

Research Article

Gossiping with Message Splitting on Structured Networks

Wenxiang Dong, Ying Yang, and Wenyi Zhang

Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China

Correspondence should be addressed to Wenyi Zhang; wenyizha@ustc.edu.cn

Received 24 February 2014; Revised 20 November 2014; Accepted 25 November 2014

Academic Editor: Javier Bajo

Copyright © 2015 Wenxiang Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gossiping of a single source with multiple messages (by splitting information into pieces) has been treated only for complete graphs, shown to considerably reduce the completion time, that is, the first time at which all network nodes are informed, compared with single-message gossiping. In this paper, gossiping of a single source with multiple messages is treated, for networks modeled as certain structured graphs, wherein upper bounds of the high-probability completion time are established through a novel “dependency graph” technique. The results shed useful insights into the behavior of multiple-message gossiping and can be useful for data dissemination in sensor networks, multihopping content distribution, and file downloading in peer-to-peer networks.

1. Introduction

Data dissemination is a fundamental issue for information diffusion in networks [1–5]. Gossiping (a.k.a. rumor spreading) is an effective way of data dissemination: imagine the situation where a rumor arises in town and people undertake gossiping-based actions to spread the rumor among the population [6]. There are two atomic types of gossiping protocols: “pull”: an uninformed node selects a message it has not possessed and requests it from a randomly selected neighboring node, and “push”: an informed node selects a message it possesses and sends it to a randomly selected neighboring node.

Gossip-based algorithms are promising solution for many of the next generation networks, due to their simplicity, robustness, flexibility, and scalability. Existing applications are numerous, such as consensus and averaging in sensor networks [7, 8], ad hoc message routing [9], peer-to-peer file distribution [10], and information diffusion in social networks [11, 12]. Extended versions of the basic gossiping protocol include algebraic gossip [13] and geographic gossip [14], but they have additional overheads on message complexity and geographic knowledge, and in this paper our attention is focused on the basic gossiping protocol.

Most analytical works on gossiping have dealt with high-probability bounds of the completion time. Supposing a static and connected network, the completion time of a gossiping

protocol is the first time at which all nodes are informed. Several important classes of network topologies have been analyzed for gossiping of a single source with a single message, including complete graphs [6], general graphs and hypercubes [15], random graphs [15, 16], random geometric graphs [17], graphs with edge expansion [18], and graphs with vertex expansion [19]. Besides, gossiping of all sources each with a different single message has been considered in [10, 20].

On the other hand, the gossiping type of a single source with multiple messages has been treated in [10], for complete graphs. In practice, the multiple messages may be obtained by splitting a file into multiple units. It has been shown in [10] that, compared with single-message gossiping, gossiping with multiple messages significantly reduces the completion time. Complete graphs lack geometry and enable direct message passing between any two nodes. In this paper, we are hence motivated to study gossiping with multiple messages on more structured graphs. Networks with various types of structure are more general and realistic than complete graphs, and the studies on structured graphs bring in new challenges and insights into the gossiping problem. The key question we seek to answer is how much benefit message splitting brings in when the message passing is restricted by the network topology.

In this paper, we focus on a “sequential pull” protocol in which each node attempts to pull its desired messages sequentially following an indexing order. Our contributions are as follows.

- (1) We develop a novel “dependency graph” analysis technique for gossip spreading and leverage it to establish a high-probability upper bound of the completion time for line graphs. In a nutshell, our result indicates that, in a line graph of n nodes and with k messages, the high-probability completion time scales at most like $\mathcal{O}((n + k) \log(n + k))$ as both n and k grow large, inferior to that for complete graphs [10], which is $\mathcal{O}(k + \log n)$, but still drastically superior to that without message splitting [17], which is $\mathcal{O}(kn)$.
- (2) We apply the result for line graphs to several other network topologies, including ring graphs, general graphs with given diameter and maximum degree, grid graphs, and random geometric graphs.
- (3) We carry out numerical experiments to corroborate our analytical results.

The remaining part of this paper is organized as follows. In Section 2 we describe the sequential pull gossiping protocol. In Section 3 we establish the high-probability upper bound of the completion time on line graphs. In Section 4 we treat other classes of network topologies. In Section 5 we present numerical results. Finally Section 6 concludes this paper.

2. Pull-Based Gossiping Protocol

In general, a connected static network is represented as an undirected graph $G = (V, E)$ consisting of a set of nodes V and a set of edges E . Every node in V wishes to obtain an entire copy of a file, which has been split into multiple units, each of which as a message for gossiping. A pair of nodes may communicate directly with each other if and only if the nodes are connected by an edge in E . Time is slotted and a message can be transferred from a sender to a receiver within a time slot, which is called *round* throughout this paper. For each node, if a message is obtained in round t , it may send that message to one of its neighbors since round $t + 1$; when obtaining all messages, a node is said to be *informed*, otherwise *uninformed*.

Similar to that in [20], we describe the protocol of pull-based gossiping with multiple messages by a random walk following a Markov chain. Initially, a single source node in G is endowed with k messages, indexed as $1, 2, \dots, k$. In each of the following rounds, every uninformed node u selects a message it has not possessed and contacts a neighbor v in its neighbor set $N(u)$ with probability q_{uv} or contacts no neighbor with probability $q_{uu} = 1 - \sum_{v \in N(u)} q_{uv}$. If v is contacted by u and does possess the requested message, then u obtains that message successfully. If the messages are requested sequentially in order by every node, that is, each uninformed node always pulls the message with the smallest index it has not possessed, the gossiping protocol is called “*sequential pull*” with multiple messages [10]. Note that “sequential pull”

may be an appropriate restriction in certain applications (e.g., with streaming/real-time nature), in which messages need to be sequentially accumulated for content reconstruction.

In particular, for a line graph $G = (V, E)$, in which n nodes indexed as $1, 2, \dots, n$ are sequentially placed from one endpoint to the other, the gossiping protocol can also be described as follows. Without loss of generality, assume that the endpoint node 1 is the source node. In each round, each uninformed node u pulls the message with the smallest index it has not possessed from its neighbor $u - 1$ with probability q_u . Throughout this paper, we focus on the case where $q_u \equiv q, 2 \leq u \leq n$. Note that “gossiping with multiple messages on line graphs” can be leveraged to model the content distribution process of a multihopping data dissemination problem for sensor networks or for ad hoc networks; for example, see Section 5.

In the subsequent analysis, the main attention is paid to the *completion time* T , which is the first time when all nodes in G are informed. Our goal is to establish a *high-probability* upper bound of T , that is, a quantity such that T does not exceed it with probability at least $1 - n^{-c}$, for some constant $c > 0$, for any sufficiently large n .

3. Upper Bound of Completion Time on Line Graphs

In this section, we derive a high-probability upper bound of the completion time on line graphs, by a novel “dependency graph” technique. To the best of our knowledge, this technique has not been identified or used for analyzing gossiping in the literature. The result for line graphs, in company with complete graphs [10], reveals the potential benefit for message splitting to accelerate content distribution. Complete graphs and line graphs may be viewed as two extreme cases of graph topology, in that complete graphs have “no geometry” without message-passing constraints and line graphs have the “strongest geometry” with the most stringent message-passing constraints. Importantly note that the message dissemination process through gossiping protocols itself does not depend on the network topology; however, our analysis of the completion time is focused on a specific class of graphs and thus has the knowledge on the network topology.

3.1. Gossiping with Single Message. Before treating gossiping with multiple messages, we consider the single-message case, which provides a baseline for the multiple-message case to compare with.

Lemma 1. *Consider a line graph of n nodes and a single message, and let T denote the first time when all nodes are informed. For any $c > 0$, choose $\varepsilon = c \log n \cdot (1 + \sqrt{1 + 2(n - 1)/(c \log n)})/(n - 1)$; then one has, for all n ,*

$$\Pr \left\{ T < \frac{n - 1}{q} (1 + \varepsilon) \right\} > 1 - n^{-c}. \quad (1)$$

Remark 1. We see that $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$, and hence the high-probability upper bound of the completion time T is $\mathcal{O}(n/q)$ for a file with unit size. So, given a file with size of

k units to be distributed on the network without message splitting, the completion time would be $\mathcal{O}(kn/q)$.

In order to prove Lemma 1, from [17, Lemma 4.3] (for a line graph, the completion time of gossiping with a single message is of the same magnitude for both “pull” and “push” based protocols), we can set $\exp(-(n-1)\varepsilon^2/(2(1+\varepsilon))) \leq n^{-c}$, which is able to be done by choosing $\varepsilon = c \log n \cdot (1 + \sqrt{1 + 2(n-1)/(c \log n)})/(n-1)$. In the following, we present a detailed proof of Lemma 1, since [17, Lemma 4.3] is only stated for the “push-” based single-message case and its proof is absent.

Proof of Lemma 1. For $2 \leq i \leq n$, let X_i denote the time that node i needs to successfully pull the message from node $i-1$ after node $i-1$ succeeds; then $T = \sum_{i=2}^n X_i$. Due to the memoryless nature of the gossiping process, $\{X_i, 2 \leq i \leq n\}$ are independent and identically distributed (i.i.d.) random variables, and all of them obey the geometric distribution with parameter q .

Let $\{Z_i, 1 \leq i \leq m\}$ be m independent Poisson trials such that $\Pr[Z_i = 1] = q$ and $\Pr[Z_i = 0] = 1 - q$, for all $1 \leq i \leq m$. Then, we have

$$\Pr[T \geq m] = \Pr\left[\sum_{i=1}^m Z_i \leq (n-1)\right], \quad (2)$$

since the event of $\{T \geq m\}$ means that at most $n-1$ Poisson trials have succeeded (i.e., with $Z_i = 1$) within m trials.

Let $\varepsilon \geq 0$ and $u = E[\sum_{i=1}^{(n-1)(1+\varepsilon)/q} Z_i]$ (the value of $(n-1)(1+\varepsilon)/q$ might not be an integer, but it does not affect the magnitude of the asymptotic upper bound of gossiping's completion time), from (2) and Chernoff Bound [21, Theorem 4.5], we have

$$\begin{aligned} \Pr\left[T \geq \frac{n-1}{q} (1+\varepsilon)\right] \\ = \Pr\left[\sum_{i=1}^{(n-1)(1+\varepsilon)/q} Z_i \leq (n-1)\right] &\leq \exp\left(-\frac{(n-1)\varepsilon^2}{2(1+\varepsilon)}\right). \end{aligned} \quad (3)$$

Take a step ahead and set $\exp(-(n-1)\varepsilon^2/(2(1+\varepsilon))) \leq n^{-c}$; then the minimum of ε is able to be obtained by $c \log n \cdot (1 + \sqrt{1 + 2(n-1)/(c \log n)})/(n-1)$. Therefore, the proof is completed. \square

3.2. Gossiping with Multiple Messages. If a file with size of k units is split into k messages and the sequential pull-based gossiping protocol is implemented, then each node may send messages to its neighbors without waiting until obtaining the entire file. This is the key idea to leverage message splitting for accelerating content distribution. In the following, we establish an upper bound of the completion time for this case.

Theorem 2. Consider a line graph of n nodes and the protocol of sequential pull-based gossiping with k messages, and let T denote the first time when all nodes are informed. Denote

$m = \min\{n-1, k\}$ and call it the degree of parallelism. Then one has, for all $c > 0$,

$$\Pr\left\{T < \frac{(n+k-2)[\log m + c \log n + \log(n+k-2)]}{q}\right\} > 1 - n^{-c}. \quad (4)$$

Remark 2. We see that a high-probability upper bound of the completion time in the multiple-message case is $\mathcal{O}[(n+k) \log(n+k)/q]$ when both n and k are large enough. Recall that, in Section 3.1, the completion time with single-message gossiping without message splitting scales like $\mathcal{O}(kn/q)$. So we observe that the potential benefit by message splitting can be significant. For example, if $k = n$, then the completion time without message splitting scales like $\mathcal{O}(n^2)$, but with message splitting it scales like $\mathcal{O}(n \log n)$.

On the other hand, a high-probability upper bound of the completion time for complete graphs has been obtained in [10], and it scales like $\mathcal{O}(k + \log n)$ with n nodes and k messages (in [10] the bound $\mathcal{O}(k + \log n)$ is in fact derived when both “pull” and “push” operations are implemented in the gossiping process). It is no wonder that gossiping on complete graphs is much more efficient than on line graphs since there is no communication constraint when all nodes can directly communicate with each other, and such a performance gap is clearly revealed in the difference between the scaling behaviors.

The fundamental reasons on the superiority of the multiple-message gossiping with message splitting over the single-message gossiping without message splitting can be explained as follows. When a file with size of k units needs to be transmitted, each node can only send data to its neighbors after receiving the entire file in the single-message case. However, in the multiple-message case, the file can be split into k messages and each node can send messages to its neighbors without waiting until obtaining the entire file. The parallelism that neighboring nodes can simultaneously obtain different parts of a file can help to accelerate the content distribution.

Before proving Theorem 2, we introduce the key concept of “dependency graph” for gossiping with multiple messages.

(1) *Dependency Graph* (See Figure 1). For $i = 2, 3, \dots, n$ and $j = 1, 2, \dots, k$, let (i, j) denote a node-message pair which indicates the event that node i is pulling message j . We say that a node-message pair is *realized* if the node has obtained the message. We may thus construct a directed graph whose nodes are all the node-message pairs and call it the *dependency graph* since it describes the dependency among all the node-message pairs. Specifically, for (i, j) , it has (at most) two incoming edges, from $(i-1, j)$ and $(i, j-1)$, respectively, meaning that node i can start to pull message j immediately after both events $(i-1, j)$ and $(i, j-1)$ are realized. Take Figure 1 for example: the event $(2, 1)$ has to be realized in the first place; for each of the remaining node-message pairs, it can be realized only after its predecessor pairs, that is, the pairs directing to it, have already been realized.

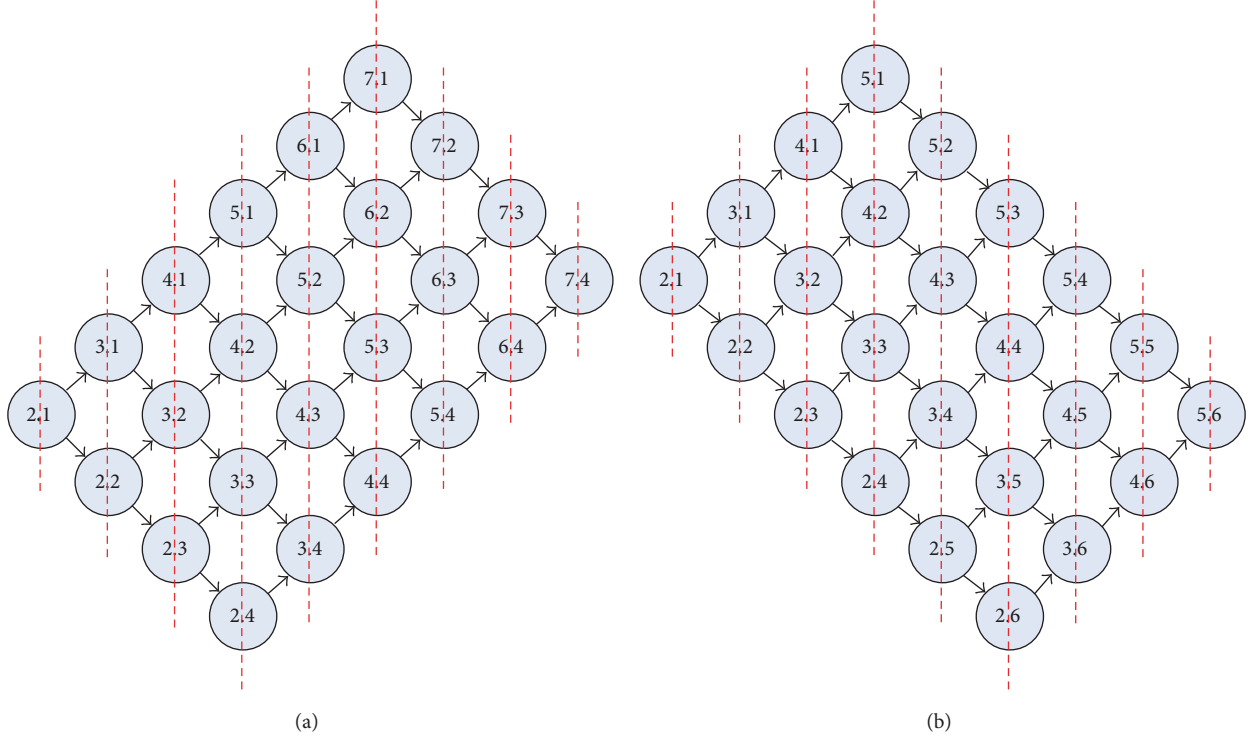


FIGURE 1: Illustration of the dependency graph of all node-message pairs. (a) End-to-end length of line graph ($n - 1 = 6$) is greater than the number of messages ($k = 4$). (b) End-to-end length of line graph ($n - 1 = 4$) is smaller than the number of messages ($k = 6$).

For the dependency graph, we group the node-message pairs into multiple groups, so that two pairs (i, j) and (k, l) are in the same group if and only if $i + j = k + l$. For a line graph of n nodes and k messages, we thus have $n+k-2$ groups, indexed by the node-message-index-sum from 3 to $n+k$. Note that a group has at most $m = \min\{n-1, k\}$ node-message pairs and the quantity $m = \min\{n-1, k\}$ is called the *degree of parallelism*.

Now consider a genie-aided gossiping schedule, which is based on the sequential pull-based gossiping protocol in Section 2, so that a genie enforces that only after all the node-message pairs in group s have been realized, the gossiping processes for the pairs in group $s+1$ can start, for each $s = 3, \dots, n+k-1$. Apparently the resulting completion time T in the genie-aided schedule is stochastically not smaller than the actual completion time. In the following, we will first upper-bound the time for the node-message pairs in a group to be realized, and with that, we will prove Theorem 2 subsequently.

(2) Completion Time for Each Group

Lemma 3. Consider the genie-aided gossiping schedule described above, and let T' denote the time for all the node-message pairs in a group to be realized. Then, for all $c > 0$,

$$\Pr \left\{ T' < \frac{\log m + c \log n + \log(n+k-2)}{q} \right\} > 1 - \frac{n^{-c}}{n+k-2}. \quad (5)$$

Remark 3. We see that a high-probability upper bound of the completion time for each group is $\mathcal{O}[\log(n+k)/q]$ when both n and k are large enough. In the following, we will see that the total completion time for the multiple-message case is the summation over all these groups and the quantity $(n+k-2)$ in the term $n^{-c}/(n+k-2)$ of (5) is the union bound factor for the summation.

Proof of Lemma 3. Let T_1 denote the first time when a node-message pair is realized; that is, a node successfully pulls its desired message. Since a node's attempt to pull from the predecessor node occurs with probability q in the gossiping protocol, we have (the value of $(\log m + c \log n + \log(n+k-2))/q$ might not be an integer, but it does not affect the magnitude of the asymptotic upper bound of T_1)

$$\begin{aligned} \Pr \left\{ T_1 \geq \frac{\log m + c \log n + \log(n+k-2)}{q} \right\} &\leq (1-q)^{\lceil \log m + c \log n + \log(n+k-2) \rceil / q} \\ &= \left((1-q)^{1/q} \right)^{\log m + c \log n + \log(n+k-2)} \\ &\leq \exp(-[\log m + c \log n + \log(n+k-2)]) \\ &= \frac{n^{-c}}{m(n+k-2)}. \end{aligned} \quad (6)$$

Now, taking the union bound over all the node-message pairs in one specific group completes the proof, since each

group has at most m node-message pairs and the realizations of those pairs are mutually independent. \square

(3) *Completion Time for Gossiping on Line Graphs.* We are now ready to prove Theorem 2.

Proof of Theorem 2. First, consider the case where $n - 1 \geq k$. In the dependency graph (see, e.g., Figure 1(a)), all node-message pairs are grouped as

$$\begin{aligned}
 \text{group } s_3 &= \{(2, 1)\}, \\
 \text{group } s_4 &= \{(3, 1), (2, 2)\}, \\
 &\vdots \\
 \text{group } s_{k+2} &= \{(k+1, 1), (k, 2), \dots, (2, k)\}, \\
 &\vdots \\
 \text{group } s_{n+2} &= \{(n, 2), (n-1, 3), \dots, (n-k+2, k)\}, \\
 &\vdots \\
 \text{group } s_{n+k-1} &= \{(n, k-1), (n-1, k)\}, \\
 \text{group } s_{n+k} &= \{(n, k)\}.
 \end{aligned} \tag{7}$$

Similarly, for the case where $n - 1 < k$ (see, e.g., Figure 1(b)), all node-message pairs are grouped as

$$\begin{aligned}
 \text{group } s_3 &= \{(2, 1)\}, \\
 \text{group } s_4 &= \{(3, 1), (2, 2)\}, \\
 &\vdots \\
 \text{group } s_{n+1} &= \{(n, 1), (n-1, 2), \dots, (2, n-1)\}, \\
 &\vdots \\
 \text{group } s_{k+2} &= \{(n, k-n+2), (n-1, k-n+3), \dots, (2, k)\}, \\
 &\vdots \\
 \text{group } s_{n+k-1} &= \{(n, k-1), (n-1, k)\}, \\
 \text{group } s_{n+k} &= \{(n, k)\}.
 \end{aligned} \tag{8}$$

For both of these two cases, note that there are totally $n + k - 2$ groups and that the number of node-message pairs in each group is at most m . Using Lemma 3, we know that, in each group, all the node-message pairs are realized within time $\lceil \log m + c \log n + \log(n + k - 2) \rceil / q$ with probability greater than $1 - n^{-c} / (n + k - 2)$.

Since in the genie-aided schedule the gossiping processes are executed group by group sequentially, with $n+k-2$ groups, we have

$$\begin{aligned}
 T &< \sum_{i=1}^{n+k-2} \left(\frac{\log m + c \log n + \log(n + k - 2)}{q} \right) \\
 &= \frac{(n + k - 2) [\log m + c \log n + \log(n + k - 2)]}{q}.
 \end{aligned} \tag{9}$$

Eventually, taking the union bound leads to the fact that (9) holds with probability greater than $1 - n^{-c}$, and thus the proof of Theorem 2 is completed. \square

4. Completion Time on Several Structured Networks

In this section, we extend the result for line graphs to several other classes of network topologies. The analysis of line graphs is instrumental for establishing completion time bounds for other classes of graphs; see, for example, [15, 17]. We define $T(G)$ as the *high-probability completion time* of gossiping with multiple messages on network graph G , if all nodes obtain all messages within $T(G)$ rounds with probability at least $1 - c_0 \cdot n^{-c}$ for some constant $c > 0$ and fixed constant $c_0 > 0$, for any sufficiently large n .

For a connected network of n nodes with maximum degree δ , assume that each node u pulls messages from any of its neighbors with equal probability $1/\delta_u$ (δ_u is the degree of u). In order to apply Theorem 2, consider the shortest path from the source node to another node u , denoting the length of the shortest path by d_u . Along this shortest path, in each round, each node pulls message from its predecessor node with probability at least $1/\delta$. So, from Theorem 2, the completion time for node u to obtain all the k messages is upper-bounded by $\mathcal{O}(\delta(d_u + k)(\log(d_u + k) + c \log n))$, with probability at least $1 - n^{-c}$. In the following, we apply the result to several classes of graphs.

(1) *Ring Graphs.* A ring graph of n nodes can be simply divided into two equal-length line graphs each of length $n/2$, and its maximum degree is $\delta = 2$. Therefore, we have the following corollary.

Corollary 4. *For a ring graph of n nodes, the sequential pull-based gossiping protocol with k messages behaves like*

$$T(G) = \mathcal{O} \left((n + 2k) \left(\log \left(\frac{n}{2} + k \right) + c \log n \right) \right), \tag{10}$$

for any $c > 0$.

(2) *General Graphs with Fixed Diameter and Maximum Degree.* For a general connected graph with diameter d and maximum degree δ , a high-probability upper bound of the completion time in the single-message case is $\mathcal{O}(\delta(d + \log n))$ [15]. Now, for any arbitrary node in the network, from our discussion above and Theorem 2, the completion time for that node to obtain all the k messages is upper bounded by

$\mathcal{O}(\delta(d+k)(\log(d+k) + (1+c)\log n))$, with probability at least $1 - n^{-1-c}$. Taking the union bound over all the n nodes in the network hence leads to the following corollary.

Corollary 5. *For a general graph of n nodes with diameter d and maximum degree δ , the sequential pull-based gossiping protocol with k messages behaves like*

$$T(G) = \mathcal{O}(\delta(d+k)(\log(d+k) + (1+c)\log n)), \quad (11)$$

for any $c > 0$.

(3) *Grid Graphs.* For a grid graph of n nodes on a $\sqrt{n} \times \sqrt{n}$ lattice, by setting $d = 2\sqrt{n}$ and $\delta = 4$ in (11), we have the following corollary.

Corollary 6. *For a grid graph of n nodes on a $\sqrt{n} \times \sqrt{n}$ lattice, the sequential pull-based gossiping protocol with k messages behaves like*

$$T(G) = \mathcal{O}((2\sqrt{n} + k)(\log(2\sqrt{n} + k) + (1+c)\log n)), \quad (12)$$

for any $c > 0$.

(4) *Random Geometric Graphs.* A random geometric graph $G(n, r)$ is constructed by placing n nodes independently and uniformly at random in square $[0, \sqrt{n}]^2$, where any pair of two nodes are connected by an edge if and only if their Euclidean distance is at most r . From [17], the diameter of the largest connected component of $G(n, r)$ is $\Theta(\sqrt{n}/r)$ and all nodes have degree smaller than $\Theta(\log n)$ when $r = \mathcal{O}(\sqrt{\log n})$. Therefore, the following corollary follows by setting $d = \Theta(\sqrt{n}/r)$ and $\delta = \Theta(\log n)$ in (11).

Corollary 7. *For a random geometric graph $G(n, r)$ with $r = \mathcal{O}(\sqrt{\log n})$, under the sequential pull-based gossiping protocol with k messages, the high-probability completion time of the largest connected component in $G(n, r)$ behaves like*

$$T(G) = \mathcal{O}\left((\log n) \left(\frac{\sqrt{n}}{r} + k\right) \left(\log\left(\frac{\sqrt{n}}{r} + k\right) + (1+c)\log n\right)\right), \quad (13)$$

for any $c > 0$.

5. Simulations

In this section, we carry out experiments to validate our analytical results against the sequential pull-based gossiping with multiple messages. We treat a *multihopping data dissemination* problem for sensor networks or for ad hoc networks, in which the content distribution can be exactly modeled by a *multiple-message gossiping* process on line graphs as well.

Consider a multihopping network, where a source node u_s wishes to spread a file with size of k units to a destination node u_t . The file is equally split into k messages, indexed as $1, 2, \dots, k$. Suppose the routing path from u_s to u_t is known

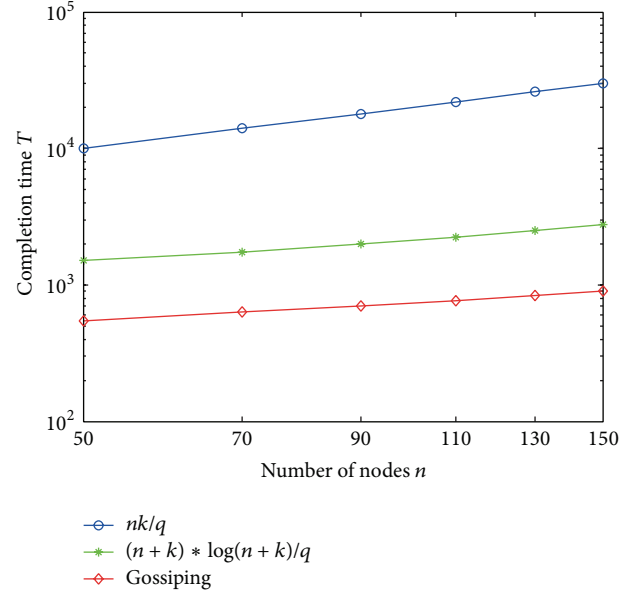


FIGURE 2: Log-log plot of gossiping's completion time versus node number ($k = 100$, $q = 0.5$).

and consists of n nodes. These nodes are sequentially indexed as $1, 2, \dots, n$ from the source to the destination. Time is slotted and a message can be transferred from a sender to a receiver within a round. In each round, each uninformed node u requests the message with the smallest index it has not possessed from its neighbor $u - 1$. All the communications between neighbors are assumed to fail with probability $1 - q$, since there may be wireless error or the requested nodes may be busy. The protocol overhead is assumed to be encapsulated by physical-layer design, which is not considered herein. Recall the gossiping protocol described in Section 2, and then we see that this multihopping data dissemination can be exactly modeled by a multiple-message gossiping process on a line graph.

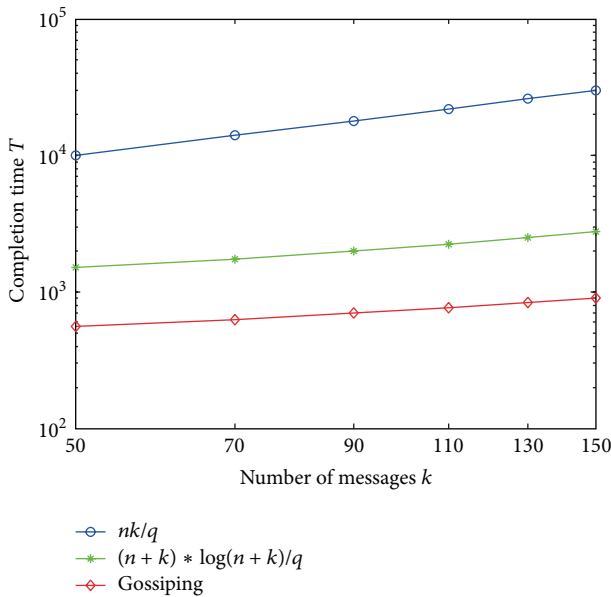
During our experiments, we let the multiple-message gossiping process be run for 1000000 times and record the completion time. The line graph $G = (V, E)$ consists of n nodes, and one endpoint is endowed with k messages. The successful probability q of communications from senders to receivers is 0.5 for all experiment runs. The pseudocode of the simulation using the multiple-message gossiping protocol is presented in Algorithm 1.

The simulation results on gossiping's completion time are demonstrated in Figures 2 and 3, where the theoretical results are also presented and all the curves are plotted in the log-log scale. In Figure 2, k is fixed to be 100 and n ranges from 50 to 150; and in Figure 3, n is fixed to be 100 and k ranges from 50 to 150. In both figures, the curve of " nk/q " presents the theoretical completion time of the naive gossiping without message splitting, the curve of " $(n+k) * \log(n+k)/q$ " presents the theoretical completion time of the multiple-message gossiping predicted by Theorem 2, and the curve of "gossiping" presents the maximum value of the completion time recorded from our experiments. From

```

for  $r = 1 \rightarrow 1000000$  do
  Initialization: a node in  $G$  is informed at round  $t = 0$ .
  while not all nodes in  $G$  are informed do
     $t = t + 1$ .
    for each uninformed node  $u$  do
       $u$  selects a neighbor  $v$  from its neighbor set  $N(u)$  with probability  $q_{uv} \equiv q$ .
       $u$  selects a message with the smallest index  $i \leq k$  it has not possessed.
       $u$  attempts to pull the message  $i$  from the neighbor  $v$ .
      if  $v$  does possess  $i$  then
         $u$  becomes informed at the beginning of round  $t + 1$ .
      end if
    end for
  end while
  Record the completion time  $T$ .
end for

```

ALGORITHM 1: Gossiping($G; k; q$).FIGURE 3: Log-log plot of gossiping's completion time versus message number ($n = 100$, $q = 0.5$).

the simulation results, we see that the benefit by message splitting is significant for accelerating data dissemination, and the upper bound of the multiple-message gossiping protocol established in Theorem 2 is validated as well. However, there is still gap between the analytical upper bound and the simulated completion time. It is an open issue to find a tighter upper bound of the completion time for gossiping with multiple messages on line graphs.

6. Conclusions

In this paper, we have investigated the problem of gossiping with multiple messages on structured networks so as to shed insight into the behavior of the multiple-message gossiping. We have developed the “dependency graph” analytical

technique and further derived an upper bound of the high-probability completion time on line graphs. The potential benefit has been revealed for message splitting to accelerate content distribution through networks, and the result for line graphs has also been further extended to several other classes of network topologies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work has been supported by National Basic Research Program of China (973 Program) through Grant 2012CB316004, SRFDP and RGC ERG Joint Research Scheme through Grant 20133402140001, National Natural Science Foundation of China through Grant 61379003, and the 100 Talents Program of Chinese Academy of Sciences.

References

- [1] S. Selvakenedy and S. Sinnappan, “An adaptive data dissemination strategy for wireless sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 3, no. 1, pp. 23–40, 2007.
- [2] C. Boldrini, M. Conti, and A. Passarella, “ContentPlace: social-aware data dissemination in opportunistic networks,” in *Proceedings of the 11th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM '08)*, pp. 203–210, October 2008.
- [3] W. Gao and G. Cao, “User-centric data dissemination in disruption tolerant networks,” in *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 3119–3127, Shanghai, China, April 2011.
- [4] Z. Li, G. Xie, K. Hwang, and Z. Li, “Churn-resilient protocol for massive data dissemination in P2P networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1342–1349, 2011.

- [5] H. Zhao and J. Zhu, "Efficient data dissemination in urban VANETs: parked vehicles are natural infrastructures," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 151795, 11 pages, 2012.
- [6] A. M. Frieze and G. R. Grimmett, "The shortest-path problem for graphs with random arc-lengths," *Discrete Applied Mathematics*, vol. 10, no. 1, pp. 57–77, 1985.
- [7] J. Tang, S. Dai, J. Li, and S. Li, "Gossip-based scalable directed diffusion for wireless sensor networks," *International Journal of Communication Systems*, vol. 24, no. 11, pp. 1418–1430, 2011.
- [8] Z. Huang and Q. Huang, "To reach consensus using uninorm aggregation operator: a gossip-based protocol," *International Journal of Intelligent Systems*, vol. 27, no. 4, pp. 375–395, 2012.
- [9] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Tech. Rep. CS-200006, Duke University, 2000.
- [10] S. Sanghavi, B. Hajek, and L. Massoulié, "Gossiping with multiple messages," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4640–4654, 2007.
- [11] P. G. Lind, L. R. Da Silva, J. S. Andrade, and H. J. Herrmann, "Spreading gossip in social networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 3, Article ID 036117, 2007.
- [12] F. Chierichetti, S. Lattanzi, and A. Panconesi, "Rumor spreading in social networks," *Theoretical Computer Science*, vol. 412, no. 24, pp. 2602–2610, 2011.
- [13] S. Deb, M. Médard, and C. Choute, "Algebraic gossip: a network coding approach to optimal multiple rumor mongering," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2486–2507, 2006.
- [14] A. D. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, 2008.
- [15] U. Feige, D. Peleg, P. Raghavan, and E. Upfal, "Randomized broadcast in networks," *Random Structures and Algorithms*, vol. 1, no. 4, pp. 447–460, 1990.
- [16] R. Elsässer, "On the communication complexity of randomized broadcasting in random-like graphs," in *Proceedings of the 18th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA '06)*, pp. 148–157, August 2006.
- [17] M. Bradonjić, R. Elsässer, T. Friedrich, T. Sauerwald, and A. Stauffer, "Efficient broadcast on random geometric graphs," in *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10)*, pp. 1412–1421, 2010.
- [18] T. Sauerwald, "On mixing and edge expansion properties in randomized broadcasting," *Algorithmica*, vol. 56, no. 1, pp. 51–88, 2010.
- [19] G. Giakkoupis and T. Sauerwald, "Rumor spreading and vertex expansion," in *Proceedings of 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*, pp. 1623–1641, 2012.
- [20] D. Shah, *Gossip Algorithms*, Now Publishers, 2009.
- [21] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.