Hop Limited Flooding over Dynamic Networks

Milan Vojnović and Alexandre Proutiere

Microsoft Research
Cambridge, UK
{milanv, aproutie}@microsoft.com

Technical Report
MSR-TR-2010-97

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

Abstract— We study the performance of hop-limited broadcasting of a message in dynamic graphs where links between nodes switch between active and inactive states. We analyze the performance with respect to the completion time, defined as the time for the message to reach a given portion of nodes, and the communication complexity, defined as the number of message forwarding per node. We analyze two natural flooding algorithms. First is a lazy algorithm where the message can be forwarded by a node only if it was first received by this node through a path shorter than the hop limit count. Second is a more complex protocol where each node forwards the message at a given time, if it could have been received by this node through a path shorter than the hop limit count. We derive exact asymptotics for the completion time and the communication complexity for large network size which reveal the effect of the hop limit count. Perhaps surprisingly, we find that both flooding algorithms perform near optimum and that the simpler (lazy) algorithm is only slightly worse than the other, more complicated algorithm.

The results provide insights into performance of networked systems that use hop limits, for example, in the contexts of peer-to-peer systems and mobile ad-hoc networks.

I. Introduction

We consider information dissemination over dynamic graphs where edges between nodes are randomly activated over time, and where two nodes communicate only when the edge connecting these nodes is active. These problems arise in many networking contexts, including peer-to-peer systems where by design, node neighborhood sets are dynamic, and delay-tolerant wireless mobile networks where connectivity between nodes changes over time due to mobility and finite radio communication range.

We are interested in analyzing gossip-like spreading algorithms to broadcast a message initially held by a set of sources to all other nodes (or to a significant proportion of nodes) in the network. More precisely, we investigate flooding algorithms where each contact between two nodes can be exploited to disseminate the message. The only constraint imposed on the algorithms is the message *time-to-live* (TTL): the message received by a node cannot have been transmitted more than a fixed number, say k, times. Another way of expressing this constraint is to say that a message held by a node is attached an *age* that corresponds to the number of hops of the path through which this message was received. Then when two nodes are in contact, messages with age strictly smaller than k only can be transmitted. This constraint is natural and limiting the hop count of messages is a standard approach for flooding algorithms used in networking. For example, in wireless mobile networks, this approach has been proposed to improve the performance of information dissemination: allowing nodes to receive the message not only via direct contacts with sources but also through one relay node (2-hop flooding scheme) significantly enhances the system performance in terms of delay and capacity [11], [21]. In general, the hop count limit is commonly used to limit the scope of broadcasting to deal with the broadcast storm problem [25], or to reduce the protocol complexity [24], [6] and has been used in many protocol designs, e.g. [23], [14], [9], [12], and even supported through standardization [8]. Surprisingly, flooding algorithms with limited hop count have not been analyzed theoretically. In this paper, we characterize the performance of such algorithms in terms of completion time (the time it takes for a given proportion of nodes to receive the message), and also evaluate their communication costs (e.g. the number of message transmissions made by a node before completion).

Two natural algorithms for hop limited flooding are studied. Our first algorithm is simply referred to as *k-hop limited flooding*. Under this algorithm when two nodes, say A and B are in contact: (i) if one node only, e.g. A, holds a copy of the message of age $j < k$, then A forwards the message to B, and the age of the message held by B is set to $j + 1$; (ii) if both nodes hold a copy of message of respective ages i and j (less than k) at A and B, then these ages are updated to $\min(i, j + 1)$ and $\min(j, i + 1)$, respectively. Our second algorithm, referred to as *lazy k-hop limited flooding*, is similar to the first algorithm except that updating the ages of messages held at nodes as described in (ii) above is not allowed. This means that the age of the message held by a node remains unchanged after the node first receives the message, and that this age corresponds to the length of the path through which the message was first received. Notice that the latter algorithm is more parsimonious as every node that became informed through a path of length k never forwards the message to other nodes. Lazy k-hop limited flooding is simpler to implement than k-hop limited flooding, as nodes that received the message do not need to update the age of their copy of the message.

Our analysis concerns dynamic graphs where for a set of n nodes, every edge connecting a pair of nodes is active at instances of a Poisson process of rate $1/(n - 1)$. This class of dynamic graphs has been considered in previous work [3], [16], [1], [4], [5], [19], and is a continuous-time analog of discrete-time dynamic graphs defined as a sequence of independent and identically distributed Erdos-Renyi random graphs. The use of these dynamic graphs where each pair of nodes is eventually in contact is natural as a model of some mobile ad-hoc networks, but also, more generally, to model peer-to-peer systems with so-called *topology independence* property [10].

Summary of our Contributions. We study the performance of hop limited flooding algorithms in the general case where the initial proportion of nodes holding a copy of the message and the targeted final proportion of informed nodes are arbitrary. But here for illustrative purposes, we summarize our results only when a single copy of the message is initially present in the network, and when all nodes should finally get the message.

- We derive tight asymptotic estimates of the completion time $\bar{t}_{n,k}$ of the k-hop limited flooding and the lazy k-hop limited flooding algorithms as the size of the network n grows large. Specifically, we show that for k-hop flooding,[1]

$$\bar{t}_{n,k} \overset{n\to\infty}{\sim} (k!n)^{1/k} \cdot \log^{1/k}(n)$$

while for lazy k-hop flooding,

$$\bar{t}_{n,k} \overset{n\to\infty}{\sim} (k!n)^{1/k} \cdot \log(n^{1/k}).$$

These show that for every fixed hop-count limit $k > 0$, both algorithms yield optimum completion time up to a polylogarithmic factor as the completion time of any k-hop limited flooding scheme under our communication model is $\Omega((k!n)^{1/k})$.[2] The above results recover the coupon collector result as a special case for $k = 1$; for general values of k, note that the scaling for lazy k-hop flooding is like with the coupon collector but with $n^{1/k}$ coupons.

• For every fixed hop count $k > 0$, the competitive ratio of the completion time under k-hop flooding and lazy k-hop flooding is polylogarithmic, specifically $\frac{1}{k}\log^{1-1/k}(n)$, which is at most $O\left(\frac{\log(n)}{\log(\log(n))}\right)$. Hence although the lazy k-hop limited flooding algorithm is simpler and more parsimonious, it yields similar completion time as the k-hop limited flooding algorithm.

• The communication costs of sources satisfy: for k-hop limited flooding,

$$\bar{c}_{n,k} \overset{n\to\infty}{\sim} (k!n)^{1/k} \cdot \frac{\Gamma\left(\frac{1}{k}\right)}{k}$$

and for lazy k-hop limited flooding,

$$\bar{c}_{n,k} \overset{n\to\infty}{\sim} (k!n)^{1/k}.$$

• Lazy k-hop limited flooding exhibits similar communication costs as k-hop limited flooding, as the corresponding competitive ratio is $k/\Gamma(1/k)$ whatever the network size n is. This ratio is equal to $2/\sqrt{\pi} \approx 1.15$ for $k = 2$ and decreases to 1 with k, for $k \geq 2$.

• The above scalings are obtained by considering ordinary differential equation (ODE) systems that are rigorous limits of our stochastic systems as the network size grows large. We obtained more refined results by direct stochastic analysis for the special cases of one- and two-hop flooding. These results not only conform to the above asymptotics but for lazy k-hop flooding we were also able to characterize the variance and concentration results for the completion time.

• Our simulation results suggest that the derived asymptotes are good approximations already for small values of network size n.

Outline of the Paper. In Section II we discuss related work. Section III presents the algorithms and network model in more detail. In Section IV, we derive the deterministic ODE systems characterizing the limiting behavior of the stochastic systems as the network size grows large. Section V-A contains main results for k-hop flooding (Theorem 2 and Theorem 3) while Section V-B contains main results for lazy k-hop flooding (Theorem 4 and Theorem 5 along with concentration results). In Section VI, we derive estimates of communication costs for k-hop limited flooding (Theorem 7) and lazy k-hop limited flooding (Theorem 8). Section VII presents some numerical experiments to illustrate our analytical results. Finally, in Section VIII we conclude. We deferred proofs of all our results to Appendix.

[1] Hereinafter, for two sequences a_n and b_n, we write $a_n \overset{n\to\infty}{\sim} b_n$ meaning that a_n/b_n goes to 1 as n grows large.

[2] More precisely, assume that all nodes are always in contact, but that in each time unit, a node can transmit the message to at most one node, and impose the k-hop limit, then any dissemination scheme has a completion time $\Omega((k!n)^{1/k})$.

II. Related Work

Broadcast and gossip problems have had a central role in the context of networked and distributed systems and have been studied under various assumptions; e.g. see [13] and references therein for a survey of the results on the broadcast and gossip over static graphs. Specific results include lower bounds such as $\lceil \log_2(n) \rceil$ for the broadcast time for any connected graph of n vertices and upper bounds for specific graphs. Epidemic-style algorithms for broadcast or gossip problems have been considered under various names such as rumor or gossip spreading with early works including [7], [20], [15]. Epidemic-style algorithms provide a lightweight and robust approach for information dissemination with optimum-order completion time. Our work is different from this line of work in that we consider gossip spreading over limited hop paths. The case of limited hop paths is of interest in practice in the context of peer-to-peer systems and mobile ad-hoc networks, which we discussed in Section I.

Our work is closely related to the recent line of work on distributed computation over dynamic graphs [16], [1], [4], [5], [19]. In [1], [4], a class of parsimonious flooding is studied where each node upon receiving the message attempts to forward it for a limited number of time slots. Therein, the dynamic graph is over discrete time where each edge is activated according to a two-state Markov chain. The authors characterize the completion time in terms of the parameters of the edge-activation process and characterize the required number of slots for the message to reach all nodes almost surely. This line of work does not impose a limit on the number of hops through which a message can reach a node and is thus different from ours. Note also that in our setting every node is guaranteed to receive the message in a finite time.

Another related work is [3] that considered the diameter of dynamic graphs in discrete and continuous time where the latter corresponds to the dynamic graph considered in the present paper. Again, this work does not impose a limit on the number of hops through which the message can reach a node. Furthermore, the work is also different with respect to the metric considered. [3] considers the expected number of paths between a pair of nodes over a time interval, whereas to evaluate e.g. the completion time, we need to characterize the probability that a path exists between a pair of nodes. Yet another example of related work is [2] where the authors analyze the delay of gossiping algorithms in dynamic graphs, but again for algorithms without any hop-count limit.

Finally, it is noteworthy that the problem that we study could also be seen as a generalization of the coupon collector problem [17][Section 3.6] and [18]. For example, the dynamics of our 1-hop limited flooding is equivalent to that of standard coupon collector problem.

III. Model and Algorithms

We consider the problem of disseminating a message to nodes that communicate through a random dynamic graph with a fixed number n of nodes. Each edge in the graph is activated at random times and is attached a weight representing the intensity of its activation process. Activation processes of edges are stochastically independent. We restrict our attention to the case where edges are activated according to a Poisson process of intensity $1/(n-1)$. For example, this holds if each node is equipped with a Poisson clock ticking at rate 1/2, and when its clock ticks, it makes contact with a node uniformly chosen at random. Alternatively, this is an assumption that may well approximate, in mobile networks, an underlying process of inter-contact times between nodes in practice (see, e.g., [3]).

Two nodes may exchange the message only when the corresponding edge is activated. We say that a node is *informed* if it holds a copy of the message, and *uninformed* otherwise. Initially, only one node or a small proportion of nodes is informed. We are interested in designing decentralized broadcast algorithms that inform all nodes as fast as possible and with as few transmissions as possible. To assess the performance of broadcast algorithms, we use several metrics. First we are interested in the *completion time* defined as the time it takes for a given proportion of nodes to be informed. Then we consider the *communication cost* defined as the total number of message transmissions per node before completion, and in particular, the maximum number of message transmissions per node until completion.

For example, if one node only, the source, is initially informed, and if other nodes can get the message through direct contact with the source, then we recover the classical coupon collector model. The average completion time then scales as $n\log(n)$ when n grows large and the maximum per node communication cost is of course equal to $n-1$. On the other hand, if every contact between two nodes are used to transmit the message (no hop count limit), the completion time is $\Theta(\log(n))$ and the maximum communication cost per node is $\Theta(\log(n))$.

We aim at analyzing broadcast algorithms for which the maximum number of transmissions per node is limited. In [1], the authors propose to limit transmissions by imposing that a node stops retransmitting the message after a fixed amount of time k. With this constraint, the dissemination does not always complete. However, if k is large enough, it then completes almost surely, and the completion time can be characterized. In this paper, we analyze algorithms that use the so-called *age* or TTL (*Time-To-Live*) of the message held at a given node to limit the intensity of the message spreading. By definition, the age of the message held at a node is the number of times the message has been transmitted to reach this node. More precisely, we consider algorithms for which the age of messages is bounded by $k \geq 1$. These algorithms are guaranteed to complete for all k.

k-hop limited flooding. When two nodes are in contact: (i) if one node only has the message and its age is $j < k$, then the message is transmitted to the other node, and the age of this new message copy is $j+1$; (ii) if both nodes have the message with ages i and j respectively, if $i < j$, the message age at the node with age j is updated to $i+1$.

Lazy k-hop limited flooding. When two nodes are in contact: if one node only has the message and its age is $j < k$, then the message is transmitted to the other node, and the age of this new message copy is $j+1$. Note that under this algorithm, the age of message held by a node corresponds to the number of hops of a path through which the message could first reach this node.

Note that the 1-hop limited flooding algorithm and its lazy version coincide and actually correspond to the coupon collector model. Next we define the metrics used to assess the performance of the algorithms depending on parameter k.

Completion times. Let $1/a_n$ be the initial proportion of informed nodes, and assume that the objective of our algorithms is to reduce as fast as possible the proportion of uninformed nodes from $1-1/a_n$ to at most $(1-1/a_n)/b_n$. Typically, we will be interested in the case where a single node is informed $a_n = n$, and where all nodes are ultimately informed[3] $b_n = n$. We denote by $t_{n,k}$ the (deterministic) time it takes for the k-hop limited flooding algorithm (or its lazy version) to reach this proportion on average. We also introduce $T_{n,k}$ as the (random) time it takes for the algorithm to reach this proportion. As we shall see later on, we believe that the system dynamics obeys a concentration principle, in the sense that almost surely $\lim_{n\to\infty} T_{n,k}/t_{n,k} = 1$, which we confirm for special cases $k = 1$ and $k = 2$ by comparison with analysis of underlying stochastic processes.

Communication cost. The communication cost is defined as the maximum number of message forwarding per node required until completion, i.e., from time 0 to $t_{n,k}$.

IV. System dynamics and asymptotics

Let us denote by $Q_i(t)$ the number of nodes holding a copy of the message with age at most i at time t. Then, under both k-hop limited flooding and lazy k-hop limited flooding algorithms, $(Q(t) = (Q_0(t), \ldots, Q_k(t)), t \geq 0)$ is a continuous-time Markov process. Next we give the transition rates of this Markov process under the two proposed algorithms. In

[3] Actually, if $b_n = n$, the target proportion of informed node is roughly equal to $1 - 1/n$, so only one node remains uninformed. It can be easily shown that the time it takes to inform the last node is always negligible compared to the time it takes to get a proportion of uninformed nodes equal to $(1 - 1/a_n)/b_n$.

what follows, for all $u \leq v$, we denote by $e_{u,v}$ the $(k+1)$-dimensional binary vector whose coordinates are all equal to 0 except for coordinates $u, u+1, \ldots, v$.

k-hop limited flooding. The transition $Q \to Q + e_{i,j-1}$, for $1 \leq i < j \leq k$, occurs when an edge between a node holding a copy of the message with age j and a node with a copy with age $i-1$ is activated. This happens at rate $(Q_{i-1} - Q_{i-2})(Q_j - Q_{j-1})/(n-1)$. Similarly, a transition $Q \to Q + e_{i,k}$ is triggered by contacts of a node that does not hold the message with a node holding a copy with age $i-1$. These contacts occur at rate $(Q_{i-1} - Q_{i-2})(n - Q_k)/(n-1)$. The transition rates of the Markov process are then (using the convention $Q_{-1} = 0$):

$$Q \to \begin{cases} Q + e_{i,j-1} & : (Q_{i-1} - Q_{i-2}) \frac{Q_j - Q_{j-1}}{n} 1_{1 \leq i < j \leq k} \\ Q + e_{i,k} & : (Q_{i-1} - Q_{i-2}) \frac{n - Q_k}{n}. \end{cases}$$

Note that as a consequence, for all $i = 1, \ldots, k$, Q_i evolves as

$$Q_i \to Q_i + 1 \ : Q_{i-1} \frac{n - Q_i}{n}.$$

In an interval of time of duration dt, the expected drift of the proportion of nodes holding a copy of the message with age less than or equal i is then $\frac{Q_{i-1}}{n}(1 - \frac{Q_i}{n})dt$. Assume now that we scale the system so as the proportion of nodes holding the packet initially converges when the system size grows large ($n \to \infty$). We then expect that when the system size grows, by virtue of the law of large numbers, the proportion of nodes holding a message with age less than or equal i evolves as:

$$\frac{d}{dt} q_i(t) = q_{i-1}(t)(1 - q_i(t)), \text{ for } i = 1, \ldots, k, \quad (1)$$

where for all t, $q_0(t) = \lim_{n \to \infty} Q_0(0)/n$ (this limit exists by assumption). We provide a theoretical justification of (1) in Theorem 1.

Lazy k-hop limited flooding. Under this algorithm, the system state changes only at contacts between a node holding a copy of the message with age less than k, and an uninformed node. The transition rates are in this case:

$$Q \to Q + e_{i,k} \ : (Q_{i-1} - Q_{i-2}) \frac{n - Q_k}{n}.$$

It is not difficult to check that for all $i = 1, \ldots, k$,

$$Q_i \to Q_i + 1 \ : Q_{i-1} \frac{n - Q_k}{n}.$$

Finally, as the system size grows large, the proportion of nodes holding a message with age of at most i evolves as:

$$\frac{d}{dt} q_i(t) = q_{i-1}(t)(1 - q_k(t)), \text{ for } i = 1, \ldots, k, \quad (2)$$

where again for all t, $q_0(t) = \lim_{n \to \infty} Q_0(0)/n$

The asymptotic system evolution when the system size grows large can be formally justified using Kurtz's theorem (e.g. [22]). Specifically, we have:

Theorem 1: Fix a finite time horizon T. Assume that $Q(0)/n$ tends to $q(0)$ almost surely as $n \to \infty$. We have:

$$\lim_{n \to \infty} \sup_{t \in [0,T]} \left\| \frac{Q(t)}{n} - q(t) \right\| = 0, \quad \text{almost surely,} \quad (3)$$

where under the k-hop limited algorithm (resp. its lazy version), $q(\cdot)$ is the unique solution of (1) (resp. (2)) with initial condition $q(0)$ and with $q_0(t) = q_0(0)$ for all t.

In the following, we use the deterministic asymptotic system behavior to derive estimates of the completion time of the proposed algorithms. More precisely, we use the following heuristic estimate of $t_{n,k}$:

$$t_{n,k} \overset{n \to \infty}{\sim} \bar{t}_{n,k} = \inf \left\{ t \geq 0 : q_k(t) = 1 - \frac{1 - 1/a_n}{b_n} \right\} \quad (4)$$

where $q(\cdot)$ is the unique solution of (1) (or (2) depending on the algorithm considered) with initial condition $q(0) = (1/a_n, 1/a_n, \ldots, 1/a_n)$, and with $q_0(t) = 1/a_n$ for all $t \geq 0$. There are two difficulties in proving (4) theoretically. First as we shall see, the completion time grows large when $n \to \infty$, whereas Kurtz's concentration result is valid over finite time horizons only, and hence does not provide a sufficient justification. Then, note that if the initial condition $Q(0)/n$ tends to 0 as $n \to \infty$, in this case, when applying Theorem 1, we would conclude that the system does not evolve at all, i.e., $q_i(t) = 0$ for all $i \geq 1$. In this paper, we do not explain how to circumvent these problems. Instead we provide direct proofs of (4) for the specific cases $k = 1$ and $k = 2$ that yield exactly the same asymptote as n grows large as that of $\bar{t}_{n,k}$.

To conclude this section, we present a simple lower bound on the performance of both k-hop limited flooding algorithm and its lazy version, assuming that (4) holds. Observe that for both algorithms:

$$\frac{d}{dt} q_i(t) \leq q_{i-1}(t), \text{ for } i = 1, \ldots, k.$$

We deduce that for all $i = 1, \ldots, k$, and all time t: $q_i(t) \leq \bar{q}_i(t)$ where $\bar{q}(\cdot)$ is the solution of the following system of ordinary differential equations $\frac{d}{dt}\bar{q}_i(t) = \bar{q}_{i-1}(t)$, for $i = 1, \ldots, k$, and $\bar{q}_0(t) = q_0(t)$ for every $t \geq 0$. We then have:

$$q_i(t) \leq \bar{q}_i(t) = \frac{1}{a_n} \sum_{j=0}^{i} \frac{t^j}{j!}. \quad (5)$$

Using this with (4) yields the following lower bound.

Lemma 1: Assume that $\lim_{n \to \infty} a_n = \infty$. Then, under both k-hop limited flooding and lazy k-hop limited flooding algorithms we have:

$$\bar{t}_{n,k} \geq (k! a_n)^{1/k} \cdot \left(1 - \frac{1}{b_n}\right)^{\frac{1}{k}} (1 - o(1)). \quad (6)$$

Note that the bound derived in the above lemma does not depend on b_n provided that $b_n \to \infty$ as $n \to \infty$, which indicates that it can be very crude. Essentially, as we prove later on, the bound captures only the time it takes for the system to evolve from a state with a negligible proportion a_n of informed nodes to a state where a strictly positive proportion $\alpha < 1$ of nodes are informed. As we shall demonstrate, starting from a fixed proportion $\alpha > 0$ of informed nodes, the time it takes for the system to approach a proportion $1/b_n$ of informed nodes grows large when $n \to \infty$.

V. Completion Time

A. Completion Time under k-hop Limited Flooding

In this section, we first provide asymptotic estimates of the completion time under the k-hop limited flooding algorithm using the approximation (4). Then we validate the approximation through a direct stochastic system analysis in the specific cases where $k = 1$ and $k = 2$.

Completion time asymptotics. To evaluate the completion time, we use the approximation $\bar{t}_{n,k}$, defined in (4). First notice that in view of (1), we have: for all $i = 1, \dots, k$,

$$q_i(t) = 1 - \left(1 - \frac{1}{a_n}\right) \exp\left(-\int_0^t q_{i-1}(s)ds\right). \quad (7)$$

Let us define $r_i(t) = \int_0^t q_i(s)ds$. Notice that $r_i(t)$ can be interpreted as the number of connection attempts by a node that received the message through at most i hops over the time interval $[0, t]$. From (7), we deduce that $r_0(t) = t/a_n$ and that $r_i(t)$, for $i = 1, \dots, k$, is recursively defined by:

$$r_i(t) = t - \left(1 - \frac{1}{a_n}\right) \int_0^t e^{-r_{i-1}(s)} ds. \quad (8)$$

To derive tight estimates of $\bar{t}_{n,k}$, we work with r_i's instead of directly working with q_i's. The completion time $\bar{t}_{n,k}$ satisfies $q_k(\bar{t}_{n,k}) = 1 - (1 - \frac{1}{a_n})\frac{1}{b_n}$, and hence:

$$r_{k-1}(\bar{t}_{n,k}) = \log(b_n). \quad (9)$$

Theorem 2: Assume that either a_n or b_n are increasing sequences. Then, the completion time $\bar{t}_{n,k}$ under k-hop limited flooding satisfies

$$\bar{t}_{n,k} \geq (k! a_n)^{1/k} \cdot \log^{1/k}(b_n) \cdot (1 - o(1)).$$

Furthermore, if a_n and b_n are such that $\log(b_n) = o((k!/k^k \cdot a_n)^{1/(k-1)})$, then

$$\bar{t}_{n,k} \leq (k! a_n)^{1/k} \cdot \log^{1/k}(b_n) \cdot (1 + o(1)).$$

Note that if b_n is bounded, the completion time scales as $C \cdot a_n^{1/k}$ as $n \to \infty$ for a constant $C > 0$ which validates the observation made in the previous section that the lower bound derived in Lemma 1 essentially captures the time for message to reach a fixed strictly positive proportion $\alpha < 1$ of nodes. Theorem 2 implies that when b_n grows to infinity, the completion time is much larger than this crude initial bound.

Note that when $k = 1$, if $a_n = b_n = n$, the completion time scales as $n \log(n)$ which is expected since the system evolves as that in the coupon collector model. This observation also shows that for $k = 1$, the approximation (4) holds, i.e., $t_{n,1} = \bar{t}_{n,1}$. Further remark that plugging $k = \log(n)$ in $(k! a_n)^{1/k} \cdot \log^{1/k}(b_n)$, we conclude that the completion time scales as $C \cdot \log(n)$ for a constant $C > 0$ (if $a_n = b_n = n$), and the algorithm becomes optimal in terms of completion time (up to a multiplicative constant).

Stochastic analysis. We present a direct stochastic analysis for special cases of one- and two-hop limited flooding. This enables us to compare with the general result of Theorem 2 which was derived using the limit ODE system.

a) One-hop flooding: The case of one-hop limited flooding is special in that it boils down to exactly the same dynamics under both k-hop limited flooding and lazy k-hop limited flooding. Moreover, this case is rather simple and amenable for detailed analysis. We will note that this case is intimately related to the well known coupon collector problem [17].

Recall that initially n/a_n nodes are informed and the remaining $n(1 - 1/a_n)$ of the nodes are uniformed. Let I and U denote the respective sets of initially informed and uninformed nodes, thus $|I| = n/a_n$ and $|U| = n(1 - 1/a_n)$. Let $X_{u,v}$ be the first time when node $u \in U$ and node $v \in I$ are in contact. Notice that $X_{u,v}$ are independent and identically distributed random variables whose distribution is exponential with mean $n - 1$. Every node $u \in U$ becomes informed at the first contact with a node from the set I, therefore, this occurs at the time $X_u = \min\{X_{u,v} : v \in I\}$. It is easy to note that X_u are independent and identically distributed random variables whose distribution is exponential with mean $a_n(n-1)/n$. The expected number of non-informed nodes at time t is given by $\mathbb{E}(\sum_{u \in U} 1_{X_u > t}) = n(1 - 1/a_n) \exp(-\frac{n}{(n-1)a_n} t)$, for every $t \geq 0$. It follows that for the completion time $t_{n,1}$ we have

$$n\left(1 - \frac{1}{a_n}\right) e^{-\frac{n}{(n-1)a_n} t_{n,1}} = n\left(1 - \frac{1}{a_n}\right) \frac{1}{b_n}.$$

Hence,

$$t_{n,1} = a_n \log(b_n) \cdot \left(1 - \frac{1}{n}\right)$$

which yields exactly the same asymptotic as given in Theorem 2 for $k = 1$.

Note that for the case of single initially informed node, the dynamics is essentially that of the coupon collector; the source node samples other nodes uniformly at random with replacement and the time it takes to contact all nodes is equivalent to sampling each node at least once. In this case, we recover the scaling $n \log(n)$ of the coupon collector problem. The case with more than one initially informed node is slightly different as now each initially uninformed node has to be sampled at least once by any of the initially informed nodes. Nevertheless, the above simple analysis reveals that the dynamics is essentially the same as for the case of a single initially informed node with the only difference that the evolution is speeded up proportional to the number of sources n/a_n.

b) Two-hop flooding: Theorem 2 identifies the asymptotically dominant term of the completion time, provided that the approximation (4) holds. In the following, we provide a justification of this approximation for the special case $k = 2$ via a direct analysis of the stochastic system.

Theorem 3: Suppose that, initially, the message is held by one node, hence $a_n = n$. Then we have:

$$t_{n,2} \overset{n\to\infty}{\sim} \sqrt{2n\log(b_n)}.$$

In particular, if the completion time $t_{n,2}$ is defined as the time at which the expected number of uninformed nodes is 1 (i.e. $b_n = n$), we have $t_{n,2} \overset{n\to\infty}{\sim} \sqrt{2n\log(n)}$.

The proof available in Appendix exploits the fact that every fixed initially uninformed node can become informed by receiving the message from the source through the set of edge-disjoint paths consisting of the single hop path connecting the source and the destination node and two hop paths passing through distinct relay nodes.

B. Completion Time under Lazy k-hop Limited Flooding

We now turn our attention to the lazy k-hop limited flooding algorithm. Again, we first provide asymptotic estimates of its completion time for any arbitrary k assuming that the approximation (4) holds, and then present a direct stochastic system analysis for the special case $k = 2$.

Completion time asymptotics. Assume that the approximation (4) holds so that $\bar{t}_{n,k}$ is an estimate of the completion time. We show that:

Theorem 4: Assume that a_n is an increasing sequence. Then, the completion time $\bar{t}_{n,k}$, for the lazy k-hop limited flooding satisfies

$$\bar{t}_{n,k} = (k!a_n)^{1/k} \cdot [\log(b_n^{1/k}) + C_k] + O(1)$$

where

$$C_k = \frac{1}{k}\left(\log(k) + \int_0^1 \frac{\sum_{j=0}^{k-2}(k-1-j)x^j}{\sum_{j=0}^{k-1} x^j}dx\right).$$

The constant C_k is equal to $\log(2)$ for $k = 2$ and diminishes to zero with k, for $k \geq 2$. Indeed, for every fixed $k \geq 1$, the result implies $\bar{t}_{n,k} \overset{n\to\infty}{\sim} (k!a_n)^{1/k} \cdot \log(b_n^{1/k})$, provided that b_n is an increasing sequence.

Stochastic analysis and concentration results. We consider the case of lazy two-hop limited flooding by direct stochastic analysis. Specifically, we consider the time to inform all nodes, $T_{n,2}$, and provide lower and upper bounds for the expected value $\mathbb{E}(T_{n,2})$. This will show that the asymptotically dominant term is exactly that of Theorem 4, for $k = 2$. We then provide an asymptotically tight estimate for the variance and a concentration result for the random completion time $T_{n,2}$.

Theorem 5: The expected completion time satisfies the following

$$-O(\frac{a_n}{n}\log(n)^2) \leq \mathbb{E}(T_{n,2}) - \sqrt{2a_n}\log(\sqrt{n}) \leq O(\sqrt{a_n}).$$

Therefore, $\mathbb{E}(T_{n,2}) \overset{n\to\infty}{\sim} \sqrt{2a_n}\log(\sqrt{n})$.

The proof of the theorem relies on analysis of an embedded Markov chain and is available in Appendix. Similar type of analysis is carried to establish the following bound for the variance of the completion time $T_{n,2}$.

Theorem 6: The variance of the completion time satisfies the following

$$\lim_{n\to\infty} \frac{\mathrm{Var}(T_{n,2})}{a_n} \leq \frac{\pi^2}{12}.$$

Furthermore, the inequality is asymptotically tight provided that a_n grows faster than $\log^2(n)$.

From the last two results we observe that if initially a single node is informed, i.e. $a_n = n$, the expected completion time scales faster than the standard deviation of $T_{n,2}$ for a logarithmic factor with n. This suggests that $T_{n,2}$ concentrates around the expected value $\mathbb{E}(T_{n,2})$ as n grows large. This can be formally claimed using Chebyshev's inequality along with the last two theorems:

Corollary 1: Suppose $a_n = n$ and let d_n be an increasing sequence of positive numbers. Then, there exists $n_0 > 0$ such that for every $n \geq n_0$, we have

$$\mathbb{P}(|T_{n,2} - \sqrt{2n}\log(\sqrt{n})| \leq \sqrt{nd_n} + O(\sqrt{n})) \geq 1 - \frac{\pi^2}{12d_n}.$$

C. Discussion

In this section we compare the completion times for k-hop limited flooding and lazy k-hop limited flooding. Since lazy k-hop limited flooding is more parsimonious in that a node that became informed by receiving the message through a path of k hops never forwards the message, it is clear that the completion time of lazy k-hop limited flooding is at least that of k-hop limited flooding. The following is a corollary of Theorem 2 and Theorem 4.

Corollary 2: For the completion times of the lazy k-hop limited flooding $\bar{t}^{\text{lazy}}_{n,k}$ and the completion of the k-hop limited flooding $\bar{t}_{n,k}$, we have the following competitive ratio

$$\frac{\bar{t}^{\text{lazy}}_{n,k}}{\bar{t}_{n,k}} \overset{n\to\infty}{\sim} \frac{1}{k}\log^{1-\frac{1}{k}}(b_n). \tag{10}$$

Furthermore, the largest (over values of k) competitive ratio is $O(\frac{\log(b_n)}{\log(\log(b_n))})$.

In particular, if the completion time is defined as the time to reach almost all nodes, we have $b_n = n$ and, then, the competitive ratio is $O(\frac{\log(n)}{\log(\log(n))})$.

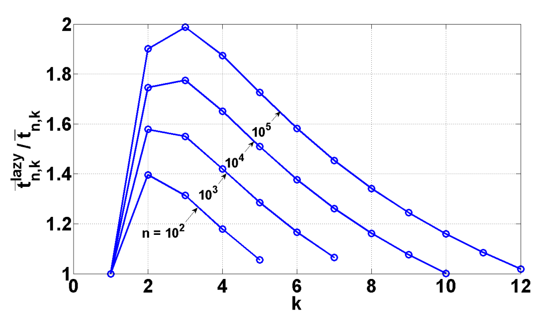

Figure 1. The competitive ratio of the completion times under lazy k-hop limited flooding and k-hop limited flooding.

Numerical example. In Figure 10, we show the ratio of the asymptotic completion times for the lazy k-hop limited flooding in Theorem 4 and k-hop limited flooding in Theorem 2. For given network size n, the ratio achieves a maximum value and then diminishes to value 1, which is reached at the diameter of the network of order $\log(n)$.

VI. Communication complexity

In this section we provide estimates of the communication complexity of the k-hop and lazy k-hop limited flooding algorithms. More precisely we are interested in quantifying the cost in terms of number of transmissions a particular node has to do so as to help the message broadcast (before the completion time). Ideally, the total transmission cost of broadcasting the message, should be fairly shared among nodes. Let us define $c_{n,k}(s)$ as the number of transmissions of a node able to forward the message from time s until the completion time. For example, $c_{n,k}(0)$ is the number of transmissions made by sources. As previously, we use the deterministic asymptotic system behavior to estimate $c_{n,k}(s)$. The latter is approximated by:

$$\overline{c}_{n,k}(s) = \int_s^{t_{n,k}} (1 - q_k(t))dt. \tag{11}$$

In the above expression, $1-q_k(t)$ is the rate at which the node meets an uniformed node, and hence forwards the message. Note that in the case of k-hop limited flooding, when two nodes holding a copy of the message meet, they do not need to actually transmit the message; they may just update the age of their copies.

Let us first briefly analyze the extreme cases $k = 1$ (coupon collector) and $k = \infty$ (no hop limit). When $k = 1$, only sources can disseminate the message, and we easily obtain:

$$\overline{c}_{n,1}(0) \overset{n\to\infty}{\sim} a_n \left(1 - \frac{1}{b_n}\right) \text{ and } \overline{c}_{n,1}(s) = 0, \text{ for } s > 0.$$

For $k = \infty$, we obtain:

$$\overline{c}_{n,\infty}(s) \overset{n\to\infty}{\sim} \log(1 + a_n e^{-s}), \text{ for } s \geq 0.$$

Indeed, if exactly one node initially holds the message (i.e. $a_n = n$), then for the latter two cases we have that the communication cost for the source is order n and order $\log(n)$, respectively.

A. k-hop Limited Flooding

We now evaluate the communication cost under k-hop limited flooding algorithm, with $k > 1$. Note that under k-hop limited flooding, when a node receives the message, the age of the message held by the node can be either k, or strictly less than k. In the former case, the node does not forward the message unless it manages to decrease the age of the message copy later. In the latter case, the node forwards the message until the completion time. Let s be an instant where the age of the message at a node becomes strictly less than k (this happens either because the node gets a copy of a message with age less than $k-1$, or because the node updates the message age). In the following theorem, we evaluate the asymptotic (as n grows large) cost $\overline{c}_{n,k}(s)$. More precisely, we consider an increasing sequence of times s_n and estimate $\overline{c}_{n,k}(s_n)$.

Theorem 7: Assume that $\lim_{n\to\infty} a_n = \infty = \lim_{n\to\infty} b_n$ and that $\lim_{n\to\infty} \log(b_n)/a_n = 0$. Under k-hop limited flooding, the communication cost for a node from a time s_n when the age of its message is strictly less than k to the completion time satisfies:
(i) If $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \alpha < \infty$,

$$\overline{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot \frac{\Gamma(\frac{1}{k}, \alpha^k)}{k} \cdot (1 + o(1)), \tag{12}$$

where $\Gamma(x, \alpha) = \int_\alpha^\infty u^{x-1}e^{-u}du$ is the incomplete Gamma function.
(ii) If $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \infty$ and $\lim_{n\to\infty} s_n/\overline{t}_{n,k} = 0$,

$$\overline{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot \frac{\exp\left(-\frac{s_n^k}{k!a_n}\right)}{k s_n^{k-1}} \cdot (1 + o(1)). \tag{13}$$

The above theorem implies that the communication cost of sources satisfies:

$$\overline{c}_{n,k}(0) = (k!a_n)^{1/k} \cdot \frac{\Gamma(\frac{1}{k})}{k} \cdot (1 + o(1)).$$

This cost is equal to $(k!a_n)^{1/k}$ up to a fixed multiplicative factor $\Gamma(1/k)/k$ (independent of n). Actually, as illustrated later in this section, for $k > 1$, $\Gamma(1/k)/k$ is an increasing function of k and is approximately equal to 0.89 for $k = 2$ and tends to 1 when k grows large.

B. Lazy k-hop Limited Flooding

We now estimate the communication costs under the lazy k-hop limited flooding algorithm. Under this algorithm, when a node receives for the first time a copy of the message, it will forward the message ever after only if the age of the message is strictly less than k. In the following theorem, we evaluate the asymptotic (as n grows large) cost $\overline{c}_{n,k}(s_n)$ for an increasing sequence s_n.

Theorem 8: Assume that $\lim_{n\to\infty} a_n = \infty = \lim_{n\to\infty} b_n$. Under lazy k-hop limited flooding, the communication cost for a node from a time s_n when the age of its message is strictly less than k to the completion time satisfies:
(i) If $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \alpha < \infty$,

$$\overline{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot (1 - \psi_k^{-1}(\alpha) + o(1)) \tag{14}$$

where the function $\psi_k(\cdot)$ is defined by $\psi_k(x) = \int_0^x du/(1 - u^k)$, for $0 \leq x < 1$.
(ii) If $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \infty$ and $\lim_{n\to\infty} s_n/\overline{t}_{n,k} = 0$,

$$\overline{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot \Theta\left(\exp\left(-\frac{k s_n}{(k!a_n)^{1/k}}\right)\right). \tag{15}$$

It should be noted that $\psi_k(0) = 0 = \psi_k^{-1}(0)$, and hence, the communication cost of sources satisfies:

$$\overline{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot (1 + o(1)).$$

C. Discussion

We next compare the communication costs under the k-hop and lazy k-hop limited flooding algorithms. We are interested in nodes having the largest communication costs, i.e., sources or nodes that receive the message early (at a time that scales as $o((k!a_n)^{1/k})$). For these nodes the communication cost is $\overline{c}_{n,k}(0)$. From the two previous theorems, we deduce that:

Corollary 3: For the communication costs of sources under k-hop limited flooding $\overline{c}_{n,k}(0)$ and under lazy k-hop limited flooding $\overline{c}_{n,k}^{\text{lazy}}(0)$, we have the following competitive ratio:

$$\frac{\overline{c}_{n,k}^{\text{lazy}}(0)}{\overline{c}_{n,k}(0)} \overset{n\to\infty}{\sim} \frac{k}{\Gamma(\frac{1}{k})}. \quad (16)$$

Remark that the above competitive ratio does not depend on n (which contrasts with the competitive ratio of the asymptotic completion times). Also notice that $\Gamma(1/k) < k$ so that the communication cost of sources under lazy k-hop limited flooding is, as expected, greater than that under k-hop limited flooding. It is noteworthy that the competitive ratio decreases with k from $2/\sqrt{\pi}$ to 1, over $k \geq 2$ (see Figure 2) and $k/\Gamma(1/k) = 1 + \gamma/k + O(1/k^2)$ where γ is the Euler-Mascheroni constant (≈ 0.58).

Overall, the lazy k-hop limited flooding is $2/\sqrt{\pi}$-competitive to k-hop limited flooding and this is achieved for $k = 2$. The communication costs under lazy k-hop and k-hop limited flooding are very close to each other, as their ratio remains between 1 and 1.15 whatever the maximum number of hops k and the network size n are!

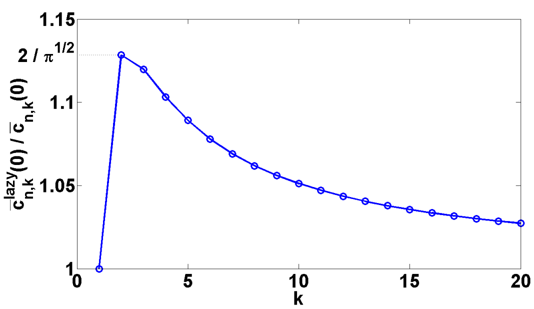

Figure 2. The competitive ratio of the communication costs for sources under lazy k-hop limited flooding and k-hop limited flooding.

VII. Numerical results

In this section we illustrate our analytical results and the behavior of flooding algorithms on asymptotically large networks by numerically solving the limiting ODE systems. We also compare our asymptotic results with empirical estimates obtained through simulation of underlying Markov processes; in particular, we show that our asymptotes for the completion time become accurate already for small values of n.

A. Asymptotic System Behavior

We illustrate the dynamics of flooding algorithms in large networks by numerically solving the ode systems identified in Section III; the example is for parameters set as follows $k = 3$, $n = 10,000$, $a_n = b_n = n$. In Figure 3 we show four distinct charts which we discuss in the following.

First, in Figure 3 (1) we show the fraction of informed nodes versus time. This indicates that the completion time under lazy k-hop limited flooding is indeed larger than under k-hop limited flooding; the lazy k-hop limited flooding progresses slower in the end phase in comparison with k-hop limited flooding. The slow down is for a factor of about $3/2$ which is consistent with Corollary 2.

Second, in Figure 3 (2) we show the rate at which nodes turn into those holding the message of age less than the hop count limit k; we call this rate *the seed arrival rate* as these nodes provide the message to other nodes. On the one hand, for k-hop limited flooding, the seed arrival rate is equal to the rate at which nodes that are either uninformed or informed through a path of length k (thus cannot forward the message) contact nodes that hold the message of age smaller or equal to $k-2$. Indeed, at time t, the seed arrival rate is $(1-q_k(t))q_{k-2}(t) + q_k(t)q_{k-2}(t) = q_{k-2}(t)$. On the other hand, for lazy k-hop limited flooding, the seed arrival rate is equal to the rate at which uninformed nodes are in contact with nodes that hold the message of age smaller or equal to $k-2$. Indeed, at time t, the seed arrival rate is $(1-q_k(t))q_{k-2}(t)$. In Figure 3 (2), we observe that under lazy k-hop limited flooding, the seed arrival rate peaks at a positive time like under flooding with no hop limits while it monotonically increases over time for k-hop limited flooding.

Third, in Figure 3 (3), we show the communication cost versus time as defined in Section VI. In particular, we note that the maximum communication cost (for sources) for the lazy k-hop limited flooding is larger for about 10% relative to k-hop limited flooding, which is consistent with Corollary 3.

Finally, in Figure 3 (4), we show the density of the communication costs per node over all nodes that become informed by the completion time. It is easy to establish that this density is $f_{n,k}(x) = q_{k-2}(s(x))/(1-q_k(s(x)))$ and $f_{n,k}(x) = q_{k-2}(s(x))$, where $s(x) = \overline{c}_{n,k}^{-1}(x)$, for k-hop limited flooding and lazy k-hop limited flooding, respectively. Figure 3 (4) illustrates that indeed there is a higher mass towards larger costs under lazy k-hop limited flooding. This is balanced with a smaller mass over small costs (the total number of message transmissions under both schemes is equal).

B. Convergence to the Asymptotics

In Figure 4 we compare empirical estimates of the completion time for k-hop limited flooding for particular values of the hop count limit $k = 3$ and $k = 4$. These empirical estimates are obtained through simulation of the underlying Markov process (Section III); we used 50 independent simulation runs and report the estimated means along with 95% confidence intervals. We observe that our analytical asymptotes are good approximations already for n as small as order 10.

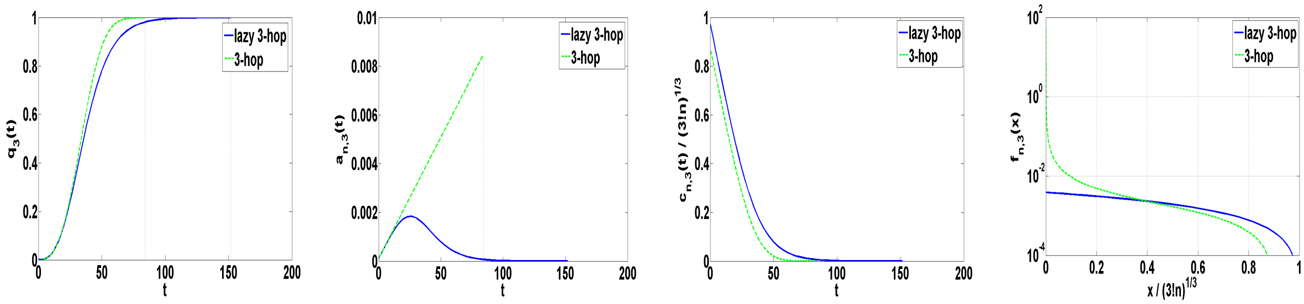

Figure 3. Flooding evolution for $k = 3$ and $n = 10,000$. From left to right: (1) fraction of informed nodes versus time, (2) arrival rate of nodes that hold message of age less than k versus time, (3) the communication cost versus time, and (4) density of per node communication costs.

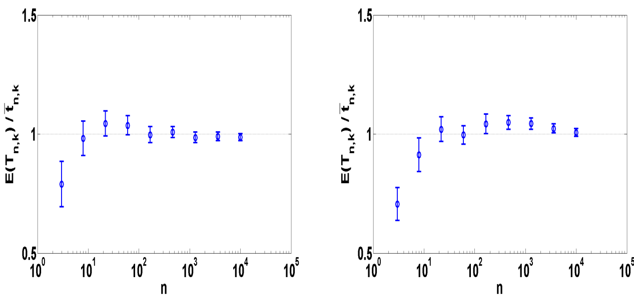

Figure 4. Empirical estimates versus analytical results for completion times: (left) 3-hop and (right) 4-hop limited flooding.

VIII. Conclusion

We provided characterizations of the completion time and the communication complexity for two natural algorithms for hop-limited message broadcast in dynamic networks. Our results reveal that an extremely simple (lazy) algorithm for hop-limited flooding provides near optimum performance that is only slightly worse than that of a more complex protocol. These results provide guidelines for designers of hop-limited systems that arise in various networking contexts, including peer-to-peer systems and mobile ad-hoc networks.

An interesting direction for future work is to pursue the same type of analysis for other classes of dynamic graphs.

References

[1] H. Baumann, P. Crescenzi, and P. Fraigniaud. Parsimonious flooding in dynamic graphs. In *Proc. of ACM PODC*, pages 260–269, 2009.

[2] A. Chaintreau, J.-Y. Le Boudec, and N. Ristanovic. The age of gossip: spatial mean field regime. In *Proc. of ACM Sigmetrics*, Seattle, WA, USA, 2009.

[3] A. Chaintreau, A. Mtibaa, L. Massoulié, and C. Diot. The diameter of opportunistic mobile networks. In *Proc. of ACM CoNEXT 2007*, New York, NY, USA, 2007.

[4] A. F. F. Clementi, A. Monti, F. Pasquale, and R. Silvestri. Broadcasting in dynamic radio networks. *Journal of Computer and System Sciences*, 75(4):213–230, 2008.

[5] A. F. F. Clementi, A. Monti, F. Pasquale, and R. Silvestri. Flooding Time in edge-Markovian Dynamic Graphs. In *Proc. of ACM PODC*, pages 213–222, 2008.

[6] S. Cui, A. M. Haimovich, O. Somekh, and H. V. Poor. Decentralized two-hop opportunistic relaying with limited channel state information. In *Proc. of IEEE ISIT*, Toronto, Canada, 2008.

[7] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *Proc. of ACM PODC*, pages 1–12, Vancouver, British Columbia, Canada, 1987.

[8] K. Fall. DTN Scope Control using Hop Limits (SCHL). Internet-Draft - http://tools.ietf.org/html/draft-fall-dtnrg-schl-00, 2000.

[9] A. El Fawal, J.-Y. Le Boudec, and K. Salamatian. Self-limiting epidemic forwarding. In *Proc. of IEEE Workshop on Autonomic Opportunistic Communications*, Helsinki, Finland, 2007.

[10] P. Gilbert, V. Ramasubramanian, P. Stuedi, and D. B. Terry. The duality between message routing and epidemic data replication. In *Proc. of Hotnets*, NYC, NY, USA, 2009.

[11] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of wireless networks. *IEEE Trans. on Networking*, 10(4):1063–6692, 2002.

[12] Z. Haas, J. Y. Halpern, and L. Li. Gossip-based ad hoc routing. In *Proc. of IEEE Infocom*, New York, NY, USA, 2002.

[13] J. Hromkovic, R. Klasing, B. Monien, and R. Peine. *Combinatorial Network Theory, D.-Z. Du and D. F. Hsu (Eds.)*, chapter Dissemination of Information in Interconnection Networks (Broadcasting & Gossiping), pages 125–222. Kluwer Academic Publishers, 1986.

[14] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proc. of ACM Mobihoc*, Hong Kong, China, 2008.

[15] R. M. Karp, C. Shindelhauer, S. Shenker, and B. Vocking. Randomized rumor spreading. In *Proc. of IEEE FOCS*, pages 565–574, 2000.

[16] F. Kuhn, N. Lynch, and R. Oshman. Distributed Computation in Dynamic Networks. In *Proc. of ACM STOC '10*, Boston, MA, USA, 2010.

[17] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[18] P. Neal. The generalised coupon collector problem. *J. Appl. Prob.*, 45:621–629, 2008.

[19] R. O'Dell and R. Wattenhofer. Information Dissemination in Highly Dynamic Graphs. In *Proc. of ACM DIALM-POMC, Cologne, Germany*, September 2005.

[20] B. Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[21] G. Sharma, R. R. Mazumdar, and N. B. Shroff. Delay and capacity trade-offs in mobile ad hoc networks: a global perspective. *IEEE/ACM Trans. on Networking*, 15, 2007.

[22] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis*. Chapman & Hall, 1995.

[23] C. Singh, A. Kumar, and R. Sundaresan. Delay and energy optimal two-hop relaying in delay tolerant networks. In *Proc. of WiOpt*, 2010.

[24] H.-Y. Wei and R. D. Gitlin. Two-hop-relay architecture for next-generation wwan/wlan integration. *IEEE Wireless Communications*, 11, 2004.

[25] S.-Y. Yao, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu. The broadcast storm problem in mobile ad hoc network. In *Proc. of ACM Mobicom*, Seattle, WA, USA, 1999.

IX. Appendix

A. Proof of Theorem 2

We first show the lower bound. Using the definition $r_i(t) = \int_0^t q_i(s)ds$ and (5), we have

$$r_i(t) \leq \frac{1}{a_n} \sum_{j=1}^{i+1} \frac{t^j}{j!}, \text{ for } t \geq 0 \text{ and } i = 1, \ldots, k. \quad (17)$$

Combining with (9) we have

$$\log(b_n) = r_{k-1}(\bar{t}_{n,k}) \leq \frac{1}{a_n} \sum_{j=1}^{k} \frac{\bar{t}_{n,k}^j}{j!} = \frac{\bar{t}_{n,k}^k}{k! a_n}(1 + o(1)).$$

Therefore,

$$\bar{t}_{n,k} \geq (k! a_n)^{1/k} \cdot \log^{1/k}(b_n) \cdot (1 - o(1))$$

which establishes the lower bound.

The upper bound can be established as follows. From (8), we have $r_0(t) = t/a_n$ and using the fact $e^{-x} \leq 1 - x + \frac{x^2}{2}$, for every $x \geq 0$, we have for $i = 1, \ldots, k$,

$$r_i(t) \geq \frac{t}{a_n} + \left(1 - \frac{1}{a_n}\right)\left(\int_0^t r_{i-1}(s)ds - \frac{1}{2}\int_0^t r_{i-1}(s)^2 ds\right).$$

Since $r_i(t)$ is non-decreasing with t for every i, we have for $i = 1, \ldots, k$,

$$r_i(t) \geq \frac{t}{a_n} + \left(1 - \frac{1}{a_n}\right)\left(\int_0^t r_{i-1}(s)ds - \frac{1}{2} r_{i-1}(t)^2 t\right).$$

By iterating the last recurrence from $i = 1$ to $i = k-1$ and repeatedly using the bound $\int_0^t s^j r_{i-1}(s)^2 ds \leq \frac{t^{j+1}}{j+1} r_{i-1}(t)^2$ for $j \geq 0$, we obtain

$$\begin{aligned} r_{k-1}(t) \quad &\geq \quad \left(1 - \frac{1}{a_n}\right)^{k-1} \cdot \frac{t^k}{k! a_n} \\ &+ \sum_{j=1}^{k-1} \left(1 - \frac{1}{a_n}\right)^{j-1} \frac{t^j}{j! a_n} - \epsilon_j(t) \end{aligned}$$

where we define

$$\epsilon_j(t) = \frac{1}{2}\left(1 - \frac{1}{a_n}\right)^{k-j} \frac{t^{k-j}}{(k-j)!} r_{j-1}(t)^2.$$

Therefore,

$$r_{k-1}(t) \geq \left(1 - \frac{1}{a_n}\right)^{k-1} \cdot \frac{t^k}{k! a_n} - \sum_{j=1}^{k-1} \epsilon_j(t). \quad (18)$$

Now, using (17), note

$$\begin{aligned} \epsilon_j(t) \quad &\leq \quad \frac{t^{k-j}}{(k-j)!} \cdot \left(\frac{1}{a_n} \sum_{i=1}^{j} \frac{t^i}{i!}\right)^2 \\ &= \quad \frac{t^k}{k! a_n} \cdot \binom{k}{j} \cdot \left(\sum_{i=0}^{j-1} \frac{j!}{(j-i)!} \frac{1}{t^i}\right) \cdot \frac{1}{a_n} \sum_{i=1}^{j} \frac{t^i}{i!}. \end{aligned}$$

Hence, for any increasing sequence t_n and every $j = 1, \ldots, k$,

$$\epsilon_j(t_n) \leq \frac{t_n^k}{k! a_n} \cdot O\left(\frac{t_n^{k-1}}{(k-1)! a_n}\right).$$

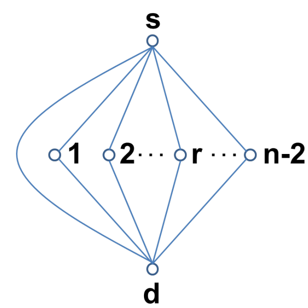

Figure 5. Two-hop limited flooding.

Therefore, from (18),

$$r_{k-1}(t_n) \geq \frac{t_n^k}{k! a_n} \cdot \left[1 - o(1) - O\left(\frac{t_n^{k-1}}{(k-1)! a_n}\right)\right].$$

For our sequence $\bar{t}_{n,k}$ we have $\frac{\bar{t}_n^{k-1}}{(k-1)! a_n} = o(1)$, and thus $r_{k-1}(\bar{t}_{n,k}) \geq \frac{\bar{t}_{n,k}^k}{k! a_n} \cdot (1 - o(1))$. Combining with (9) we obtain

$$\bar{t}_{n,k} \leq (k! a_n)^{1/k} \cdot \log^{1/k}(b_n) \cdot (1 + o(1))$$

which completes the proof.

B. Proof of Theorem 3

We consider a system of n nodes that are elements of the set V. Let $s \in V$ be the source and let us consider an arbitrary node in $V \setminus \{s\}$, say this node is d. Let $X_i(t)$ be equal to 1 if node i is informed at time t and is equal to 0, otherwise. We have

$$\begin{aligned} \mathbb{E}(1 - X_d(t)) \quad &= \quad \mathbb{P}(\text{node } d \text{ is uninformed at time } t) \\ &= \quad \mathbb{P}(A_{sd}(t) \cap \{\cap_{r \in V \setminus \{s,d\}} A_{s,r,d}(t)\}) \end{aligned}$$

where $A_{s,d}(t)$ is the event that there is no contact between source s and destination d in $[0, t]$ and $A_{s,r,d}(t)$ is the event that there exists no path from the source s through the node r to the destination d in time interval $[0, t]$. See Figure 5 for an illustration of edge-disjoint paths connecting the source and the destination node.

The events $A_{s,d}(t), A_{s,r,d}(t)$, for $r \in V \setminus \{s, d\}$, are independent and the probabilities of the events $A_{s,r,d}(t)$, for $r \in V \setminus \{s, d\}$, are identical, hence

$$\mathbb{E}(1 - X_d(t)) = \mathbb{P}(A_{s,d}(t))\mathbb{P}(A_{s,r,d}(t))^{n-2}$$

where r denotes an arbitrary node in $V \setminus \{s, d\}$.

Since each edge is activated at instances of a Poisson process with rate $1/(n-1)$ we have

$$\begin{aligned} \mathbb{P}(A_{s,d}(t)) \quad &= \quad \mathbb{P}(N(0,t] = 0) \\ \mathbb{P}(A_{s,r,d}(t)) \quad &= \quad \mathbb{P}(N(0,t] < 2) \end{aligned}$$

where $N(0, t]$ is a Poisson process of rate $1/(n-1)$.

Therefore,

$$\mathbb{E}(1 - X_d(t)) = \mathbb{P}(N(0,t] = 0) \cdot \mathbb{P}(N(0,t] < 2)^{n-2}. \quad (19)$$

We have established that

$$\begin{aligned}\mathbb{E}(1-X_d(t)) &= e^{-\frac{t}{n-1}}\left(e^{-\frac{t}{n-1}}+\frac{t}{n-1}e^{-\frac{t}{n-1}}\right)^{n-2}\\ &= \left[e^{-\frac{t}{n-2}}\left(1+\frac{t}{n-1}\right)\right]^{n-2}.\end{aligned}$$

Note that for large $t_{n,2}$ such that $t_{n,2}/n$ is small, we have

$$\begin{aligned}& e^{-\frac{t_{n,2}}{n-2}}\left(1-\frac{t_{n,2}}{n-1}\right)\\ =& \left(1-\frac{t_{n,2}}{n-2}+\frac{1}{2}\left(\frac{t_{n,2}}{n-2}\right)^2+O\left(\left(\frac{t_{n,2}}{n-2}\right)^3\right)\right)\\ & \cdot\left(1+\frac{t_{n,2}}{n-2}-\frac{t_{n,2}}{(n-1)(n-2)}\right)\\ =& 1-\frac{1}{2}\left(\frac{t_{n,2}}{n-2}\right)^2+O\left(\frac{t_{n,2}}{(n-2)^2}\right).\end{aligned}$$

Therefore,

$$\mathbb{E}(1-X_d(t_{n,2}))\le\exp\left(-\frac{t_{n,2}^2}{2(n-2)}[1-O(1/t_{n,2})]\right)$$

where the inequality is tight for large n. Therefore, we obtained

$$\mathbb{E}(1-X_d(t_{n,2}))\overset{n\to\infty}{\sim}e^{-\frac{t_{n,2}^2}{2n}}.$$

Hence, defining $t_{n,2}$ such that $\mathbb{E}(1-X_1(t_{n,2}))=(1-1/n)/b_n$, we have

$$t_{n,2}\overset{n\to\infty}{\sim}\sqrt{2n\log(b_n)}$$

which establishes the asserted result.

Remark The above analysis can be straightforwardly extended to more general case of $1\le m<n$ sources. For this more general case, similar analysis would be pursued as for Eq. (19), by considering instead

$$\begin{aligned}& \mathbb{E}(1-X_d(t))\\ =& e^{-\frac{m}{n-1}t}\left(\frac{m}{n-1}\int_0^t e^{-\frac{m}{n-1}s}\mathbb{P}(N(0,t-s]=0)ds\right.\\ & \left.+e^{-\frac{m}{n-1}t}\right)^{n-2}.\end{aligned}$$

C. Proof of Theorem 4

We consider the system under k-hop limited flooding that evolves according to (2). Let us define $u(t)$, for $t\ge 0$, by $u(0)=0$ and

$$\frac{d}{dt}u(t)=1-\tilde{q}_k(u(t)),\ \text{for } t\ge 0,$$

where $\tilde{q}_i(u(t)):=q_i(t)$, for $0\le i\le k$. $u(t)$ can be interpreted as the fraction of uploads made by a source node over the time interval $[0,t]$.

From (2), $\frac{d}{du}\tilde{q}_i(u)=\tilde{q}_{i-1}(u)$, for $i=1,2,\ldots,k$, whose solution for the initial value $\tilde{q}_i(0)=1/a_n$, for $1\le i\le k$, is given by

$$\tilde{q}_i(u)=\frac{1}{a_n}\sum_{j=0}^{i}\frac{u^j}{j!},\ \text{for } i=0,1,\ldots,k.$$

It follows that $u(t)$ satisfies $u(0)=0$ and

$$\frac{d}{dt}u(t)=1-\frac{1}{a_n}\sum_{j=0}^{k}\frac{u(t)^j}{j!},\ \text{for } t\ge 0. \tag{20}$$

Let $u_{n,k}$ be the value of $u(t)$ for time t equal to the completion time $\bar{t}_{n,k}$, i.e. $u_{n,k}$ is such that

$$1-\tilde{q}_k(u_{n,k})=\left(1-\frac{1}{a_n}\right)\frac{1}{b_n}.$$

It is of convenience to define $v_{n,k}=u_{n,k}/(k!a_n)^{1/k}$ and let $p_k(x)$ be the k-th order polynomial $p_k(x)=1-\tilde{q}_k((k!a_n)^{1/k}\cdot x)$, i.e.

$$p_k(x)=1-\frac{1}{a_n}\sum_{j=0}^{k}\frac{(k!a_n)^{j/k}}{j!}x^j.$$

We then have that $v_{n,k}$ is given by

$$p_k(v_{n,k})=\left(1-\frac{1}{a_n}\right)\frac{1}{b_n}. \tag{21}$$

From (20), the completion time $t_{n,k}$ is given by

$$\bar{t}_{n,k}=(k!a_n)^{1/k}\cdot\int_0^{v_{n,k}}\frac{1}{p_k(x)}dx. \tag{22}$$

In the remainder of the proof we use the last identity to estimate $\bar{t}_{n,k}$. We will show that the second factor in (22) is asymptotically a function of the gap between the point $\bar{v}_{n,k}$ at which the integrand goes to infinity and the point $v_{n,k}$.

The value $\bar{v}_{n,k}$ is a unique null-point of the polynomial $p_k(x)$ in the interval $[0,\infty)$ and is of multiplicity 1. Indeed, since $p_k(x)$ is decreasing over $[0,\infty)$, we have $p(v_{n,k})>0$ and $p_k(\bar{v}_{n,k})=0$, hence, $v_{n,k}<\bar{v}_{n,k}$. It is not difficult to observe that $v_{n,k}=1-o(1)$ and $\bar{v}_{n,k}=1-o(1)$ that we will use later in the proof.

We will next show that the following holds

$$\bar{t}_{n,k}=(k!a_n)^{1/k}\cdot\frac{1}{k}\left[\log\left(\frac{\bar{v}_{n,k}}{\bar{v}_{n,k}-v_{n,k}}\right)+C'_k\right]+O(1) \tag{23}$$

where

$$C'_k=\int_0^1\frac{\sum_{j=0}^{k-2}(k-1-j)x^j}{\sum_{j=0}^{k-1}x^j}dx.$$

Eq. (23) will follow from (22) once we establish that

$$\begin{aligned}& \int_0^{v_{n,k}}\frac{1}{p_k(x)}dx\\ =& \frac{1}{k}\cdot\left[\log\left(\frac{\bar{v}_{n,k}}{\bar{v}_{n,k}-v_{n,k}}\right)+C'_k\right]+O(1/(k!a_n)^{1/k}).\end{aligned}$$

In order to show the latter, we use the partial fraction decomposition to note that there exist polynomials $f_{k-1}(x)=\sum_{j=0}^{k-1}c_jx^j$ and $g_{k-2}(x)=\sum_{j=0}^{k-2}d_jx^j$ and A_n such that

$$p_k(x)=(\bar{v}_{n,k}-x)f_{k-1}(x)$$

and

$$\frac{1}{p_k(x)}=\frac{A_n}{\bar{v}_{n,k}-x}+\frac{g_{k-2}(x)}{f_{k-1}(x)}.$$

We next evaluate

$$\begin{aligned}&\int_0^{v_{n,k}} \frac{1}{p_k(x)}dx \\ =\ & \int_0^{v_{n,k}} \frac{A_n}{\bar{v}_{n,k}-x}dx + \int_0^{v_{n,k}} \frac{g_{k-2}(x)}{f_{k-1}(x)}dx \\ =\ & A_n \log\left(\frac{\bar{v}_{n,k}}{\bar{v}_{n,k}-v_{n,k}}\right) + \int_0^{v_{n,k}} \frac{g_{k-2}(x)}{f_{k-1}(x)}dx.\end{aligned}$$

By straightforward calculus, it can be obtained

$$c_j = \frac{1-\frac{1}{a_n}\sum_{i=0}^{j}\frac{(k!a_n)^{i/k}}{i!}\bar{v}_{n,k}^i}{\bar{v}_{n,k}^{j+1}},\ \text{for } 1 \le j \le k-1, \quad (24)$$

$$d_j = A_n \sum_{i=j+1}^{k-1} c_i \bar{v}_{n,k}^{i-j},\ \text{for } 0 \le j \le k-2$$

and

$$A_n = \frac{1}{\sum_{j=0}^{k-1} c_j \bar{v}_{n,k}^j}. \quad (25)$$

From (25) and (24),

$$A_n = \frac{\bar{v}_{n,k}}{k - \frac{1}{a_n}\sum_{j=0}^{k-1}\sum_{i=0}^{j}\frac{(k!a_n)^{i/k}}{i!}\bar{v}_{n,k}^i}.$$

Note that

$$\begin{aligned}&\frac{1}{a_n}\sum_{j=0}^{k-1}\sum_{i=0}^{j}\frac{(k!a_n)^{i/k}}{i!}\bar{v}_{n,k}^i \\ =\ & \frac{k}{(k!a_n)^{1/k}}\left(1+O(1/(k!a_n)^{1/k})\right).\end{aligned}$$

Hence

$$A_n = \frac{1}{k}\left(1+O(1/(k!a_n)^{1/k})\right).$$

Furthermore, observe that for every $0 \le x \le v_{n,k}$,

$$\frac{g_{k-2}(x)}{f_{k-1}(x)} = \frac{1}{k}\frac{\sum_{j=0}^{k-2}(k-1-j)x^j}{\sum_{j=0}^{k-1}x^j} + O(1/(k!a_n)^{1/k})$$

hence,

$$\int_0^{v_{n,k}} \frac{g_{k-2}(x)}{f_{k-1}(x)}dx = \frac{1}{k}\cdot C_k' + O(1/(k!a_n)^{1/k}).$$

So far, we showed that (23) holds. It remains only to estimate the gap $\bar{v}_{n,k} - v_{n,k}$.

Bounding the gap $\bar{v}_{n,k} - v_{n,k}$. In (23), we showed that the asymptotically dominant term of the completion time $t_{n,k}$ is a product of $(k!a_n)^{1/k}$ and a function of the gap $\bar{v}_{n,k} - v_{n,k}$ which we estimate in the following.

Lemma 2: The gap $\bar{v}_{n,k} - v_{n,k}$ satisfies

$$\bar{v}_{n,k} - v_{n,k} = \frac{1}{kb_n}(1+o(1)).$$

Proof: We first show the lower bound. Since $p_k(x)$ is a concave decreasing function on $[0,\infty)$ we have

$$\left(1-\frac{1}{a_n}\right)\frac{1}{b_n} = p_k(v_{n,k}) - p_k(\bar{v}_{n,k}) \le -p_k'(\bar{v}_{n,k})(\bar{v}_{n,k} - v_{n,k})$$

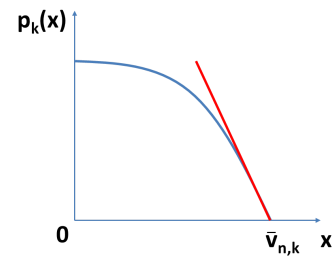

Figure 6. Bounding $p_k(x)$.

See Figure 6 for an illustration.

Combined with

$$\begin{aligned}-p_k'(\bar{v}_{n,k}) &= \frac{(k!a_n)^{1/k}}{a_n}\sum_{j=1}^{k-1}\frac{(k!a_n)^{j/k}}{j!}\bar{v}_{n,k}^k \\ &= k(1+o(1))\end{aligned}$$

we have

$$\bar{v}_{n,k} - v_{n,k} \ge \frac{1}{kb_n}(1-o(1)).$$

We next show the upper bound as follows

$$\begin{aligned}\left(1-\frac{1}{a_n}\right)\frac{1}{b_n} &= p_k(v_{n,k}) - p_k(\bar{v}_{n,k}) \\ &= \frac{1}{a_n}\sum_{j=1}^{k}\frac{(k!a_n)^{j/k}}{j!}[\bar{v}_{n,k}^j - v_{n,k}^j] \\ &\ge \frac{1}{a_n}\sum_{j=2}^{k}\frac{(k!a_n)^{j/k}}{j!}[jv_{n,k}^j\cdot(\bar{v}_{n,k}-v_{n,k})] \\ &\ge kv_{n,k}^k\cdot(\bar{v}_{n,k}-v_{n,k})\end{aligned}$$

where the first inequality is by convexity of x^j, for every $j = 1,\dots,k$. Since $v_{n,k} = 1-o(1)$, we have

$$\bar{v}_{n,k} - v_{n,k} \le \frac{1}{kb_n}(1+o(1)).$$

■

Finally, using the last lemma with (23) we establish

$$\bar{t}_{n,k} = (k!a_n)^{1/k}\left[\log(b_n^{1/k}) + \frac{\log(k)+C_k'}{k}\right] + O(1)$$

which completes the proof.

D. Comparison with Exact Solution for Lazy 2-hop

For the case $k=2$ one can exactly solve the integral in (22) which yields the following solution:

$$\begin{aligned}\bar{t}_{n,2} &= \sqrt{2a_n}\cdot\frac{1}{2\sqrt{1-\frac{1}{2a_n}}}\cdot \\ &\quad \cdot\log\left(\frac{1-\frac{1}{a_n}+v_{n,2}\left(\sqrt{1-\frac{1}{2a_n}}\right)}{1-\frac{1}{a_n}-v_{n,2}\left(\sqrt{1-\frac{1}{2a_n}}+\frac{1}{\sqrt{2a_n}}\right)}\right).\end{aligned}$$

Recall that $v_{n,k}$ is given by $p_2(v_{n,2}) = (1-1/a_n)/b_n$ which amounts to solving a quadratic equation whose positive solution yields

$$v_{n,2} = \sqrt{1 - \frac{1}{b_n} - \frac{1}{2a_n} + \frac{1}{a_n b_n}} - \frac{1}{\sqrt{2a_n}}.$$

It is easy to derive

$$\bar{t}_{n,2} \overset{n\to\infty}{\sim} \sqrt{2a_n}\log(\sqrt{4b_n})$$

which indeed conforms to our general result in Theorem 4. See Figure 7 for an example with $a_n = b_n = n$.

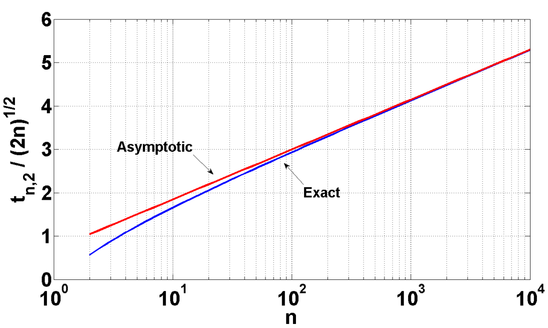

Figure 7. Asymptotic versus exact completion time for $k = 2$.

E. Proof of Theorem 5

We consider the system with $1 \le m < n$ initially informed nodes. Note that $m = n/a_n$ is a sequence of n but for simplicity of notation, in the following, we omit the subscript n and simply write m.

We consider the system evolution over phases $s = 0, 1, \ldots, n-m-1$ where phase s corresponds to the time interval over which there are $n-m-s$ uninformed nodes. The number of uninformed nodes is non-increasing over time and it decrements by 1 at a contact of a uninformed node and an informed node. Let $0 = T_0 < T_1 < \cdots < T_{n-m-1}$ denote instances such that $[T_s, T_{s+1})$ is the interval of phase s. A new phase is initiated if either (1) an uninformed node contacts a source node or (2) an uniformed node contacts a node that became informed by an earlier contact to a source node.

Let Y_s denote the number of nodes that became informed by a contact to a source node at some time $0 < t \le T_s$. It is not difficult to observe that Y_s is a discrete-time Markov chain satisfying the following: $Y_0 = 0$ and

$$Y_{s+1} = Y_s + X_s, \quad \text{for } s = 0, \ldots, n-m-1, \tag{26}$$

where X_s is a Bernoulli random variable, conditional on the value Y_s, with mean

$$\mathbb{P}(X_s = 1 | Y_s) = 1 - \mathbb{P}(X_s = 0 | Y_s) = \frac{m}{m + Y_s}.$$

Recall that a new phase begins at a contact of an informed node with either a source node or a node that became informed at an earlier contact with a source. Hence, the duration of phase s, denoted with $\tau_s = T_{s+1} - T_s$ is a random variable, which conditional on value Y_s, is a minimum of $(n-m-s)(m+Y_s)$ exponential random variables each with mean $n-1$.

Indeed, we have

$$T_{n,2} = \sum_{s=0}^{n-m-1} \tau_s$$

where given Y_s τ_s is an exponential random variable with mean $(n-1)/[(n-m-s)(m+Y_s)]$. In particular, the expected completion time is given by

$$\mathbb{E}(T_{n,2}) = (n-1)\sum_{s=0}^{n-m-1} \frac{1}{n-m-s}\mathbb{E}\left(\frac{1}{m+Y_s}\right). \tag{27}$$

Notice that for the expected completion time, it suffices to know the expected value of a function where the expectation is with respect to the distribution of Y_s, for $s = 1, 2, \ldots, n-m-1$. In the remainder of the proof, we estimate $\mathbb{E}(1/(m+Y_s))$. tion time, it suffices to know the expected value of a function where the expectation is with respect to the distribution of Y_s, for $s = 1, 2, \ldots, n-m-1$. In the remainder of the proof, we estimate $\mathbb{E}(1/(m+Y_s))$.

An auxiliary lemma. We note a number of properties about the random variable Y_s that are used later in the proof, in the following lemma.

Lemma 3: The random variable Y_s satisfies

$$m + \mathbb{E}(Y_s) \ge \sqrt{2ms + m^2 + 1}, \text{ for } s > 0 \tag{28}$$

$$m + \mathbb{E}(Y_s) \le \sqrt{2ms} + m + 1 + H_s \tag{29}$$

$$\mathrm{Var}(Y_s) \le \sqrt{2ms + m^2 + 1} - m - 1 \tag{30}$$

$$\mathbb{E}\left(\frac{1}{m+Y_s}\right) \le \frac{1}{m + \mathbb{E}(Y_s)} + \frac{1}{2ms}. \tag{31}$$

where $H_s = \sum_{i=1}^{s} \frac{1}{i}$.

Proof: We first show (28), then (30) and then 29. Eq. (31) follows as a by-product of (29).

Proof of inequality (28). From (26), $Y_0 = 0$, $Y_1 = 1$, and for $s = 1, \ldots, n-m-1$,

$$\mathbb{E}(Y_{s+1}) = \mathbb{E}(Y_s) + \mathbb{E}\left(\frac{m}{m+Y_s}\right).$$

Let $z_s := \mathbb{E}(Y_s) + m$. Note, by Jensen's inequality, $\mathbb{E}(m/(m+Y_s)) \ge m/(m+\mathbb{E}(Y_s)) = m/z_s$. Therefore,

$$z_{s+1} \ge z_s + \frac{m}{z_s}, \quad s = 0, 1 \ldots, n-m-1.$$

Notice that this is equivalent to $z_{s+1} \ge f(z_s)$ where $f(x) = x + m/x$ is an increasing function for $x > \sqrt{m}$. Since $z_0 = m$ and z_s is non-decreasing, we $z_s \ge \underline{z}_s$, for $s = 0, \ldots, n-m-1$, where

$$\underline{z}_{s+1} = \underline{z}_s + \frac{m}{\underline{z}_s}.$$

Now, let $\underline{z}(s)$ be given by $\underline{z}(1) = m+1$ and $d\underline{z}(s)/ds = m/\underline{z}(s)$, for $1 \le s \le t$. The latter differential system has the solution $\underline{z}(t) = \sqrt{2mt + m^2 + 1}$. It is easy to check that $z_s \ge \underline{z}_s \ge \underline{z}(s)$ for every $s = 1, 2, \ldots, n-m-1$, and hence, the result follows.

Proof of inequality (29). We first note the following identity, satisfied by the first and the second moment of Y_s.

Claim 1: For $s = 0, \ldots, n-m-1$,

$$\mathbb{E}(Y_s^2) = 2ms - (2m-1)\mathbb{E}(Y_s).$$

Proof: From (26), for $s = 0, \ldots, n-m-1$,

$$Y_{s+1}^2 = Y_s^2 + 2Y_sX_s + X_s^2 = Y_s^2 + 2Y_sX_s + X_s.$$

Taking expectations on both sides, we have

$$\begin{aligned} \mathbb{E}(Y_{s+1}^2) - \mathbb{E}(Y_s^2) &= 2m\mathbb{E}\left(\frac{Y_s}{m+Y_s}\right) + m\mathbb{E}\left(\frac{1}{m+Y_s}\right) \\ &= 2m - m(2m-1)\mathbb{E}\left(\frac{1}{m+Y_s}\right) \\ &= 2m - (2m-1)(\mathbb{E}(Y_{s+1}) - \mathbb{E}(Y_s)). \end{aligned}$$

Summing over $s = 1, \ldots, t-1$, we have

$$\begin{aligned} \mathbb{E}(Y_t^2) - 1 &= 2m(t-1) - (2m-1)(\mathbb{E}(Y_t) - 1) \\ &= 2mt - (2m-1)\mathbb{E}(Y_t) - 1 \end{aligned}$$

from which the claim follows. ■

By the claim,

$$\begin{aligned} \mathrm{Var}(Y_s) &= \mathbb{E}(Y_s^2) - \mathbb{E}(Y_s)^2 \\ &= 2ms - (2m-1)\mathbb{E}(Y_s) - \mathbb{E}(Y_s)^2. \end{aligned}$$

Combining with (29), we obtain (30).

Proof of inequality (29). By limited Taylor development of the function $1/(m+x)$ around $\mathbb{E}(Y_s)$, we obtain

$$\begin{aligned} &\mathbb{E}\left(\frac{1}{m+Y_s}\right) - \frac{1}{m+\mathbb{E}(Y_s)} \\ \leq\ & \frac{1}{(m+\mathbb{E}(Y_s))^3}\mathbb{E}((Y_s - \mathbb{E}(Y_s))^2). \qquad (32) \end{aligned}$$

From the latter fact combined with (28) and (30), we establish (31).

Now, from (26) and (31), we have

$$\mathbb{E}(Y_{s+1}) \leq \mathbb{E}(Y_s) + \frac{m}{m+\mathbb{E}(Y_s)} + \frac{1}{2s}.$$

Let $z_t = m + \mathbb{E}(Y_t)$ to rewrite the last inequality as

$$z_{s+1} \leq z_s + \frac{m}{z_s} + \frac{1}{2s}.$$

By (28), $z_s \geq \sqrt{2ms}$. Hence,

$$z_{s+1} - z_s \leq \sqrt{\frac{m}{2}}\frac{1}{\sqrt{s}} + \frac{1}{2s}.$$

Summing both sides over $s = 1$ to $s = t-1$, we have

$$\begin{aligned} z_t - z_1 &\leq \sqrt{\frac{m}{2}}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}} + \frac{1}{2}\sum_{s=1}^{t-1}\frac{1}{s} \\ &\leq \sqrt{\frac{m}{2}}\left(1 + \int_1^t \frac{ds}{\sqrt{s}}\right) + \frac{1}{2}H_t \\ &= -\sqrt{\frac{m}{2}} + \sqrt{2mt} + \frac{1}{2}H_t \end{aligned}$$

Since $z_1 = m+1$, the inequality (29) follows. ■

Upper bound. We next upper bound the expected completion time. From (27), we havetion time. From (27), we have

$$\mathbb{E}(T_{n,2}) = A_n + B_n$$

where

$$\begin{aligned} A_n &= (n-1)\sum_{s=0}^{n-m-1}\frac{1}{n-m-s}\frac{1}{m+\mathbb{E}(Y_s)} \\ B_n &= (n-1)\sum_{s=0}^{n-m-1}\frac{1}{n-m-s}\cdot \\ &\quad \cdot\left[\mathbb{E}\left(\frac{1}{m+Y_s}\right) - \frac{1}{m+\mathbb{E}(Y_s)}\right]. \end{aligned}$$

We proceed by upper bounding the terms A_n and B_n. Using (28), we can write

$$\begin{aligned} A_n &\leq \frac{n-1}{m(n-m)} + \frac{n-1}{\sqrt{2m}}\sum_{s=1}^{n-m-1}\frac{1}{n-m-s}\frac{1}{\sqrt{s}} \\ &= \frac{n-1}{m(n-m)} + \frac{n-1}{2\sqrt{2m(n-m)}}\cdot \\ &\quad \cdot\sum_{s=1}^{n-m-1}\frac{1}{(\sqrt{n-m}-\sqrt{s})\sqrt{s}} \\ &\quad + \frac{1}{(\sqrt{n-m}+\sqrt{s})\sqrt{s}}. \end{aligned}$$

By elementary analysis, we have

$$\begin{aligned} &\sum_{s=1}^{n-m-1}\frac{1}{(\sqrt{n-m}-\sqrt{s})\sqrt{s}} \\ =\ & \frac{1}{\sqrt{n-m}}\sum_{s=1}^{n-m-1}\frac{1}{\sqrt{s}} + \frac{1}{\sqrt{n-m}-\sqrt{s}} \\ \leq\ & \frac{1}{\sqrt{n-m}}\left(1 + \int_1^{n-m-1}\frac{ds}{\sqrt{s}}\right. \\ & + \frac{1}{\sqrt{n-m}-\sqrt{n-m-1}} \\ & \left. + \int_1^{n-m-1}\frac{ds}{\sqrt{n-m}-\sqrt{s}}\right) \\ =\ & \frac{1}{\sqrt{n-m}}\left(1 + \frac{1}{\sqrt{n-m}-\sqrt{n-m-1}}\right) \\ & + 2\log\frac{\sqrt{n-m}-1}{\sqrt{n-m}-\sqrt{n-m-1}} \end{aligned}$$

and

$$\begin{aligned} &\sum_{s=1}^{n-m-1}\frac{1}{(\sqrt{n-m}+\sqrt{s})\sqrt{s}} \\ \leq\ & \frac{1}{\sqrt{n-m}+1} + \int_1^{n-m-1}\frac{ds}{(\sqrt{n-m}+\sqrt{s})\sqrt{s}} \\ =\ & \frac{1}{\sqrt{n-m}+1} + 2\log\frac{\sqrt{n-m}+\sqrt{n-m-1}}{\sqrt{n-m}+1}. \end{aligned}$$

Putting the pieces together, we obtain

$$\begin{aligned} A_n \leq\ & \frac{n-1}{m(n-m)} \\ & + \frac{n-1}{\sqrt{2m(n-m)}} \log \frac{\left(1+\sqrt{1+\frac{1}{n-m}}\right)\left(1-\frac{1}{\sqrt{n-m}}\right)}{\left(1-\sqrt{1-\frac{1}{n-m}}\right)\left(1+\frac{1}{\sqrt{n-m}}\right)} \\ & + \frac{n-1}{2\sqrt{2m(n-m)}}\left[\frac{1}{\sqrt{n-m}}\cdot \right. \\ & \left. \cdot\left(1+\frac{1}{\sqrt{n-m}-\sqrt{n-m-1}}\right)+\frac{1}{\sqrt{n-m}+1}\right]. \end{aligned}$$

From the last inequality, it is not difficult to observe that

$$A_n \leq \sqrt{2a_n}\log(\sqrt{n}) + \Theta(\sqrt{a_n}). \quad (33)$$

We next upper bound the term B_n. Using (31), we have

$$\begin{aligned} B_n \leq\ & \frac{n-1}{m(n-m)} + \frac{n-1}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)s} \\ =\ & \frac{n-1}{m(n-m)} + \frac{n-1}{m(n-m)}H_{n-m-1} \end{aligned}$$

where $H_i = \sum_{j=1}^{i} 1/j$. From the last inequality, we observe

$$B_n \leq \frac{a_n \log(n)}{n} + O(1). \quad (34)$$

From (33) and (34), we have

$$\mathbb{E}(T_{n,2}) \leq \sqrt{2a_n}\log(\sqrt{n}) + \Theta(\sqrt{a_n}) \quad (35)$$

which establishes the upper bound.

Lower bound. From (27) and Jensen's inequality

$$\mathbb{E}(T_{n,2}) \geq (n-1)\sum_{s=0}^{n-m-1}\frac{1}{n-m-s}\frac{1}{z_s}$$

where $z_s = m + \mathbb{E}(Y_s)$.

Note

$$\begin{aligned} & (n-1)\sum_{s=0}^{n-m-1}\frac{1}{n-m-s}\frac{1}{z_s} \\ \geq\ & \frac{n-1}{\sqrt{2m}}\sum_{s=1}^{n-m-1}\frac{1}{n-m-s}\frac{1}{\sqrt{s}}\frac{1}{1+\frac{m+1+\frac{1}{2}H_s}{\sqrt{2ms}}} \\ \geq\ & \frac{n-1}{\sqrt{2m}}\sum_{s=1}^{n-m-1}\frac{1}{n-m-s}\frac{1}{\sqrt{s}}\left(1-\frac{m+1+\frac{1}{2}H_s}{\sqrt{2ms}}\right) \end{aligned}$$

where the fist inequality follows from (29), the second inequality follows by the fact $1/(1+x) \geq 1-x$, for $x \geq 0$.

In view of the last above inequality, we have

$$\mathbb{E}(T_{n,2}) \geq A_n - B_n$$

where

$$\begin{aligned} A_n =\ & \frac{n-1}{\sqrt{2m}}\sum_{s=1}^{n-m-1}\frac{1}{n-m-s}\frac{1}{\sqrt{s}} \\ B_n =\ & \frac{n-1}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)s}(m+1+\frac{1}{2}H_s). \end{aligned}$$

By similar arguments as for the upper bound, it can be showed that

$$A_n \geq \frac{n-1}{\sqrt{2m(n-m)}}\log\frac{\left(1-\frac{1}{\sqrt{n-m}}\right)\left(1+\sqrt{1-\frac{1}{n-m}}\right)}{\left(1+\frac{1}{\sqrt{n-m}}\right)\left(1-\sqrt{1-\frac{1}{n-m}}\right)}.$$

Using the facts

$$\frac{1-\frac{1}{\sqrt{n-m}}}{1+\frac{1}{\sqrt{n-m}}} \geq \left(1-\frac{1}{\sqrt{n-m}}\right)^2$$

and

$$\frac{1+\sqrt{1-\frac{1}{n-m}}}{1-\sqrt{1-\frac{1}{n-m}}} \geq \frac{1}{1-\sqrt{1-\frac{1}{n-m}}} \geq n-m$$

we obtain

$$A_n \geq \sqrt{\frac{2}{m}}\frac{n-1}{\sqrt{n-m}}\left(\log\sqrt{n-m}+\log\left(1-\frac{1}{\sqrt{n-m}}\right)\right).$$

It is now easy to observe

$$A_n \geq \sqrt{2a_n}\log\sqrt{n} + O(\sqrt{a_n/n}).$$

Furthermore,

$$\begin{aligned} B_n \leq\ & \frac{n-1}{2m}\left(m+1+\frac{1}{2}H_{n-m-1}\right)\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)s} \\ =\ & \frac{n-1}{m(n-m)}\left(m+1+\frac{1}{2}H_{n-m-1}\right)H_{n-m-1} \end{aligned}$$

Hence, it can be observed that

$$B_n \leq O(\frac{a_n}{n}\log(n)^2).$$

The asserted lower bound follows.

F. Proof of Theorem 6

Since conditional on $Y_0, Y_2, \ldots, Y_{n-m-1}$, the completion time $T_{n,k}$ is a sum of independent random variables whose respective distributions are exponential with means $(n-1)/[(n+m-s)(m+Y_s)]$, $s = 0, \ldots, n-m-1$, we have

$$\mathrm{Var}(T_{n,2}) = (n-1)^2\sum_{s=0}^{n-m-1}\frac{1}{(n-m-s)^2}\mathbb{E}\left(\frac{1}{(m+Y_s)^2}\right).$$

Let us define A_n and B_n such that

$$\begin{aligned} A_n =\ & (n-1)^2\sum_{s=0}^{n-m-1}\frac{1}{(n-m-s)^2}\frac{1}{(m+\mathbb{E}(Y_s))^2} \\ B_n =\ & (n-1)^2\sum_{s=0}^{n-m-1}\frac{1}{(n-m-s)^2}\left[\mathbb{E}\left(\frac{1}{(m+Y_s)^2}\right)\right. \\ & \left. -\frac{1}{(m+\mathbb{E}(Y_s))^2}\right] \end{aligned}$$

and then

$$\mathrm{Var}(T_{n,2}) = A_n + B_n. \quad (36)$$

Upper bound. We first upper bound the term A_n. By the facts $\mathbb{P}(Y_0 = 0) = 1$ and $m + \mathbb{E}(Y_s) \geq \sqrt{2ms}$, where the latter is from (28), we have

$$A_n \leq \left(\frac{n-1}{m(n-m)}\right)^2 + \frac{(n-1)^2}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s}.$$

By elementary calculus,

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

Hence,

$$\begin{aligned}&\tfrac{(n-1)^2}{2m}\textstyle\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s}\\ &= \tfrac{1}{2m}\left(\tfrac{n-1}{n-m}\right)^2\left(2H_{n-m-1} + (n-m)\textstyle\sum_{s=1}^{n-m-1}\frac{1}{s^2}\right)\\ &= \tfrac{\pi^2}{12}a_n + \Theta\left(\tfrac{a_n\log(n)}{n}\right)\end{aligned} \tag{37}$$

and, therefore,

$$\lim_{n\to\infty}\frac{A_n}{a_n} \leq \lim_{n\to\infty}\sum_{s=1}^{n-m-1}\frac{1}{s^2} = \frac{\pi^2}{12}.$$

We next consider the term B_n. By limited Taylor development of the function $1/(m+x)^2$ around $\mathbb{E}(Y_s)$ we have

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{(m+Y_s)^2}\right) - \frac{1}{(m+\mathbb{E}(Y_s))^2}\\ \leq\ &\frac{3}{(m+\mathbb{E}(Y_s))^4}\mathbb{E}((Y_s-\mathbb{E}(Y_s))^2)\\ \leq\ &\frac{3}{(2m)^{3/2}}\frac{1}{s^{3/2}}\end{aligned}$$

where the last inequality follows from (30) and (28). It follows

$$B_n \leq \frac{3(n-1)^2}{(2m)^{3/2}}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s^{3/2}}. \tag{38}$$

Noting that

$$\begin{aligned}\frac{1}{(n-m-s)^2 s^{3/2}} &= \frac{1}{(n-m)^2}\frac{1}{s^{3/2}}\\ &+\frac{1}{(n-m)^2}\frac{1}{(n-m-s)\sqrt{s}}\\ &+\frac{1}{(n-m)}\frac{1}{(n-m-s)^2\sqrt{s}}\end{aligned}$$

we have

$$\begin{aligned}B_n \leq\ &\frac{3}{(2m)^{3/2}}\left\{\left(\frac{n-1}{n-m}\right)^2\cdot\right.\\ &\cdot\left[\sum_{s=1}^{n-m-1}\frac{1}{s^{3/2}} + \sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)\sqrt{s}}\right]\\ &\left.+\frac{(n-1)^2}{n-m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2\sqrt{s}}\right\}.\end{aligned}$$

It is not difficult to check that

$$\begin{aligned}&\sum_{s=1}^{n-m-1}\frac{1}{s^{3/2}} = O(1)\\ &\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)\sqrt{s}}\\ =\ &O(\sqrt{(n-m)}\log(n-m))\\ &\frac{(n-1)^2}{n-m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2\sqrt{s}}\\ =\ &O(\sqrt{(n-m)}\log(n-m)).\end{aligned}$$

It thus follows

$$B_n \leq O\left(\frac{a_n\sqrt{a_n}\log(n)}{n}\right). \tag{39}$$

We have established that

$$\mathrm{Var}(T_{n,2}) \leq \frac{\pi^2}{12}a_n + o(a_n).$$

Lower bound. From (36) and Jensen's inequality, we have $\mathrm{Var}(T_{n,2}) \geq A_n$, and thus it suffices to lower bound the term A_n. Using (29),

$$\begin{aligned}A_n \geq\ &\frac{(n-1)^2}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s}\cdot\\ &\cdot\frac{1}{\left(1+(m+1+H_s)/\sqrt{2ms}\right)^2}\\ \geq\ &\frac{(n-1)^2}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s}\left(1-2\frac{m+1+H_s}{\sqrt{2ms}}\right)\\ \geq\ &\frac{(n-1)^2}{2m}\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s}\\ &-\frac{(n-1)^2}{\sqrt{2}m^{3/2}}(m+1+H_{n-m-1})\cdot\\ &\cdot\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s^{3/2}}\\ =\ &\frac{\pi^2}{12}a_n + \Theta\left(\frac{a_n\log(n)}{n}\right)\\ &-\frac{(n-1)^2}{\sqrt{2}m^{3/2}}(m+1+H_{n-m-1})\cdot\\ &\cdot\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s^{3/2}}\end{aligned}$$

where the equality is by (37). By similar arguments as in obtaining (39) from (38), we have

$$\begin{aligned}&\frac{(n-1)^2}{\sqrt{2}m^{3/2}}(m+1+H_{n-m-1})\sum_{s=1}^{n-m-1}\frac{1}{(n-m-s)^2 s^{3/2}}\\ =\ &a_n\frac{\log(n)}{\sqrt{a_n}}(1+o(1)).\end{aligned}$$

It follows that $\mathrm{Var}(T_{n,2})/a_n \geq \frac{\pi^2}{12} - \frac{\log(n)}{\sqrt{a_n}}(1+o(1))$, which completes the proof.

G. Proof of Theorem 7

We use here the notations introduced in the proof of Theorem 2. We have:

$$\bar{c}_{n,k}(s_n) = \int_{s_n}^{t_{n,k}} (1 - q_k(u))du = \left(1 - \frac{1}{a_n}\right) \int_{s_n}^{t_{n,k}} \exp(-r_{k-1}(u))du.$$

We deduce from the inequalities on $r_{k-1}(t)$ derived in the proof of Theorem 2 that, if $\log(b_n) = o(a_n)$,

$$\bar{c}_{n,k}(s_n) = \left(1 - \frac{1}{a_n}\right)(k!a_n)^{1/k} \int_{s_n/(k!a_n)^{1/k}}^{t_{n,k}/(k!a_n)^{1/k}} \exp(-u^k)du \cdot (1 + o(1)). \tag{40}$$

(i) Assume that $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \alpha < \infty$. From (40), we deduce that

$$\bar{c}_{n,k}(s_n) = (k!a_n)^{1/k} \int_{\alpha}^{\infty} \exp(-u^k)du \cdot (1 + o(1)).$$

Then (12) follows from the fact:

$$\int_{\alpha}^{\infty} \exp(-u^k)du = \frac{1}{k} \int_{\alpha^k}^{\infty} v^{\frac{1}{k}-1} e^{-v} dv = \frac{1}{k} \Gamma\left(\frac{1}{k}, \alpha^k\right).$$

Now in view of the assumptions made in (ii), (13) follows from (40) and the fact that for x large:

$$\int_{x}^{\infty} \exp(-u^k)du = \frac{e^{-x^k}}{kx^{k-1}}(1 + o(1)).$$

H. Proof of Theorem 8

We use here the notation introduced in the proof of Theorem 4. In addition, we introduce $v(t) = u(t)/(k!a_n)^{1/k}$ and $v_n = v(s_n)$. Let us first prove (14). We have:

$$\frac{s_n}{(k!a_n)^{1/k}} = \int_0^{v_n} \frac{1}{p_k(x)}dx \geq \int_0^{v_n} \frac{1}{1 - x^k}dx = \psi_k(v_n).$$

Hence

$$v_n \leq \psi_k^{-1}\left(\frac{s_n}{(k!a_n)^{1/k}}\right).$$

Now

$$\bar{c}_{n,k}(s_n) = (k!a_n)^{1/k}(v_{n,k} - v_n) \geq (k!a_n)^{1/k}\left(v_{n,k} - \psi_k^{-1}\left(\frac{s_n}{(k!a_n)^{1/k}}\right)\right).$$

Moreover, we have

$$\frac{s_n}{(k!a_n)^{1/k}} \leq \int_0^{v_n} \frac{1}{1 - d_n - x^k}dx = \frac{1}{(1 - d_n)^{1-1/k}} \int_0^{v_n/(1-d_n)^{1/k}} \frac{dx}{1 - x^k} \leq \frac{1}{(1 - d_n)^{1-1/k}}\psi_k(v_n/(1 - d_n)^{1/k}),$$

where

$$d_n = \frac{1}{a_n} \sum_{j=0}^{k-1} \frac{(k!a_n)^{j/k}}{j!}.$$

Hence

$$v_n \geq (1 - d_n)^{1/k}\psi_k^{-1}\left(\frac{s_n(1 - d_n)^{1-1/k}}{(k!a_n)^{1/k}}\right),$$

and

$$\bar{c}_{n,k}(s_n) \geq (k!a_n)^{1/k}\left(v_{n,k} - (1 - d_n)^{1/k}\psi_k^{-1}\left(\frac{s_n(1 - d_n)^{1-1/k}}{(k!a_n)^{1/k}}\right)\right).$$

Now since $v_{n,k} \to 1$ and $d_n \to 0$ as $n \to \infty$, if $\lim_{n\to\infty} s_n/(k!a_n)^{1/k} = \alpha$,

$$\bar{c}_{n,k}(s_n) = (k!a_n)^{1/k}(1 - \psi_k^{-1}(\alpha) + o(1)).$$

We now turn our attention to (15). As in the proof of Theorem 4, we can show that

$$s_n = (k!a_n)^{1/k} \cdot \frac{1}{k}\left[\log\left(\frac{\bar{v}_{n,k}}{\bar{v}_{n,k} - v_n}\right) + C'_k\right] + O(1).$$

We deduce that

$$v_n = \bar{v}_{n,k}\left(1 - \exp\left(-\frac{ks_n}{(k!a_n)^{1/k}} + O(1)\right)\right).$$

Now

$$\bar{c}_{n,k}(s_n) = (k!a_n)^{1/k}\left[v_{n,k} - \bar{v}_{n,k} + \bar{v}_{n,k}\exp\left(-\frac{ks_n}{(k!a_n)^{1/k}} + O(1)\right)\right].$$

Remark that as shown previously, $\bar{v}_{n,k} - v_{n,k} = \frac{1}{kb_n}(1 + o(1))$. Also remark that since $\lim_{n\to\infty} s_n/t_{n,k} = 0$,

$$\frac{1}{b_n} = o\left(\exp\left(-\frac{ks_n}{(k!a_n)^{1/k}}\right)\right).$$

We then conclude that

$$\bar{c}_{n,k}(s_n) = (k!a_n)^{1/k} \cdot \exp\left(-\frac{ks_n}{(k!a_n)^{1/k}} + O(1)\right).$$