University of Groningen

Analysis of Accelerated Gossip Algorithms

Liu, J.; Anderson, B.D.O.; Cao, M.; Morse, A.S.

Link to publication in University of Groningen/UMCG research database

Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

WeAIn4.4

Analysis of Accelerated Gossip Algorithms

J. Liu, B. D. O. Anderson, M. Cao and A. S. Morse

Abstract— This paper investigates accelerated gossip algorithms for distributed computations in networks where shift-registers are utilized at each node. By using tools from matrix analysis, we prove the existence of the desired acceleration and establish the fastest rate of convergence in expectation for two-register symmetric gossip. Some classes of networks with regular graph topologies are studied in detail to validate the analytical results by comparison with existing empirical data. We also analyze convergence of second moment and provide a necessary condition for convergence in multi-register symmetric gossip. The proposed approach has the potential to be applied to the more challenging open problem of asymmetric gossip.

I. INTRODUCTION

While sensor networks have been utilized in a wide range of applications, a central theme of research that has remained the focus over the past decade is the design of efficient distributed computation algorithms, especially for the scenarios where sensors are constrained by limited sensing, computation and communication capabilities. In this context, much attention has been given to the distributed consensus problem in which all sensors are required to agree, using distributed averaging, on the same estimate of some variable of interest. The existing distributed algorithms that solve this problem include those known as gossip algorithms, which can be classified into probabilistic ones and deterministic ones, see for example [1] and [2] respectively. The authors of [3] propose a randomized gossip algorithm where at each time instant a single randomly chosen pair of nodes in a network update their values together to the mean of their current values. Such an algorithm is easy to implement and only requires simple computations at each node. However, the convergence rate of the algorithm is relatively slow, which is a critical drawback and needs to be improved.

A number of papers have studied this issue of slow convergence of gossip algorithms and various strategies have been studied to improve the convergence rate. In [4], the authors establish the necessary and sufficient conditions for the convergence of distributed linear iterations and propose computational methods for obtaining the fastest iteration. In

[5], a lower bound on the convergence rate for a class of network consensus algorithms is given through a graphical approach. A consensus propagation algorithm is proposed in [6] to achieve better scaling properties than pairwise averaging. The authors of [7] introduce simple and numerically stable algorithms with better worst-case performance. In [8], a probabilistic counting mechanism is utilized to improve the convergence rate and in [9] sensors' location information is exploited to boost the convergence. Among all the proposed acceleration strategies, there is one particular approach that has motivated the research in this paper. In [10] it has been demonstrated through simulations that by installing shift-registers to sensors and thus enabling the utilization of computational results in each sensor's finite memory, substantial acceleration (up to 10 folds) can be achieved for the stochastic gossip algorithms. However, the theoretical explanation for the observed improvement is not complete in [10]. Follow-up works, e.g. [11], have tried to establish the necessary and sufficient conditions for convergence, but no one has been able to describe the accelerated convergence rate of gossip algorithms with shift-registers in a rigorous and precise fashion.

The main contribution of this paper is that we are able to thoroughly analyze the two-register gossip algorithm, first proposed in [10], under the symmetric assumption. The fastest rate of convergence in expectation is provided in closed-form, which depends on the given probabilistic strategy according to which sensors are activated to gossip together. In section II, we review the original gossip algorithm studied in [3] and the accelerated gossip algorithms introduced in [10]. In section III, we first establish a complete analysis of convergence in expectation for two-register symmetric gossip; then we validate the theoretical results by exploring some specific regular graphs and using the experimental results in [10]; finally we study second moment convergence and establish a necessary condition of speeding up convergence for general multi-register symmetric gossip.

II. GOSSIP ALGORITHM

We use a graph $\mathbb{G}$ to describe the topology of a given sensor network with n nodes. Let each node in the network be a vertex of $\mathbb{G}$ and each link be an edge. Let $\mathcal{N} = \{1, \ldots, n\}$ denote the vertex set of $\mathbb{G}$ and $\mathcal{E}$ be the edge set. Let $x_i(0)$ denote the initial value at node i and $x_{\text{ave}} = \frac{1}{n} \sum_i x_i(0)$ be the average of all the initial values. The goal of gossip is to compute x_{ave} via a distributed algorithm. In this paper, we use the synchronous model; in other words, all nodes act according to the same time sequence $\{1, 2, \ldots\}$. For each $i \in \{1, \ldots, n\}$, let $x_i(t)$ denote the value of node i at time t.

At any time $t \in \{1, 2, ...\}$, there is one and only one node to be activated; each node has an equal probability $\frac{1}{n}$ of being activated. If node i is activated, then with probability p_{ij} node i chooses node j to update their values together, where $p_{ij} > 0$ only if $i \neq j$ and $(i, j) \in \mathcal{E}$; the values of all the remaining nodes remain the same. Let P be an $n \times n$ stochastic matrix whose ijth entry is p_{ij}. For clarity of expression, we assume that P has 1 as a simple eigenvalue and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. A gossip problem is called *symmetric* if for all i and j, there holds $p_{ij} = p_{ji}$.

Consider any probabilistic algorithm characterized by matrix P, the set of node value update rules can be written in state form. Toward this end, we use the symbol $\mathcal{Q}$ to denote a suitably defined set, indexing the class of all pairs of nodes defined on n vertices: $\mathcal{Q} = \{(i, j) : i \neq j, \ i, j = 1, ..., n\}$. For each pair of nodes $(i, j) \in \mathcal{Q}$, let $A_{(i,j)}$ be the random matrix that describes the update rule when nodes i and j are activated, then

$$x(t + 1) = A_{\sigma(t)} x(t) \tag{1}$$

where x is the state vector $x = [x_1 \ x_2 \ \cdots \ x_n]'$ and $\sigma : \{0, 1, ...\} \rightarrow \mathcal{Q}$ is a switching signal whose value at time t, is the index representing the randomly chosen pair of nodes at time t. For simplicity, we will adopt the notation A_{ij} instead of $A_{(i,j)}$ in the sequel. Letting the mean of the (*i.i.d.*) matrices $A_{\sigma(t)}$ be denoted by $\bar{A}$, we have $E[x(t)] = \bar{A}^t x(0)$.

We first review the original gossip algorithm studied in [3]. At time t, let node i be activated with probability $\frac{1}{n}$ and suppose it chooses some neighboring node j, with probability p_{ij}, to update their values together to the average of their current values, then $A_{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - e_j)'$ where e_i is the $n \times 1$ unit vector with the ith component equal to 1 and I is the n-dimensional identity matrix. Obviously the A_{ij}'s are doubly stochastic matrices, so therefore is $\bar{A}$. It has been proved in [3] that the system (1) converges in expectation to the average value x_{ave} if and only if $\bar{A}$ has 1 as a simple eigenvalue and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. It is also pointed out in [4] that the spectral radius of $\bar{A} - \mathbf{1}\mathbf{1}'/n$, denoted by $\rho(\bar{A} - \mathbf{1}\mathbf{1}'/n)$, is equal to the *asymptotic convergence factor*, where $\mathbf{1}$ denotes the n-dimensional all ones column vector and the asymptotic convergence factor is defined by

$$r(\bar{A}) = \sup_{x(0) \neq x_{\text{ave}}\mathbf{1}} \lim_{t \to \infty} \left(\frac{\|x(t) - x_{\text{ave}}\mathbf{1}\|_2}{\|x(0) - x_{\text{ave}}\mathbf{1}\|_2} \right)^{1/t}$$

The conclusion has been generalized to any expectation matrix, which we will discuss later. Therefore, for any randomized gossip algorithm, each node's value converges in expectation to the same value (which may not be x_{ave}) if and only if its expectation matrix has 1 as a simple eigenvalue and all the remaining eigenvalues strictly less than 1 in magnitude; the rate of convergence in expectation is governed by the second largest eigenvalue in magnitude. Furthermore, the authors of [3] establish a sufficient condition for second moment convergence, which is that $E[A_{ij}' A_{ij}]$ has all but one eigenvalue less than 1; the rate of convergence in mean

square is governed by $\bar{A}$'s second largest eigenvalue, which is nonnegative since $\bar{A}$ is positive semidefinite.

Cao *et al.* [10] introduce a technique which uses memory in the form of shift-registers to accelerate the original gossip algorithm. Each node has the same number of registers, the first of which stores the sensor's current value, and the remainder of which store earlier values corresponding to the sensor. The algorithm changes how a pair of nodes update their values once they decide to update together. For each node $i \in \{1, ..., n\}$, let x_{ir} denote the value stored in its rth register. In the case where each sensor is provided with two registers, the algorithm can be described as follows. Let nodes i and j be the pair of nodes updating their current values together at time t, then

$$\begin{cases} x_{i1}(t + 1) = \omega(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t)) + (1 - \omega)x_{i2}(t) \\ x_{i2}(t + 1) = x_{i1}(t) \\ x_{j1}(t + 1) = \omega(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t)) + (1 - \omega)x_{j2}(t) \\ x_{j2}(t + 1) = x_{j1}(t) \end{cases} \tag{2}$$

where $1 \leq \omega < 2$ is a constant; the values of the registers of all the other nodes remain the same. In the case where each sensor is provided with $m > 2$ registers, the accelerated gossip algorithm at each node, when nodes i and j are the pair to update their current values together, is generalized as follows:

$$\begin{cases} x_{i1}(t + 1) = \omega_1(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t)) + \sum_{r=2}^{m} \omega_r x_{ir}(t) \\ x_{ir}(t + 1) = x_{i(r-1)}(t), \qquad r = 2, ..., m \\ x_{j1}(t + 1) = \omega_1(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t)) + \sum_{r=2}^{m} \omega_r x_{jr}(t) \\ x_{jr}(t + 1) = x_{j(r-1)}(t), \qquad r = 2, ..., m \end{cases} \tag{3}$$

where ω_r $(r = 1, ..., m)$ are constants satisfying $\sum_{r=1}^{m} \omega_r = 1$; the values of the registers of all the other nodes remain the same. The experiments in [10] use random geometric graphs with four different configurations of shift-registers. The first three, denoted by D2, D4 and D8, consist of 2, 4 and 8 registers respectively for which only the first and last register are used to compute the new value of the first register (*i.e.*, $\omega_r = 0$ when $r \neq 1, m$). For the fourth one, denoted by X4, all of the registers are used to compute the new value of the first register. The results of the experiments show that by employing more registers and choosing a clever combination of coefficients, the algorithms can improve the convergence rate substantially.

III. ANALYSIS

In this section, we provide a theoretical analysis for the accelerated algorithms so that we can find the fastest rate of convergence in expectation and the corresponding optimal coefficients for symmetric gossip. We also establish convergence in mean square. First we consider the case where each node is provided with two registers.

Boyd *et al.* [3] write the expectation matrix of the original algorithm as $\bar{A} = I - \frac{1}{2n}D + \frac{1}{2n}(P + P')$ where D is a diagonal matrix with entries $d_{ii} = 1 + \sum_{j=1}^{n} p_{ji}$. Notice that the symmetry of $\bar{A}$ does not depend on P being symmetric. However, for symmetric P, it can be easier to calculate certain quantities, for example $D = 2I$ and $\bar{A} = (1 - \frac{1}{n})I + \frac{1}{n}P$. Now we consider the changes that arise with the accelerated algorithm. The probabilities stay the same, but the new update equations are (2). Define the enlarged state vector as $z = [x_{11} \cdots x_{n1} x_{12} \cdots x_{n2}]'$. The matrices corresponding to A_{ij} and $\bar{A}$ of dimension $n \times n$ are then replaced by matrices of dimension $2n \times 2n$. Denote the enlarged matrices by B_{ij} and $\bar{B}$, we have

$$B_{ij} = \left[\begin{array}{c} \sum_{k \neq i,j} e_k e_k' + \frac{\omega}{2}(e_i + e_j)(e_i + e_j)' \\ e_i e_i' + e_j e_j' \\ \\ (1 - \omega)(e_i e_i' + e_j e_j') \\ \sum_{k \neq i,j} e_k e_k' \end{array} \right]$$

$$\bar{B} = \left[\begin{array}{cc} I + \frac{\omega}{2}\frac{P+P'}{n} + (\frac{\omega}{2} - 1)\frac{D}{n} & (1 - \omega)\frac{D}{n} \\ \frac{D}{n} & I - \frac{D}{n} \end{array} \right]$$

For arbitrary ω, it is not hard to check that the row sums of B_{ij} and $\bar{B}$ are all equal to 1; however, not all entries are nonnegative because of the upper right block.

A. Convergence in Expectation

In this subsection, we prove the existence of the desired convergence in expectation and study the behavior of $\rho_2(\bar{B})$, the second largest magnitude of any eigenvalue of $\bar{B}$. As a first step, we need the following fact.

Lemma 1: Suppose A is an $n \times n$ matrix, and let B be a $2n \times 2n$ matrix given by

$$B = \left[\begin{array}{cc} A & aI \\ bI & cI \end{array} \right]$$

where (to avoid trivialities) $ab \neq 0$. Consider the eigenvalue equation $B[\alpha_1 \ \alpha_2]' = \lambda[\alpha_1 \ \alpha_2]'$ with α_1 and α_2 not both zero, then $A\alpha_1 = \mu\alpha_1$ where α_1 is necessarily nonzero and

$$\mu = \lambda - \frac{ab}{\lambda - c} \tag{4}$$

Conversely, given μ and $\alpha_1 \neq 0$ satisfying $A\alpha_1 = \mu\alpha_1$, and with λ_i ($i = 1, 2$), the two solutions of (4), and $\alpha_{2i} = b\alpha_1/(\lambda_i - c)$, there holds $B[\alpha_1 \ \alpha_{2i}]' = \lambda_i[\alpha_1 \ \alpha_{2i}]'$.

The simple proof is omitted. For symmetric P, there holds

$$\bar{B} = \left[\begin{array}{cc} (1 + \frac{\omega-2}{n})I + \frac{\omega}{n}P & \frac{2(1-\omega)}{n}I \\ \frac{2}{n}I & (1 - \frac{2}{n})I \end{array} \right]$$

Let $A = (1 + \frac{\omega-2}{n})I + \frac{\omega}{n}P$, $a = \frac{2(1-\omega)}{n}$, $b = \frac{2}{n}$, and $c = 1 - \frac{2}{n}$. Lemma 1 implies that the $2n$ eigenvalues of $\bar{B}$ are determined by the n eigenvalues of A with the relation (4). If we denote the n eigenvalues of P as $1 = \lambda_1(P) > \lambda_2(P) \geq \lambda_3(P) \geq \cdots \geq \lambda_n(P) > -1$, the n eigenvalues of A can also be written as a non-increasing sequence with the values $\mu_i(A) = 1 + \frac{\omega-2}{n} + \frac{\omega}{n}\lambda_i(P)$. Thus, the $2n$ eigenvalues of $\bar{B}$ are determined by the n eigenvalues of P. For each $\mu_i(A)$, we can obtain two eigenvalues of $\bar{B}$, denoted by λ_{i1} and λ_{i2}.

We assume that $\lambda_{i1} \geq \lambda_{i2}$ if they are both real. When $\mu_i(A)$ satisfies the condition that equation (4) has two real roots, we get

$$\lambda_{i1,2} = 1 + \frac{1}{n}\left[\frac{1}{2}\omega(1 + \lambda_i(P)) - 2 \right]$$
$$\pm \frac{1}{2n}\sqrt{\omega^2(1 + \lambda_i(P))^2 - 16(\omega - 1)} \tag{5}$$

First, we notice that $d\lambda_{i1}/d\lambda_i(P) > 0$, which implies that the real eigenvalues λ_{i1} (i belongs to a subset of $\{1, 2, ..., n\}$) form a non-increasing sequence. Second, from (5) we can get $\lambda_{i2} \geq 0$, which indicates that all the real eigenvalues of $\bar{B}$ are nonnegative. Third, λ_{11} and λ_{12} are always real when $1 < \omega < 2$; $\lambda_{11} = 1$ is the largest real eigenvalue of $\bar{B}$. Furthermore, if λ_{21} and λ_{22} are real as well, the second largest real eigenvalue of $\bar{B}$ is the larger one between λ_{12} and λ_{21}. On the other hand, if for some $\mu_i(A)$, the roots of equation (4) are complex, the magnitude of the two eigenvalues are the same,

$$|\lambda_{i1,2}| = |\lambda_{i1}| = |\lambda_{i2}|$$
$$= \sqrt{1 - \frac{4}{n} + \frac{\omega}{n}(1 + \lambda_i(P)) + \frac{2\omega}{n^2}(1 - \lambda_i(P))} \tag{6}$$

Lemma 2: When $1 < \omega < 2$, $\bar{B}$ has 1 as a simple eigenvalue and all the remaining $2n - 1$ eigenvalues are strictly less than 1 in magnitude.

Proof: For any complex eigenvalue, $d|\lambda_{i1,2}|/d\omega > 0$ and $d|\lambda_{i1,2}|/d\lambda_i(P) \geq 0$, which indicates that $|\lambda_{i1,2}|$ increases as either ω or $\lambda_i(P)$ increases, then an upper bound for $|\lambda_{i1,2}|$ is given from (6) with $\omega = 2$ and $\lambda_i(P) = 1$, $|\lambda_{i1,2}| \leq (1 - \frac{2}{n})^{1/2} < 1$. Recall that all the real eigenvalues of $\bar{B}$ are nonnegative, and 1 is the largest one. In addition, the second largest real eigenvalue is always smaller than 1 when $1 < \omega < 2$. Therefore, 1 is a simple eigenvalue of $\bar{B}$. ∎

Lemma 3: Given an $n \times n$ matrix M and vectors $c, d \in \mathbb{R}^n$, the equation $\lim_{t \to \infty} M^t = dc'/c'd$ holds if and only if $c'M = c'$, $Md = d$, and $\rho(M - dc'/(c'd)) < 1$, where $\rho(\cdot)$ denotes the spectral radius of a matrix. Moreover, $\rho(M - dc'/(c'd))$ is equal to the asymptotic convergence factor.

This is Theorem 2 of [4]. Note that $\rho(M - dc'/(c'd)) = \rho_2(M)$ and $\rho(M - dc'/(c'd)) < 1$ implies that 1 is a simple eigenvalue of M.

Theorem 1: Given the sensor network with the state evolution equation $z(t+1) = B_{\sigma(t)}z(t)$ using the symmetric accelerated gossip algorithm (2), each node's value will converge in expectation to the desired value $z_{\text{ave}} = \frac{1}{n}\sum_i x_{i1}(0)$ if each node's two registers are initialized so that $x_{i1}(0) = x_{i2}(0)$ for all $i = 1, ..., n$.

Proof: It can be shown that

$$B_{ij}\left[\begin{array}{c} \mathbf{1} \\ \mathbf{1} \end{array} \right] = \left[\begin{array}{c} \mathbf{1} \\ \mathbf{1} \end{array} \right] \tag{7}$$

and (Lemma 1 in [10])

$$\left[\begin{array}{cc} \frac{1}{2-\omega}\mathbf{1}' & \frac{1-\omega}{2-\omega}\mathbf{1}' \end{array} \right] B_{ij} = \left[\begin{array}{cc} \frac{1}{2-\omega}\mathbf{1}' & \frac{1-\omega}{2-\omega}\mathbf{1}' \end{array} \right] \tag{8}$$

Because $\bar{B} = E[B_{\sigma(t)}] = \sum_{ij} \frac{1}{n}p_{ij}B_{ij}$ is a convex combination of B_{ij}, equations (7) and (8) also hold for $\bar{B}$.
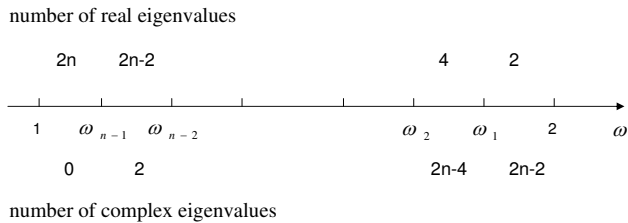
Fig. 1. Distribution of $\bar{B}$'s real and complex eigenvalues

Fig. 2. Overview of $\rho_2(\bar{B}(\omega))$

We have proved in Lemma 2 that under the symmetric gossip constraint, $\bar{B}$ has all but one eigenvalue less than 1 in magnitude. According to Lemma 3, it can be easily checked that, if $x_{i1}(0) = x_{i2}(0)$, the desired convergence in expectation is achieved. ∎

In the sequel we will write the second largest magnitude of any eigenvalue of $\bar{B}$ as a function of ω, $\rho_2(\bar{B}(\omega))$, and then find the optimal point of $\omega \in (1, 2)$, denoted by ω^*, which minimizes the function.

First we explore the distribution of the number of real and complex eigenvalues of $\bar{B}$. From (5) we know that whether λ_{i1} and λ_{i2} are real or not depends on the sign of a second order polynomial $f(\omega) = (1 + \lambda_i(P))^2\omega^2 - 16\omega + 16$. Since $f(1) \geq 0$ and $f(2) \leq 0$, $f(\omega)$ has a unique zero point between 1 and 2, which is

$$\omega(\lambda_i(P)) = \frac{8 - 4\sqrt{4 - (1 + \lambda_i(P))^2}}{(1 + \lambda_i(P))^2}$$

Therefore, for each $\lambda_i(P)$, if $1 < \omega \leq \omega(\lambda_i(P))$, λ_{i1} and λ_{i2} are real; if $\omega(\lambda_i(P)) < \omega < 2$, λ_{i1} and λ_{i2} are complex. In addition, $d\omega(\lambda_i(P))/d\lambda_i(P) > 0$, which indicates that $\omega(\lambda_i(P))$ is an increasing function of $\lambda_i(P)$; in particular, $\omega(1) = 2$. If we denote $\omega(\lambda_i(P))$ as ω_{i-1}, $i = 2, ..., n$, we get $1 < \omega_{n-1} \leq \omega_{n-2} \leq \cdots \leq \omega_1 < 2$. If $1 < \omega \leq \omega_{n-1}$, $\bar{B}$ has $2n$ real eigenvalues; $\cdots$; if $\omega_1 < \omega < 2$, $\bar{B}$ has 2 real eigenvalues and $2n - 2$ complex eigenvalues. See Figure 1.

Lemma 4: When $1 < \omega \leq \omega_1$, $\rho_2(\bar{B}(\omega)) = \lambda_{21}(\omega)$.

The proof of Lemma 4 can be found in the full length version of this paper.

When $\omega_1 < \omega < 2$, $\bar{B}$ only has two real eigenvalues, 1 and λ_{12}. The largest magnitude of any complex eigenvalue is determined by $\lambda_2(P)$. So the second largest magnitude of any eigenvalue of $\bar{B}$ is the larger one between λ_{12} and $|\lambda_{21}|$. It can be checked that $\lambda_{12}(\omega) \geq |\lambda_{21}(\omega)|$ if and only if

$$g(\omega) = 4\omega^2 + (3n - \lambda_2(P)n + 2\lambda_2(P) - 18)\omega + 16 - 4n \geq 0$$

Since $g(\omega_1) < 0$ and $g(2) > 0$, $g(\omega)$ has a unique zero point between ω_1 and 2, denoted by ω_0. Therefore, when $\omega_1 < \omega < \omega_0$, $|\lambda_{21}| > \lambda_{12}$; when $\omega_0 \leq \omega < 2$, $\lambda_{12} \geq |\lambda_{21}|$. This leads to the following lemma.

Lemma 5: When $\omega_1 < \omega < \omega_0$, $\rho_2(\bar{B}(\omega)) = |\lambda_{21}(\omega)|$; when $\omega_0 \leq \omega < 2$, $\rho_2(\bar{B}(\omega)) = \lambda_{12}(\omega)$.

Combining Lemmas 4 and 5, we see that $\rho_2(\bar{B}(\omega))$ is continuous on the interval $(1, 2)$, and corresponds to a real eigenvalue for the intervals $(1, \omega_1]$ and $[\omega_0, 2)$. Figure 2 provides an overview of $\rho_2(\bar{B}(\omega))$. The dashed curve
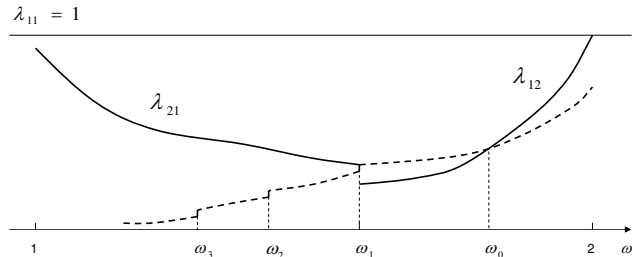
denotes the largest magnitude of any complex eigenvalue; it is discontinuous at $\omega_1, \omega_2, ...$ because $\{\lambda_i(P)\}$ is a discrete set.

Theorem 2: The minimum of $\rho_2(\bar{B}(\omega))$ on the interval $(1, 2)$ is unique, and the value of the optimal point of ω at this minimum is

$$\omega^* = \omega_1 = \frac{8 - 4\sqrt{4 - (1 + \lambda_2(P))^2}}{(1 + \lambda_2(P))^2} \quad (9)$$

Proof: When $1 < \omega \leq \omega_1$, λ_{21} is a real eigenvalue. Because of Lemma 4 and the fact that $\lambda_{21}(\omega)$ is a decreasing function of ω, $\lambda_{21}(\omega_1)$ gives a lower bound of $\rho_2(\bar{B}(\omega))$ when $1 < \omega \leq \omega_1$. When $\omega_1 < \omega < 2$, λ_{21} is a complex eigenvalue. Because $|\lambda_{21}(\omega)|$ is an increasing function of ω, $|\lambda_{21}(\omega_1)|$ gives a lower bound of $\rho_2(\bar{B}(\omega))$ when $\omega_1 < \omega < 2$. In addition, $\rho_2(\bar{B}(\omega))$ is continuous at ω_1. Therefore ω_1 is the unique point which minimizes $\rho_2(\bar{B}(\omega))$. ∎

Notice that $\rho_2(\bar{A}) = \lambda_2(\bar{A})$ since $\bar{A}$ is symmetric positive-semidefinite. Since it can be checked that $\rho_2(\bar{B}(\omega_1)) < \rho_2(\bar{A})$, we get the following result.

Corollary: Under the symmetric constraint, the accelerated gossip algorithm utilizing two shift-registers at each node has faster rate of convergence in expectation than the original gossip algorithm.

Here we define the *acceleration ratio* as

$$\eta = \frac{\log \rho_2(\bar{B}(\omega_1))}{\log \rho_2(\bar{A})}$$

to measure the speedup of the accelerated algorithm in symmetric gossip. For large n,

$$\eta \approx \frac{2\lambda_2(P) - 2 + 2\sqrt{4 - (1 + \lambda_2(P))^2}}{1 - \lambda_2(P)^2}$$

then η becomes a function only depending on $\lambda_2(P)$. It can be checked that $\eta > 1$ when $-1 < \lambda_2(P) < 1$.

B. Experimental Validation

According to (9), ω^* depends on the second largest eigenvalue of P. We are now interested in how the value of ω^* varies with different symmetric P. We assume that each node, if activated, communicates with its neighbors with equal probability, which is the assumption also made in the experiments in [10]. Imposing these requirements means that we are exploring in more detail what happens with regular graphs. For each integer $d \in \{2, 3, ..., n-1\}$, $A = dP$ is the

adjacency matrix of a d-degree regular graph, and we have $\lambda_i(P) = \mu_i(A)/d$. We consider two special cases of regular graphs: the complete graph K_n with $d = n-1$ and the cycle C_n with $d = 2$. First, the spectrum of K_n consists of $n-1$ with multiplicity one and -1 with multiplicity $n-1$. Then $\lambda_2(P) = -1/(n-1)$, which reaches the minimum of $\lambda_2(P)$. For $n \geq 10000$, $\lambda_2(P) \simeq 0$ and $\omega^* \simeq 1.0718$. Second, the spectrum of C_n consists of the numbers $2\cos(2\pi i/n)$, $i = 1,...,n$ [12]. Then $\lambda_2(P) = \cos\frac{2\pi}{n}$. For $n \geq 10000$, $\lambda_2(P) \simeq 1$ and $\omega^* \to 2$.

The authors of [13] proved that for any fixed d and for any infinite family of d-regular graphs G_i, $\liminf \mu_2(G_i) \geq 2\sqrt{d-1}$, and conjectured that almost all d-regular graphs G on n vertices satisfy $\mu_2(G) \leq 2\sqrt{d-1}+o(1)$ as n tends to infinity, which has been proved by Friedman [14]. We take $2\sqrt{d-1}$ as an estimate of $\mu_2(A)$, then $\lambda_2(P) \simeq 2\sqrt{d-1}/d$. Now we consider a different class of regular graphs, two-dimensional grid graphs. Grid graphs can be seen as a crude approximate of the topology of a sensor network (a random geometric graph) when sensors are uniformly distributed and sensors have the same sensing radius. If we ignore the boundary effect and substitute $d = 4$ into the equation, we get $\lambda_2(P) = 0.866$ and then $\omega^* = 1.47$, which agrees with the experimental result that ω^* should lie between 1.4 and 1.5, see TABLE II in [10].

While [10] contains comprehensive experimental results, the analysis developed in this paper has not yet reached the point where a full set of comparisons can be made. This is because the tools developed in this paper are tailored for symmetric P's while most of the graphs considered in [10] are random geometric graphs, and rarely have a symmetric P. This explains why the analytical values of ω^* in the cases of K_n and C_n are quite different from the experimental results. Even though we conjecture that the ω^* of non-regular graphs is almost the same as that of d-regular graphs if $\bar{d} = d$, where $\bar{d}$ denotes the average degree, TABLE II in [10] indicates that $\bar{d} \ll n-1$ and probably $\bar{d}$ is obviously larger than 2. When $d = 15$ and $d = 45$ (which can be viewed as loose estimations of $\bar{d}$ in the experiments), ω^* equals to 1.2038 and 1.1364 respectively.

C. Convergence of Second Moment

In this subsection, we investigate the convergence in mean square, which in some ways is more important than convergence in expectation. Let $y(t) = z(t) - z_{\text{ave}}\mathbf{1}$ and consider its evolution,

$$y(t+1) = B_{\sigma(t)}z(t) - z_{\text{ave}}B_{\sigma(t)}\mathbf{1} = B_{\sigma(t)}y(t)$$

The first equation holds because of the fact that $\mathbf{1}$ is an eigenvector for all $B_{\sigma(t)}$. Thus, y evolves according to the same linear system as z. Then we can get

$$E[y(t+1)'y(t+1)|y(t)] = y(t)'E[B'_{\sigma(t)}B_{\sigma(t)}]y(t)$$

The matrices $B_{\sigma(t)}$ are identically distributed, we shorten $B_{\sigma(t)}$ to B for convenience. Since $B^T B$ is symmetric positive-semidefinite,

$$y(t)'E[B'B]y(t) \leq \lambda_1(E[B'B])\|y(t)\|^2 \qquad (10)$$

where $\lambda_1(E[B'B])$ is the largest eigenvalue of $E[B'B]$. Repeatedly using (10), we obtain the bound

$$E[y(t)'y(t)] \leq \lambda_1^t(E[B'B])\|y(0)\|^2 \qquad (11)$$

If $y(t)$ is constrained to be orthogonal to the eigenvector of $E[B'B]$ corresponding to the maximum eigenvalue, then in (11) we have λ_2 (the second largest eigenvalue) rather than λ_1. Therefore, we get a sufficient condition for the convergence of second moment. However, it can be easily checked that when $\omega = 1$ and $P = P'$, $\lambda_1(E[B'B])$ is greater than 1; the same property holds for ω near to 1 by continuity. Also, we cannot always have $\lambda_2(E[B'B])$ strictly less 1. In the sequel, we will present a new approach to investigate the second moment convergence.

Suppose that N is a nonsingular matrix; observe that if Nz and Ny converge, then so do z and y, and conversely. Thus, we can study Nz and Ny rather than z and y. Consider the new system involving N, the evolution of $\hat{y} = Ny$ becomes

$$\hat{y}(t+1) = N^{-1}BN\hat{y}(t)$$

Thus B is replaced by $N^{-1}BN$, and importantly, $B'B$ is replaced by $N'B'N'^{-1}N^{-1}BN$. Let $C = N^{-1}BN$, the eigenvalue properties of C are the same as those of B; but the eigenvalue properties of $C'C$ are not the same as those of $B'B$. We are interested in the eigenvalues of $E[C'C]$, and in particular its second largest eigenvalue. If we select N as a diagonal block matrix,

$$N = \left[\begin{array}{cc} \alpha I & 0 \\ 0 & \alpha^{-1}I \end{array} \right]$$

where $\alpha = \sqrt[4]{\omega - 1}$, then

$$C = \left[\begin{array}{c} \sum_{k \neq i,j} e_k e_k' + \frac{\omega}{2}(e_i + e_j)(e_i + e_j)' \\ \sqrt{\omega - 1}(e_i e_i' + e_j e_j') \\ \\ -\sqrt{\omega - 1}(e_i e_i' + e_j e_j') \\ \sum_{k \neq i,j} e_k e_k' \end{array} \right]$$

The determinant of N is 1. It leaves the two diagonal block entries of B the same; it causes the two block off-diagonal entries of B to be equal and opposite in sign. Then we can get the expression for $E[C'C]$ involving P and D, in particular, when $P = P'$, $E[C'C]$ equals

$$\left[\begin{array}{cc} (1 - \frac{4}{n} + \frac{2\omega}{n} + \frac{\omega^2}{n})I + \frac{\omega^2}{n}P & -\frac{\omega\sqrt{\omega-1}}{n}(P + I) \\ -\frac{\omega\sqrt{\omega-1}}{n}(P + I) & (1 - \frac{4}{n} + \frac{2\omega}{n})I \end{array} \right]$$

It is immediate that when $\omega = 1$, the largest eigenvalue of $E[C'C]$ equals to 1. In addition, its second largest eigenvalue is strictly less than 1; it follows by continuity that for ω near to 1, the same property will hold. Therefore, we prove the convergence of second moment for ω close to 1.

D. Multi-Register Symmetric Gossip

In order to achieve faster convergence, we analyze general multi-register symmetric gossip. We take X4 of [10] as an example, in which all of the 4 registers are used to compute the new value of the first register. We define the state vector as a $4n$-dimensional column vector ordering the

875

entries in the same way as we did in the two-register case. The matrices corresponding to A_{ij} and $\bar{A}$ are now replaced by $4n$-dimensional D_{ij} and $\bar{D}$. Using a similar analysis as presented earlier, we obtain

$$\bar{D} = \begin{bmatrix} A_1 & \frac{2\omega_2}{n}I & \frac{2\omega_3}{n}I & \frac{2\omega_4}{n}I \\ bI & cI & 0 & 0 \\ 0 & bI & cI & 0 \\ 0 & 0 & bI & cI \end{bmatrix}$$

where $A_1 = (1 + \frac{\omega_1 - 2}{n})I + \frac{\omega_1}{n}P$, $b = \frac{2}{n}$, and $c = 1 - \frac{2}{n}$. A_1 is symmetric because P is. The $4n$ eigenvalues of $\bar{D}$ are determined by the n eigenvalues of A_1 through the equation

$$\mu = \lambda - \frac{2\omega_2}{n}\frac{b}{\lambda - c} - \frac{2\omega_3}{n}\frac{b^2}{(\lambda - c)^2} - \frac{2\omega_4}{n}\frac{b^3}{(\lambda - c)^3}$$

where λ denotes the eigenvalue of $\bar{D}$ and μ denotes the corresponding eigenvalue of A_1. It is easy to extend the results to the general multi-register symmetric gossip problem. That is, for the m-register accelerated algorithm (3) with $P = P'$, the mn eigenvalues of $mn \times mn$ matrix $\bar{D}$ are determined by the n eigenvalues of $n \times n$ matrix A_1 through the equation

$$\mu = \lambda - \sum_{r=2}^{m} \frac{2\omega_r}{n}\left(\frac{b}{\lambda - c}\right)^{r-1} \tag{12}$$

while the eigenvalues of A_1 are determined by the eigenvalues of P through the relation

$$\mu_i(A_1) = 1 + \frac{\omega_1 - 2}{n} + \frac{\omega_1}{n}\lambda_i(P), \quad i = 1, ..., 2n$$

In particular, $\lambda_1(P) = 1$ and $\mu_1(A_1) = 1 + 2(\omega_1 - 1)/n$. Although it becomes much harder to get an explicit expression for λ from equation (12) when $m > 2$, we can establish a necessary condition of the combination for ω_r's to ensure the convergence in expectation. To obtain a contradiction, suppose $\lambda = 1 + \epsilon$, where $\epsilon \geq 0$, the righthand side of (12) can be then expressed as a function of ϵ

$$h(\epsilon) = 1 + \epsilon - \sum_{r=2}^{m} \frac{2^r \omega_r}{n(2 + n\epsilon)^{r-1}}$$

Notice that $h(0) = \mu_1(A_1)$, which is the largest eigenvalue of A_1. Because when $\epsilon > 0$, $\lambda = 1 + \epsilon$ cannot be an eigenvalue of $\bar{D}$, we must have $h'(0) \geq 0$. Therefore, we get a necessary condition for ω_r's,

$$1 + \sum_{r=2}^{m} \omega_r(r - 1) \geq 0 \tag{13}$$

If only the first and last register are used to compute the new value of the first register, $\omega_r = 0$, $r = 2, ..., m - 1$, and $\omega_m = 1 - \omega_1$. Then the necessary condition (13) becomes

$$\omega_1 \leq 1 + \frac{1}{m - 1} \tag{14}$$

It can be checked that the simulation result of X4 satisfies (13) and the experimental results of D2, D4, D8 all satisfy (14). See TABLE II in [10]. In addition, condition (14) indicates that when only the first and last register are used, as the number of registers m increases, the allowable range of ω_1 becomes smaller and smaller. For example, $1 \leq \omega_1 \leq 1.032$ when $m = 32$.

IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we have discussed the accelerated gossip algorithm using two shift-registers on each node. We investigate the spectrum of the enlarged expectation matrix and find the optimal coefficient and the fastest rate of convergence in expectation which depends on the probability matrix P. The theoretical results are verified by looking into some special classes of regular graphs. We have also established convergence of second moment and studied multi-register symmetric gossip algorithms through the similar approach based on matrix analysis and established a necessary condition for the combination of the coefficients. Currently we are looking at the more challenging case where the probability matrix P is asymmetric. The technical difficulty is that the expectation of the system update matrix can no longer be written as a clean block matrix as we have done for the symmetric case, and thus it is hard to use the spectrum analysis technique that we have heavily relied on in this paper. Tools from the convergence analysis of infinite sequences of nonnegative matrices may prove to be useful for our research in the future.

REFERENCES

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Gossip algorithms: design, analysis and applications. In *Proc. of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pages 1653–1664, 2005.

[2] M. Cao, A. S. Morse, and B. D. O. Anderson. Reaching a consensus in a dynamically changing environment - a graphical approach. *SIAM Journal on Control and Optimization*, 47(2):575–600, 2008.

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[4] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1):65–78, 2004.

[5] M. Cao, D. A. Spielman, and A. S. Morse. A lower bound on convergence of a distributed network consensus algorithm. In *Proc. of the 44th IEEE CDC*, pages 2356–2361, 2005.

[6] C. C. Moallemi and B. V. Roy. Consensus propagation. *IEEE Transactions on Information Theory*, 52(11):4753–4766, 2006.

[7] A. Olshevsky and J. Tsitsiklis. Convergence rates in distributed consensus and averaging. In *Proc. of the 45th IEEE CDC*, pages 3387–3392, 2006.

[8] D. Mosk-Aoyama and D. Shah. Computing separable function via gossip. In *Proc. of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 113–122, 2005.

[9] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright. Geographic gossip: efficient aggregatiton for sensor networks. In *Proc. of the 5th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 69–76, 2006.

[10] M. Cao, D. A. Spielman, and E. M. Yeh. Accelerated gossip algorithms for distributed computation. In *Proc. of the 44th Annual Allerton Conference on Communication, Control, and Computation*, pages 952–959, 2006.

[11] B. Johansson and M. Johansson. Faster linear iterations for distributed averaging. In *Proc. of the 17th International Federation of Automatic Control World Congress (IFAC)*, pages 2861–2866, 2008.

[12] D. M. Cvetkovic, M. Doob, and H. Sachs. *Spectra of Graphs: Theory and Application*, page 72. Academic Press, New York, 1979.

[13] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[14] J. Friedman. A proof of Alon's second eigenvalue conjecture. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing (STOC)*, pages 720–724, 2003.