Exploring complex networks by walking on them

Shi-Jie Yang
Department of Physics, Beijing Normal University, Beijing 100875, China
(Received 10 June 2004; revised manuscript received 8 October 2004; published 5 January 2005)

We carry out a comparative study of the problem of a walker searching several typical complex networks. The search efficiency is evaluated for various strategies. Having no knowledge of the global properties of the underlying networks and the optimal path between any two given nodes, it is found that the best search strategy is the self-avoiding random walk. The preferentially self-avoiding random walk does not help in improving the search efficiency further. In return, topological information of the underlying networks may be drawn by comparing the results of the different search strategies.

DOI: 10.1103/PhysRevE.71.016107 PACS number(s): 89.75.Hc, 05.40.Fb, 89.75.Fb, 89.75.Da

I. INTRODUCTION

In the past few years, much scientific interest has been devoted to the characterization and modelling of a wide range of complex systems that can be described as networks [1,2]. Systems such as the Internet [3–5] or the World Wide Web [6], social communities [7], food webs [8], and biological interacting networks [9–11] are represented as a graph, in which nodes represent the population individuals and links the physical interactions among them. Among these networks, most have complex topological properties and dynamical features that cannot be accounted for by classical graph modelling. It has been demonstrated that many of these real-world networks pose small-world and clustering property [12]. On the other hand, scale-free (SF) degree distributions seem to emerge frequently as dominant features governing the topology of real-world networks [13]. These global properties imply a large connectivity heterogeneity and a short average distance between nodes, which have considerable impact on the behavior of physical processes taking place on top of the network. A number of models have been developed to understand the structure and functions of underlying real-world networks. For instance, scale-free networks have been shown to be resilient to random damage [14–16] while also shown to be fragile under intentional attacks targeting the nodes with high degree. It is also prone to epidemic spreading (null epidemic threshold) [17–19].

While a number of recent works have concentrated on the properties of the power-law networks and how they are dynamically generated, another interesting problem is to find efficient algorithms for searching within these particular kinds of graphs [20,21]. In the most general distributed search context, one may have very little information about the location of the target. An interesting example is provided by the recent emergence of peer-to-peer networks, which have gained enormous popularity with users wanting to share their computer files. In such networks, the name of the target file may be known, but due to the network's *ad hoc* nature, the node holding the file is not known until a real-time search is performed. File-sharing systems that do not have a central server include GNUTELLA and FREENET. Files are found by forwarding queries to one's neighbors until the target is found. Recent measurements of GNUTELLA networks [22] and simulated FREENET networks [23] show that they have power-law degree distributions.

Search 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. However, the searcher does not presumedly know the shortest path to reach the target. The searcher even has no idea of the general topology of the network that will be searched. It is important to design appropriate search strategies in order to acquire high efficiency and, in the meantime, get an overall idea of the underlying network. Random walk becomes important in the extreme opposite case where only local connectivity is known at each node. It is theoretically interesting to probe how the structural heterogeneity affects the nature of the diffusive and relaxation dynamics of the random walk [20,24,25]. In return, random walk is also suggested to be a useful tool in studying the structure of networks.

Much is known about random walks on both regular and random networks [26,27]. Recently, there have been several studies of random walks on small-world networks (SWN) [28–30] as well as on the SF's [20,21]. In this work, we systemically carry out comparative studies of random walks for several typical complex networks. We suppose at every step, the walker only know neighbors of its present node. So if the target is at one neighbor of the present node where the walker stays, this round of search is over. The search strategies adopted by the walker include the following: random walk (RW), no-back (NB) walk, no-triangle-loop (NTL) walk, no-quadrangle-loop (NQL) walk, and self-avoiding (SA) random walk. For the RW, the walker may hop to a neighbor node by randomly taking one of the links with equal probability. It forgets all information about its past. The NB walk implies that a random walker, if possible, will not return to the node it was situated at the previous step. Similarly, NTL and NQL random walks mean that the walker will try to avoid walking in loops, with three or four edges, respectively, unless there is no other choice. We mention that the NQL walk also includes the NTL, which means it eliminates quadrangle loops as well as triangle loops. Finally, the SA random walk implies that the walker is the smarter. It tries to avoid revisiting the node that it has already visited in a run of the search. Surely, the SA walk also includes the NTL and NQL walks.

Let the walker start from one of the nodes and set in turn the other nodes as the target. For every pair of given nodes,

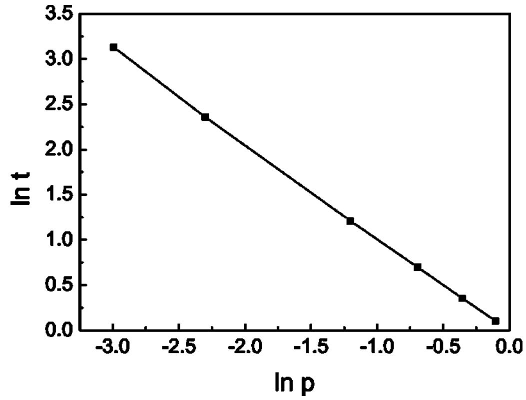

FIG. 1. Log-Log plot of average search time (walking steps) t versus link probability p of a random graph. The total number of nodes $N=1000$. The slope of the line is -1. The behavior of all the walk strategies is similar.

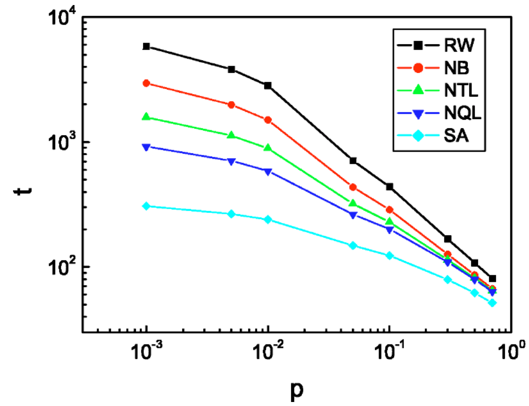

FIG. 2. (Color online) Dependence of average search times on the short-cut probability p on the WS model for RW, NB, NTL, NQL, and SA walks, respectively. The total number of nodes $N=300$.

we perform 200 runs of simulations and take average of the search times. It is found that the search time is only slightly dependent on the starting nodes, regardless of the large variance in node degree for some networks. The overall mean search time is again average over the whole network. It is found that search efficiency of each walk strategy varies widely with the topology of the underlying networks. In general, the self-avoiding random walk is the most efficient search strategy.

We also perform a preferentially self-avoiding (PSA) random walk on these networks, in which the walker is prone to a near neighbor with a higher degree. In this case, the walker must know the degree of its present node, as well as the degrees of its near neighbors. Contrary to one's intuition, it does not promote the efficiency of the search processes. In many cases, it lowers the efficiency and only greatly increases the computer running time.

The paper is organized as follows. In Sec. II, we study the search processes on random graph networks. Section III concerns SWN's and in Sec. IV, properties of scale-free networks are investigated. Summary is included in Sec. V.

II. RANDOM-GRAPH NETWORKS

We define a random graph as N labelled nodes and every pair of the nodes being connected with probability p. Consequently the total number of edges is a random variable with the expected value $pN(N-1)/2$. The degree k_i of a node i follows a binomial distribution with parameters $N-1$ and p:

$$P(k_i=k)=C_k^{N-1}p^k(1-p)^{N-1-k}. \qquad (1)$$

We perform random walks only on the largest cluster. The walker starts in turn from a node to reach a given node on the network. Average is taken over 200 runs. We note that the search time is only slightly dependent on the starting nodes. Figure 1 shows a log-log plot of average search times for various p, given a total number of nodes $N=1000$. It exhibits a power-law relation $t \propto p^{-\gamma}$, with exponent $\gamma=1$. In fact, we have performed all the above mentioned search strategies, and found the difference is rather small. It means that for the random graph network, clustering effect and short path effect are not obvious. As a whole, a random walk is also the optimal walk.

III. THE WATTS–STROGATZ MODEL

The small-world networks proposed by Watts–Strogatz (WS) are structures of much recent interest. It combines aspects from regular and completely random lattices. Such structure may be devised by adding, in a random way, links to an ordered lattice. A major feature is that even at a very low density of additional links, the chemical distance (minimal distance between two points) drastically decreases from its original value on the underlying regular lattice. Namely, it exhibits the small-world characteristics. The WS model possesses inherent loops and therefore displays other properties than the disordered treelike structure, namely, it has a large clustering coefficient.

We start from N sites on a ring, where each of the sites is connected by links to its four nearest and next-nearest neighbors. For all pairs of sites (i,k), we add a link with probability p. As in Sec. II, we perform random walks for various strategies on the WS network for $N=300$ and found great difference against the random graph network. Figure 2 shows that the random walk is now a bad search strategy. The no-back walk, no-triangle-loop and no-quadrangle-loop walks, as well as self-avoiding walk run more efficiently. The NTL and NQL walks eliminate repeating visits in loops for clustered nodes. Hence the improvement of search efficiency for NTL and NQL walks as to the RW implies that high clusters of nodes are popular in the underlying SWN's. On the other hand, the SA random walk can further largely decrease the search time, indicating that a small-world property of the underlying network has important effect. This can be understood by noting that the SA walker may effectively travel to remote nodes by taking the short cuts in the small-world network and hence avoid being trapped in a small region where it must repeatedly visit the same nodes because of no other choice. In other words, the short-cuts help the walker to speed up the iteration.

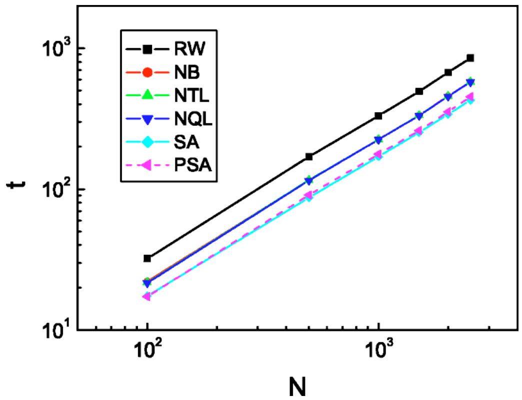

FIG. 3. (Color online) Average search times for the BA model. All the straight lines are parallel to each other. In particular, the lines for NB, NTL, and NQL collapse into one. The search time of the PSA walk (dashed line) is slightly higher than that of the SA walk.

IV. SCALE-FREE NETWORKS

A large number of real networks, including metabolic networks, the protein interaction network, the World Wide Web, and even some social networks, exhibit the scale-free topology, in which the vertex has a power-law degree distribution $p(k) \propto k^{-\gamma}$ typically with scaling exponent $2<\gamma<3$ [1,2]. With the pioneer work of Barabási and Albert (BA) [13], dozens of scale-free models have been constructed [31–36]. The fundamental ingredients in the BA model and its variants are the network growth and the preferential attachment: vertices are added one after another to the network, and edges are more prone to be connected to vertices with large k. Other models with nongrowing algorithm also exist [37,38]. Below we will focus on two typical SF networks: one is the original BA model; the other is a clustered SF model. We study the effects of clustering on the dynamical processes.

A. The Barabási–Albert model

The algorithm of the BA model is the following:

(1) Growth, starting with a small number m_0 of nodes, at every step, we add a new node with $m=2$ edges that link the new node to m different modes already present in the system.

(2) Preferential attachment, when choosing the nodes to which the new node connects, we assume that the probability Π that a new node will be connected to node i depends on the degree k_i of node i, such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{2}$$

The scale-free network generated in such a way has an exponent $\gamma=3$. It is well known that the BA model has small clustering coefficient C, which decreases with the network size, following approximately a power law $C \sim N^{-0.75}$.

Figure 3 is the average time for various search strategies on the BA models of different size N. It is notable that all the linear lines are almost parallel. In particular, the lines for NB, NTL, and NQL collapse into one, implying clustering effects are quite limited. On the other hand, the improvement of search efficiency for the SA random walk is related to the small-world feature of the SF network. In Fig. 3 we also plot the search time for a preferentially self-avoiding random walk (dashed line). For the PSA walk, we design that the hopping probability of the walker to a near neighbor is proportional to the degree of this neighbor. Hence a star node is in a favored position. However, one can see that this walk strategy does not work as good as the primitive self-avoiding random walk. At first sight, this result is somehow contrary to one's intuition. It is also at variance with a conclusion obtained in Ref. [20], which claims that the optimal search strategy is the PSA walk. For this point, we argue that although from star nodes a walker may conveniently get to more nodes, these star nodes are also more frequently revisited from other nodes. Hence PSA walk may double count the advantage of star nodes. In other words, the star nodes are too frequently visited. It becomes more difficult to reach a target on nodes of small degree. The same conclusion applies to other networks. It should be mentioned that Ref. [20] assumed that the walker knows not only the degree of the nearest neighbors, but also the information of the near neighbors' neighbors. This is a more strict condition than ours. We are not certain whether the discrepancy between us originates from this additional assumption.

To reveal the characteristics of the large connectivity heterogeneity of scale-free networks, we study the mean first-passage time (MFPT) $\langle T_j \rangle$ of a given node j with K_j near neighbors. 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

$$P_{ij}(t+1) = \sum_k \frac{A_{kj}}{K_k} P_{ik}(t), \tag{3}$$

where A_{kj} is the adjacency matrix element with $A_{kj}=1$ if there is a link between node j and node k, and $A_{kj}=0$ otherwise. According to Ref. [21], one can obtain that

$$\langle T_j \rangle \propto 1/K_j. \tag{4}$$

Namely, targets on nodes with larger degrees are more easily found out than on nodes with smaller degrees. At a mean field level, they are preferred in receiving information from the whole network. It is notable that MFPT is independent of the explicit topological structure of the underlying networks. In Fig. 4 we depict numerical simulations of the MFPT according to the degree distribution for a system of size $N=1000$. After average over the search times for a given degree k (the lower panel), the fitted line shows that it is indeed inversely proportional to the corresponding node degree.

We note that some nodes with a small number of neighbors also have small MFPT (Fig. 4, upper panel). These nodes play important roles in dynamical processes of information propagation. It is called the *random-walk betweenness* [21], which is not necessarily the same as the shortest-path betweenness extensively discussed in the literature.

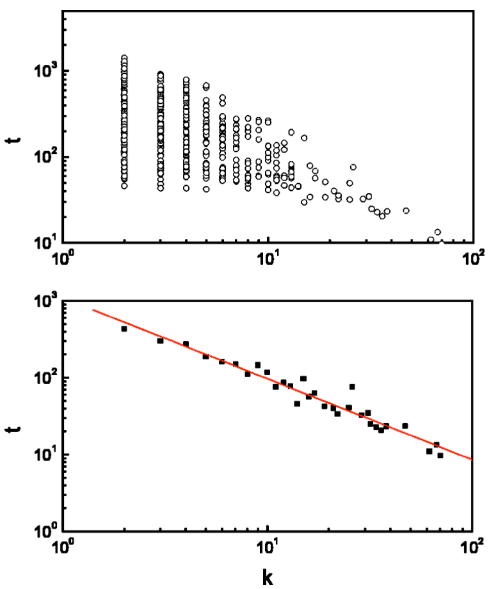

FIG. 4. (Color online) Upper panel, MFPT for the BA model for targets with large connectivity heterogeneity. Lower panel, average over the search times of given k's. A fitted linear relation is obtained with a slope of -1.

B. Clustered scale-free models

One major deficit of the BA model is the lack of high clustering coefficient, which is present in most practical networks. In particular, the clustering coefficient usually has a degree-dependent power-law form $C(k) \propto k^{-\gamma}$, with $\gamma \sim 1$ for most cases [10,34,39–41]. Lots of models have been proposed to account for this hierarchical feature [31–36]. As an example, we will consider the growing model (the deactivation model) introduced by Klemm and Eguíluz in which nodes are progressively deactivated with a probability inversely proportional to their connectivity [36]. Analytical arguments and numerical simulations have led to the claim that, under general conditions, the deactivated model allowing a core of m active nodes, generating a network with average degree $\langle k \rangle = 2m$ and degree probability distribution $P(k) = 2m^2 k^{-3}$. The scale-free properties are associated to a high clustering coefficient.

The model starts from a completely connected graph of $m=2$ active nodes and proceeds by adding new nodes one by one. Each time a node is added, (1) it is connected to all active nodes in the network; (2) one of the active nodes is selected and set inactive with probability

$$p_d(k_i) = \frac{\left[\sum_{j\in A}(a+k_j)^{-1}\right]^{-1}}{a+k_i}; \qquad (5)$$

and (3) the new node is set active. The sum in Eq. (5) runs over the set of active nodes A. $a=2$ is a model parameter.

Figure 5 is a plot as that in Fig. 3. In contrast to the BA model, one finds that NTL walk and NQL walk all improve the search efficiency. This is resulted from the large clustering phenomena of the deactivation model. As discussed in Sec. III, the further reduction of search time for the self-avoiding random walk reflects the small-world property of this model.

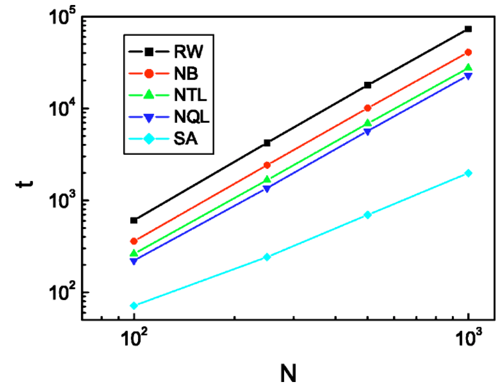

FIG. 5. (Color online) Average search times for the deactivation model with high clustering coefficient. The power-law relations inhibit in all search processes for RW, NB, NTL, NQL, and SA walks, respectively.

V. SUMMARY

We have played random walks on complex architectures from random-graph, small-world to scale-free networks. The walker may take different walking strategies to promote search efficiency. It is found that the self-avoiding random walk is the most efficient search strategy if the walker is not aware of the global structure of the underlying network. NTL and NQL walks can be adopted to probe the clustering phenomena of the networks by comparing with the results of RW and NB walks. This comparison is intriguing because NTL and NQL prevent the walker from lingering in a local part of the network if the nodes are highly clustered. A SA walker may rapidly get out of the possible trapping through the short cuts of the underlying network to remote parts. Hence the SA walk can be employed as a tool to probe the small-world property by checking if it can further reduce the average search time. In the framework of our study, where at most the degrees of near neighbors are known, the preferentially SA walk does not help to improve the search process further. The possible reason is discussed.

In conclusion, we find that dynamical processes on networks are greatly dependent on the topological features of the networks. In return, it is useful to explore the network topology by comparing various walking strategies. The explicit relations between various random walks and the topologies of complex networks deserve further studies.

ACKNOWLEDGMENT

This work was supported by a grant from Beijing Normal University.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[4] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[5] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[6] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[7] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[8] J. M. Montoya and R. V. Solé, J. Theor. Biol. **214**, 405 (2002).
[9] A. Wagner, Mol. Biol. Evol. **18**, 1283 (2001).
[10] H. Jeong, S. Mason, A. L. Barabási, and Z. N. Oltai, Nature (London) **411**, 41 (2001).
[11] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. Kepler, Adv. Complex Syst. **5**, 43 (2003).
[12] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[13] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[14] R. A. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[15] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[16] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[17] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[18] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[19] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **26**, 521 (2002).
[20] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys. Rev. E **64**, 046135 (2001).
[21] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[22] Clip2 Company, Gnutella, http://www.clip2.com/gnutella.html
[23] T. Hong, in *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*, edited by A. Oram (O'Reilley, Sebastopol, CA, 2001), Chap. 14, pp. 203–241.
[24] S. Jespersen, I. M. Sokolov, and A. Blumen, Phys. Rev. E **62**, 4405 (2000); B. Tadić, Eur. Phys. J. B **23**, 221 (2001).
[25] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[26] M. N. Barber and B. W. Ninham, *Random and Restricted Walks* (Gordon and Breach, New York, 1970).
[27] B. D. Hughes, *Random Walks and Random Environments* (Clarendon, Oxford, 1996), Vols. 1 and 2.
[28] J. Lahtinen, J. Kertész, and K. Kaski, Phys. Rev. E **64**, 057105 (2001).
[29] S. A. Pandit and R. E. Amritkar, Phys. Rev. E **63**, 041104 (2001).
[30] E. Almaas, R. V. Kulkarni, and D. Stroud, Phys. Rev. E **68**, 056105 (2003).
[31] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001); Phys. Rev. Lett. **86**, 5632 (2001).
[32] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[33] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **65**, 066122 (2002).
[34] E. Ravasz *et al.*, Science **297**, 1551 (2002).
[35] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[36] K. Klemm and V. M. Eguíluz, Phys. Rev. E **65**, 057102 (2002); **65**, 036123 (2002).
[37] G. Caldarelli, A. Capocci, P. DeLos Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[38] B. J. Kim, A. Trusina, P. Minnhagen, and K. Sneppen, AO/0403006 (unpublished).
[39] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[40] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[41] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A **311**, 590 (2002).