CHAOS 18, 023103 (2008)

Disorder and decision cost in spatial networks

Massimiliano Zanin,[1] Javier M. Buldú,[2] P. Cano,[3] and S. Boccaletti[4,5]
[1]Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[2]Departamento de Física, Universidad Rey Juan Carlos, Tulipán s/n, 28933 Móstoles, Madrid, Spain
[3]Music Technology Group, Universitat Pompeu Fabra, 08003 Barcelona, Spain
[4]CNR-Istituto dei Sistemi Complessi, Via Madonna del Piano, 10, 50019 Sesto Fiorentino (Florence), Italy
[5]The Italian Embassy in Tel Aviv, Trade Tower, 25 Hamered Street, Tel Aviv, Israel

(Received 17 December 2007; accepted 5 March 2008; published online 16 April 2008)

We introduce the concept of decision cost of a spatial graph, which measures the disorder of a given network taking into account not only the connections between nodes but their position in a two-dimensional map. The influence of the network size is evaluated and we show that normalization of the decision cost allows us to compare the degree of disorder of networks of different sizes. Under this framework, we measure the disorder of the connections between airports of two different countries and obtain some conclusions about which of them is more disordered. The introduced concepts (decision cost and disorder of spatial networks) can easily be extended to Euclidean networks of higher dimensions, and also to networks whose nodes have a certain fitness property (i.e., one-dimensional). © *2008 American Institute of Physics*. [DOI: 10.1063/1.2901916]

When analyzing structures tailored by men, complex networks theory has been a useful instrument in order to study their complexity. In many works related to this field, the projected network is obtained by detecting connections between their fundamental units (i.e., nodes), disregarding any information about their spatial distribution. In this paper, we study the importance of the Euclidean position of nodes for a key property of the network: the *decision cost*, that is, the difficulty for a blind agent to make its way from a starting node to a target node. We demonstrate how this parameter is a good indicator of the disorder of a spatial net, and we apply this measure to two different airport networks.

INTRODUCTION

The experiment of Milgram[1] exploring communication through a social network has been one of the most influential works concerning complex networks, specifically social complex networks. Milgram tried to reveal the complex properties of the acquaintance network between individuals, and asked several people in Nebraska to reach a stranger target in Massachusetts. The rule was that each individual should send a letter to an acquaintance whom he/she would estimate to be closer to the target person. Surprisingly, some of the letters arrived at the selected person after a short number of steps. This experiment revealed the small-world nature of certain networks, in this case a network of acquaintances, which can be traveled from one corner to the other through a short number of connections. It was more than 30 years later that Watts and Strogatz[2] (1998) defined the particular properties of small-world networks, reactivating the study of complex networks, a field that has profusely grown during recent years.[3]

After the seminal work of Watts and Strogatz, studies focused on the structural properties of these kinds of networks, which lie between regular and random structures, and gradually shifted to analyze the influence of the topology on the dynamics and the evolution of the networks. The degree distribution, the clustering coefficient, or the assortativity have become usual parameters to evaluate the structural properties of a network, which, in addition, reveal some information about navigation, structural resilience, or modularity within the graph.[4] Specifically, navigation through a complex network is an interesting topic due to its importance on the design of freight transport networks, Internet-based recommender systems, or even the streets distribution of a city. As an example, it has been shown that two-dimensional small-world networks reduce their delivery time between any two nodes when a critical value of the clustering is fulfilled.[5]

In the current work, we are concerned with the structure and navigation through spatial networks, i.e., networks that have a representation in Euclidean space. Several works have dealt with the structural properties of different spatial networks, such as railway or subway networks,[6,7] the network of streets within a city,[8,10] or the network of connections between airports.[11] In most of these works, however, the projected network consisted of a set of nodes (stations, airports, or streets) connected by links, where the spatial position of the nodes was disregarded. In this way, two airports have a distance of one if a direct flight exists between them, no matter if they are 100 or 1000 kilometers apart. In the following, we will introduce a fundamental distinction when considering a spatial network: we will take into account not only connections between nodes but the position of them in Euclidean space. In this way, we are able to define a structural parameter that will help us to decide about the way of navigating through the network: the *decision cost* (DC). As we will explain in detail, the decision cost also quantifies the disorder of the network taking into account not only the heterogeneity of the degree distribution but the amount of randomness of the spatial distribution. Therefore, it is a param-

1054-1500/2008/18(2)/023103/6/$23.00

eter to be considered in the design of spatial networks, since it gives hints about navigability strategies within the graph.

DECISION COST IN SPATIAL NETWORKS

Approaches to calculate the disorder of a complex network have focused on the analysis of degree heterogeneity.[12–14] This is a good indicator of the disorder of the network, which has also been used to evaluate the resilience of a network to random failures.[13] Nevertheless, the position of nodes has often been disregarded, in general due to the fact that a node cannot be associated with a position in the Euclidean space (e.g., in social networks). However, in spatial networks the position of nodes can be crucial when evaluating the disorder of the network. If we take a regular two-dimensional lattice with all nodes connected with its closest neighbors and we move nodes randomly while keeping their connections, it is clear that the disorder of the network will increase. Therefore, it is necessary to include the position of the nodes when evaluating the disorder of spatial networks.

Rosvall[14] evaluated the navigation properties of a network in terms of calculating the information required to move from a node to another through the shortest path; for each step, the agent must choose the right exit link, then the information is defined by the Shannon entropy like $\log_2 k_n$, where k_n is the degree of the nth node. Although this measure gives some hints about how complex is moving through the net, there is no way to extract how difficult is to find the shortest path between two points, once the position of the target point is known. Furthermore, the distances between nodes are not considered in the model. Another recently proposed measure[9] consists of counting the number of bits of information needed to transmit a message to any given node of the network, or conversely, to evaluate the departing node when a message is received.

Suppose that we have an agent that knows where it is at a certain time, where it should go (the destination node), and, furthermore, it has information about the position of the nodes connected with itself. Such an agent, however, has no information about the overall structure of the network, and therefore it must decide a navigation strategy according to a limited (local) information. The strategy through which the agent moves is the following: at every node of the path, it will move to the node closest to the destination. It is clear that this strategy can lead to the incorrect solution when the network is not regular, and generally will not result in the shortest path.

We can now define the shortest-path decision cost ($\mathrm{DC}_{\mathrm{sp}}$) as *the information needed to define the shortest path between two nodes, modifying the trivial solution given by our agent*:

$$\mathrm{DC}_{\mathrm{sp}} = -\frac{\sum_{n-1} \log_2 (1/\#c)}{n-1}, \tag{1}$$

where n is the number of nodes of the path, and $\#c$ is the logical order of the outgoing link (1 the closest to the objective, 2 the second closest, etc.).

The classical way to evaluate the quantity of information requested to successfully reach a target node consists in simply counting the number of questions that an agent should ask to find the correct path (i.e., for each link, the agent asks if that link is the correct one in the shortest path to reach the target). Although it is a more straightforward way, it lacks an important feature: the normalization of the decision cost, and therefore, the ability of comparing networks of different sizes.

For the whole network, the $\overline{\mathrm{DC}_{\mathrm{sp}}}$ will be

$$\overline{\mathrm{DC}_{\mathrm{sp}}} = \frac{\sum_{i=1}^{N^*} \mathrm{DC}_{\mathrm{sp}}(i)}{N^*}, \tag{2}$$

where N^* is the set of shortest path, with three or more nodes. Since we must calculate the shortest path between each pair of nodes, the computational cost of the $\mathrm{DC}_{\mathrm{sp}}$ is, in the worst case, of $O(N^2)$, N being the total number of nodes.

Let us look at an example of the decision cost of a network. In Fig. 1, the second and third nets were created from the first one by moving nodes E and C; moreover, nodes B and D have been brought closer to keep the *characteristic path length* of the network. If we consider that the *degree distribution* and the *clustering coefficient* are the same, the three nets have the same statistical properties. From the point of view of path decision cost, on the contrary, our three nets are very different.

On the first net, any shortest path we choose can be covered by going to the neighbor node that is closer to the objective. As a result, the decision cost $\mathrm{DC}_{\mathrm{sp}}=0$. In the second one, when going from A to D, the most logical step would be move to B, because it is the node closer to D. Nevertheless, this option will not lead to the shortest path. The same occurs from C to B, leading to a decision cost $\mathrm{DC}_{\mathrm{sp}}$ greater than zero. In the third network, we have four problematic paths: $A \to D$, $C \to B$, $B \to C$, $D \to A$. At each of these paths, we must avoid the most logical link and follow the second option, so from Eq. (1), $\mathrm{DC}_{\mathrm{sp}}$ is always $1/2$.

Despite the fact that the definition of decision cost is focused in spatial networks, it also applies to any network with a certain fitness value associated with its nodes. In this case, the fitness parameter is the parameter evaluated by the agent to move through the network, i.e., on its way to reaching a target node with a certain fitness, the agent moves to the connected nodes which have the closest fitness to the target node.

FROM ORDER TO DISORDER

In the following, we evaluate the disorder of spatial networks of different sizes by computing their decision costs. We depart from a regular two-dimensional (2D) network of $N \times N$ components coupled through a nearest-neighbor mechanism. In this way, each node has four links, specifically one with the upper node, one with the lower node, and one with each lateral node (with the exception of the boundary nodes). This particular network has a decision cost of $\mathrm{DC}_{\mathrm{sp}}=0$ since all nodes are placed regularly and the movement toward the target node will always lead to the shortest path. From this starting point, we put the network out of

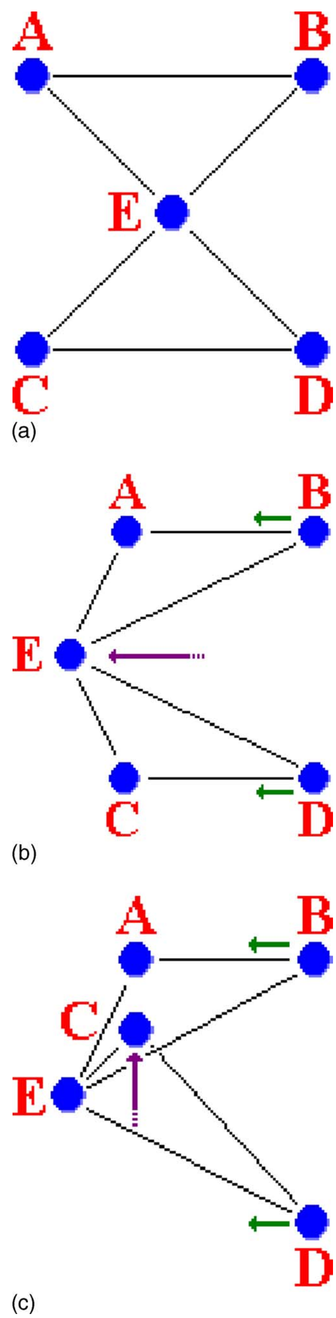


FIG. 1. (Color online) An example of path-searching complexity: a simple net (a) and two evolutions [(b) and (c)]. Depending on the spatial distribution of nodes, the ease of finding the shortest path may change considerably.

order by introducing a rewiring probability p at each link. Figure 2(a) shows the evolution of the decision cost as a function of the rewiring probability p for different sizes of an $N \times N$ network (where $N=3,\ldots,12$). We can see how the decision cost (disorder) of the network increases as the random links are introduced, as could be expected, with the highest cost (disorder) corresponding to $p=1$. On the other hand, decision cost increases with the size of the network, i.e., the bigger the network, the more information we need to travel inside it.

When the decision cost is normalized by the value of the randomly connected network DC_{ran}, i.e., $p=1$, we obtain Fig. 2(b). We observe how the normalized decision cost DC_{norm}

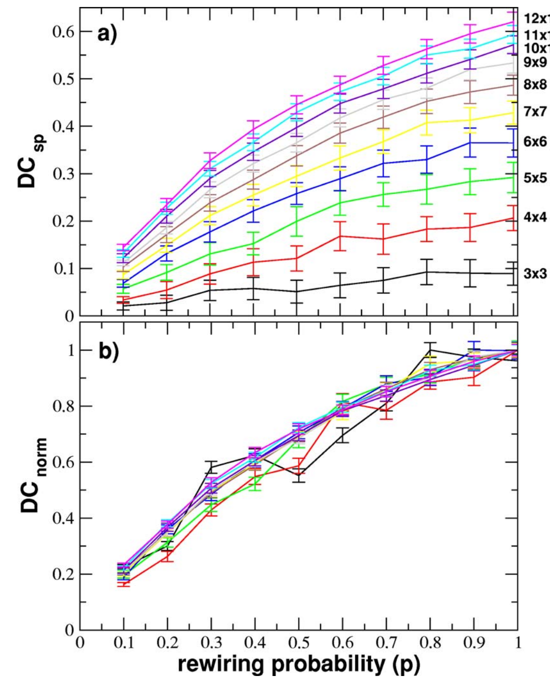


FIG. 2. (Color online) Decision cost DC_{sp} (a) and normalized decision cost DC_{norm} (b) of $N \times N$ spatial networks as a function of the rewiring probability p. The number of nodes ranges from 9 (3×3) to 144 (12×12). Each point corresponds to the mean value of 20 simulations, with its corresponding error bars.

shows the same trend in all cases, no matter the size of the network. This fact indicates that the normalized decision cost is a suitable parameter in order to compare the disorder of spatial networks of different sizes.

At this point, we would like to mention some aspects about the decision cost of a network:

(1) It measures the degree of disorder of a spatial network, specifically the disorder of its connections.
(2) It is a good indicator to decide a strategy to navigate in the network (when the graph has low values of the decision cost, the most recommended option is to move in the direction of the target node, while other strategies are more suitable for navigating graphs with high values of the decision cost).
(3) One can define a normalized decision cost DC_{norm}, which could be used to compare the degree of disorder of different networks.
(4) The decision cost is of special interest for small to medium networks, where indicators such as the mean shortest path or the clustering coefficient are very sensitive to the initial conditions.
(5) Note that a random network does not have the highest decision cost. Despite the fact that we use the random network to normalize, we can find networks with DC_{norm} higher than the unity.

Concerning the latter point, it is interesting to identify the parameters that modify the decision cost of a spatial network. As we have shown in Fig. 2, the rewiring probability and the number of nodes increase the decision cost of a regular spatial network (i.e., a network with nodes equally dis-

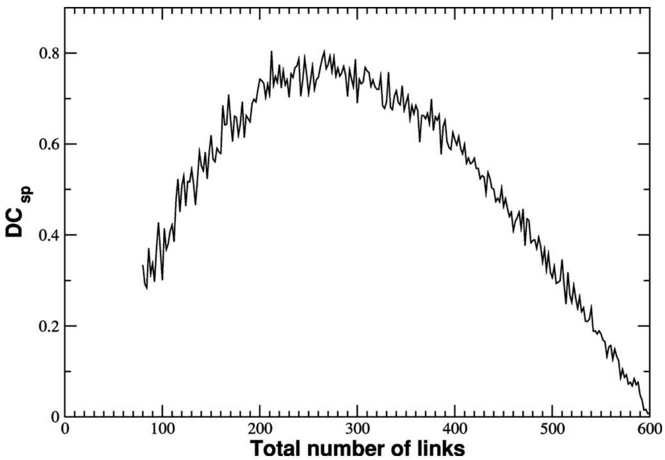
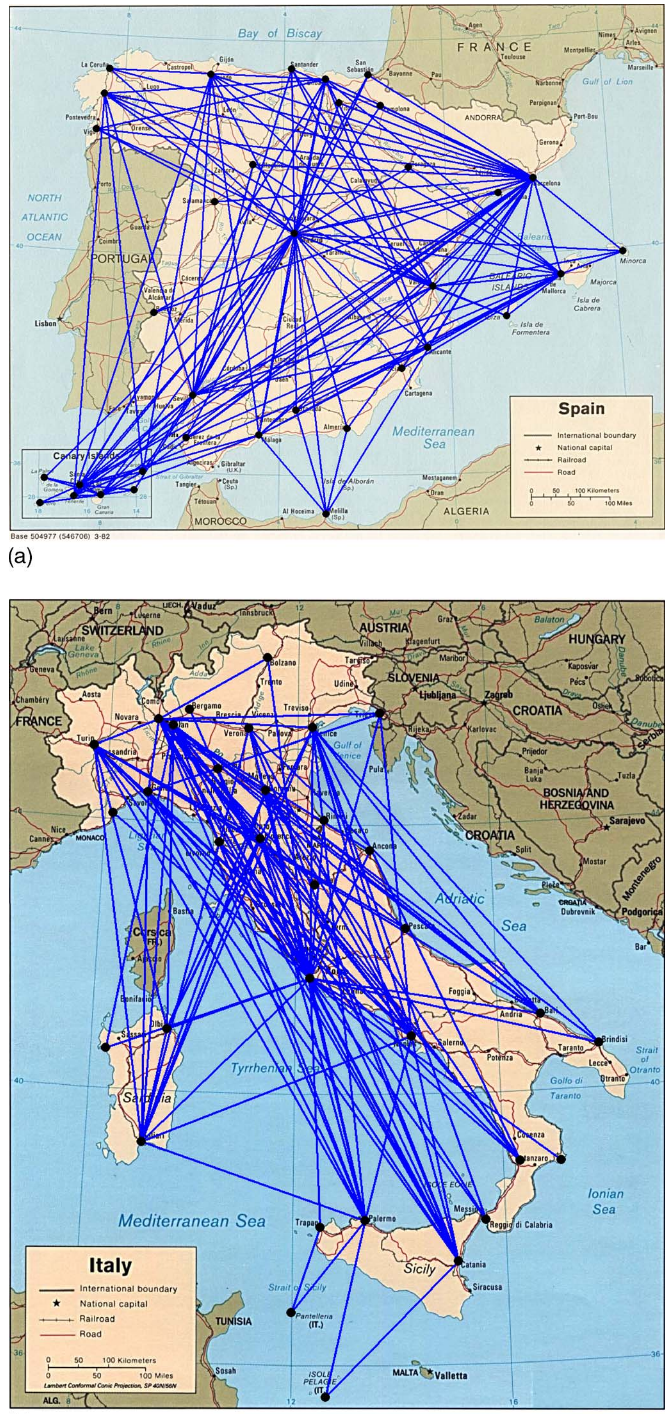


FIG. 3. Decision cost DC_{sp} of a 5×5 spatial network as a function of the number of links. At each step, we increase the number of links and recalculate the decision cost. The initial network has all links randomly distributed ($p=1$). Each point corresponds to the mean value of 50 simulations.



(a)

(b)

FIG. 4. (Color online) Graphical representation of the Spanish and Italian national airport network. Connections are obtained from domestic flights between airports.

tributed). However, up to now we have kept a constant number of links, but what happens if we increase it? The answer to this question can be inferred by analyzing Fig. 3, where we increase the number of links of a 5×5 regular network with its initial links randomly distributed. We can see that decision cost (i.e., the disorder) does not have a monotonic evolution. When few links are added, the decision cost of the network increases since we introduce them randomly. Nevertheless, there exists a critical value, related to the size of the network, at which the decision cost begins to decrease. This behavior is due to the fact that when too many links are included, the mean distance between nodes decreases drastically, arriving at the lowest value when all nodes are mutually connected.

DECISION COST IN REAL NETWORKS

We have analyzed the decision cost of two networks obtained from Italian and Spanish domestic flights, i.e., the direct connections between their main national airports. Figure 4 shows both networks plotted over their corresponding country maps, while in Fig. 5 we plot two typical distributions when analyzing the structure of a complex network: the cumulative degree distribution $P_c(k)$ [(a) and (d)] and the clustering coefficient distribution $C(k)$ [(b) and (e)]. We can observe that, in both cases, long tails are observed at $P_c(k)$ despite fluctuations appearing due to the low number of nodes ($N_{\text{Spain}}=35$ and $N_{\text{Italy}}=34$). The decay in $C(k)$ indicates the absence of degree correlations. We also include the node betweenness $B_{\text{node}}(k)$ as a function of the node degree [(c) and (f)], which measures the number of shortest paths that go through a certain airport of degree k. Details of other statistical values are given in Table I.

But let us focus on the main topic of the paper, namely the decision cost as a way to evaluate the disorder of a network. We compute the decision cost for both the Italian and Spanish networks obtaining $DC_{sp}^{\text{Italy}}=1.46$ and $DC_{sp}^{\text{Spain}}=1.21$, respectively. It is worth discussing the meaning of both values. As explained above, the decision cost measures the disorder of the network by evaluating the quantity of information requested to move optimally through the network. The fact that $DC_{sp}\neq0$ indicates that moving to the closest airport is not the best strategy to arrive at a desired place. This is a common feature in both airport networks, since worldwide airport connections are constructed around "central" airports that act as hubs, in what is called "Hub and Spoke" strategy.[15] Therefore, traveling through a hub reduces the covered distance, despite the fact that the hub may not be the closest airport to the target place.

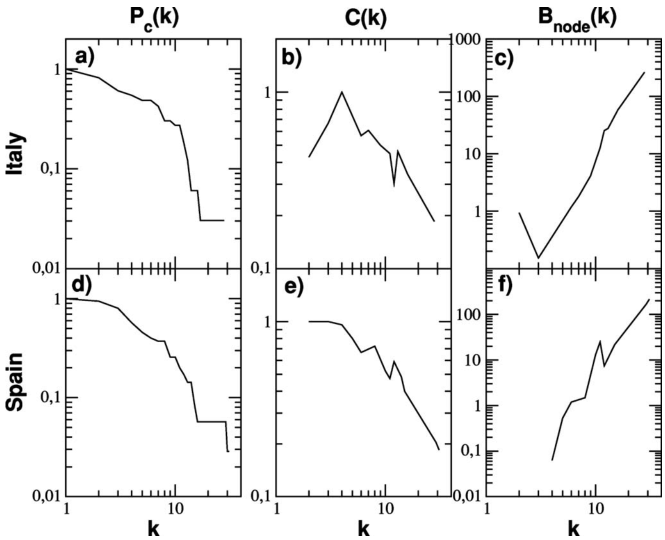


FIG. 5. Cumulative degree distribution $P_c(k)$, clustering coefficient $C(k)$, and node betweenness $B_{\text{node}}(k)$ for the Italian (upper plots) and Spanish (lower plots) airport network.

When looking at both values, a question arises: Is the Italian network more disordered than the Spanish one (since $\text{DC}_{\text{sp}}^{\text{Italy}} > \text{DC}_{\text{sp}}^{\text{Spain}}$)? In other words, what is the significance of the value of the decision cost? It is clear that networks with different numbers of nodes and links will have different decision costs, and a direct comparison between both networks is inadvisable.

In order to compare networks of different size, or to measure the percentage of disorder of a certain network, we must normalize the value of the decision cost. In the preceding section, we saw that, in regularly distributed networks, normalization through the randomly connected network leads to a normalized decision cost with a common shape, no matter the number of nodes considered in the network. However, in real networks, nodes are not regularly distributed and the reference network must be defined in a different way.

With this aim, we have defined a randomly connected airport network departing from an unconnected network where nodes occupy the real position of the airports. Next, we connect nodes randomly, all of them with at least one link, until we arrive at a total number of links equal to that of the real network. In this way, we obtain a network with exactly the same spatial distribution and the same number of connections but with the links randomly distributed. We calculate the decision cost of the random network DC_{ran} and we use it to normalize the decision cost of the real network ($\text{DC}_{\text{norm}} = \text{DC}_{\text{sp}} / \text{DC}_{\text{ran}}$), obtaining $\text{DC}_{\text{norm}}^{\text{Italy}} = 1.07$ and $\text{DC}_{\text{norm}}^{\text{Spain}} = 1.17$. Note the difference between both networks: despite the Italian network having higher decision cost, its normalized value is lower than the Spanish network. This fact indicates that the higher decision cost of the Italian network is due to a more disordered distribution of nodes, i.e., the location of the airports. However, by normalization, we skip the influence of the node distribution and account for the disorder introduced by the network connections. For this particular comparison, we can see how, in the Italian network, connections between airports have been created more efficiently than in the Spanish network.

TABLE I. Summary of several network parameters: Number of nodes n, number of links m, mean geodesic path $\bar{d}$, diameter $d_{\max}$ of the network, global clustering coefficient C, decision cost DC_{sp}, and decision cost of the randomly connected DC_{ran} network.

	National airport networks	
	Italy	Spain
n	33	35
m	105	123
$\bar{d}$ ($d_{\max}$)	1.92 (3)	1.84 (3)
C	0.418	0.738
DC_{sp}	1.46	1.21
DC_{ran}	1.366	1.035

CONCLUSIONS

In spatial networks, the position of the nodes in the Euclidean space is intrinsically related with the decision cost of traveling through the network. In this paper, we have introduced the concept of decision cost DC_{sp}, which measures the disorder of a given network taking into account not only the connections between nodes but their position in a two-dimensional map. Departing from a regular 2D network, we introduce a rewiring probability and evaluate its influence in the decision cost. The influence of the network size is also evaluated and we show that normalization of the decision cost allows us to compare the degree of disorder of networks of different sizes. We evaluate the decision cost of the connections between airports of two different countries and we obtain some conclusions about which of them is more disordered. We believe that the concept of decision cost can be applied not only to networks of higher dimensions but also to networks where nodes have a certain fitness property. By projecting the fitness in a one-dimensional map, it is possible to define a distance between nodes and, therefore, to evaluate the decision cost of the network.

[1] S. Milgram, "The small world problem," Psychol. Today **1**, 61 (1967).
[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," Nature (London) **393**, 440 (1998).
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Phys. Rep. **424**, 175 (2006).
[4] M. E. J. Newman, "The structure and function of complex networks," SIAM Rev. **45**, 167 (2003).
[5] J. M. Kleinberg, "Navigation in a small world," Nature (London) **406**, 845 (2000).
[6] P. Sen, S. Dagupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, "Small-world properties of the Indian Subway," Phys. Rev. E **67**, 036106 (2003).
[7] V. Latora and M. Marchiori, "Is the Boston subway a small-world network?" Physica A **314**, 109 (2002).
[8] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, "Networks and cities: An information perspective," Phys. Rev. Lett. **94**, 028701 (2005).
[9] K. Sneppen, A. Trusina, and M. Rosvall, "Hide-and-seek on complex networks," Europhys. Lett. **69**, 853 (2005).
[10] A. Cardillo, S. Scellato, V. Latora, and S. Porta, "Structural properties of planar graphs of urban street patterns," Phys. Rev. E **73**, 066107 (2006).
[11] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).

[12]R. V. Solé and S. Valverde, "Information theory of complex networks: On evolution and architectural constraints," Lect. Notes Phys. **650**, 189 (2004).
[13]B. Wang, H. Tang, C. Guo, and Z. Xiu, "Entropy optimization of scale-free networks' robustness to random failures," Physica A **363**, 591 (2006).
[14]M. Rosvall, P. Minnhagen, and K. Sneppen, "Navigating networks with limited information," Phys. Rev. E **71**, 066111 (2005).
[15]E. Merenne, *Géographie des Transports* (Nathan, Paris, 1995).