Title	Estimating mean first passage time of biased random walks with short relaxation time on complex networks
Author(s)	Lee, Zhuo Qi; Hsu, Wen-Jing; Lin, Miao
Citation	Lee, Z. Q., Hsu, W.-J., & Lin, M. (2014). Estimating Mean First Passage Time of Biased Random Walks with Short Relaxation Time on Complex Networks. PLoS ONE, 9(4), e93348-.
Date	2014
URL	http://hdl.handle.net/10220/19549
Rights	

Estimating Mean First Passage Time of Biased Random Walks with Short Relaxation Time on Complex Networks

Zhuo Qi Lee, Wen-Jing Hsu, Miao Lin*

School of Computer Engineering, Nanyang Technological University, Singapore

Abstract

Biased random walk has been studied extensively over the past decade especially in the transport and communication networks communities. The mean first passage time (MFPT) of a biased random walk is an important performance indicator in those domains. While the fundamental matrix approach gives precise solution to MFPT, the computation is expensive and the solution lacks interpretability. Other approaches based on the Mean Field Theory relate MFPT to the node degree alone. However, nodes with the same degree may have very different local weight distribution, which may result in vastly different MFPT. We derive an approximate bound to the MFPT of biased random walk with short relaxation time on complex network where the biases are controlled by arbitrarily assigned node weights. We show that the MFPT of a node in this general case is closely related to not only its node degree, but also its local weight distribution. The MFPTs obtained from computer simulations also agree with the new theoretical analysis. Our result enables fast estimation of MFPT, which is useful especially to differentiate between nodes that have very different local node weight distribution even though they share the same node degrees.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zqlee1@e.ntu.edu.sg

Introduction

Scale-free node degree distribution, small network diameter, large clustering coefficients – these are common properties found to be present in complex networks arising from seemingly disparate fields such as biology, computer science, cosmology, etc. [1,2]. It is widely believed that there should be common underlying principles behind the formation of these networks that resulted in the observed properties. As such, complex networks have received much research attention during the past decade.

One of the studies pertaining to complex networks is the network efficiency and capacity analysis [3–7]. In these studies, answers to the questions such as 'how fast can a message be delivered to a given destination' and 'how many packets may be generated in the system before congestion arises' are essential in understanding the performance of a network [5,7,8]. While there are existing results on routing and flow balancing in networks with certain topologies [9], they usually assumed the underlying topology or the knowledge of link formation mechanism. There are cases where the assumptions do not hold, for example, the animal foraging strategy [10] and the web searching process as depicted by the PageRank algorithm [11]. Approaches based on random walk can be applied when the detailed information of network formation mechanism is absent. Quantities such as *stationary distribution* and *mean first passage time* (MFPT) are important as they can be used to answer the questions about the performance of networks as mentioned above.

The concept of random walk has also been applied in social networks. Even-Dar et al. [12] studied the process of spreading influences in social networks by means of Voter Model and

showed that the pathways in which the influences propagate are equivalent to series of random walks. Thus, the MFPT to a node A, yields the expected time for the other nodes in the network to be influenced by node A. Selecting a node with low MFPT for spreading the news could result in fast information propagation.

In fact, MFPT can be calculated by using the fundamental matrix method [13]. However, the computation involves multiple matrix multiplications. When the method is applied on large scale networks with millions of nodes, the computation becomes practically infeasible. Moreover, the solutions obtained from the fundamental matrix approach are too generic and hard to interpret. For instance, it is unclear which factors, be it node degree, eigenvalue, local connectivity, or others, govern the MFPT by just looking at the solution expression. Further research is needed to better characterize MFPT and to reduce the computational cost.

In [10], Condamin et al. showed a mean first passage time analysis using the pseudo Green function. They related MFPT to the network size and diameter. The general applicability of their result to non-fractal networks has been discussed in [14]. Fronczak et. al [8] applied the mean field theory to study the MFPT based on the Erdos-Renyi (ER) random graphs and networks generated by using the Barabasi-Albert (BA) preferential attachment model. Lau et. al [14] showed asymptotic analysis of the first passage time of unbiased random walk for a class of networks with *short relaxation time* by using the mean field theory. However, the solution relates MFPT to the node degree alone. There are cases where the nodes share the same node degree while having vastly different local topology such as those depicted in Figure 1.
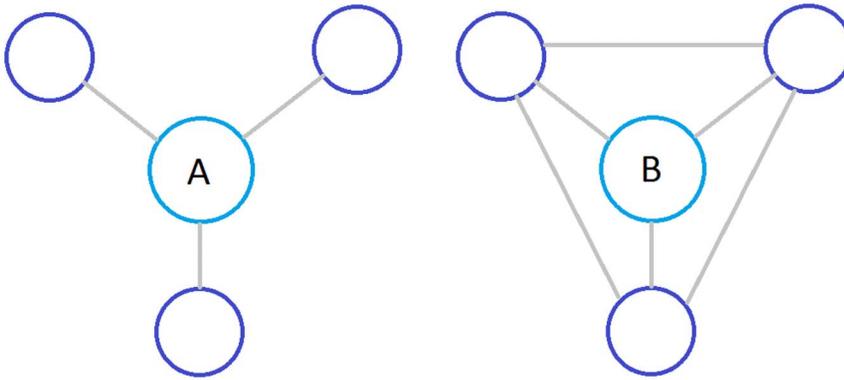
Figure 1. Nodes with same degree may have very different local connectivity. The figure shows two examples where a node with degree 3 may be part of a sparsely-connected star network or a densely-connected clique.
doi:10.1371/journal.pone.0093348.g001

Random walks with short relaxation time are also known as random walks with *non-compact exploration* [15–17]. In [15], Bénichou et al. presented the conditions for which a random walk on fractal network falls into the compact or non-compact exploration regime. In [16], Hwang et al. presented a fundamental result on the relationship between the node degree and the MFPT of heterogeneous networks, i.e., networks with power-law degree distribution. They focused on uniform random walks without self-loops. It was shown that the general trend of MFPT with respect to the node degree exhibit cross-over behaviour that is governed by the *spectral dimension d_s* and the exponent of the degree distribution γ. For the case $d_s < 2$, which is known as the compact exploration regime (see also [15]), the MFPT was found to be independent of node degree. For $d_s > 2$, the MFPT is related to the node degree via a power-law function where the exponent is determined by d_s and γ.

While the focus of [16] is to show the general trend of MFPT with regard to node degree and spectral dimension, we focus on explaining the differences in MFPT for nodes with same degree when the underlying random walk falls into the non-compact exploration regime. We generalize the FPT analysis as discussed in [14] for a class of random walks with short relaxation time where the nodes have arbitrary weights. In addition, we analyse the first passage time at an improved level of precision by incorporating exact solutions to the stationary distribution. This enables a more detailed view of the neighbourhood around the target node. The new expression allows us to differentiate between the case where the neighbourhood of a node is sparsely connect and the case where it is densely connected as depicted in Figure 1. We will show that the MFPT of a node is closely related to the local weight distribution around it. By changing the node weights, the spectral dimension and hence the relaxation time will be affected, we have also conducted simulations to study how the spectral dimension is affected by varying the weight assignments. They will be discussed in more detail in Discussions section.

Analysis

Outline

In this section, we will show the detailed analysis for approximating the FPT decay rate and MFPT of random walks with short relaxation time. First, we will show the stationary distribution for random walk with arbitrary node weight assignment scheme. Then, we apply a flow-based heuristics to estimate the quasi-stationary distribution of the random walk when

a sink node is introduced. Next, we show that the FPT distribution follows an exponential decay, and further show that the decay exponent is related to the quasi-stationary distribution of the neighbours of the target node. Finally, we obtain the MFPT by approximating the integration over the FPT distribution.

Stationary and Quasi-stationary Distribution

A complex network is modelled by a connected and undirected network $G = (V, E, W)$, where V denotes the set of nodes with $|V| = N$, E denotes the set of edges (i,j), and W denotes the weights assigned to the nodes. An edge (i,j) represents the existence of relationship between nodes i and j. For the sake of simplifying expression, we assume without loss of generality that self-loops are present for all nodes, i.e., $(i,i) \in E$ for $i \in [1, N]$. We define a biased random walk on G with arbitrary positive weights assigned on nodes. A transition from node i to a neighbouring node j is based on the following transition rule:

$$Pr(i \rightarrow j) = \frac{w_j}{\sum_{h \in n(i)} w_h} \tag{1}$$

where w_i denotes the weight of node i, and $n(i) = \{h | (i,h) \in E\}$ denotes the neighbourhood of node i. $Pr(i \rightarrow j)$ denotes the probability of a random walker at node i moving to node j at the next time step. When self-loop is present, the staying probability of a node is not a constant but is dependent on the local weight distribution. Let the transition matrix be denoted by $P = [Pr(i \rightarrow j)]$, and $P_i(t)$ denote the probability of the random walker appearing at node i exactly at timestep t. The master equation is given by:

$$P_i(t+1) = \sum_{j \in n(i)} P_j(t) Pr(j \rightarrow i) \tag{2}$$

Recall that a Markov Chain (MC) is said to be *regular* if $\exists q \in \mathbb{Z}^+$, $\forall (i,j) P^{(q)}(i \rightarrow j) > 0$ where $P^{(q)}(i \rightarrow j)$ denotes the (i,j) entry of the q-th order transition matrix P^q. The probability distribution of a regular MC will converge to a unique stationary distribution π regardless of the starting position as $t \rightarrow \infty$. The *relaxation time τ* of a Markov Chain is the time for the state probability distribution to be close to the stationary distribution, i.e., the standard deviation of P^τ is bounded by $1/e$ where e denotes the Euler Number (see
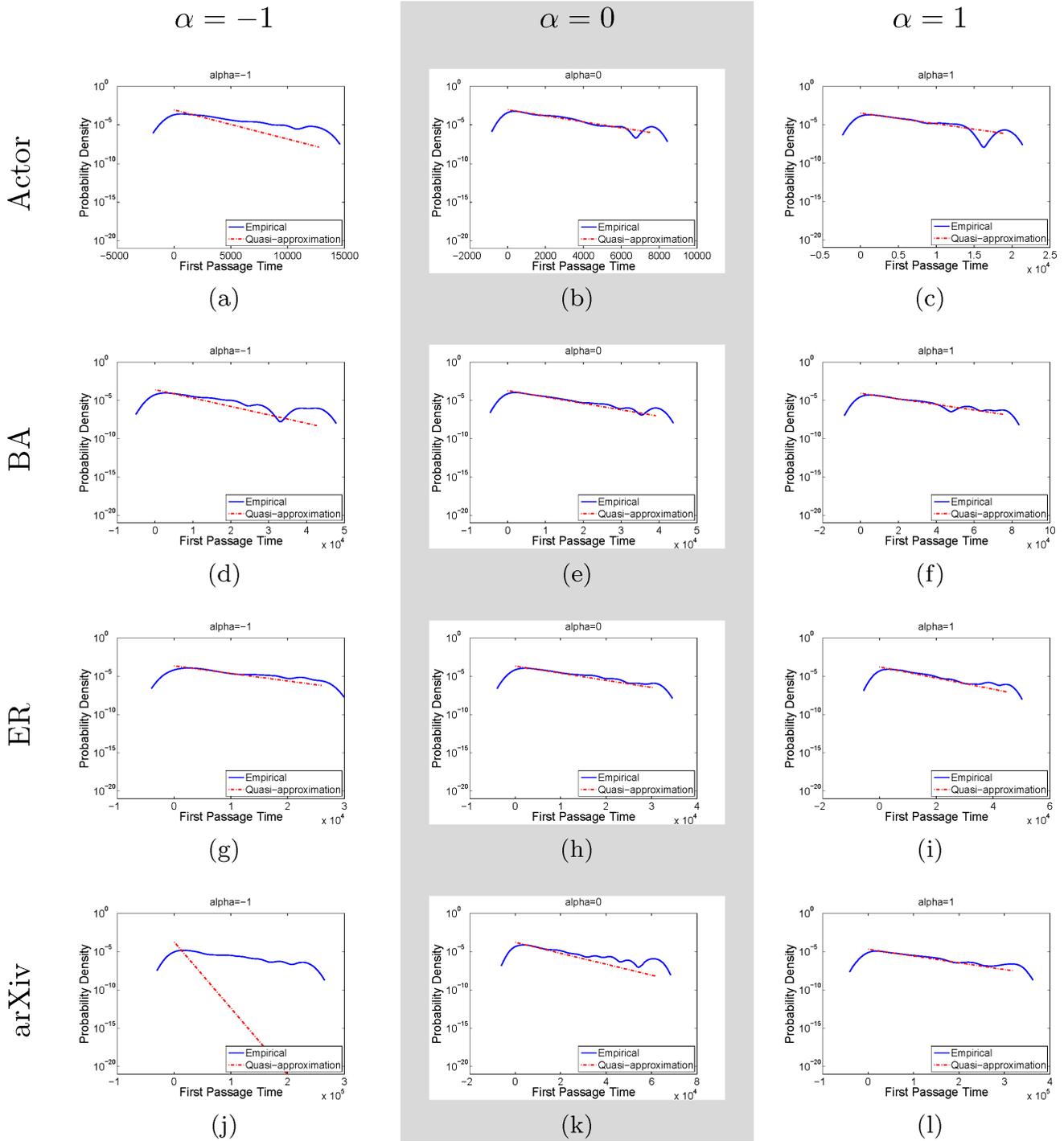
Figure 2. Plots of empirical first passage time distribution against theoretical prediction according to the approximate bound given by Ineq.(13) for different networks and weighting factors. Each row corresponds to a network in the following order: Actor, BA, ER, and arXiv. The columns, from left to right, correspond to $\alpha = -1$, 0, 1 respectively. For most cases, the tail of the first passage time distribution can be predicted fairly accurately except for Figure 2(j), which is due to the high relaxation time as shown in Table 2.
doi:10.1371/journal.pone.0093348.g002

[18] Chapter 12). The relaxation time of a Markov Chain is given by $\frac{1}{1-\lambda_2}$, where λ_2 is the second largest eigenvalue [18].

In the networks that we consider, since every node has a self-loop and the graph is strongly connected, the corresponding Markov Chain is regular. In the following, we show the exact form

of stationary distribution for the biased random walk as defined in Eq.(1).

Let w_i be the weight of node i, w'_i be the neighbourhood weight of node i given by $w'_i = w_i \sum_{j \in n(i)} w_j$, and $\langle w' \rangle = \frac{1}{N} \sum_i w'_i$. The stationary distribution of the biased random walk is given by

Table 1. Summary of networks studied.

Network	# nodes	# edges	CC	Dm	Source
arXiv	4158	15501	0.5569	17	[21]
BA	4158	31136	0.0174	5	[1]
ER	4158	21020	0.0022	7	[24]
Actor	968	13324	0.6751	9	[1]

C denotes the average clustering coefficients and Dm denotes the network diameter.
doi:10.1371/journal.pone.0093348.t001

$$\pi_i = \lim_{t \to \infty} P^t(i) = \frac{w_i \sum_{j \in n(i)} w_j}{\sum_x w_x \sum_{y \in n(x)} w_y} = \frac{w_i'}{N \cdot \langle w' \rangle} \tag{3}$$

The expression can be verified by applying the equilibrium condition on the master equation as given by Eq. (2).

Let v_d denote the sink node and let G_{v_d} denote the resultant graph from designating node v_d as sink node. Let $\Phi(t)$ denote the corresponding random walk probability distribution and $\phi_v(t)$ represent the probability for a random walker to be present at node v at time t. The following set of rules describes the new random walk on G_{v_d}:

$$\begin{cases} \phi_v(0) = \delta_{v_s} & \text{(initial condition)} \\ \phi_{v_d}(t) = 0 & \text{(sink node)} \\ \phi_v(t) = \sum_{u \in n(v)} \phi_u(t-1) \frac{w_v}{\sum_{r \in n(v)} w_r} & \text{(transition rule)} \end{cases} \tag{4}$$

where v_s denotes the source node and $\delta_{v_s} = 1$ for v_s, and 0 otherwise. For connected network where v_d is reachable by every node, $\phi_v(t)$ tends to zero when t tends to infinity. However, for random walks with short relaxation time [14,19], the conditional probability distribution $P_i(t|T > t)$ (i.e. conditioned on the survival of the random walker) will converge as $t \gg \tau$, where T denotes the time to reach the sink node. As such, the converged conditional probability distribution is called the *quasi-stationary distribution* denoted as Ψ_v. For $t \gg \tau$, we have the following approximation:

$$\Psi_v \approx \frac{\phi_v(t)}{\phi_{total}(t)} \tag{5}$$

where $\phi_{total}(t) = \sum_{v \in V} \phi_v(t)$ denotes the total survival probability at time step t.

We approximate the quasi-stationary distribution of G_{v_d} around the sink node with a flow-based heuristic [14] described in the following. We begin with the stationary distribution π in G. Under the stationary distribution, based on Eq.(3), the probability of the random walker traversing an edge (u,v) is $P_\pi(u,v) = \frac{w_u w_v}{N \langle w' \rangle}$. Next, we treat an undirected edge as a combination of in-link and out-link. We remove the out-links from v_d and hence making v_d a sink node. Thus, the equilibrium will be broken and nodes $u \in n(v_d)$ will have their 'flow' constantly drawn by v_d. Finally, we approximate the quasi-stationary distribution of such nodes by discounting the probability of utilizing the edge (u,v_d):

$$\Psi_u \approx (1 - \frac{P_\pi(u,v_d)}{\pi_u})\pi_u = (1 - Pr(u \to v_d))\pi_u \tag{6}$$

Asymptotic First Passage Time Analysis

To obtain FPT, we re-designate the destination node v_d as a sink node such that the random walk process terminates once the random walker moves into the sink node. The time taken for it to be absorbed into the sink node is then the same as the FPT. Let $F(v_d,t|v_s)$ denote the probability of visiting node v_d at timestep t for the first time with v_s as the starting node. Then by definition, we have

$$F(v_d,t|v_s) = \phi_{total}(t-1) - \phi_{total}(t) \tag{7}$$

i.e. the first passage probability at time t is given by the difference in the total survival probability between time t and $t-1$.

On the other hand, we can also obtain the first passage probability by using the transition rule.

Table 2. Pearson's correlation coefficient for MFPT.

Network	Correlation			Relaxation time τ		
	$\alpha = -1$	0	1	$\alpha = -1$	0	1
arXiv	0.2619	0.7435	**0.9269**	555.56	555.56	1666.7
BA	**0.9241**	**0.9812**	**0.9940**	3.5311	2.2060	2.0178
ER	0.4860	**0.9846**	**0.9973**	16.502	3.7327	2.8417
Actor	0.0064	**0.9231**	**0.9851**	476.19	61.728	26.041

Overall, the correlations are high whenever the relaxation time is low. For the BA network, the relaxation time is consistently low and thus the correlation is extremely good.
doi:10.1371/journal.pone.0093348.t002

Figure 3. Comparison of MFPTs obtained from simulation against that obtained from using the bound shown in Ineq. (16). The column on the left compares the empirical MFPT to the results obtained by (i) using the bound in Ineq. (16); (ii) using the result presented in [8]. The column on the right shows the scatter plot and correlation between the empirical results and our proposed theoretical results. The rows correspond to the Actor, arXiv, and BA network respectively. While the result presented by Fronczak et al. [8] gives the general trend of MFPTs with respect to node degree, we find that the MFPTs for a given node degree are distributed across a wide range and cannot be fitted with a function of the node degree alone. The scatter plots show a strong correspondence between the empirical results and our proposed theoretical results. This is further supported by the high Pearson correlation coefficients which are shown on top of the scatter plots.
doi:10.1371/journal.pone.0093348.g003

Table 3. Spectral dimension d_s w.r.t. weighting factor α.

		d_s		
α	Actor	arXiv	BA	ER
−50	0.042	0.016	0.026	0.110
−25	0.068	0.028	0.142	0.173
−5	0.360	0.300	3.760	0.990
−4	0.476	0.360	4.500	1.537
−3	0.590	1.735	5.066	2.220
−2	2.000	2.561	∞	3.526
−1	1.932	3.440	∞	∞
0	1.500	2.505	∞	∞
1	1.456	1.060	∞	∞
2	1.400	0.667	∞	∞
3	1.025	0.420	∞	∞
4	0.920	0.229	∞	∞
5	0.880	0.220	∞	4.378
25	0.180	0.295	∞	0.536
−50	0.107	0.090	∞	0.220

The data is obtained by conducting simulation and fitting the exponent of RTO probability according to the definition given in Eq.(17). We mainly consider the range [−5,5] as they are mostly considered in the literature. The entries {−50, −25, 25, 50} are used to examine the effect on d_s for large α. The entries with $d_s = \infty$ are obtained by the observation that the RTO probability stabilized fairly quickly in less than 10 time steps. The data in this table is plotted in Figure 4.
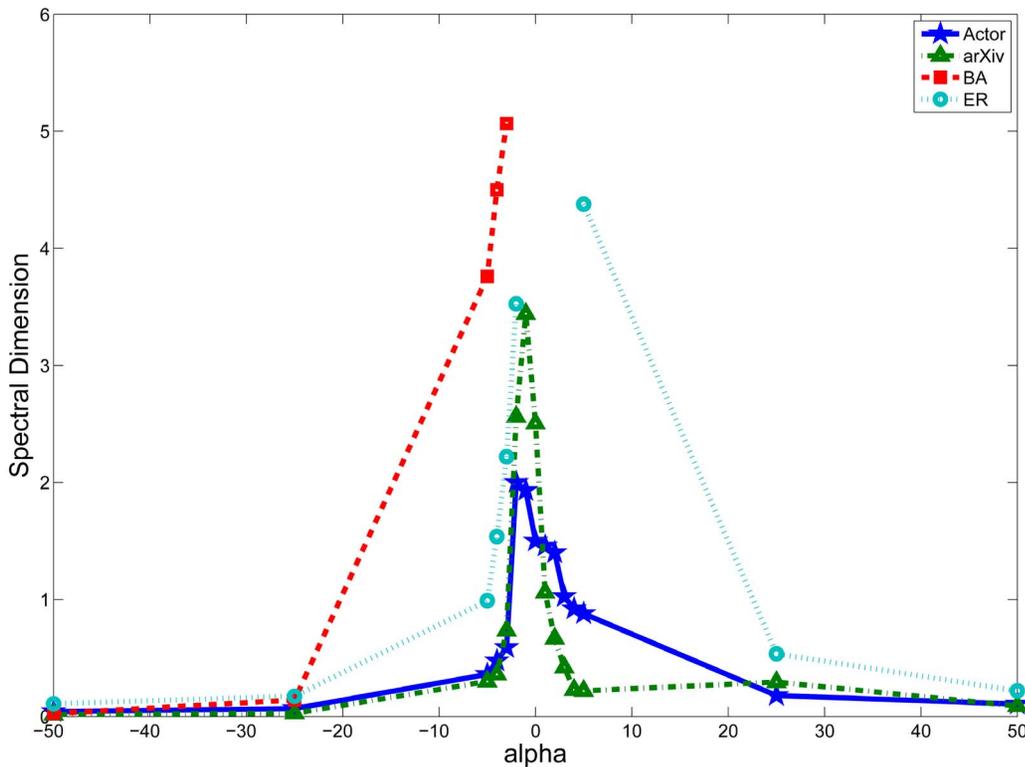doi:10.1371/journal.pone.0093348.t003

Figure 4. Relationship between spectral dimension d_s and the weighting factor α. The data used for drawing this figure is tabulated in Table 3. The lines corresponding to BA and ER network appear disconnected as they have $d_s = \infty$ for certain values of α and cannot be adequately shown in the figure. The spectral dimension generally peaks in the interval [−1, 1] and drops significantly for α of greater magnitude. This is especially true for the BA and ER networks (from infinity to a finite value). In the extreme case, by setting $\alpha = \infty$, the 'random walk' is no longer random as the node with largest degree will always be chosen at every step. Similar reasoning also applies for $\alpha = -\infty$. Therefore, towards both extremes, we would expect the random walk to become more localized and hence falls into the compact exploration regime.
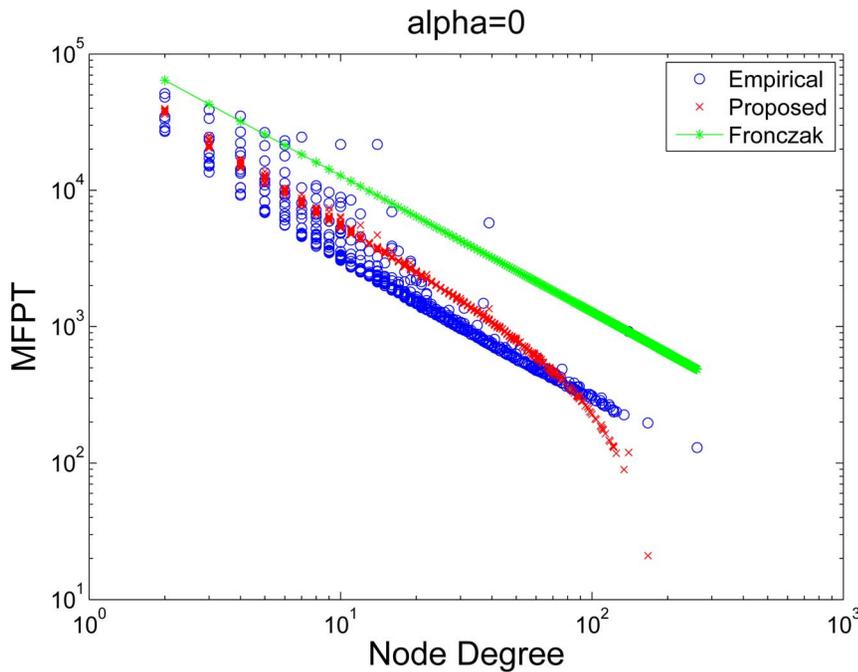doi:10.1371/journal.pone.0093348.g004

Figure 5. MFPT of the Actor network when $\alpha = 0$. Even though the spectral dimension d_s is less than 2, the MFPT is not found to be independent of node degree. Instead, the MFPT exhibits a power-law relationship with respect to the node degree. The disparity arises probably as a result of self-loops, which affects both the RTO probability distribution and the estimated d_s.
doi:10.1371/journal.pone.0093348.g005

$$F(v_d, t | v_s) = \sum_{u \in n(v_d)} \phi_u(t-1) \frac{w_{v_d}}{\sum_{r \in n(u)} w_r} \qquad (8)$$

For $t < \tau$, the first passage probability is dependent on the source-sink distance. Here, we focus on the asymptotic behaviour of the random walker for $t \gg \tau$. As such, we can apply Approx.(5) and after simplifying, we get:

$$F(v_d, t | v_s) \approx \phi_{total}(t-1) \sum_{u \in n(v_d)} \Psi_u \frac{w_{v_d}}{\sum_{r \in n(u)} w_r} \qquad (9)$$

Let $\beta_{v_d} = \sum_{u \in n(v_d)} \Psi_u \frac{w_{v_d}}{\sum_{r \in n(u)} w_r}$. By combining Eq.(7) and Approx.(9), after simplification, we obtain the following recursive approximation:

$$\phi_{total}(t) \approx (1 - \beta_{v_d}) \phi_{total}(t-1) \qquad (10)$$

By definition, $\phi_{total}(0) = 1$ as the random walker has not started moving and thus the survival probability is 1. Solving the recursive formula yields:

$$\phi_{total}(t) \approx (1 - \beta_{v_d})^t \qquad (11)$$

$$\leq \exp(-\beta_{v_d} t) \qquad (12)$$

Substituting the result back to Approx.(9) yields:

$$F(v_d, t | v_s) \gtrsim \beta_{v_d} e^{-\beta_{v_d}(t-1)} \qquad (13)$$

Thus, the first passage probability is approximately bounded below by an exponential function for $t \gg \tau$ and the decay rate is given by β_{v_d}. To calculate the decay rate, we apply Approx.(6):

$$
\begin{aligned}
\beta_{v_d} &= \sum_{u \in n(v_d)} \Psi_u \frac{w_{v_d}}{\sum_{r \in n(u)} w_r} \\
&\approx \frac{w_{v_d}}{N \langle w' \rangle} \sum_{u \in n(v_d)} \left(1 - \frac{w_{v_d}}{\sum_{r \in n(u)} w_r} \right) w_u \\
&= \frac{w_{v_d}}{N \langle w' \rangle} \left(\sum_{u \in n(v_d)} w_u - w_{v_d} \sum_{u \in n(v_d)} \frac{w_u}{\sum_{r \in n(u)} w_r} \right) \qquad (14) \\
&= \frac{w'_{v_d}}{N \langle w' \rangle} \left(1 - \frac{w_{v_d} \sum_{u \in n(v_d)} Pr(u \to u)}{\sum_{u \in n(v_d)} w_u} \right) \\
&= \pi_{v_d} \left(1 - Pr(v_d \to v_d) \sum_{u \in n(v_d)} Pr(u \to u) \right)
\end{aligned}
$$

where the last step is just a simplification of the expression by using the definition as shown in Eq.(1) as we assumed earlier that every node in the network has a self-loop. Nevertheless, for the cases where self-loop is absent, the expression $Pr(v \to v)$ can be

substituted for $\frac{w_v}{\sum_{u \in n(v)} w_u}$, as self-loop played no role in our derivation except for the regularity of the random walk.

The result in Approx.(14) suggests that the decay rate of the first passage time distribution depends mainly on the sink node's stationary distribution and the transition probabilities around the sink node.

Mean First Passage Time

After obtaining the first passage time distribution, we can estimate the mean first passage time by using the approximation:

$$\langle T \rangle = \sum_1^\infty t F(v_d, t | v_s)$$

$$\approx \int_1^\infty t F(v_d, t | v_s) dt \qquad (15)$$

$$= \int_1^c t F(v_d, t | v_s) dt + \int_c^\infty t F(v_d, t | v_s) dt$$

for some $c \gg \tau$. In Approx.(15), the integral is separated into two parts – from 1 to c and from c to ∞ for short-term and long-term behaviour respectively. While the first passage time analysis is based on the assumption $t \gg \tau$, our result is an upper bound on the first passage probability for small t. This is justified by the observation that the first passage time probability starts from zero initially, increases to a peak, then decreases exponentially (see Figure 2 and [15] for non-compact case) with the decay rate as presented previously. Thus, Ineq.(13) is an overestimation of the first passage probability for small t. Hence, we provide a lower bound of mean first passage time as follows:

$$\langle T \rangle > \int_1^\infty t \beta_{v_d} e^{-\beta_{v_d}(t-1)} dt$$

$$= \int_1^\infty (1 - t\beta_{v_d}) e^{-\beta_{v_d}(t-1)} dt + \int_1^\infty e^{-\beta_{v_d}(t-1)} dt \qquad (16)$$

$$= 1 + \frac{1}{\beta_{v_d}}$$

Although Ineq.(16) is a lower bound of the mean first passage time, results from simulations show that it is strongly correlated to the actual MFPT.

In order to calculate the MFPT, an edge will be followed exactly twice to obtain the weight of the connected neighbour. Thus, assuming the network is represented in adjacency list format, the time complexity of the approximation is given by $O(N+M)$. The time complexity to convert a network from adjacency matrix representation into adjacency list representation is bounded by $O(N^2)$. Hence, the overall time complexity of the approach is bounded by $O(N^2+M)$ for adjacency matrix representation. Nevertheless, in terms of computational efficiency, the approach is a great improvement from the fundamental matrix method [13] which has time complexity of $O(N^3)$ or $O(N^{2.376})$ depending on the actual implementation of matrix operations [20]. Thus, our approach can be used to estimate MFPTs of the network quickly especially for large scale networks.

Results

We verify the theoretical results on two real world networks and two artificially generated networks: (i)arXiv General Relativity and

Quantum Cosmology collaboration network, obtained from the Stanford SNAP website [21], (ii) Barabási-Albert (BA) preferential attachment network [1,22], (iii) Erdos-Renyi (ER) random graph, and (iv) Actor collaboration network from the Barabási lab [1]. Because the analysis only applies to connected graphs, we used the largest component of the arXiv network. The two generated networks are chosen to test the theoretical results on networks with different structure and edge density. The ER network is generated with $p=0.0011$, while the BA network is generated with $m=N_0=7$. The Actor network is constructed from the first 1000 records of the database, where each record consists of the actors who collaborated in the same movie. Table 1 summarizes the networks that we have examined.

We apply the node weight assignment scheme given by $w_i = k_i^\alpha$, where k_i is degree of node i and α is an integer. This node weight assignment scheme is mainly studied in the network traffic community such as [5,7]. For our experiments, we mostly consider the range of α in $[-1,1]$ except for the Actor network as other values with greater magnitude will result in random walks with exceedingly long walk lengths.

The experiments are conducted as described below. Firstly, for a given network, we choose 20 nodes randomly as the source nodes. The weighting factor α is then fixed and the node weights are computed accordingly. For each source node, 250 times of full random walk simulations are conducted independently. A full random walk simulation starts with the random walker at the source node and terminates when every other node has been visited at least once. The first passage time to each node is recorded. The procedure is then repeated for other values of α.

To obtain the first passage time distribution, we applied Gaussian Kernel Density Estimation on the first passage time statistics collected from the simulations. We randomly selected different source-sink pairs for each network and different values of α, and the results are plotted in Figure 2. As shown in the figure, for most cases, we can predict the tail of the first passage time distribution fairly accurately except for the case when $\alpha = -1$ in arXiv network. With reference to Table 2, we find that for $\alpha = -1$, the relaxation time for the corresponding random walk is very high, and thus for certain sink nodes with high absorption rates, our theoretical result may not be applicable.

In Figure 3, we compare the empirical MFPT to that predicted by Ineq. (16). Since we relaxed several expressions during the derivation, the values predicted by Ineq. (16) may not be of the same scale as the empirical result. Therefore, we renormalized the predicted values *(P-set)* with respect to the empirical result *(E-set)*. The renormalization scheme is described as follows. First, we sort both the P-set and E-set in non-increasing order. Then we rescale the middle 90% of the P-set with respect to that of the E-set, i.e. obtain the rescale parameters (shearing and scaling) by ignoring both the upper and lower 5 percentile of both sets. Finally, the renormalization is applied to the whole P-set.

As shown in Figure 3, we can observe that while the result by Fronczak et al. predicted the general trend of MFPT with respect to node degree, our result further refined the predicted values by examining local weight distribution. By zooming into a greater level of detail, our approach has revealed a useful relationship between local connectivity and the MFPT especially as highlighted in Figure 3(a),(c), and (e). The weighting factor α controls whether high degree nodes should receive greater attention or vice-versa. Thus, for small α, high degree nodes should be reached less often and hence greater MFPT. Surprisingly, even for $\alpha = -2$, there are cases where high degree nodes can be reached fairly quickly as depicted in Figure 3(a). By examining the structure of the Actor network, we find that many high degree nodes are connected with

bridge nodes of low degree. On the other hand, the MFPT of the BA network does not show this kind of pattern. We believe that this is due to the fact that the BA network possesses a much simpler structure than the Actor network.

To investigate the overall performance of the bound given in Ineq. (16), we also calculated Pearson's correlation coefficient, and the results are summarized in Table 2. Overall, the correlations are high whenever the relaxation time is low. For the BA network, the relaxation time is consistently low and thus the correlation is extremely good.

Discussions

In [15,16], it was found that the random walk exploration can be divided into compact and non-compact regimes based on their spectral dimension. For the compact case, the random walker spends longer time travelling around certain neighbourhood and hence the relaxation time is long, while the non-compact case is the other way round. In our study, we focus on the non-compact case, i.e., random walk with short relaxation time. While the structure plays a part in deciding the dynamics of random walk, the weighting factor α also affects the mixing rates. To examine how much the assumption of short relaxation time holds when the weighting factor α is changed, we also investigated the relationship between α and the spectral dimension of the corresponding random walk d_s. We numerically estimate d_s by using the following relation [17]:

$$P_o(t) \sim t^{-d_s/2} \tag{17}$$

where $P_o(t)$ denotes the Return-to-origin (RTO) probability at time t [17]. To obtain the RTO probabilities, we conducted simulations with 100,000 random walkers which are placed randomly at $t=0$. Each random walker walks independently for 100,000 time steps, and the fraction of random walker returning to their respective starting node is recorded at each time step. Table 3 and Figure 4 summarizes the result.

As shown in Figure 4, the spectral dimension generally peaks in the interval $[-1, 1]$, which explains the applicability of our result. While for the case of uniform random walk, it was shown that d_s of both BA networks and random graphs are infinity [23], we found finite d_s for $\alpha \neq 0$. In the extreme case, by setting $\alpha = -\infty$, the 'random walk' is no longer random as the node with smallest degree will always be chosen at every step. For this scenario, the network will be broken into cycles where leaf nodes form the smallest cycle. Similar phenomenon also applies for $\alpha = \infty$. Therefore, towards both extremes, we would expect the random walk to become more localized and hence falls into the compact exploration regime where our result may not be applicable.

We also observe several disparities between our results and that of [16]. For instance, as shown in Figure 5 and Table 3, when $\alpha = 0$, we found that the spectral dimension of the Actor network is 1.500. However, the MFPT is found to be following a power-law relationship with respect to node degree instead of being independent of node degree. Similar results have been obtained for the arXiv network for $\alpha = 1$, for which $d_s = 1.060$, as shown in Figure 3(c). We believe that the disparities arise from the following facts: (i) we considered random walks with self-loops where the staying probabilities are proportional to the node weights; and (ii) the random walks are biased by weight assignments. The self-loop changes the RTO probability distribution and thus also affects the estimated d_s. The nodes are weighted differently, therefore the cross-over threshold for d_s may not be the same as that of [16]. Further research is needed to better understand the effects of self-loops and node weights on the spectral dimension of a random walk.

In summary, we have shown the exact form of stationary distribution for a class of biased random walks on networks where the nodes are assigned arbitrary weights. By using this result, we have presented a new method that gives improved estimation of MFPT for random walks with short relaxation time. We have verified that the decay rate of the first passage time distribution can be estimated fairly accurately and the MFPTs are found to be better revealed by local weight distributions. Given its low computational cost, our method enables quick inspection of the MFPT for large scale networks. This is especially true for cases where the ranking rather than the actual values of MFPT of nodes is more important. For instance, to contain virus outbreaks, the new method can be used to quickly rank the nodes based on estimated MFPT and judiciously apply security measures on the nodes that are ranked highly. Our result can also be readily extended to the case of cyclic search [8] where the random walker will scan the direct neighbours of current node for the target as opposed to blindly following the transition rule.

Acknowledgments

Author Contributions

Conceived and designed the experiments: ZQL WJH ML. Performed the experiments: ZQL. Analyzed the data: ZQL. Contributed reagents/materials/analysis tools: ZQL. Wrote the paper: ZQL WJH.

References

1. Barabsi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.
2. Krioukov D, Kitsak M, Sinkovits R, Rideout D, Meyer D, et al. (2012) Network Cosmology. Nature Scientific Reports 2.
3. de Martino D, Dall'Asta L, Bianconi G, Marsili M (2009) Congestion phenomena on complex networks. Physical Review E 79: 015101.
4. Yan G, Zhou T, Hu B, Fu ZQ, Wang BH (2006) Efficient routing on complex networks. Physical Review E 73: 046108.
5. Ling X, Hu MB, Jiang R, Wu QS (2010) Global dynamic routing for scale-free networks. Phys Rev E 81: 016113.
6. Chen S, Huang W, Cattani C, Altieri G (2012) Traffic Dynamics on Complex Networks: A Survey. Mathematical Problems in Engineering.
7. Dan Wang ZL (2012) General local routing on complex networks. In: Communications and Infor- mation Processing.
8. Fronczak A, Fronczak P (2009) Biased random walks in complex networks: The role of local navigation rules. Physical Review E 80: 016107.
9. Kleinberg J (2006) Complex networks and decentralized search algorithms. In: In Proceedings of the International Congress of Mathematicians (ICM).
10. Condamin S, Bénichou O, Tejedor V, Voituriez R, Klafter J (2007) First-passage times in complex scale-invariant media. Nature 450: 77–80.
11. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web. Technical Report 1999–66, Stanford InfoLab.
12. Even-Dar E, Shapira A (2007) A note on maximizing the spread of influence in social networks. In: Proceedings of the 3rd international conference on Internet and network economics. Berlin, Heidelberg: Springer-Verlag, WINE'07, 281–286.
13. Grinstead CM, Snell JL (2003) Introduction to Probability. AMS.
14. Lau HW, Szeto KY (2010) Asymptotic analysis of first passage time in complex networks. EPL (Europhysics Letters) 90: 40005.
15. Bénichou O, Chevalier C, Klafter J, Meyer B, Voituriez R (2010) Geometry-controlled kinetics. Nature Chemistry 2: 472–477.

16. Hwang S, Lee DS, Kahng B (2012) First passage time for random walks in heterogeneous networks. Phys Rev Lett 109: 088701.

17. Hwang S, Yun CK, Lee DS, Kahng B, Kim D (2010) Spectral dimensions of hierarchical scale-free networks with weighted shortcuts. Phys Rev E 82: 056110.

18. David A Levin YP, Wilmer EL (2009) Markov Chains and Mixing Times. AMS.

19. Erik van Doorn PP (2011) Quasi-stationary distributions. Memorandum, University of Twente, AE Enschede, The Netherlands. Available at http://doc.utwente.nl/77521/1/memo1945.pdf. Accessed 2014 March 10.

20. Coppersmith D, Winograd S (1990) Matrix multiplication via arithmetic progressions. Journal of Symbolic Computation 9: 251–280.

21. Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. ACM Trans Knowl Discov Data 1.

22. Barabási AL, Albert R, Jeong H (1999) Mean-field theory for scale-free random networks. Physica A 272: 173–187.

23. Samukhin AN, Dorogovtsev SN, Mendes JFF (2008) Laplacian spectra of, and random walks on, complex networks: Are scale-free architectures really important? Phys Rev E 77: 036115.

24. Gilbert EN (1959) Random graphs. Annals of Mathematical Statistics 30: 1141–1144.