

## Towards Design Principles for Optimal Transport Networks

G. Li,<sup>1</sup> S. D. S. Reis,<sup>2</sup> A. A. Moreira,<sup>2</sup> S. Havlin,<sup>3</sup> H. E. Stanley,<sup>1</sup> and J. S. Andrade, Jr.<sup>2</sup>

<sup>1</sup>Center for Polymer Studies, Department of Physics, Boston University,  
590 Commonwealth Avenue, Boston, Massachusetts 02215, USA

<sup>2</sup>Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Ceará, Brazil

<sup>3</sup>Department of Physics, Minerva Center, Bar Ilan University, Ramat Gan 52900, Israel

(Received 9 July 2009; published 6 January 2010)

We investigate the navigation problem in lattices with long-range connections and subject to a cost constraint. Our network is built from a regular two-dimensional ( $d = 2$ ) square lattice to be improved by adding long-range connections (shortcuts) with probability  $P_{ij} \sim r_{ij}^{-\alpha}$ , where  $r_{ij}$  is the Manhattan distance between sites  $i$  and  $j$ , and  $\alpha$  is a variable exponent. We introduce a cost constraint on the total length of the additional links and find optimal transport in the system for  $\alpha = d + 1$  established here for  $d = 1$  and  $d = 2$ . Remarkably, this condition remains optimal, regardless of the strategy used for navigation, being based on local or global knowledge of the network structure, in sharp contrast with the results obtained for unconstrained navigation using global or local information, where the optimal conditions are  $\alpha = 0$  and  $\alpha = d$ , respectively. The validity of our results is supported by data on the U.S. airport network.

DOI: 10.1103/PhysRevLett.104.018701

PACS numbers: 89.75.Hc, 02.50.-r, 05.40.Fb, 89.75.Fb

The interplay between topology and dynamics in complex systems represents the focus of many studies in different fields of research with important scientific and technological applications. Because of their enormous potential to represent the intricate topology of numerous systems in nature, complex networks [1,2] have recently been used as substrates in combination with a plethora of dynamical models to describe the behavior of biological, social, chemical, physical, and technological networks [3]. Much attention has been dedicated to the problem of navigation in complex network geometries [4–14]. In most cases, the influence of the underlying network geography on the performance of the transport process is investigated assuming that only local information is available for navigation [4,5,7,8,15].

For many navigation problems of interest in science and technology, global rather than local information is required, i.e., any source node  $s$  possesses the knowledge of the entire network topology. In this situation, the average shortest path  $\langle \ell \rangle$  from source to target becomes the relevant navigation variable to be optimized. For example, in a subway network, such as in Manhattan, the travel routes should be planned or changed in such a way as to minimize the travel time for a given limited reconstruction cost. This task is performed by considering the whole structure of the network in terms of its nodes and links, namely, by knowing the location of all subway stations, their connections, and the shortest path between any two stations. If we now consider an underlying network of streets and avenues over which one has to plan or improve an existing subway network, and if the aim is to minimize the average travel time between its stations, the search for an optimal strategy to add new connections in the network for a given budget should therefore play a key role. Here we show that the imposition of a cost constraint, which to

the best of our knowledge has not been considered for optimal navigation, represents a crucial ingredient in the design and development of efficient navigation networks.

Consider the case of an existing transport network which needs improvements [16]. The financial cost to build up a large number of new direct connections between distant stops (i.e., non-neighboring sites) can make it prohibitive, since only limited resources are normally available for this task. These types of problems can be modeled in the following way, which is similar to the formulation presented in Ref. [4]. In a two-dimensional regular square lattice, with all  $N = L^2$  sites present, each site  $i$  is connected with its four nearest neighbors. The sites represent the stops and the bonds represent the routes of the transport system (see Fig. 1). In our model, pairs of sites  $ij$  are then randomly chosen to receive long-range connections with

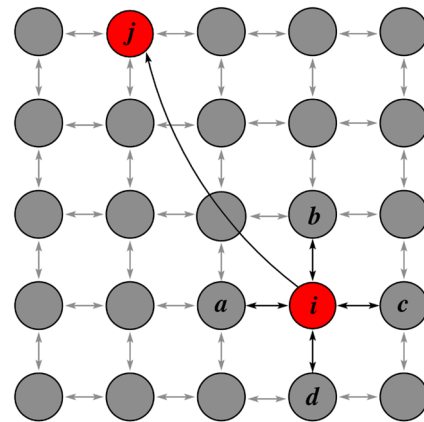


FIG. 1 (color online). Connections of a single node  $i$ . Each node  $i$  has four short-range connections to its nearest neighbors ( $a$ ,  $b$ ,  $c$ , and  $d$ ). A long-range directed connection may be placed to a random chosen node  $j$ .

probability proportional to  $r_{ij}^{-\alpha}$ , where  $r_{ij}$  is the Manhattan distance between sites  $i$  and  $j$ , namely, the distance measured as the number of connections separating the nodes in the underlying regular lattice. Finally, and distinct from Ref. [4], the addition of long-range connections to the system stops when their total length (cost),  $\sum r_{ij}$ , reaches a given value  $\Lambda$ . Since  $\alpha$  controls the average length of the long-range connections, we obtain that, for a fixed value of  $\Lambda$ , and small values of  $\alpha$ , longer connections, but fewer in number can be added due to the imposed total length limit. We therefore expect that an optimal navigation condition must be revealed as a tradeoff between the length and the number of connections added to the system.

Optimal navigation with local knowledge and the presence of long-range links in a lattice network without constraints was recently studied by Kleinberg [4]. Here we show for the first time that a rather different behavior can be observed when realistic constraints on total length are imposed on the process of adding long-range connections, regardless if navigation is based on local or global knowledge of the network structure. To better demonstrate the competition between total length and number of links, we generate a single network realization ( $L = 256$ ) for a given value of  $\alpha$  and compute the shortest-path length  $\ell$  from each node in the network to its central node. This calculation is performed as follows: once we choose the root node (e.g., the central one), we visit all its neighbors, including the neighboring nodes connected by long-range connections. These visited nodes are classified as shell one nodes, meaning that they are only one time step away from the root node. After that, we visit all the neighbors of these nodes not visited before and classify them as shell two nodes. Following this procedure for all network nodes, we obtain the  $\ell$  values (time) for each node to be reached from the root node. Figure 2 shows the contour plots representation of the  $\ell$  values performed for four different values of the parameter  $\alpha$ . For  $\alpha = 1$  and 2 the number of long-range connections is small. As a result, only a few little islands sparsely dispersed in these networks are really close to their central nodes (only a few short and/or long-range connections away). For  $\alpha = 3$  the added long-range links are shorter, but more numerous, thus substantially decreasing the shortest path over the whole network. For  $\alpha = 4$ , due to the very short length size of the added connections, only a limited region surrounding the central node displays a reduced shortest path. Sites which are further away from the origin have significantly larger path length  $\ell$  to the origin.

We extract more quantitative information about this navigation problem by performing extensive simulations for different values of  $\alpha$  and different system sizes. In each case, the average shortest path  $\langle \ell \rangle$  is calculated over all realizations, considering all the shortest distances between each pair of nodes. We assume that the total length (cost) is proportional to the total length of the links in the underlying network, i.e.,  $\Lambda = AL^2$ , where  $A$  is a constant. That

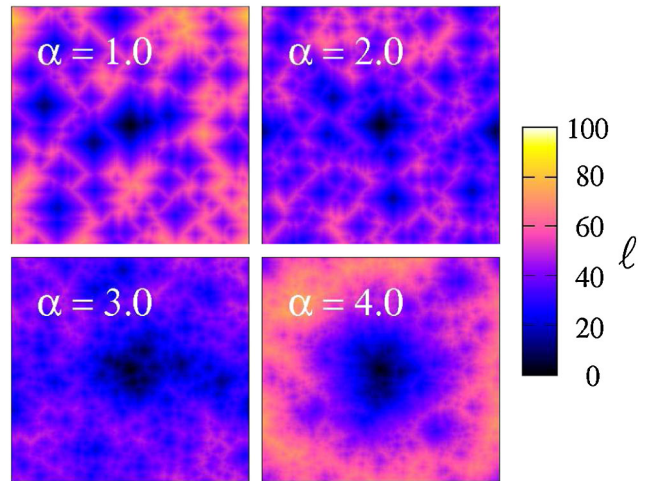


FIG. 2 (color online). Shortest-path length  $\ell$  from each node to the central node in the network for different values of  $\alpha$ . In this case we impose a constraint in the length of the long-range connections. The sum of the length of these connections is limited,  $\Lambda = \sum r_{ij} = N$ , where  $N$  is the number of nodes in the underlying lattice. The network model is constructed from a square lattice with  $L^2$  nodes, with  $L = 256$ . We can clearly observe that the best condition for shortest-path length is obtained for  $\alpha = 3$ .

is, the budget to improve the system is a fraction of the cost of the current network (without long-range connections) [17].

The results presented in Fig. 3 clearly indicate the presence of a minimum  $\langle \ell \rangle$  for different system sizes at the same value of the exponent  $\alpha = 3$ , where optimal

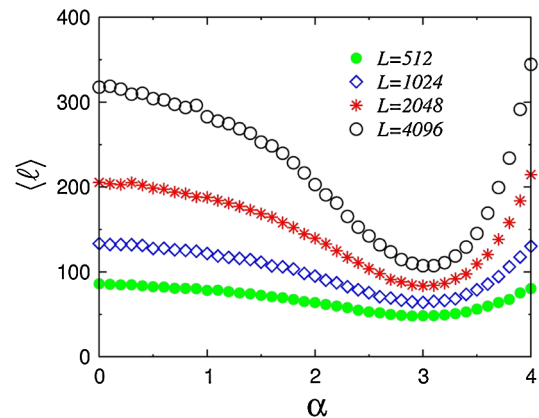


FIG. 3 (color online). Average shortest-path length  $\langle \ell \rangle$  as a function of  $\alpha$ . There is a constraint in the total length of the long-range connections,  $\Lambda = \sum r_{ij} = L^2$ , where  $L$  is the size of the underlying square lattice. We find that the optimal shortest path is achieved for  $\alpha = 3$ . With the restriction on total length, the number of long-range connections is not fixed (e.g., with  $\alpha = 0$ , large long-range connections become frequent, which reduces the total number of long-range connections.) To obtain these results, we simulated 10000 realizations for  $L = 512$ , 3500 realizations for  $L = 1024$  and 2048, and 25 realizations for  $L = 4096$ .

navigation is achieved with (global) knowledge of the shortest paths and cost limitations. The way in which  $\langle \ell \rangle$  scales with system size  $L$ , however, seems to follow rather different behaviors, depending on the value of  $\alpha$ . We tested two possible forms for  $\langle \ell \rangle$  vs  $N$ , a power law and a power of  $\log_{10} N$ . As shown in the main plot of Fig. 4(a), our results for  $\alpha \neq 3$  suggest that the shortest path  $\langle \ell \rangle$  follows a power law with system size  $L$ . This is supported by the plot of successive slopes  $\delta_S$  obtained from  $\log_{10} \langle \ell \rangle$  versus  $\log_{10} L$ , which are almost invariant, as shown in the inset of Fig. 4(a). In contrast, in the case of  $\alpha = 3$ , the increase with  $L$  of  $\langle \ell \rangle$  appears to be less rapid than a power law. Interestingly, the successive slopes  $\gamma_S$  obtained from  $\log_{10} \langle \ell \rangle$  versus  $\log_{10}(\log_{10} L)$ , as presented in Fig. 4(b), indicate that  $\langle \ell \rangle$  increases as a power of the logarithm of  $L$ ,  $\langle \ell \rangle \sim \log_{10}^{\gamma_S} L$ , rather than a power of  $L$ , only for  $\alpha = 3$ . This provides clear support for the fact that in the optimal condition,  $\alpha = 3$ , the transport will improve even further as  $L$  increases, as suggested by Fig. 3.

We also studied our model for a one-dimensional lattice and observed similar behavior. The optimal condition we obtained in this case at  $\alpha = 2$  (data not shown) leads us to conjecture that the optimal value is obtained at  $\alpha = d + 1$ , where  $d$  is the dimension of the underlying lattice. Note that the Kleinberg result was extended to fractals [5], where the optimal exponent is found to be  $\alpha = d_f$ , namely, the fractal dimension of the substrate. The  $\langle \ell \rangle$  dependence on  $L$  for different  $\alpha$  for the Kleinberg model was recently derived analytically [7].

In the following, we present analytical arguments showing that  $\alpha = 3$  is indeed the only case where logarithmic scaling of  $\langle \ell \rangle$  with  $L$  can occur, while for  $\alpha \neq 3$  a power-

law with  $L$  should exist. By arbitrarily fixing the cost parameter to  $\Lambda = AL^2$ , we obtain that  $\rho \sim \langle r \rangle^{-1}$  [18], where  $\rho$  is the expected density and  $\langle r \rangle$  is the average length of the added long-range connections. Since  $\langle r \rangle \sim \int_2^L r^{2-\alpha} dr$  [19], it follows that for  $2 \leq \alpha < 3$ ,  $\rho \sim L^{\alpha-3}$  and for  $\alpha < 2$ ,  $\langle r \rangle$  is limited by the network size leading to  $\rho \sim L^{-1}$ . Thus, for all values of  $\alpha < 3$  the density of the long-range links added, due to the constraint, decreases as a power law with  $L$ . As a consequence of this power-law decrease in density,  $\langle \ell \rangle$  must increase as a power of  $L$ . To see this we argue that  $\langle \ell \rangle$  is bounded by the relation  $\langle \ell \rangle > \rho^{-1/d}$ . The right-hand side  $\rho^{-1/d}$  appears for the case of the small world model, where  $\alpha = 0$ , with a fixed concentration of links  $\langle \ell \rangle \sim \rho^{-1/d} \ln L$  [2]. Since for the case  $0 < \alpha < 3$ ,  $\langle \ell \rangle$  decreases with increasing  $\alpha$ , the bound  $\langle \ell \rangle > L^{(3-\alpha)/d}$  is rigorous and  $\langle \ell \rangle$  in this range must scale as a power of  $L$ . For  $\alpha > 3$  and sufficiently large networks,  $\langle r \rangle$  is finite and the density becomes independent of the system size, i.e.,  $\rho \sim L^0$ . Thus, the effect of the constraint  $\Lambda$  on navigation should become negligible. However, the finite value of  $\langle r \rangle$  suggests that long-range links can be neglected and therefore  $\langle \ell \rangle$  should scale as a power of  $L$ . Thus, it follows that only for  $\alpha = 3$ ,  $\langle \ell \rangle$  can scale logarithmically with  $L$ , as suggested by our numerical simulations (see Fig. 4).

It is important to note that our global navigation scheme with  $\langle \ell \rangle$  can be considered as a lower bound to any other transport navigation process. For example, a strategy based on purely local knowledge of the network structure will necessarily perform worse than any other with global information. In Ref. [4], for example, the greedy algorithm is introduced as a paradigm based on local information, where the traveler, when leaving a node, chooses to move to the one among its neighbors which has the smallest Manhattan distance to the target. Kleinberg found that  $\alpha = 2$  is the optimal value in the navigation with the

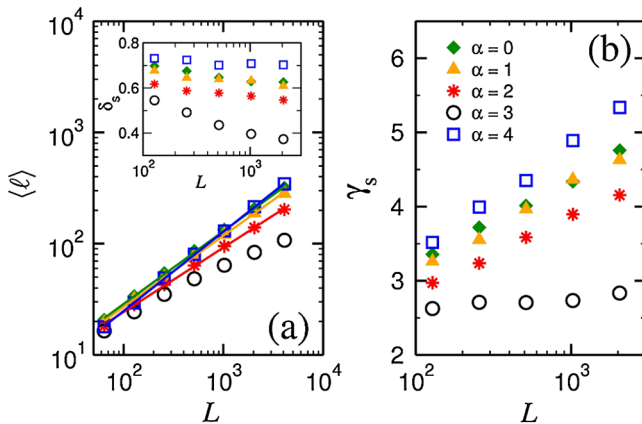


FIG. 4 (color online). In (a) we show the average shortest-path length  $\langle \ell \rangle$  as a function of the lattice size  $L$  for the square lattice. The constraint in the total length of the long-range connections is  $\Lambda = L^2$ . The curve with  $\alpha = 3$  increases slower with  $L$  compared to any other value of  $\alpha$ . In the inset, the plot of the successive slopes  $\delta_S$  obtained from  $\log_{10} \langle \ell \rangle$  versus  $\log_{10} L$  reinforces the display of power-law behavior of  $\langle \ell \rangle$  with  $L$  for  $\alpha \neq 3$ . The plot of the successive slopes  $\gamma_S$  obtained from  $\log_{10} \langle \ell \rangle$  versus  $\log_{10}(\log_{10} L)$  shown in (b) indicates that  $\langle \ell \rangle$  increases as a power of the logarithm of  $L$  for the optimal condition  $\alpha = 3$ .

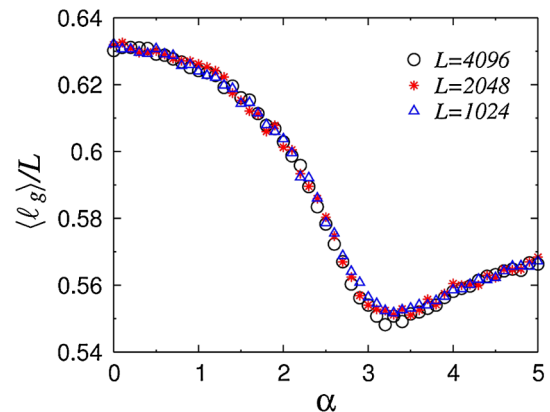


FIG. 5 (color online). The characteristic average delivery time  $\langle \ell_g \rangle / L$  as a function of  $\alpha$  for navigation with the greedy algorithm. The cost  $\Lambda$  involved to add long-range connections changes the behavior of the density of long-range connections. As a result of that, a minimum is observed at  $\alpha \approx 3$ . Each data point is a result of 4000 simulations and the cost  $\Lambda$  is fixed at  $L^2$ .

greedy algorithm [4]. We next ask, what would be the optimal  $\alpha$  for the greedy algorithm when cost restriction  $\Lambda = AL^2$  is imposed? We find also for the greedy algorithm that the optimal value is  $\alpha = 3$ . This is shown in Fig. 5, where we plot the average delivery length  $\langle \ell_g \rangle$  that a message travels with only local information of the system geometry. The message is sent from the source node  $s$  to the target node  $t$  through a network generated with the constraint  $\Lambda = L^2$ . Remarkably, the presence of a minimum also at  $\alpha \approx 3$  shows that the type of information (local or global) used by the message holder to pass it through the system during the navigation process becomes unimportant if the network is constructed under length (cost) limitations. However, the two mechanisms display very different and distinct behaviors regarding the scaling with system size. While we observe logarithmic growth for the optimal condition  $\alpha = 3$ , in the case of global information, the time to reach the source, with the greedy algorithm and with cost constraint, appears to increase linearly with size for all values of  $\alpha$ . The linearity of  $\langle \ell_g \rangle$  with  $L$  is observed in the scaling collapse (Fig. 5) of the curves of  $\langle \ell_g \rangle / L$  vs  $\alpha$ .

In summary, our results suggest that, regardless of the strategy used by the traveler, based on local or global knowledge of the network structure, the best transportation condition is obtained with an exponent  $\alpha = d + 1$ , where  $d$  is the topological dimension of the underlying lattice. Our results hold for  $d = 1$  and 2, but simulations in three dimensions are still necessary to further support this conjecture. The results recently reported by Bianconi *et al.* [20] on the U.S. airport network yield an exponent  $\alpha = 3$ , which is similar to our optimal exponent for  $d = 2$ . The fact that the probability of a flight connection within U.S. decays as a power law with the distance between airports,  $r^{-\alpha}$ , where  $\alpha = 3.0 \pm 0.2$ , seems to reveal the optimized aspect of the network under the conditions of geographical availability (for customer satisfaction) and cost limitations (for airline company profit). The result  $\alpha = 3$  is in sharp contrast with the results obtained for unconstrained systems with global and local information, where the optimal conditions are  $\alpha = 0$  [9] and  $\alpha = d$  [4,5], respectively. The contrast between the optimal results is even more dramatic. While in the unconstrained case the mean length of a link diverges, we find that when cost is considered the mean length is finite. In the case where the traveler has global knowledge of the network, and is able to identify the shortest path for navigation, we obtain a slow (logarithmic) growth with size for the transit time at the optimal condition. A different picture is obtained if the traveler has only local knowledge of the network. For example, in the case where the transportation path is decided based on the Manhattan distance to the target, we obtain a linear growth of the transit time with system size, for all values of the exponent  $\alpha$ . Finally, our results suggest that the idea of introducing a cost constraint in the navigation problem offers a different theoretical framework to understand the

evolving topologies of other important complex network structures in nature, such as subways, trains, or the Internet. Of course, at this point we must emphasize that our approach represents only one specific model within a larger family of models where design principles can be tested to improve the performance of the transport system. In the case of airport networks, for example, other variables than the particular cost function that we adopted can be used for realistic optimization purposes.

We thank CNPq, CAPES, FUNCAP, FINEP, ONR, DTRA, the Israel Science Foundation, and the European Project EPIWORK for financial support, and D. ben-Avraham and R. Cohen for useful discussions.

- 
- [1] R. Albert and A.-L. Barabási, *Nature (London)* **401**, 130 (1999); D. Watts and S. Strogatz, *Nature (London)* **393**, 440 (1998).
  - [2] M. Barthélemy and L. A. N. Amaral, *Phys. Rev. Lett.* **82**, 3180 (1999); **82**, 5180 (1999).
  - [3] H. Jeong *et al.*, *Nature (London)* **407**, 651 (2000); S. Lawrence and C. L. Giles, *Nature (London)* **400**, 107 (1999); C. L. Giles *et al.*, *Science* **280**, 1815 (1998).
  - [4] J. M. Kleinberg, *Nature (London)* **406**, 845 (2000); in *Proceedings of the 32nd ACM Symposium on Theory of Computing* (ACM, New York, 2000), pp. 163–170.
  - [5] M. R. Roberson and D. ben-Avraham, *Phys. Rev. E* **74**, 017101 (2006).
  - [6] M. T. Gastner and M. E. J. Newman, *Phys. Rev. E* **74**, 016117 (2006).
  - [7] S. Carmi *et al.*, *Phys. Rev. Lett.* **102**, 238702 (2009).
  - [8] C. Caretta Cartozo and P. De Los Rios, *Phys. Rev. Lett.* **102**, 238703 (2009).
  - [9] K. Komidis *et al.*, *Europhys. Lett.* **82**, 48005 (2008).
  - [10] C. F. Moukarzel and M. Argollo de Menezes, *Phys. Rev. E* **65**, 056709 (2002).
  - [11] R. Guimera *et al.*, *Phys. Rev. Lett.* **89**, 248701 (2002).
  - [12] L. Danon, A. Arenas, and A. Diaz-Guilera, *Phys. Rev. E* **77**, 036103 (2008).
  - [13] M. C. Santos *et al.*, *Phys. Rev. E* **77**, 041101 (2008).
  - [14] M. Boguna and D. Krioukov, *Phys. Rev. Lett.* **102**, 058701 (2009).
  - [15] V. Latora and M. Marchiori, *Physica (Amsterdam)* **314A**, 109 (2002).
  - [16] H. Youn, M. T. Gastner, and H. Jeong, *Phys. Rev. Lett.* **101**, 128701 (2008).
  - [17] Here, we show the case  $A = 1$  but we obtained similar results for several values in the range  $0 < A < 1$ .
  - [18] The density of long-range connections is defined as  $\rho = N_l / L^2$ , where  $N_l$  is the total number of long-range connections present in the system. The cost limitation imposes that  $N_l = \Lambda / \langle r \rangle$ , and since  $\Lambda = AL^2$ , we obtain that  $\rho = A \langle r \rangle^{-1}$ .
  - [19] The average length of long-range connections is given by  $\langle r \rangle = \sum_{r=2}^L r N_r r^{-\alpha} / \sum_{r=2}^L N_r r^{-\alpha}$ , where  $N_r = 4r$  is the number of nodes that are at a Manhattan distance  $r$  from the source in a square lattice. Approximating the sum by an integral, we obtain  $\langle r \rangle \sim \int_2^L r^{2-\alpha} dr$ .
  - [20] G. Bianconi, P. Pin, and M. Marsilli, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11 433 (2009).