



Analysis of accelerated gossip algorithms[☆]



Ji Liu^{a,1}, Brian D.O. Anderson^b, Ming Cao^c, A. Stephen Morse^a

^a Yale University, USA

^b The Australian National University and National ICT Australia Ltd., Australia

^c University of Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 7 January 2011

Received in revised form

25 June 2012

Accepted 24 October 2012

Available online 12 February 2013

Keywords:

Cooperative control

Distributed averaging

Convergence rate

ABSTRACT

Gossiping is a distributed process whose purpose is to enable the members of a group of $n > 1$ autonomous agents to asymptotically determine in a decentralized manner, the average of the initial values of their scalar gossip variables. This paper analyzes the accelerated gossip algorithms, first proposed in Cao, Spielman, and Yeh (2006), in which local memory is exploited by installing shift-registers at each agent. For the two-register case, the existence of the desired convergence is established under a symmetry assumption by separately studying the convergence in expectation and in mean square. In particular, the optimal rate of convergence in expectation is derived which is faster than that of the standard gossip algorithm, and a sufficient condition on the adjustable parameter for the convergence in mean square is provided. These theoretical results are validated for some classes of networks by comparison with existing empirical data. More general multi-register cases are also discussed.

© 2013 Elsevier Ltd. All rights reserved.

1. 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 among the members of a group of sensors or mobile autonomous agents, especially for the scenarios where sensors or agents are constrained by limited sensing, computation, and communication capabilities. One particular type of distributed information processing which has received much attention lately is called distributed averaging (Xiao & Boyd, 2004). In its simplest form, distributed averaging deals with a network of $n > 1$ agents and the constraint that each agent i is able to communicate only

with certain other agents called agent i 's neighbors. Neighbor relations are described by a simple, connected graph \mathbb{A} in which vertices correspond to agents and edges indicate neighbor relations. Thus the neighbors of an agent i have the same labels as the vertices in \mathbb{A} which are adjacent to vertex i . Initially, each agent has or acquires a real number y_i which might be a measured temperature or something similar. The distributed averaging problem is to devise a protocol which will enable each agent to compute the average $y_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n y_i$ using only information acquired from its neighbors.

The distributed averaging problem can be solved in many ways. A well studied approach to the problem is for each agent to use a linear iterative update rule of the general form

$$x_i(t+1) = w_{ii}x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij}x_j(t), \quad x_i(0) = y_i \quad (1)$$

where t is a discrete time index, $x_i(t)$ is agent i 's current estimate of y_{avg} , the w_{ij} are real-valued weights, and \mathcal{N}_i is the set of labels of the neighbors of agent i . By introducing an n -vector x whose i th entry is x_i , such a system of equations admits a state space model of the form

$$x(t+1) = Wx(t), \quad t \geq 0 \quad (2)$$

where W is a real-valued $n \times n$ matrix whose row and column sums all equal one. In Xiao and Boyd (2004) several methods are proposed for choosing the w_{ij} with the goal of obtaining algorithms with improved convergence rates. Algorithms of this type, which require each agent to communicate with all of its neighbors on each iteration, are sometimes called broadcast algorithms.

An alternative approach to distributed averaging, which typically does not involve broadcasting, exploits a form of “gossiping”

[☆] The work of J. Liu and A.S. Morse was supported by the US Air Force Office of Scientific Research, the US Army Research Office, and the National Science Foundation. The work of B.D.O. Anderson was supported by the Australian Research Council under DP-0877562 and DP-110100538, and by National ICT Australia, which is funded by the Australian Government as represented by the Department of Broadband Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. B.D.O. Anderson was also supported by the US Air Force Research laboratory grant number FA2386-10-1-4102. The work of M. Cao was supported by grants from the Dutch Organization for Scientific Research (NWO), the Dutch Technology Foundation (STW) and the European Research Council (ERC-StG-2012-307207). The material in this paper was partially presented at the 48th IEEE Conference on Decision and Control (CDC) and the 28th Chinese Control Conference (CCC), December 16–18, 2009 in Shanghai, China. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Ian R. Petersen.

E-mail addresses: ji.liu@yale.edu (J. Liu), brian.anderson@anu.edu.au (B.D.O. Anderson), m.cao@rug.nl (M. Cao), as.morse@yale.edu (A.S. Morse).

¹ Tel.: +1 203 432 4078; fax: +1 203 432 7481.

(Boyd, Ghosh, Prabhakar, & Shah, 2006). In a standard gossiping process, a pair of agents with labels i and j are said to *gossip* at time t if both $x_i(t+1)$ and $x_j(t+1)$ are set equal to the average of $x_i(t)$ and $x_j(t)$. Each agent is allowed to gossip with at most one neighbor at one time. Under appropriate assumptions, algorithms which possess this simple property can be shown to solve the distributed averaging problem. Generally not every pair of agents is allowed to gossip. The edges of \mathcal{A} specify which gossip pairs are allowable. The actual sequence of gossip pairs which occurs during a specific gossip sequence might be determined either probabilistically (Boyd et al., 2006; Kempe, Dobra, & Gehrke, 2003) or deterministically (Liu, Mou, Morse, Anderson, & Yu, 2011; Mehyar, Spanos, Pongsa-japan, Low, & Murray, 2007), depending on the problem of interest. It is the former type of problem to which this paper is addressed.

As mentioned previously, the goal of gossiping is to compute the average of the initial values of the x_i , henceforth called *gossip variables*. Of particular interest is the rate at which a sequence of gossip variables converge to the same value at y_{avg} . Recent results concerning convergence rates appear in Anderson, Yu, and Morse (2010), He, Morse, Liu, and Mou (2011), Mou, Yu, Anderson, and Morse (2010) and Yu et al. (to be submitted) for deterministic periodic gossiping and in Liu, Morse, Anderson, and Yu (2010); Liu et al. (2011), Nedić, Olshevsky, Ozdaglar, and Tsitsiklis (2009) and Olshevsky and Tsitsiklis (2009) for deterministic aperiodic gossiping. In Boyd et al. (2006), the convergence rate question is addressed for gossip algorithms in which the sequence of gossip pairs is determined probabilistically. A modified probabilistic gossip algorithm intended to speed up convergence is proposed in Cao et al. (2006) without proof of correctness, but with convincing experimental results. The algorithm has recently been analyzed in Liu, Anderson, Cao, and Morse (2009). This paper presents a more comprehensive treatment of the work in Liu et al. (2009).

In Boyd et al. (2006), Boyd et al. propose a standard gossip algorithm where at each clock time a single randomly chosen pair of agents in the network update their values of gossip variables together to the average of their current values. Such an algorithm is easy to implement and only requires simple computations at each agent. However, the convergence rate of the algorithm is relatively slow, which is a critical drawback and it is clearly desirable that it should be improved. A number of papers have studied this issue of slow convergence of the standard gossip algorithm and various strategies have been proposed to improve the convergence rate (Bénézit, Dimakis, Thiran, & Vetterli, 2010; Cao et al., 2006; Dimakis, Sarwate, & Wainwright, 2006). For a survey on the most recent works on this issue see Dimakis, Kar, Moura, Rabbat, and Scaglione (2010).

Among all the proposed acceleration strategies, there is one particular approach that has motivated the research in this paper. In Cao et al. (2006), Cao et al. have demonstrated through simulations that by installing shift-registers at each agent to enable utilization of computational results in each agent's finite memory, substantial acceleration (up to 10-fold) can be achieved for probabilistic gossip algorithms. However, there is no theoretical explanation in Cao et al. (2006) for this observed improvement. In fact, except for the preliminary version of this paper (Liu et al., 2009), no one has yet been able to describe the accelerated convergence of the gossip algorithms using shift-registers, first proposed in Cao et al. (2006), in a rigorous and precise fashion. The main contribution of this paper is to thoroughly analyze the two-register case of the accelerated gossip algorithms in Cao et al. (2006) under a symmetry assumption by separately studying the convergence in expectation and in mean square. The fastest rate of convergence in expectation is derived in closed form which depends on the given probabilistic strategy according to which agent pairs are activated. A sufficient condition on the adjustable parameter for the convergence in mean square is also provided.

These theoretical results are validated for some classes of networks by comparison with the empirical data in Cao et al. (2006).

Also related to this paper is the work of Aysal, Oreshkin, and Coates (2009), Johansson and Johansson (2008), Muthukrishnan, Ghosh, and Schultz (1998) and Oreshkin, Coates, and Rabbat (2010) in which several clever ideas have emerged for speeding up the convergence of broadcast algorithms by augmenting (1) with delayed values of $x_i(t)$. In a pioneering paper (Muthukrishnan et al., 1998), Muthukrishnan et al. propose augmented update equations of the form

$$x_i(t+1) = g \sum_{j \in \mathcal{N}_i} w_{ij} x_j(t) + (1-g)x_i(t-1)$$

$x_i(1) = x_i(0) = y_i$ where g is a real, constant design parameter not depending on i . In matrix terms one thus has

$$x(t+1) = gWx(t) + (1-g)x(t-1), \quad t \geq 1. \quad (3)$$

It has been shown in Muthukrishnan et al. (1998) that there exists a value of $g \in (1, 2)$ dependent on W with which a faster rate at which $x(t)$ converges is obtained compared with (2). This result is inspired by the work of Golub and Varga (1961) and derived using matrix iterative analysis (Varga, 1962). Recently the more sophisticated iteration

$$x(t+1) = gWx(t) + (1-g)(\theta_3 Wx(t) + \theta_2 x(t) + \theta_1 x(t-1)), \quad t \geq 1 \quad (4)$$

has been proposed by Aysal et al. (2009) to further speed up convergence where θ_1 , θ_2 , and θ_3 are prespecified real-valued constants, but without proof of correctness. The system (4) has been analyzed lately by Oreshkin et al. (2010) under special spectrum assumptions.

Although exploiting the same over-relaxation technique as the preceding augmented broadcast algorithms, the accelerated gossip algorithms discussed in this paper admit probabilistic time-varying systems in which the update matrix depends on time, and consequently we prove convergence both in expectation and in mean square; by way of contrast the systems (3) and (4) are deterministic and time-invariant. Therefore there is a fundamental difference between this paper and the work of Aysal et al. (2009), Johansson and Johansson (2008), Muthukrishnan et al. (1998) and Oreshkin et al. (2010). Although both our work and the work of Aysal et al. (2009) and Oreshkin et al. (2010) are aiming at the same goal of providing theoretical explanations about why shift-registers can help accelerate convergence, different system models have been set up and the results obtained in one cannot be applied directly to the other.

The remainder of this paper is structured as follows. In Section 2, we review the standard gossip algorithm studied in Boyd et al. (2006) and the accelerated gossip algorithms introduced in Cao et al. (2006). In Section 3, we first establish a complete analysis of the convergence in expectation for the two-register case under a symmetry assumption (Section 3.1). Then we study the convergence in mean square and provide a sufficient condition under the same assumption (Section 3.2). Moreover, we also establish a necessary condition for speeding up convergence in more general multi-register cases (Section 3.3). Finally, we validate the theoretical results in Section 4 by exploring some classes of network graphs and comparing them with the experimental results in Cao et al. (2006).

2. Gossiping

As mentioned in the introduction, gossiping is a form of distributed computation whose purpose is to calculate the average value of a set of numbers or measurements. The type of gossiping we want to consider involves a group of $n > 1$ agents labeled 1 to n . Each agent i has control over a real-valued scalar quantity x_i called a *gossip variable* which the agent is able to update from time to time. We say that a *gossip* occurs at time $t \in \{0, 1, 2, \dots\}$

between agents i and j if both agents communicate with each other and update their gossip variables at time t ; the values of both agents' gossip variables at time $t + 1$ are determined by a pre-specified gossip algorithm. If agent i does not gossip at time t , its gossip variable does not change. Generally not every pair of agents is allowed to gossip. We use a simple, undirected n -vertex graph \mathbb{A} , called an *allowable gossip graph*, to specify which gossip pairs are allowable. In other words, a gossip between agents i and j is *allowable* if and only if (i, j) is an edge in \mathbb{A} . We use the symbol \mathcal{E} to denote the edge set of \mathbb{A} . Agent j is a *neighbor* of agent i if $(i, j) \in \mathcal{E}$. Each agent is allowed to gossip with at most one of its neighbors at one time. Let $x_i(0)$ be the initial value of agent i 's gossip variable and let $x_{\text{ave}} = \frac{1}{n} \sum_{i=1}^n x_i(0)$ denote the average of the initial values of all n gossip variables. The goal of gossiping is for the n agents to reach a consensus in the sense that all n gossip variables ultimately reach the same value at x_{ave} in the limit as $t \rightarrow \infty$. For this to be possible, no matter what the initial values of the gossiping variables are, it is clearly necessary that \mathbb{A} is a connected graph. We assume that this is so.

2.1. Gossip algorithms

At each time $t \in \{0, 1, 2, \dots\}$, we assume that exactly one agent in the group is activated and that each agent has an equal probability $\frac{1}{n}$ of being activated. If agent i is activated, then with probability p_{ij} , agent i chooses agent j to gossip, where $p_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$. Let P be an $n \times n$ matrix whose ij th entry is p_{ij} . We call such P a *probability matrix* which characterizes a probabilistic gossip algorithm. It is clear that P is a "stochastic matrix" where, by a *stochastic matrix*, we mean a nonnegative square matrix whose row sums all equal one. Since P is stochastic, it has an eigenvalue 1 and its spectral radius is 1 (Horn & Johnson, 1985). Following Boyd et al. (2006), we assume that P has eigenvalue 1 with multiplicity 1 and all the remaining $n - 1$ eigenvalues are strictly less than 1 in magnitude. Such a matrix can always be found if the underlying graph \mathbb{A} is connected and nonbipartite (Boyd et al., 2006). It is worth emphasizing that this assumption imposed on P also implies that \mathbb{A} is a connected graph.² Let us agree to call a probabilistic gossip algorithm *symmetric* if P is a symmetric matrix (i.e., $p_{ij} = p_{ji}$ for all $i, j \in \{1, 2, \dots, n\}$). For most of this paper, a symmetry assumption will be made.

Consider a probabilistic gossip algorithm characterized by a probability matrix P . Suppose that the set of agents' update rules consists of distributed linear iterations. Then the gossiping process can be modeled by a discrete time linear system in a state form. Toward this end, for each $(i, j) \in \mathcal{E}$, let $A_{(i,j)}$ be the matrix characterizing the update rule when agents i and j are the only pair to gossip. Then

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

where $x = [x_1 \ x_2 \ \dots \ x_n]'$ is the vector of gossip variables and $\sigma : \{0, 1, 2, \dots\} \rightarrow \mathcal{E}$ is a switching signal whose value at time t , is the index of the edge representing the randomly chosen pair of agents gossiping at time t . For simplicity, we will adopt the notation A_{ij} instead of $A_{(i,j)}$ in the sequel. Let \bar{A} denote the mean of the independent and identically distributed (i.i.d.) matrices $A_{\sigma(t)}$. Then we have

$$E[x(t)] = \prod_{k=0}^t E[A_{\sigma(k)}] x(0) = \bar{A}^t x(0).$$

Thus $E[x(t)]$ converges if \bar{A}^t converges as $t \rightarrow \infty$. Following Boyd et al. (2006), we term this *convergence in expectation*. The

conditions on \bar{A} for this to happen have been stated in Boyd et al. (2006). Let $\|\cdot\|$ denote the Euclidean norm on \mathbb{R}^n . Suppose that convergence in expectation occurs and that the limiting value of $x(t)$ is x_∞ . If in addition $E[\|x(t) - x_\infty\|^2] \rightarrow 0$ as $t \rightarrow \infty$, we say that $x(t)$ *converges in mean square*. We are interested in having both types of convergence and indeed ensuring that x_∞ is actually $x_{\text{ave}} \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^n$ is a column vector whose entries are all ones. Convergence in expectation ensures that the mean of the density of $x(t)$ converges, and convergence in mean square ensures that the variance of those densities tends to zero. The notion of carrying out a separate examination of these two types of convergence and then amalgamating them to obtain the desired overall convergence result is drawn from Boyd et al. (2006).

We first review the standard gossip algorithm studied in Boyd et al. (2006). At time $t \geq 0$, suppose that agent i is activated (with probability $\frac{1}{n}$) and chooses its neighbor j (with probability $p_{ij} > 0$) to gossip. Then both i and j update their values of gossip variables at time $t + 1$ to the average of their values at time t . In other words

$$\begin{cases} x_i(t+1) = \frac{1}{2}x_i(t) + \frac{1}{2}x_j(t) \\ x_j(t+1) = \frac{1}{2}x_i(t) + \frac{1}{2}x_j(t). \end{cases} \quad (5)$$

The values of all the other agents remain the same:

$$x_k(t+1) = x_k(t), \quad k \neq i, j. \quad (6)$$

Then the update matrix A_{ij} can be written as

$$A_{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - e_j)'$$

where e_i is the unit n -vector with the i th entry equal to 1 and I is the $n \times n$ identity matrix. The mean of $A_{\sigma(t)}$ equals

$$\bar{A} = E[A_{\sigma(t)}] = \sum_{i,j} \frac{1}{n} p_{ij} A_{ij}.$$

For all $i, j \in \{1, 2, \dots, n\}$, each A_{ij} is a "doubly stochastic matrix" where, by a *doubly stochastic matrix*, we mean a nonnegative square matrix whose column and row sums all equal one, and so is \bar{A} . It has been shown in Boyd et al. (2006) that if all n agents in the group adhere to the standard gossip algorithm (5) and (6), then all n gossip variables converge to x_{ave} in expectation if and only if \bar{A} has eigenvalue 1 with multiplicity 1 and all the remaining $n - 1$ eigenvalues are strictly less than one in magnitude. The authors of Boyd et al. (2006) also establish a necessary and sufficient condition for convergence in mean square which is that $E[A_{\sigma(t)} \otimes A_{\sigma(t)}]$ has all but one eigenvalue less than one in magnitude, where \otimes stands for the Kronecker product. Moreover, a more computationally convenient sufficient condition for convergence in mean square is provided in Boyd et al. (2006) which is that $E[A'_{\sigma(t)} A_{\sigma(t)}]$ has all but one eigenvalue less than one. Since it can be verified that $E[A'_{\sigma(t)} A_{\sigma(t)}] = \bar{A}$, the rate of convergence in mean square is determined by the second largest eigenvalue of \bar{A} which is nonnegative since \bar{A} is positive semidefinite.

In Cao et al. (2006), Cao et al. introduce a technique which uses memory in the form of shift-registers to accelerate the standard gossip algorithm. Each agent is provided with the same number of registers, the first of which stores the agent's current value of gossip variable, and the remainder of which store earlier values corresponding to the agent. The algorithm changes how a pair of agents updating their gossip variables once they decide to gossip. For each $i \in \{1, 2, \dots, n\}$, let x_{ir} denote the value stored in agent i 's r th register.

In the case when each agent is provided with just two registers, the first of which stores the agent's current value and the second of which stores the agent's value before the latest update, the

² Suppose \mathbb{A} is a disconnected graph with k components. Then it is easy to verify that P has eigenvalue 1 with multiplicity k .

algorithm can be described as follows. Suppose i and j are the only pair of agents to gossip at time t . Then

$$\begin{cases} x_{i1}(t+1) = \omega \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + (1-\omega)x_{i2}(t) \\ x_{j2}(t+1) = x_{i1}(t) \\ x_{j1}(t+1) = \omega \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + (1-\omega)x_{j2}(t) \\ x_{i2}(t+1) = x_{j1}(t) \end{cases} \quad (7)$$

where $1 \leq \omega < 2$ is an adjustable parameter independent of i and j . The values of the registers of all the other agents remain the same:

$$x_{kr}(t+1) = x_{kr}(t), \quad k \neq i, j, \quad r = 1, 2. \quad (8)$$

In the case when each agent is provided with $m > 2$ registers, the accelerated gossip algorithm is generalized as follows. Suppose agents i and j are the only pair to gossip at time t . Then

$$\begin{cases} x_{i1}(t+1) = \omega_1 \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + \sum_{r=2}^m \omega_r x_{ir}(t) \\ x_{ir}(t+1) = x_{i(r-1)}(t), \quad r = 2, 3, \dots, m \\ x_{j1}(t+1) = \omega_1 \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + \sum_{r=2}^m \omega_r x_{jr}(t) \\ x_{jr}(t+1) = x_{j(r-1)}(t), \quad r = 2, 3, \dots, m \end{cases} \quad (9)$$

where $\omega_r, r \in \{1, 2, \dots, m\}$, are adjustable parameters satisfying $\sum_{r=1}^m \omega_r = 1$. The values stored in the registers of all the other agents remain the same:

$$x_{kr}(t+1) = x_{kr}(t), \quad k \neq i, j, \quad r = 1, 2, \dots, m. \quad (10)$$

The experiments in Cao et al. (2006) 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., $\omega_r = 0$ for $r \neq 1, m$). 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 adjustable parameters, the algorithms can speed up convergence substantially.

3. Analysis

In this section, we provide a theoretical analysis of the accelerated gossip algorithms proposed in Cao et al. (2006). Throughout this section, we assume that $P = P'$ except when otherwise noted. For the case when each agent is provided with two registers, we derive the fastest rate of convergence in expectation and the corresponding optimal parameter ω ; a sufficient condition on ω ensuring convergence in mean square is also provided. We begin with some quantities associated with the standard gossip algorithm.

In Boyd et al. (2006), Boyd et al. write the expectation matrix \bar{A} of the standard gossip algorithm as

$$\bar{A} = I - \frac{1}{2n}D + \frac{1}{2n}(P + P')$$

where D is a diagonal matrix with entries $d_{ii} = \sum_{j=1}^n (p_{ij} + p_{ji}) = 1 + \sum_{j=1}^n p_{ji}$. Note that the symmetry of \bar{A} does not depend on P being symmetric. However, for symmetric P , it is easier to calculate certain quantities. For example, $D = 2I$ and

$$\bar{A} = \left(1 - \frac{1}{n}\right)I + \frac{1}{n}P.$$

Let $\lambda_2(P)$ denote the second largest eigenvalue of P which is necessarily real since P is symmetric. Then the second largest

eigenvalue of \bar{A} equals

$$\lambda_2(\bar{A}) = 1 - \frac{1}{n}(1 - \lambda_2(P))$$

which is also necessarily real.

Now we consider the changes that arise with the accelerated gossip algorithm in the two-register case. The probabilities stay the same, but the new update equations are listed in (7). Define the enlarged state vector as

$$z = [x_{11} \ x_{21} \ \cdots \ x_{n1} \ x_{12} \ x_{22} \ \cdots \ x_{n2}]'.$$

The matrices corresponding to A_{ij} and \bar{A} of dimension $n \times n$ are then replaced by matrices of dimension $2n \times 2n$. Let B_{ij} and \bar{B} denote the corresponding enlarged matrices. Then

$$B_{ij} = \begin{bmatrix} \sum_{k \neq i,j} e_k e_k' + \frac{\omega}{2}(e_i + e_j)(e_i + e_j)'(1 - \omega)(e_i e_i' + e_j e_j') \\ e_i e_i' + e_j e_j' \end{bmatrix}$$

and

$$\bar{B} = \sum_{i,j} \frac{1}{n} p_{ij} B_{ij} = \begin{bmatrix} I + \frac{\omega}{2n}(P + P') + \frac{\omega - 2}{2n}D & \frac{1 - \omega}{n}D \\ \frac{1}{n}D & I - \frac{1}{n}D \end{bmatrix}.$$

With $P = P'$, there holds $D = 2I$ and

$$\bar{B} = \begin{bmatrix} \left(1 + \frac{\omega - 2}{n}\right)I + \frac{\omega}{n}P & \frac{2(1 - \omega)}{n}I \\ \frac{2}{n}I & \left(1 - \frac{2}{n}\right)I \end{bmatrix}. \quad (11)$$

It is easy to see that for any ω , the row sums of B_{ij} and \bar{B} all equal one. However, not all entries of B_{ij} and \bar{B} are nonnegative when $1 < \omega < 2$ because of the upper right block. Therefore B_{ij} and \bar{B} are not stochastic matrices.

3.1. Convergence in expectation

In this subsection, we establish the existence of the desired convergence in expectation of the accelerated gossip algorithm in the two-register case. It is shown that the use of the second register provides a speed-up in convergence. We study the behavior of the second largest magnitude of any eigenvalue of \bar{B} , denoted by $\rho_2(\bar{B})$, and identify the value of ω maximizing the speed-up.

First observe that with $\omega = 1$, the rate of convergence in expectation remains the same as the standard gossip algorithm since only one register is actually used. We shall next consider how $\rho_2(\bar{B})$ varies as a function of ω when $1 < \omega < 2$. Before doing this, we show the existence of the desired convergence of the accelerated gossip algorithm. Toward this end, we need the following lemma.

Lemma 1. Suppose M is an $n \times n$ real matrix. Let L be a $2n \times 2n$ matrix given by

$$L = \begin{bmatrix} M & aI \\ bI & cI \end{bmatrix}$$

where $a, b, c \in \mathbb{R}$ and $ab \neq 0$. If there exist $\lambda \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}^n$ such that α, β are not both zero vectors and

$$L \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

then $\alpha \neq 0, \lambda \neq c$, and $M\alpha = \mu\alpha$ where

$$\mu = \lambda - \frac{ab}{\lambda - c}. \quad (12)$$

Conversely, if there exist $\mu \in \mathbb{R}$ and $\alpha \in \mathbb{R}^n$ such that $\alpha \neq 0$ and $M\alpha = \mu\alpha$, then

$$L \begin{bmatrix} \alpha \\ \beta_i \end{bmatrix} = \lambda_i \begin{bmatrix} \alpha \\ \beta_i \end{bmatrix}, \quad i = 1, 2$$

where λ_1, λ_2 are the two solutions of (12) and $\beta_i = \frac{b}{\lambda_i - c}\alpha$.

The simple proof is omitted.

Lemma 1 implies that the $2n$ eigenvalues of the $2n \times 2n$ matrix

$$\bar{B} = \begin{bmatrix} A & aI \\ bI & cI \end{bmatrix}$$

where $a = \frac{2}{n}(1 - \omega)$, $b = \frac{2}{n}$, $c = 1 - \frac{2}{n}$, and

$$A = \left(1 + \frac{\omega - 2}{n}\right)I + \frac{\omega}{n}P \quad (13)$$

are determined by the n eigenvalues of the $n \times n$ matrix A with (12) which is equivalent to a quadratic equation

$$\lambda^2 - (\mu + c)\lambda + \mu c - ab = 0. \quad (14)$$

In addition, the n eigenvalues of A are determined by the n eigenvalues of P because of (13). Thus the $2n$ eigenvalues of \bar{B} are determined by the n eigenvalues of P . Let $\lambda_i(P)$ denote the i th largest eigenvalue of P , $i \in \{1, 2, \dots, n\}$. Since P is a symmetric stochastic matrix and we have assumed that P has eigenvalue 1 with multiplicity 1 and all the other $n - 1$ eigenvalues are strictly less than one in magnitude, we have

$$1 = \lambda_1(P) > \lambda_2(P) \geq \lambda_3(P) \geq \dots \geq \lambda_n(P) > -1.$$

Let $\mu_i(A)$ denote the i th largest eigenvalue of A , $i \in \{1, 2, \dots, n\}$. From (13), the n eigenvalues of A can also be written as a non-increasing sequence with the values

$$\mu_i(A) = 1 + \frac{\omega - 2}{n} + \frac{\omega}{n}\lambda_i(P). \quad (15)$$

In particular, the largest eigenvalue of A is $\mu_1(A) = 1 + \frac{2}{n}(\omega - 1)$. For each $\mu_i(A)$, we can obtain two eigenvalues of \bar{B} which are the two roots of (14), denoted by λ_{i1} and λ_{i2} . We assume that $\lambda_{i1} \geq \lambda_{i2}$ if they are both real. Thus when $\mu_i(A)$ satisfies the condition that Eq. (14) has two real roots, the corresponding two real eigenvalues of \bar{B} are

$$\lambda_{i1,2} = \frac{\mu_i(A) + c}{2} \pm \sqrt{\left(\frac{\mu_i(A) - c}{2}\right)^2 + ab}. \quad (16)$$

Replace $\mu_i(A)$ by $\lambda_i(P)$ using (15), we get

$$\lambda_{i1,2} = 1 + \frac{1}{n} \left(\frac{1}{2} \omega (1 + \lambda_i(P)) - 2 \right) \pm \frac{1}{2n} \sqrt{\omega^2 (1 + \lambda_i(P))^2 - 16(\omega - 1)}. \quad (17)$$

First, we notice from (16) that $\partial \lambda_{i1} / \partial \mu_i(A) > 0$, which implies that the real eigenvalues λ_{i1} (i belongs to a subset of $\{1, 2, \dots, n\}$) form a non-increasing sequence. Second, from (17) we have $\lambda_{i2} \geq 0$ when $1 < \omega < 2$, which implies that all the real eigenvalues of \bar{B} are nonnegative. Third, λ_{11} and λ_{12} are always real when $1 < \omega < 2$. In particular, $\lambda_{11} = 1$ is the largest real eigenvalue of \bar{B} and

$$\lambda_{12} = 1 - \frac{4}{n} + \frac{2\omega}{n} \quad (18)$$

which can be computed using (17) with $\lambda_1(P) = 1$. Furthermore, if λ_{21} and λ_{22} are real as well, the second largest real eigenvalue of \bar{B} is the larger of λ_{12} and λ_{21} .

On the other hand, if for some $\mu_i(A)$, the corresponding two eigenvalues of \bar{B} (i.e., the two roots of (14)) are complex, the

magnitudes of these two eigenvalues are the same:

$$\begin{aligned} |\lambda_{i1,2}| &= |\lambda_{i1}| = |\lambda_{i2}| = \sqrt{\mu c - ab} \\ &= \sqrt{1 - \frac{4}{n} + \frac{\omega}{n}(1 + \lambda_i(P)) + \frac{2\omega}{n^2}(1 - \lambda_i(P))}. \end{aligned} \quad (19)$$

Now we have the explicit expressions for the magnitudes of all eigenvalues of \bar{B} , including both real and complex eigenvalues.

Lemma 2. Suppose P is an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. Let \bar{B} be defined by (11). If $1 < \omega < 2$, then \bar{B} has an eigenvalue 1 and all the remaining $2n - 1$ eigenvalues are strictly less than 1 in magnitude.

Proof of Lemma 2. We consider the complex eigenvalues and real eigenvalues of \bar{B} separately. For any complex eigenvalue $\lambda_{i1,2}$, using (19) it can be checked that $\partial |\lambda_{i1,2}| / \partial \omega > 0$ and $\partial |\lambda_{i1,2}| / \partial \lambda_i(P) \geq 0$. This implies that $|\lambda_{i1,2}|$ increases as either ω or $\lambda_i(P)$ increases. Since $1 < \omega < 2$ and $-1 < \lambda_i(P) \leq 1$, an upper bound of $|\lambda_{i1,2}|$ is given by (19) with $\omega = 2$ and $\lambda_i(P) = 1$. With these values, the right side of (19) evaluates as $(1 - \frac{2}{n})^{\frac{1}{2}}$. Thus $|\lambda_{i1,2}| \leq (1 - \frac{2}{n})^{\frac{1}{2}} < 1$. Next we turn to the real eigenvalues. Recall that following (17), we have established that all the real eigenvalues of \bar{B} are nonnegative and $\lambda_{11} = 1$ is the largest. It can be verified that the second largest real eigenvalue is always smaller than 1 when $1 < \omega < 2$. Therefore 1 is a simple eigenvalue of \bar{B} . \square

We are now in a position to show the convergence in expectation of the accelerated gossip algorithm (7) and (8).

Theorem 1. Suppose that the enlarged vector of gossip variables $z(t)$ evolves according to

$$z(t+1) = B_{\sigma(t)} z(t), \quad t \geq 0 \quad (20)$$

where $B_{\sigma(t)}$ is determined by the accelerated gossip algorithm (7) and (8). Let P be an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. If $1 < \omega < 2$, then all the entries of $z(t)$ converge in expectation to the average

$$z_{\text{ave}} = \frac{1}{n} \sum_{i=1}^n x_{i1}(0)$$

if each agent i 's two registers are initialized so that $x_{i1}(0) = x_{i2}(0)$, $i \in \{1, 2, \dots, n\}$.

The proof of Theorem 1 depends on the following lemma.

Lemma 3. Given any matrix $M \in \mathbb{R}^{n \times n}$ and vectors $c, d \in \mathbb{R}^n$ such that $c'd \neq 0$. The equation

$$\lim_{t \rightarrow \infty} M^t = \frac{dc'}{c'd}$$

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.

This lemma is Theorem 2 in Xiao and Boyd (2004). Note that $\rho(M - dc'/(c'd))$ equals the second largest magnitude of any eigenvalue of M , and thus $\rho(M - dc'/(c'd)) < 1$ implies that 1 is a simple eigenvalue of M .

Proof of Theorem 1. Since the matrices $B_{\sigma(t)}$, $t \in \{0, 1, 2, \dots\}$, are independent and identically distributed, from (20) we have

$$E[z(t)] = \prod_{i=0}^t E[B_{\sigma(i)}] z(0) = \bar{B}^t z(0)$$

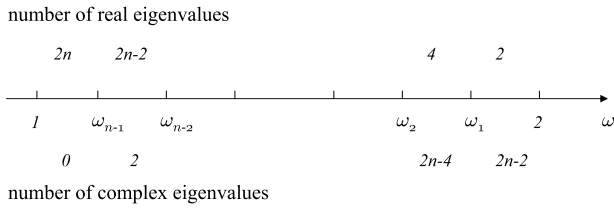


Fig. 1. Distribution of the number of \bar{B} 's real and complex eigenvalues.

where

$$\bar{B} = E[B_{\sigma(t)}] = \sum_{i,j} \frac{1}{n} p_{ij} B_{ij}.$$

It can be verified that

$$B_{ij} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (21)$$

and (Lemma 1 in Cao et al. (2006))

$$[\mathbf{1}' \quad (1-\omega)\mathbf{1}'] B_{ij} = [\mathbf{1}' \quad (1-\omega)\mathbf{1}']. \quad (22)$$

Because \bar{B} is a convex combination of B_{ij} 's, Eqs. (21) and (22) also hold for \bar{B} . We have shown in Lemma 2 that one is a simple eigenvalue of \bar{B} and all the other eigenvalues are strictly less than one in magnitude. Thus by Lemma 3,

$$\lim_{t \rightarrow \infty} \bar{B}^t = \frac{1}{(2-\omega)n} \begin{bmatrix} 1 \\ 1 \end{bmatrix} [\mathbf{1}' \quad (1-\omega)\mathbf{1}'].$$

If $x_{i1}(0) = x_{i2}(0)$, then

$$\lim_{t \rightarrow \infty} E[z(t)] = \lim_{t \rightarrow \infty} \bar{B}^t z(0) = \frac{1}{n} \sum_{i=1}^n x_{i1}(0).$$

This completes the proof. \square

In the sequel we will write the second largest magnitude of any eigenvalue of \bar{B} as a function of ω , denoted by $\rho_2(\bar{B}(\omega))$, and then find the optimal value 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 (17) it is clear 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

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

Then 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 implies 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} for all $i \in \{2, 3, \dots, n\}$, we have

$$1 < \omega_{n-1} \leq \omega_{n-2} \leq \dots \leq \omega_1 < 2.$$

If $1 < \omega \leq \omega_{n-1}$, \bar{B} has $2n$ real eigenvalues; if $\omega_{n-1} < \omega \leq \omega_{n-2}$, \bar{B} has $2n-2$ real eigenvalues and 2 complex eigenvalues; \dots ; if $\omega_2 < \omega \leq \omega_1$, \bar{B} has 4 real eigenvalues and $2n-4$ complex eigenvalues; if $\omega_1 < \omega < 2$, \bar{B} has 2 real eigenvalues and $2n-2$ complex eigenvalues. See Fig. 1 for the distribution of the number of \bar{B} 's real and complex eigenvalues.

Lemma 4. Suppose that P is an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n-1$ eigenvalues strictly less than 1 in magnitude. Let \bar{B} be defined by (11) and

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

where $\lambda_2(P)$ is the second largest eigenvalue of P . If $1 < \omega \leq \omega_1$, then

$$\begin{aligned} \rho_2(\bar{B}(\omega)) &= \lambda_{21}(\omega) = 1 + \frac{1}{n} \left(\frac{1}{2} \omega (1 + \lambda_2(P)) - 2 \right) \\ &\quad + \frac{1}{2n} \sqrt{(1 + \lambda_2(P))^2 \omega^2 - 16\omega + 16}. \end{aligned} \quad (23)$$

Proof of Lemma 4. We consider real eigenvalues and complex eigenvalues separately. Since $1 < \omega \leq \omega_1$, the second largest real eigenvalue of \bar{B} is the larger one between λ_{12} and λ_{21} , which are given by (18) and (23) respectively. It can be checked that $\lambda_{21}(\omega) > \lambda_{12}(\omega)$ when $1 < \omega \leq \omega_1$. Thus the second largest real eigenvalue of \bar{B} is λ_{21} . It can also be checked that $\partial \lambda_{21}(\omega) / \partial \omega < 0$ when $1 < \omega \leq \omega_1$, which implies that $\lambda_{21}(\omega)$ decreases as ω increases. In addition, when $\omega = \omega_1$, $\lambda_{21}(\omega)$ and $\lambda_{22}(\omega)$ are equal and $\lambda_{21}(\omega)$ reaches the minimum

$$\lambda_{21}(\omega_1) = \sqrt{1 - \frac{4}{n} + \frac{\omega_1}{n} (1 + \lambda_2(P)) + \frac{2\omega_1}{n^2} (1 - \lambda_2(P))}. \quad (24)$$

Next we turn to complex eigenvalues. Recall that the magnitude of complex eigenvalues increases as either ω or $\lambda_i(P)$ increases, thus the right side of (24) is an upper bound of $|\lambda_{i1,2}|$ when $1 < \omega \leq \omega_1$. Therefore when $1 < \omega \leq \omega_1$, $\lambda_{21}(\omega)$ is always greater than the largest magnitude of any complex eigenvalue, which completes the proof. \square

Our next step is to consider the case when $\omega_1 < \omega < 2$. In this case, \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 of λ_{12} and $|\lambda_{21}|$:

$$\rho_2(\bar{B}(\omega)) = \max\{\lambda_{12}(\omega), |\lambda_{21}(\omega)|\}.$$

We repeat the two expressions as follows:

$$\lambda_{12}(\omega) = 1 - \frac{2}{n} (2 - \omega)$$

and

$$|\lambda_{21}(\omega)| = \sqrt{1 - \frac{4}{n} + \frac{\omega}{n} (1 + \lambda_2(P)) + \frac{2\omega}{n^2} (1 - \lambda_2(P))}.$$

It can be verified that $\lambda_{12}(\omega) \geq |\lambda_{21}(\omega)|$ if and only if

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

Note that $|\lambda_{21}(\omega)| > \lambda_{12}(\omega)$ when $\omega = \omega_1$ and $\lambda_{12}(\omega) > |\lambda_{21}(\omega)|$ when $\omega = 2$. So we have $g(\omega_1) < 0$ and $g(2) > 0$. Because $g(\omega)$ is a second order polynomial, $g(\omega)$ has a unique zero point between ω_1 and 2. Let ω_0 denote this zero point. Then when $\omega_1 < \omega < \omega_0$, $|\lambda_{21}| > \lambda_{12}$; when $\omega_0 \leq \omega < 2$, $\lambda_{12} \geq |\lambda_{21}|$. We are led to the following lemma.

Lemma 5. Suppose that P is an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n-1$ eigenvalues strictly less than 1 in magnitude. Let \bar{B} be defined by (11) and

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

where $\lambda_2(P)$ is the second largest eigenvalue of P . Then when $\omega_1 < \omega < \omega_0$, $\rho_2(\bar{B}(\omega)) = |\lambda_{21}(\omega)|$, and when $\omega_0 \leq \omega < 2$, $\rho_2(\bar{B}(\omega)) = \lambda_{12}(\omega)$.

Combining Lemmas 4 and 5, it is clear 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 $[\omega_0, 2)$. We provide an

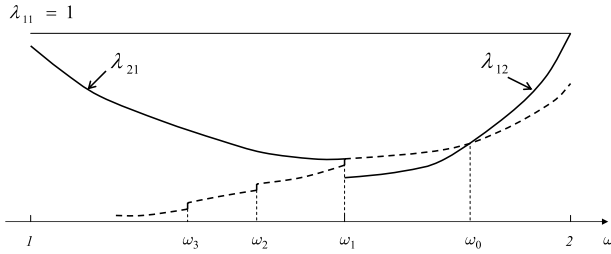


Fig. 2. The magnitude of some eigenvalues of $\bar{B}(\omega)$.

overview of $\rho_2(\bar{B}(\omega))$ in Fig. 2 by plotting the curve of the magnitude of those specific eigenvalues of $\bar{B}(\omega)$ which determine the value of $\rho_2(\bar{B}(\omega))$. The dashed curve denotes the largest magnitude of any complex eigenvalue; it is discontinuous at $\omega_1, \omega_2, \dots, \omega_{n-1}$ because $\{\lambda_1(P), \lambda_2(P), \dots, \lambda_n(P)\}$ is a discrete set. Note that when $\omega = 2$, $\lambda_{12} = \lambda_{11} = 1$ and thus convergence cannot be reached.

Theorem 2. Suppose that P is an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. Let B be defined by (11) and $\lambda_2(P)$ be the second largest eigenvalue of P . Then when $1 < \omega < 2$, the minimum of $\rho_2(\bar{B}(\omega))$ is unique; the value of ω at this minimum is

$$\omega^* = \omega_1 = \frac{8 - 4\sqrt{4 - (1 + \lambda_2(P))^2}}{(1 + \lambda_2(P))^2} \quad (25)$$

and the minimum of $\rho_2(\bar{B}(\omega))$ is

$$\rho_2(\bar{B}(\omega_1)) = 1 + \frac{1}{n} \left(\frac{4 - 2\sqrt{4 - (1 + \lambda_2(P))^2}}{1 + \lambda_2(P)} - 2 \right). \quad (26)$$

Proof of Theorem 2. 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 the minimum 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 and $\lambda_{12}(\omega) > |\lambda_{21}(\omega)|$ on $\omega_1 < \omega < 2$, $|\lambda_{21}(\omega_1)|$ gives the infimum 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 $\rho_2(\bar{B}(\omega))$. \square

Note that $\rho_2(\bar{A}) = \lambda_2(\bar{A})$ since \bar{A} is symmetric and positive semidefinite. Recall that the rate of convergence in expectation of the standard gossip algorithm with symmetric P is

$$\lambda_2(\bar{A}) = 1 - \frac{1}{n}(1 - \lambda_2(P)). \quad (27)$$

It can be verified that $\rho_2(\bar{B}(\omega_1)) < \rho_2(\bar{A})$ when $-1 < \lambda_2(P) < 1$. We are led to the following result:

Corollary 1. Suppose that P is an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n - 1$ eigenvalues strictly less than 1 in magnitude. Then the accelerated gossip algorithm utilizing two shift-registers at each agent with $\omega = \omega^*$ as defined in (25) has a faster rate of convergence in expectation than the standard gossip algorithm.

3.2. Convergence in mean square

In this subsection, we investigate the convergence in mean square of the accelerated algorithm (7) and (8). Let $y(t) = z(t) - z_{\text{ave}}\mathbf{1}$ where $z(t) = [x_{11} \ x_{21} \ \dots \ x_{n1} \ x_{12} \ x_{22} \ \dots \ x_{n2}]'$ and $z_{\text{ave}} = \frac{1}{n} \sum_{i=1}^n x_{i1}(0)$. Then

$$\begin{aligned} y(t+1) &= z(t+1) - z_{\text{ave}}\mathbf{1} \\ &= B_{\sigma(t)}z(t) - z_{\text{ave}}B_{\sigma(t)}\mathbf{1} \\ &= B_{\sigma(t)}y(t). \end{aligned}$$

The second equation holds because of the fact that $\mathbf{1}$ is an eigenvector of all $B_{\sigma(t)}$ for eigenvalue one. Following the argument of Boyd et al. (2006), we obtain the bound

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

If $y(t)$ were constrained to be orthogonal to the eigenvector of $E[B'B]$ for the largest eigenvalue, we would have

$$E[y(t)'y(t)] \leq \lambda_2^t(E[B'B])\|y(0)\|^2. \quad (28)$$

Therefore we would obtain a sufficient condition for the convergence in mean square; the corresponding convergence rate is $\lambda_2(E[B'B])$. However, it is easy to verify that when $\omega = 1$ and $P = P'$, $\lambda_1(E[B'B])$ is greater than one; the same property holds for ω near to one by continuity. Also examples reveal that we cannot always have $\lambda_2(E[B'B])$ strictly less one.

In the sequel we will present a modification of this approach to establish the convergence in mean square.³

Suppose that N is a nonsingular matrix. Observe that if $N^{-1}z(t)$ and $N^{-1}y(t)$ converge, then so do $z(t)$ and $y(t)$, and conversely. Thus we can study $N^{-1}z(t)$ and $N^{-1}y(t)$ rather than $z(t)$ and $y(t)$. We retain B as shorthand for $B_{\sigma(t)}$. The evolution of $\hat{y}(t) = N^{-1}y(t)$ becomes

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

Compared with the evolution of $y(t)$, 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 $B'B$. In fact, we are interested in the eigenvalues of $E[C'C]$, and in particular its second largest eigenvalue. If we select N as a triangular block matrix

$$N = \begin{bmatrix} \alpha I & \gamma I \\ 0 & \beta I \end{bmatrix}$$

where α, β , and γ are real-valued constants, then

$$N^{-1} = \begin{bmatrix} \frac{1}{\alpha}I & -\frac{\gamma}{\alpha\beta}I \\ 0 & \frac{1}{\beta}I \end{bmatrix}$$

and

$$C = N^{-1}BN = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where

$$C_{11} = \sum_{k \neq i,j} e_k e_k' + \frac{\omega}{2}(e_i + e_j)(e_i + e_j)' - \frac{\gamma}{\beta}(e_i e_i' + e_j e_j')$$

$$\begin{aligned} C_{12} &= \frac{\gamma}{\alpha} \frac{\omega}{2}(e_i + e_j)(e_i + e_j)' \\ &\quad + \left(\frac{\beta}{\alpha}(1 - \omega) - \frac{\gamma^2}{\alpha\beta} \right)(e_i e_i' + e_j e_j') \end{aligned}$$

$$C_{21} = \frac{\alpha}{\beta}(e_i e_i' + e_j e_j')$$

$$C_{22} = \sum_{k \neq i,j} e_k e_k' + \frac{\gamma}{\beta}(e_i e_i' + e_j e_j').$$

Let $r = \frac{\beta}{\alpha}$ and $s = \frac{\gamma}{\beta}$. More can be said:

³ In Boyd et al. (2006), another criterion for convergence in mean square is derived for the standard gossip algorithm based on the expected value of the Kronecker product. However, rewriting $E[B \otimes B]$ in an elegant form has so far eluded us.

Lemma 6. Let r and s be any two non-zero real numbers such that

$$r^2(1-s)(1+s-\omega) = 1. \quad (29)$$

Then $N'v$ is an eigenvector of $E[C'C]$ for eigenvalue 1 where v is a column vector such that

$$v' = \begin{bmatrix} \frac{1}{2-\omega} \mathbf{1}' & \frac{1-\omega}{2-\omega} \mathbf{1}' \end{bmatrix}.$$

Proof of Lemma 6. With symmetric P , it can be shown that

$$E[C'C] = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}$$

where

$$E_{11} = \left(1 - \frac{2}{n} + \frac{2}{n} \left(s^2 + \frac{1}{r^2}\right)\right) I + \frac{2}{n} \left(\frac{\omega^2}{2} - s\omega\right) (P + I)$$

$$E_{12} = E_{21} = \frac{2}{n} \left(rs^3 - rs(1-\omega) + \frac{s}{r}\right) I + \frac{2}{n} \left(rs \frac{\omega^2}{2} - rs^2\omega + \frac{r}{2}\omega(1-\omega)\right) (P + I)$$

$$E_{22} = \frac{2}{n} (r^2s^4 - 2r^2s^2(1-\omega) + r^2(1-\omega)^2 + s^2) I + \frac{2}{n} \left(r^2s^2 \frac{\omega^2}{2} - r^2s^3\omega + r^2s\omega(1-\omega)\right) (P + I) + \left(1 - \frac{2}{n}\right) I.$$

Thus

$$E[C'C]N'v = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} \begin{bmatrix} a\mathbf{1} \\ b\mathbf{1} \end{bmatrix} = \begin{bmatrix} a\Phi\mathbf{1} \\ b\Psi\mathbf{1} \end{bmatrix} \quad (30)$$

where $a = \frac{\alpha}{2-\omega}$, $b = \frac{\gamma+\beta(1-\omega)}{2-\omega}$, and

$$\begin{aligned} \Phi &= 1 - \frac{2}{n} + \frac{2}{n} \left(s^2 + \frac{1}{r^2}\right) + \frac{4}{n} \left(\frac{\omega^2}{2} - s\omega\right) \\ &\quad + \frac{2}{r(1-s)n} \left(rs^3 - rs(1-\omega) + \frac{s}{r}\right) \\ &\quad + \frac{4}{r(1-s)n} \left(rs \frac{\omega^2}{2} - rs^2\omega + \frac{r}{2}\omega(1-\omega)\right) \\ \Psi &= 1 - \frac{2}{n} + \frac{2}{n} (r^2s^4 - 2r^2s^2(1-\omega) + r^2(1-\omega)^2 + s^2) \\ &\quad + \frac{4}{n} \left(r^2s^2 \frac{\omega^2}{2} - r^2s^3\omega + r^2s\omega(1-\omega)\right) \\ &\quad + \frac{2}{n} r(1-s) \left(rs^3 - rs(1-\omega) + \frac{s}{r}\right) \\ &\quad + \frac{4}{n} r(1-s) \left(rs \frac{\omega^2}{2} - rs^2\omega + \frac{r}{2}\omega(1-\omega)\right). \end{aligned}$$

It can be verified that $\Phi = \Psi$ if and only if $r^2(1-s)(1+s-\omega) = 1$. Moreover, $\Phi = \Psi = 1$ when (29) holds. Then from (30), we have $E[C'C]N'v = N'v$, which implies that $N'v$ is an eigenvector of $E[C'C]$. \square

Note that given any pair of non-zero real numbers r and s , there always exist $\alpha, \beta, \gamma \in \mathbb{R}$ such that $r = \frac{\beta}{\alpha}$ and $s = \frac{\gamma}{\beta}$. In addition, from (29) it is clear that

$$\omega - 1 < s < 1. \quad (31)$$

Recall that $y(t) = z(t) - z_{\text{ave}}\mathbf{1}$. With the initial conditions $x_{i1}(0) = x_{i2}(0)$, $i \in \{1, 2, \dots, n\}$, Eq. (22) ensures that $v'y(t) = 0$ for all

$t \geq 0$. Consequently, $\hat{y}(t) \perp N'v$ for all $t \geq 0$. Thus similar to (28), provided that 1 is the largest eigenvalue of $E[C'C]$, we have

$$E[\hat{y}(t)\hat{y}(t)] \leq \lambda_2^t(E[C'C])\|\hat{y}(0)\|^2.$$

The fact that 1 is the largest eigenvalue of $E[C'C]$ will be established with the aid of the following lemma.

Lemma 7. Suppose M is an $n \times n$ real matrix. Let L be a $2n \times 2n$ matrix given by

$$L = \begin{bmatrix} a_1I + a_2(M+I) & a_3I + a_4(M+I) \\ a_3I + a_4(M+I) & a_5I + a_6(M+I) \end{bmatrix}$$

where $a_i \in \mathbb{R}$, $i \in \{1, 2, \dots, 6\}$, and I is the $n \times n$ identity matrix. If μ is an eigenvalue of M , then there are two corresponding eigenvalues of L , which are the two roots of the quadratic equation

$$\lambda^2 + b\lambda + c = 0 \quad (32)$$

with $b = -(a_1 + a_5) - (a_2 + a_6)(1 + \mu)$ and $c = -(a_3^2 - a_1a_5) - (2a_3a_4 - a_1a_6 - a_2a_5)(1 + \mu) - (a_4^2 - a_2a_6)(1 + \mu)^2$.

The straightforward proof is omitted.

From the proof of Lemma 6, it is easy to see that $E[C'C]$ has the same form as L . By Lemma 7, for each eigenvalue μ of P ($-1 < \mu \leq 1$), there are two corresponding eigenvalues of $E[C'C]$ which are the two roots of (32):

$$\lambda = \frac{1}{2} \left((a_1 + a_5) + (a_2 + a_6)(1 + \mu) \pm \sqrt{\Delta} \right) \quad (33)$$

where

$$\begin{aligned} \Delta &= (4a_3^2 + (a_1 - a_5)^2) + (8a_3a_4 + 2(a_1 - a_5)(a_2 - a_6)) \\ &\quad \times (1 + \mu) + (4a_4^2 + (a_2 - a_6)^2)(1 + \mu)^2 \\ &= 4(a_3 + a_4(1 + \mu))^2 + ((a_1 - a_5) + (a_2 - a_6)(1 + \mu))^2. \end{aligned}$$

For the matrix $E[C'C]$, we have

$$\begin{cases} a_1 = 1 - \frac{2}{n} + \frac{2}{n} \left(s^2 + \frac{1}{r^2}\right) \\ a_2 = \frac{2}{n} \left(\frac{\omega^2}{2} - s\omega\right) \\ a_3 = \frac{2}{n} \left(rs^3 - rs(1-\omega) + \frac{s}{r}\right) \\ a_4 = \frac{2}{n} \left(rs \frac{\omega^2}{2} - rs^2\omega + \frac{r}{2}\omega(1-\omega)\right) \\ a_5 = 1 - \frac{2}{n} + \frac{2}{n} (r^2s^4 - 2r^2s^2(1-\omega) + r^2(1-\omega)^2 + s^2) \\ a_6 = \frac{2}{n} \left(r^2s^2 \frac{\omega^2}{2} - r^2s^3\omega + r^2s\omega(1-\omega)\right). \end{cases} \quad (34)$$

It is clear that $\lambda \geq 0$ since $E[C'C]$ is symmetric and positive semidefinite. When we choose the plus sign and $\mu = 1$ in (33), it can be verified that $\lambda = 1$. In addition, it can be shown from (33) that a sufficient condition under which $\lambda = 1$ is the unique largest eigenvalue of $E[C'C]$ consists of the following two inequalities:

$$\begin{cases} a_2 + a_6 \geq 0 \\ 8a_3a_4 + 2(a_1 - a_5)(a_2 - a_6) \leq 8a_4^2 + 2(a_2 - a_6)^2. \end{cases} \quad (35)$$

The computation shows that $a_2 + a_6 \geq 0$ if and only if

$$\omega \leq 4 - 2\sqrt{2} \quad (36)$$

and

$$\omega - \sqrt{\omega^2 - 8\omega + 8} \leq 4s \leq \omega + \sqrt{\omega^2 - 8\omega + 8}. \quad (37)$$

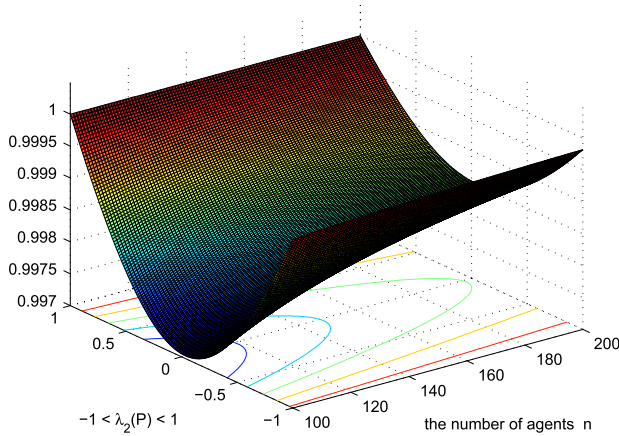


Fig. 3. The rate of convergence in mean square with $\omega = 4 - 2\sqrt{2}$.

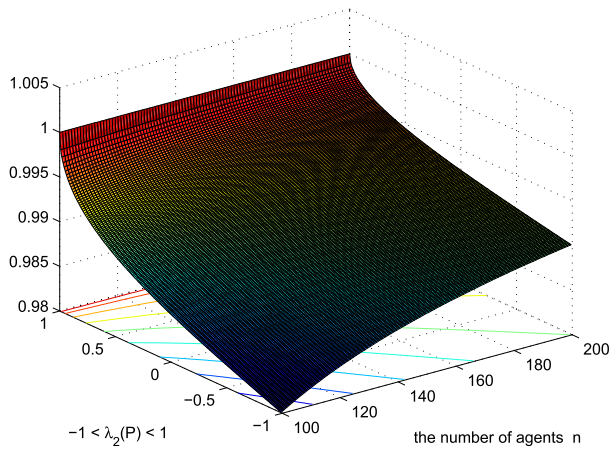


Fig. 4. The rate of convergence in expectation.

When (36) holds, it can be checked that (37) holds as well and it gives a tighter bound on s than (31). Let $s = \frac{\omega}{4}$ with which (37) holds and let

$$r = ((1-s)(1+s-\omega))^{-\frac{1}{2}}$$

so that (29) holds. With these values, it can be verified that the second inequality in (35) also holds. Thus when $1 < \omega \leq 4 - 2\sqrt{2}$, $E[C'C]$ has one as the largest eigenvalue and all the other eigenvalues are strictly less than one. We have proved the following.

Proposition 1. Suppose that the enlarged vector of gossip variables $z(t)$ evolves according to $z(t+1) = B_{\sigma(t)}z(t)$, $t \geq 0$ where $B_{\sigma(t)}$ is determined by the accelerated gossip algorithm (7) and (8). Let P be an $n \times n$ symmetric probability matrix with an eigenvalue 1 and all the remaining $n-1$ eigenvalues strictly less than 1 in magnitude. If $1 < \omega \leq 4 - 2\sqrt{2}$, then $z(t)$ converges in mean square.

From (33), it is clear that an upper bound of the rate of convergence in mean square is

$$\lambda_2[C'C] = \frac{1}{2} \left((a_1 + a_5) + (a_2 + a_6)(1 + \lambda_2(P)) + \sqrt{\Delta} \right) \quad (38)$$

where

$$\Delta = 4(a_3 + a_4(1 + \lambda_2(P)))^2 + ((a_1 - a_5) + (a_2 - a_6)(1 + \lambda_2(P)))^2.$$

Here the a_i , $i \in \{1, 2, \dots, 6\}$, are given by (34) with $s = \frac{\omega}{4}$ and $r = ((1-s)(1+s-\omega))^{-\frac{1}{2}}$. Recall that the fastest rate of convergence in expectation is

$$\rho_2(\bar{B}(\omega^*)) = 1 + \frac{1}{n} \left(\frac{4 - 2\sqrt{4 - (1 + \lambda_2(P))^2}}{1 + \lambda_2(P)} - 2 \right).$$

It is of interest to compare the two convergence rates. Toward this end, set $\omega = 4 - 2\sqrt{2}$. Then both $\lambda_2[C'C]$ and $\rho_2(\bar{B}(\omega_1))$ depend only on n and $\lambda_2(P)$. We plot them in Figs. 3 and 4 respectively as a function of $\lambda_2(P)$ and n with $-1 < \lambda_2(P) < 1$ and $100 \leq n \leq 200$. It is shown from the figures that the rate of convergence in mean square is slower than the rate of convergence in expectation.

3.3. Multi-register cases

In order to possibly achieve faster convergence, we analyze more general multi-register cases of the accelerated gossip algorithms in Cao et al. (2006) from the standpoint of convergence in expectation. We take X4 of Cao et al. (2006) as an example in which all of the 4 registers are used to compute the new value of the first register. Suppose agents i and j are the pair to gossip at time $t \geq 0$. Then the accelerated gossip algorithm is given as follows:

$$\begin{cases} x_{i1}(t+1) = \omega_1 \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + \sum_{r=2}^4 \omega_r x_{ir}(t) \\ x_{ir}(t+1) = x_{i(r-1)}(t), \quad r = 2, 3, 4 \\ x_{j1}(t+1) = \omega_1 \left(\frac{1}{2}x_{i1}(t) + \frac{1}{2}x_{j1}(t) \right) + \sum_{r=2}^4 \omega_r x_{jr}(t) \\ x_{jr}(t+1) = x_{j(r-1)}(t), \quad r = 2, 3, 4 \end{cases}$$

where ω_r , $r \in \{1, 2, 3, 4\}$, are adjustable parameters satisfying $\sum_{r=1}^4 \omega_r = 1$. We define the state vector as a $4n$ -dimensional column vector ordering the 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 B_{ij} and \bar{B} of dimension $4n \times 4n$. Using a similar analysis as presented earlier, we obtain

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

where $b = \frac{2}{n}$, $c = 1 - \frac{2}{n}$, and

$$A_1 = \left(1 + \frac{\omega_1 - 2}{n} \right) I + \frac{\omega_1}{n} P$$

A_1 is symmetric because P is. Then the $4n$ eigenvalues of \bar{B} 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 an eigenvalue of \bar{B} and μ denotes an eigenvalue of A_1 . Also, the eigenvalues of A_1 are determined by the eigenvalues of P through the relation

$$\mu_i(A_1) = 1 + \frac{\omega_1 - 2}{n} + \frac{\omega_1}{n} \lambda_i(P), \quad i \in \{1, 2, \dots, n\}. \quad (39)$$

It is easy to extend the preceding results to more general cases. That is, for the m -register accelerated algorithm (9) and (10) with $P = P'$, the mn eigenvalues of the $mn \times mn$ matrix \bar{B} are determined by the n eigenvalues of $n \times n$ matrix A_1 through the equation

$$\mu = \lambda - \sum_{r=2}^m \frac{2\omega_r}{n} \left(\frac{b}{\lambda - c} \right)^{r-1} \quad (40)$$

while the eigenvalues of A_1 are determined by the eigenvalues of P through the relation (39). In particular, $\lambda_1(P) = 1$ and $\mu_1(A_1) = 1 + \frac{2}{n}(\omega_1 - 1)$. Although it becomes much harder to get an explicit expression for λ from (40) when $m > 2$, we can establish a necessary condition on the combination of the ω_r for the convergence in expectation.

Set $\lambda = 1 + \epsilon$. Then the right hand side of (40) can be written as a function of ϵ :

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

Note that $h(0) = \mu_1(A_1)$ which is the largest eigenvalue of A_1 and that

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

When $\epsilon = 0$, $\lambda = 1$ is an eigenvalue of \bar{B} . Thus in order to ensure the convergence in expectation, $\lambda = 1 + \epsilon$ cannot be an eigenvalue of \bar{B} when $\epsilon > 0$, which implies that $h'(0) \geq 0$. Therefore, we get a necessary condition for the ω_r :

$$1 + \sum_{r=2}^m \omega_r (r-1) \geq 0. \quad (41)$$

If only the first and last registers are used to compute the new value of the first register, the $\omega_r = 0$, $r \in \{2, 3, \dots, m-1\}$, and $\omega_m = 1 - \omega_1$. Then the necessary condition (41) becomes

$$\omega_1 \leq 1 + \frac{1}{m-1}. \quad (42)$$

It is easy to verify that the simulation results of X4 in Cao et al. (2006) satisfy (41) and that of D2, D4, and D8 all satisfy (42). See TABLE II in Cao et al. (2006). In addition, the condition (42) implies that when only the first and last registers are used, the allowable range of ω_1 becomes smaller and smaller as the number of registers m increases. For example, $1 < \omega_1 \leq 1.032$ when $m = 32$. Also note that from (42), $\omega_1 \leq 2$ when $m = 2$. But it has been shown in Section 3.1 that when $\omega_1 = 2$, the convergence in expectation cannot be reached. Thus (42) is not a sufficient condition.

4. Experimental validation

In this section we will validate the theoretical results derived in the preceding section by exploring some classes of graphs and comparing them with the experimental results in Cao et al. (2006). Toward this end, we assume that each agent, if activated, gossips with its neighbors with equal probability. This assumption is also made in the experiments in Cao et al. (2006). Then for any symmetric gossip algorithm, an allowable gossip graph \mathbb{A} uniquely determines a probability matrix P . Moreover, such P matrices have the following property: For each integer $d \in \{2, 3, \dots, n-1\}$, $A = dP$ is the adjacency matrix of a d -degree regular graph whose eigenvalues are

$$d = \mu_1(A) > \mu_2(A) \geq \mu_3(A) \cdots \geq \mu_n(A) > -d.$$

It is clear that $\lambda_i(P) = \frac{\mu_i(A)}{d}$, $i \in \{1, 2, \dots, n\}$. From (25), ω^* depends on $\lambda_2(P)$, the second largest eigenvalue of P . We are now interested in how ω^* varies with different $\lambda_2(P)$.

We first consider two special classes of regular graphs: complete graphs \mathbb{K}_n (with $d = n-1$) and cycle graphs \mathbb{C}_n (with $d = 2$). For any complete graph \mathbb{K}_n , $n \geq 2$, the spectrum of its adjacency matrix consists of $n-1$ with multiplicity 1 and -1 with multiplicity $n-1$ (Cvetkovic, Doob, & Sachs, 1979). So in this case $\lambda_2(P) = -\frac{1}{n-1}$. In particular, when $n \geq 10000$, $\lambda_2(P) \approx 0$ and thus $\omega^* \approx 1.0718$. For any cycle graph \mathbb{C}_n , $n \geq 3$, the spectrum of its adjacency matrix consists of the numbers $2 \cos(2\pi i/n)$, $i \in \{1, 2, \dots, n\}$

(Cvetkovic et al., 1979). So in this case $\lambda_2(P) = \cos \frac{2\pi}{n}$. In particular, when $n \geq 10000$, $\lambda_2(P) \approx 1$ and thus $\omega^* \approx 2$. Next we consider another class of regular graphs \mathbb{C}_n^2 , the second power of cycle graphs \mathbb{C}_n .⁴ The adjacency matrix of \mathbb{C}_n^2 is a circulant matrix and its eigenvalues are given by $\mu_j = w_j + w_j^2 + w_j^{n-2} + w_j^{n-1}$, $j \in \{0, 1, \dots, n-1\}$, where w_j are the n roots of the unity which satisfy $w_j^n = 1$ (Biggs, 1974).

In Alon (1986), Alon has shown that for any fixed d and any infinite family of d -regular graphs \mathbb{G}_i , $i \in \{1, 2, \dots\}$, $\liminf \mu_2(A_i) \geq 2\sqrt{d-1}$, where A_i is the adjacency matrix of \mathbb{G}_i . He also conjectured that almost all d -regular graphs with n vertices satisfy $\mu_2(A) \leq 2\sqrt{d-1} + o(1)$ as n tends to infinity. The conjecture has subsequently been proved by Friedman (2003). Thus we take $2\sqrt{d-1}$ as an estimate of $\mu_2(A)$. Then

$$\lambda_2(P) \approx \frac{2\sqrt{d-1}}{d}. \quad (43)$$

A different class of regular graphs, namely two-dimensional grid graphs (with $d = 4$), can be seen as a crude approximation to the topology of a sensor network in the case when sensors are uniformly distributed and sensors have the same sensing radius (Penrose, 2003). If we ignore the boundary effect and substitute $d = 4$ into (43), we have $\lambda_2(P) \approx 0.866$ and thus $\omega^* \approx 1.47$, which agrees with the experimental result that ω^* should lie between 1.4 and 1.5. See TABLE II in Cao et al. (2006).

While Cao et al. (2006) 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 while most of the graphs considered in Cao et al. (2006) are random geometric graphs (Penrose, 2003) which rarely have a symmetric P . This explains why the analytical values of ω^* in the cases of \mathbb{K}_n and \mathbb{C}_n are quite different from the experimental results in Cao et al. (2006).

From the preceding, we have derived the values or approximations of $\lambda_2(P)$ where P is the probability matrix which characterizes a gossip algorithm for some classes of regular graphs in the case when each agent, if activated, gossips with its neighbors with equal probability. Let r_s denote the convergence rate of the standard gossip algorithm, r_e denote the optimal rate of convergence in expectation of the two-register accelerated gossip algorithm, and r_m denote an upper bound of the rate of convergence in mean square of the two-register accelerated gossip algorithm with $\omega = 4 - 2\sqrt{2}$. Recall that r_s given in (27), r_e given in (26), and r_m given in (38) all depend on $\lambda_2(P)$ and the number of agents n . Then it is possible to calculate or estimate these convergence rates for some specific allowable gossip graphs. The computation results are summarized in Table 1. From the table, it is clear that $r_e < r_s < r_m$. But it is worth emphasizing that r_m is just an upper bound of the convergence in mean square. The optimization of r_m , as what we have already done for r_e , has so far eluded us and is a subject for future research.

5. Conclusions and future work

In this paper, we study the accelerated gossip algorithms proposed in Cao et al. (2006) under a symmetry assumption. For the two-register case, we investigate the spectrum of the enlarged expectation matrix and derive the fastest rate of convergence in expectation which depends on the probability matrix P . We also establish a sufficient condition for the convergence in mean square. The theoretical results are verified by looking into some special classes of regular graphs and comparing with existing empirical

⁴ The k th power of a graph \mathbb{G} is a graph with the same set of vertices as \mathbb{G} and an edge between two vertices if and only if there is a path in \mathbb{G} of length at most k between them.

Table 1

	r_s	r_e	r_m
\mathbb{K}_{50}	0.9796	0.9705	0.9941
\mathbb{K}_{100}	0.9899	0.9853	0.9970
\mathbb{K}_{500}	0.9980	0.9971	0.9994
\mathbb{C}_{50}	0.9998	0.9966	0.9999
\mathbb{C}_{100}	0.9999	0.9991	0.9999
\mathbb{C}_{50}^2	0.9996	0.9947	0.9998
\mathbb{C}_{100}^2	0.9999	0.9986	0.9999
\mathbb{G}_{50}	0.9973	0.9874	0.9987
\mathbb{G}_{100}	0.9987	0.9937	0.9994
\mathbb{G}_{500}	0.9997	0.9987	0.9999

data. Through a similar approach based on matrix analysis, a necessary condition for the multi-register cases is also provided.

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

- Alon, N. (1986). Eigenvalues and expanders. *Combinatorica*, 6(2), 83–96.
- Anderson, B. D. O., Yu, C., & Morse, A. S. (2010). Convergence of periodic gossiping algorithms. In *Perspectives in mathematical system theory, control, and signal processing* (pp. 127–138). Springer.
- Aysal, T. C., Oreshkin, B. N., & Coates, M. J. (2009). Accelerated distributed average consensus via localized node state prediction. *IEEE Transactions on Signal Processing*, 57(4), 1563–1576.
- Bénézit, F., Dimakis, A. G., Thiran, P., & Vetterli, M. (2010). Order-optimal consensus through randomized path averaging. *IEEE Transactions on Information Theory*, 56(10), 5150–5167.
- Biggs, N. (1974). *Algebraic graph theory*. New York: Cambridge University Press.
- Boyd, S., Ghosh, A., Prabhakar, B., & Shah, D. (2006). Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6), 2508–2530.
- Cao, M., Spielman, D. A., & Yeh, E. M. (2006). Accelerated gossip algorithms for distributed computation. In *Proceedings of the 44th annual Allerton conference on communication, control, and computing* (pp. 952–959).
- Cvetkovic, D. M., Doob, M., & Sachs, H. (1979). *Spectra of graphs: theory and application*. New York: Academic Press.
- Dimakis, A. G., Kar, S., Moura, J. M. F., Rabbat, M. G., & Scaglione, A. (2010). Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11), 1847–1864.
- Dimakis, A. G., Sarwate, A. D., & Wainwright, M. J. (2006). Geographic gossip: efficient aggregation for sensor networks. In *Proceedings of the 5th international conference on information processing in sensor networks* (pp. 69–76).
- Friedman, J. (2003). A proof of Alon's second eigenvalue conjecture. In *Proceedings of the 35th annual ACM symposium on theory of computing* (pp. 720–724).
- Golub, G. H., & Varga, R. (1961). Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods—part I. *Numerische Mathematik*, 3, 147–156.
- He, F., Morse, A. S., Liu, J., & Mou, S. (2011). Periodic gossiping. In *Proceedings of the 18th IFAC world congress* (pp. 8718–8723).
- Horn, R. C., & Johnson, C. R. (1985). *Matrix analysis*. New York: Cambridge University Press.
- Johansson, B., & Johansson, M. (2008). Faster linear iterations for distributed averaging. In *Proceedings of the 17th IFAC world congress* (pp. 2861–2866).
- Kempe, D., Dobra, A., & Gehrke, J. (2003). Gossip-based computation of aggregate information. In *Proceedings of the 44th annual IEEE symposium on foundations of computer science* (pp. 482–491).
- Liu, J., Anderson, B. D. O., Cao, M., & Morse, A. S. (2009). Analysis of accelerated gossip algorithms. In *Proceedings of the 48th IEEE conference on decision and control* (pp. 871–876).
- Liu, J., Morse, A. S., Anderson, B. D. O., & Yu, C. (2010). The contraction coefficient of a complete gossip sequence. In *Three decades of progress in control sciences* (pp. 275–290). Springer.
- Liu, J., Mou, S., Morse, A. S., Anderson, B. D. O., & Yu, C. (2011). Deterministic gossiping. *Proceedings of the IEEE*, 99(9), 1505–1524.
- Mehyar, M., Spanos, D., Pongsajaporn, J., Low, S. H., & Murray, R. M. (2007). Asynchronous distributed averaging on communication networks. *IEEE/ACM Transactions on Networking*, 15(3), 512–520.
- Mou, S., Yu, C., Anderson, B. D. O., & Morse, A. S. (2010). Deterministic gossiping with a periodic protocol. In *Proceedings of the 49th IEEE conference on decision and control* (pp. 5787–5791).
- Muthukrishnan, S., Ghosh, B., & Schultz, M. H. (1998). First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory of Computing Systems*, 31, 331–354.
- Nedić, A., Olshevsky, A., Ozdaglar, A., & Tsitsiklis, J. N. (2009). On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, 54(11), 2506–2517.
- Olshevsky, A., & Tsitsiklis, J. N. (2009). Convergence speed in distributed consensus and averaging. *SIAM Journal Control and Optimization*, 48(1), 33–55.
- Oreshkin, B. N., Coates, M. J., & Rabbat, M. G. (2010). Optimization and analysis of distributed averaging with short node memory. *IEEE Transactions on Signal Processing*, 58(5), 2850–2865.
- Penrose, M. (2003). *Random geometric graphs*. New York: Oxford University Press.
- Varga, R. (1962). *Matrix iterative analysis*. Englewood Cliffs: Prentice Hall.
- Xiao, L., & Boyd, S. (2004). Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1), 65–78.
- Yu, C., Anderson, B. D. O., Mou, S., Liu, J., He, F., & Morse, A. S. (2013). Gossiping periodically. *IEEE Transactions on Automatic Control* (to be submitted).



Ji Liu was born in Shanghai, China. He received his B.S. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006 and his M.S. degree in electrical engineering from Yale University, New Haven, CT, in 2007, where he is currently working toward a Ph.D. in electrical engineering. His current research interests are in the areas of distributed control and social networks.



Brian D.O. Anderson was born in Sydney, Australia, and educated at Sydney University in mathematics and electrical engineering, with a Ph.D. in electrical engineering from Stanford University in 1966. He is a Distinguished Professor at the Australian National University and Distinguished Researcher in National ICT Australia (NICTA). His awards include the IFAC Quazza Medal in 1999, the IEEE Control Systems Award of 1997, the 2001 IEEE James H. Mulligan, Jr. Education Medal, and the Bode Prize of the IEEE Control System Society in 1992, and a number of other medals and best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, Honorary Fellow of the Institution of Engineers, Australia, and a Foreign Associate of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Université Catholique de Louvain, Belgium, and ETH, Zürich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science. He served as the first President of NICTA, and was a member of company boards, including Cochlear Ltd, the world's major supplier of bionic ears, and a member of the Prime Minister's Science Council under three prime ministers. His current research interests are in distributed control, sensor networks and econometric modeling.



Ming Cao is currently an assistant professor with the Faculty of Mathematics and Natural Sciences at the University of Groningen, The Netherlands. He received his Bachelor's degree in 1999 and Master's degree in 2002 from Tsinghua University, Beijing, China, and his Ph.D. in 2007 from Yale University, New Haven, CT, USA, all in electrical engineering. From September 2007 to August 2008, he was a Research Associate with the Department of Mechanical and Aerospace Engineering at Princeton University, Princeton, NJ, USA. He worked as a Research Intern during the summer of 2006 with the Mathematical Sciences Department at IBM T.J. Watson Research Center, NY, USA. His main research interest is in autonomous agents and multi-agent systems, mobile sensor networks and evolutionary network games. He is a member of the IFAC technical committee on networked systems, an associate editor for *Systems and Control Letters* and a member of the conference editorial board of the IEEE Control Systems Society.



A. Stephen Morse received his B.S.E.E. degree from Cornell University, his M.S. degree from the University of Arizona, and his Ph.D. from Purdue University. Following three years at the Office of Control Theory and Application (OCTA), NASA Electronics Research Center, Cambridge, MA, he joined the Yale University faculty, where he is currently the Dudley Professor of Engineering. His main interest is in network synthesis, optimal control, multi-variable control, adaptive control, urban transportation, vision-based control, hybrid and nonlinear systems, and, most recently, coordination and control of large grouping of mobile autonomous agents.

Dr. Morse received the George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society in 1993 and 2005, the American Automatic Control Council's Best Paper Award (twice), and the IEEE Technical Field Award for Control Systems in 1999. He is a life fellow of the IEEE, and a member of the National Academy of Engineering and the Connecticut Academy of Science and Engineering.