The Infection Propagation Time of Graphs

Tassos Dimitriou* Sotiris Nikoletseas† Paul Spirakis†

Abstract

Consider k particles, 1 red and $k-1$ white, chasing each other on the nodes of a graph G. If the red one catches one of the white, it "infects" it with its color. The newly red particles are now available to infect more white ones. When is it the case that all white will become red? It turns out that this simple question is an instance of information propagation between random walks and has important applications to mobile computing where a set of mobile hosts acts as an intermediary for the spread of information.

In this paper we model this problem by k concurrent random walks, one corresponding to the red particle and $k-1$ to the white ones. The *infection time* T_k of infecting *all* the white particles with red color is then a random variable that depends on k, the initial position of the particles, the number of nodes and edges of the graph, as well as on the structure of the graph.

In this work we develop a set of probabilistic tools that we use to obtain upper bounds on the expected value of T_k for general graphs and important special cases. We easily get that an upper bound on the expected value of T_k, ET_k, is the worst case (over all initial positions) *expected meeting time* m^* of two random walks multiplied by $\Theta(\log k)$. We demonstrate that this is, indeed, a tight bound; i.e. there is a graph G (a special case of the "lollipop" graph), a range of values $k < n$ (such that $\sqrt{n} - k = \Theta(\sqrt{n})$) and an initial position of particles achieving this bound.

When G is a clique or has nice expansion properties, we prove *much smaller bounds* for ET_k. We have evaluated and validated all our results by large scale experiments which we also present and discuss here. In particular, the experiments demonstrate that our analytical results for these expander graphs are tight.

Note: Due to lack of space, we have included a clearly marked Appendix.

*Athens Institute of Technology, Greece. E-mail: `tdimitr@math.uoa.gr`

†Computer Technology Institute and Department of Computer Engineering & Informatics, University of Patras, Greece. E-mails: `nikole@cti.gr, spirakis@cti.gr`.

1 Introduction, Problem Definition and Motivation

Properties of *interacting particles* (moving on a finite graph) are of interest within a number of areas of Science, primarily Physics in the early days, but increasingly Biology, the Social Sciences and Computer Science today, since they model interesting phenomena like magnetism, spatial competition, tumor growth, spread of infection, economic systems and mobile communication. The book by Liggett ([12]) and Chapter 14 of the Aldous and Fill book ([1]) are good references.

We consider here a special case of interacting particles: The motion of each of the particles is a random walk on the graph. The walks are *concurrent*; for simplicity, the initial positions of the particles are those of the steady-state probability distribution of the walks.

Let k the number of particles. Let $G(V, E)$ the graph of the walks. Let $|V| = n \geq k, |E| = m$. Initially, one particle is *red* and all others are *white*. Let $\pi()$ the steady-state distribution of each of the walks. The particles interact according to the following *Infection Rule:* When a red particle meets (at a graph node) one (or more) white particle(s), the white particle(s) turn red.

Initially, each vertex of G can have at most one particle (thus $k \leq n$).

Definition 1 Let T_k be the least time instant at which all particles are red.

Note that T_k is a random variable. We are interested here in the expected value of T_k, $E_\phi T_k$, where ϕ is the initial distribution of particles.

Definition 2 We call *Infection Time*, w.r.t. an initial probability distribution ϕ, the value of $E_\phi T_k$.

Note that $\pi()$ always exists when the random walk is of *continuous time* with *transition rates* $q_{vx} = \frac{1}{d_v}$ if $\{v, x\}$ is an edge (d_v is the degree of vertex v), and $q_{vx} = 0$ if not, because, then, the random walk is aperiodic and ergodic.

In fact, we start by considering continuous time walks. Note that the discrete-time walks defined by the transition probability $p_{uv} = \frac{1}{d_u}$ if $\{u, v\} \in E$, and $p_{uv} = 0$ else, have the same stationary distribution mean hitting times as in the continuous case ([1]). The main goal of this work is to study the Infection Times of various (unweighted, undirected, finite) graphs.

Motivation. The problem studied here is an example of (i) the spread of a virus (i.e. the red colour) in computer networks (ii) information spreading (such as rumor spreading, gossiping) (iii) a considerable fraction of the communication time of stations in ad-hoc mobile networks, under the existence of a (mobile) infrastructure as a virtual intermediary pool for messages.

For (i) above, we note that virus attacks in computer networks pose several key problems regarding intrusion propagation. Various models have been proposed for studying the effective detection and defeat of attacks, see e.g. [15]. Intrusion Propagation (the process of spread of such attacks) has mostly been investigated under gossip or epidemiological models (see [10]). Also, [14] investigates the analytic and experimental behaviour of several protocols for the attack propagation problem in networks under a new model of intrusion propagation introduced there.

For (ii), [8] studies the gossip problem, where n persons wish to distribute n rumors among themselves (where however which persons communicate at each time is fixed in advance by the designer of the algorithm). See also [3], that considers rumor spreading where, in contrast to [14], the choice of who receives information at each time is under the control of an adversary.

We elaborate in more detail on (iii). Recall that an ad-hoc mobile network is a collection of mobile hosts, with wireless communication capabilities, forming a temporary network without the aid of any established fixed infrastructure. In [4, 5, 6] we have studied the problem of basic communication, i.e. to send information from a sender mobile user, MH_S, to a receiver mobile user, MH_R. For such dynamically changing networks of high mobility we have proposed protocols for basic communication which exploit *the accidental meetings* of the mobile hosts and the co-ordinated (by the protocol) motion of a *small part* of the network.

We abstract the 3D network area by a *motion-graph*, whose vertices model cubes of volume close to that of the transmission sphere of hosts, and whose edges connect adjacent cubes.

The protocol works as follows: The nodes of the support move fast enough to visit (in sufficiently short time) the entire motion graph. When some support node is within communication range of a sender, it notifies the sender that it may send its message(s). The messages are then stored "somewhere within the support structure". When a receiver comes within range of a support node, the messages are then forwarded to the receiver.

Clearly the size, k, and the shape of the support affects performance. In [4, 5, 6] we study two basic alternatives for the support structure: a) the "snake" support, where the nodes of the support move always remaining pair-wise adjacent (i.e., forming a chain of k nodes). Essentially the support moves as a "snake", with the head doing a random walk on the motion graph and each other node executing the simple protocol "move where the node preceding me was before". b) a different approach is to allow each member of Σ to perform an *independent* random walk on the motion graph, i.e., the members of Σ can be viewed as *"runners"* running on G. *When two runners meet, they exchange any information* given to them by senders encountered.

In [4, 5, 6] we perform, using Markov Chain techniques, a rigorous average case analysis for the expected communication time of the snake protocol. The infection propagation time of graphs in this paper can be used to estimate the expected communication time in the runners case. We expect that, because of the "parallelism" of runners, *the runners protocol significantly outperforms the snake protocol* (a fact experimentally validated in [4, 5, 6]).

2 New Results and Related Work

We first show an *upper bound* on the expected infection propagation time ET_k for any $k > 2$ and for any undirected graph G, where we consider continuous time walks.

We demonstrate that this bound is *tight* on a Lollipop graph for certain values of k and an initial position of particles.

Then, we turn into discrete time walks and derive *much smaller* bounds on ET_k for *the clique and for expander graphs.* Note that our model of continuous time walks is just the "continuization" of the corresponding discrete time chains. Thus the results hold in both cases.

Previous work and Comparison. Coppersmith et al ([7]) evaluated the expected *meeting time* of *two* random walks (i.e. ET_2). They showed that in the worst case (and assuming discrete time), $ET_2 = (4/27 + o(1))n^3$. Tetali and Winkler ([18]) gave some earlier bounds on ET_2. Sunderam and Winkler ([17]) examined a related but different problem: each node of a clique is a processor having some information piece. They estimate techniques to minimize the time when all processors know all pieces of the information. There, processors do not move and messages are exchanged in discrete, synchronized rounds.

Works on other models of interacting particles (e.g. the anti-voter and voter models), can be found in Ch. 14 of [1] and in ([12]). Infection models were studied well in the past, under the direction of population biology. Their targets are similar but the main difference is the missing of the *graph* as space of motions. See [16] and references there.

So, our work *extends* the results of [7, 18] to $k > 2$ concurrent random walks on a finite undirected graph.

Roadmap. Our first results on sections 3, 4 and 5 refer to *continuous time* concurrent random walks. The results on special graphs (sections 6, 7) refer to discrete time walks. In section 8, we present and discuss our experimental findings. Finally, in the Appendix we present the Figures with the experimental findings and also a detailed proof of Lemma 4.

3 A Tool: Exponential tails of hitting times and meeting times

Definition 3 Let $A \subseteq V$ and $i \in V$. Define $E_i T_A$ to be the expected value of the first hitting time to A (i.e. the first time to arrive at a node of A) of a random walk on G, starting from vertex i. Let T_A be the corresponding random variable. Let $t_A^* = \max_{i \in V} E_i T_A$.

Definition 4 Let E an event about a random walk on G. We denote by $\Pr_\phi(E)$ the probability of E when the starting position of the walk is according to the distribution $\phi()$.

The first result that follows can be found in [1]. We explain it here by elementary means (to also enhance readability of the results that follow): For any initial distribution $\phi()$, any time $s > 0$ and integer $m \geq 1$ it is

$$\begin{aligned} \Pr_\phi\{T_A > ms \text{ given that } T_A > (m-1)s\} &= \Pr_\theta\{T_A > s\} \text{ for some distribution } \theta \\ &\leq \max_{i \in V} \Pr_i\{T_A > s\} \leq \frac{t_A^*}{s} \end{aligned}$$

by the Markov inequality. So, by induction on m, we get:

$$\Pr_\phi\{T_A > js\} \leq \left(\frac{t_A^*}{s}\right)^j \qquad \text{i.e.} \qquad \Pr_\phi\{T_A > t\} \leq \left(\frac{t_A^*}{s}\right)^{\lfloor t/s \rfloor}$$

where $t > 0$. In continuous time, if we choose $s = et_A^*$ we get

Lemma 1 (SubExponentiality Lemma 1a) *([1])* For any $t : 0 < t < \infty$ it is

$$\sup_\phi \Pr_\phi\{T_A > t\} \leq exp(-\lfloor t/et_A^* \rfloor)$$

The machinery stated above can be used also for the meeting times:

Definition 5 Let $M_{i,j}$ be the first time that two independent copies of a random walk on G meet given that they start from i, j. Let $m^* = \max_{i,j} M_{i,j}$.

Working exactly as above, we then get:

Lemma 2 (SubExponentiality Lemma 1b) *([1])*

$$\Pr\{M_{i,j} > t\} \leq exp(-\lfloor t/et_m^* \rfloor)$$

Remark 1: Note that the above Lemmata are useful only when t is large, i.e. above et_m^* or et_A^*. When t is small then the Lemmata hold trivially because they just say that a tail probability is less than or equal to 1.

Remark 2: In fact, because our space is *finite*, many quantities which converge to 0 as $t \to \infty$, must converge exponentially fast, by iterating over worst-case initial states.

This general property is an implication of the *submultiplicativity property* of our Markov Chains (Random Walks), which is as follows: If X_t is the vertex of the walk at time t and we define a deviation from stationarity measure $d(t)$ as

$$d(t) = \max_{i,j} \left| \Pr_i(X_t) - \Pr_j(X_t) \right|$$

then one can show that

$$d(s+t) \leq d(s)d(t) \quad \forall s, t \geq 0$$

Definition 6 For a continuous time walk on G let T_j be the *first hitting time* vertex j, i.e. the first time for the walk to be at vertex j. Let $M_{i,j}$ be the meeting time of two independent copies of the walk, started at vertices i and j, and $EM_{i,j}$ its expected value.

Another interesting fact about the relation of hitting times and meeting times, for continuous time walks, is the following:

Lemma 3 (Aldous and Fill [1], Chapter 14, Proposition 5)

$$\max_{i,j} EM_{i,j} \leq \max_{i,j} E_i T_j$$

This can be easily seen when i, j are vertices of degree 1 in the $n+1$-vertex star graph.

4 A tight upper bound on infection time for general graphs

Consider, instead of our process (call it P_1) of infection, another process of the same walks governed by the rule:

"When particles meet, they coalesce into clusters and the cluster thereafter sticks together and moves as a single random walk. If the cluster contains red particles then it is colored red".

Definition 7 Let $C_{k,n}$ be a random variable, which is the time at which all particles coalesce into one single cluster.

Clearly, $\forall t > 0$, it is $\Pr\{C_{k,n} > t\} \geq \Pr\{T_k > t\}$, and thus

$$E_\phi C_{k,n} \geq E_\phi T_k \tag{1}$$

for all initial position distributions ϕ. But $EC_{k,n}$ can be bounded as follows:

Lemma 4 $EC_{k,n} \leq em^*(2 + \log k)$,
where e is the basis of the natural logarithms.

Proof: (sketch). We proceed exactly as in Aldous and Fill, Chapter 14, Proposition 11 ([1]) with the only difference being that initially we have k particles instead of n. The proof uses the second of the subexponentiality Lemmas. (see Appendix) $\diamond$

Hence, by (1) we get

Theorem 1 $ET_k \leq em^*(2 + \log k)$

Although this bound is crude for a lot of graphs, there are graphs and positions of particles achieving it. Consider a special case of the Lollipop graph, with a clique of size $n - \sqrt{n}$ and a path of length $l = \sqrt{n}$ extending out of it. Initially, there is one red particle in the clique. There are k white particles, all being at the furthest k vertices of the path. Here $k << \sqrt{n}$. In the sequel, take k so that $\sqrt{n} - k = \Theta(\sqrt{n})$. For a graphical representation of an instance of this graph where $n = 16, k = 3$, see Figure 1 in the next page.

Clearly the worst case expected meeting time is $m^* = \Theta(l^2) = \Theta(n)$, i.e. the average time for the random walk of a white particle to enter the clique, by elementary random walk properties (see Feller, [9]).

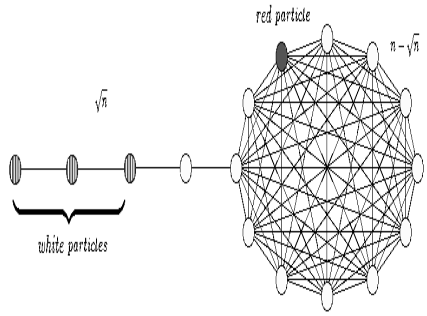

Figure 1: The Lollipop graph we consider here

Let the white particles be labelled (from the furthest to the clique inwards) as $i_1, \ldots, i_k$. Let their times to enter the clique be the random variables $S_1, \ldots, S_k$, respectively. Note that it is very hard for the red particle to exit the clique. This can happen only when it gets to the vertex where the path and the clique meet (i.e. once every $\Theta(n)$ time), but even then the probability of exiting the clique is $\frac{1}{n-\sqrt{n}+1}$. On the average, thus, the red particle stays inside the clique for $\Theta(n^2)$ expected time. Let $S = \max\{S_1, \ldots, S_k\}$. Clearly then, the infection time is $T_{k+1} = S + \Theta(1)$ w.h.p., where $\Theta(1)$ accounts for the last white particle meeting another particle (which will be red by then). Thus,

$$E(T_{k+1}) = E(S) + \Theta(1) \tag{2}$$

From the subexponentiality property of individual hitting times (Lemma 1) we have that $\forall i = 1, \ldots, k$ and for

$$s^* = \max_{i=1,\ldots,k} E(S_i) = \Theta(n)$$

we get

$$\Pr\{T_i > \lambda e s^*\} \leq e^{-\lambda}, \quad \lambda = 1, 2, 3, \ldots \tag{3}$$

So,

$$\Pr\{S > \lambda e s^*\} \leq k e^{-\lambda}, \quad \lambda = 1, 2, 3, \ldots \tag{4}$$

i.e

$$\Pr\left\{\frac{S}{es^*} > x\right\} \leq kee^{-x}, \quad x \geq 0 \tag{5}$$

i.e.

$$\Pr\left\{\frac{S}{es^*} - \log(ek) > x\right\} \leq e^{-x}, \quad x \geq 0 \tag{6}$$

This says that the distribution of $\frac{S}{es^*} - \log(ek)$ is "stochastically smaller" than the exponential distribution, implying

$$E\left(\frac{S}{es^*} - \log(ek)\right) \leq 1$$

and hence

$$E(S) \leq (2 + \log k) e s^* \tag{7}$$

In fact, for the one-dimensional random walks this bound is tight! In the continuous time, the number of +1 and -1 steps of a single walk in the line are independent Poisson(t) variables (of mean 1). Then a bound as in (3) holds with equality, and from

$$\Pr\{S < t\} = \Pr\{T_1 < t\} \cdot \Pr\{T_2 < t\} \cdots \Pr\{T_k < t\}$$

(because of the independence of the walks), also relations similar to (4), (5) hold with equality. Thus,

Lemma 5 $E(S) = \Theta(s^* \log k)$

But $s^* = \Theta(m^*)$ for the $\sqrt{n}$ line as we said. So,

Lemma 6 $E(S) = \Theta(m^* \log k)$

Thus, by (2) and (7), we get:

Theorem 2 In the Lollipop graph and for k such that $\sqrt{n} - k = \Theta(\sqrt{n})$, there is a set of initial positions of particles so that $E(T_{k+1}) = \Theta(m^* \log k)$. Thus, the bound of Theorem 1 is tight.

Note however than even in the lollipop graph, for large k, the infection time is much smaller than this upper bound. This is demonstrated in the experiments section and we believe that this can also be analytically shown.

5 Infection times on the Clique

As our experiments also indicate (see Section 8), the bound stated in Theorems 1 and 2 is not tight for "close to regular graphs". In the next sections we derive *much better* bounds for such graphs. The experiments we have conducted and which we discuss in Section 8 and in the Appendix, demonstrate that our analytical results are tight. The case of the clique K_n is instructive, as the following theorem demonstrates:

Theorem 3 *Consider $k \le n$ particles, one red, $k-1$ white, sitting on the nodes of a K_n. Then the expected time when all particles become red after randomly walking on the K_n is bounded by*

$$\frac{4n}{k-1}(\ln k + O(1)).$$

Denote by R_t and W_t the number of red and white particles at time t, where $R_t + W_t = k$ and $R_0 = 1$. Our goal is to show that for $t^* = O((n/k)\ln k)$ the expected number of red particles is k. It will be helpful in the analysis to consider each round of particle's movement as a two stage process. In the first stage of round t we move all the red particles to a random neighbor of their current location. Then we move the white particles. If a white particle falls into the position of a red one, it becomes infected and changes its color. At the end of the second stage we count all the red particles and we obtain R_{t+1}. Observe that this model is exactly the same as the original one but is simpler to analyze and give us the following intermediate result.

Lemma 7 *Let R_t the number of red particles at the end of round t. Then at the end of round $t+1$ the expected difference $\Delta_t = R_{t+1} - R_t$ satisfies*

$$\Delta_t \ge \frac{R_t(k-R_t)}{n}\left(1 - \frac{R_t}{2n}\right)$$

Proof: Let $r = R_t$ the number of red particles. To simplify the analysis we allow each particle to stay in the current node with probability $1/n$. Thus the next node can be any of the nodes $1, \ldots, n$ with equal probability $1/n$. Now consider what happens when the red particles move.

If r is small then all these particles will land almost surely to different nodes. But when r is large the red particles may form clusters which in effect decrease the population of red particles that may infect the white ones. Thus we need to compute the number r' of different clusters of red nodes and treat each cluster as one red "super-particle" that can propagate its color.

So what is r'? We can compute r' by treating the r particles as balls which have to be thrown in n bins, where each bin corresponds to one of the possible n neighbors. The number of different clusters will simply be the expected number of occupied bins in the balls experiment, which is by definition n minus the number of empty bins. Now the expected number of empty bins is easily found to be $n(1-\frac{1}{n})^r$, which is bounded from above by $n(1-\frac{r}{n}+\frac{r^2}{2n^2})$. Thus we get that the number of different super-particles r' is

$$r' \geq nn\left(1-\frac{r}{n}+\frac{r^2}{2n^2}\right) = r - \frac{r^2}{2n} = r\left(1-\frac{r}{2n}\right).$$

All these super-particles are now available in stage two to infect a white particle if it happens to land in the same node occupied by the cluster. The probability that a white particle changes its color to red is at least r'/n, thus on the average we expect $W_t(r'/n) = (k-r)(r'/n)$ of them to become infected. We conclude that at the end of round $t+1$, the expected number of red particles will be

$$R_{t+1} \geq R_t + \frac{(k-R_t)R_t}{n}\left(1-\frac{R_t}{2n}\right)$$

or equivalently since $R_t \leq k \leq n$

$$\Delta_t \geq \frac{R_t(k-R_t)}{2n}$$

and the proof of the lemma is complete. $\diamond$

This lemma will help us estimate the number of rounds before all particles become red. An easy but not so accurate bound follows by observing that Δ_t is always at least $(k-1)/2n$. Thus during successive rounds we make at least $(k-1)/2n$ progress towards infecting all particles, therefore the total number of rounds required is at most $2n$. A more accurate counting however gives the desired result.

Define T_i to be the time when $R_{T_i} = i$, that is the moment when the number of red particles is exactly i. Thus in the time interval $[T_i, T_{i+1})$, the expected number of red particles is at least i and smaller than $i+1$. In this time interval the expected progress Δ_i in *all* rounds between T_i and T_{i+1} is therefore at least

$$\Delta_i \geq \frac{i(k-i-1)}{2n}.$$

But then the expected number of rounds between T_i and T_{i+1} is at most $1/\Delta_i$ since the number of red particles increases by one. Thus overall, the number of required rounds T_k is

$$\begin{aligned} T_k &\leq \sum_{i=1}^{k-2}\frac{2n}{i(k-i-1)} + \frac{2n}{k-1} = \frac{2n}{k-1}\sum_{i=1}^{k-2}\left(\frac{1}{i}+\frac{1}{k-i-1}\right) + \frac{2n}{k-1} \\ &< \frac{4n}{k-1}\left(\sum_{i=1}^{k-2}\frac{1}{i}+\frac{1}{2}\right) < \frac{4n}{k-1}(\ln k + O(1)) \end{aligned}$$

thus completing the proof of the theorem.

One immediate generalization of the random walk on a clique is when we let p denote the probability of visiting a random neighbor from the current node. Clearly, Theorem 3 corresponds to the case where $p = 1/n$. Working in an entirely analogous way we obtain

Theorem 4 *Consider $k \leq n$ particles, one red, $k-1$ white, sitting on the nodes of a K_n, where the probability of visiting any node is p. Then the expected time when all particles become red is*

$$O\left(\frac{(1/p)^2}{nk}\ln k\right).$$

6 The case of Graphs with good Expansion

A graph $G = (V, E)$ informally is an expander if the number of neighbors of any given set of nodes S is not "too small". More formally

Definition 8 *The expansion v of a graph $G = (V, E)$ is*

$$v = \min_{|S| \leq |V|/2} \left\{\frac{N(S)}{S}\right\}$$

where $N(S)$ is the set of vertices in $V - S$ which are adjacent to some vertex in S.

Expander graphs have many useful properties but perhaps the most important one is that a random walk on the nodes of the graph is unlikely to remain trapped for a long time in a small portion of the state space. So if $P = [p_{u,v}]_{n \times n}$ denotes the one-step transition matrix of the walk and π_0 the initial distribution, then the distribution at time t, π_t, quickly converges to a unique stationary distribution π. The speed of convergence is given by the following lemma:

Lemma 8 (Expander mixing lemma, [2]) *Denote by $\Delta(t)$ the relative pointwise distance between π_t and π at time t. If G has expansion v then*

$$\Delta(t) < \frac{(1 - \frac{v^2}{d})^t}{\min \pi_u}$$

where d is the maximum degree of the graph.

An immediate corollary of the mixing lemma is that the time t required for $\Delta(t)$ to drop below some amount ϵ is at most $\frac{d}{v^2}(\ln\frac{1}{\epsilon} + \ln\frac{1}{\min \pi_i})$. Thus if the expansion is not too small, convergence to the stationary distribution is guaranteed to be fast.

An instructive case is when the graph is d-regular. For random d-regular graphs or for explicit constructions of families of d-regular graphs it is known that the expansion is a small constant. Thus the time to approximate the stationary distribution, which in this case is $1/n$, by a small amount $\epsilon = n^{-2}$ is at most $O(d \log n)$.

This observation will help us get a good upper bound for the infection time of k particles on a d-regular graph G with good expansion properties. In particular we will reduce this case to the clique case presented in Section 5 by "simulating" each step of the walk on a clique with $O(d \log n)$ steps of a walk on the d-regular graph.

More precisely we move the k particles on G by breaking their walks into epochs of duration $O(d \log n)$. We let each particle perform a $O(d \log n)$ steps walk inside each epoch and we consider what happens at the end of this walk. In particular, we allow white particles to be infected only if they meet a red one at the end of the epoch. This gives an upper bound on the infection time because white particles may truly be infected in some intermediate time.

Each epoch has the effect of simulating one step at the clique, since by the corollary to expander mixing lemma, the probability of visiting any particular neighbor is equal (by an n^{-2} amount) to the stationary distribution $1/n$.

Applying Theorem 4, we conclude that the expected number of epochs required to infect all the red particles is at most $O((n/k)\log k)$. Combined with the $O(d \log n)$ length of each epoch we find that the infection time of a d-regular graph with good expansion properties is at most

$$O\left(\frac{nd}{k}\log k \log n\right) = O\left(\frac{m}{k}\log k \log n\right)$$

where m is the number of edges of the graph.

Observe that the same bound (to within a polylogaritmic factor) applies for more general expander graphs. For such graphs the expansion v is bounded below by $1/\log^{O(1)} n$, where n is the number of nodes of the graph, thus each epoch never requires more than $O(d \log^{O(1)} n)$ steps to reach the stationary distribution.

7 Future Work

The results of the previous section suggest that graphs with good expansion are likely to have small infection times. Does this mean that a graph with bad expansion is likely to have large infection times? Intuition suggests a yes answer since in this case the walks do not diffuse rapidly and particles are likely to remain trapped for longer times into small regions of the space.

A notion that captures nicely the probability of "getting stuck" to any subset S of the state space is the *conductance* of the graph which is essentially a weighted analog of expansion.

The conductance (see [11]) induced by a random walk on a graph $G = (V, E)$ is

$$\Phi = \min_S \Phi_S,$$

where for any subset S of V such that $\sum_{u \in S} \pi_u \leq 1/2$

$$\Phi_S = \sum_{u \in S, v \in \bar{S}} \pi_u P_{u,v} \;\; / \;\; \sum_{u \in S} \pi_u$$

The quantity Φ_S may be interpreted as the probability, in the stationary distribution, of making a transition from S to $\bar{S}$ conditional on being on S. Thus the minimum of such Φ_S is the conductance of the Markov chain which essentially identifies the places of the state space where one is likely to get trapped.

Consider now such a bad set S and place the initial red particle there. If the conductance Φ is small (and S was chosen to have small conductance) this set forms a constriction in the state space which makes it hard to pass from S to $\bar{S}$ and reach the white particles. Intuitively, we expect the existence of such bad sets to lead to big delays in crossing the cut from S to $\bar{S}$. In particular, the expected time to cross the cut will be proportional to $1/\Phi$.

We are currently working on such a characterization of graphs with good infection times in terms of the expansion of the random graph. It would be nice to see a lemma relating the two quantities thus providing an alternative, useful characterization of graphs with good expansion.

8 Experiments

We implemented the experiments in C++ using classes and data structures of the LEDA Platform ([13]). We investigated four classes of graphs: lollipops, cliques, 4-regular graphs and $G_{n,p}$

graphs. For each graph type we conducted experiments for various graph sizes (n, the number of vertices) and numbers of particles (k). For each instance we repeated the experiment enough times to get good average results and smooth time curves. For the expander graphs we tested $n \in \{100, 200, 400, 1600, 3200\}$, $k \in \{2, 10, 20, 25, 30, 40, 50, 70, 100\}$, carrying out 200 repetitions in each case, while for the lollipop we tested $n \in \{100, 200\}$, $k \in \{2, 10, 20, 25, 30, 40, 50, 70, 100\}$, conducting 100 repetitions. In the figures (see Appendix) we display the findings for the largest graph size (n) in each case.

The infection stochastic process has been implemented evolving in discrete time rounds. Initially, each particle is placed on a vertex of the graph studied. In the beginning of each round every particle moves randomly and uniformly to one of the neighbouring vertices of the vertex it currently resides. If during a round a red particle resides on the same node with some white particles, then the white particles become red. The process ends when all particles become red. We measure in each case the expected number of rounds needed for all particles to become red.

In the "lollipop" graph case, we had to experimentally evaluate m^*, the maximum of the expected meeting times of two particles, over all possible pairs of starting vertices of the particles. To obtain a realistic value of m^*, we took the average of the meeting times of the two particles in 80 repetitions, for every vertex pair, and then the maximum of all these mean values.

The main experimental findings are the following: (i) As shown in Fig. 2, in the lollipop the crude upper bound is indeed tight for small k, while it is trivial for larger values of k. (ii) The bound we prove for the clique is very tight (see Fig. 3 where the analytical and the experimental curves almost coincide). The same is happening in the regular graphs (the distance between the two curves in Fig. 4 is only due to the fact that the degree (and thus the expansion) is very small. (iii) A similar behaviour is exhibited in the $G_{n,p}$ case, as shown in Fig. 5, 6 and 7. This similarity in cliques, regular graphs and $G_{n,p}$ graphs is due to their common expansion properties. iv) As shown in Fig. 6, the infection time in lollipops is much bigger compared to that of the expander graphs (note that in this Figure we display the lollipop infection time divided by 100, for its curve to fit in the Figure).

Acknowledgements: We thank the anonymous referee who, in an earlier version, pointed out an erroneous use of Lemma 2 and mentioned the Lollipop as a worst case graph. We warmly thank H. Efthimiou, whose contribution to the experimental part has been valuable.

References

[1] D. Aldous and J. Fill: *Reversible Markov Chains and Random Walks on Graphs.* Unpublished manuscript. http://stat- www.berkeley.edu/users/aldous/book.html, 1999.

[2] N. Alon. Eigenvalues and Expanders. *Combinatorica*, 6:83–96, 1986.

[3] J. Aspnes and W. Hurwood: Spreading rumors rapidly despite an adversary. In *Proc. 15th ACM Symposium on Principles of Distributed Computing* PODC96, (1996).

[4] I. Chatzigiannakis, S. Nikoletseas and P. Spirakis: Distributed Communication and Control Algorithms for Ad- hoc Mobile Networks. In the Journal of Parallel and Distributed Computing (JPDC), Special Issue on Mobile Ad-hoc Networking and Computing, to appear in 2002.

[5] I. Chatzigiannakis and S. Nikoletseas: Design and Analysis of an Efficient Communication Strategy for Hierarchical and Highly Changing Ad-hoc Mobile Networks. In ACM/Baltzer Mobile Networks Journal (MONET), Special Issue on Parallel Processing Issues in Mobile Computing, Guest Editors: A. Zomaya and M. Kumar, to appear in 2002.

[6] I. Chatzigiannakis, S. Nikoletseas, and P. Spirakis: An Efficient Communication Strategy for Ad-hoc Mobile Networks In *Proc. 15th International Symposium on Distributed Computing* – DISC'01. Lecture Notes in Computer Science, Volume 2180 (Springer-Verlag, 2001), pp. 285–299. Also Brief Announcement in *Proc. 20th Annual Symposium on Principles of Distributed Computing* – PODC'01, pp. 320–322, ACM Press 2001.

[7] D. Coppersmith, P. Tetali and P. Winkler: Collisions among random walks on a graph. In SIAM J. Disc. Math. 6:3 (1993), pp. 363-374.

[8] S. Even and B. Monien: On the number of rounds needed to disseminate information. In *Proc. 1st ACM Symposium on Parallel Algorithms and Architectures* SPAA89, (1989).

[9] W. Feller. An Introduction to Probability Theory and its Applications. *John Wiley*, New York, 1957.

[10] J. Kephard and S. White: Directed-Graph Epidemiological Models of Computer Viruses In *Proc. IEEE Symposium on Security and Privacy*, 1991. Also IBM Technical Report.

[11] M. R. Jerrum and A. Sinclair: Conductance and the rapid mixing property of Markov chains: The approximation of the permanent resolved. In *Proc. 20th ACM Symposium on Theory of Computing (STOC)*, pages 235–244, 1988.

[12] T. Liggett: Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer Verlag 1999.

[13] K. Mehlhorn and S. Naher: LEDA: A Platform for Combinatorial and Geometric Computing. Cambridge University Press. (1999).

[14] S. Nikoletseas, G. Prasinos, P. Spirakis and C. Zaroliagis, "Attack Propagation in Networks", in the Theory of Computing Systems (TOCS) Journal, Special Issue on the Thirteenth (13th) Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA 2001), to appear.

[15] R. Ostrofsky and M. Yung: How to Withstand Mobile Virus Attacks In *Proc. 10th ACM Symposium on Principles of Distributed Computing* PODC91, (1991), pp. 5159.

[16] J. R. Norris: Markov Chains. Cambridge University Press, 1997.

[17] V.S. Sunderam and P. Winkler, In Disc. Appl. Math. 42 (1993), pp. 75-86.

[18] P. Tetali and P. Winkler: Simultaneous reversible Markov chains. In Combinatorics, Paul Erds is Eighty Vol. 1, D. Miklos, V.T. Sos and T. Szonyi (editors), Janos Bolyai Mathematics Institute, Budapest, 1993, pp. 433-452.

Appendix

Prof of Lemma 4: Order the initial positions of the particles arbitrarily as $i_1, \ldots, i_k$. First, let the k particles perform independent random walks forever, with the particles starting at positions i, j first meeting at time $M_{i,j}$ (say). Then, when two particles meet, let them cluster and follow the future path of the lowest-labelled particle. Similarly, when two clusters meet, let them cluster and follow the future path of the lowest-labelled particle in the combined cluster. Then it follows that

$$C_{k,n} \leq \max_j M_{i_1,j}$$

Let $m^* = \max_{i,j} EM_{i,j}$. By the second subexponentiality lemma then

$$\Pr\{M_{i,j} > t\} \leq \exp\left(-\left\lfloor \frac{t}{em^*} \right\rfloor\right) \tag{8}$$

So, by (8), we get

$$EC_{k,n} = \int_0^\infty \Pr\{C_{k,n} > t\} dt \leq \int_0^\infty \min\left\{1, \sum_{j \in \{i_2,\ldots,i_k\}} \Pr\{M_{i,j} > t\}\right\} dt$$

and thus

$$EC_{k,n} \leq \int_0^\infty \min\left\{1, k\exp\left(-\frac{t}{em^*}\right)\right\} dt = em^*(2 + \log k)$$

$\diamond$

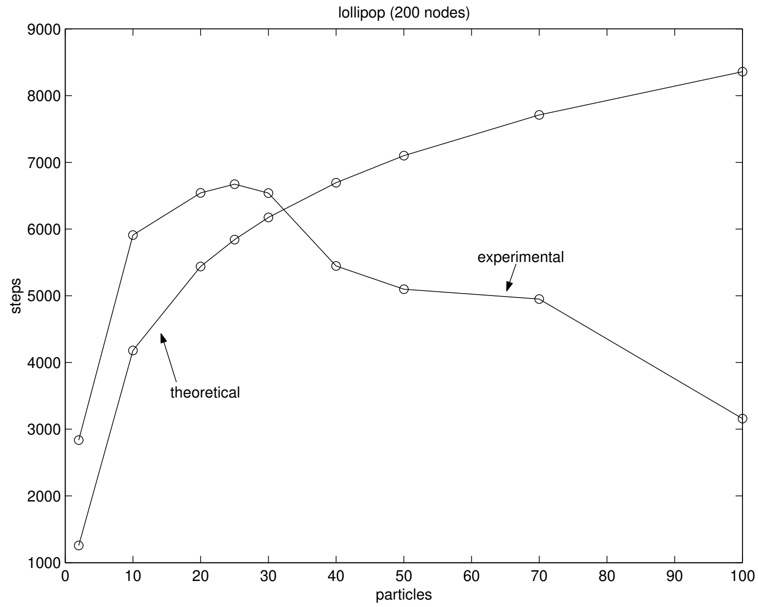

Figure 2: The lollipop case

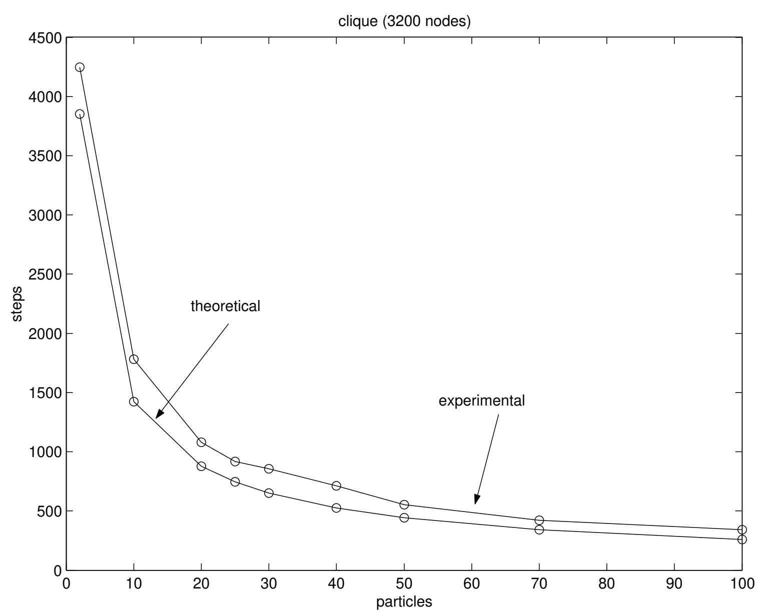

Figure 3: The clique case

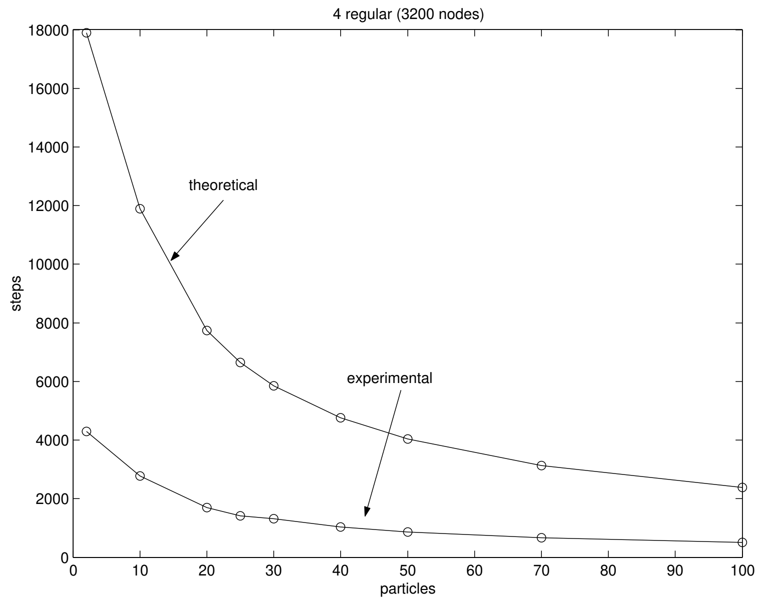

Figure 4: The regular graphs case

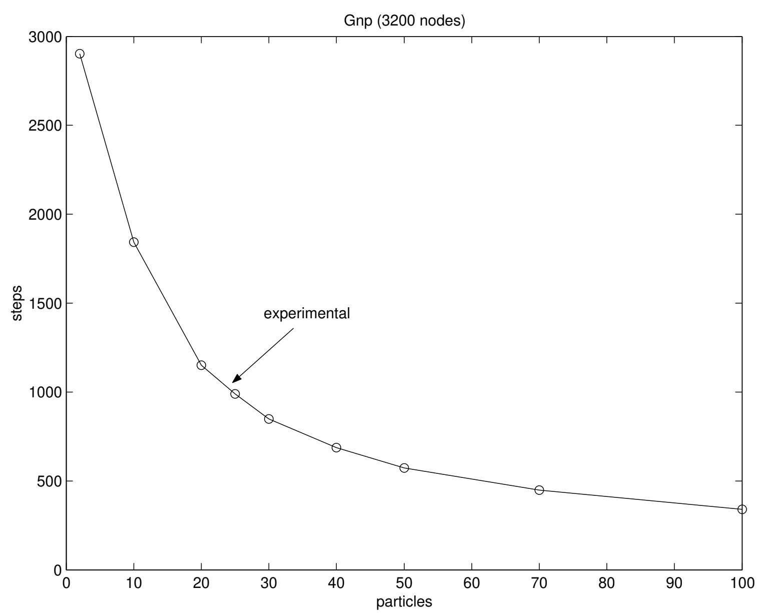

Figure 5: The $G_{n,p}$ graphs case

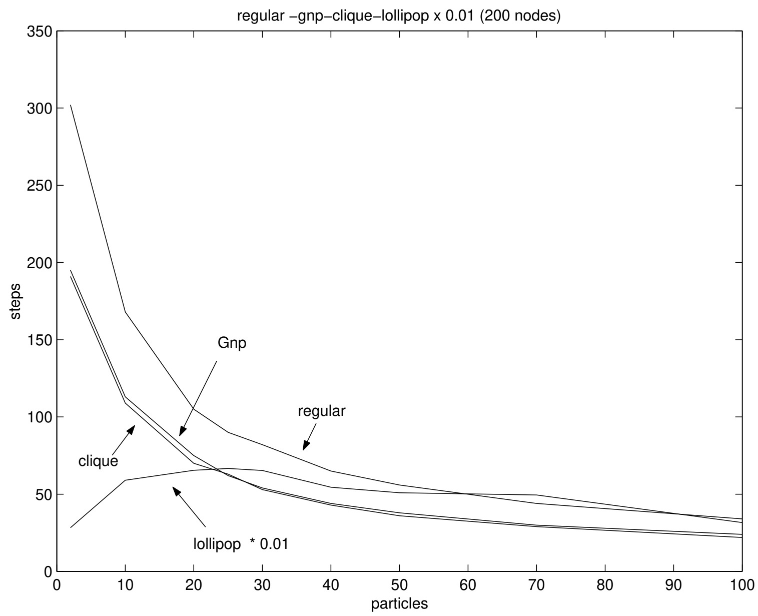

Figure 6: Comparison

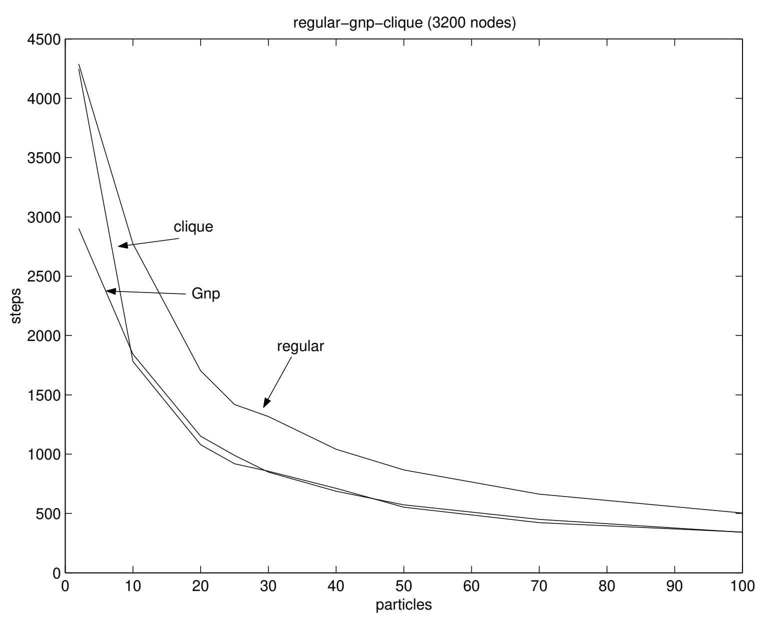

Figure 7: Expander graphs comparison