Network exploration using true self-avoiding walks

Yup Kim,* Seokjong Park, and Soon-Hyung Yook†

Department of Physics and Research Institute for Basic Sciences, Kyung Hee University, Seoul 130-701, Korea

(Received 27 February 2016; published 14 October 2016)

We study the mean first passage time (MFPT) of true self-avoiding walks (TSAWs) on various networks as a measure of searching efficiency. From the numerical analysis, we find that the MFPT of TSAWs, τ^{TSAW}, approaches the theoretical minimum $\tau^{\mathrm{th}}/N = \frac{1}{2}$ on synthetic networks whose degree-degree correlations are positive. On the other hand, for biased random walks (BRWs) we find that the MFPT, τ^{BRW}, depends on the parameter α, which controls the degree-dependent bias. More importantly, we find that the minimum MFPT of BRWs, $\tau^{\mathrm{BRW}}_{\mathrm{min}}$, always satisfies the inequality $\tau^{\mathrm{BRW}}_{\mathrm{min}} > \tau^{\mathrm{TSAW}}$ on any synthetic network. The inequality is also satisfied on various real networks. From these results, we show that the TSAW is one of the most efficient models, whose efficiency approaches the theoretical limit in network explorations.

DOI: 10.1103/PhysRevE.94.042309

I. INTRODUCTION

Due to its simplicity and theoretical importance, the random walk (RW) and its variants on a network have been intensively investigated to uncover various topological properties of networks as well as the dynamical properties of RWs on networks themselves [1–9]. They also play an important role as an efficient tool for sampling or exploring the network when only the local information is available [10–12]. It has been shown that the sampling efficiency can be drastically enhanced if the walks avoid multiple visits of nodes [11,12].

In addition, due to the explosive growth of human activities through communication networks, finding the efficient searching algorithm on a network such as the Internet and mobile communication networks is a very important problem in practice to send a message to the destination in a short time. Among the many dynamical processes, RW-based processes provide very simple, efficient, and widely applicable methods in information searching and routing on a network when there is no complete knowledge about the network or when the topology of the network is frequently changed [13–16]. Like the network sampling, the suppression of revisitation enhances the searching efficiency. In the normal RW, the probability of finding a walker at a node of degree k is proportional to k [4]. This increases revisitations of high-degree nodes and decreases the searching efficiency. To suppress such revisitation of high-degree nodes, biased random walk (BRW) models were suggested as alternative strategies for information search on a network with only local information [14,16].

In this sense, the self-avoiding walk (SAW) [17] is more efficient than the RW and BRW for exploring the network because it avoids the nodes already visited in a more stringent way. Even though SAWs are more efficient than RWs and BRWs, there is an intrinsic disadvantage in simply applying SAWs to information search on a network, because in SAWs the walker can be trapped if all the connected neighbors have already been visited. The typical length scale of SAWs for trapping is characterized by the attrition length. The average attrition length, $\langle L\rangle$, of SAWs on complex networks with N nodes is known to scale as $\langle L\rangle \sim N^{\delta}$ with $\delta < 1$ [18,19].

Recently, a modified SAW model in which SAWs are combined with normal RWs was introduced for direct application to the searching problem on a network [20]. However, if we coarse-grain the network with a typical length $\langle L\rangle$, then each node in the coarse-grained network corresponds to a subnetwork composed of $\langle L\rangle \sim N^{\delta}$ nodes. Due to the sublinear scaling of $\langle L\rangle$ the coarse-grained network is also infinite when $N \to \infty$, and the walker takes RW to move one node to the other on the coarse-grained network. Thus, the enhancement in the searching efficiency of the modified SAWs originates mainly from the finite-size effect.

In this paper, we use the true SAWs (TSAWs) [17,21] as the strategy for information search on complex networks to truly suppress the revisitation of nodes and compare the results with those of BRWs. For the quantitative analysis of the searching efficiency, we measure the mean first-passage time (MFPT) and show that the MFPT of TSAWs approaches the *theoretical limit*. Furthermore, TSAWs are non-Markovian processes. Thus, the history of walks significantly affects the searching efficiency. Especially, how many hubs are visited in the past and how often the walker passes these hubs are very crucial in exploring a network. Thus finding the effect of degree-degree correlation is another important quest in understanding the efficiency of searching strategies.

The paper is organized as follows. In Sec. II, we define the TSAW and BRW as well as the underlying synthetic networks. Section III reports the numerical results. The summary and discussion are provided in Sec. IV.

II. MODEL

A. True self-avoiding walk

The TSAW is defined as a stochastic process in which the probability that a walker hops to the next node is proportional to a negative exponential of the number of visitations. TSAWs on a network with N nodes are implemented as follows. Initially a node i is randomly selected and a walker is placed at i. At each time step t, if there are nearest neighbors of the current position which have not yet been visited, then the walker hops to a node chosen randomly from the neighbors

*ykim@khu.ac.kr
†syook@khu.ac.kr

not yet visited. If all neighbors have already been visited, then we uniformly choose a node among those neighboring nodes which have been visited least often in the past [17,21].

B. Biased random walk

In BRWs on a network the hopping probability from a node i to one of the connected nodes j is defined as [22,23]

$$\pi_{ij} = \frac{k_j^{\alpha}}{\sum_{\ell=1}^{k_i} k_{\ell}^{\alpha}}. \tag{1}$$

Here k_j is the degree of node j and $\sum_{\ell=1}^{k_i}$ represents the sum over the connected neighbors of node i. The exponent α is the control parameter. $\alpha = 0$ corresponds to the normal RWs. $\alpha > 0\,(< 0)$ means that the walker prefers to move to the nodes of a high (low) degree.

C. Synthetic networks and degree-degree correlation

In order to study the effect of the underlying topology on the searching efficiency, we consider three synthetic network models: the Erdös-Rényi (ER) model for random networks [24], the Barabási and Albert (BA) model [25], and the configuration model (CM) [26]. The degree distribution $P(k)$ of the ER model is binomial distribution. In the limit $N \to \infty$, $P(k)$ of the ER model becomes a Poisson distribution, $P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$, when the network is sparse [24]. Here $\langle k \rangle$ is the average degree. The BA model is characterized by the power-law degree distribution, i.e., $P(k) \sim k^{-\gamma}$ with $\gamma = 3$. The network whose degree distribution satisfies the power law is called a scale-free (SF) network. $P(k)$ of the CM can be any distribution, but we use $P(k) \sim k^{-\gamma}$ to generate a SF network with various values of γ.

The degree-degree correlation of a network with M edges is generally measured by the Pearson coefficient [27], defined as

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \tag{2}$$

where j_i and k_i are the degrees of the nodes at the ends of the ith edge, with $i = 1,\ldots,M$. If $r > 0$ ($r < 0$), then the network is said to be assortative (disassortative) and the network is neutral when $r = 0$. For the systematic generation of the correlated networks, we use the edge exchange method [28].

After applying the edge exchange method, we extract only the largest connected component (LCC) and check $P(k)$, r, and the size of the LCC, N. If N, $P(k)$, and r of the LCC are identical with the preassigned conditions, we use the LCC as the underlying topology for the simulation. The obtained networks belong to some peculiar subset of network ensembles generated by the ER model, BA model, or CM.

III. MEAN FIRST-PASSAGE TIME

Let τ_{ij} be the time at which the walker starting at node i visits node j for the first time. Then the MFPT, τ, is defined as the average of τ_{ij} over all possible pairs of nodes:

$$\tau \equiv \frac{1}{N(N-1)} \sum_{i,j} \tau_{ij}. \tag{3}$$

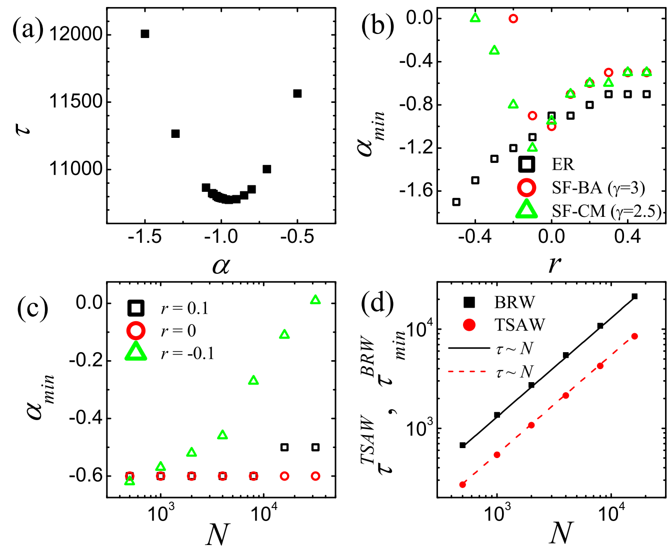

FIG. 1. (a) Plot of τ vs α for BRWs on SF networks generated by CM with $N = 10^4$, $\gamma = 2.5$, $r = 0$, and $\langle k \rangle = 47$. (b) Plot of the obtained $\alpha_{\min}$ for various values of r in three types of networks. $\langle k \rangle = 40$ for ER and BA networks and $\langle k \rangle = 46.5$ for the CM with $\gamma = 2.5$. (c) Plot of $\alpha_{\min}$ of BRW vs N for CM with $\gamma = 2.5$ and $\langle k \rangle = 10$. (d) $\tau_{\min}^{\rm BRW}$ and $\tau^{\rm TSAW}$ vs N on the CM with $r = 0.1, \gamma = 2.5$, and $\langle k \rangle = 10$. Solid and dashed lines in (d) represent the relation $\tau \sim N$.

To measure τ_{ij} we place the walker on a node i at $t = 0$. At each time step, the walker takes a walk according to the model defined in Sec. II until all nodes are visited. If a node j is visited for the first time at t, then we set $\tau_{ij} = t$. Repeat the procedure for all i (=1, 2,..., N). Since the TSAW and BRW are stochastic processes, τ_{ij} are averaged over 50 realizations of the walks for each i. We calculate τ using Eq. (3) for single-network generation and average τ over 100 different networks.

As shown in Fig. 1(a), the τ of BRWs depends on α and is a convex function of α. In fact, for all the considered networks, the τ of BRWs is always a convex function of α as in Ref. [16] (which are not shown).

Thus, there is a value $\alpha_{\min}$ at which the τ of BRWs becomes the minimum $\tau_{\min}^{\rm BRW}$. For example, we obtain $\tau_{\min}^{\rm BRW} \simeq 1.08N$ ($N = 10^4$) at $\alpha_{\min} \simeq -0.95$ on SF networks generated by the CM with $\gamma = 2.5$ and $r = 0$ [see Fig. 1(a)]. We obtain $\alpha_{\min}$ for other values of r and N through the measurement of the τ of BRW for various values of α as in Fig. 1(a). The results are shown in Fig. 1(b). The obtained $\alpha_{\min}$ is quite close to the theoretical expectation $\alpha_{\min} = -1$ for $r = 0$ but deviates significantly from $\alpha_{\min} = -1$ when $r \neq 0$, which agrees with the results in Ref. [16]. This implies that $\tau_{\min}^{\rm BRW}$ is affected by the degree-degree correlation because τ depends on α [29]. Thus the degree-degree correlation is a crucial factor in determining the searching efficiency. Furthermore, $\alpha_{\min}$ also depends on N if $r \neq 0$ [see Fig. 1(c)]. In Fig. 1(d), we display $\tau^{\rm TSAW}$ and $\tau_{\min}^{\rm BRW}$ versus N in the CM with $\gamma = 2.5$ and $r = 0.1$ (>0). The data in Fig. 1(d) show that the inequality $\tau_{\min}^{\rm BRW} > \tau^{\rm TSAW}$ holds for any N. The same inequality is found for other values of γ, $\langle k \rangle$, and r. Since $\tau^{\rm BRW} - \tau^{\rm TSAW}$ becomes much larger than $\tau_{\min}^{\rm BRW} - \tau^{\rm TSAW}$ when α deviates from $\alpha_{\min}$, to reach the best

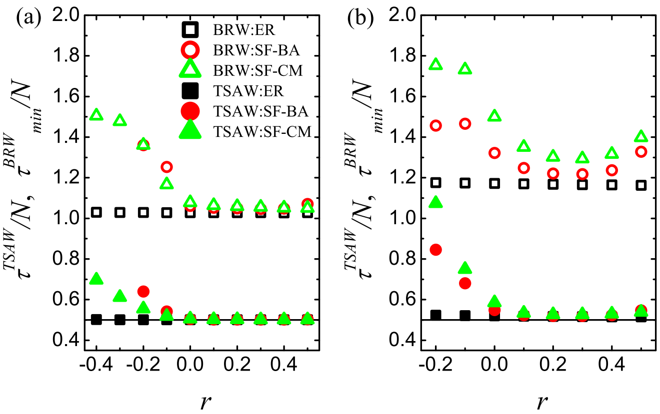

FIG. 2. Plots of $\tau^{\mathrm{BRW}}_{\min}/N$ of BRWs (open symbols) and τ^{TSAW}/N of TSAWs (filled symbols) with (a) $\langle k\rangle = 40$ and (b) $\langle k\rangle = 10$. The solid line represents $\tau/N = 1/2$, which is the theoretical minimum with $N = 10^4$.

searching efficiency by BRWs, α should be carefully adjusted to be $\alpha_{\min}$.

In Fig. 2 the measured $\tau^{\mathrm{BRW}}_{\min}/N$ at $\alpha_{\min}$ and τ^{TSAW}/N for various r's are displayed. Here τ^{TSAW} is the τ of TSAWs. For $\langle k\rangle = 40$, we find that $\tau^{\mathrm{BRW}}_{\min}$ on both BA and CM networks decreases as r increases and saturates to $\tau^{\mathrm{BRW}}_{\min}/N \simeq 1$, while $\tau^{\mathrm{BRW}}_{\min}/N \simeq 1$ on ER networks regardless of r as shown in Fig. 2(a). τ^{TSAW} also decreases as r increases and saturates to a constant which is very close to the *theoretical minimum* $\tau^{\mathrm{th}}/N = \frac{1}{2}$. This limit is obtained when all nodes in a network are visited only once without revisitation. Large values of $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} for $r < 0$ can be understood from the following heuristic arguments. In BRWs on networks, the probability of finding a walker at a node of degree k is proportional to $k^{-\alpha+1}$ [4,29]. Thus, the walker more frequently visits nodes of large k, regardless of the value of α when $r = 0$. If $r < 0$, then nodes of large k have a strong tendency to be connected with nodes of small k. As a result, the walker at a node of small k coming from nodes of large k has a relatively high probability of moving back to a node of large k for BRWs. This increases the revisitation probability of nodes of large k and the walker is more easily trapped in a small part of a network. Similarly, for TSAWs if the walker arrives at a dead-end through a node of large k, then the revisitation probability of a large-degree node increases, which causes a large τ^{TSAW}. On the other hand, if $r > 0$, then each node tends to be connected with nodes of a similar degree. This means that the walker at a node has almost the same probability of moving to one of its neighboring nodes, regardless of the value of α, which reduces the revisitation probability of a specific node. Due to the homogeneous degree distribution in the ER network, r does not affect the searching efficiency of either BRWs or TSAWs on ER networks. More importantly, the data clearly show that $\tau^{\mathrm{BRW}}_{\min} > \tau^{\mathrm{TSAW}}$ for all r. Moreover, the value of τ for TSAWs approaches the theoretical limit $\tau^{\mathrm{th}}/N = \frac{1}{2}$ as r increases [see Fig. 2(a)]. This clearly indicates that the TSAW provides the most optimal performance for exploring the complex network.

For relatively small $\langle k\rangle$ (=10), we find that both $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} increase compared to those for larger $\langle k\rangle$ as shown in Fig. 2(b). This increase in τ's is a natural consequence because the number of possible paths connecting two nodes decreases as $\langle k\rangle$ decreases. In addition, the data in Fig. 2(b) show that $\tau^{\mathrm{BRW}}_{\min}$ on SF networks is a convex function of r. The increase in $\tau^{\mathrm{BRW}}_{\min}$ for $r \geqslant 0.4$ might come from the algorithmic limit of the edge exchange methods [28], which causes some biased sampling of unnatural topologies when the change of r from the original network, $|\Delta r|$, becomes large. However, we find that the inequality $\tau^{\mathrm{BRW}}_{\min} > \tau^{\mathrm{TSAW}}$ is still valid for any r, regardless of this numerical artifact. Furthermore, τ^{TSAW} for $r > 0$ still approaches a theoretical minimum $\tau^{\mathrm{th}}/N = \frac{1}{2}$ and the difference $\tau^{\mathrm{BRW}}_{\min} - \tau^{\mathrm{TSAW}}$ becomes larger than that for large $\langle k\rangle$. We also verify that the inequality $\tau^{\mathrm{BRW}}_{\min} > \tau^{\mathrm{TSAW}}$ holds for other values of $\langle k\rangle$.

We also measure $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} on several real networks. The results are listed in Table I with additional topological properties. As reported in Table I, $\tau^{\mathrm{BRW}}_{\min} > \tau^{\mathrm{TSAW}}$ on any networks, and the difference between $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} becomes maximum for the road network in California. One possible source of this large difference between $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} for the road network in California might come from the small value of $\langle k\rangle$ ($\simeq$2.8) as addressed in Fig. 2. The data in Table I show some tendency for τ^{TSAW} to decrease as $\langle k\rangle$ increases. Moreover, τ^{TSAW} approaches $\tau^{\mathrm{th}}/N = \frac{1}{2}$ as r and $\langle k\rangle$ increase.

IV. SUMMARY AND DISCUSSION

We study the MFPT of TSAWs as a measure of the searching efficiency on complex networks. From the numerical analysis, we find that the inequality $\tau^{\mathrm{BRW}}_{\min} > \tau^{\mathrm{TSAW}}$ holds for any underlying topology. Especially, we find that τ^{TASW} approaches the theoretical limit, $\tau^{\mathrm{th}}/N = 1/2$, when $\langle k\rangle$ is large enough. For a network with relatively small $\langle k\rangle$ the measured τ^{TSAW} is slightly larger than that for large $\langle k\rangle$, but it still remains at a value close to the theoretical

TABLE I. List of measured $\tau^{\mathrm{BRW}}_{\min}$ and τ^{TSAW} on real networks.

	N	$\langle k\rangle$	r	$\alpha_{\min}$	$\tau^{\mathrm{BRW}}_{\min}/N$	τ^{TSAW}/N
www (google) [30]	855802	10.03	−0.06	−0.33	4.11	1.29
Coauthor-AstroPh [31]	17903	22.0	0.20	−0.58	2.15	0.68
Gnutella [32]	62561	4.72	−0.09	−0.55	2.35	0.88
Yeast [33]	2224	5.94	−0.11	−0.46	2.71	0.90
Amazon [34]	410236	11.89	−0.02	−0.58	1.92	0.61
Wordnet [35]	75606	3.18	−0.09	−0.10	3.14	1.71
Road (CA) [30]	1957027	2.8	0.12	0.4	15.15	1.45

minimum for $r > 0$. The similar behavior of τ^{TSAW} is also observed on various real networks.

Furthermore, to obtain the best efficiency of BRWs, α should be very carefully tuned to be $\alpha = \alpha_{\min}$ for each network because $\alpha_{\min}$ depends on the topological properties of the underlying network such as r and N. This implies that $\alpha_{\min}$ should be readjusted if some topological properties of underlying network are changed. The tuning of α is composed of the measurement of τ^{BRW} for various values of α, which cannot be accomplished without the exploration of the network using BRWs. Thus, the tuning of α is practically infeasible or is not efficient for large real networks. In contrast to BRW, TSAW does not require any tuning of parameters. These results clearly show that the TSAW is one of the most efficient and simple models to explore any complex networks using only the local information without any tuning of parameters.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A2057791) and by the NRF (Republic of Korea) grant funded by the Ministry of Education (Republic of Korea) (NRF-2015R1D1A1A01060081).

[1] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, New York, 2008).
[2] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
[3] E. Almaas, R. V. Kulkarni, and D. Stroud, Phys. Rev. E **68**, 056105 (2003).
[4] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[5] S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature **450**, 77 (2007).
[6] S. Lee, S.-H. Yook, and Y. Kim, Physica A **387**, 3033 (2008).
[7] S. Lee, S.-H. Yook, and Y. Kim, Eur. Phys. J. B **68**, 277 (2009).
[8] M. Rosvall, P. Minnhagen, and K. Sneppen, Phys. Rev. E **71**, 066111 (2005).
[9] V. Zlatić, A. Gabrielli, and G. Caldarelli, Phys. Rev. E **82**, 066109 (2010).
[10] S. Yoon, S. Lee, S.-H. Yook, and Y. Kim, Phys. Rev. E **75**, 046114 (2007).
[11] S.-H. Yook, Y.-K. Yun, and Y. Kim, J. Korean Phys. Soc. **56**, 990 (2010).
[12] L. D. F. Costa and G. Travieso, Phys. Rev. E **75**, 016102 (2007).
[13] S. Lee, S.-H. Yook, and Y. Kim, Physica A **385**, 743 (2007).
[14] S. Lee, S.-H. Yook, and Y. Kim, Phys. Rev. E **80**, 017102 (2009).
[15] Y. Kim, H. Park, W. Choi, and S.-H. Yook, Eur. Phys. J. B **88**, 192 (2015).
[16] M. Bonaventura, V. Nicosia, and V. Latora, Phys. Rev. E **89**, 012803 (2014).
[17] N. Madras and G. Slade, *The Self-Avoiding Walk* (Birkhäuser, Boston, 1993).
[18] C. P. Herrero, Phys. Rev. E **71**, 016103 (2005).
[19] C. P. Herrero, Eur. Phys. J. B **56**, 71 (2007).
[20] V. M. L. Millán, V. Cholvi, L. López, and A. F. Anta, Networks **60**, 71 (2012).
[21] D. J. Amit, G. Parisi, and L. Peliti, Phys. Rev. B **27**, 1635 (1983).
[22] S.-J. Yang, Phys. Rev. E **71**, 016107 (2005).
[23] W.-X. Wang, B.-H. Wang, C.-Y. Yin, Y.-B. Xie, and T. Zhou, Phys. Rev. E **73**, 026111 (2006).
[24] P. Erdös and A. Rényi, Publ. Math. **6**, 290 (1959).
[25] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[26] M. Molloy and B. Reed, Random Struct. Alg. **6**, 161 (1995).
[27] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[28] R. Xulvi-Brunet and I. M. Sokolov, Phys. Rev. E **70**, 066102 (2004).
[29] A. Fronczak and P. Fronczak, Phys. Rev. E **80**, 016107 (2009).
[30] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, Internet Math **6**, 29 (2009).
[31] M. E. J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).
[32] R. Matei, A. Iamnitchi, and P. Foster, IEEE Internet Comput. **6**, 50 (2002).
[33] S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, Nucleic Acids Res. **31**, 2443 (2003).
[34] J. Leskovec, L. Adamic, and B. Adamic, ACM Transact. Web (ACM TWEB) **1**, 2 (2007).
[35] Edited by C. Fellbaum, *WordNet: An Electronic Lexical Database* (MIT Press, Cambridge, MA, 1998).