

# Cover time for random walks on arbitrary complex networks

Benjamin F. Maier\*

*Robert Koch-Institute, Nordufer 20, D-13353 Berlin and  
Department of Physics, Humboldt-University of Berlin, Newtonstraße 15, D-12489 Berlin*

Dirk Brockmann

*Robert Koch-Institute, Nordufer 20, D-13353 Berlin and  
Institute for Theoretical Biology, Humboldt-University of Berlin, Philippstr. 13, D-10115 Berlin*

We present an analytical method for computing the mean cover time of a random walk process on arbitrary, complex networks. The cover time is defined as the time a random walker requires to visit every node in the network at least once. This quantity is particularly important for random search processes and target localization in network topologies. Based on the global mean first passage time of target nodes we derive an estimate for the cumulative distribution function of the cover time based on first passage time statistics. We show that our result can be applied to various model networks, including Erdős–Rényi and Barabási–Albert networks, as well as various real-world networks. Our results reveal an intimate link between first passage and cover time statistics in networks in which structurally induced temporal correlations decay quickly and offer a computationally efficient way for estimating cover times in network related applications.

## I. INTRODUCTION

Random walks have been studied extensively for more than a century and emerged as an efficient descriptive model for spreading and diffusion processes in physics, biology, social sciences, epidemiology, and computer science [1–6]. Because of their wide applicability and relevance to dynamic phenomena, random walk processes on complex networks have become a topic of interest [1]. In this context, the calculation of a resistor network’s total resistance [5], synchronization phenomena in networks of coupled oscillators [4], the global spread of infectious diseases on the global air traffic network [7] and ranking the importance of single websites in the world wide web [8] are just a few examples of systems that have been investigated based on concepts derived from random walk theory.

Especially important is the understanding of temporal aspects of stochastic processes and how different network structures influence the equilibration process. Consequently, a lot of theoretical work focused on understanding the connection between network structure and relaxation time scales or first passage times (FPTs), i.e. the time it takes a single walker to travel from one node to another. Both, relaxation and first passage times quantify different aspects but fail to capture the characteristic time a walker requires on average to visit *every* node in a network, which is captured by the *cover time* of the process. This quantity, however, has important practical applications from biology to computer science, for instance, for estimating how long it will take to distribute a chemical or a certain commodity to every node in a network. Researchers have been able to obtain analytical, asymptotic results for the cover time for model networks, e.g. complete graphs, Erdős–Rényi (ER) and Barabási–Albert (BA) networks. Despite these results, analytical results, approximate results or heuristics concerning the mean cover time of real-world networks is lacking

and it is unclear how a real-world network’s mean cover time is related to other temporal features of random walks on networks and how higher order structures of a real network may impact the mean cover time.

In the following, we present a theoretical approach that predicts the cover time on arbitrary complex networks using only FPT statistics. We show that for networks on which random walks equilibrate quickly, the cover time can be estimated accurately by the maximum of a set of FPTs drawn from the ensemble of FPT distributions of all target nodes. Our method’s predictions are in excellent agreement with results provided by computer simulations for a variety of real-world networks, as well as for ER networks, BA networks, complete graphs and the configuration model (uniform degree random networks). We furthermore show that our method produces results with significant deviations for lattice-like networks (e.g. subway networks [9] and two-dimensional regular lattices) that violate the assumption of rapid relaxation.

## II. THEORY

### A. Random walks and first passage times

We investigate a simple discrete time random walk on complex networks. Given an unweighted, undirected, network composed of  $N$  nodes,  $E$  links and adjacency matrix  $A_{vu}$  with  $A_{vu} = 1$  if node  $u$  and  $v$  are connected and  $A_{vu} = 0$  otherwise, a walker starts at  $t = 0$  on any node  $u$ . Then it chooses one of  $u$ ’s neighbors  $v \in U$  at random and jumps to  $v$ , completing the transition from time  $t = 0$  to time  $t = 1$ . This process is repeated indefinitely and governed by the master equation

$$P_v(t+1) = \sum_{u=1}^N W_{vu} P_u(t),$$

where  $P_u(t)$  is the probability that the walker is at node  $u$  at time  $t$ ,  $k_u = \sum_{v=1}^N A_{vu}$  is the degree of node  $u$  and  $W_{vu} =$

\* bfmaier@physik.hu-berlin.de

$A_{vu}/k_u$  is the transition probability of the walker going from node  $u$  to node  $v$  in one time step. We assume that the network has a single component, so every node can be reached, in principle, from every other node. Under relatively general conditions this process will approach the equilibrium

$$P_v^* = k_v/2E.$$

Central questions for random walks are often connected to first passage times (FPT), e.g. the mean first passage time  $\tau_{vu}$  (MFPT) between two nodes  $u$  and  $v$ . This time is defined as the mean number of steps it takes a random walker starting at node  $u$  to first arrive at node  $v$ . Another important quantity is the global mean first passage time of node  $v$  (GMFPT),

$$\tau_v = \frac{1}{N-1} \sum_{u \neq v} \tau_{vu}. \quad (1)$$

The GMFPT can be used as a measure of centrality for node  $v$  since a node which is quickly reachable from the whole network may be interpreted to be “important”. Those passage times are well analyzed and often efficiently computable from network properties. One of those is the graph Laplacian, which emerges naturally when investigating diffusion on networks and is defined as

$$L_{vu} = k_v \delta_{vu} - A_{vu},$$

where  $\delta_{vu}$  denotes Kronecker’s delta. Using this operator, the MFPT between two nodes can be computed by inversion of a reduced Laplacian (note that the operator above is singular and hence not invertible) [5]. A method described in [10] computes the GMFPT from the graph Laplacian’s spectrum. Given its eigenvalues in increasing order  $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$  and corresponding orthonormal eigenvectors  $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{iN})^T$ , one can compute node  $v$ ’s exact GMFPT as

$$\tau_v^{\text{ex}} = \frac{N}{N-1} \sum_{i=2}^N \frac{1}{\lambda_i} \left( 2E \mu_{iv}^2 - \mu_{iv} \sum_n k_n \mu_{in} \right). \quad (2)$$

A computationally more efficient method estimates the GMFPT by its lower bound, which is given by

$$\tau_v^{\text{es}} \geq \frac{N \langle k \rangle}{k_v} \frac{1}{1 - \langle k \rangle^{-1}}, \quad (3)$$

if the process relaxates quickly, e.g. in time  $t \ll N$ , as derived in [11]. In the same study it was shown that this result holds remarkably well for Erdős–Rényi (ER) and Barabási–Albert (BA) networks, as well as for a variety of real-world networks.

Another temporal characteristic with practical relevance is the mean cover time  $T_u$ , defined as the mean number of steps it takes a random walker starting at node  $u$  to visit every other node at least once. For various network models simple heuristics concerning the asymptotic scaling of the mean cover time with growing node number  $N$  have been derived [12–14]. For ER, BA and fully connected graphs it was shown that

$$\langle T \rangle \sim Q N \log N,$$

with network specific prefactor  $Q$ . Such scaling relationships are useful for comparative analyses, e.g. when networks for different sizes of the same class are compared. They are less helpful when actual expected cover times need to be computed for empirical networks where  $N$  is fixed and comparative or relative statements are insufficient.

Unfortunately, a general procedure for estimating the actual cover time for arbitrary complex networks, as well as the connection between the mean cover time and FPT observables is lacking. In the following we present a method that estimates the cover time using passage time statistics.

## B. Cover Time

Recently it has been found that if a random walk process equilibrates quickly, i.e. the initial concentration of random walkers approaches the equilibrium concentration in a small number of time steps  $t \ll N$ , the information about the start node is lost [11] and the first passage time at destination  $v$  is (for larger times  $\tau$ ) distributed asymptotically according to

$$p_v(\tau) \propto \exp(-\tau/\tau_v)$$

where  $\tau_v$  is the GMFPT of Eq. (1).  $\tau_v$  can differ between nodes and depends on the topological features of the network only. Note that in the following paragraphs we will often refer to the FPT decay rate  $\beta_v = \tau_v^{-1}$  instead of the GMFPT, simplifying the notation.

In order to find the mean cover time from the collection of distributions  $p_v(\tau)$  we proceed as illustrated in Fig. 1. Excluding the starting node  $u$ , we pick a FPT  $t_v$  for each target node  $v$  from their respective FPT distribution  $p_v$  at random, resulting in a set of  $N-1$  FPT that we call  $\mathcal{F}$ . Consequently, the cover time  $T_u$  is given as the maximum element of  $\mathcal{F}$ . In order to find the distribution of this maximum, we compute the probability that a time  $T$  is equal to the maximum of this set as the probability that no element of  $\mathcal{F}$  is larger than  $T$ , yielding

$$\begin{aligned} P_u(T) &= P_u(\text{“}T \text{ is the maximum time”}) \\ &= P_u(t_v \leq T \ \forall t_v \in \mathcal{F}) = \prod_{v \neq u} P_u(t_v \leq T). \end{aligned} \quad (4)$$

We further approximate our result by assuming a continuous time distribution, easing the computations without significantly changing the outcome. Then, the probability that any time  $t_v \in \mathcal{F}$  is lower than or equal to  $T$  is

$$\begin{aligned} P(t_v \leq T) &= \int_0^T dt \ p_v(t) = \int_0^T dt \ \beta_v \exp(-\beta_v t) \\ &= 1 - \exp(-\beta_v T). \end{aligned} \quad (5)$$

Eqs. (5) and (4) yield the cumulative distribution function for the cover time,

$$P_u(T) = \prod_{v \neq u} \left( 1 - \exp(-\beta_v T) \right).$$

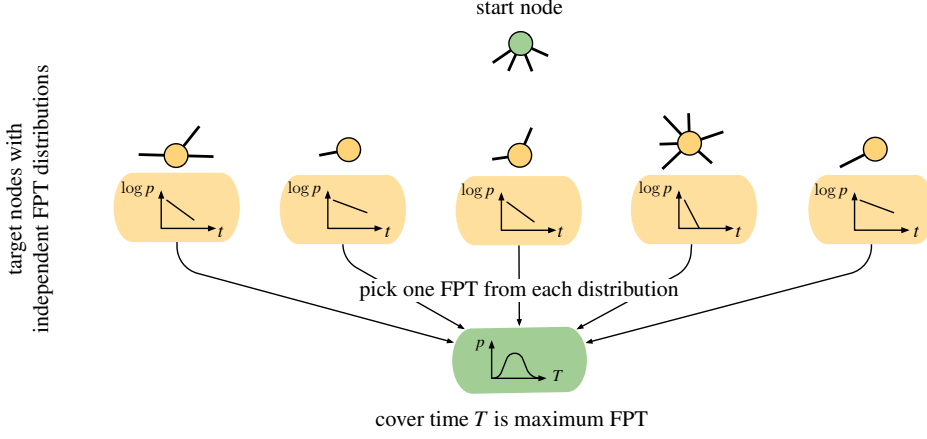


Figure 1. Illustration of our approach. Each node of the network is viewed as an independent entity that can be visited by the walker starting at the green node. Each target node  $v$  has its own first passage time (FPT) distribution which is asymptotically distributed as  $\sim \exp(-\tau/\tau_v)$  with its global mean FPT  $\tau_v$ . In order to compute the cover time, we draw one FPT  $\tau$  from every target's distribution. Then the cover time is given as the maximum time of all drawn FPTs.

and the expected cover time

$$T_u = \int_0^\infty dT \left[ 1 - \prod_{v \neq u} (1 - \exp(-\beta_v T)) \right]. \quad (6)$$

We can find the global mean cover time by averaging over target nodes  $u$  as

$$\langle T \rangle = \int_0^\infty dT \left[ 1 - P(T) \frac{1}{N} \sum_{u=1}^N \frac{1}{1 - \exp(-\beta_u T)} \right] \quad (7)$$

with

$$P(T) = \prod_{v=1}^N (1 - \exp(-\beta_v T)). \quad (8)$$

However, as argued in App. B, without introducing too much error for  $N \gg 1$ , we will make use of a simpler integral to find

$$\begin{aligned} \langle T \rangle &= \int_0^\infty dT [1 - P(T)] = \int_0^\infty dT \left[ 1 - \prod_{v=1}^N (1 - \exp(-\beta_v T)) \right] \\ &= \sum_{S \in \mathcal{P}^*(\mathcal{V})} (-1)^{|S|+1} \left( \sum_{v \in S} \beta_v \right)^{-1}. \end{aligned} \quad (9)$$

Here,  $\mathcal{V}$  is the set of all nodes and  $\mathcal{P}^*(\mathcal{V})$  is the set of all possible subsets of  $\mathcal{V}$  excluding the empty set. Even though one can solve the integral Eq. (9) analytically to obtain the result above, in practice it is more feasible to solve the integral numerically than iterating over  $\mathcal{P}^*(\mathcal{V})$  which has  $2^N - 1$  elements and hence becomes very large rather quickly.

Now, the estimation of the global mean cover time reduces to an efficient estimation of the FPT decay rates  $\beta_v$ . There

are two ways to estimate the decay rates with the GMFPTs as described in Sec. II A. Using the estimation of the lower bound Eq. (3), the estimated global mean cover time is given as

$$\langle T^{\text{es}} \rangle \geq \int_0^\infty dT \left[ 1 - \prod_{v=1}^N \left( 1 - \exp \left( -T k_v \frac{1 - \langle k \rangle^{-1}}{N \langle k \rangle} \right) \right) \right]. \quad (10)$$

The advantage of this method is that only the network's degree sequence  $k_v$  needs to be known in order to estimate the global mean cover time. However, this method can obviously only account for a lower bound. Secondly, we can compute the exact GMFPTs using Eq. (2). Then the computed global mean cover time is

$$\langle T^{\text{ex}} \rangle = \int_0^\infty dT \left[ 1 - \prod_{v=1}^N \left( 1 - \exp \left( -\frac{T}{\tau_v^{\text{ex}}} \right) \right) \right]. \quad (11)$$

### C. Cover time of networks with equal GMFPTs

Let us consider a network where all nodes have approximately the same GMFPT  $\tilde{\tau}$  and the structure is sufficiently random that we can estimate the mean cover time using Eq. (6). We find

$$\begin{aligned} \langle T \rangle(\tilde{\tau}) &= \int_0^\infty dT \left[ 1 - (1 - \exp(-T/\tilde{\tau}))^{N-1} \right] \\ &= \tilde{\tau} \left[ \gamma + \psi(N) \right], \end{aligned} \quad (12)$$

where  $\gamma \approx 0.57722$  is the Euler-Mascheroni constant and  $\psi(z) = \Gamma'(z)/\Gamma(z)$  the polygamma function.

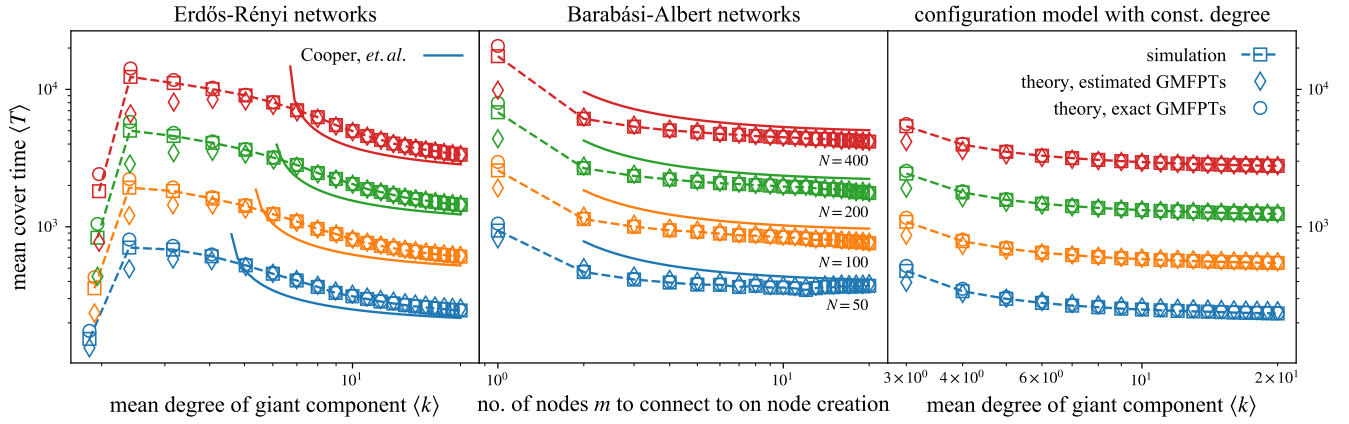


Figure 2. The mean cover time  $\langle T \rangle$  of the giant component of Erdős-Rényi (ER) networks (left), Barabási-Albert (BA) networks (center), and networks generated with the configuration model and nearly constant node degree (right). Shown are averages measured from 1000 simulations per data point ( $\square$ ) and both estimations of the mean cover time using ( $\diamond$ ) estimated GMFPTs from the target nodes' degrees Eq. (10) and ( $\circ$ ) exact GMFPTs Eq. (11) computed from the graph Laplacian's spectrum. Dashed lines are simulation results and a guide to the eye. Further displayed are the asymptotic results derived by Cooper, et. al. for the ER [12] and BA [13] network model.

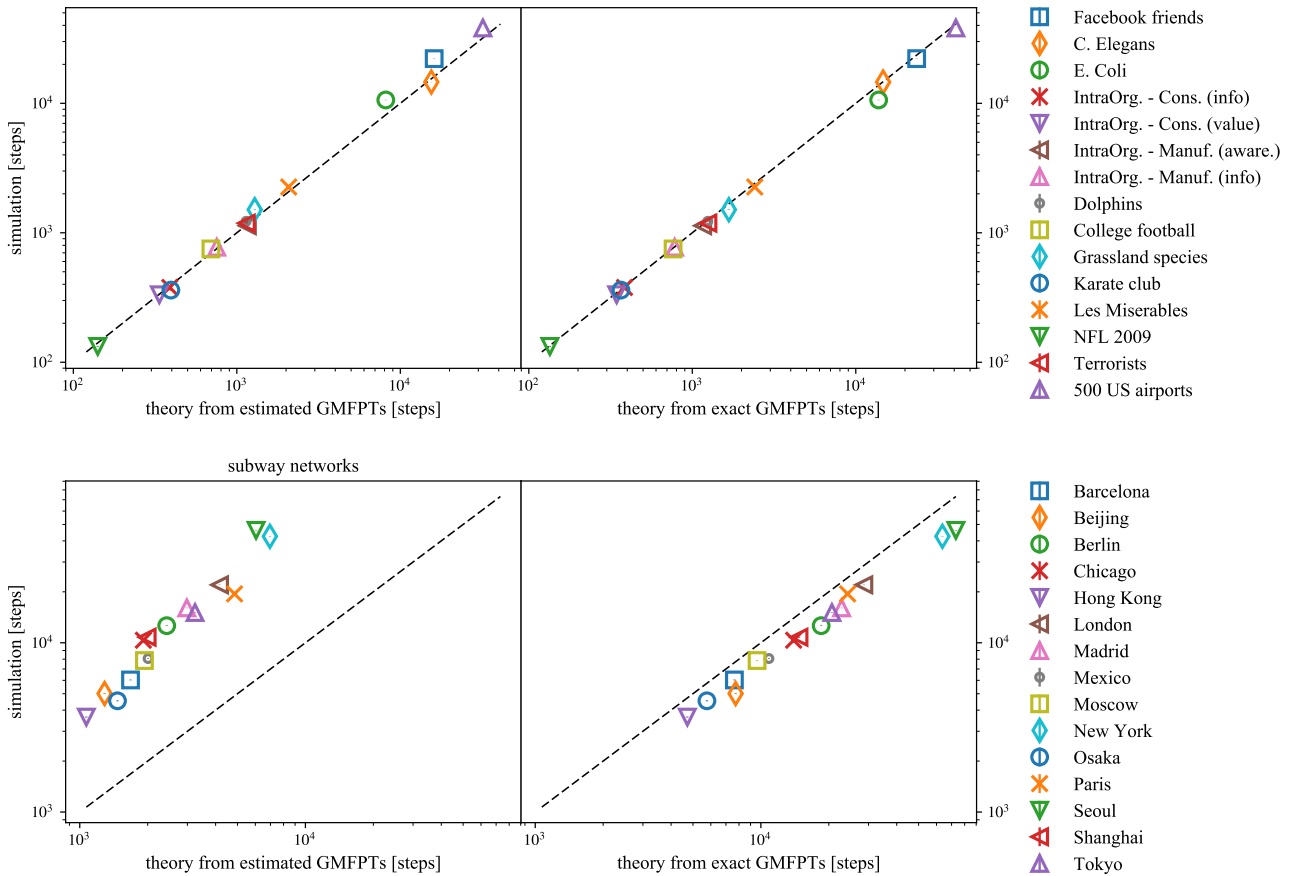


Figure 3. Mean cover times of simple random walks on the giant component of various networks (data sources and relative errors given in Tab. I and Tab. II). The estimated cover times are compared to the measured cover time (from 50 simulations for each data point). The dashed lines represent the ideal case  $\langle T^{\text{sim}} \rangle = \langle T^{\text{es/ex}} \rangle$ . Theoretical results are computed from (left) estimated GMFPTs from the target nodes' degrees Eq. (10) and (right) exact GMFPTs computed from the graph Laplacian's spectrum Eq. (11).

Network	$N$	$\langle k \rangle$	$\frac{\langle T^{\text{sim}} \rangle}{N \log N}$	$\frac{\langle T^{\text{es}} \rangle}{N \log N}$	$1 - \frac{\langle T^{\text{sim}} \rangle}{\langle T^{\text{es}} \rangle}$	$\frac{\langle T^{\text{ex}} \rangle}{N \log N}$	$1 - \frac{\langle T^{\text{sim}} \rangle}{\langle T^{\text{ex}} \rangle}$
First author's facebook friends network [15]	329	11.9	11.61	8.45	0.37	12.36	0.061
C. Elegans neural network [16]	297	14.6	8.64	9.15	0.06	8.69	0.006
E. Coli protein interaction [17]	329	2.8	5.57	4.27	0.30	7.24	0.231
Intra-org. contacts - Cons. (info) [18]	43	15.3	2.34	2.41	0.03	2.38	0.018
Intra-org. contacts - Cons. (value)	44	16.0	2.00	2.02	0.01	2.07	0.036
Intra-org. contacts - Manuf. (awareness)	77	25.5	3.39	3.47	0.02	3.46	0.021
Intra-org. contacts - Manuf. (info)	76	23.3	2.35	2.29	0.03	2.37	0.009
Social interaction in dolphins [19]	62	5.1	4.79	4.46	0.07	4.86	0.015
American college football [20]	115	10.7	1.37	1.27	0.08	1.40	0.017
Food web of grassland species [21]	75	3.0	4.66	3.97	0.17	5.17	0.099
Zachary's Karate club [22]	34	4.5	3.01	3.29	0.09	3.05	0.015
Interactions in "Les Misérables" [23]	77	6.6	6.75	6.21	0.09	7.21	0.063
Matches of the NFL 2009 [24]	32	13.2	1.20	1.27	0.06	1.21	0.015
Network of associations between terrorists [25]	62	4.9	4.63	4.47	0.04	4.87	0.049
Connections between 500 largest US airports [26]	500	11.9	12.29	10.30	0.19	13.18	0.067

Table I. Mean cover times of simple random walks on the giant component of various networks, in units of the cover time on a complete graph with equal node count. Displayed is the number of nodes  $N$ , the mean degree  $\langle k \rangle$  of the giant component and the measured mean cover time  $\langle T^{\text{sim}} \rangle$  extracted from 50 simulations per network with one walker starting on every node. Additionally shown are both theoretical estimations of the cover time using (a) estimated GMFPTs  $\langle T^{\text{es}} \rangle$  from the target nodes' degrees, Eq. (10) and (b) exact GMFPTs  $\langle T^{\text{ex}} \rangle$  computed from the graph Laplacian's spectrum Eq. (11). Both estimations are given with their relative error to the simulated mean cover time. Note that all networks have been symmetrized and an unweighted link  $(u, v)$  has been created if a weight between two nodes was  $w_{uv} > 0$ . For the Intra-organizational networks, we created a link if both nodes put something else than "I do not know this person" in their questionnaire. These values are additionally shown in Fig. 3 (top).

An example for networks fulfilling the conditions above is the *configuration model* where all nodes have identical degree. This includes, e.g. the complete graph. The cover time of the complete graph scales as  $\langle T \rangle \sim N \log N$  for large  $N$  [14], a result which is reproduced by Eq. (12) since the GMFPT for each node is  $\tilde{\tau} = N - 1$  (see App. A) and  $\psi(N) \sim \log N$  for large  $N$ .

### III. RESULTS

We compared the predictions of Eqs. (10) and (11) with simulation results for single component ER, BA and real-world networks, as well as Eq. (12) for random networks with a strongly peaked degree distribution. On every node we placed a walker at time  $t = 0$ . Subsequently, we let each walker do a random walk as described in Sec. II A. Each walker proceeded until it visited each node at least once, completing total coverage and marking cover time  $T_u$ . Then  $\langle T \rangle$  was computed as the average of all  $T_u$ . For a more detailed description of the numerical methods as well as the used code, see App. C.

For both ER and BA networks, we generated networks with  $N = 50, 100, 200, 400$  nodes, ER networks with integer mean degree  $\langle k \rangle \in [1, 20]$ , and BA networks with parameter  $m \in [1, 20]$  (new links per node at creation). In order to test Eq. (12), we generated networks using the configuration model with  $N = 50, 100, 200, 400$  nodes and degree sequence  $k_u = k \ \forall u \in \mathcal{V}$ , scanning integer degrees  $k \in [3, 20]$ .

After creation of each network, we extracted the giant component, deleted self-links and multiple occurrences of unique links, ran the random walks and estimated the cover time using Eqs. (10), (11) and Eq. (12), respectively, for 1000 networks each. The theoretic results are in excellent agreement with the simulation results, as can be seen in Fig. 2. The relative error made by our method decreases with increasing number of nodes  $N$  as well as increasing mean degree  $\langle k \rangle$  and quickly reaches values below 1%. Unsurprisingly, our method performs better compared to the results of [12, 13] due to the asymptotic nature of the latter.

We furthermore simulated random walks on the giant component of 15 real-world networks, listed in Tab. I. Some of those are directed networks that we converted to undirected networks by converting every directed link to an undirected link. For weighted networks we assigned an undirected link  $(u, v)$  if a weight was  $w_{uv} > 0$ . For the intra-organizational networks from [18], employees had to fill out questionnaires regarding their relationships to co-workers. Here, we assigned an undirected link  $(u, v)$  if both  $u$  and  $v$  marked something else than "I do not know this person". As can be seen in Tab. I, our method produces results that are very close to the simulated values (mostly relative errors of  $\sim 1\% - 10\%$ ). One exception is the computed cover time for the E. coli protein interaction network [17] with a relatively high relative error of  $\sim 23\%$ . Generally, the more exact result of GMFPTs calculated via the graph Laplacian gives results with lower relative error than using lower bound GMFPTs, as expected.

Additionally, we performed simulations on two-

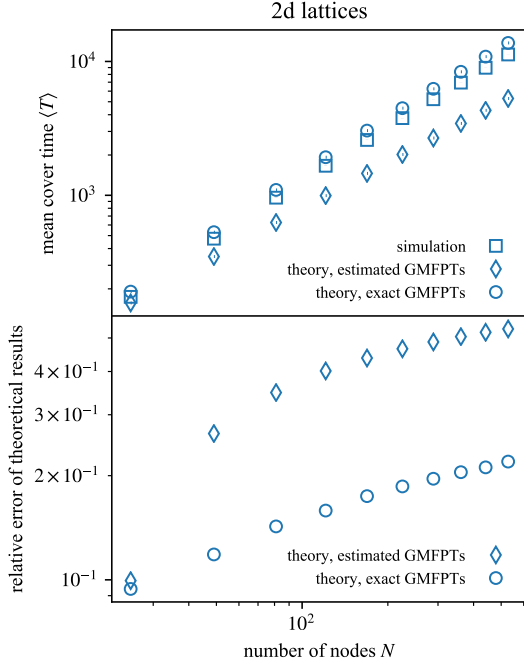


Figure 4. Example of our method yielding results with rather large deviations from simulations. Here, we show the mean cover time for various two-dimensional square lattices as ( $\square$ ) an average over 1000 simulations for each data point, ( $\diamond$ ) theoretical result from estimated GMFPTs from the target nodes' degrees Eq. (10) and ( $\circ$ ) theoretical result from exact GMFPTs computed from the graph Laplacian's spectrum Eq. (11). For both theoretic estimations the relative error is shown in the bottom plots; it is increasing with increasing system size.

dimensional square lattices of  $N \in [5^2, 25^2]$  nodes. For those networks the GMFPT is roughly equal for all nodes, however, information of the starting node is very important for picking the right FPT of targets. Simply picking one FPT from a distribution is not a valid procedure anymore, due to high correlation of FPTs with lattice distance. Hence, we suspect that our method will not perform well for lattice-like networks. Indeed, as can be seen in Fig. 4, the relative error between simulation and theoretic estimations increases with increasing  $N$ , up to  $\sim 50\%$  for the estimated GMFPTs and  $\sim 20\%$  for the exact GMFPTs. Similar results are obtained for lattice-like real-world networks such as subway networks [9] (shown in Tab. II and Fig. 3). Here, the estimation from estimated GMFPTs systematically underestimates the cover time while using exact GMFPTs yielded an overestimation of the cover time by  $\sim 20\% - 40\%$ .

#### IV. CONCLUSIONS

We studied the cover time of simple discrete time random walks on complex networks with  $N$  nodes. Treating each target node as independent from the start node, we were able to find the cumulative distribution function of the cover time

by finding the maximum of  $N$  drawn FPTs from the  $N$  target nodes' FPT distributions which solely depend on their GMFPT. Using this method, the complexity of finding the mean cover time of an arbitrary complex network is heavily decreased since the problem is practically reduced to finding the nodes' GMFPTs using simple estimations or spectral methods, which is computationally much more feasible than simulating random walks starting on every node, especially for large networks.

We showed that this procedure yields reliable estimations of the mean cover time for a variety of networks where random walks decay quickly, namely ER networks, BA networks, random graphs with heavily peaked degree distributions, and a collection of real-world networks. We furthermore showed that for lattice-like networks our method does not produce reliable results since the information about the start node is important for picking the right FPT from the FPT distribution of the target nodes. While the spatial correlation of nodes is responsible for this deviation, it still remains unclear how to determine under which conditions our method is safely applicable. This is a task for future investigations.

Finally, note that even though we derived our results for unweighted networks, they can be easily made applicable to weighted networks and subclasses of directed networks, as they only depend on the calculation of GMFPTs. Those, in turn, depend solely on the transition matrix  $W_{vu} = A_{vu} / \sum_u A_{vu}$  which is similarly defined for weighted and directed networks.

#### Appendix A: The GMFPT of a complete graph

Suppose a random walker starts at any node  $u$ . The probability to reach any other node of the network in one time step is  $p = 1/(N-1)$ . Looking at a single target node  $v$  we want to calculate the probability that  $v$  is first passaged at time  $t$ , which is given as

$$p_t = (1-p)^{t-1} p.$$

Hence, the GMFPT for every target node is

$$\tilde{\tau} = \sum_{t=1}^{\infty} t p_t = -p \frac{\partial}{\partial p} \frac{1}{p} = \frac{1}{p} = N-1.$$

#### Appendix B: Approximation of mean cover time integral

In the following, we show that instead of solving integral Eq. (7), one can safely use Eq. (9). We do so by calculating

Network	$N$	$\langle k \rangle$	$\frac{\langle T^{\text{sim}} \rangle}{N \log N}$	$\frac{\langle T^{\text{es}} \rangle}{N \log N}$	$1 - \frac{\langle T^{\text{sim}} \rangle}{\langle T^{\text{es}} \rangle}$	$\frac{\langle T^{\text{ex}} \rangle}{N \log N}$	$1 - \frac{\langle T^{\text{sim}} \rangle}{\langle T^{\text{ex}} \rangle}$
Barcelona	128	2.2	9.7	2.70	2.6	12.29	0.21
Beijing	104	2.2	10.4	2.67	2.9	16.00	0.35
Berlin	170	2.1	14.4	2.78	4.2	21.19	0.32
Chicago	141	2.1	14.8	2.74	4.4	20.02	0.26
Hong Kong	82	2.1	10.1	2.96	2.4	13.07	0.23
London	266	2.3	14.8	2.80	4.3	19.26	0.23
Madrid	209	2.3	14.4	2.67	4.4	20.42	0.29
Mexico	147	2.2	11.0	2.74	3.0	14.80	0.26
Moscow	134	2.3	12.0	2.95	3.1	14.66	0.18
New York	433	2.2	16.2	2.65	5.1	24.25	0.33
Osaka	108	2.3	9.0	2.91	2.1	11.40	0.21
Paris	299	2.4	11.4	2.85	3.0	14.22	0.20
Seoul	392	2.2	19.7	2.58	6.6	31.24	0.37
Shanghai	148	2.1	14.6	2.69	4.4	19.96	0.27
Tokyo	217	2.4	12.9	2.78	3.6	17.69	0.27

Table II. Same procedure as in Tab. I, but for subway networks of big cities, taken from [9]. These values are additionally shown in Fig. 3 (bottom).

the total difference between both as

$$\begin{aligned} \Theta &= \int_0^\infty dT [1 - P(T)] - \\ &\quad - \int_0^\infty dT \left[ 1 - P(T) \frac{1}{N} \sum_{v=1}^N \frac{1}{1 - \exp(-\beta_v T)} \right] \\ &= \int_0^\infty dT P(T) [Q(T) - 1], \end{aligned} \quad (\text{B1})$$

defining  $Q(T) = \frac{1}{N} \sum_{v=1}^N (1 - \exp(-\beta_v T))^{-1}$ . Note that the cover time cdf  $P(T)$  is given by Eq. (8), s.t. both

$$\begin{aligned} \lim_{T \rightarrow 0} P(T) &= 0 \\ \lim_{T \rightarrow 0} P(T) Q(T) &= 0 \end{aligned}$$

and

$$\begin{aligned} \lim_{T \rightarrow \infty} P(T) &= 1 \\ \lim_{T \rightarrow \infty} P(T) Q(T) &= 1, \end{aligned}$$

meaning that for both integration limits, the integrand approaches 0. In the following we assume that the distribution of decay rates is relatively homogeneous in the region of small rates (implying that there are some rates that are of the same order as  $\beta_{\min}$ ). This is a relatively safe assumption for most network models and real-world networks as in most cases there are more nodes with small degree (hence small decay rates) than nodes with high degree (hence high decay rates). Now suppose the integration approaches a time

where  $T \approx \beta_{\min}^{-1}$ , implying that there are still some terms  $1 - \exp(-T\beta_v) \ll 1$ , such that  $P(T) \xrightarrow{N \gg 1} 0$ . Furthermore, there will already be a majority of terms  $1 - \exp(-T\beta_v) \rightarrow 1$  which leads to  $Q(T)$  approaching  $Q(T) \xrightarrow{N \gg 1} 1$ . Hence, we can safely assume that for a network with a larger number of nodes the integrand approaches zero at all times while the global mean cover time grows quickly and thus the relative error of Eq. (9) is approaching

$$\frac{\Theta}{\langle T \rangle} \xrightarrow{N \gg 1} 0.$$

### Appendix C: Simulations and numerical evaluations of the mean cover time

The code used for the random walk simulations is a standard implementation of discrete time random walks on networks and available online as a C++/Python/Matlab package, see <https://github.com/benmaier/cNetworkDiff>.

In order to evaluate the mean cover time, the integrals Eqs. (10) and (11) were solved numerically. Note that for the numerical integration, an upper integration bound of infinity can be problematic when the decay region of the integrand is unknown. Hence, we chose an upper integration limit  $T_{\max}$  where  $1 - P(T_{\max}) \leq 10^{-10}$ . The code is available online as a Python package, see <https://github.com/benmaier/nwDiff>. For solving the integral Eq. (11) we computed the GMFPT of each node via Eq. (2) with eigenvalues and -vectors computed using the NumPy implementation [27] of the standard algorithm for eigenvalue and -vector computation of Hermitian matrices [28].

- 
- [1] N. Masuda, M. A. Porter, and R. Lambiotte, “Random walks and diffusion on networks,” *arXiv:1612.03281 [cond-mat, physics:physics]*, Dec. 2016. arXiv: 1612.03281.
- [2] H. C. Berg, *Random walks in biology*. Princeton, NJ: Princeton University Press, 1993.
- [3] J. Klafter and I. M. Sokolov, *First steps in random walks: from tools to applications*. Oxford ; New York: Oxford University Press, 2011. OCLC: ocn714724924.
- [4] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge, UK ; New York: Cambridge University Press, 2008. OCLC: ocn231581094.
- [5] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.
- [6] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*. Universitext, Berlin: Springer, 1992. OCLC: 246776666.
- [7] F. Iannelli, A. Koher, D. Brockmann, P. Hövel, and I. M. Sokolov, “Effective distances for epidemics spreading on complex networks,” *Physical Review E*, vol. 95, p. 012313, Jan. 2017.
- [8] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” Technical Report 1999-66, Stanford InfoLab, Nov. 1999.
- [9] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, “A long-time limit for world subway networks,” *Journal of The Royal Society Interface*, 2012.
- [10] Y. Lin, A. Julaiti, and Z. Zhang, “Mean first-passage time for random walks in general graphs with a deep trap,” *The Journal of Chemical Physics*, vol. 137, p. 124104, Sept. 2012.
- [11] H. W. Lau and K. Y. Szeto, “Asymptotic analysis of first passage time in complex networks,” *EPL (Europhysics Letters)*, vol. 90, no. 4, p. 40005, 2010.
- [12] C. Cooper and A. Frieze, “The cover time of sparse random graphs,” *Random Structures and Algorithms*, vol. 30, pp. 1–16, Jan. 2007.
- [13] C. Cooper and A. Frieze, “The cover time of the preferential attachment graph,” *Journal of Combinatorial Theory, Series B*, vol. 97, no. 2, pp. 269–290, 2007.
- [14] L. Lovász, “Random Walks on Graphs: A Survey,” in *Combinatorics, Paul Erdős is Eighty* (D. Miklós, V. T. Sós, and T. Szőnyi, eds.), vol. 2, pp. 353–398, Budapest: János Bolyai Mathematical Society, 1996.
- [15] B. F. Maier, “B.F. Maier’s FB friends network, <https://github.com/benmaier/BFMaierFBnetwork>,” Apr. 2017.
- [16] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, pp. 440–442, June 1998.
- [17] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, “Network motifs in the transcriptional regulation network of *Escherichia coli*,” *Nature Genetics*, vol. 31, no. 1, pp. 64–68, 2002.
- [18] R. Cross and A. Parker, *The Hidden Power of Social Networks*. Boston, MA: Harvard Business School Press, 2004.
- [19] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, “The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations,” *Behavioral Ecology and Sociobiology*, vol. 54, pp. 396–405, Sept. 2003.
- [20] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 7821–7826, June 2002.
- [21] H. A. Dawah, B. A. Hawkins, and M. F. Claridge, “Structure of the Parasitoid Communities of Grass-Feeding Chalcid Wasps,” *Journal of Animal Ecology*, vol. 64, no. 6, pp. 708–720, 1995.
- [22] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *Journal of Anthropological Research*, vol. 33, pp. 452–473, 1977.
- [23] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*. Reading, MA: Addison-Wesley, 1993.
- [24] C. Aicher, A. Z. Jacobs, and A. Clauset, “Learning latent block structure in weighted networks,” *Journal of Complex Networks*, vol. 3, pp. 221–248, June 2015.
- [25] V. Krebs, “Mapping networks of terrorist cells,” *Connections*, vol. 24, pp. 43–52, 2002.
- [26] V. Colizza, R. Pastor-Satorras, and A. Vespignani, “Reaction-diffusion processes and metapopulation models in heterogeneous networks,” *Nature Physics*, vol. 3, pp. 276–282, Apr. 2007.
- [27] E. Jones, T. Oliphant, P. Peterson, and others, *SciPy: Open source scientific tools for Python*. 2001.
- [28] G. Strang, *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/Cole, 2nd ed ed., 1980.