Information Dissemination via Gossip: Applications to Averaging and Coding

Damon Mosk-Aoyama
Stanford University
damonma@cs.stanford.edu

Devavrat Shah
MIT
devavrat@mit.edu

Abstract

We study distributed algorithms, also known as *gossip* algorithms, for information dissemination in an arbitrary connected network of nodes. Distributed algorithms have applications to peer-to-peer, sensor, and ad hoc networks, in which nodes operate under limited computational, communication, and energy resources. These constraints naturally give rise to "gossip" algorithms: schemes in which nodes repeatedly communicate with randomly chosen neighbors, thus distributing the computational burden across all the nodes in the network.

We analyze the information dissemination problem under the gossip constraint for arbitrary networks, and find that the information dissemination time of a gossip algorithm is strongly related to the isoperimetric properties of the underlying graph. This characterization allows us to formulate the problem of finding the fastest information dissemination algorithm as a concave maximization problem over the convex set of graph-conformant doubly stochastic matrices.

Next, we use these results for two seemingly unrelated important questions: *distributed averaging* and *coding based information dissemination*. For averaging, we analyze an algorithm based on a classic result of Flajolet and Martin [7]. Information dissemination based on coding was introduced by Deb and Médard [6]. They showed the virtue of coding by analyzing a coding algorithm for a complete graph. Although their scheme generalizes to arbitrary graphs, the analysis does not. We present an analysis of this algorithm for arbitrary graphs, which suggests that for a large class of graphs, such as grid-like graphs, coding-based algorithms do not seem to improve performance.

Finally, we apply our results to several classes of graphs: complete graphs, expander graphs, and grid graphs.

1 Introduction

With the development of peer-to-peer, sensor, and wireless ad hoc networks, there has been recent interest in distributed algorithms for information dissemination and fault-tolerant computation. This is due primarily to the following operational characteristics, which constrain such networks: (i) the network may not have a centralized entity for facilitating computation, communication, and time synchronization; (ii) the network topology may not be completely known to the nodes of the network; (iii) nodes may join or leave the network (even expire), so that the network topology itself may change; and (iv) in the case of sensor networks, the computational power and energy resources may be very limited. These constraints motivate the design of simple decentralized algorithms for computation, in which each node exchanges information with only a few of its immediate neighbors in a time unit (or round). The goal in this setting is to design algorithms so that the desired communication and computation are performed as quickly and efficiently as possible.

We first study the problem of distributed information dissemination: given a network of n nodes, each node wishes to disseminate its own information to all the other nodes as quickly as possible via a *gossip* algorithm. This problem is defined in detail in Section 1.1. We analyze a class of randomized algorithms, and find that the time required for dissemination of the information is related to isoperimetric properties, analogous to conductance, of a probability matrix that specifies the algorithm. This characterization allows us to pose the question of finding an optimal algorithm as the problem of finding a (graph-conformant) doubly stochastic probability matrix with maximum conductance. It turns out that this problem involves the maximization of a concave function over a convex set, and hence it can be solved easily.

We apply these results to analyze two seemingly unrelated gossip algorithms for two different questions. The first question involves distributed averaging. Distributed averaging arises in many applications, such as the coordination of autonomous agents, distributed estimation, distributed data fusion on ad hoc networks, and decentralized optimization. This problem has received a lot of attention [11, 4]. We analyze an averaging algorithm, suggested in a sequence of previous papers [5, 3, 16], which is based on a classic result of Flajolet and Martin [7]. In particular, we show that the averaging time of this algorithm is strongly related to the information dissemination time. We give an example of a graph on which the distributed averaging algorithm based on [7] is better than the optimal iterative algorithm of [4, 11].

The second question concerns the problem of information dissemination via network coding. Recently, Deb and Médard [6] proposed a gossip algorithm for information dissemination using random linear codes, and showed that for a complete graph the algorithm performs much better than a randomized gossip algorithm. Their scheme works for arbitrary graphs. However, its analysis for arbitrary graphs is not so straightforward. Using the insights from our analysis of our basic information dissemination algorithm, we analyze the coding-based information dissemination algorithm. Our results show that the coding-based algorithm does not improve performance for grid-like graphs.

1.1 Setup and model

Information dissemination. We consider the following model for information dissemination in a network. Let $G = (V, E)$ be a connected graph, with $|V| = n$ nodes. We assume that each node $i \in V$ begins the protocol with a single distinct message, m_i. If the edge set E contains an edge (i, j), then the nodes i and j can exchange information during the algorithm.

The protocol is asynchronous[1], and proceeds in a sequence of rounds. A natural model for asynchronous operation is as follows. Each node has an independent clock. The clock ticks according to a Poisson process of rate 1. When the clock at node i ticks, node i is said to become active.

An alternative characterization of this process involves a single global clock ticking according to a Poisson process of rate n. A clock tick corresponds to the start of a round. At each tick of the global clock, exactly one of the nodes is chosen to become active. This choice is made independently and uniformly at random over V.

When node i becomes active, it chooses one of its neighbors, say j, as a communication partner. We consider a simple randomized scheme for choosing the communication partner of a node when it becomes active. On becoming active, node i contacts node j with probability P_{ij}. Thus, an $n \times n$

[1] All the results of this paper do not change qualitatively if the protocol is assumed to be synchronous. In particular, the convergence time of algorithm becomes n times faster than that under the asynchronous model.

matrix $P = [P_{ij}]$ characterizes the algorithm. When node i contacts j, they exchange messages, with i sending all[2] of its messages to j, and receiving all of the messages that j has.

We are interested in the number of rounds required for every node to receive all of the n messages. Under the above protocol, the dissemination of a fixed message m is not affected by the presence of other messages at the nodes that are transmitting m. Hence, we will focus our attention on the dissemination time of a single message from the initial node to all the other nodes. Now, for a particular message m_i, let S_t be the set of nodes that have the message after round t of the protocol. With S_0 denoting the initial set of nodes that contain m_i at the outset of the protocol, we have $S_0 = \{i\}$.

Definition 1. For any $\epsilon \in (0, 1)$, the ϵ-spreading time of a communication matrix P, denoted by $T_{\text{spr}}(\epsilon, P)$, is

$$T_{\text{spr}}(\epsilon, P) = \sup_{\substack{S_0=\{i\},\\ i\in V}} \inf\{t : \Pr(|S_t| < n) \leq \epsilon\}.$$

This definition captures the worst case, over all nodes i, of the number of rounds required for every other node to receive the message m_i that originates at i.

Averaging. The setup is similar to that for information dissemination. In this setting, each node i has a positive integer x_i initially. Let $x(0) = (x_i)$ denote the n-dimensional vector containing the initial values at the nodes. The goal is to compute the average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ at each node. We do not consider any fixed protocol. The precise protocol of interest will be described in Section 3. In general, the performance of a protocol is measured by the averaging time T_{ave}, which is defined as follows.

Definition 2. For any $\epsilon, \delta \in (0, 1)$, the ϵ-averaging time of a protocol, denoted by $T_{\text{ave}}(\epsilon, \delta)$, is defined as follows. For any $x(0)$, each node i converges to an estimate $S(x(0))$ in time $T_{\text{ave}}(\epsilon, \delta)$ with probability at least δ. The estimate, $S(x(0))$, has the property that

$$\Pr\left(\frac{|S(x(0)) - \bar{x}|}{|\bar{x}|} > \epsilon\right) \leq \epsilon.$$

Information dissemination via network coding. Again, the setting is very similar to that of the information dissemination problem. The difference comes from the protocol used to transmit messages. We give the details of this in Section 4. The performance of a protocol is measured in terms of the information spreading time.

1.2 Previous results

In this section, we briefly present a summary of previously known results. The questions considered in this paper have been studied extensively in various contexts. Hence, by no means do we claim to be complete in presenting all the related previous results.

Information dissemination. This question has been studied in various contexts for more than two decades. Notably, the results of [9] established that when the graph is complete, the information

[2]This may require the capacity of links between two nodes to be $\Theta(n)$ times the capacity required for a single message. We will discuss effect of limited capacity links in Section A.

spreading time is $\Theta(n \log n)$[3] for $\epsilon = 1/n$. For other related results, we refer the reader to [17, 18, 10, 11]. We take note of the somewhat related recent work of Ganesh, Massoulie, and Towsley [8] about the spread of epidemics in the network.

Averaging. This question has recently received a lot of attention. Notably, the results of Kempe, Dobra, and Gehrke [11] showed the existence of an averaging algorithm with optimal averaging time of $\Theta(n \log n)$ for $\epsilon = \delta = 1/n$ for a complete graph. However, their results did not extend for arbitrary graphs. In [4], Boyd et al. generalized the results for averaging to arbitrary graphs. They analyzed a large class of averaging algorithms, and found the averaging time to be related to the mixing time of a random walk related to the algorithm. They also found an optimal averaging algorithm as a solution to a semidefinite program. Their results on the averaging time also provide bounds on the information dissemination time (a lot weaker than the results of this paper).

Information dissemination via coding. Network coding has been studied in a number of recent papers, such as [1, 14, 13, 12]. More recently, Deb and Médard [6] showed that a coding-based gossip algorithm for information dissemination can spread information faster than the randomized gossip algorithm of [9] in a complete graph. Their algorithm easily generalizes to arbitrary graphs. However, their method of analysis does not extend.

1.3 Main results

Consider the following definitions, which are similar to the popular notion of conductance that is used in the analysis of the mixing times of Markov chains [20].

Definition 3. For a nonempty proper subset $S \subset V$ of vertices, the *uniform ergodic flow* across the cut (S, S^c), denoted $F_P(S, S^c)$, is

$$F_P(S, S^c) = \sum_{i \in S, j \in S^c} (P_{ij} + P_{ji}).$$

Definition 4. The *uniform conductance*, denoted $\Phi^u_P(S)$, of a nonempty proper subset $S \subset V$ is

$$\Phi^u_P(S) = \frac{F_P(S, S^c)}{|S|}.$$

Let $\mathcal{C}$ be the set of nonempty proper subsets $S \subset V$ such that S induces a connected subgraph on G. We make use of the following two quantities in our analysis.

$$F^k_P = \min_{\substack{S \in \mathcal{C}, \\ |S| = k}} F_P(S, S^c)$$

$$\Phi^k_P = \min_{\substack{S \in \mathcal{C}, \\ |S| \leq k}} \Phi^u_P(S)$$

Note that because of the symmetry in the definition of F_P, we have $F_P(S, S^c) = F_P(S^c, S)$ for any subset $S \subset V$, which implies that $F^k_P = F^{n-k}_P$. Analogous to the standard notion of *conductance* used in the literature [20], we define the uniform conductance of P as $\Phi^u_P \triangleq \Phi^{n/2}_P$.

[3]The time is evaluated in our model. Since our model is different from the original work, the results stated in this paper may look different from the original work.

Information dissemination. The information dissemination algorithm described in Section 1.1, with the communication matrix P, performs as follows. We use the following notation for vectors, all of which have dimension n: $\vec{1}$ has 1 as every entry; b^k has the entries $b^k_k = 1$ and $b^k_i = 0$ for $i \neq k$; and $\min\{x, y\}$ denotes the element-wise minimum of the two vectors x and y.

Theorem 1. *For any $\epsilon \in (0,1)$, the ϵ-spreading time, $T_{spr}(\epsilon, P)$ is bounded as*

$$L(\epsilon, P) \;\leq\; T_{spr}(\epsilon, P) \;\leq \mu_n + 8\sqrt{\frac{\log \epsilon^{-1}\mu_n}{F^*}},$$

where $\mu_n = n\left(\sum_{k=1}^{n-1} F_P^{k^{-1}}\right)$, $F^* = n^{-1}\min_{k=1}^{\lfloor n/2 \rfloor} F_P^k$, and

$$L(\epsilon, P) = \max_{k=1,\ldots,n} \min\{t : q(t) \geq (1-\epsilon)\vec{1}, \textit{when } q(0) = b^k\}$$

with $q(t+1) = \min\{\left(I + \frac{1}{n}\left(P + P^T\right)\right) q(t), \vec{1}\}$ for all $t \geq 0$.

Averaging. As discussed in Section 3, we can use the framework of the information dissemination algorithm from Section 1.1 to estimate the average of a collection of integers, one at each node in the network.

Theorem 2. *For any $2 \times 10^{-3} < \epsilon < 1$ and $\delta \in (0,1)$, there is an asymptotic distributed averaging algorithm based on matrix P such that*

$$T_{ave}(\epsilon, \delta) \leq T_{spr}(\delta/m(\epsilon), P),$$

where $m(\epsilon) \in [c_1\epsilon^{-3}, c_2\epsilon^{-3}]$ for some universal constants $0 < c_1 < c_2$ for all $\epsilon \in (0,1)$.

The Section 3.3 shows the implication of the above result with help of an example.
Information dissemination via coding. The coding-based information dissemination algorithm (described in detail in Section 4) performs as follows.

Theorem 3. *For any $\epsilon \in (0,1)$ and large enough n, under the gossip algorithm based on Random Linear Coding (over the finite field $\mathbf{F}_q$), using the matrix P,*

$$T_{spr}(\epsilon, P) \;\leq\; \hat{\mu}_n + 8\sqrt{\frac{\log \epsilon^{-1}\hat{\mu}_n}{\Phi_P^{n-1}}}, \textit{ where } \hat{\mu}_n = 2\left(1 - \frac{1}{q}\right)^{-1}\left(\sum_{k=1}^{n-1} \frac{k}{\Phi_P^k}\right). \tag{1}$$

1.4 Organization

The rest of the paper is organized as follows. In Section 2, we analyze the basic information dissemination gossip algorithm presented in Section 1.1. We characterize the information dissemination time as closely related to properties of a random walk related to the algorithm. Using this characterization, we study an optimal information dissemination algorithm in Section 2.3. We use the results of Section 2 to analyze a distributed averaging algorithm based on the results of [7] in Section 3. In Section 4, we analyze the network coding based information dissemination algorithm. In Section 5, we apply our results for three graphs of interest: grid graph, expander graphs, and complete graphs. Finally, we present our conclusions.

1.5 Preliminaries on Geometric random variables

Consider a sequence of independent Geometric random variables $G_1, \ldots, G_k$ with parameters $p_1, \ldots, p_k$, where $p_i, 1 \leq i \leq k$ are small. Now consider independent Exponential random variables $X_1, \ldots, X_k$ where X_i is of rate $\theta_i = \log(1-p_i)^{-1}$. It is straightforward to see that X_i+2 stochastically dominates G_i and $X_i - 1$ is stochastically dominated by G_i. Define, $S_k = \frac{1}{k}\sum_{i=1}^{k} G_i$ and $\hat{S}_k = 2 + \frac{1}{k}\sum_{i=1}^{k} X_i$. Then, $\hat{S}_k$ stochastically dominates S_k and $\hat{S}_k - 3$ stochastically dominates S_k. Thus, to obtain bounds on $\Pr(S_k > l)$ it is sufficient obtain bounds on $\Pr(\hat{S}_k > l)$. We state the following result.

Lemma 4. *For $\hat{S}_k$ as defined above, let $\hat{\mu}_k = E[\hat{S}_k]$. By definition,*

$$\hat{\mu}_k \quad = \quad 2 + \frac{1}{k}\sum_{i=1}^{k}\frac{1}{\theta_i}. \tag{2}$$

Let $\theta^* = \min_i \theta_i$. Then,

$$\Pr(\hat{S}_k > (1+\epsilon)\hat{\mu}_k) \leq \exp\left(-\frac{k\epsilon^2\theta^*\hat{\mu}_k}{32}\right).$$

Before we present the proof of Lemma 4, we present a straightforward corollary using above discussion.

Corollary 5. *For S_k as defined above, let $\mu_k = E[S_k]$. Then, for $\epsilon > 0$ and $\theta^* = \min_i \log(1-p_i)^{-1}$,*

$$\Pr(S_k > (1+\epsilon)(\mu_k + 3)) \leq \exp\left(-\frac{k\epsilon^2\theta^*\mu_k}{32}\right).$$

Proof of Lemma 4. Consider the following. Let $\delta = \lambda\theta^* > 0$ and $t > \mu_k$,

$$\begin{aligned}
\Pr(\hat{S}_k > t) &= E[\mathbf{1}_{\{k\hat{S}_k - kt\}}] \\
&\leq E[\exp(\delta(k\hat{S}_k - kt))] \\
&\leq \exp(-\delta kt)\prod_{i=1}^{k} E[\exp(\delta(X_i+2))] \\
&= \exp(-\delta k(t-2))\prod_{i=1}^{k} E[\exp(\delta X_i)] \\
&= \exp(-\delta k(t-2))\prod_{i=1}^{k}\left(1-\frac{\delta}{\theta_i}\right)^{-1}, \qquad (3)
\end{aligned}$$

where the last equality follows from the well-known fact that for an Exponential random variable, X, of rate θ the $E[\exp(\delta X)] = \left(1-\frac{\delta}{\theta}\right)^{-1}$. For δ really small compared to each θ_i, for $1 \leq i \leq k$, we obtain that

$$\begin{aligned}
\prod_{i=1}^{k}\left(1-\frac{\delta}{\theta_i}\right)^{-1} &= \exp\left(-\sum_{i=1}^{k}\log(1-\delta/\theta_i)\right) \\
&\overset{(a)}{\leq} \exp\left((1+0.25\epsilon)\sum_{i=1}^{k}\delta/\theta_i\right) \\
&= \exp(k\delta\hat{\mu}_k + 0.25k\delta\epsilon\hat{\mu}_k - 2k\delta - 0.5k\epsilon\delta), \qquad (4)
\end{aligned}$$

where (a) uses $\log(1-x) \geq -(1+0.25\epsilon)x$ for $x \leq \epsilon/8$. Hence, it is sufficient to have $\delta/\theta^* = \lambda = \epsilon/8$. From (3) and (4), we obtain

$$\Pr(\hat{S}_k > t) \quad \leq \quad \exp(-k\delta t + k\delta\hat{\mu}_k(1+0.25\epsilon) - 0.5k\delta\epsilon). \tag{5}$$

Hence, for $t = (1+\epsilon)\hat{\mu}_k$, we obtain

$$\Pr(\hat{S}_k > (1+\epsilon)\hat{\mu}_k) \quad \leq \quad \exp(-0.75k\delta\epsilon\hat{\mu}_k + 0.5k\delta\epsilon). \tag{6}$$

Since, $\mu_k > 1$ we obtain

$$\Pr(\hat{S}_k > (1+\epsilon)\hat{\mu}_k) \quad \leq \quad \exp(-0.25k\delta\epsilon\hat{\mu}_k). \tag{7}$$

Replacing $\delta = \lambda\theta^* = \epsilon\theta^*/8$, we obtain

$$\Pr(\hat{S}_k > (1+\epsilon)\hat{\mu}_k) \quad \leq \quad \exp\left(-\frac{k\epsilon^2\theta^*\hat{\mu}_k}{32}\right). \tag{8}$$

□

2 Information dissemination

In this section, we present analysis of the information dissemination gossip algorithm presented in Section 1.1, and prove Theorem 1 (Lemmas 6 and 8). We also give an additional upper bound on the information dissemination time that is based on graph structure (see Lemma 7 below).

2.1 Upper bounds

We study two basic approaches to providing upper bounds on the information dissemination time. The first one uses the uniform ergodic flow property of the communication matrix P (see Lemma 6). The second one is based on an analysis of paths that a message can take as it spreads across the nodes in the network (see Lemma 7).

2.1.1 Flow-based bound

We prove the following lemma, which gives the same bound as Theorem 1.

Lemma 6. *For any $\epsilon \in (0,1)$,*

$$T_{spr}(\epsilon, P) \leq \mu_n + 8\sqrt{\frac{\log \epsilon^{-1}\mu_n}{F^*}},$$

where $\mu_n = n\left(\sum_{k=1}^{n-1} F_P^{k^{-1}}\right)$ and $F^* = n^{-1}\min_{k=1}^{\lfloor n/2 \rfloor} F_P^k$.

Proof. Consider the dissemination of a fixed message m through the network. Let S_t denote the set of nodes containing m after round t. As the algorithm proceeds, the size of S_t increases, from $|S_0| = 1$ to n. Under the asynchronous protocol, by definition $|S_{t+1}| - |S_t| \in \{0,1\}$. That is, the increase in $|S_t|$ is a Bernoulli random variable.

Consider a particular time t. The size of S_t increases in round $t+1$ if a node $i \in S_t$ becomes active and chooses a node $j \in S_t^c$ as its communication partner, or vice versa. This happens with probability $P(S_t) \stackrel{\triangle}{=} \frac{\sum_{i \in S_t, j \in S_t^c}(P_{ij}+P_{ji})}{n}$. If $|S_t| = k$, then, from the definitions above,

$$\begin{aligned} P(S_t) &= \frac{F_P(S_t, S_t^c)}{n} \\ &\geq \frac{F_P^k}{n} \stackrel{\triangle}{=} p_k. \end{aligned} \tag{9}$$

Thus, the probability that $|S_t|$ increases in round $t+1$ is at least p_k when $k = |S_t|$. Since $F_P^k = F_P^{n-k}$, $p_k = p_{n-k}$.

Define $T = \inf\{t : |S(t)| = n\}$. From the above discussion, we obtain that T is stochastically dominated by the random variable $\hat{T} = \sum_{k=1}^{n-1} G_k$, where the G_k are independent Geometric random variables with corresponding parameters p_k. Now,

$$E[T] \leq E[\hat{T}] = \sum_{k=1}^{n-1} \frac{1}{p_k} = n \left(\sum_{k=1}^{n-1} {F_P^k}^{-1} \right). \tag{10}$$

To obtain an upper bound on T that holds with probability $1-\epsilon$, we apply Corollary 5. Let $p^* = \min_{k=1}^n \log(1-p_k)^{-1}$. From Corollary 5, for $\lambda > 0$, we have

$$\Pr\left(\hat{T} > (1+\lambda)(E[\hat{T}]+3)\right) \leq \exp\left(-\frac{\lambda^2 p^* E[\hat{T}]}{32}\right). \tag{11}$$

For the choice of $\lambda = \sqrt{\frac{32 \log \epsilon^{-1}}{p^* E[\hat{T}]}}$, we obtain from (11) that

$$\Pr\left(\hat{T} > E[\hat{T}] + \sqrt{\frac{32 \log \epsilon^{-1} E[\hat{T}]}{p^*}} + 3 + \sqrt{\frac{288 \log \epsilon^{-1}}{p^* E[\hat{T}]}}\right) \leq \epsilon. \tag{12}$$

It follows from (9)-(12) and the fact that $E[\hat{T}] = \Omega(n)$ that for large enough n,

$$\Pr\left(\hat{T} > E[\hat{T}] + 8\sqrt{\frac{\log \epsilon^{-1} E[\hat{T}]}{p^*}}\right) \leq \epsilon. \tag{13}$$

Finally, observe that $p^* = \min_{k=1}^{n-1} \log(1-p_k)^{-1} \geq \min_{k=1}^{n-1} p_k = \min_{k=1}^{\lfloor n/2 \rfloor} p_k = \min_{k=1}^{\lfloor n/2 \rfloor} (F_P^k/n) = F^*$. Hence, (13) implies the statement of Lemma 6. □

2.1.2 Path-based bound

Define the graph $G_P = (V, E_P)$ by the edge set $E_P = \{(i,j) \in E : P_{ij} + P_{ji} > 0\}$. For any edge $e = (i,j) \in E_P$, let $p(e) = (P_{ij} + P_{ji})/n$ and $w(e) = 1/p(e)$. For two nodes i and j, let $\mathcal{Q}_{ij}$ denote the set of simple paths between i and j in G_P. Now, consider a path $Q = (e_1, \ldots, e_\ell) \in \mathcal{Q}_{ij}$. Let $w(Q) = \sum_{k=1}^{\ell} w(e_k)$ and $\hat{w}(Q) = w(Q) + 40\sqrt{\log n w(Q)(\max_{k=1}^{\ell} w(e_k))}$. We present a second upper bound on the ϵ-spreading time of the information dissemination algorithm described in Section 1.1. This approach is based on the path structure of G_P, and yields the following result.

Lemma 7. *Under the information dissemination process with the communication matrix P, let*

$$Q^* = \arg \max_{i,j \in V, i \neq j} \min_{Q \in \mathcal{Q}_{ij}} \hat{w}(Q).$$

Then, $T_{spr}\left(\frac{1}{n}, P\right) \leq \hat{w}(Q^*)$.

Proof. Under the information dissemination algorithm, any edge $e = (i,j) \in E_P$ is chosen for communication in any round with probability $p(e) = (P_{ij} + P_{ji})/n$. We focus our attention on a message originating at node i. For another node $j \neq i$, let T_{ij} be the time it takes for the message to be transmitted from i to j.

Now, let $Q = (e_1, \ldots, e_\ell)$ be a simple (acyclic) path between i and j in the graph G_P. At any point during the dissemination process, j will have the message if all of the edges on Q have been chosen in the order $e_1, \ldots, e_\ell$. Next, we consider the number of rounds needed for this event to occur, which we will denote by T_Q. Then, T_Q is stochastically dominated as

$$T_Q \quad \leq \quad \sum_{k=1}^{\ell} G(k) \quad \stackrel{\triangle}{=} \quad \hat{T}, \tag{14}$$

where the $G(k)$ are independent Geometric random variables with corresponding parameters $p(e_k)$. From (14),

$$E[T_Q] \leq E[\hat{T}] = \sum_{k=1}^{\ell} w(e_k). \tag{15}$$

To obtain an upper bound on the information dissemination time that holds with high probability, we use Corollary 5. Let $p^* = \min_{k=1}^{\ell} \log(1 - p(e_k))^{-1} \geq \min_{k=1}^{\ell} p(e_k)$. Now, from Corollary 5, we can obtain that

$$\Pr\left(\hat{T} > E[\hat{T}] + 40\sqrt{\frac{\log n E[\hat{T}]}{p^*}}\right) \quad \leq \quad n^{-4}. \tag{16}$$

From the definition of $\hat{w}(Q)$, (14) and (16), we obtain

$$\Pr\left(T_Q \geq \hat{w}(Q)\right) \quad \leq \quad n^{-4}. \tag{17}$$

The inequality in (17) holds for all $Q \in \mathcal{Q}_{ij}$. Hence, for $Q^{ij} = \arg\min_{Q \in \mathcal{Q}_{ij}} \hat{w}(Q)$,

$$\Pr\left(T_{ij} \geq \hat{w}(Q^{ij})\right) \quad \leq \quad n^{-4}. \tag{18}$$

Let $Q^* = \arg\max_{i,j \in V, i \neq j} \hat{w}(Q^{ij})$. Then, using the union bound for all $O(n^2)$ node pairs, we obtain that for $T = \sup_{ij} T_{ij}$,

$$\Pr\left(T \geq \hat{w}(Q^*)\right) \quad \leq \quad n^{-2}. \tag{19}$$

This completes the proof of Lemma 7. ∎

2.2 Lower bound

We obtain the following lower bound on $T_{\rm spr}(\epsilon, P)$, as claimed in Theorem 1.

Lemma 8. *Let $L(\epsilon, P)$ be as defined in Theorem 1. Then, for any $\epsilon \in (0,1)$, $T_{spr}(\epsilon, P) \geq L(\epsilon, P)$.*

Proof. As before, we examine the dissemination of a message m that starts at one node k. Let $r(t) = [r_i(t)]$ be vector of size n, with $r_i(t)$ denoting the probability that node i has the message after t rounds. Initially, $r(0) = b^k$. Consider a particular node i. After round $t+1$, the node i will have m if it had m after t rounds, or if in round $t+1$ it communicates with a node j that had m after t rounds. This leads to the following equation for $r_i(t+1)$, in which we use the assumption that $P_{jj} = 0$ for all $j \in V$.

$$r_i(t+1) \quad = \quad r_i(t) + (1 - r_i(t))\left(\frac{1}{n}\sum_{j \in V} P_{ij} r_j(t) + \sum_{j \in V} \frac{1}{n} r_j(t) P_{ji}\right)$$

By dropping the $(1 - r_i(t))$ factor, we obtain an upper bound on the probability that node i has the message after $t+1$ rounds.

$$r_i(t+1) \leq r_i(t) + \frac{1}{n}\sum_{j \in V} (P_{ij} + P_{ji}) r_j(t)$$

This upper bound may be rewritten in vector form as follows. Throughout this work, we adopt the convention that an inequality $x \leq y$ involving two n-dimensional vectors x and y means that $x_i \leq y_i$ for all $i = 1, \ldots, n$.

$$r(t+1) \leq \left(I + \frac{1}{n}\left(P + P^T\right)\right) r(t) \tag{20}$$

Separately, $r(t+1)$ is a probability vector. Hence,

$$r(t+1) \quad \leq \quad \vec{1}. \tag{21}$$

In light of bounds (20) and (21), define vector $q(\cdot)$ as follows. For $t \geq 0$,

$$q(t+1) \quad = \quad \min\left\{\left(I + \frac{1}{n}\left(P + P^T\right)\right) q(t), \vec{1}\right\}, \; q(0) = b^k,$$

where the minimum is taken element-wise. That is, for two vectors x and y of dimension n, the vector $z = \min\{x, y\}$ is defined by $z_i = \min\{x_i, y_i\}$ for all $i = 1, \ldots, n$. Now, the inequality in (20)-(21) implies that if $q(0) = r(0) = b^k$, then

$$r(t) \quad \leq \quad q(t). \tag{22}$$

By definition, for $t \geq T_{\rm spr}(\epsilon, P)$, $r(t) \geq (1 - \epsilon)\vec{1}$. Hence, by definition of $L(\epsilon, P)$ and (22) we obtain that $T_{\rm spr}(\epsilon, P) \geq L(\epsilon, P)$. □

2.3 Information dissemination: optimization

Recall Theorem 1. It suggests that information dissemination time is upper bounded by quantity $O((n \log n)/\Phi^u_P)$. Thus, as the conductance Φ^u_P increase by changing P, the upper bound decreases. Based on this, we make the following assumption.

Assumption 1. Information dissemination time is monotonically decreasing function of conductance, Φ^u_P, of probability matrix P related to the algorithm.

Under Assumption 1, the minimization of information dissemination time is equivalent to maximization of conductance Φ^u_P over probability matrices P. Define $Q = \frac{1}{2}\left(P + P^T\right)$. Then, $\Phi^u_Q = \Phi^u_P$ by definition. Hence, we can restrict our attention towards maximizing conductance over set of doubly stochastic matrices. Next, we claim the following (this may be very well-known, however we could not find a reference).

Lemma 9. *The conductance, Φ^u_Q, is concave as function of doubly stochastic matrix Q.*

Proof. Consider any two doubly stochastic matrices, $Q_1 \neq Q_2$. For $\alpha \in (0,1)$, let $Q = \alpha Q_1 + (1-\alpha)Q_2$. Let S_1, S_2 and S be subset of size $\leq n/2$ such that

$$\Phi^u_{Q_1} = \Phi^u_{Q_1}(S_1), \quad \Phi^u_{Q_2} = \Phi^u_{Q_2}(S_2) \text{ and } \Phi^u_Q = \Phi^u_Q(S).$$

Now consider the following.

$$\begin{aligned} \alpha\Phi^u_{Q_1} + (1-\alpha)\Phi^u_{Q_2} &= \alpha\Phi^u_{Q_1}(S_1) + (1-\alpha)\Phi^u_{Q_2}(S_2) \\ &\overset{(a)}{\leq} \alpha\Phi^u_{Q_1}(S) + (1-\alpha)\Phi^u_{Q_2}(S) \overset{(b)}{=} \Phi^u_{\alpha Q_1 + (1-\alpha)Q_2}(S) \\ &= \Phi^u_Q, \end{aligned} \tag{23}$$

where (a) and (b) follow from definition. The (23) proves the Lemma 9. □

Now, set of doubly stochastic matrices is convex and bounded. Hence, from Lemma 9, the maximization of conductance is equivalent to maximization of a concave function over a bounded convex set. It is well-known that it can be solved easily by methods like simple gradient methods. Further, such optimization can be easily done in a distributed manner. Thus, under Assumption 1, optimal information dissemination algorithm can be easily found.

3 Averaging

We now consider the application of our analysis to the problem of computing the average of a set of positive integers via a distributed algorithm. Suppose that each of the n nodes in the network has a single positive integer, and our goal is to calculate the average of these values. Boyd et al. analyze a randomized gossip algorithm that iteratively computes the average of values at the nodes of a network [4]. The approach for averaging that we study here is a variant of an algorithm by Flajolet and Martin for estimating the number of distinct elements in a multiset [7]. It was suggested by several groups of authors [5, 3, 16], although the running time of this algorithm in a setting such as the one considered in this work was not analyzed.

Our goal is to estimate the sum of the integers, which we will denote by S, and n, the number of nodes in the network. The ratio of these quantities will then serve as an estimate of the average

of the integer values, which can be expressed as $\bar{x} = S/n$. For a fixed $\epsilon \in (2 \times 10^{-3}, 1)$, we seek an estimate $\hat{x}$ of the average that is in the closed interval $[(1-\epsilon)\bar{x}, (1+\epsilon)\bar{x}]$ (the lower bound on the accuracy of the estimate arises from properties of the analysis of the counting algorithm that we employ).

To this end, suppose that we obtain estimates $\hat{S}$ and $\hat{n}$ of S and n, respectively, such that $\hat{S} \in [(1-\epsilon_1)S, (1+\epsilon_1)S]$ and $\hat{n} \in [(1-\epsilon_1)n, (1+\epsilon_1)n]$, where $\epsilon_1 > 0$. Then, the estimate $\hat{x} = \hat{S}/\hat{n}$ of the average will be in the interval $[\bar{x}(1-\epsilon_1)/(1+\epsilon_1), \bar{x}(1+\epsilon_1)/(1-\epsilon_1)]$. We set $\epsilon_1 = \epsilon/3$, which ensures that an estimate in this interval will also be in the interval $[(1-\epsilon)\bar{x}, (1+\epsilon)\bar{x}]$ when $\epsilon < 1$. Since estimating the number of nodes is a special case of estimating the sum when all the integers are 1, we focus our attention on the following task: *given integers $x_1, \ldots, x_n$ at the n nodes, compute an estimate of the sum $S = \sum_{i=1}^{n} x_i$ at all the nodes.*

3.1 Algorithm: FMA

We describe the Flajolet-Martin Algorithm (FMA) [7] in this section. They introducted the idea of stochastic averaging, which is applied to this setting. Assume that all integers can be represented using L bits, so that each integer is in the range $(0, 2^L)$. Each node maintains m bitmaps $B^1, \ldots, B^m$ of length $L+1$, where m is a parameter to be set below. Furthermore, we assume that the nodes have access to a random hash function $h : \{0, 1, \ldots, 2^L - 1\} \to \{0, 1, \ldots, 2^L - 1\}$, which, for any input, produces an output distributed uniformly at random over the integers $0, 1, \ldots, 2^L - 1$.

Consider a node i with integer x_i. For all $k = 1, \ldots, x_i$, the node generates a random integer Y_k, which is independent of all other random variables and distributed uniformly at random over $\{0, 1, \ldots, 2^L - 1\}$. Let $r_k = h(Y_k) \bmod m$ and $q_k = \lfloor h(Y_k)/m \rfloor$. We write $q_k(\ell)$ to denote bit ℓ in the binary representation of q_k, adopting the convention that bits are numbered from 0, so that 0 is the least-significant bit and $L-1$ is the most-significant bit. The function ρ is defined as follows.

$$\rho(q) = \begin{cases} \min\{\ell \geq 0 : q(\ell) = 1\}, & \text{if } q > 0 \\ L, & \text{if } q = 0 \end{cases}$$

Note that $\rho(q)$ is the index of the least-significant bit set to 1 in the binary representation of q when $q > 0$. The node i initializes its bitmaps by setting all of the bits in the m bitmaps to 0. For $k = 1, \ldots, x_i$, it sets bit $\rho(q_k)$ in the bitmap B^{r_k}, which we denote by $B^{r_k}(\rho(q_k))$, to 1.

Now, the nodes in the network use the basic randomized information dissemination protocol with a communication matrix P to compute the bitwise OR of the bitmaps. Because the number of bitmaps m may potentially be large and it is desirable for the algorithm to execute each round in constant time, we assume that in every round the OR of two bitmaps is computed. That is, when a node i with bitmaps $B_i^1, \ldots, B_i^m$ contacts a node j with bitmaps $B_j^1, \ldots, B_j^m$ in any round, each node sets bitmap k to $B_i^k \vee B_j^k$ for one value $k \in \{1, \ldots, m\}$, where $\vee$ denotes the bitwise OR operation on the two bitmap operands. We assume that the value k is chosen to be $k = t \bmod m+1$, where t is the round number.

After all nodes have computed the OR of all m bitmaps, they calculate an estimate of S, denoted as S', as follows. Given a bitmap B, let $Z_B = \{\ell \geq 0 : B(\ell) = 0\}$. Define a function R (that maps a bitmap to an integer in $\{0, \ldots, L\}$) as

$$R(B) = \begin{cases} \min Z_B, & \text{if } Z_B \neq \emptyset \\ L, & \text{if } Z_B = \emptyset \end{cases}$$

Then, an estimate S' of the sum S is

$$S' = \frac{m}{\varphi}\left(2^{\frac{\sum_{k=1}^{m} R(B^k)}{m}}\right),$$

where $\varphi \approx 0.77351$ is a constant defined in [7].

3.2 Analysis

We prove the following result about FMA, which implies Theorem 2.

Lemma 10. *For any $\delta \in (0,1)$ and $\epsilon \in (2 \times 10^{-3}, 1)$, the FMA algorithm, based on matrix P, computes S', an estimate of S, such that $S' \in [(1-\epsilon/2)S, (1+\epsilon/2)S]$ with probability at least $1-\epsilon/4$; and the algorithm takes time $T_{spr}(\delta/m, P)$ with probability at least $1-\delta$ where $m \in [c_1\epsilon^{-3}, c_2\epsilon^{-3}]$ for some universal constants $0 < c_1 < c_2$.*

Proof. The algorithm FMA stops when all nodes have computed the OR of all the m bitmaps. This is the same as the time it takes for m independent information dissemination process to complete. Let these random times be $T_1, \ldots, T_m$. By definition,

$$\Pr\left(T_i > T_{\text{spr}}(\delta/m, P)\right) \quad \leq \quad \delta/m. \tag{24}$$

Hence, by union-bound and (24), we obtain

$$\Pr\left(\sup_{i=1}^{m} T_i > T_{\text{spr}}(\delta/m, P)\right) \quad \leq \quad \delta. \tag{25}$$

The (25) implies that the algorithm FMA computes estimate S' by the time $T_{\text{spr}}(\delta/m, P)$ with probability at least $1-\delta$. Thus, to prove Lemma 10, we need to show that for $m \in [c_1\epsilon^{-3}, c_2\epsilon^{-}3]$, the $S' \in [(1-\epsilon)S, (1+\epsilon)S]$.

In [7], authors showed that for the expectation $E(S')$ and the standard deviation $\sigma(S')$ of S' satisfy the following relationships when S is sufficiently large.

$$\left|\frac{E(S')}{S} - (1+\alpha(m))\right| < 10^{-5} \text{ and } \left|\frac{\sigma(S')}{S} - \beta(m)\right| < 10^{-5} \tag{26}$$

The functions $\alpha(m)$ and $\beta(m)$ are such that there exist universal constants a_1, a_2 such that

$$\alpha(m) \in [a_1/m, a_2/m], \text{ and } \beta(m) \in [a_1/\sqrt{m}, a_2/\sqrt{m}].$$

In view of the bias in the estimate S', for simplicity of proof, we consider a new estimate $\hat{S} = S'/(1+\alpha(m))$. Then, for $m \geq 8a_2/\epsilon$ and $\alpha(m) \geq 0$,

$$S' \quad \in \quad \left[(1-\epsilon/4)\hat{S}, (1+\epsilon/4)\hat{S}\right]. \tag{27}$$

That is, S' and $\hat{S}$ are roughly the same for large m. Now, we bound probability of error in $\hat{S}$. Let $\epsilon_1 = \epsilon/4$. By definition,

$$\hat{S} \notin [(1-\epsilon_1)S, (1+\epsilon_1)S] \quad \Rightarrow \quad |S' - (1+\alpha(m))S| > \epsilon_1(1+\alpha(m))S. \tag{28}$$

Now, (28) and (26) implies the following.

$$\begin{aligned} |S' - E(S')| &\geq |S' - (1+\alpha(m))S| - |(1+\alpha(m))S - E(S')| \\ &\geq [\epsilon_1(1+\alpha(m)) - 10^-5]S. \end{aligned} \tag{29}$$

Recall that by choice of $\epsilon_1 = \epsilon/2 > 10^-3$, the RHS of (29) is positive. Again, using (26),

$$S \geq \frac{\sigma(S')}{\beta(m) + 10^{-5}}. \tag{30}$$

Now, using Chebyshev's inequality along with (28)-(30), we obtain

$$\Pr(\hat{S} \notin [(1-\epsilon_1)S, (1+\epsilon_1)S]) \leq \Pr\left(|S' - E(S')| > \frac{\epsilon_1(1+\alpha(m)) - 10^{-5}}{\beta(m) + 10^{-5}}\sigma(S')\right) \tag{31}$$

$$\leq \left(\frac{\beta(m) + 10^{-5}}{\epsilon_1(1+\alpha(m)) - 10^{-5}}\right)^2. \tag{32}$$

Now, $\beta(m) \leq a_2/\sqrt{m}$. Then, for $m \geq 64a_2^2\epsilon^{-3}$ ($\geq 4a_2\epsilon^{-1}$ required for (27)), $\beta(m) \leq \epsilon^{-1.5}/8$. Further, $\alpha(m) > 0$ and $\epsilon > 2 \times 10^{-3} \Rightarrow 10^{-5} < \epsilon^{1.5}/8$. Combining this with (32), we obtain

$$\Pr(\hat{S} \notin [(1-\epsilon_1)S, (1+\epsilon_1)S]) \leq \epsilon/4. \tag{33}$$

From (25), (33) and recalling that $\epsilon_1 = \epsilon/4$, for $m \geq 64a_2^2\epsilon^{-3}$

$$\Pr(S' \notin [(1-\epsilon/2)S, (1+\epsilon/2)S]) \leq \epsilon/4. \tag{34}$$

Appropriate selection of constants c_1, c_2 yields the Lemma 10. □

To see how Lemma 10 implies the proof of Theorem 2, consider the following. In the special case that all the integers at the nodes are 1, we obtain an estimate n' of the number of nodes n such that $\Pr(n' \notin [(1-\epsilon/2)n, (1+\epsilon/2)n]) \leq \epsilon/4$ as well. These two inequalities yield an upper bound on the probability that $\hat{x} = S'/n'$ is not in the interval $[(1-\epsilon)\bar{x}, (1+\epsilon)\bar{x}]$.

$$\Pr(\hat{x} \notin [(1-\epsilon)\bar{x}, (1+\epsilon)\bar{x}]) \leq \Pr(S' \notin [(1-\epsilon/2)S, (1+\epsilon/2)S]) + \Pr(n' \notin [(1-\epsilon/2)n, (1+\epsilon/2)n]) \leq \epsilon.$$

3.3 Implication

To show the strength of the result, we consider a simple circle graph: n nodes are placed on a circle with each node connected to two other nodes, one on the left and the other on the right. The second largest eigenvalue on this graph is $1-\Theta(n^{-2})$. Hence, the optimal averaging algorithm based on [4] will require $\Omega(n^3 \log \epsilon^{-1})$ time to estimate the average within precision ϵ for any $\epsilon \in (0,1)$. However, the above algorithm will require $O(n^2\epsilon^{-3} + n \log \delta^{-1}) \approx O(n^2)$ time. Thus, the algorithm based on [7] improves performance over the optimal algorithm of [4] by an order of magnitude. It is easy to see that $\Omega(n^2)$ is the minimal time required for communicating even a piece of information from one end to the other end of the ring graph under the asynchronous time model. Thus, in this sense, the averaging algorithm based on [7] is *absolutely optimal*.

4 Information dissemination via coding

Recently, Deb and Médard [6] have shown that the use of random linear coding can help in efficiently spreading messages in a complete graph. However, in the arbitrary graph topology it is not clear how the coding-based algorithm performs. The analysis methods of [6] are specialized for complete graphs, and do not extend to arbitrary graphs. In this section, we present an analysis of the coding-based scheme of [6] for arbitrary graphs. The remainder of the section is organized as follows: first, we present a coding-based gossip scheme for information dissemination over an arbitrary graph. It is a natural modification of the scheme presented in [6] for complete graphs. Then, we present a bound on the time required to spread information under this scheme.

4.1 Coding based gossip algorithm

The coding-based gossip algorithm using a communication matrix P is a natural extension of the basic gossip algorithm described in Section 1.1. As in the setting of Section 1.1, each node starts with its unique message with the goal of spreading its message to all the other nodes.

The algorithm is asynchronous and runs in rounds. In any round, exactly one of the nodes is chosen to become active. This choice is made uniformly at random over V, and independently of all the other random choices. When node i becomes active, it contacts one of its neighbors, say j, with probability P_{ij}. Both nodes, i and j, transmit a code based on their current information to each other, according to the random linear coding (RLC) protocol explained below. When each node has received "enough" coded messages, they can decode (see below) them to obtain all n original messages.

Random Linear Coding (RLC) Protocol. This is exactly the same setup as in [6]. Each message is a vector over a finite field, $\mathbf{F}_q$ of size $q \geq n$. Let each message be a vector of size $r \in \mathbf{Z}$. In particular, let the initial message at node i be $m_i \in \mathbf{F}_q^r$, for $1 \leq i \leq n$. We assume that all the n messages, $\{m_i : 1 \leq i \leq n\}$, are linearly independent. Let $M = \{m_1, \ldots, m_n\}$ denote the set of n message vectors. During the execution of the gossip algorithm, each node collects linear combinations of message vectors in M. When each node has n linearly independent such vectors, it can recover all the messages in M successfully.

Now, consider a round, say t, during the execution of gossip algorithm. Suppose that node i becomes active and contacts node j in this round. Let $S_i(t)$ and $S_j(t)$ be the set of all coded messages at nodes i and j at the beginning of round t. By definition, for $f_l \in S_i(t),\ 1 \leq l \leq |S_i(t)|$, $f_l \in \mathbf{F}_q^r$ and

$$f_l \quad = \quad \sum_{k=1}^{n} a_{l_k} m_k, \quad a_{l_k} \in \mathbf{F}_q. \tag{35}$$

The protocol ensures that node i knows the coefficients a_{l_j} (see [6] for details). Similarly, let $S_j(t) = \{g_1, \ldots, g_{|S_j(t)|}\}$. Now as part of the protocol, node i transmits a random coded message with payload $e_{ij} \in \mathbf{F}_q^r$, where

$$e_{ij} \quad = \quad \sum_{f_l \in S_i(t)} \beta_l f_l, \quad \beta_l \in \mathbf{F}_q, \text{ and } \Pr(\beta_l = \beta) = \frac{1}{q}, \quad \forall \beta \in \mathbf{F}_q. \tag{36}$$

The message e_{ij} can be re-written as follows.

$$e_{ij} = \sum_{f_l \in S_i(t)} \beta_l f_l = \sum_{f_l \in S_i(t)} \beta_l \sum_{k=1}^{n} a_{l_k} m_k = \sum_{k=1}^{n} \left(\sum_{1 \le l \le |S_i(t)|} a_{l_k} \right) m_k = \sum_{k=1}^{n} \theta_k m_k, \quad (37)$$

where $\theta_k = \sum_{1 \le l \le |S_i(t)|} a_{l_k} \in \mathbf{F}_q$. For the purpose of decoding, along with e_{ij}, node i transmits $(\theta_1, \ldots, \theta_n)$ to node j. Analogous to e_{ij}, node j transmits to i a random coded message with payload $e_{ji} \in F_q^r$ and the associated n scalars for decoding purposes. Next, we recall the following key result, which will be used crucially in our analysis.

Lemma 11 (Lemma 2.1, [6]). *Let $S_i(t)^-$ and $S_j(t)^-$ denote the subspaces spanned by the code-vectors $S_i(t)$ and $S_j(t)$ respectively. Let $S_i(t)^+$ and $S_j(t)^+$ be subspaces spanned by code-vectors $S_i(t) \cup \{e_{ji}\}$ and $S_j(t) \cup \{e_{ij}\}$. Then,*

$$\Pr\left(dim(S_i(t)^+) > dim(S_i(t)^-) | S_j(t)^- \not\subseteq S_i(t)^- \right) \ge 1 - \frac{1}{q}.$$

4.2 Analysis

The performance of the gossip algorithm presented in the previous section is described by Theorem 3. Next, we present the proof of Theorem 3.

Proof of Theorem 3. We first present some definitions and notations. Let $t \in \mathbf{Z}_+$ denote the round number of algorithm.

Message space. The subspace spanned by messages at node i in the beginning of time t is denoted by $S_i(t)^-$ and that at the end of round t is denoted by $S_i(t)^+$. Note that, $S_i(t)^+ = S_i(t+1)^-$.

Type. Two nodes, i and j, are called of the same type at time t, if $S_i(t)^- = S_j(t)^-$, that is, the subspace spanned by the messages at nodes i and j are identical. For example, if both nodes have message sufficient to decode all n messages, then subspace spanned by both of them will be the same, that is they are of the same type.

Maximal type-size. Now, consider any type. All the nodes are divided into different equivalent type-class. At time t, let $Y(t)$ be the size of the largest type class, also denoted by maximal type-size.

Dimension increase. When a node i transmits random linear code to node j such that $S_i(t)^- \subseteq /S_j(t)^-$, from Lemma 11, $\dim(S_j(t)^+) \ge \dim(S_j(t)^-) + 1$ with probability $1 - 1/q$. Now, suppose at time t, two nodes i and j are not of the same type. Then it must be that either (a) $S_i(t)^- \not\subseteq S_j(t)^-$ or (b) $S_j(t)^- \not\subseteq S_i(t)^-$. Thus, when two nodes of different type contact each other, at least dimension of one node increases by 1 with probability $1 - 1/q$.

Stopping condition. The information spreading time is equivalent to $\min\{t : \dim(S_i(t)^-) = n, \ \forall i\}$. Initially, at $t = 0$ $\dim(S_i(0)^-) = 1 \ \forall i$. Thus, information spreads to all nodes when overall dimension increase is $n(n-1)$. Let D_k be the smallest time such that net dimension increase is at least k. By definition, $D_0 = 0$ and information spreading time is the same as $D_{n(n-1)}$.

Now, define $t_k = \min\{t : Y(t) \ge k\}$ and $T_k = \min\{j : D_j = t_k\}$. In words, t_k is the first time when any maximal type-size becomes at least k and T_k is the net dimension increase at time t_k. By definition, $T_1 = 0$. We state the following result.

Lemma 12. *For any $1 \le k \le n$, $T_k \ge k(k-1)$.*

Proof. Consider time t_k when the first time any maximal type-size becomes k. At this time, there is a type such that corresponding type class has k nodes. Let they be, $i_1, \ldots, i_k$. Since they are of the same type, it must be that $S_{i_1}(t_k)^- = \cdots = S_{i_k}(t_k)^-$. By definition, $m_{i_l} \in S_{i_l}(t)^-$, for all $t \geq 0$. Hence, for all $1 \leq l \leq k$, $\text{span}(m_{i_1}, \ldots, m_{i_k}) \subseteq S_{i_l}(t_k)^-$. That is, $\dim(S_{i_l}(t_k)^-) \geq k$ for $1 \leq l \leq k$. Thus, net dimension increase is $k(k-1)$ by time t_k. That is, $T_k \geq k(k-1)$. This completes the proof of Lemma 12. □

We note that, $T_n = n(n-1)$ and is the time when all nodes have received enough message to decode the original messages.

Probability of dimension increase. Consider at time t. Let there be nodes of $l \leq n$ types. Let these type classes be $C_1(t), \ldots, C_l(t)$. Now consider one of these l type classes, say $C_1(t)$. The probability of a node in $C_1(t)$ exchanging a code with a node not in $C_1(t)$ is given by

$$P(C_1(t)) \;=\; \frac{1}{n}\left(\sum_{i \in C_1(t); j \notin C_1(t)} P_{ij} + P_{ji}\right) \;=\; \frac{|C_1(t)|}{n}\Phi_P^u(C_1(t)). \tag{38}$$

For $t \in [t_k, t_{k+1})$, for $k \in \mathbf{Z}_+$, $|C_r(t)| \leq k$ for $1 \leq r \leq l$. Hence, (38) can be bounded below as,

$$P(C_1(t)) \;\geq\; \frac{|C_1(t)|}{n}\Phi_P^k. \tag{39}$$

The (39) is true for all $C_r(t), 1 \leq r \leq l$. Hence, we obtain that probability of a pair of nodes from different type sets exchange codes at time $t \in [T_k, T_{k+1})$ is given by

$$P^k \;\geq\; \sum_{r=1}^{l} P(C_r(t)) \;\geq\; \sum_{r=1}^{l} \frac{|C_r(t)|}{n}\Phi_P^k \;=\; \Phi_P^k. \tag{40}$$

Now, when nodes from different type exchange code, as noted before, with probability $1 - 1/q$ net dimension increase at least by 1. Thus, when in the time interval $[t_k, t_{k+1})$, the dimension increase by 1 can be upper bounded an independent Geometrical random variable with parameter $p_k \stackrel{\triangle}{=} \left(1 - \frac{1}{q}\right)\Phi_P^k$. When the net dimension increase is $n(n-1)$, all the nodes have received enough coded message. That is, the information spreading time T_{spr} can be stochastically upper bounded as $T_{\text{spr}} \leq \sum_{l=1}^{n(n-1)} G_l$. where G_l are independent Geometric random variables with parameter p_k when $l \in [T_k, T_{k+1})$. By definition, p_k is monotonically decreasing in k. Hence, the smaller the T_k values are, the worse the stochastic upper bound above on T_{spr} is. Using Lemma 12, the worst upper bound on T_{spr} is as follows:

$$T_{\text{spr}} \;\leq\; \sum_{k=1}^{n-1}\sum_{l=1}^{2k} G_l(k) \stackrel{\triangle}{=} \hat{T}, \tag{41}$$

where $G_l(k)$ are independent Geometrical random variables with parameter p_k. From (41), it is straightforward that

$$E[T_{\text{spr}}] \;\leq\; E[\hat{T}] \;=\; 2\left(1 - \frac{1}{q}\right)^{-1}\sum_{k=1}^{n-1}\frac{k}{\Phi_P^k}. \tag{42}$$

To obtain the bound with probability $1-\epsilon$, we use Corollary 5. Let $p^* = \min_{k=1}^{n-1} \log(1-p_k)^{-1} \geq \min_{k=1}^{n-1} p_k = \min_{k=1}^{n-1} \left(1-\frac{1}{q}\right) \Phi_P^k$. By definition, Φ_P^k is monotonically decreasing in k. Hence,

$$p^* \quad = \quad \left(1-\frac{1}{q}\right) \Phi_P^{n-1}. \tag{43}$$

Now, from Corollary 5, for $\lambda > 0$,

$$\Pr\left(\hat{T} > (1+\lambda)(E[\hat{T}]+3)\right) \quad \leq \quad \exp\left(-\frac{\lambda^2 p^* E[\hat{T}]}{32}\right). \tag{44}$$

The (44) suggests that for the choice of $\lambda = \sqrt{\frac{32\log\epsilon^{-1}}{p^* E[\hat{T}]}}$, we obtain

$$\Pr\left(\hat{T} > E[\hat{T}] + \sqrt{\frac{32\log\epsilon^{-1} E[\hat{T}]}{p^*}} + 3 + \sqrt{\frac{288\log\epsilon^{-1}}{p^* E[\hat{T}]}}\right) \quad \leq \quad \epsilon. \tag{45}$$

From (45) and $E[\hat{T}] = \Omega(n^2)$, we obtain that for large enough n,

$$\Pr\left(\hat{T} > E[\hat{T}] + 8\sqrt{\frac{\log\epsilon^{-1} E[\hat{T}]}{p^*}}\right) \quad \leq \quad \epsilon. \tag{46}$$

Now, (41)-(43) and (46) immediately imply the statement of Theorem 3. □

5 Applications

We study here the application of our preceding results to several types of graphs. In particular, we consider complete graphs, constant-degree expander graphs, and grid graphs. To obtain upper bounds on the time required to disseminate all the messages in the network to all the nodes, we study the communication matrix P that describes the natural random walk on each of these graphs. That is, the probability P_{ij} that node i contacts a node $j \neq i$ when i becomes active is $1/d_i$, where d_i is the degree of i. In addition, we apply the lower bound on the information dissemination time to the case of doubly stochastic communication matrices.

As a general tool, we use the following corollary of Lemma 6.

Corollary 13. *For any $\epsilon \in (0,1)$, $T_{spr}(\epsilon, P) = O\left(\frac{n\left(\log n + \sqrt{\log\epsilon^{-1}\log n}\right)}{\Phi_P^u}\right)$.*

Proof. From Lemma 6, we have the upper bound $\mu_n = n\left(\sum_{k=1}^{n-1} {F_P^k}^{-1}\right)$ on the expected number of rounds needed to transmit the message to all the nodes. Using the fact that $F_P^k = F_P^{n-k}$, we obtain $\mu_n \leq 2n\left(\sum_{k=1}^{\lfloor n/2 \rfloor} {F_P^k}^{-1}\right)$.

For any subset $S \subset V$ of vertices with $|S| = k \leq n/2$, we can bound the uniform ergodic flow across the cut (S, S^c) in terms of the uniform conductance Φ_P^u.

$$F_P(S,S^c) = \sum_{i\in S, j\in S^c} (P_{ij}+P_{ji}) = k \sum_{i\in S, j\in S^c} \frac{P_{ij}+P_{ji}}{|S|} = k\Phi_P^u(S) \geq k\Phi_P^u$$

This implies that $F_P^k \geq k\Phi_P^u$, and so $F^* = n^{-1} \min_{k=1}^{\lfloor n/2 \rfloor} F_P^k \geq \Phi_P^u/n$. The lower bound on F_P^k leads to an upper bound on μ_n.

$$\mu_n \leq 2n \left(\sum_{k=1}^{\lfloor n/2 \rfloor} F_P^{k^{-1}} \right) \leq 2n \left(\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k\Phi_P^u} \right) = \frac{2n}{\Phi_P^u} \left(\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k} \right) \leq \frac{2nH_{\lfloor n/2 \rfloor}}{\Phi_P^u}$$

We now obtain from Lemma 6 the following upper bound on the ϵ-spreading time of the information dissemination algorithm.

$$T_{\mathrm{spr}}(\epsilon, P) \leq \frac{2nH_{\lfloor n/2 \rfloor}}{\Phi_P^u} + 8\sqrt{\frac{2\log \epsilon^{-1} n^2 H_{\lfloor n/2 \rfloor}}{{\Phi_P^u}^2}} = \frac{2nH_{\lfloor n/2 \rfloor}}{\Phi_P^u} + \frac{8n}{\Phi_P^u}\sqrt{2\log \epsilon^{-1} H_{\lfloor n/2 \rfloor}}$$

As the harmonic number $H_{\lfloor n/2 \rfloor}$ satisfies $H_{\lfloor n/2 \rfloor} = \Theta(\log n)$, we have $T_{\mathrm{spr}}(\epsilon, P) = O(n(\log n + \sqrt{\log \epsilon^{-1} \log n})/\Phi_P^u)$. □

For $\epsilon = n^{-c}$, where c is a positive constant, Corollary 13 gives the upper bound $T_{\mathrm{spr}}(n^{-c}, P) = O((n \log n)/\Phi_P^u)$ on the ϵ-spreading time, and as a result every node receives every message in $O((n \log n)/\Phi_P^u)$ time with high probability.

5.1 Complete graph

On a complete graph, the natural random walk corresponds to the transition matrix P with $P_{ii} = 0$ for $i = 1, \ldots, n$, and $P_{ij} = 1/(n-1)$ for $j \neq i$. This regular structure allows us to directly evaluate the uniform conductance of any nonempty subset $S \subset V$ with $|S| = k \leq n/2$.

$$\Phi_P^u(S) = \frac{F_P(S, S^c)}{|S|} = \frac{\sum_{i \in S, j \in S^c} (P_{ij} + P_{ji})}{|S|} = \frac{|S||S^c| \left(\frac{2}{n-1}\right)}{|S|} = \frac{2|S^c|}{n-1} \geq \frac{2\left(\frac{n}{2}\right)}{n-1} = \frac{n}{n-1}$$

This implies that $\Phi_P^u \geq n/(n-1)$. Applying Corollary 13, we obtain $T_{\mathrm{spr}}(\epsilon, P) = O(n(\log n + \sqrt{\log \epsilon^{-1} \log n}))$, and $T_{\mathrm{spr}}(n^{-c}, P) = O(n \log n)$ when $c > 0$ is a constant. We conclude that every node receives every message in $O(n \log n)$ time with high probability, an upper bound that matches the results of [9].

5.2 Expander graph

Expander graphs have been used for numerous applications, and explicit constructions are known for constant-degree expanders [19]. We consider here an undirected graph in which the maximum degree of any vertex is d, where d is a constant. Suppose that the edge expansion of the graph is

$$\min_{\substack{S \subset V, \\ 1 \leq |S| \leq n/2}} \frac{|C(S, S^c)|}{|S|} = \alpha,$$

where $C(S, S^c)$ is the set of edges in the cut (S, S^c), and $\alpha > 0$ is a constant.

Since the degree of each node in the graph is bounded, the transition matrix P for the natural random walk on this expander satisfies $P_{ij} \geq 1/d$ for all $i \neq j$. For a nonempty subset $S \subset V$ of

vertices with $|S| = k \leq n/2$, the uniform conductance of S can be bounded from below in terms of α.

$$\Phi_P^u(S) = \frac{F_P(S, S^c)}{|S|} = \frac{\sum_{i \in S, j \in S^c}(P_{ij} + P_{ji})}{|S|} \geq \frac{2|C(S, S^c)|}{d|S|} \geq \frac{2\alpha}{d}$$

Thus, $\Phi_P^u \geq 2\alpha/d$. Corollary 13 now implies the same asymptotic upper bound as in the case of the complete graph, $T_{\text{spr}}(\epsilon, P) = O(n(\log n + \sqrt{\log \epsilon^{-1} \log n}))$. This suggests that the expansion properties of a constant-degree expander are sufficient to ensure that information can be disseminated in an expander as rapidly as in a complete graph, in an asymptotic sense. An interesting question for further study is whether expanders in which the degree is not constant, such as random graphs generated according to the preferential connectivity model [15], have ϵ-spreading times of the same asymptotic order.

5.3 Grid

We now consider a d-dimensional grid graph on n nodes, where $k = n^{1/d}$ is an integer. Each node in the grid can be represented as a d-dimensional vector $x = (x_i)$, where $x_i \in \{1, \ldots, k\}$ for $1 \leq i \leq d$. There is one node for each distinct vector of this type, and so the total number of nodes in the graph is $k^d = (n^{1/d})^d = n$. For any two nodes x and y, there is an edge (x, y) in the graph if and only if, for some $i \in \{1, \ldots, d\}$, $|x_i - y_i| = 1$, and $x_j = y_j$ for all $j \neq i$.

In [2], it is shown that the isoperimetric number for this grid graph is

$$\min_{\substack{S \subset V, \\ 1 \leq |S| \leq n/2}} \frac{|C(S, S^c)|}{|S|} = \Theta\left(\frac{1}{k}\right) = \Theta\left(\frac{1}{n^{1/d}}\right).$$

By the definition of the edge set, the degree of each node in the graph is at most $2d$. This gives a lower bound of $P_{ij} \geq 1/(2d)$ on the transition probability of the natural random walk for $i \neq j$.

As in the case of expander graphs, we obtain a lower bound on the uniform conductance of any nonempty subset of vertices $S \subset V$ in terms of the isoperimetric number.

$$\Phi_P^u(S) = \frac{F_P(S, S^c)}{|S|} = \frac{\sum_{i \in S, j \in S^c}(P_{ij} + P_{ji})}{|S|} \geq \frac{2|C(S, S^c)|}{2d|S|} = \Omega\left(\frac{1}{dn^{1/d}}\right)$$

This implies that $\Phi_P^u = \Omega(1/(dn^{1/d}))$. Applying Corollary 13, we obtain that, for the transition matrix P corresponding to the natural random walk on a d-dimensional grid graph with n nodes, $T_{\text{spr}}(\epsilon, P) = O(dn^{(1+1/d)}(\log n + \sqrt{\log \epsilon^{-1} \log n}))$.

5.4 Application of lower bound

We now consider applying the lower bound in Lemma 8 to a class of communication matrices P. Specifically, we obtain a lower bound on the ϵ-spreading time for a doubly stochastic matrix P.

First, we note that the vector $q(\cdot)$ defined in the proof of Lemma 8 satisfies $q(t) \leq A^t q(0)$ for $t \geq 1$, where the matrix A is defined as follows.

$$A = I + \frac{1}{n}(P + P^T)$$

This implies that $\|q(t)\| \leq \|A\|^t \|q(0)\| = \|A\|^t$, where $\|A\|$ denotes the spectral norm of the matrix A. Since P is doubly stochastic, $\|P\|, \|P^T\| \leq 1$, and so we can use the triangle inequality to obtain an upper bound on the norm of A.

$$\|A\| \leq \|I\| + \frac{1}{n}\|P + P^T\| \leq 1 + \frac{1}{n}(\|P\| + \|P^T\|) \leq 1 + \frac{2}{n}$$

By the definition of the ϵ-spreading time $T_{\text{spr}}(\epsilon, P)$ and the analysis in the proof of Lemma 8, for $t \geq T_{\text{spr}}(\epsilon, P)$ we must have $q(t) \geq (1-\epsilon)\vec{1}$, which implies that $\|q(t)\| \geq (1-\epsilon)\sqrt{n}$. On the other hand, substituting the upper bound above on $\|A\|$ into the upper bound on $\|q(t)\|$ yields

$$\|q(t)\| \leq \left(1 + \frac{2}{n}\right)^t \leq \exp\left(\frac{2t}{n}\right).$$

For $t < n\log(n(1-\epsilon)^2)/4$, then, we have $\|q(t)\| < (1-\epsilon)\sqrt{n}$. We conclude that $T_{\text{spr}}(\epsilon, P) = \Omega(n\log(n(1-\epsilon)^2))$ for all doubly stochastic matrices P.

6 Conclusion

In this paper, we considered the question of information dissemination via gossip algorithms. We found that the information dissemination time of the randomized gossip algorithms that we considered is strongly related to the isoperimetric properties of the probability matrix that describes the algorithm. This characterization led to the formulation of an optimal information dissemination algorithm as a solution to a concave optimization problem over a convex set.

We applied these results to two applications. First, we used these results to analyze an averaging algorithm based on a classic result of [7]. This allowed us to conclude that, in some cases, this averaging algorithm better by an order of magnitude better than other averaging algorithms considered recently [11, 4]. Second, we used a similar method to analyze a coding-based gossip algorithm for arbitrary graphs, and obtain a tight performance bound. This shows that coding-based gossip is not useful (nor harmful) for grid-type graphs. Finally, we evaluated our results in the context of various graphs of interest.

Acknowledgements

We thank Ashish Goel, Chandra Nair, and Balaji Prabhakar for useful discussions and suggestions.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.

[2] M. C. Azizoğlu and Ö. Eğecioğlu. The isoperimetric number of d-dimensional k-ary arrays. *International Journal of Foundations of Computer Science*, 10(3):289–300, 1999.

[3] M. Bawa, A. Gionis, H. Garcia-Molina, and R. Motwani. The price of validity in dynamic networks. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, pages 515–526, 2004.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Gossip algorithms: Design, analysis and applications. In *Proceedings of IEEE INFOCOM 2005*, 2005.

[5] J. Considine, F. Li, G. Kollios, and J. Byers. Approximate aggregation techniques for sensor databases. In *Proceedings of the 20th International Conference on Data Engineering*, pages 449–460, 2004.

[6] S. Deb and M. Médard. Algebraic gossip: A network coding approach to optimal multiple rumor mongering. In *Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

[7] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31(2):182–209, 1985.

[8] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *IEEE INFOCOM*, 2005.

[9] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized rumor spreading. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 565–574, 2000.

[10] D. Kempe and J. Kleinberg. Protocols and impossibility results for gossip-based communication mechanisms. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 471–480, 2002.

[11] D. Kempe, J. Kleinberg, and A. Demers. Spatial gossip and resource location protocols. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 163–172, 2001.

[12] R. Koetter and M. Médard. Beyond routing: An algebraic approach to network coding. In *Proceedings of IEEE INFOCOM 2002*, pages 122–130, 2002.

[13] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, 11(5):782–795, 2003.

[14] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49(2):371–381, 2003.

[15] M. Mihail, C. Papadimitriou, and A. Saberi. On certain connectivity properties of the Internet topology. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 28–35, 2003.

[16] S. Nath, P. B. Gibbons, S. Seshan, and Z. R. Anderson. Synopsis diffusion for robust aggregation in sensor networks. In *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, pages 250–262, 2004.

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

[18] R. Ravi. Rapid rumor ramification: Approximating the minimum broadcast time. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, pages 202–213, 1994.

[19] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 3–13, 2000.

[20] A. Sinclair. *Algorithms for Random Generation and Counting: A Markov Chain Approach*. Birkhäuser, Boston, 1993.

A Relation between $T_{\mathbf{spr}}$ and link-capacity

The analysis of $T_{\rm spr}$ assumes that when two nodes i and j communicate with each other during the course of algorithm, they can instantly exchange all information of each other. This requires link capacity of $\Theta(n)$ between node-pairs. However, capacity of link between node can be constrained. In such a situation, one can translate the above results in a straightforward manner as follows: let each link have capacity of C. Then, the information dissemination time, $T^c_{\rm spr}$, is upper bounded by the uncapacitated information dissemination time, $T_{\rm spr}$, as

$$T^c_{\rm spr} \leq O(n/c)T_{\rm spr}.$$