Reachability on scale-free networks

Yagnik Goswami and M. S. Santhanam
Indian Institute of Science Education and Research,
Dr. Homi Bhabha Road, Pune 411008, India.
(Dated: January 7, 2015)

Random walk on discrete lattices is a fundamental model in physics that forms the basis for our understanding of transport and diffusion processes in physical systems. In this work, we study unreachability in networks, i.e., the number of nodes not visited by any walkers until some finite time. We show that for the case of multiple random walkers on scale-free networks, the fraction of sites not visited is well approximated by a stretched exponential function. We also discuss some preliminary results for distinct sites visited on time-varying networks.

PACS numbers: 05.45.-a, 03.67.Mn, 05.45.Mt

I. INTRODUCTION

Random walk was introduced more than 100 years ago by Karl Pearson [1] and over the decades this idea has formed the basis for our understanding of transport and diffusion processes in physical systems [2]. Random walk model continues to be relevant today in many applications ranging from physics and computer sciences to biology [3]. In particular, random walks on disordered lattice such as complex networks [4] form the basis for modern applications such as web search engines [5] and recommender systems [6]. In addition, search and discovery on complex networks [7], information spreading on networks [8], clustering algorithms [9] and extreme events on networks [10] are other areas in which random walks play a crucial role.

One class of relevant application is in the design of algorithms for efficient search and discovery process on complex networks. What route should a packet of information on information networks travel from origin to its destination? How to quickly find new webpages on the world wide web? All these questions are similar in spirit to the general problem of search for information or other resources on complex networks. The process of discovery of resources through random walking on networks is useful as a benchmark to against which to measure the efficiency of other search and discover algorithms. In many practical networked systems several variants of random walks are used as algorithm to explore a network, TCP/IP being a specific example [11].

As a statistical physics problem, random walks on discrete lattices has been studied for several decades especially to understand first passage time distribution [12], distance covered by walkers, extreme events [10], distinct sites visited and site occupancy, flux-fluctuation relations [14] and node centrality in complex networks [8]. Most of these measures have direct bearing on the questions related to efficient search and information diffusion on networks. Many results for random walks on regular d-dimensional lattices are known [13], such as for instance mean number of distinct sites visited by W random walkers [15]. On regular d-dimensional lattices and for short times, the number of distinct sites visited $f(t) \propto t^d$ and asymptotically behaves $f(t) \propto W t$ for $d > 3$. In general, Ref. [15] identifies three distinct temporal regimes with different behaviours for $f_r(t)$.

On the contrary, if the underlying topology is not a regular lattice but instead forms a complex network, not many results are known for distinct sites visited by multiple random walkers. Distinct visited sites are those visited by at least one walker until time $t$. For the case of a single walker on scale-free networks, it has been numerically shown that the fraction of distinct sites visited in $t$ time steps is $f_r(t) = t$ and as $t \to \infty$, $f_r(t) \to 1$ due to finite size of network [16]. In a small world network, $f_r(t)$ displays a cross-over from $f_r(t) = \sqrt{t}$ to $f_r(t) = t$ behaviour depending on whether the walker has managed to hit a short-cut in the SW network or not [17]. Another related point of view is to consider the probability for the paths taken by the random walkers on networks and it has been shown that only three distinct distributions are possible [18].

Recently, exact distribution for the number of distinct sites visited by many random walkers in a one dimensional lattice has been obtained [20]. What happens in the case of multiple random walkers on a network? In this work, we study unreachability $f_{wr}(t)$, i.e., the fraction of sites not visited by any of the $W$ random walkers on scale-free network until time $t$. We show that this fraction $f_{wr}(t)$ can be modelled as a stretched exponential function. If $N_r$ is the number of nodes reached in time $t$, then $f_{wr}(t)$ is related to fraction of distinct nodes $f_r(t) = N_r/N$ visited by the random walkers in time $t$ through the relation $f_{wr}(t) = 1 - f_r(t)$.

Further, we discuss the results obtained in the case when the networks are rewired. It is known that many real-life networks are time-varying [19] and if the time scales over which network varies is comparable to the time scale of dynamics taking place on them then it becomes necessary to account for temporal variation of networks as well. We consider a case in which some specific percentage of nodes are rewired and we show that network reachability is enhanced by the process of rewiring the edges.
with different network sizes as zeroth node. This figure shows the result for six cases random walkers were placed on the same node designated. Let us see the distribution, many statistical quantities related to the random walkers in scale-free networks with degree distribution, we consider a scale-free network with \( N \) nodes and \( E \) edges on which \( W \) independent random walkers are executing random walk. This dynamics can be described by a master equation of the form,

\[
P_{ij}(n+1) = \sum_k A_{kj} P_{ik}(n)
\]

where \( P_{ij}(n) \) is the occupation probability for a walker starting \( i \)-th node at time \( n=0 \) and reaching \( j \)-th node at time \( n+1 \). In this, \( A_{ij} \) represents the elements of an adjacency matrix which carries the information about connectivity of the network, \( A_{ij} = 1 \) if \( i \)-th and \( j \)-th nodes are connected and 0 otherwise. Starting from this equation, many statistical quantities related to the random walk such as the first passage time can be calculated.

In Fig. 1, we show the fraction of sites not reached by random walkers in scale-free networks with degree distribution \( P(k) \sim k^{-\gamma}, \) where \( \gamma \approx 2.2 \). At time \( t=0 \), all the random walkers were placed on the same node designated as zeroth node. This figure shows the result for six cases with different network sizes \( N \) and number of walkers \( W \). For all the cases, we note that the scaled parameter of interest is \( x = Wt/N \) and \( f_{ur}(t) = 1 - N_r/N \). This choice of scaling the variables leads to a data collapse as shown in Fig. 1. The unreachability is well approximated by a stretched exponential function of the form,

\[
f_{ur}(t) = e^{-B x^\beta}
\]

where \( x = Wt/N \) and \( B \approx 0.75 \) and \( \beta \approx 0.8 \) and these numerical values have been obtained through a fit to the data shown in Fig. 1. As \( x << 1 \), we can expand the right hand side to obtain \( f_{ur}(t) \approx 1 - Bx^\beta \). Thus, the fraction of distinct sites visited is \( f_r(t) \approx x^\beta \). We note that in Ref. 16 it was argued based on numerical results that for the case of one walker the distinct sites reached is given by \( f_r(t) = t/N \) for \( t << N \), implying that the parameter \( \beta = 1 \). Physically, the stretched exponential function conveys that the distinct number of sites visited will equal the number of sites in a network only at infinite time. This is also in agreement with the asymptotic result that the first passage time distribution for a random walker displays an exponentially decaying tail [21]. In Fig. 2 the same results as in Fig. 1 are shown, but now as a function of \( x^\beta \). As expected, in the semi-log scale, we obtain a linearly decaying curve confirming the scaling observed for unreachability. Thus, unreachability is a finite time effect in connected networks since as \( t \to \infty \) all nodes are eventually reached.

II. RANDOM WALKS ON NETWORKS

Random walks on networks is a straightforward generalisation of one-dimensional lattice random walk to a network setting. Thus, a walker on \( i \)-th node with degree \( d_i \) at discrete time \( n \) can choose to hop to any of its neighbours with the probability \( 1/d_i \). In this work, we consider a scale-free network with \( N \) nodes and \( E \) edges on which \( W \) independent random walkers are executing random walk. This dynamics can be described by a master equation of the form,

\[
P_{ij}(n+1) = \sum_k A_{kj} P_{ik}(n)
\]

where \( P_{ij}(n) \) is the occupation probability for a walker starting \( i \)-th node at time \( n=0 \) and reaching \( j \)-th node at time \( n+1 \). In this, \( A_{ij} \) represents the elements of an adjacency matrix which carries the information about connectivity of the network, \( A_{ij} = 1 \) if \( i \)-th and \( j \)-th nodes are connected and 0 otherwise. Starting from this equation, many statistical quantities related to the random walk such as the first passage time can be calculated.

In Fig. 1 we show the fraction of sites not reached by random walkers in scale-free networks with degree distribution \( P(k) \sim k^{-\gamma}, \) where \( \gamma \approx 2.2 \). At time \( t=0 \), all the random walkers were placed on the same node designated as zeroth node. This figure shows the result for six cases with different network sizes \( N \) and number of walkers \( W \). For all the cases, we note that the scaled parameter of interest is \( x = Wt/N \) and \( f_{ur}(t) = 1 - N_r/N \). This choice of scaling the variables leads to a data collapse as shown in Fig. 1. The unreachability is well approximated by a stretched exponential function of the form,

\[
f_{ur}(t) = e^{-B x^\beta}
\]

where \( x = Wt/N \) and \( B \approx 0.75 \) and \( \beta \approx 0.8 \) and these numerical values have been obtained through a fit to the data shown in Fig. 1. As \( x << 1 \), we can expand the right hand side to obtain \( f_{ur}(t) \approx 1 - Bx^\beta \). Thus, the fraction of distinct sites visited is \( f_r(t) \approx x^\beta \). We note that in Ref. 16 it was argued based on numerical results that for the case of one walker the distinct sites reached is given by \( f_r(t) = t/N \) for \( t << N \), implying that the parameter \( \beta = 1 \). Physically, the stretched exponential function conveys that the distinct number of sites visited will equal the number of sites in a network only at infinite time. This is also in agreement with the asymptotic result that the first passage time distribution for a random walker displays an exponentially decaying tail [21]. In Fig. 2 the same results as in Fig. 1 are shown, but now as a function of \( x^\beta \). As expected, in the semi-log scale, we obtain a linearly decaying curve confirming the scaling observed for unreachability. Thus, unreachability is a finite time effect in connected networks since as \( t \to \infty \) all nodes are eventually reached.

III. SITES VISITED ON A REWIRED NETWORK

In this section, we present preliminary results related to the number of distinct sites visited on a periodically rewired network. This problem is relevant in the case of search on networks that are subject to rewiring of the edges. Consider the case of web graph whose nodes (web pages) get rewired almost all the time. In practice, it is useful to consider snapshots of a network like web graph at periodic intervals to monitor how random walk properties are modified by the changes in the network structure.

In the spirit of this argument, we consider the dynamics of \( W \) random walkers on a scale-free network and the rewiring algorithm proceeds as follows. Consider a fully connected scale-free network with \( N \) nodes.
We rewire $p$ percentage of its nodes after every $T$ time steps. For every $t = nT$, where $n > 0$ is an integer, we randomly choose two nodes, with probabilities $g_1(k)$ and $g_2(k)$ respectively, and exchange their edges. The probability of choosing a node with degree $k$ is taken to be $g_i(k) = k^{\alpha_i}$, $i = 1, 2$. In this, $\alpha_i = 0$ denotes random choice of nodes without regard to their degree. For $\alpha_i > 0$, hubs are preferentially chosen and for $\alpha_i < 0$ leaf nodes are preferentially chosen for rewiring.

In Figure 3 we show the result for the fraction of nodes reached in $t$ when the network is rewired after every $T$ time steps. At each of the rewiring instants, $p = 0.7\%$ of the nodes were rewired. The value of $\alpha_1 = 1$ and $\alpha_2 = -3$ implies that every rewiring involves exchange of edges between one hub and one leaf node. As the figure reveals, more rewiring leads to more reachability of the network in comparison with the case of static network without rewiring. As $T$ increases the time interval between two consecutive rewiring instances increases. Hence, as we would expect, reachability is less and tends to merge with the static network result.

In general, increasing the value of rewiring percentage $p$ also increases the reachability of the network in comparison with that of the static network. Indeed, even if $\alpha_1 = \alpha_2 = 0$ leading to random choice of nodes of rewiring, there is enhancement in the reachability of the network. This shown in Fig. 4 which reveals that periodic random rewiring also leads to enhanced reachability. However, if targetted rewiring is performed by selectively choosing hub and leaf, then reachability is even more enhanced compared to the random rewiring case.

Finally, we comment on the physical basis of the effect arising due to rewiring of the network. In a scale-free network there are few hubs but a large number of leaf nodes. Random walkers are most likely to be exploring the domain of large number of leaf hubs. However, when a rewiring is effected, two nodes are chosen and their edges are exchanged. If the nodes happen to be a hub and a leaf node, then random walkers on leaf node suddenly find themselves on a hub and can quickly explore territory previously unavailable when they were on a leaf node. As more rewiring takes place this effect leads to enhanced reachability on the network. When nodes are randomly chosen the enhancement is still seen but not as much as in the case of hub-leaf rewiring. This is because in a scale-free network most nodes being leaf nodes, any exchange of edges between them does not lead to enhanced reachability. For the same reason, rewiring is unlikely to lead to any enhancements in reachability for Erdos-Renyi networks [22].

In summary, we have studied the question of reachability of network nodes when multiple and independent random walkers are exploring a scale-free network. We show through numerical simulations that the unreachability of the network, defined as the fraction of sites not visited until time $t$, can be represented by a stretched exponential function with $x = Wt/N$ being the scaled parameter. As Fig. 1 shows, we obtain a good data collapse for various network sizes and numbers of walkers. Further, we also discuss some preliminary results on the effect of rewiring on the reachability of network nodes. We show that rewiring leads to enhancement in the number of distinct sites visited by random walkers.

[1] Karl Pearson, Nature 72, 294 (1905).
[2] Oliver Ibe, Elements of random walk and diffusion pro-
cesses, (wiley, 2013); George Weiss, Aspects and applications of the random walk, (North-Holland, 1994).
[3] Edward A Codling, Michael J Plank and Simon Benhamou, J. R. Soc. Interface 5, 813 (2008).
[4] S. Boccaletti et. al., Phys. Rep. 424, 175 (2006).
[5] Mark Levene, An Introduction to Search Engines and Web Navigation, (Wiley, 2010).
[6] L. Lu et. al., Phys. Rep. 519, 1 (2012).
[7] O. Bnichou, C. Loverdo, M. Moreau, and R. Voituriez, Rev. Mod. Phys. 83, 81 (2011).
[8] J. D. Noh and H. Reiger, Phys. Rev. Lett. 92, 118701 (2004); Ernesto Estrada, Naomichi Hatano, Michele Benzi, Phys. Rep. 514, 89 (2012).
[9] S. A. Tabrizi et. al., Physica A 392, 5772 (2013).
[10] V. Kishore, M. S. Santhanam and R. E. Amritkar, Phys. Rev. Lett. X XXX (2011); D. De Martino et. al., Phys. Rev. E 79, 015101(R) (2009); P. Echenique et. al., EPL 71, 325 (2005).
[11] D. Methi and K. Ramasamy, Network Routing : Algorithms, Protocols and Architectures, (Morgan Kaufmann, 2007).
[12] Y. Lin and Z. Zhang, Sci. Rep. 4 5365 (2014); S. Redner, A Guide to First-Passage Processes, (Cambridge University Press, 2007)
[13] Barry D. Hughes, Random walks and random environments, (Oxford Univ Press, 1995).
[14] M. Argollo de Menezes and A.L. Barabsi, Phys. Rev. Lett. 92, 028701 (2004);
[15] H. Larraalde et. al., Phys. Rev. A 45, 7128 (1992).
[16] S. Lee, S-H Yook and Y. Kim, Physica A 387, 3033 (2008).
[17] E. Almaas, R. V. Kulkarni and D. Stroud, Phys. Rev. E 68, 056105 (2003).
[18] T. J. Perkins, E. Foxall, L. Glass and R. Edwards, Nature Comm. 5, 5121 (2014).
[19] Peter Holme and Jari Saramaki, Phys. Rep. 519, 97 (2012).
[20] A. Kundu, S.N. Majumdar, and G. Schehr, Phys. Rev. Lett. 110, 220602 (2013).
[21] H. W. Lau and K. Y. Szeto, EPL 90 40005 (2010).
[22] Mark Newman, Networks: An Introduction, (Oxford Univ Press, 2010).