A random walk model for infection on graphs [*]

A. Ganesh
Department of Mathematics
University of Bristol
University Walk, Clifton
Bristol BS8 1TW, U.K.
A.Ganesh@bristol.ac.uk

M. Draief
Department of Electrical and Electronic
Engineering
Imperial College London
South Kensington Campus
London SW7 2AZ, U.K.
M.Draief@imperial.ac.uk

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 graph including regular graphs and the classical Erdös-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.

Keywords

random graphs, complex networks, communication systems, wireless networks.

1. INTRODUCTION

In recent years there have 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 lauching computer-virus outbreaks that take advantage of human mobility [2, 22, 24].

Over the past couple years, there have been indeed reports of malicious code that take advantage of bluetooth vulnerabilities such as the Cabir worm that was detected during the World Athletics Championship [25] and another at a company that has been reported by CommWarrior [26]. Despite their small scales, these incidents bode more threats taking advantage of events and locations where individuals gather in close proximity [32, 33].

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 an number of random individuals chosen uniformly at random among the population [10]. Recently there have been an increasing interest in understanding the impact of the network topology on the spread of epidemics in networks with fixed nodes, see [14, 17].

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 relies on the use of a non-rigorous mean-field approximations (similar to the classical Kephart-White model [21]) that incorporates the mobility patterns of users. In [30], 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 et al. [28, 29] 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 [6] 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 others range. As in the case of static networks [31], 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. 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: they stay at each vertex for an exponentially distributed time with unit mean, and then move to a neighbour of that vertex chosen uniformly at random. The infection can pass from an infected to a susceptible individual only if they are both at the same vertex, and the probability of its being passed over a time interval of length τ is $1 - \exp(-\beta\tau)$, where $\beta > 0$ is a parameter called the infection rate. We shall consider 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 [11]. 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 finite network. 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 [5]. In fact many distributed algorithms use random walks

as a building block. Applications in networks include token management [9], load balancing [19], small-world routing [13, 23], search [18], information propagation and gathering [20], network topology monitoring [16] and group communication in ad-hoc networks [12]. The paper [27] provides a survey of the properties of a random walk on a finite graph.

There have been fewer work related to multiple interacting mobile agents on a finite network. In [8] the authors propose a number of these dynamics and study their asymptotic properties on regular graphs. Aldous [1] 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 [9] 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.

Organisation of the paper. The rest of the paper is organised as follows. In Section 2, we present our model and the family of networks we will consider. Besides we state our main results that relate the coincidence time of the two walkers to the stationary distribution of a random walk on a graph. In Section 3 we give a detailed proof of our main result on the probability of infection for regular graphs, the Erdös-Rényi graph and power-law networks. Section 4 summarises our contribution and suggests further extensions of our work.

2. MODELS AND RESULTS

We now describe the model precisely. Let $X_t, Y_t \in V$ denote the positions of the susceptible and infected individuals respectively at time t. We model $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ as independent continuous-time Markov chains (CTMCs) on the finite state space V, with the same transition rate matrix given by

$$q_{xy} = \begin{cases} \frac{1}{\text{degree}(x)} & \text{if } (x, y) \in E, \\ 0 & \text{if } y \neq x \text{ and } (x, y) \notin E, \\ -1 & \text{if } y = x \end{cases}$$

where $\text{degree}(x)$ is the number of neighbours of x (nodes y such that $(x, y) \in E$) in the graph G.

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.,

$$\tau(t) = \int_0^t 1(X_s = Y_s)ds. \tag{1}$$

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

$$\mathbb{E}(I(t) \mid \tau(t)) = 1 - \exp(-\beta\tau(t)), \tag{2}$$

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.

Observe that the Markov chains X_t, Y_t have invariant distribution π given by

$$\pi_x = \frac{\text{degree}(x)}{\sum_{v \in V} \text{degree}(v)} \qquad (3)$$

and that they are reversible, i.e., $\pi_x q_{xy} = \pi_y q_{yx}$ for all $x, y \in V$.

We consider the case when these chains are started independently in the stationary distribution and provide estimates on the coincidence time and the infection probability, for arbitrary graphs.

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

$$\mathbb{E}[\tau(t)] = \sum_{v \in V} \pi_v^2 t$$

$$\gamma(t) \leq 1 - \exp\left(-\beta t \sum_{v \in V} \pi_v^2\right).$$

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

$$\mathbb{P}(X_s = Y_s) = \sum_{v \in V} \mathbb{P}(X_s = Y_s = v)$$

$$= \sum_{v \in V} \pi_v^2,$$

because X_s and Y_s are independent, and are in stationarity. Hence, it is immediate from (1) that

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

$$= \sum_{v \in V} \pi_v^2 t.$$

Next, taking expectations in (2) with respect to the conditioning random variable $\tau(t)$, we have

$$\gamma(t) = 1 - \mathbb{E}[\exp(-\beta \tau(t))]$$

$$\leq 1 - \exp(-\beta \mathbb{E}[\tau(t)])$$

$$= 1 - \exp\left(-\beta t \sum_{v \in V} \pi_v^2\right),$$

where the inequality follows from Jensen's inequality. □

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 (whp) if $\mathbb{P}(A_n)$ tends to 1 as n tends to infinity.

Examples

Complete graphs. Consider the complete graph on n nodes, namely the graph in which there is an edge between every pair of nodes. Thus, $\text{degree}(v) = n - 1$ 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 $\text{degree}(v) = r$ 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 $r \geq 2$. (If $r = 1$, then G is a matching and is not connected.) 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., [3, 15]. This observation has led to the development of generative models for graphs with power-law tails [3, 4] as well as random-graph models possessing this property [7]. For definiteness, we work with the model proposed in [7], but we believe that similar results will hold for the other models as well.

In the model of [7], each node v is associated with a positive weight w_v, and edges are present independently with probabilities related to the weights by

$$\mathbb{P}((u, v) \in E) = \frac{w_u w_v}{W} \text{ where } W = \sum_{x \in V} w_x. \qquad (4)$$

We assume that $W \geq w_{\max}^2$, so that the above defines a probability. It can be verified that $\mathbb{E}[\text{degree}(v)] = w_v$ 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 [7], parametrised by the mean degree d, the maximum degree m, and the exponent $\gamma > 2$ of the weight distribution:

$$w_i = m\Big(1 + \frac{i}{i_0}\Big)^{-\frac{1}{\gamma-1}}, \quad i = 0, 1, \ldots, n-1, \qquad (5)$$

where

$$i_0 = n\Big(\frac{d(\gamma-2)}{m(\gamma-1)}\Big)^{\gamma-1}. \qquad (6)$$

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:

$$d \geq \delta > 0, \ d = o(m), \ m \leq \sqrt{nd}, \ \frac{m}{d} = o\Big(n^{\frac{1}{\gamma-1}}\Big). \qquad (7)$$

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 (4) defines valid probabilities. The last assumption ensures that i_0, defined in (6), tends to infinity.

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 (4) and (5), and satisfying the assumptions in (7). 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, m or d.

3. PROOF OF THEOREM 2

If the graph G is regular, then, by (3), $\pi_v = 1/n$ for all $v \in V$. Hence, the claim of the theorem follows from Theorem 1.

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

$$\sum_{v \in V} \pi_v^2 = \frac{\sum_{v \in V} \text{degree}(v)^2}{\Big(\sum_{v \in V} \text{degree}(v)\Big)^2}. \qquad (8)$$

Define

$$D = \sum_{v \in V} \text{degree}(v) = \sum_{u,v \in V} A_{uv}, \qquad (9)$$

where $A_{uv} = 1((u, v) \in E)$, and

$$\begin{aligned} X_v &= \text{degree}(v)(\text{degree}(v) - 1) \\ &= \sum_{i \neq j} A_{vi} A_{vj} \\ D_2 &= \sum_{v \in V} X_v. \end{aligned} \qquad (10)$$

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 (8) and Theorem 1, and the fact that $\sum_{v \in V} \text{degree}(v)^2 = D_2 + D$ and $D = \sum_{v \in V} \text{degree}(v)$, 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 $\{w_i, i = 0, \ldots, n-1\}$.

For notational convenience, we define

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

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

Next, consider power-law graphs with weight sequence specified by (5) and (6). Since i_0 tends to infinity by assumption, we have for such graphs that

$$\begin{aligned} \overline{w^k} &= \frac{m^k}{n} \sum_{i=0}^{n-1} \Big(1 + \frac{i}{i_0}\Big)^{-\frac{k}{\gamma-1}} \\ &\sim \frac{m^k}{n} \int_0^n \Big(1 + \frac{x}{i_0}\Big)^{-\frac{k}{\gamma-1}} dx \\ &= m^k \frac{i_0}{n} \int_0^{n/i_0} (1 + x)^{-\frac{k}{\gamma-1}} dx. \end{aligned} \qquad (11)$$

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

$$\overline{w^k} \sim \begin{cases} \frac{(\gamma-2)^k}{(\gamma-1)^{k-1}(\gamma-1-k)} d^k, & \text{if } \gamma > k+1, \\[2mm] \frac{(k-1)^k}{k^{k-1}} d^k \log \frac{m}{d}, & \text{if } \gamma = k+1, \\[2mm] \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(k+1-\gamma)} d^{\gamma-1} m^{k+1-\gamma}, & \text{if } 2 < \gamma < k+1. \end{cases} \qquad (12)$$

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

LEMMA 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 (9). Then, we have*

$$\mathbb{E}[D] = n\overline{w},$$

$$\text{Var}(D) = 2\Big(n\overline{w} - \Big(\frac{\overline{w^2}}{\overline{w}}\Big)^2\Big) - \Big(\frac{\overline{w^2}}{\overline{w}} - \frac{\overline{w^4}}{n\overline{w^2}}\Big), \quad (13)$$

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

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

$$\mathbb{E}[D] = n^2 p$$
$$\text{Var}(D) = (2n-1)np(1-p) \sim 2n^2 p(1-p), \quad (14)$$

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

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

.

Proof. It is immediate from (9) 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 (13). Next, rewrite (9) 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

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

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

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

Upon simplifying, this yields the second equality in (13). Now, using the fact that $\overline{w^k} = (np)^k$ for Erdős-Rényi graphs $G(n, p)$, we readily obtain (14).

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 (11) that

$$\overline{w^4} \sim \begin{cases} \frac{(\gamma-2)^4}{(\gamma-1)^3(\gamma-5)}d^4, & \text{if } \gamma > 5, \\[2mm] \frac{81}{64}d^4 \log\frac{m}{d}, & \text{if } \gamma = 5, \quad (15) \\[2mm] \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(5-\gamma)}d^{\gamma-1}m^{5-\gamma}, & \text{if } 2 < \gamma < 5, \end{cases}$$

while

$$\overline{w^2} \sim \begin{cases} \frac{(\gamma-2)^2}{(\gamma-1)(\gamma-3)}d^2, & \text{if } \gamma > 3, \\[2mm] \frac{1}{2}d^2 \log\frac{m}{d}, & \text{if } \gamma = 3, \quad (16) \\[2mm] \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}(3-\gamma)}d^{\gamma-1}m^{3-\gamma}, & \text{if } 2 < \gamma < 3, \end{cases}$$

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

By (13), it suffices to show that

$$\Big(\frac{\overline{w^2}}{\overline{w}}\Big)^2 = o(nd) \text{ and } \frac{\overline{w^4}}{\overline{w^2}} = o(n^2 d) \quad (17)$$

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

Suppose first that $\gamma > 3$. Then, by (16) and the fact that $\overline{w} = nd$,

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

where the last equality follows by (7), and the fac that $d \leq n$.

Now let $\gamma = 3$. Then, by (16) and the fact that $\overline{w} = nd$,

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

where the last equality follows by (7). On the other hand, if $2 < \gamma < 3$, then, by (16),

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

where we have used the inequality $m \leq \sqrt{nd}$ from (7) to obtain the second equality. To obtain the last equality, note that it follows from (7) that $m = o(n)$ and hence that $d = o(n)$ as well. We have thus established the first equality in (17) for all $\gamma > 2$. The proof of the second equality is similar and is omitted. This completes the proof of the lemma. $\square$

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-laws random graphs satisfying the assumptions of Theorem 2, then the sum of node degrees D concentrates at its expected value in the sense that $D \sim \mathbb{E}[D]$ with high probability (whp).*

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

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

We have the following:

LEMMA 2. *Let D_2 be defined as in (10). We then have*

$$\mathbb{E}[D_2] = n\overline{w^2} - \left(\frac{\overline{w^2}}{\overline{w}}\right)^2$$

$$\mathrm{Var}(D_2) \leq 4n\overline{w^3} + 2n\overline{w^2} + 4n\frac{(\overline{w^2})^2}{\overline{w}}. \qquad (18)$$

Proof. 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 (18). Next, for distinct nodes $u, v \in V$, we have

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

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) iuj and kvl 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 $\sum_i \mathbb{E}[A_{iu}] = \mathbb{E}[\text{degree}(u)] = w_u$. Hence, we see from the above that

$$0 \leq \mathrm{Cov}(X_u, X_v) \leq 4\frac{w_u^2 w_v^2}{W}. \qquad (19)$$

Similarly, we obtain

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

Using the fact that distinct edges are independent, we get

$$\mathrm{Var}(X_u) \leq 4w_u^3 + 2w_u^2. \qquad (20)$$

Now, by (10), (19) and (20)

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

Computing the above sums yields the second part of (18). $\square$

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.

LEMMA 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 $D_2 \sim \mathbb{E}[D_2] \sim n^3 p^2$ whp.*

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

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

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

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

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 lemma. $\square$

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

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

Proof. 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, $c_1, c_2, \ldots$ will denote generic positive constants, not necessarily the same from line to line. Recall that $\overline{w} \sim d$.

Suppose first that $\gamma \geq 4$. Then, by (12), $\overline{w^3} = O(d^3 \log \frac{m}{d})$ and $\overline{w^2} \sim c_1 d^2$. Therefore, by Lemma 2,

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

and

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

where the last equality holds because of the assumption in (7) that $d \geq \delta$ for some constant $\delta > 0$. Thus, we see that

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

since $m \leq n$.

Suppose next that $3 \leq \gamma < 4$. Then, by (12), $\overline{w^3} = O(d^{\gamma-1} m^{4-\gamma})$, while $\overline{w^2} \sim c_1 d^2$ if $3 < \gamma < 4$ and $\overline{w^2} \sim c_2 d^2 \log \frac{m}{d}$ if $\gamma = 3$. Therefore, by Lemma 2,

$$
\begin{aligned}
\mathbb{E}[D_2] &\geq c_1 nd^2 - c_2 d^2 \log^2 \frac{m}{d} \\
&\geq c_1 nd^2 - c_2 d^2 \log^2 n = \Omega(nd^2), \quad (21)
\end{aligned}
$$

whereas,

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

We have used the assumption that d is uniformly bounded away from zero to obtain the second inequality above. Since we also assumed in (7) that $d = o(m)$, we have

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

for all $\gamma < 4$. Hence, $\mathrm{Var}(D_2) = O(nd^{\gamma-1} m^{4-\gamma})$. Combining this with (21), we get

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

We have used (7) to obtain the second equality above and the fact that $\gamma \geq 3$ to obtain the last equality. Moreover, $\mathbb{E}[D_2] \sim cnd^2$ for $3 < \gamma < 4$, whereas $\mathbb{E}[D_2] \sim cnd^2 \log(m)$ for $\gamma = 3$.

Finally, suppose that $2 < \gamma < 3$. Then, by (12), $\overline{w^3} = O(d^{\gamma-1} m^{4-\gamma})$ and $\overline{w^2} \sim c_1 d^{\gamma-1} m^{3-\gamma}$, so that, by Lemma 2,

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

Now, by (7), $(m/d)^{3-\gamma} = o(n^{(3-\gamma)/(\gamma-1)}) = o(n)$ since $\gamma > 2$. Consequently,

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

On the other hand,

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

Hence,

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

Now, by (7), 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 $\frac{\mathrm{Var}(D_2)}{\mathbb{E}[D_2]^2} = o(1)$. Note that $\mathbb{E}[D_2] \sim cnd^{\gamma-1} m^{3-\gamma}$, for $2 < \gamma < 3$.

Using Chebyshev's inequality, this establishes the claim of the lemma. $\qquad \square$

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

4. CONCLUSION AND FUTURE WORK

In this work we have presented a simple model for the spread of epidemics where individuals are mobile. In this framework we were interested in the setting where there are two individuals one infected and one healthy both performing random walks on the network. Our preliminary investigation highlights the effect of the topology on the spread of an epidemic, motivated by networking phenomena such as worms and viruses, failures, and dissemination of information. Under this natural model, we provided an explicit relationship between the structure over which the walks are performed and the coincidence time of the two walkers. To this end we analysed both homogeneous (regular, complete and Erdös-Rènyi graphs) and heterogeneous (power-law graphs) networks. We pinpointed the existence of a phase transition for the coincidence time in the case of power-law networks depending on the parameter of the power-law degree distribution. We also derived bounds on the probability of infection.

As a final remark, we propose some several interesting directions to pursue the work presented here. In our present model individuals are supposed to start their walks in stationary regime. This can be relaxed since the networks we study are expanders and thus random walks on such networks have nice mixing properties as illustrated in [17] through the computation of the isoperimetric constant of the underlying graphs. As the probability that more than two particles meet in the same node is small compared to pair meetings we anticipate that similar results can be derived when considering k walkers as long as k is small with respect to n the number of sites in the network.

5. REFERENCES

[1] D. Aldous, "Meeting times for independent Markov chains", *Stochastic Processes and their Applications*, 38, 185-193, 1991.

[2] P. Wang, M.C. González, C.A. Hidalgo and A.-L.Barabási, "Understanding the spreading patterns of mobile phones viruses", *Science* **324**, 1071-1076, 2009.

[3] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science* **286**, 509–512, 1999.

[4] B. Bollobás and O. Riordan, "The diameter of a scale-free random graph", *Combinatorica*, **4**, 5–34, 2004.

[5] M. Bui, T. Bernard, D. Sohier,. and A. Bui, "Random walks in distributed computing: A survey", in *Proc. IICS*, 1-14, 2004.

[6] A. Chaintreau, P. Hui, J. Scott, R. Gass, J. Crowcroft,

and C. Diot, "Impact of human mobility on opportunistic forwarding algorithms", *IEEE Transactions on Mobile Computing*, **6 (6)**, 606-620, 2007.

[7] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees", *Internet Mathematics*, **1**, 91–114, 2003.

[8] C. Cooper, A. Frieze and T. Radzik, "Multiple random walks in random regular graphs", manuscript.

[9] D. Coppersmith, P. Tetali and P. Winkler, "Collisions among random walks on a graph", *SIAM Journal on Discrete Mathematics*, 6(3), 1993.

[10] D.J. Daley and J. Gani , *Epidemic Modelling: an Introduction*, Cambridge University Press, Studies in mathematical biology, 2001.

[11] N. Datta and T. C. Dorlas, "Random walks on a complete graph: a model for infection", *J. Appl. Prob.*, **41**, 1008–1021, 2004.

[12] S. Dolev, E. Schiller, and J. Welch, "Random walk for self-stabilizing group communication in ad hoc networks", *IEEE Trans. Mob. Comput.*, 5(7), 893-905, 2006.

[13] M. Draief and A. Ganesh, "Efficient Routing in Poisson small-world networks", *Journal of Applied Probability*, 43(3), 678-686, 2006.

[14] M. Draief, A. Ganesh and L . Massoulié, "Thresholds for virus spread on networks", *Annals of Applied Probability*, 18(2), 359-378, 2008.

[15] M. Faloutsos, P. Faloutsos and C. Faloutsos, "On power-law relationships of the Internet topology", in *Proceedings ACM Sigcomm*, 1999.

[16] A. J. Ganesh, A.-M. Kermarrec, E. Le Merrer and L. Massoulié, "Peer counting and sampling in overlay networks based on random walks", *Distributed Computing*, 20(4), 267-278, 2007.

[17] A. Ganesh, L. Massoulié and D. Towsley, "The effect of network topology on the spread of epidemics", in *Proc. INFOCOM*, 1455-1466, 2005.

[18] C. Gkantsidis, M. Mihail, and A. Saberi, "Hybrid search schemes for unstructured peer-to-peer networks", in *Proc. INFOCOM*, 1526-1537, 2005.

[19] D. R. Karger and M. Ruhl, "Simple efficient load balancing algorithms for peer-to-peer systems", in *Proc. SPAA*, 36-43, 2004.

[20] D. Kempe, J. Kleinberg and A. Demers, "Spatial gossip and resource location protocols", in *Proc. STOC*, 163-172, 2001.

[21] J. Kephart and S. White, " Directed-graph epidemiological models of computer viruses", in *Proceedings of the IEEE Computer Symposium on Research in Security and Privacy*, 343-359, 1991.

[22] J. Kleinberg, "The Wireless Epidemic", *Nature*, **449**, 287-288, 2007.

[23] J. Kleinberg. *The small-world phenomenon: an algorithm perspective*, in *Proc. STOC*, 163-170, 2000.

[24] N. Leavitt, "Mobile phones: the next frontier for hackers?", *Computer, IEEE Computer Society*, **38 (4)**, 20-23, 2005.

[25] Mobile phone virus Cabir, http://www.dancewithshadows.com/tech/mobile-phone-virus-cabir.asp.

[26] http://www.f-secure.com/v-descs/commwarrior.shtml.

[27] L. Lovász, "Random Walks on Graphs: a survey", in *Combinatorics, Bolyai Society Mathematical Studies*, 2, 1-46, 1993.

[28] M. Nekovee, "Worm epidemics in wireless ad hoc networks", *New Journal of Physics*, **9**, 189, 2007.

[29] C.J. Rhodes, M, Nekovee, "The opportunistic transmission of wireless worms between mobile devices", *arXiv*, arXiv:0802.2685v1, 2008.

[30] J. W. Mickens and B. D. Noble, "Modeling epidemic spreading in mobile environments", in *Proceedings of the 4th ACM workshop on Wireless security*, 77 - 86, 2005.

[31] B. Pittel, "On spreading a rumor", *SIAM Journal on Applied Mathematics*, **47 (1)**, 213 - 223, 1987.

[32] G. Yan, L. Cuellar, S. Eindenbenz, H. Flores, N, Hengartner, V. Vu, "Bluetooth worm propagation: mobility pattern matters!", in *Proceedings of the 2nd ACM symposium on Information, computer and communications security, ASIACCS'07*, 32 - 44, 2007.

[33] J. Su, K. K. W. Chan, A.G. Miklas, K. Po, A. Akhavan, S. Saroiu, E. de Lara, A. Goel, "A preliminary investigation of worm infections in a bluetooth environment," in *Proceedings of the 4th ACM workshop on Recurring malcode, WORM'06*, 9-16, 2006.