Diffusion at the Boundary: Measures to Characterize the Diffusion Process in Complex Networks*

Lucas Antiqueira,[1] Damián Blasi,[2] and Borys Wróbel[3]

[1] *Institute of Physics of São Carlos, University of São Paulo, São Carlos, SP, Brazil*†
[2] *Instituto Balseiro, Comision Nacional de Energia Atomica, San Carlos de Bariloche, Argentina*‡
[3] *Evolutionary Systems Laboratory, Institute of Oceanography, Polish Academy of Sciences, Sopot, and Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University in Poznan, Poland*§

(Dated: October 31, 2010)

Many propagation processes on networks can be explained by assuming a simple random walk dynamics. Inspired by the well-known relation between number of steps and mean squared distance in pure diffusive process, we studied diffusion over complex networks, using several models of networks: random, small-world, scale-free and geographical. Because the rate at which the process reaches larger distances is also of interest, we additionally studied the spreading rate, which is simply a measure of the random walker efficiency to spread away from the node that started the process. Our results thus far show that many important features of the network topology are displayed by these measures. The results to come involve the scaling of these measures with network size and density.

Keywords: complex networks, random walk, diffusion, boundary

I. INTRODUCTION

Many real-world systems that exhibit network structure hold non-trivial properties in between those of regular and random networks. To capture this fact they received the name of complex networks [1–3]. Several different complex network models have been proposed. One such model is the small-world [4], which departs from a purely random structure [5] by presenting high local clustering while distances between nodes are relatively small. Networks with a power-law degree distribution (scale-free networks [6]) provide another model suitable for systems where a small percentage of the nodes monopolize most of the links. Other models were proposed to have the so-called community structure [7], where groups of nodes are tightly connected to each other in comparison with the rest of the network.

A miscellany of temporal processes were studied using these models, where frequently the dynamics is driven by local interactions between each node and its neighbors. One may consider, for example, the information flow on the Internet [8] and in the brain [9], culture dissemination [10], rumor spreading [11] and the transmission of diseases like flu over the network of acquaintances [12] or HIV and other STDs over the network of sexual contacts [13]. Issues such as the characteristic time in which a propagation process reaches all nodes or which elements are efficient as transmitters are critical to the control and understanding of these phenomena.

The propagation dynamics is represented in this paper by the random walk [14], a well-known Markovian process frequently used to model navigation and exploration processes on networks. Although simple, the random walk is a powerful tool to understand the general diffusive dynamics on specific types of networks. We then investigate some essential properties of diffusion on networks to find out how fast random walkers go away from the origin and if they have a tendency to spread away from the origin at specific network positions (we call them *boundaries*). We then analyze several different network topologies: random [5], small-world [4], scale-free [6] and geographical [15] networks. The next section provides the necessary definitions, followed by a brief description of network models used in simulation experiments. Results of the experiments are provided along with a discussion of observed propagation features. Finally, we present general conclusions of our analysis.

II. DEFINITIONS AND PROPOSED MEASURES

A network can be represented as a *graph*: a tuple $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N} = \{1, \ldots, N\}$ is the set of N nodes and $\mathcal{E} = \{e_1, \ldots, e_E\}$ is the set of E edges. Each edge e_i is a set $\{a, b\}$ that represents the undirected connection between nodes a and b. A graph can be described by an adjacency matrix A of size $N \times N$, whose elements A_{ab} are equal to 1 whenever there exists an edge i so that $e_i = \{a, b\}$ (otherwise $A_{ab} = 0$). A *walk* on a graph is defined as the string of nodes $\mathcal{W}(n_1, n_W) = (n_1, \ldots, n_W)$ where each pair of subsequent nodes is connected by an edge. The length of a walk is its number of edges, $W - 1$. A *path* is a walk where each node in the string $\mathcal{W}$ is visited only once. The *minimum path* $\mathcal{W}_{min}(a, b)$ between nodes

*Working paper prepared for the Complex Systems Summer School (CSSS), Santa Fe Institute (SFI), USA, 2010.

†Electronic address: lantiq@ursa.ifsc.usp.br
‡Electronic address: blasid@ib.cnea.gov.ar
§Electronic address: bwrobel@iopan.gda.pl

a and b is the path of smallest length $\ell(a,b)$ starting at node a and ending at b. We now define (i) the *distance* between nodes a and b as the length of the minimum path $\ell(a,b)$, (ii) the *distance matrix* D with elements $D_{ij} = \ell(i,j)$, and (iii) the *squared distance matrix* $D'_{ij} = D^2_{ij}$. In this paper we only deal with connected graphs: a path can be found for any pair of nodes. Finally, we define the *boundary* $\mathcal{B}(i,h)$ of a node i:

$$\mathcal{B}(i,h) = \{j : d(i,j) = h\} \tag{1}$$

which is the set of all nodes at distance h from i, called also the hierarchical (or concentric) level h of node i [16]. The cardinality of the set $\mathcal{B}(i,h)$, $|\mathcal{B}(i,h)|$, is always less than N save for the trivial case of a graph consisting of a single node when the equality holds.

A discrete time *random walk* on a graph is driven by the stochastic matrix S defined as

$$S_{ij} = \frac{A_{ij}}{\sum_{k=1}^{N} A_{ik}} \tag{2}$$

where S_{ij} is the probability of a random walker going from node i to j in one time step. The number of paths existing between nodes i and j after t walker steps is $A(t)_{ij}$, where matrix $A(t) = A^t$, the t-th power of A. We then define the *Brownian span* $S(t)$ as a matrix with elements $S(t)_{ij}$, the normalized number of paths of length t starting at node i and ending at j:

$$S(t)_{ij} = \frac{A(t)_{ij}}{\sum_{k=1}^{N} A(t)_{ik}} \tag{3}$$

In other words, $S(t)_{ij}$ is the probability of a random walker going from node i to j in t time steps. The *average squared distance* $\langle d^2(i,t)\rangle$ of the random walker to the origin node i after t time steps is given by

$$\langle d^2(i,t)\rangle = \mathrm{diag}(S(t)D')_i \tag{4}$$

where $\mathrm{diag}(X)_i$ gives the i-th element of the diagonal of X. This quantity is interesting because in pure diffusive processes [17]

$$\sqrt{\langle d^2(i,t)\rangle} \propto t^{1/2} \tag{5}$$

This formula has a probabilistic interpretation: the exponent of t is related to the chance a stochastic process has to progressively go farther from the diffusion origin.

As far as we know, the relationship between the average squared distance and t for complex networks has not been yet derived analytically. In fact, this problem is closely related to finding the distance distribution of a network, a task that has been solved only for random networks under certain assumptions [18]. We therefore propose to measure this relationship numerically in order to evaluate the tendency of the walkers to go away from the origin in different graph structures.

As time goes by, the above defined quantity is more strongly influenced by the nodes lying at the current boundary. Nevertheless, the precise information about the spreading process at the boundary is not reflected by this measure. To tackle this issue, we measure the probability of a walker going away the origin given that it is currently at boundary $\mathcal{B}(i,h)$

$$\alpha(i,h) = \frac{\sum_{j\in\mathcal{B}(i,h)} k_j^+(i,h)}{\sum_{j\in\mathcal{B}(i,h)} (k_j - k_j^+(i,h))} \tag{6}$$

where k_j is the number of edges attached to node j (degree) and $k_j^+(i,h)$ is the number of edges linking node j to the next boundary $h+1$ having node i as reference (origin) node. In other words, α measures the *spreading rate* of the diffusion at a given boundary by considering the proportion of links leading the current boundary to the next in relation to the number of links connecting the current boundary to itself or to a previous one.

III. NETWORK MODELS

There is a large number of network models, but two are commonly believed to capture some of the essential properties of real world systems: the small-world and scale-free models. In our analysis we also include the classic random graph model and a model with a spatially constrained topology, the geographical model.

In *Random networks*, introduced by Erdős and Rényi [5] and Gilbert [19], nodes are connected to each other with probability p, and the node degree follows the Poisson distribution with the average $\langle k\rangle = p(N-1)$ in the limit $N \to \infty$, where N is the total number of nodes.

Scale-free networks, on the other hand, have a power-law degree distribution: a few nodes have many connections (the *hubs*) and a large number of nodes is poorly connected. Barabási and Albert [6] proposed a way to generate scale-free networks that includes two mechanisms: growth and preferential attachment. The network starts with m_0 nodes. A single node is added at each time step and connected to m nodes already in the network. These m new edges are created with a preference for nodes with large degrees: a new node i is connected to node j with probability $P(i \to j) = k_j / \sum_u k_u$. The average degree in a scale-free network is $\langle k\rangle = 2m$.

Small-world networks are networks in which the distances between nodes are small (and similar to the distances in a random network of the same size) and neighboring nodes belong to highly connected clusters. In the small-world model devised by Watts and Strogatz [4], the starting structure is regular, with every node connected to its 2η closest neighbors. Then, a rewiring process takes place: each edge is rewired with probability p, interconnecting distant nodes. The parameter p allows to go from a regular network ($p = 0$) to a random network ($p = 1$). A small-world network has average degree $\langle k\rangle = 2\eta$.

Geographical networks intend to capture the fact that real world systems (eg. highway and urban networks) are spatially constrained: nodes close to each other in space

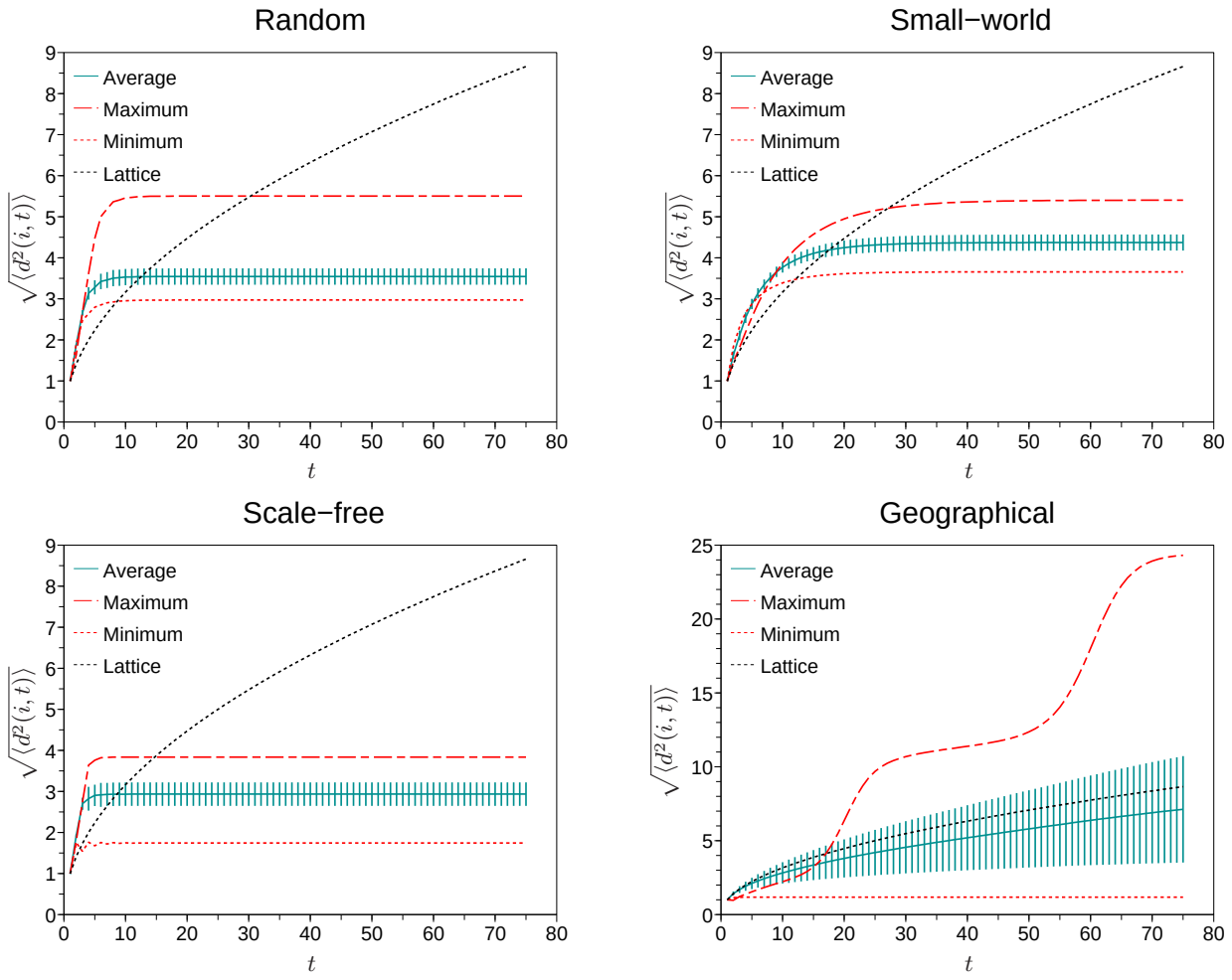

FIG. 1: Average squared distances of random walkers to the origin as a function of discrete time t for different network topologies. Solid lines correspond to the mean for all nodes in the network averaged over 50 independent network realizations, standard deviations are shown as vertical bars. Red dashed lines correspond to the maximum and minimum distances. The known relation $t^{1/2}$ for diffusive processes is shown in blue. Note that the scale of the y-axis is different for geographical networks.

have a greater chance of being connected than nodes far away from one another [15]. Although many models allow for spacial constraints, we will focus here on a simple model with a very small chance for direct connections between spatially distant nodes, which prevents the emergence of the small-world effect. In our model, nodes are randomly embedded in a 2D space of size $L \times L$ and each node i is connected to other nodes distant at most δ from i. In this model, finite border effects emerge and the average node degree can be approximated as $\langle k \rangle \approx \pi\delta^2 N/L^2$.

IV. RESULTS AND DISCUSSION

To evaluate the diffusion propagation at the boundary for complex networks with different topologies, we generated 50 realizations of each network model with $N = 1000$ nodes and $E = 4000$ edges (ie. average degree $\langle k \rangle = 8$). For each node in each network, we computed the square root of the average squared distance (Eq. 4) and the spreading rate at the boundary (Eq. 6).

The diffusion process on random and scale-free networks has similar features: *super-diffusion* (the average square distance growing more rapidly than $t^{1/2}$) during the first 10 time steps followed by the stationary regime (Fig. 1). Even though small-world networks also show a super-diffusive behavior, the stationary regime is reached about twice slower than in random and scale-free networks, which is a signature of the greater average clustering in this kind of networks. The minimum curve of the scale-free network is reflecting the existence of hubs: highly connected nodes from where any node is reached in a few steps. Geographical networks show a very different behavior: sub-diffusion at first and then very large standard deviations. A possible explanation for the minimum and maximum distance curves in this kind of network is the absence of long-range links: the diameter (the largest distance between any two nodes) of the network is rela-

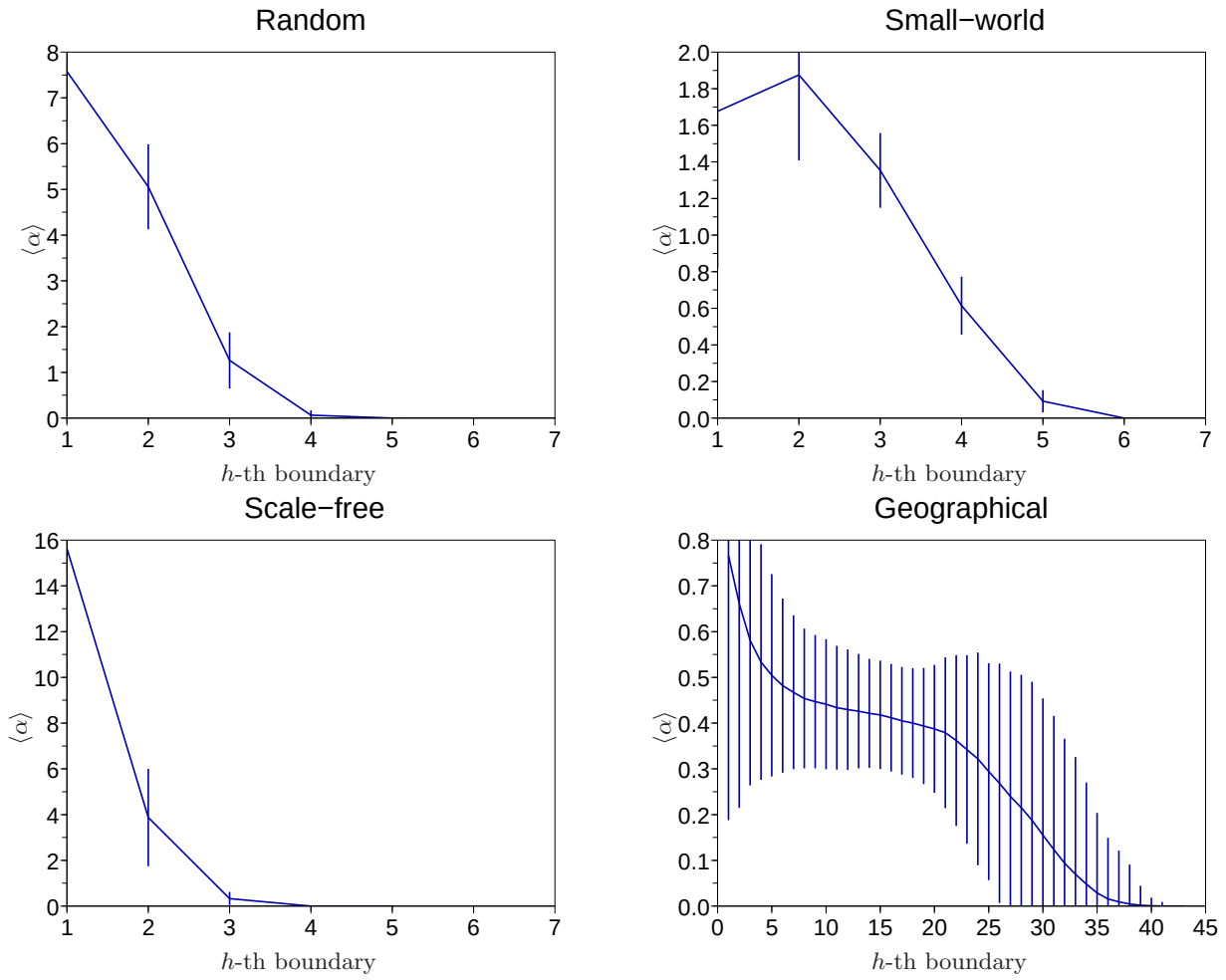

FIG. 2: The average spreading rate $\langle\alpha\rangle$ for each network type as a function of boundary level h. Solid lines correspond to the mean for all nodes in the network averaged over 50 independent network realizations with standard deviations shown as vertical bars. Note that the scale of the y-axis is different for all networks (x-axis is scaled differently only for geographical networks).

tively large, so the walkers may reach large distances or may stay around the origin for a long time without reaching other network regions. When a walker stays trapped in a region and later reaches other regions, a distance curve with inflection points like the maximum curve in Fig. 1 is observed.

As was mentioned above, the spreading rate is tightly linked to the distance distribution and we can observe that the characteristic features of small-world and scale free networks are clearly present. In scale-free networks a large number of nodes have few links and a small group monopolizes most of them. Therefore, high spreading rate at the beginning (see Fig. 2) can be explained as follows: the most probable scenario for the process is to have an origin on a weakly connected node, and in the following step, a densely connected node is reached, reaching a boundary where there are much more outgoing than incoming links. When comparing random and small-world networks (both having similar distances between nodes) it is possible to argue that the high clustering in the small-world structure is responsible for the high dissimilarity in the measure on the first four boundaries (Fig. 2). In these networks two immediate neighbors of a given node are likely to be connected, which gives rise to many links lying inside the same boundary, thus reducing the spreading rate.

For geographical networks, the spreading rate is below 1 even for the first few boundaries, which means that the diffusion tends to stay near the origin. This behavior can be explained using, once again, the fact that geographical networks present a local connectivity pattern, without long-range links. Another consequence of this feature is that boundary levels h can be much larger for geographical networks than for other models. It is also interesting that that when the diffusion reaches further boundaries, a plateau near $\langle\alpha\rangle \sim 0.4$ is found for h around 10–20, showing that, after leaving the origin neighborhood, the diffusion tends to stay in a region of similar connectivity. Finally, for more distant diffusion boundaries, the spreading rate decays again, since

boundaries are now closer to the network border (recall that these networks are embedded in 2D space).

V. CONCLUDING REMARKS

In this work we describe the diffusion process on networks with different topologies using two measures of propagation: (i) the average distance of random walkers to the origin at time t and (ii) the spreading rate calculated using the relative number of outgoing links in a given boundary h. The results of our simulations show that small-world, random and scale-free networks can be put in this order according to the increasing spreading rate and decreasing average distance in a stationary regime, which is reached later in the case of small-world networks. Geographical networks behave differently since they do not have hubs nor long-range links to shorten the distances between nodes. In this kind of networks the diameter is large and walkers may stay in one region for a long time before they diffuse to other, more distant regions. This results in larger average and maximum distances, in the presence of inflection points in single distance curves and in the maintenance of non-zero spreading rate for much farther boundary levels. We believe that the measures developed here reflect the clustering coefficient, distance and community structure of the networks, therefore showing how these properties influence a diffusion process on networks. We are currently working to show how the proposed measures scale with the number of nodes and links in the networks.

Acknowledgments

The authors would like to thank Tom Carter and Greg Leibon for helpful discussions. Lucas Antiqueira is grateful to FAPESP (Brazil) for sponsorship #06/61743-7.

[1] R. Albert and A. L. Barabási, Reviews of Modern Physics **74**, 47 (2002).
[2] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[3] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Reviews of Modern Physics **80**, 1275 (2008).
[4] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[5] P. Erdős and A. Rényi, Publicationes Mathematicae (Debrecen) **6**, 290 (1959).
[6] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[7] S. Fortunato, Physics Reports **486**, 75 (2010).
[8] J. L. Iribarren and E. Moro, Physical Review Letters **103**, 038702 (2009).
[9] E. Salinas and T. J. Sejnowski, Nature Reviews Neuroscience **2**, 539 (2001).
[10] R. Axelrod, Journal of Conflict Resolution **41**, 203 (1997).
[11] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, Physica A **374**, 457 (2007).
[12] L. Ancel Meyers, M. E. Newman, M. Martin, and S. Schrag, Emerging Infectious Diseases **9**, 204 (2003).
[13] J. H. Jones and M. S. Handcock, Nature **423**, 605 (2003).
[14] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
[15] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, Advances in Physics **56**, 167 (2007).
[16] L. da F. Costa and F. N. Silva, Journal of Statistical Physics **125**, 841 (2006).
[17] S. Chandrasekhar, Reviews of Modern Physics **15**, 1 (1943).
[18] V. Blondel, J. Guillaume, J. Hendrickx, and R. Jungers, Physical Review E **76**, 066101 (2007).
[19] E. N. Gilbert, Annals of Mathematical Statistics **30**, 1141 (1959).