Physica A 389 (2010) 587–594

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Mixing search on complex networks

Cun-Lai Pu *, Wen-Jiang Pei

Information Science and Engineering College, Southeast University, China

ARTICLE INFO

Article history:
Received 26 June 2009
Received in revised form 30 September 2009
Available online 9 October 2009

PACS:
89.75.Hc
02.50.-r
05.40.Fb

Keywords:
Random walk
Complex networks
First passage time
Search

ABSTRACT

In this article, we derive the first passage time (FPT) distribution and the mean first passage time (MFPT) of random walks from multiple sources on networks. On the basis of analysis and simulation, we find that the MFPT drops substantially when particle number increases at the first stage, and converges to the shortest distance between the sources and the destination when particle number tends to infinite. Given the fact that a Brownian particle from a high-degree node often needs a large number of steps to reach an expected low-degree node, which is the bottleneck for a single random walk, we propose a mixing search model to improve the efficiency of search processes by using random walks from multiple sources to continue the searches from high-degree nodes to destinations. We compare our model with the mixing navigation model proposed by Zhou on complex networks and find that our model converges much faster with lower hardware cost than Zhou's model. Moreover, simulations on scale-free networks show that the search efficiency of our model is much higher than that of a single random walk, and comparable to that of multiple random walks which have much higher hardware cost than our model. Finally, we discuss the traffic cost of our model, and propose an absorption strategy for our model to recover the additional walkers in networks. Simulations indicate that this strategy reduces the traffic cost of our model effectively.

1. Introduction

In recent years, network structure and dynamics have attracted tremendous interest from physicists [1,2], mainly because they want to understand and characterize the underlying mechanisms of complex networks such as the Internet, social networks, and biological networks. One of the most important dynamics taking place on networks is search or navigation [3–14]. The studies of network search can be traced back to the famous Milgram's small-world experiment. The results of this experiment revealed that not only do short paths exist between any pair of nodes, but such paths can easily be found using only local information [2]. However, Kleinberg [3] pointed out that the efficient navigability is a fundamental property of only some small-world structures. In large-scale networks, there are many paths between any pair of nodes. Whether a node can find the efficient or even the shortest paths to any other node depends on the network topology, knowledge of the network topology acquired by the nodes and the search algorithm employed. Without any knowledge on topology, a single random walk search [4] was proved to be inefficient. The search efficiency can be improved by using some local information, such as the geographical location of target [3], the degrees of neighboring nodes [5,6], and local betweenness centrality (LBC) [7]. In the extreme case, if all nodes know how to deliver the message along the shortest paths, the highest efficiency can be achieved. To achieve this, all the nodes in networks should be expensive routers which are able to calculate the best paths after handling a large amount of external real-time information [8]. In addition, the routers require a lot of storage capacity to store next hops of the best paths [9]. Apparently, it is uneconomical and impractical, especially in huge-size ad hoc networks. So, it is of great significance and challenge to design efficient network search algorithms.

* Corresponding address: Information Science and Engineering College, Southeast University, 210096 Nan-Jing, Jiang-Su, China.
E-mail address: pucunlai@yahoo.cn (C.-L. Pu).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.10.007

Random walks of Brownian particles are often used to describe search processes on networks. Traditionally, much attention has been focused on parameters such as access time or hitting time, commute time, cover time and so on [15,16]. Recent studies show that random walks on complex networks can reveal a variety of characteristics of the underlying networks, such as the diameter [28], centrality [18], community structure [17], etc. Random walk has also been exploited to solve search problem on networks, such as sending message and locating target node [19], tracking moving object [20], building dynamic routing on networks [21]. Besides, the statistical problems of a single random walker have been generated into multiple random walks gradually [22,23], although some may be very difficult. In this article, we first derive the distribution of the first passage time (FPT) and the mean first passage time (MFPT) of random walks from multiple sources on networks. After this, we propose a mixing search (MS) model based on random walks from multiple sources and implement our model on complex networks of different topologies. To describe the behavior of our model, we measure the average search time of the whole network and compare our model with the mixing navigation (MN) model proposed by Zhou [24]. Results demonstrate that our model gets higher efficiency with lower hardware cost on complex networks. We also compare our model with single and multiple random walks on scale-free networks. Simulations confirm that the search efficiency of our model is much higher than that of a single random walk and comparable to that of multiple random walks. However, the hardware cost of our model is much lower than that of multiple random walks. Finally, we present an absorption strategy for our model to handle the additional walkers in networks, which reduces the traffic cost of our model effectively.

The outline of the article is as follows. In Section 2, we derive the distribution of FPT and MFPT of random walks from multiple sources on networks. In Section 3, we present the MS model based on the study of part two and implement it on complex networks of different topological characteristics. We also give the absorption strategy in this part. In Section 4, we present our conclusions.

2. Random walks from multiple sources

The random walk of a Brownian particle is a first-order Markov process. On an arbitrary finite network which is connected, supposing a particle start from node s to node d in t steps, the transition probability $P_{sd}(t)$ is as follows [4]:

$$P_{sd}(t) = \sum_{j_1,\ldots,j_{t-1}} \frac{A_{sj_1}}{K_s} \cdot \frac{A_{j_1j_2}}{K_{j_1}} \cdots \frac{A_{j_{t-1}d}}{K_{j_{t-1}}} \tag{1}$$

where A_{ij} are the entries of the adjacent matrix A, and $A_{ij} = 1$ if there is a link from i to j, otherwise, $A_{ij} = 0$. K_i is the degree of node i. The probability transfer matrix of the whole network can be described as follows:

$$P = K^{-1}A \tag{2}$$

where K is a diagonal matrix and $K = \mathrm{diag}(K_1, K_2, \ldots, K_N)$. Then, we suppose there are many particles distributed on k different nodes at the initial time. The k nodes are labeled 1 through k, $0 < k < N$. The number of particles on node i is n_i. At $t = 0$, all the particles start out walking randomly to the destination labeled d, $k < d \leq N$. After some time T_d, the process ends when the first particle arrive at d. Obviously, T_d is the FPT which is a random variable. In order to calculate the distribution of T_d, we need to use two other random variables: X_{id} and $Y_{id}(n_i)$.X_{id} represents the FPT of a single particle from i to d.$Y_{id}(n_i)$ is the passage time of the first arriving particle at d, which is one of the n_i particles starting from i simultaneously and independently at the initial time. In fact, the FPT is equal to the time taken by the first arriving particle during the road, no matter how the initial distribution of particles is. Assume all the particles walk independently, we can obtain that:

$$Y_{id}(n_i) = \min\{X_{id}^1, X_{id}^2, \ldots, X_{id}^{n_i}\} \tag{3}$$

where X_{id}^m is the mth particle starting from i to d. Similarly, T_d can be calculated as follows:

$$\begin{aligned} T_d &= \min\{Y_{1d}(n_1), Y_{2d}(n_2), \ldots, Y_{kd}(n_k)\} \\ &= \min\{X_{1d}^1, X_{1d}^2, \ldots, X_{1d}^{n_1}, X_{2d}^1, X_{2d}^2, \ldots, X_{2d}^{n_2}, \ldots, X_{kd}^1, X_{kd}^2, \ldots, X_{kd}^{n_k}\}. \end{aligned} \tag{4}$$

Then, we can obtain that:

$$\begin{aligned} Prob\{T_d > t\} &= Prob\{\min\{X_{1d}^1, X_{1d}^2, \ldots, X_{1d}^{n_1}, \ldots, X_{kd}^1, X_{kd}^2, \ldots, X_{kd}^{n_k}\} > t\} \\ &= \prod_{i=1}^{n_1} Prob\{X_{1d} > t\} \prod_{i=1}^{n_2} Prob\{X_{2d} > t\} \cdots \prod_{i=1}^{n_k} Prob\{X_{kd} > t\}. \end{aligned} \tag{5}$$

Due to the same independent walks of particles, Eq. (5) can be simplified as follows:

$$Prob\{T_d > t\} = \prod_{i=1}^{k} (Prob\{X_{id} > t\})^{n_i}. \tag{6}$$

Since the process ends up when the first arriving particle appears on d, we can treat d as an absorbing state. The new transfer matrix is as follows:

$$Q(d) = (P_1, \ldots, P_{d-1}, 0, P_{d+1} \ldots, P_N) \tag{7}$$