

A random walk model for infection on graphs: spread of epidemics & rumours with mobile agents

Moez Draief · Ayalvadi Ganesh

Received: 13 January 2010 / Accepted: 27 July 2010 / Published online: 17 August 2010
© Springer Science+Business Media, LLC 2010

Abstract We address the question of understanding the effect of the underlying network topology on the spread of a virus and the dissemination of information when users are mobile performing independent random walks on a graph. To this end, we propose a simple model of infection that enables to study the coincidence time of two random walkers on an arbitrary graph. By studying the coincidence time of a susceptible and an infected individual both moving in the graph we obtain estimates of the infection probability. The main result of this paper is to pinpoint the impact of the network topology on the infection probability. More precisely, we prove that for homogeneous graphs including regular graphs and the classical Erdős–Rényi model, the coincidence time is inversely proportional to the number of nodes in the graph. We then study the model on power-law graphs, that exhibit heterogeneous connectivity patterns, and show the existence of a phase transition for the coincidence time depending on the parameter of the power-law of the degree distribution. We finally undertake a preliminary analysis for the case with k random walkers and provide upper bounds on the convergence time for both the complete graph and regular graphs.

Keywords Random walks on graphs · Power-law graphs · Erdős–Rényi graph · Regular graphs · Infection time

A preliminary version of this paper has appeared as Ganesh and Draief (2009).

M. Draief (✉)

Department of Electrical Engineering, Imperial College London, South Kensington, SW7 2AZ, UK
e-mail: m.draief@imperial.ac.uk

A. Ganesh

Department of Mathematics, Bristol University, Bristol, BS8 1TW, UK
e-mail: a.ganesh@bristol.ac.uk

1 Introduction

In recent years, there has been a surge of hand-held wireless computing devices such as PDAs together with the proliferation of new services. These portable computing devices are equipped with a short-range wireless technology such as WiFi or Bluetooth. Despite providing a great deal of flexibility this ability to wirelessly connect to other devices, and to transfer data on the move, attracted the attention of virus writers who exploit such features for launching computer-virus outbreaks that take advantage of human mobility (Wang et al. 2009; Kleinberg 2007; Leavitt 2005).

Over the past couple of years, there have been indeed reports of malicious code that takes advantage of bluetooth vulnerabilities such as the Cabir worm that was detected during the World Athletics Championship (Mobile 2010) and another at a company that has been reported by CommWarrior (<http://www.f-secure.com/v-descs/commwarrior.shtml>). Despite their small scale, these incidents bode more threats taking advantage of events and locations where individuals gather in close proximity (Yan et al. 2007; Su et al. 2006).

In much of the literature on mathematical epidemiology, the members of the population are assumed to occupy fixed locations and the probability of infection passing between a pair of them in a fixed time interval is taken to be some function of the distance between them. Mean-field (aka homogeneous mixing) models are a special case where an infected individual can potentially infect a number of random individuals chosen uniformly at random among the population (Daley and Gani 2001). Recently, there has been an increasing interest in understanding the impact of the network topology on the spread of epidemics in networks with fixed nodes, see Draief (2006), Draief and Massoulié (2010). There has however been little analytical work to date on how mobility patterns may affect the outcomes of the processes concerned.

In this work, we consider a different model in which the agents are mobile and can only infect each other if they are in sufficiently close proximity. The model is motivated both by certain kinds of biological epidemics, whose transmission may be dominated by sites at which individuals gather in close proximity (e.g. workplaces or public transport for a disease like SARS, cattle markets for foot-and-mouth disease, etc.) and by malware spreading between wireless devices via Bluetooth connections, for example.

Related work In what follows, we briefly describe some of the relevant related work on modelling epidemic spreading in mobile environments. To our knowledge, the first attempts to model virus spreading in mobile networks rely on the use of non-rigorous mean-field approximations (similar to the classical Kephart–White model Kephart and White 1991) that incorporate the mobility patterns of users. In Mickens and Noble (2005), the authors derive a threshold for the persistence of the epidemic by computing the average number of neighbours of a given node. Using a similar approach but with different mobility patterns, Nekovee (2007), Rhodes and Nekovee (2008) explore the evolution of the number of devices that are infected in terms of the contact rate between users.

A related line of work studying the dissemination of information in opportunistic networks (Chaintreau et al. 2007) focuses on the following analogous problem: Suppose that a set of mobile agents with wireless communication capabilities, forming a temporary network without the aid of a fixed infrastructure, are interested in a

piece of information that is initially held by one user. The information is transmitted between users who happen to be in each other's range. As in the case of static networks (Pittel 1987), one may be interested in the time it takes for the rumour to be known to all users. To this end we need to understand how information is transmitted between an informed and an ignorant user. This is also relevant to patterns of opinion formation and propagation of trends on social networks (Amini et al. 2009). Our work gives some insight on the impact of the network structure on the likelihood of successfully transmitting the rumour.

Our contribution In contrast to the previous work which has focused on Euclidean models and homogeneous mobility patterns, in this work we consider a model wherein the different locations that a user can reach have varying popularity. More precisely, we consider a simple and stylised mathematical model of the spread of infection as follows. There is a finite, connected, undirected graph $G = (V, E)$ on which the individuals perform independent random walks following a certain rate transition matrix Q . The infection can pass from an infected to a susceptible individual only if they are both at the same vertex. We assume that the probability of the virus being passed over a time interval of length τ is $1 - \exp(-\beta\tau)$, where $\beta > 0$. The parameter β is known the infection rate. We shall focus a single infected and a single susceptible individual and ask what the probability is that the susceptible individual becomes infected by time t . This probability has been studied in the case of a complete graph in Datta and Dorlas (2004). Here, we extend their results to a much wider class of graphs.

It is simplistic to consider just a single infective and a single susceptible individual. Nevertheless, insights gained from this setting are relevant in the “sparse” case, where the number of both infected and susceptible individuals is small and inter-contact times are fairly large. In that case, it is not a bad approximation to consider each pair of individuals in isolation. The “dense” setting will require quite different techniques and is not treated here.

Random walks on graphs Motivated by applications in Physics, Biology, Social Sciences and Computer Science, there has been an ever increasing interest in analysing the properties of interacting particles or agents moving on a given network (Draief et al. 2005). In particular, random walks play a central role in computer science, spanning a wide range of areas in both theory and practice, including distributed computing (Bui et al. 2004). In fact many distributed algorithms use random walks as a building block. Applications in networks include token management (Coppersmith et al. 1993), load balancing (Karger and Ruhl 2004), small-world routing (Draief and Ganesh 2006; Kleinberg 2000), search Gkantsidis et al. (2005), information propagation and gathering (Kempe et al. 2001), network topology monitoring (Ganesh et al. 2007) and group communication in ad-hoc networks (Dolev et al. 2006). The paper Lovász (1993) provides a survey of the properties of a random walk on a finite graph.

There has been fewer work related to multiple interacting mobile agents on a finite network. In Cooper et al. (2009) the authors propose a number of these dynamics and study their asymptotic properties on regular graphs. Aldous (1991) derives an upper bound for the expected meeting time of two independent copies of Markov chain as a function of the hitting time for a single chain. Coppersmith et al. (1993)

provide an upper bound for the expected meeting time of a variant of the problems of two random walks on a general graph whereby an adversary tries to keep the tokens apart. Finally, Dimitriou et al. (2006) addresses a similar problem to ours. The authors analyse the dynamics of interacting random walks $k - 1$ of which are healthy and one is infectious where the infection is transmitted as soon as an infectious and a healthy particle meet at given site. This model is related to the broadcast model presented in Cooper et al. (2009). They derive upper bounds for the time it takes to infect all the healthy particles using standard results on the coalescence time of multiple random walks (Aldous and Fill 2002).

Organisation of the paper The rest of the paper is organised as follows. In Section 2, we present our model and state our main results that relate the coincidence time of the two walkers to the stationary distribution of a (general) random walk on a graph which enables us to upper bound the probability of infection at a given time t . In Section 3, we analyse the case of the standard continuous random walk and instantiate our results for regular graphs, the Erdős–Rényi graph and power-law networks. In Section 4, we present a more detailed analysis for the complete graph and regular graphs in the case of $k \geq 2$ walkers providing tighter bounds than the ones derived in Dimitriou et al. (2006) for both models.

2 Model

We now describe the model precisely. We assume that the particles are mobile on a graph $G = (V, E)$ and that their movements are characterised by independent continuous-time Markov chains (CTMCs) on the finite state space $V = 1, \dots, n$, with the same transition rate matrix given by $Q = (q_{ij})_{i,j \in V}$ where $\sum_{j \in V} q_{ij} = 0$ for all $i \in V$. More precisely, we assume that if a given particle occupies node i then it jumps to node j at rate $q_{i,j} \geq 0$. We assume that the graph G is connected and that $q_{i,j} > 0$ if $(i, j) \in E$ (the edge (i, j) is present in the graph G) and 0 otherwise.

It is well known that if Q is the transition matrix of an ergodic Markov chain then there exists an invariant distribution π on V such that $\pi Q = (0, \dots, 0)$ and

$$\lim_{t \rightarrow \infty} \|\mathbb{P}(X_t \in \cdot \mid X_0 = i) - \pi\| = 0,$$

where

$$\|v - \eta\| = \frac{1}{2} \sum_{k \in V} |v_k - \eta_k| = \max_{A \subseteq V} |v_A - \eta_A|$$

is the *total variation distance* between v and η two probability measures on V .

From now on we assume that the Markov chain with transition rate matrix Q is reversible, i.e. $\pi_i q_{ij} = \pi_j q_{ji}$ (for a detailed account on reversible Markov chains see (Aldous and Fill 2002, Chapter 3). Considering $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$ the eigenvalues of $-Q$, and using (Ganesh et al. 2007, Lemma 1), we have that,

$$\|\mathbb{P}(X_t \in \cdot \mid X_0 = i) - \pi\| \leq \frac{1}{2\sqrt{\pi_i}} e^{-\lambda_2 t}. \quad (1)$$

We now consider two independent random walkers $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ moving between the nodes of the graph according the matrix Q where X_t and Y_t

correspond to the positions of the infected and susceptible individuals respectively at time t . A first natural object to study in this framework is the meeting time between the two walks defined by

$$T_M = \min\{t \geq 0, X_t = Y_t\}$$

which is finite regardless of the initial condition. In Aldous and Fill (2002, Proposition 5, Chapter 14), it is shown that the worst-case mean meeting time

$$\tau_M = \max_{i,j \in V} \mathbb{E}(T_M \mid X_0 = i, Y_0 = j) \leq \max_{i,j} \mathbb{E}(H_j \mid X_0 = i),$$

where H_j is the time it takes the (single) chain $(X_t)_t$ to hit node j starting from node i .

We define the coincidence time up to time t , denoted $\tau(t)$, as the total time up to t during which both walkers are at the same vertex, i.e.,

$$\tau(t) = \int_0^t \mathbf{1}_{(X_s=Y_s)} ds. \quad (2)$$

Let $I(t)$ denote the indicator function that the initial susceptible becomes infected by time t . Then, conditional on $\tau(t)$, we have

$$\mathbb{E}(I(t) \mid \tau(t)) = 1 - \exp(-\beta\tau(t)), \quad (3)$$

where $\beta > 0$ is the infection rate. Let $\gamma(t) = \mathbb{E}(I(t))$ be the probability that the initial susceptible becomes infected by time t .

We are interested in estimating the coincidence time $\tau(t)$ and the infection probability $\gamma(t)$ for different families of graphs. We consider the case when these chains are started independently in the stationary distribution and provide estimates of the coincidence time and an upper bound for the infection probability for general mobility models.

Theorem 1 Suppose X_0 and Y_0 are chosen independently according to the invariant distribution π . Then, we have

$$\begin{aligned} \mathbb{E}[\tau(t)] &= \sum_{v \in V} \pi_v^2 t \\ \gamma(t) &\leq 1 - \exp\left(-\beta t \sum_{v \in V} \pi_v^2\right). \end{aligned}$$

Proof Observe that, for all $s \geq 0$,

$$\begin{aligned} \mathbb{P}(X_s = Y_s) &= \sum_{v \in V} \mathbb{P}(X_s = Y_s = v) \\ &= \sum_{v \in V} \pi_v^2, \end{aligned}$$

because X_s and Y_s are independent, and are in stationarity. Hence, it is immediate from Eq. 2 that

$$\begin{aligned}\mathbb{E}[\tau(t)] &= \int_0^t \sum_{v \in V} \pi_v^2 ds \\ &= \sum_{v \in V} \pi_v^2 t.\end{aligned}$$

Next, taking expectations in Eq. 3 with respect to the conditioning random variable $\tau(t)$, we have

$$\begin{aligned}\gamma(t) &= 1 - \mathbb{E}[\exp(-\beta\tau(t))] \\ &\leq 1 - \exp(-\beta\mathbb{E}[\tau(t)]) \\ &= 1 - \exp\left(-\beta t \sum_{v \in V} \pi_v^2\right),\end{aligned}$$

where the inequality follows from Jensen's inequality. \square

We now introduce some terminology and define some examples of graph models that we shall consider. For two functions $f(\cdot)$ and $g(\cdot)$ on the natural numbers, we write $f(n) \sim g(n)$ to mean that their ratio tends to 1 as n tends to infinity. We write $f(n) = O(g(n))$ if $f(n)/g(n)$ remains bounded by a finite constant, $f(n) = o(g(n))$ if $f(n)/g(n)$ tends to zero, and $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$. For a sequence of events A_n indexed by $n \in \mathbb{N}$, we say that they occur with high probability (w.h.p.) if $\mathbb{P}(A_n)$ tends to 1 as n tends to infinity.

Example: uniform random walk If we consider the case where the rate transition matrix is given by

$$q_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E, \\ 0 & \text{if } (i, j) \notin E, \\ -d_i & \text{if } i = j \end{cases}$$

where d_i is the degree of i , i.e. the number of neighbours of node i in the graph G then it is known that the stationary distribution is the uniform distribution over V , $\pi_i = 1/n$ for all $i \in V$, and a direct application of Theorem 1 yields

$$\mathbb{E}[\tau(t)] = t/n, \quad \gamma(t) \leq 1 - e^{-\beta t/n}.$$

Moreover, the two walks need not start from the stationary distribution. In fact, it is not difficult to see that in this case, if the graph is connected, then the second eigenvalue of Q above is positive, and in fact bounded away from zero, for all n , using the fact that $x^T Q x = -\sum_{(i,j) \in E} (x_i - x_j)^2$. Therefore, after $O(\log(n))$, using Eq. 1, the distribution of the walks coincide with the uniform distribution w.h.p.

3 Standard continuous-time random walk

In this section, we choose a matrix Q in order to allow a non-uniform probability of being at a given node in stationary regime. More precisley, the particles perform

independent continuous-time Markov chains (CTMCs) on the finite state space V , with the same transition rate matrix given by

$$q_{ij} = \begin{cases} \frac{1}{d_i} & \text{if } (i, j) \in E, \\ 0 & \text{if } (i, j) \notin E, \\ -1 & \text{if } i = j \end{cases}$$

Moreover, they both initially start in positions distributed according to the invariant distribution π given by

$$\pi_i = \frac{d_i}{\sum_{v \in V} d_v}. \quad (4)$$

Note that if X_t and Y_t are not started from the stationary distribution, then by Eq. 1, after $\frac{c \log n}{\lambda_2}$, c constant but large, the distributions of X_t and Y_t coincide with the uniform distribution on V w.h.p. Moreover, by Cheeger's inequality (see Mohar 1997), we have that

$$\lambda_2 \geq \frac{\eta(G)^2}{2\Delta(G)},$$

where $\Delta(G)$ is the maximal degree of the nodes in G and $\eta(G)$ is the *isoperimetric constant* or *expansion* of a graph G defined by

$$\eta(G) = \inf_{U \subset V, |U| \leq n/2} \frac{E(U, U^c)}{|U|},$$

$E(U, U^c)$ denotes the number of edges having one vertex in U and the other in its complement, U^c (i.e., crossing the cut (U, U^c)); and $|U|$ the number of vertices or size of U . In particular if the graph is an expander, then the spectral gap λ_2 is large, and the (uniform) random walk is rapidly mixing (Mohar 1997). In particular, the infection does not take place, w.h.p., before the two walks approach the stationary distribution. In what follows, we will consider graphs which are expanders as illustrated in Ganesh et al. (2005) through the computation of the isoperimetric constant for Erdős–Rényi and power-law graphs and Friedman's proof Alon's second eigenvalue conjecture for regular graphs (Friedman 2003).

3.1 Examples

We now introduce a number of families of graphs of interest.

Complete graphs Consider the complete graph on n nodes, namely the graph in which there is an edge between every pair of nodes. Thus, $d_v = n - 1$ and $\pi_v = 1/n$ for all $v \in V$, so we have by Theorem 1 that $\mathbb{E}[\tau(t)] = t/n$. This result should be intuitive by symmetry. Lemma 1 also gives us an upper bound on the infection probability, $\gamma(t) \leq 1 - \exp(-\beta t/n)$. Roughly speaking, this says that it takes time of order n/β for the susceptible individual to become infected; for $t \ll n/\beta$, the probability of being infected is vanishingly small. Again, this is consistent with intuition.

Regular graphs A graph $G = (V, E)$ is said to be r -regular if $d_v = r$ for all $v \in V$. Thus, a complete graph is regular with $r = n - 1$. It is readily verified that $\pi_v = 1/n$

for all $v \in V$ if G is r -regular for any $r \geq 3$. Hence, if G is connected, we have the same estimates for $\tau(t)$ and $\gamma(t)$ as for the complete graph, which is a special case corresponding to $r = n - 1$.

The next examples we consider will be families of random graphs widely used in practice to model networks.

Erdős–Rényi random graphs The Erdős–Rényi graph $G(n, p)$ is defined as a random graph on n nodes, wherein each edge is present with probability p , independent of all other edges. We consider a family of such random graphs indexed by n , and take p to be a function of n chosen so that $np > c \log n$ for some constant $c > 1$. We also condition on the graph being connected. For p as above, the probability of connectivity tends to 1 as n tends to infinity, so conditioning on connectivity does not alter any of the estimates we shall derive later for the coincidence time on such graphs. In this model, the node degrees are identically distributed Binomial random variables with parameters $(n - 1, p)$. In particular, they concentrate around the mean value of np , and have exponentially decaying tails away from this value. Thus, while Erdős–Rényi graphs are not exactly regular, they exhibit considerable homogeneity in node degrees.

Power-law random graphs In contrast to the above graph models, many real-world networks exhibit considerable heterogeneity in node degrees, and have empirical degree distributions whose tails decay polynomially; see, e.g., Barabási and Albert (1999), Faloutsos et al. (1999). This observation has led to the development of generative models for graphs with power-law tails (Barabási and Albert 1999; Bollobás and Riordan 2004) as well as random-graph models possessing this property (Chung and Lu 2003). For definiteness, we work with the model proposed in Chung and Lu (2003), but we believe that similar results will hold for the other models as well.

In the model of Chung and Lu (2003), each node v is associated with a positive weight w_v , and edges are present independently with probabilities related to the weights by

$$\mathbb{P}((u, v) \in E) = \frac{w_u w_v}{W} \text{ where } W = \sum_{x \in V} w_x. \quad (5)$$

We assume that $W \geq w_{\max}^2$, so that the above defines a probability. It can be verified that $\mathbb{E}[d_v] = w_v$ and so this model is also referred to as the expected degree model. The model allows self-loops. The Erdős–Rényi graph $G(n, p)$ is a special case corresponding to the choice $w_v = np$ for all $v \in V$. If the weights are chosen to have a power-law distribution, then so will the node degrees. The following 3-parameter model for the ordered weight sequence is proposed in Chung and Lu (2003), parametrised by the mean degree d , the maximum degree m , and the exponent $\gamma > 2$ of the weight distribution:

$$w_i = m \left(1 + \frac{i}{i_0} \right)^{-\frac{1}{\gamma-1}}, \quad i = 0, 1, \dots, n-1, \quad (6)$$

where

$$i_0 = n \left(\frac{d(\gamma-2)}{m(\gamma-1)} \right)^{\gamma-1}. \quad (7)$$

Note that $W = \sum_{i=0}^{n-1} w_i \sim nd$.

We consider a sequence of such graphs indexed by n . The maximum expected degree m and the average expected degree d may, and indeed typically will, depend on n . In models of real networks, we can typically expect d to remain bounded or to grow slowly with n , say logarithmically, while m grows more quickly, say as some fractional power of n . In this paper, we only assume the following:

$$d \geq \delta > 0, \quad d = o(m), \quad m \leq \sqrt{nd}, \quad \frac{m}{d} = o\left(n^{\frac{1}{\gamma-1}}\right). \quad (8)$$

Here, δ is a constant that does not depend on n . In other words, the average expected degree is uniformly bounded away from zero. The third assumption simply restates the requirement that $w_0^2 \leq W$, so that Eq. 5 defines valid probabilities. The last assumption ensures that i_0 , defined in Eq. 7, tends to infinity.

3.2 Application to graphs

We now describe our results about these models.

Theorem 2 *Consider a sequence of graphs $G = (V, E)$ indexed by $n = |V|$. On each graph, consider two independent random walks with initial condition X_0, Y_0 chosen independently from the invariant distribution π for the random walk on that graph.*

We have $\mathbb{E}[\tau(t)] = t/n$ for regular graphs, including the complete graph, on n nodes.

For Erdős–Rényi random graphs $G(n, p)$ conditioned to be connected, and having $np \geq c \log n$ for some $c > 1$, we have $\mathbb{E}[\tau(t)] \sim t/n$, as n tends to infinity.

Finally, consider a sequence of power-law random graphs defined via Eqs. 5 and 6, and satisfying the assumptions in Eq. 8. Then, we have the following:

$$\frac{n\mathbb{E}[\tau(t)]}{t} \sim \begin{cases} c, & \text{if } \gamma > 3, \\ c(\log m), & \text{if } \gamma = 3, \\ c(m.d)^{3-\gamma}, & \text{if } 2 < \gamma < 3, \end{cases}$$

where $c > 0$ is a constant that may depend on γ , but not on n .

The theorem illustrates a discrepancy between the dynamics of the system of two particles for regular graphs, Erdős–Rényi random graphs and power-law graphs for $\gamma > 3$ on one hand, and power-law graphs $\gamma \leq 3$ on the other hand. For the former, the particles spend $O(t/n)$ amount of time together up to time t . In fact, as we will illustrate it in Section 4 for the complete and regular graphs, the healthy particle is infected w.h.p. before time $\frac{\beta+2}{\beta}n$. As for power-law graphs with $\gamma \leq 3$, the particles spend considerably more time together, at high degree nodes, and we anticipate that the infection occurs much faster.

The remainder of this section is dedicated to proving Theorem 2. If the graph G is regular, then, by Eq. 4, $\pi_v = 1/n$ for all $v \in V$. Hence, the claim of the theorem follows from Theorem 1. A more thorough analysis is provided in Section 4 for regular and complete graphs, where we present some initial findings for the case where there are k particles: one infected and $k - 1$ healthy.

In order to estimate $\mathbb{E}[\tau(t)]$, we need to compute

$$\sum_{v \in V} \pi_v^2 = \frac{\sum_{v \in V} d_v^2}{\left(\sum_{v \in V} d_v\right)^2}. \quad (9)$$

To this end, define

$$D = \sum_{v \in V} d_v = \sum_{u, v \in V} A_{uv}, \quad (10)$$

where $A_{uv} = \mathbf{1}_{((u,v) \in E)}$ is the indicator that u and v are connected in G , and

$$X_v = d_v(d_v - 1) = \sum_{i \neq j} A_{vi} A_{vj} \quad \text{and} \quad D_2 = \sum_{v \in V} X_v. \quad (11)$$

We will derive the first and second moments of the variables D and D_2 . It then suffices to use Chebyshev's inequality to establish concentration results for both variables D and D_2 . By Eq. 9 and Theorem 1, and the fact that $\sum_{v \in V} d_v^2 = D_2 + D$ and $D = \sum_{v \in V} d_v$, we will have an estimate of the coincidence time that holds whp.

We begin by computing the mean and variance of D in the expected degree model with arbitrary weight sequence $\{w_i, i = 0, \dots, n-1\}$.

For notational convenience, we define

$$\bar{w} = \frac{1}{n} \sum_{i=0}^{n-1} w_i, \quad \bar{w}^k = \frac{1}{n} \sum_{i=0}^{n-1} w_i^k, \quad k = 2, 3, \dots$$

We obtain Erdős–Rényi graphs $G(n, p)$ by setting $w_i = np$ for all i , and so, $\bar{w}^k = (np)^k$ for such graphs.

Next, consider power-law graphs with weight sequence specified by Eqs. 6 and 7. Since i_0 tends to infinity by assumption, we have for such graphs that

$$\begin{aligned} \bar{w}^k &= \frac{m^k}{n} \sum_{i=0}^{n-1} \left(1 + \frac{i}{i_0}\right)^{-\frac{k}{\gamma-1}} \\ &\sim \frac{m^k}{n} \int_0^n \left(1 + \frac{x}{i_0}\right)^{-\frac{k}{\gamma-1}} dx \\ &= m^k \frac{i_0}{n} \int_0^{n/i_0} (1+x)^{-\frac{k}{\gamma-1}} dx. \end{aligned} \quad (12)$$

Now, straightforward calculations yield that $\bar{w} \sim d$ for all $\gamma > 2$, whereas, for $k \geq 2$, we have

$$\bar{w}^k \sim \begin{cases} \frac{(\gamma-2)^k}{(\gamma-1)^{k-1}(\gamma-1-k)} d^k, & \text{if } \gamma > k+1, \\ \frac{(k-1)^k}{k^{k-1}} d^k \log \frac{m}{d}, & \text{if } \gamma = k+1, \\ \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(k+1-\gamma)} d^{\gamma-1} m^{k+1-\gamma}, & \text{if } 2 < \gamma < k+1. \end{cases} \quad (13)$$

We can now compute the mean and variance of D , the sum of node degrees.

Proposition 1 Consider a random graph $G = (V, E)$ specified by the expected degree model with an arbitrary weight sequence $\{w_v, v \in V\}$ satisfying $W \geq w_{\max}^2$, where

$W = \sum_{v \in V} w_v$. Let the sum of node degrees, D , be defined as in Eq. 10. Then, we have

$$\begin{aligned}\mathbb{E}[D] &= n\bar{w}, \\ \text{Var}(D) &= 2\left(n\bar{w} - \left(\frac{\bar{w}^2}{\bar{w}}\right)^2\right) - \left(\frac{\bar{w}^2}{\bar{w}} - \frac{\bar{w}^4}{n\bar{w}^2}\right),\end{aligned}\quad (14)$$

where $n = |V|$ is the total number of nodes.

In particular, for large n , if G is the Erdős–Rényi random graph $G(n, p)$, then

$$\begin{aligned}\mathbb{E}[D] &= n^2 p \\ \text{Var}(D) &= (2n - 1)np(1 - p) \sim 2n^2 p(1 - p),\end{aligned}\quad (15)$$

whereas, if G is a power law random graph satisfying the assumptions of Theorem 2, then

$$\mathbb{E}[D] = nd \quad \text{and} \quad \text{Var}(D) \sim 2nd.$$

The proof is provided in the [Appendix](#) of the paper. The following corollary is now an easy consequence of Chebyshev's inequality.

Corollary 1 *If G_n , $n \in \mathbb{N}$ is a sequence either of Erdős–Rényi random graphs or of power-law random graphs satisfying the assumptions of Theorem 2, then the sum of node degrees D concentrates at its expected value in the sense that $D \sim \mathbb{E}[D]$ w.h.p.*

We now establish a similar concentration result for the sum of squared degrees. To this end, recall that

$$\begin{aligned}X_v &= d_v(d_v - 1) = \sum_{i \neq j} A_{vi} A_{vj} \\ D_2 &= \sum_{v \in V} X_v.\end{aligned}$$

We have the following:

Proposition 2 *Let D_2 be defined as in Eq. 11. We then have*

$$\begin{aligned}\mathbb{E}[D_2] &= n\bar{w}^2 - \left(\frac{\bar{w}^2}{\bar{w}}\right)^2 \\ \text{Var}(D_2) &\leq 4n\bar{w}^3 + 2n\bar{w}^2 + 4n\frac{(\bar{w}^2)^2}{\bar{w}}.\end{aligned}\quad (16)$$

We now specialise the results to Erdős–Rényi and power-law random graphs, showing that D_2 concentrates near its expected value with high probability.

Proposition 3 *Suppose $G(n, p)$ is a sequence of Erdős–Rényi random graphs indexed by n (where p depends on n but this is not made explicit in the notation), and that np is uniformly bounded away from zero. Then*

$$D_2 \sim \mathbb{E}[D_2] \sim n^3 p^2, \quad \text{w.h.p.}$$

Proof We have, by Proposition 2 and the fact that $\overline{w^k} = (np)^k$ for the Erdős–Rényi random graph $G(n, p)$, that

$$\mathbb{E}[D_2] = n^2(n-1)p^2 \sim n^3 p^2, \quad \text{Var}(D_2) \leq 8n^4 p^3 + 2n^3 p^2.$$

Hence, by Chebyshev's bound, we obtain for all $\varepsilon > 0$ that,

$$\begin{aligned} \mathbb{P}(|D_2 - \mathbb{E}[D_2]| > \varepsilon \mathbb{E}[D_2]) &\leq \frac{\text{Var}(D_2)}{\varepsilon^2 \mathbb{E}[D_2]^2} \\ &\leq \frac{1}{\varepsilon^2(n-1)^2 p} + \frac{1}{\varepsilon^2 n(n-1)^2 p^2}. \end{aligned}$$

Now, by the assumption that np is bounded away from zero, $(n-1)^2 p$ and $n(n-1)^2 p^2$ tend to infinity as n tends to infinity. Thus, $\mathbb{P}(|D_2 - \mathbb{E}[D_2]| > \varepsilon \mathbb{E}[D_2])$ tends to zero for all $\varepsilon > 0$. This establishes the claim of the Proposition. \square

Using similar arguments, albeit with more involved computation, we derive an analogous result for power-law graphs. The proof is provided in the [Appendix](#) of the paper.

Proposition 4 Suppose $G_n, n \in \mathbb{N}$ is a sequence of power-law random graphs satisfying the assumptions in Theorem 2, with $\gamma > 2$. Then, $D_2 \sim \mathbb{E}[D_2]$ whp, and

$$\mathbb{E}[D_2] \sim \begin{cases} cnd^2, & \text{if } \gamma > 3, \\ cnd^2(\log m), & \text{if } \gamma = 3, \\ cnd^{\gamma-1}m^{3-\gamma}, & \text{if } 2 < \gamma < 3, \end{cases}$$

To complete the proof of Theorem 2, it suffices to combine the results of Lemmas 1 and 4 with the fact that $\sum_v \pi_v^2 = \frac{D_2 + D}{D^2}$.

4 Further analysis of the complete and regular graphs

In this section, we focus our attention on the complete graph for which the uniform and the standard continuous-time random walks coincide. Due to the absence of topology, i.e. a walk can jump to any other node uniformly at random, we can explore this model in more detail. First, let us analyze the case of two random walks, one being infectious and the other healthy. The dynamic of the system can be described through a three state Markov chain: state 1 corresponds to the case where the two walks are apart, the state 2 corresponds to when the two walks coincide, and state 3 corresponds to the infection of the healthy walk by the infected one. The rate matrix of the corresponding continuous-time Markov chain is then given by

$$Q = \begin{pmatrix} \frac{-2}{n-1} & \frac{2}{n-1} & 0 \\ \frac{2(n-2)}{n-1} - \left(\beta + \frac{2(n-2)}{n-1}\right) & \beta & \beta \\ 0 & 0 & 0 \end{pmatrix},$$

and the corresponding transition matrix $P_t = e^{tQ}$. The eigenvalues of Q are 0, $-\lambda_1$ and $-\lambda_2$ with

$$\lambda_1 = \frac{1}{2} \left(\beta + 2 + \sqrt{(\beta + 2)^2 - 8\beta/(n-1)} \right)$$

$$\lambda_2 = \frac{1}{2} \left(\beta + 2 - \sqrt{(\beta + 2)^2 - 8\beta/(n-1)} \right),$$

with eigenvectors $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 - (n-1)\lambda_1/2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 - (n-1)\lambda_2/2 \\ 0 \end{pmatrix}$, respectively.

The infection probability after time t is therefore given by

$$\begin{aligned} \gamma(t) &= (P_t)_{1,3} = (1 \ 0 \ 0) e^{tQ} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= 1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_1 t} - \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-\lambda_2 t} \\ &\sim 1 - \exp\left(-\frac{2}{(\beta + 2)^2} \frac{\beta t}{n}\right) \end{aligned}$$

yielding a tighter bound than the one derived using Theorem 1.

Let $m_3(i)$ be the mean hitting time of state 3, starting from state i , for $i = 1, 2$. Note that $m_3(1)$ is the mean time for the healthy node to get infected. By the first-step analysis (Norris 1997, Theorem 3.3.3), we have

$$\begin{aligned} \frac{2}{n-1} m_3(1) &= \frac{2}{n-1} m_3(2) + 1 \\ \left(\beta + \frac{2(n-2)}{n-1} \right) m_3(2) &= \frac{2(n-2)}{n-1} m_3(1) + 1. \end{aligned}$$

From this, it follows, that the average time to infection $\mathbb{E}T_{\text{Inf}}$ is given by

$$\mathbb{E}(T_{\text{Inf}}) = m_3(1) \sim \frac{\beta + 2}{2\beta} n. \quad (17)$$

To conclude this section we provide a first analysis of the case with k random particles $k \geq 2$ where initially one particle is infectious and the remaining $k - 1$ are healthy. In what follows we provide an upper bound on the time to infection of the whole population. First it is easy to note that at time t a given particle can occupy any of the n nodes of the graph. Therefore by combining Eq. 17 and the union bound, the average time to infection $\mathbb{E}(T_{\text{Inf}}^{(k)})$ is bounded by

$$\mathbb{E}(T_{\text{Inf}}^{(k)}) \leq \sum_{j=2}^k (j-1) \frac{\beta + 2}{2\beta} n = \frac{\beta + 2}{4\beta} k(k-1)n.$$

We now derive a better bound. First, note that at any given time, on the complete graph, each of the particles is located uniformly at random on the graph. If we assume that there are i infectious particles and $(k - i)$ healthy ones, then for $k \leq \sqrt{n}$, it is not difficult to show that the probability that there is at least an encounter between one

healthy and one infectious particles is equivalent to $\frac{i(k-i)}{n}$. To simplify the analysis, we assume that if there are healthy and infectious particles located at the same node then there is at most one infection that occurs before any of the k particles jumps to another node which happens with probability $\frac{\beta}{\beta+k}$.

The above assumption slows down the infection process and thus yields an upper bound. It is not difficult to note that the time until there is a new particle infected is stochastically dominated by $\sum_{i=1}^N E_i$ where E_i are i.i.d. r.v. distributed according to an exponential distribution with parameter k and N is a geometric random variable with parameter $\frac{\beta}{\beta+k} \frac{i(k-i)}{n}$. Hence,

$$\begin{aligned} \mathbb{E} \left(T_{\text{Inf}}^{(k)} \right) &\leq \frac{(\beta+k)n}{\beta} \sum_{i=1}^{k-1} \frac{1}{i(k-i)} \\ &= \frac{(\beta+k)n}{\beta k} \sum_{i=1}^{k-1} \left(\frac{1}{i} + \frac{1}{k-i} \right) \\ &\sim \frac{2(\beta+k)n \log(k)}{\beta k}, \end{aligned}$$

and for $k = n^{1/2-\epsilon}$, we have $\mathbb{E} \left(T_{\text{Inf}}^{(k)} \right) \leq \frac{2n \log(k)}{\beta}$.

The above analysis can be extended to the case of r -regular graphs ($r \geq 3$), using Cooper et al. (2009, Theorem 22), one can derive that the probability of a meeting between a healthy and an infectious particle, at the stage where there are i infectious particles is asymptotically (in n) equal to $\frac{r-2}{r-1} \frac{i(k-i)}{n}$ yielding the following result.

Theorem 3 *The time to infection, starting from 1 infected particle and $k-1$ healthy ones, all performing standard continuous-random walks on an r -regular graph, $r \geq 3$ and $k = o(\sqrt{n})$, satisfies*

$$\mathbb{E} \left(T_{\text{Inf}}^{(k)} \right) \leq 2 \frac{r-2}{r-1} \frac{\beta+k}{\beta k} n \log(k).$$

Appendix

Proof of Proposition 1 It is immediate from Eq. 10 that

$$\begin{aligned} \mathbb{E}[D] &= \sum_{u,v \in V} \mathbb{P}((u,v) \in E) \\ &= \sum_{u,v \in V} \frac{w_u w_v}{W} = W, \end{aligned}$$

which establishes the first equality in Eq. 14. Next, rewrite Eq. 10 as

$$D = 2 \sum_{i=1}^n \sum_{j=i+1}^n A_{ij} + \sum_{i=1}^n A_{ii},$$

and observe from the independence of the edges that

$$\begin{aligned}\mathrm{Var}(D) &= 4 \sum_{i=1}^n \sum_{j=i+1}^n \mathrm{Var}(A_{ij}) + \sum_{i=1}^n \mathrm{Var}(A_{ii}) \\ &= 2 \sum_{i=1}^n \sum_{j=1}^n \mathrm{Var}(A_{ij}) - \sum_{i=1}^n \mathrm{Var}(A_{ii}).\end{aligned}$$

Now, $\mathrm{Var}(A_{uv}) = \mathbb{P}((u, v) \in E)(1 - \mathbb{P}((u, v) \in E))$, and so,

$$\mathrm{Var}(D) = 2 \sum_{i=1}^n \sum_{j=1}^n \left(\frac{w_i w_j}{W} - \frac{w_i^2 w_j^2}{W^2} \right) - \sum_{i=1}^n \left(\frac{w_i^2}{W} - \frac{w_i^4}{W^2} \right).$$

Upon simplifying, this yields the second equality in Eq. 14. Now, using the fact that $\overline{w^k} = (np)^k$ for Erdős–Rényi graphs $G(n, p)$, we readily obtain Eq. 15.

Next, suppose G is a power-law graph (more precisely, G_n is a sequence of power law graphs) satisfying the assumptions of Theorem 2. It follows from Eq. 12 that

$$\overline{w^4} \sim \begin{cases} \frac{(\gamma-2)^4}{(\gamma-1)^3(\gamma-5)} d^4, & \text{if } \gamma > 5, \\ \frac{81}{64} d^4 \log \frac{m}{d}, & \text{if } \gamma = 5, \\ \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(5-\gamma)} d^{\gamma-1} m^{5-\gamma}, & \text{if } 2 < \gamma < 5, \end{cases} \quad (18)$$

while

$$\overline{w^2} \sim \begin{cases} \frac{(\gamma-2)^2}{(\gamma-1)(\gamma-3)} d^2, & \text{if } \gamma > 3, \\ \frac{1}{2} d^2 \log \frac{m}{d}, & \text{if } \gamma = 3, \\ \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(3-\gamma)} d^{\gamma-1} m^{3-\gamma}, & \text{if } 2 < \gamma < 3, \end{cases} \quad (19)$$

and $\overline{w} \sim d$ for all $\gamma > 2$.

By Eq. 14, it suffices to show that

$$\left(\frac{\overline{w^2}}{\overline{w}} \right)^2 = o(nd) \text{ and } \frac{\overline{w^4}}{\overline{w}^2} = o(n^2 d) \quad (20)$$

in order to show that $\mathrm{Var}(D) \sim 2n\overline{w} \sim 2nd$.

Suppose first that $\gamma > 3$. Then, by Eq. 19 and the fact that $\overline{w} = nd$,

$$\frac{1}{nd} \left(\frac{\overline{w^2}}{\overline{w}} \right)^2 = O\left(\frac{d}{n}\right) = o(1),$$

where the last equality follows from Eq. 8, and the fact that $d \leq n$.

Now let $\gamma = 3$. Then, by Eq. 19 and the fact that $\overline{w} = nd$,

$$\frac{1}{nd} \left(\frac{\overline{w^2}}{\overline{w}} \right)^2 = O\left(\frac{d}{n} \log \frac{m}{d}\right) = O\left(\frac{m}{n} \frac{d}{m} \log \frac{m}{d}\right) = o(1),$$

where the last equality follows by Eq. 8. On the other hand, if $2 < \gamma < 3$, then, by Eq. 19,

$$\begin{aligned} \frac{1}{nd} \left(\frac{\overline{w^2}}{\overline{w}} \right)^2 &= O\left(\frac{d^{2\gamma-5} m^{6-2\gamma}}{n} \right) \\ &= O\left(\left(\frac{d}{n} \right)^{\gamma-2} \right) = o(1), \end{aligned}$$

where we have used the inequality $m \leq \sqrt{nd}$ from Eq. 8 to obtain the second equality. To obtain the last equality, note that it follows from Eq. 8 that $m = o(n)$ and hence that $d = o(n)$ as well. We have thus established the first equality in Eq. 20 for all $\gamma > 2$. The proof of the second equality is similar and is omitted. This completes the proof of the proposition. \square

Proof of Proposition 2 We first note that

$$\begin{aligned} \mathbb{E}[X_v] &= \sum_{i \neq j} \frac{w_i w_j w_v^2}{W^2} = w_v^2 \left(1 - \frac{1}{W^2} \sum_{i \in V} w_i^2 \right) \\ &= w_v^2 \left(1 - \frac{\overline{w^2}}{n \overline{w^2}} \right). \end{aligned}$$

Therefore,

$$\mathbb{E}[D_2] = \sum_{v \in V} \mathbb{E}[X_v] = n \overline{w^2} - \left(\frac{\overline{w^2}}{\overline{w}} \right)^2,$$

which is the first part of Eq. 16. Next, for distinct nodes $u, v \in V$, we have

$$\begin{aligned} \text{Cov}(X_u, X_v) &= \sum_{i \neq j} \sum_{k \neq l} \text{Cov}(A_{iu} A_{ju}, A_{kv} A_{lv}) \\ &= 4 \sum_{i \neq v, l \neq u} \text{Cov}(A_{iu} A_{uv}, A_{uv} A_{lv}) \\ &= 4 \mathbb{E}[A_{u,v}] (1 - \mathbb{E}[A_{u,v}]) \sum_{i \neq v} \mathbb{E}[A_{iu}] \sum_{l \neq u} \mathbb{E}[A_{lv}]. \end{aligned}$$

The second equality above holds because, by the independence of edges, the indicator random variables $A_{iu} A_{ju}$ and $A_{kv} A_{lv}$ corresponding to the open triangles (or 2-stars) $iu j$ and $kv l$ are independent unless two of the edges are the same; the only way this can happen is if (u, v) is a common edge and there are 4 possible node labellings corresponding to each such edge set. Now, recall that $\mathbb{E}[A_{u,v}] = w_u w_v / W$ and $\sum_i \mathbb{E}[A_{iu}] = \mathbb{E}[d_u] = w_u$. Hence, we see from the above that

$$0 \leq \text{Cov}(X_u, X_v) \leq 4 \frac{w_u^2 w_v^2}{W}. \quad (21)$$

Similarly, we obtain

$$\begin{aligned}\mathrm{Var}(X_u) &= \sum_{i \neq j} \sum_{k \neq l} \mathrm{Cov}(A_{iu} A_{ju}, A_{ku} A_{lu}) \\ &= 4 \sum_j \sum_{i \neq j} \sum_{l \neq i, j} \mathrm{Cov}(A_{iu} A_{ju}, A_{ju} A_{lu}) + 2 \sum_{i \neq j} \mathrm{Var}(A_{iu} A_{ju}) \\ &\leq 4 \sum_j \sum_{i \neq j} \sum_{l \neq i, j} \mathbb{E}[A_{iu} A_{ju} A_{lu}] + 2 \sum_{i \neq j} \mathbb{E}[A_{iu} A_{ju}].\end{aligned}$$

Using the fact that distinct edges are independent, we get

$$\mathrm{Var}(X_u) \leq 4w_u^3 + 2w_u^2. \quad (22)$$

Now, by Eqs. 11, 21 and 22

$$\begin{aligned}\mathrm{Var}(D_2) &= \sum_{u \in V} \mathrm{Var}(X_u) + \sum_{u \neq v} \mathrm{Cov}(X_u, X_v) \\ &\leq \sum_{u \in V} (4w_u^3 + 2w_u^2) + \sum_{u, v \in V} 4 \frac{w_u^2 w_v^2}{W}.\end{aligned}$$

Computing the above sums yields the second part of Eq. 16. \square

Proof of Proposition 4 We will show that $\mathrm{Var}(D_2) = o(\mathbb{E}[D_2]^2)$, so that the claim follows by Chebyshev's bound, as in the proof of the previous lemma. We will consider separately the parameter ranges $\gamma \geq 4$, $3 \leq \gamma < 4$ and $2 < \gamma < 3$, where γ is the exponent in the power law describing the degree distribution.

In the following, c_1, c_2, \dots will denote generic positive constants, not necessarily the same from line to line. Recall that $\bar{w} \sim d$.

Suppose first that $\gamma \geq 4$. Then, by Eq. 13, $\bar{w}^3 = O(d^3 \log \frac{m}{d})$ and $\bar{w}^2 \sim c_1 d^2$. Therefore, by Lemma 2,

$$\mathbb{E}[D_2] \sim c_1 n d^2$$

and

$$\mathrm{Var}(D_2) = O\left(nd^3 \log \frac{m}{d} + nd^2\right) = O\left(nd^3 \log \frac{m}{d}\right),$$

where the last equality holds because of the assumption in Eq. 8 that $d \geq \delta$ for some constant $\delta > 0$. Thus, we see that

$$\frac{\mathrm{Var}(D_2)}{\mathbb{E}[D_2]^2} = O\left(\frac{1}{nd} \log \frac{m}{d}\right) = o(1),$$

since $m \leq n$.

Suppose next that $3 \leq \gamma < 4$. Then, by Eq. 13, $\bar{w}^3 = O(d^{\gamma-1} m^{4-\gamma})$, while $\bar{w}^2 \sim c_1 d^2$ if $3 < \gamma < 4$ and $\bar{w}^2 \sim c_2 d^2 \log \frac{m}{d}$ if $\gamma = 3$. Therefore, by Lemma 2,

$$\begin{aligned}\mathbb{E}[D_2] &\geq c_1 n d^2 - c_2 d^2 \log^2 \frac{m}{d} \\ &\geq c_1 n d^2 - c_2 d^2 \log^2 n = \Omega(nd^2),\end{aligned} \quad (23)$$

whereas,

$$\begin{aligned}\text{Var}(D_2) &\leq c_1 n d^{\gamma-1} m^{4-\gamma} + c_2 n d^2 \log \frac{m}{d} + c_3 n d^3 \log^2 \frac{m}{d} \\ &\leq c_1 n d^{\gamma-1} m^{4-\gamma} + c_2 n d^3 \log^2 \frac{m}{d}, \\ &= c_1 n d^{\gamma-1} m^{4-\gamma} \left(1 + \left(\frac{d}{m} \right)^{4-\gamma} \log^2 \frac{m}{d} \right).\end{aligned}$$

We have used the assumption that d is uniformly bounded away from zero to obtain the second inequality above. Since we also assumed in Eq. 8 that $d = o(m)$, we have

$$(d/m)^{4-\gamma} \log^2(m/d) = o(1).$$

for all $\gamma < 4$. Hence, $\text{Var}(D_2) = O(n d^{\gamma-1} m^{4-\gamma})$. Combining this with Eq. 23, we get

$$\frac{\text{Var}(D_2)}{\mathbb{E}[D_2]^2} = O\left(\frac{1}{nd} \left(\frac{m}{d}\right)^{4-\gamma}\right) = O\left(\frac{1}{nd} n^{(4-\gamma)/(\gamma-1)}\right) = o(1).$$

We have used Eq. 8 to obtain the second equality above and the fact that $\gamma \geq 3$ to obtain the last equality. Moreover, $\mathbb{E}[D_2] \sim c n d^2$ for $3 < \gamma < 4$, whereas $\mathbb{E}[D_2] \sim c n d^2 \log(m)$ for $\gamma = 3$.

Finally, suppose that $2 < \gamma < 3$. Then, by Eq. 13, $\overline{w^3} = O(d^{\gamma-1} m^{4-\gamma})$ and $\overline{w^2} \sim c_1 d^{\gamma-1} m^{3-\gamma}$, so that, by Lemma 2,

$$\begin{aligned}\mathbb{E}[D_2] &\geq c_1 n d^{\gamma-1} m^{3-\gamma} - c_2 (d^{\gamma-2} m^{3-\gamma})^2 \\ &\geq c_1 n d^{\gamma-1} m^{3-\gamma} \left(1 - \frac{c_2}{n} \left(\frac{m}{d} \right)^{3-\gamma} \right).\end{aligned}$$

Now, by Eq. 8, $(m/d)^{3-\gamma} = o(n^{(3-\gamma)/(\gamma-1)}) = o(n)$ since $\gamma > 2$. Consequently,

$$\mathbb{E}[D_2] = \Omega(n d^{\gamma-1} m^{3-\gamma}).$$

On the other hand,

$$\begin{aligned}\text{Var}(D_2) &\leq c_1 n d^{\gamma-1} m^{4-\gamma} + c_2 n d^{\gamma-1} m^{3-\gamma} \\ &\quad + c_3 n d^{2\gamma-3} m^{6-2\gamma} \\ &\leq c_1 n d^{\gamma-1} m^{4-\gamma} \left(1 + \frac{c_2}{m} + c_3 \left(\frac{d}{m} \right)^{\gamma-2} \right) \\ &= O(n d^{\gamma-1} m^{4-\gamma}).\end{aligned}$$

Hence,

$$\begin{aligned}\frac{\text{Var}(D_2)}{\mathbb{E}[D_2]^2} &= O\left(\frac{n d^{\gamma-1} m^{4-\gamma}}{n^2 d^{2\gamma-2} m^{6-2\gamma}}\right) \\ &= O\left(\frac{1}{nm} \left(\frac{m}{d}\right)^{\gamma-1}\right)\end{aligned}$$

Now, by Eq. 8, and the fact that $\gamma > 2$ we have $(m/d)^{\gamma-1} = o(n)$. Since the maximum degree m is assumed to grow as a power of n , we have $\frac{\text{Var}(D_2)}{\mathbb{E}[D_2]^2} = o(1)$. Note that $\mathbb{E}[D_2] \sim c n d^{\gamma-1} m^{3-\gamma}$, for $2 < \gamma < 3$.

Using Chebyshev's inequality, this establishes the claim of the proposition. \square

References

- Aldous D (1991) Meeting times for independent Markov chains. *Stoch Process Their Appl* 38:185–193
- Aldous D, Fill J (2002) Reversible Markov chains and random walks on graphs. Monograph available at <http://www.stat.berkeley.edu/~aldous/RWG>
- Amini H, Draief M, Lelarge M (2009) Marketing in a random network. In: *Network Central and Optimization, LNCS, NET-COOP'08*. Springer, Berlin, pp 17–25
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
- Bollobás B, Riordan O (2004) “The diameter of a scale-free random graph. *Combinatorica* 4: 5–34
- Bui M, Bernard T, Sohier D, Bui A (2004) Random walks in distributed computing: a survey. In: *Proc. IICS*, 1–14
- Chaintreau A, Hui P, Scott J, Gass R, Crowcroft J, Diot C (2007) Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans Mob Comput* 6(6):606–620
- Chung F, Lu L (2003) The average distances in random graphs with given expected degrees. *Internet Math* 1:91–114
- Cooper C, Frieze A, Radzik T (2009) Multiple random walks in random regular graphs. *SIAM J Discrete Math* 22(4):1738–1761
- Coppersmith D, Tetali P, Winkler P (1993) Collisions among random walks on a graph. *SIAM J Discrete Math* 6(3):363–374
- Daley DJ, Gani J (2001) Epidemic modelling: an introduction. Cambridge University Press, *Studies in mathematical biology*
- Datta N, Dorlas TC (2004) Random walks on a complete graph: a model for infection. *J Appl Probab* 41:1008–1021
- Dimitriou T, Nikolettseas S, Spirakis P (2006) The infection time of graphs. *Discrete Appl Math* 154:2577–2589
- Dolev S, Schiller E, Welch J (2006) Random walk for self-stabilizing group communication in ad hoc networks. *IEEE Trans Mob Comput* 5(7):893–905
- Draief M (2006) Epidemic processes on complex networks. *Physica A: Statistical Mechanics and its Applications* 363(1):120–131
- Draief M, Ganesh A (2006) Efficient routing in poisson small-world networks. *J Appl Probab* 43(3):678–686
- Draief M, Mairesse J, O’Connell, N (2005) Queues, stores, and tableaux. *J Appl Probab* 42(4):1145–1167
- Draief M, Massoulié L (2010) Epidemics and rumours in complex networks. London mathematical society series, vol 369. Cambridge University Press
- Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the Internet topology. In: *Proceedings ACM Sigcomm*
- Friedman J (2003) A proof of Alons second eigenvalue conjecture. In: *Proceedings of the thirty-fifth annual ACM symposium on theory of computing (STOC’03)*, pp 720–724
- Ganesh A, Draief M (2009) A random walk model for infection on graphs. *ACM Proc. Valuetools’09*, article 34
- Ganesh AJ, Kermarrec A-M, Le Merrer E, Massoulié L (2007) Peer counting and sampling in overlay networks based on random walks. *Distrib Comput* 20(4):267–278
- Ganesh A, Massoulié L, Towsley D (2005) The effect of network topology on the spread of epidemics. In: *Proc. INFOCOM*, pp 1455–1466
- Gkantsidis C, Mihail M, Saberi A (2005) Hybrid search schemes for unstructured peer-to-peer networks. In: *Proc. INFOCOM*, pp 1526–1537
- Karger DR, Ruhl M (2004) Simple efficient load balancing algorithms for peer-to-peer systems. In: *Proc. SPAA*, pp 36–43
- Kempe D, Kleinberg J, Demers A (2001) Spatial gossip and resource location protocols. In: *Proc. STOC*, pp 163–172
- Kephart J, White S (1991) Directed-graph epidemiological models of computer viruses. In: *Proceedings of the IEEE computer symposium on research in security and privacy*, pp 343–359
- Kleinberg J (2007) The wireless epidemic. *Nature* 449:287–288
- Kleinberg J (2000) The small-world phenomenon: an algorithm perspective. In: *Proc. STOC*, pp 163–170
- Leavitt N (2005) Mobile phones: the next frontier for hackers? *IEEE Computer Society Press*, Los Alamitos, CA, vol 38, no 4, pp 20–23

- Lovász L (1993) Random Walks on Graphs: a survey. In: Combinatorics, Bolyai Society Mathematical Studies vol 2, pp 1–46
- Mickens JW, Noble BD (2005) Modeling epidemic spreading in mobile environments. In: Proceedings of the 4th ACM workshop on Wireless security, pp 77–86
- Mohar B (1997) Some applications of Laplace eigenvalues of graphs. In: Hahn G, Sabidussi G (eds) Graph symmetry. Kluwer Academic Press, Dordrecht, pp 225–275
- Mobile phone virus Cabir. <http://www.dancewithshadows.com/tech/mobile-phone-virus-cabir.asp>
- Nekovee M (2007) Worm epidemics in wireless ad hoc networks. New J Phys 9:189
- Norris J (1997) Markov chains. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press
- Pittel B (1987) On spreading a rumor. SIAM J Appl Math 47(1):213–223
- Rhodes CJ, Nekovee M (2008) The opportunistic transmission of wireless worms between mobile devices. Physica A: Statistical Mechanics and Its Applications 387(27):6837–6844
- Su J, Chan KKW, Miklas AG, Po K, Akhavan A, Saroiu S, de Lara E, Goel A (2006) A preliminary investigation of worm infections in a bluetooth environment. In: Proceedings of the 4th ACM workshop on Recurring malware, WORM'06, pp 9–16
- Wang P, González M, Hidalgo C, Baranási A-L (2009) Understanding the spreading patterns of mobile phone viruses. Science 324(5930):1071–1076
- Yan G, Cuellar L, Eindhoven S, Flores H, Hengartner N, Vu V (2007) Bluetooth worm propagation: mobility pattern matters! In: Proceedings of the 2nd ACM symposium on Information, computer and communications security, ASIACCS'07, pp 32–44



Moez Draief graduated from the Ecole Polytechnique (Paris) in 2000. He then completed a DEA in Probability Theory at the University Paris VI. He undertook a PhD in the LIAFA (Theoretical Computer Science Group), University Paris VII.

From October 2004 to January 2007, he was a Marie Curie research fellow at the Statistical Laboratory and a lecturer in Part III (Certificate of Advanced Study in Mathematics), Cambridge University.

His research interests are: applied probability and discrete mathematics, queueing theory and stochastic networks, and the application of these to the analysis of distributed algorithms and complex networks such as Peer-to-Peer and ad hoc networks.



Ayalvadi Ganesh received his BTech in EE from IIT Madras in 1988, MS and PhD in EE from Cornell University in 1991 and 1995 respectively. His PhD thesis was on the use of large deviation techniques in queueing theory. He was with Edinburgh University, Birkbeck College, London, UK, and Hewlett–Packards Basic Research Institute in Mathematical Sciences (BRIMS) and Microsoft Research before joining the Mathematics Department of Bristol University. He was also a Fellow of Kings College, Cambridge, from 2000 to 2004.

His research interests are in the mathematical modelling of communication and computer networks, and in decentralised algorithms for such networks. Specific interests include large deviations and applications to queueing theory and statistics, random graph models and stochastic processes on graphs, and decentralised algorithms for resource allocation in the Internet and in wireless networks.