Random Walks on Complex Networks

Jae Dong Noh
Department of Physics, Chungnam National University, Daejeon 305-764, Korea

Heiko Rieger
Theoretische Physik, Universität des Saarlandes, 66041 Saarbrücken, Germany

(Received 29 July 2003; published 18 March 2004)

We investigate random walks on complex networks and derive an exact expression for the mean first-passage time (MFPT) between two nodes. We introduce for each node the random walk centrality C, which is the ratio between its coordination number and a characteristic relaxation time, and show that it determines essentially the MFPT. The centrality of a node determines the relative speed by which a node can receive and spread information over the network in a random process. Numerical simulations of an ensemble of random walkers moving on paradigmatic network models confirm this analytical prediction.

DOI: 10.1103/PhysRevLett.92.118701 PACS numbers: 89.75.Hc, 05.40.Fb, 05.60.Cd

From biology over computer science to sociology the world is abundant in networks. Watts and Strogatz [1] demonstrated that many of these real-world networks pose characteristic features such as the small world and clustering property. Since classical network models do not display these features, new models have been developed to understand the structure underlying real-world networks [2,3]. The important finding is that there exists a scale-free (SF) network characterized by a power-law degree distribution [4], that is also a characteristic for the World Wide Web (WWW) [4], and for many other networks in various disciplines [2].

The SF network has a heterogeneous structure. For instance, the WWW analyzed in Ref. [4] includes nodes with degrees ranging from $\mathcal{O}(1)$ to $\mathcal{O}(10^3)$. The heterogeneity leads to intriguing properties of SF networks. In the context of percolation, SF networks are stable against random removal of nodes while they are fragile under intentional attacks targeting on nodes with high degree [5,6]. Statistical mechanical systems on SF networks also display interesting phase transitions [7–9]. In a transport process, each node in SF networks does not contribute equally likely. The importance of each node in such a process is measured with the betweenness centrality [10], which has a broad power-law distribution [11].

In this Letter, we study a random walk on general networks with a particular attention to SF networks. The random walk is a fundamental dynamic process [12]. It is theoretically interesting to study how the structural heterogeneity affects the nature of the diffusive and relaxation dynamics of the random walk [13]. Those issues will be studied further elsewhere [14]. The random walk is also interesting since it could be a mechanism of transport and search on networks [15–17]. Those processes would be optimal if one follows the shortest path between two nodes under considerations. Among all paths connecting two nodes, the shortest path is given by the one with the smallest number of links [18]. However the shortest path can be found only after global connectivity is known at each node, which is improbable in practice. The random walk becomes important in the extreme opposite case where only local connectivity is known at each node. We also suggest that the random walk is a useful tool in studying the structure of networks.

In the context of transport and search, the mean first-passage time (MFPT) is an important characteristic of the random walk. We will derive an exact formula for the MFPT of a random walker from one node i to another node j, which will be denoted by $\langle T_{ij} \rangle$, in arbitrary networks. In the optimal process it is just given by the number of links in the shortest path between two nodes, and both motions to one direction and to the other direction are symmetric. However, a random walk motion from i to j is not symmetric with the motion in the opposite direction. The asymmetry is characterized with the difference in the MFPT's. It is revealed that the difference is determined by a potential-like quantity which will be called the *random walk centrality* (RWC). The RWC links the structural heterogeneity to the asymmetry in dynamics. It also describes centralization of information wandering over networks.

We consider an arbitrary *finite* network (or graph) which consists of nodes $i = 1, \ldots, N$ and links connecting them. We assume that the network is connected [i.e., there is a path between each pair of nodes (i, j)], otherwise we simply consider each component separately. The connectivity is represented by the adjacency matrix $\mathbf{A}$ whose element $A_{ij} = 1(0)$ if there is a link from i to j (we set $A_{ii} = 0$ conventionally). In the present Letter, we restrict ourselves to an *undirected* network, namely $A_{ij} = A_{ji}$. The degree, the number of connected neighbors, of a node i is denoted by K_i and given by $K_i = \sum_j A_{ij}$.

The stochastic process in discrete time that we study is a random walk on this network described by a master

equation. The transition probabilities are defined by the following rule: A walker at node i and time t selects one of its K_i neighbors with *equal* probability to which it hops at time $t+1$, thus the transition probability from node i to node j is A_{ij}/K_i [20]. Suppose the walker starts at node i at time $t=0$, then the master equation for the probability P_{ij} to find the walker at node j at time t is

$$P_{ij}(t+1) = \sum_k \frac{A_{kj}}{K_k} P_{ik}(t). \tag{1}$$

The largest eigenvalue of the corresponding time evolution operator is 1 corresponding to the stationary distribution $P_j^\infty = \lim_{t\to\infty} P_{ij}(t)$, i.e., the infinite time limit [21]. An explicit expression for the transition probability $P_{ij}(t)$ to go from node i to node j in t steps follows by iterating Eq. (1)

$$P_{ij}(t) = \sum_{j_1,\ldots,j_{t-1}} \frac{A_{ij_1}}{K_i} \cdot \frac{A_{j_1 j_2}}{K_{j_1}} \cdots \frac{A_{j_{t-1} j}}{K_{j_{t-1}}}. \tag{2}$$

Comparing the expressions for P_{ij} and P_{ji} one sees immediately that

$$K_i P_{ij}(t) = K_j P_{ji}(t). \tag{3}$$

This is a direct consequence of the undirectedness of the network. For the stationary solution, Eq. (3) implies that $K_i P_j^\infty = K_j P_i^\infty$, and therefore one obtains

$$P_i^\infty = \frac{K_i}{\mathcal{N}}, \tag{4}$$

with $\mathcal{N} = \sum_i K_i$. Note that the stationary distribution is, up to normalization, equal to the degree of the node i—the more links a node has to other nodes in the network, the more often it will be visited by a random walker.

How fast is the random walk motion? To answer to this question, we study the MFPT. The first-passage probability $F_{ij}(t)$ from i to j after t steps satisfies the relation

$$P_{ij}(t) = \delta_{t0}\delta_{ij} + \sum_{t'=0}^{t} P_{jj}(t-t') F_{ij}(t'). \tag{5}$$

The Kronecker delta symbol insures the initial condition $P_{ij}(0) = \delta_{ij}$ [$F_{ij}(0)$ is set to zero]. Introducing the Laplace transform $\tilde{f}(s) \equiv \sum_{t=0}^\infty e^{-st} f(t)$, Eq. (5) becomes $\tilde{P}_{ij}(s) = \delta_{ij} + \tilde{F}_{ij}(s)\tilde{P}_{jj}(s)$, and one has

$$\tilde{F}_{ij}(s) = [\tilde{P}_{ij}(s) - \delta_{ij}]/\tilde{P}_{jj}(s). \tag{6}$$

In finite networks the random walk is recurrent [12], so the MFPT is given by $\langle T_{ij}\rangle = \sum_{t=0}^\infty t F_{ij}(t) = -\tilde{F}'_{ij}(0)$.

Since all moments $R_{ij}^{(n)} \equiv \sum_{t=0}^\infty t^n \{P_{ij}(t) - P_j^\infty\}$ of the exponentially decaying relaxation part of $P_{ij}(t)$ are finite, one can expand $\tilde{P}_{ij}$ as a series in s as

$$\tilde{P}_{ij}(s) = \frac{K_j}{\mathcal{N}(1-e^{-s})} + \sum_{n=0}^\infty (-1)^n R_{ij}^{(n)} \frac{s^n}{n!}. \tag{7}$$

Inserting this series into Eq. (6) and expanding it as a power series in s, we obtain that

$$\langle T_{ij}\rangle = \begin{cases} \frac{\mathcal{N}}{K_j}, & \text{for } j = i \\ \frac{\mathcal{N}}{K_j}[R_{jj}^{(0)} - R_{ij}^{(0)}], & \text{for } j \neq i \end{cases}. \tag{8}$$

A similar expression is derived in Ref. [12] for the MFPT of the random walk in *periodic* lattices.

It is very interesting to note that the average return time $\langle T_{ii}\rangle$ does not depend on the details of the global structure of the network. It is determined only by the total number of links and the degree of the node. Since it is inversely proportional to the degree, the heterogeneity in connectivity is well reflected in this quantity. In a SF network with degree distribution $P(K) \sim K^{-\gamma}$, the MFPT to the origin T_o also follows a power-law distribution $P(T_o) \sim T_o^{-(2-\gamma)}$. The MFPT to the origin distributes uniformly in the special case with $\gamma = 2$.

Random walk motions between two nodes are asymmetric. The difference between $\langle T_{ij}\rangle$ and $\langle T_{ji}\rangle$ for $i \neq j$ can be written as [using Eq. (8)]

$$\langle T_{ij}\rangle - \langle T_{ji}\rangle = \mathcal{N}\left(\frac{R_{jj}^{(0)}}{K_j} - \frac{R_{ii}^{(0)}}{K_i}\right) - \mathcal{N}\left(\frac{R_{ij}^{(0)}}{K_j} - \frac{R_{ji}^{(0)}}{K_i}\right),$$

where the last term vanishes due to Eq. (3). Therefore we obtain

$$\langle T_{ij}\rangle - \langle T_{ji}\rangle = C_j^{-1} - C_i^{-1}, \tag{9}$$

where C_i is defined as

$$C_i \equiv \frac{P_i^\infty}{\tau_i}, \tag{10}$$

where $P_i^\infty = K_i/\mathcal{N}$ and the characteristic relaxation time τ_i of the node i is given by

$$\tau_i = R_{ii}^{(0)} = \sum_{t=0}^\infty \{P_{ii}(t) - P_i^\infty\}. \tag{11}$$

We call C_i the *random walk centrality* since it quantifies how central a node i is located regarding its potential to receive information randomly diffusing over the network. To be more precise: Consider two nodes i and j with $C_i > C_j$. Assume that each of them launches a signal simultaneously, which is wandering over the network. Based on Eq. (9), one expects that the node with larger RWC will receive the signal emitted by its partner earlier. Hence, the RWC can be regarded as a measure for effectiveness in communication between nodes. In a homogeneous network with translational symmetry, all nodes have the same value of the RWC. On the other hand, in a heterogeneous network the RWC has a distribution, which leads to the asymmetry in the random dynamic process.

The RWC is determined by the degree K and τ. The order of magnitude of the characteristic relaxation time τ is related to the second largest eigenvalue (*nota bene* [21]) of the time evolution operator in (1): $P_{ii}(t) = P_i^\infty + \sum_{\lambda=2}^N a_i^{(\lambda)} b_i^{(\lambda)} \Lambda_\lambda^t$, where $a^{(\lambda)}$ and $b^{(\lambda)}$ are the left and right

eigenvectors, respectively, of the time evolution operator belonging to the eigenvalue Λ_λ. If we order the eigenvalues according to the modulus ($|\Lambda_2| \geq |\Lambda_3| \geq \ldots \geq |\Lambda_N|$) the asymptotic behavior is $P_{ii}(t) - P_i^\infty \sim a_i^{(2)} b_i^{(2)} \Lambda_2^t$ and $\tau_i \approx a_i^{(2)} b_i^{(2)} / |\ln|\Lambda_2||$. Thus the relaxation time τ_i has a node dependence only through the weight factor, which is presumably weak. On the other hand, the degree dependence is explicit.

We examined the distribution of the RWC in the Barabási-Albert (BA) network [22]. This is a model for a growing SF network; at each time step, a new node is added creating m links with other nodes which are selected with the probability proportional to their degree. We grew the network, solved the master equation numerically with the initial condition $P_i(t = 0) = \delta_{ik}$, and calculated the relaxation time τ_k for each k. Figure 1(a) shows the plot of τ vs K in the BA network of $N = 10^4$ nodes grown with the parameter $m = 2$. The degree is distributed broadly over the range $2 \leq K \lesssim 400$. On the other hand, the relaxation time turns out to be distributed very narrowly within the range $1 \lesssim \tau \lesssim 2$. We also studied BA networks of different sizes, but did not find any significant broadening of the distribution of τ. So the RWC distribution is mainly determined by the degree distribution. In Fig. 1(b) we show the plot of C vs K in the same BA network. It shows that the RWC is roughly proportional to the degree. Note, however, that the RWC is not increasing monotonically with the degree due to the fluctuation of τ as seen in Fig. 1(a).

The RWC is useful when one compares the random walk motions between two nodes, e.g., i and j with $C_i > C_j$. On average a random walker starting at j arrives at i before another walker starting at i arrives at j. Now consider an intermediate node k, which may be visited by both random walkers. Since $\langle T_{ij} \rangle > \langle T_{ji} \rangle$, it is likely that a random walker starting at node k will arrive at node i earlier than at node j. Although this argument is not exact since we neglected the time spent on the journey to the intermediate node, it indicates that nodes with larger RWC may be typically visited earlier than nodes with smaller RWC by the random walker. If we interpret the random walker as an information messenger, nodes with larger RWC are more efficient in receiving information than nodes with smaller RWC.

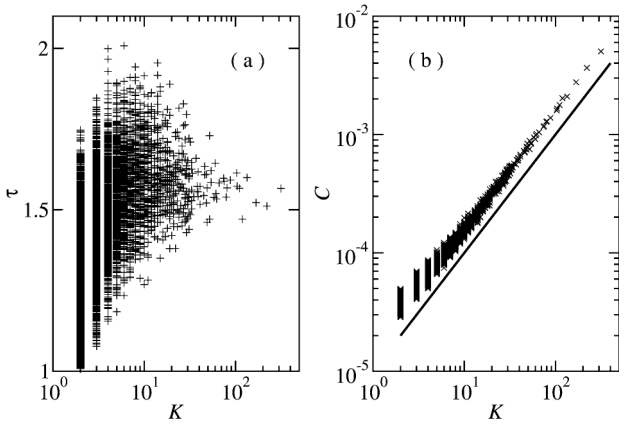

FIG. 1. (a) τ vs K and (b) C vs K calculated in the Barabási-Albert network with $N = 10\,000$ and $m = 2$. The straight line in (b) has the slope 1.

We performed numerical simulations to study the relation between the RWC and this efficiency. To quantify it, we consider a situation where initially all nodes in a network are occupied by different random walkers. They start to move at time $t = 0$, and we measure n_i, the fraction of walkers which have passed through the node i, as a function of time t. It is assumed that the walkers do not interact with each other. They may be regarded as a messenger delivering information to each node it visits. Then, with the information distribution uniformly initially, n_i is proportional to the amount of information acquired by each node. The argument in the previous paragraph suggests that typically nodes with larger values of RWC have larger value of n_i at any given time.

The BA network [22] and the hierarchical network of Ravasz and Barabási [23] were considered in the simulations. The hierarchical network is a deterministic network growing via iteration; at each iteration the network is multiplied by a factor M. The emergent network is scale free when $M \geq 3$. Since it is a deterministic network, several structural properties are known exactly [24]. We measured n_i in the BA network with $m = 2$ and $N = 512$ nodes and in the hierarchical network with $M = 5$ and $N = 5^4$ nodes for $0 \leq t \leq 2048$, which are presented in the left and the right column of Fig. 2, respectively. The

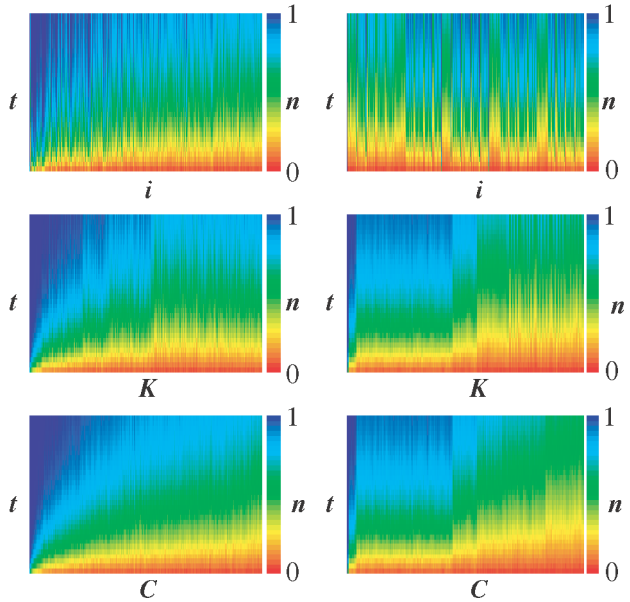

FIG. 2 (color online). Time evolution of the fraction of walkers n that pass through a node as a function of (from top to bottom) the node index i, the node degree K and the RWC C of the BA network (left column) and the hierarchical network (right column). The value of n at each time t is represented in the gray scale/color code depicted at the right border of each plot.

value of n_i is indicated in gray scale (color code) according to the reference shown in Fig. 2.

The time evolution of n_i is presented in three different ways. In the first row, the nodes are arranged in ascending order of the node index i. In the BA network, the node index corresponds to the time step at which the node is added to the network. The indexing scheme for the hierarchical network is explained in Ref. [24]. In the second row, the nodes are arranged in descending order of the degree K and in the third row they are arranged in descending order of the RWC C. At a given time t, the plot in the first row shows that n is nonmonotonous and very irregular as a function of the node index. As a function of the degree it becomes smooth, but still nonmonotonic tendencies remain. However, as a function of the RWC, it becomes much smoother and almost monotonous. We calculated for each node i the time τ_i' at which n_i becomes greater than $1/2$. In the BA network, among all node pairs (i, j) satisfying $\tau_i' < \tau_j'$, only 3% violate the relation $C_i > C_j$, whereas the number of pairs that violate the relation $K_i > K_j$ is 5 times larger.

In summary, we studied the random walk processes in complex networks. We derive an exact expression for the mean first-passage time [see Eq. (8)]. The MFPT's between two nodes differ for the two directions in general heterogeneous networks. We have shown that this difference is determined by the random walk centrality C defined in Eq. (10). Among random walk motions between two nodes, the walk to the node with larger value of C is faster than the other. Furthermore, it is argued that in a given time interval nodes with larger values of C are visited by more random walkers which were distributed uniformly initially. We confirmed this by numerical simulations on the BA and the hierarchical network. One may regard the random walkers as informations diffusing through the network. Our results imply that information does not distribute uniformly in heterogeneous networks; the information is centralized to nodes with larger values of C. The nodes with high values of C have the advantage of being aware of new information earlier than other nodes. On the other hand, it also implies that such nodes are heavily loaded within an information distribution or transport process. If the network has a finite capacity, the heavily loaded nodes may cause congestions [17]. Therefore much care should be taken of the nodes with high C values in network management. In the current work, we consider the random walks on undirected networks. The generalization to directed networks would be interesting. And in order to study congestion, the random walk motions with many *interacting* random walkers would also be interesting. We leave such generalizations to a future work.

Acknowledgement: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and by the European Community's Human Potential Programme under Contract No. HPRN-CT-2002-00307, DYGLAGEMEM.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[5] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[6] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[8] S. N. Dorogovtsev, A.V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **66**, 016104 (2002).
[9] F. Iglói and L. Turban, Phys. Rev. E **66**, 036140 (2002).
[10] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001).
[11] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001);K.-I. Goh, E. S. Oh, H. Jeong, B. Kahng, and D. Kim, Proc. Natl. Acad. Sci. U.S.A. **99**, 12583 (2002).
[12] R. D. Hughes, *Random Walks: Random Walks and Random Environments* (Clarendon, Oxford, 1995), Vol. 1.
[13] S. Jespersen, I. M. Sokolov, and A. Blumen, Phys. Rev. E **62**, 4405 (2000); B. Tadić, Eur. Phys. J. B **23**, 221 (2001); H. Zhou, cond-mat/0302030.
[14] J. D. Noh and H. Rieger (unpublished).
[15] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys. Rev. E **64**, 046135 (2001).
[16] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[17] P. Holme, cond-mat/0301013.
[18] In this definition, all links are assumed to be equivalent. Evolution of the shortest path in weighted networks is discussed in Ref. [19].
[19] J. D. Noh and H. Rieger, Phys. Rev. E **66**, 066127 (2002).
[20] In weighted networks, the hopping probability may be written as $\tilde{A}_{ij}/\tilde{K}_i$, where $\tilde{A}_{ij} = w_{ij}A_{ij}$ and $\tilde{K}_i = \sum_j \tilde{A}_{ij}$ with a weight $w_{ij} > 0$. All results in this Letter remain valid as long as the weight is symmetric, i.e., $\tilde{A}_{ij} = \tilde{A}_{ji}$.
[21] The limit exists if and only if the network contains an odd loop. Then all other eigenvalues of the time evolution operator satisfy $|\Lambda| < 1$, otherwise there exists an eigenvalue $\Lambda = -1$, for which the infinite time limit does not exist. In such cases, one may redefine the RW model setting $A_{ii} = 1$ to make the limit exist.
[22] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[23] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002); E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[24] J. D. Noh, Phys. Rev. E **67**, 045103(R) (2003).