Rev. E **66**, 026118 (2002)