

Broadcast Gossip Algorithms for Consensus

Tuncer Can Aysal, *Member, IEEE*, Mehmet Ercan Yildiz, *Student Member, IEEE*, Anand D. Sarwate, *Member, IEEE*, and Anna Scaglione, *Senior Member, IEEE*

Abstract—Motivated by applications to wireless sensor, peer-to-peer, and ad hoc networks, we study distributed broadcasting algorithms for exchanging information and computing in an arbitrarily connected network of nodes. Specifically, we study a broadcasting-based gossiping algorithm to compute the (possibly weighted) average of the initial measurements of the nodes at every node in the network. We show that the *broadcast gossip algorithm* converges almost surely to a consensus. We prove that the random consensus value is, in expectation, the average of initial node measurements and that it can be made arbitrarily close to this value in mean squared error sense, under a balanced connectivity model and by trading off convergence speed with accuracy of the computation. We provide theoretical and numerical results on the mean square error performance, on the convergence rate and study the effect of the “mixing parameter” on the convergence rate of the broadcast gossip algorithm. The results indicate that the mean squared error strictly decreases through iterations until the consensus is achieved. Finally, we assess and compare the communication cost of the broadcast gossip algorithm to achieve a given distance to consensus through theoretical and numerical results.

Index Terms—Broadcasting, distributed average consensus, gossip algorithms, sensor networks.

I. INTRODUCTION

A fundamental problem in decentralized networked systems is that of having nodes reach a state of agreement [3]–[9]. Distributed agreement is a fundamental problem in ad hoc network applications, including distributed consensus and synchronization problems [6]–[8], [10], distributed coordination of mobile autonomous agents [4], [5], and distributed data fusion in sensor networks [3], [9], [11]. It is also a central topic for load balancing (with divisible tasks) in parallel computers [12]. Vicsek *et al.* provided a variety of simulation results which demonstrate that the simple distributed algorithms allow

Manuscript received March 17, 2008; accepted January 23, 2009. First published February 24, 2009; current version published June 17, 2009. The associate editor coordinating the review of this paper and approving it for publication was Dr. Zhi Tian. This work was supported by the NSF Grant CCF 0729074. This work was presented in part at the IEEE Information Theory Workshop, Porto, Portugal, May 2008, and the Conference on Decision and Control, Cancun, Mexico, December 2008.

T. C. Aysal and M. E. Yildiz are with the Communications Research in Signal Processing Group, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: tca27@cornell.edu; mey7@cornell.edu).

A. D. Sarwate was with the Sensory Information Processing and Communication Group, Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA. He is now with the Information Theory and Applications Center, University of California at San Diego, La Jolla, CA 92093-0447 USA (e-mail: asarwate@ucsd.edu).

A. Scaglione was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. She is now with Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: ascaglione@ucdavis.edu).

Digital Object Identifier 10.1109/TSP.2009.2016247

all nodes to eventually agree on a parameter [6]. The work in [13] provided the theoretical explanation for behavior observed in these reported simulation studies. This paper focuses on a prototypical example of agreement in asynchronous networked systems, namely, the randomized average consensus problem in a wireless broadcast communication network.

A. Related Work

Gossip-based algorithms to achieve consensus over a set of agents were initially introduced by Tsitsiklis [14] and have recently received renewed attention from several other researchers [3]–[5], [15]–[20]. Randomized average consensus gossiping is an asynchronous protocol where a node chosen uniformly at random wakes up, contacts a neighbor randomly within its *connectivity radius*, and exchanges a state variable to produce a computation update. If each update results in a pairwise average of their values, the operation preserves both the total sum, and hence also the mean, of the node values. In [15], it was shown that this algorithm converges to a consensus if the graph is strongly connected on the average. Because the transmitting node must send a packet to the chosen neighbor and then wait for the neighbor’s packet, this scheme is vulnerable to packet collisions and yields a communication complexity (measured by number of radio transmissions to drive the estimation error to within $\Theta(N^{-\alpha})$, for any $\alpha > 0$) on the order of $\Theta(N^2)$ over random geometric graphs [15].

The geographic gossip algorithm proposed in [21] combines gossip with geographic routing to improve the convergence rate of random gossiping. Similar to the standard gossip algorithm, a node randomly wakes up, chooses a node randomly in the *whole network*, rather than in its *neighborhood*, and performs a pairwise averaging with this node. Geographic gossiping increases the diversity of every pairwise averaging operation. The authors show that the communication complexity is in the order of $O(N^{3/2} \sqrt{\log(N)})$, which is an improvement with respect to the standard gossiping algorithm. More recently, a variety of the algorithm that “averages around the way” has been shown to converge in $O(N \log N)$ transmissions [22]. Moreover, we note that we assume that the messages involve real numbers; the effects of message quantization in gossip and consensus algorithms, is an active area of research [17], [19], [20], [23]–[27].

B. Primary Motivations

Geographic-type gossiping improves upon the convergence speed of the standard gossip by increasing the diversity of pairwise exchanges. However, the problem of packet loss is exacerbated by the requirement that messages must be sent on long routes, creating congestion issues. Moreover, it does not mitigate the major bottleneck associated with the fact that the messages between two peers need to be routed and exchanged to

perform two updates. Finally, successfully setting up a two-way route exacerbates the problem by requiring information about the location of the nodes in the network.

The wireless medium has the advantage of being inherently broadcast and, at the cost of one transmission, one can reach several terminals. Our objective in this paper is to analyze a broadcasting-based gossip algorithm that enables *all nodes in range* to perform an update by exploiting the wireless medium, and thereby avoiding the need of complex routing and problematic pairwise exchange operations.

C. Summary of Main Contributions

To overcome the drawbacks of the standard packet based gossip algorithms, we study a broadcast based gossiping algorithm for wireless sensor networks. In the studied algorithm, a node in the network wakes up uniformly at random according to the asynchronous time model and broadcasts its value. This value is successfully received by the nodes in the predefined radius of the broadcasting node, i.e., connectivity radius. The nodes that have received the broadcasted value update their own state value and the remaining nodes sustain their value. It is shown here that by iterating this procedure, this type of gossiping algorithm is capable of achieving consensus over the network with probability one. We also show that the random consensus value is, in expectation, equal to the desired value, i.e., the average of initial node measurements. Because the sum of the node state values is not preserved at each iteration, the broadcast gossiping algorithm converges to a value that is in the neighborhood of the desired average.

The question that motivates this paper is investigating if it is possible to avoid the partner selection process altogether, analyzing a broadcast communication protocol where each random transmission triggers an update by all nodes within range, without a mechanism of reply in place to maintain the network average [1]. Fagnani and Zampieri have concurrently studied the convergence to consensus characteristics of general randomized algorithms which do not necessarily converge to the initial average (such as asymmetric gossip, broadcast gossip and packet-drop gossip) [28]. In particular, the authors have shown that random consensus algorithms in general achieve probabilistic consensus, and discussed their mean squared error characteristics (Proposition 4.4 and Corollary 3.2). Thus, Lemma 2 and Theorem 1 of this paper are special cases of the general results given in [28]. However, the study of this paper focuses on aspects specific to broadcast consensus over wireless sensor networks, and expands on our preliminary studies [1], [2]. In this paper, we provide an in depth study of broadcast gossip algorithms' speed of convergence and mean squared error characteristics. Our results also address the choice of the *mixing parameter* and its effect on both the mean square error and the convergence rate, which provides insight for implementation.

More specifically, we provide theoretical and simulation results on the mean square error and communication cost performance of the broadcast gossip algorithm. Moreover, we study the effect of the so called *mixing parameter* on the convergence rate and limiting mean square error through theoretical results and numerical experiments. In addition, we derive the optimal

mixing parameter when approached from the convergence rate perspective. Although the convergence time of our algorithm is commensurate with the standard pairwise gossip algorithms, we present simulations showing that for more modest network sizes our algorithm converges to consensus faster than other algorithms based on pairwise averages or routing.

D. Paper Organization

The remainder of this paper is organized as follows. Section II introduces the average consensus problem and the graph and time models adopted in this paper. The studied broadcast gossip algorithm is introduced in Section III and its convergence characteristics are studied in Section IV. In Section V, we derive the optimal mixing parameter considering the worst-case convergence rate and analyze the effects of various network parameters on the optimal value. The MSE characterization and communication complexity analysis are given in Section VI along with the convergence rate expression. Finally, we conclude with some discussion and future directions in Section VII.

II. GRAPH AND TIME MODELS

In the following, we briefly discuss the graph and time models adopted in this paper. Then, we describe briefly the distributed average consensus problem.

A. Graph Model

We model our wireless sensor network as a random geometric graph $G(N, R)$, where the N sensor locations are chosen uniformly and independently in a unit square area, and each pair of nodes is connected if their Euclidean distance is smaller than some transmission radius R named *connectivity radius* [29]. For our analysis, we assume that a communication within this transmission radius always succeeds. It is well known that in order to have a fully connected network asymptotically while minimizing interference, the connectivity radius R has to scale like $\Theta(\sqrt{\log(N)/N})$ [15], [29]. The N -node topology of $G(N, R)$ is represented by the $N \times N$ adjacency matrix Φ , where for $i \neq j$, $\Phi_{ij} = 1$ if nodes i and j are in their neighborhood, and $\Phi_{ij} = 0$, otherwise. Moreover, we define $\mathcal{N}_i = \{j \in \{1, 2, \dots, N\} : \Phi_{ij} \neq 0\}$ and D as a diagonal matrix with entries $D_{ii} = \mathcal{N}_i$. Finally, the Laplacian of a graph is defined as $L = D - \Phi$.

B. Time Model

We use the asynchronous time model, which is well matched to the distributed nature of sensor networks [15], [21]. In this model, each sensor node is assumed to have a clock which ticks independently according to a rate μ Poisson process. Consequently, the inter-tick times are exponentially distributed and independent across nodes and over time. This process is equivalent to a single clock whose ticking times form a Poisson process of rate $N\mu$. Let Z_t be the arrival times of this global process. In expectation, there are approximately $N\mu$ clock ticks per unit of absolute time but we will always measure time in number of ticks of this (virtual) global clock. We therefore think of time as discretized with the interval $[Z_t; Z_{t+1}]$ corresponds to the t th timeslot. We can adjust time units relative to the communication

time so that only one broadcast event occurs in the network at each time slot with high probability.

C. Average Consensus

At time slot $t \geq 0$, each node $i = 1, 2, \dots, N$ has an estimate $x_i(t)$ of the global average, and we use $x(t)$ to denote the N -vector of these estimates. The ultimate goal is to use the minimal amount of communication to drive the estimate $x(t)$ as close as possible to the average vector $\bar{x}(0)\mathbf{1}$, where $\mathbf{1}$ is the vector of all 1's and

$$\bar{x}(0) = \frac{1}{N} \sum_{i=1}^N x_i(0). \quad (1)$$

Because our algorithms are randomized, the quantity $x(t)$ for $t > 0$ is a random vector even though we assume $x(0)$ is deterministic.

III. BROADCAST BASED GOSSIPING

Informally, the asynchronous broadcast gossip algorithm is described as follows. Suppose node i 's clock is the t th that ticked. Then, node i broadcasts its own state value which is received by all neighboring nodes within distance R from it. Once the broadcasted value is received, the neighboring nodes set their values equal to the (weighted) average of their current value and the value broadcasted by the node i . Formally, node i activates and the following events occur:

- node i broadcasts wirelessly its current state value, $x_i(t)$;
- the broadcasted value is successfully received by the nodes that are within the radius R ;
- all nodes in the set of node i 's neighbors \mathcal{N}_i receive the broadcasted value $x_i(t)$, and update their state values according to the following equation:

$$x_k(t+1) = \gamma x_k(t) + (1-\gamma)x_i(t), \quad \forall k \in \mathcal{N}_i \quad (2)$$

with $\gamma \in (0, 1)$ denoting the *mixing parameter*;

- the remaining nodes in the network, including i , update their state values as

$$x_k(t+1) = x_k(t), \quad \forall k \notin \mathcal{N}_i. \quad (3)$$

This procedure takes place at every clock tick.

Let $x(t)$ denote the vector of values at the end of the t th ticking event. Then

$$x(t+1) = W(t)x(t) \quad (4)$$

where the random matrix $W(t)$, with probability $1/N$ is (assuming that the i th clock ticks)

$$W_{jk}^{(i)} = \begin{cases} 1, & j \notin \mathcal{N}_i, k = j \\ \gamma, & j \in \mathcal{N}_i, k = j \\ 1 - \gamma, & j \in \mathcal{N}_i, k = i \\ 0, & \text{elsewhere} \end{cases} \quad (5)$$

where $W^{(i)}$ denotes the weight matrix corresponding to the case where node i 's clock ticks.

The following lemma discusses two important properties of the weight matrices.

Lemma 1: The weight matrices $\{W^{(i)} : i = 1, 2, \dots, N\}$ satisfy the following:

- $\mathbf{1}$ is a right eigenvector of all $W^{(i)}$, i.e., $W^{(i)}\mathbf{1} = \mathbf{1}$, $\forall i$;
- $\mathbf{1}^T$ is not a left eigenvector of any $W^{(i)}$, i.e., $\mathbf{1}^T W^{(i)} \neq \mathbf{1}^T$, $\forall i$.

Proof: Let us consider the first claim. It suffices to show that all rows of all $W^{(i)}$ matrices sum to unity. We have

$$\sum_{k=1}^N W_{jk}^{(i)} = \mathbf{1}\{j \in \mathcal{N}_i\}(\gamma + (1-\gamma)) + \mathbf{1}\{j \notin \mathcal{N}_i\}1 = 1 \quad (6)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Thus, the proof of first item is complete.

We now turn to the second claim. Note that

$$\sum_{j=1}^N W_{jk}^{(i)} = 1 + (1-\gamma)|\mathcal{N}_i|, \quad \text{for } k = i. \quad (7)$$

Since R is chosen to make the graph connected, we have $|\mathcal{N}_i| \geq 1$, which implies that $\sum_{j=1}^N W_{jk}^{(i)}|_{k=i} > 1$. This in turn shows that for all i there exists at least one column, namely the $k = i$ column, with sum different than one, which implies that $\mathbf{1}^T W^{(i)} \neq \mathbf{1}^T$ for all i . ■

The above Lemma reveals that $c\mathbf{1}$ for some $c \in \mathbb{R}$ is a fixed point of the broadcasting gossip algorithm, so $W^{(i)}c\mathbf{1} = c\mathbf{1}$ for all i . If the algorithm converges to a consensus, the network will not leave the consensus state. However, it also shows that the sum (and therefore the average) of the vector of node values is not preserved at each step.

Suppose node k is transmitting at time t . It is easy to check that the discrepancy between the sum at the next and current time-slots is nonzero whenever $x_k(t) \neq |\mathcal{N}_k|^{-1} \sum_{i \in \mathcal{N}_k} x_i(t)$. Suppose, that $x_k(t)$ is closer to the maximum of its neighbors than the minimum, or $|x_k(t) - \min_{i \in \mathcal{N}_k} x_i(t)| > |x_k(t) - \max_{i \in \mathcal{N}_k} x_i(t)|$. Then the sum difference between time-slots is bounded:

$$\left| \sum_{i=1}^N (x_i(t+1) - x_i(t)) \right| \leq (1-\gamma)|\mathcal{N}_k| \left| x_k(t) - \min_{i \in \mathcal{N}_k} x_i(t) \right|. \quad (8)$$

Clearly, the difference between the states sum at consecutive iterations is small if the node state values are close to each other.

Let us denote the mean of i.i.d. $W(t)$ as $\mathbb{E}\{W(t)\} = W$. The following lemma gives some properties of the average weight matrix that would prove useful for the remainder of the paper. Of note is that the following is a specific case of the general weight matrix given [28].

Lemma 2: The average weight matrix W is given by

$$W = I - \frac{1-\gamma}{N} \text{diag}\{\Phi\mathbf{1}\} + \frac{1-\gamma}{N} \Phi \quad (9)$$

and, for all γ , satisfies the following equation:

$$W\mathbf{1} = \mathbf{1}, \quad \mathbf{1}^T W = \mathbf{1}^T, \quad \rho(W - J) < 1 \quad (10)$$

where $\rho(\cdot)$ denotes the spectral radius of its argument and $J = (N)^{-1}\mathbf{1}\mathbf{1}^T$.

Proof: See Appendix A. ■

The Lemma shows that, unlike the individual weight matrices, $\mathbf{1}$ is both a left and right eigenvector of the average weight matrix. Moreover, the spectral radius of the weight matrix is less than unity, a property that will prove to be useful throughout the rest of the paper.

IV. CONVERGENCE OF BROADCAST GOSSIPING

In this section, we will study the convergence of the asynchronous broadcast gossip algorithms considering a slightly more general setting, where the consensus algorithm is governed by a product of identically distributed random matrices with the only restriction that each matrix is stochastic but not doubly stochastic. Of note is that the almost sure convergence result shown in this section, i.e., Theorem 1, is focused on and specific to broadcast gossip algorithms and is a special cases of the almost sure convergence result presented in [28].

A. Convergence in the Expectation

We consider the convergence in expectation of the broadcasting gossip algorithm. The next result reveals that, although the sum is not preserved per iteration, it is preserved in expectation. We consider the initial state as deterministic, and hence all expectations are averaging the mixing matrices only.

Proposition 1: The limiting random vector obtained through broadcast gossip iterations (if it exists) is, in expectation, equal to the average of initial node measurements, i.e.

$$\mathbb{E} \left\{ \lim_{t \rightarrow \infty} x(t) \right\} = \frac{1}{N} \mathbf{1} \mathbf{1}^T x(0). \quad (11)$$

Proof: By the Lebesgue dominated convergence theorem [30], we have

$$\mathbb{E} \left\{ \lim_{t \rightarrow \infty} x(t) \right\} = \lim_{t \rightarrow \infty} \mathbb{E} \{x(t)\}. \quad (12)$$

Moreover, since the matrices $W(t)$ are independent and identically distributed (i.i.d.), we have

$$\lim_{t \rightarrow \infty} \mathbb{E} \{x(t+1)\} = \lim_{t \rightarrow \infty} W^t x(0). \quad (13)$$

Thus it suffices to prove that $\lim_{t \rightarrow \infty} W^t = (N)^{-1} \mathbf{1} \mathbf{1}^T$. From [31], we know that this statement will hold if $W \mathbf{1} = \mathbf{1}$, $\mathbf{1}^t W = \mathbf{1}^T$, and $\rho(W - J) < 1$. Since Lemma 2 indicates that these conditions are satisfied, the proof is complete. ■

The proposition indicates that the expectation of the limiting random vector of the broadcasting gossip algorithm, given a certain initial state vector, is equal to the vector whose entries are equal to average of the initial states.

B. Convergence in the Second Moment

To study the convergence of the algorithm, we analyze the convergence of the vector $\beta(t)$ defined as the vector of deviations of the components of $x(t)$ from their average at the t th iteration. This can be expressed in component form as $\beta_i(t) = x_i(t) - \bar{x}(t)$, or as

$$\beta(t) = x(t) - Jx(t) = (I - J)x(t). \quad (14)$$

Let $\lambda_i(A)$ denote the i th largest eigenvalue of a matrix A . In the following, we present a sufficient condition guaranteeing the convergence of the expectation of the deviation vector norm to zero.

Lemma 3: The expectation of the norm of the deviation vector of the broadcast gossip, i.e., $\mathbb{E}\{\|\beta(t)\|_2^2\}$, converges to zero if

$$\lambda_1(\mathbb{E}\{W(t)^T(I - J)W(t)\}) < 1 \quad (15)$$

where I denotes the identity matrix.

Proof: Utilizing the properties of $W(t)$ matrices, we find that the deviation vector $\beta(t)$ obeys the following recursion with probability one:

$$\beta(t+1) = (W(t) - JW(t))\beta(t). \quad (16)$$

Note that this iteration is different from the one tracking the distance to initial node measurements average in gossip-based algorithms which preserve the sum, and this difference impacts all our proof methodologies.

Let $Y(t) = (W(t) - JW(t))$ so that $\beta(t+1) = Y(t)\beta(t)$. Now, taking the expected norm of $\beta(t+1)$ given $\beta(t)$ and using the fact that $\|u\|_2^2 = u^T u$ for $u \in \mathbb{R}^N$, yields

$$\begin{aligned} \mathbb{E}\{\|\beta(t+1)\|_2^2 | \beta(t)\} \\ = \beta(t)^T \mathbb{E}\{Y(t)^T Y(t)\} \beta(t) \end{aligned} \quad (17)$$

$$\leq \lambda_1(\mathbb{E}\{Y(t)^T Y(t)\}) \cdot \|\beta(t)\|_2^2 \quad (18)$$

where the last line follows from the fact that all $Y(t)^T Y(t)$ matrices are symmetric and the Rayleigh-Ritz theorem [32]. Then, repeatedly conditioning and using the linear iteration obtained above, we have

$$\mathbb{E}\{\|\beta(t)\|_2^2\} \leq \lambda_1^t(\mathbb{E}\{Y(t)^T Y(t)\}) \cdot \|\beta(0)\|_2^2. \quad (19)$$

Thus, now one can see that $\lim_{t \rightarrow \infty} \mathbb{E}\{\|\beta(t)\|_2^2\} = 0$ if $\lambda_1(\mathbb{E}\{Y(t)^T Y(t)\}) < 1$. Algebraic manipulations reduce this sufficient condition to the one stated in the Lemma. ■

It is important to emphasize that the Lemma 3 gives a sufficient condition for any consensus protocol that does not preserve network sum. Moreover, note that the condition $\lambda_1(\mathbb{E}\{W(t)^T(I - J)W(t)\}) < 1$ is different than the convergence condition obtained for the standard pairwise gossip algorithms where one only need to have $\lambda_2(\mathbb{E}\{W(t)\}) < 1$ to ensure the second-order convergence to the initial node measurements average [15], [21]. Note, however, that the sufficiency condition derived for the broadcast gossip algorithms reduces to the one for average-preserving gossip algorithms when $\mathbf{1}^T W^{(i)} = \mathbf{1}^T$, $\forall i$.

In the following, we show that the broadcasting gossip algorithm satisfies the sufficiency condition required to achieve consensus in the second moment.

Proposition 2: The broadcast gossip algorithms satisfies the fact that $\lambda_1(\mathbb{E}\{W(t)^T(I - J)W(t)\}) < 1$.

Proof: First, note that the eigenvalue of interest is the maximum eigenvalue of the expectation over positive semidefinite matrices since $(W^{(i)})^T(I - J)W^{(i)} = ((I - J)W^{(i)})^T((I - J)W^{(i)})$. This indicates that $\lambda_1(\mathbb{E}\{W(t)^T(I - J)W(t)\}) \geq 0$. Moreover, let $W' =$

$\mathbb{E}\{W(t)^T W(t)\}$ and $W'' = \mathbb{E}\{W(t)^T J W(t)\}$, and observe the following

$$\lambda_1(w' - W'') = \max_{\|u\|_2=1} u^T W' u - u^T W'' u \quad (20)$$

where the above follows from the variational definition of eigenvalues (note that $W' - W''$ is a symmetric matrix). Recall that $\max_{\|u\|_2} u^T W' u = 1$ for $u = u_1 = 1/\sqrt{N}\mathbf{1}$ which is the eigenvector corresponding to the unit eigenvalue. Of note is that for all $\{u : u \in \mathbb{R}^N, \|u\|_2 = 1, u \neq u_1\}$, we have $u^T W' u < 1$ which implies that $\lambda_1(\mathbb{E}\{W(t)^T (I - J)W(t)\}) < 1$. Since $u^T W'' u \geq 0$ for all $u \in \mathbb{R}^N$ (note that the expectation is taken over positive semidefinite matrices). Thus, the task reduces to show that for $u = u_1$, we still have $\lambda_1(W' - W'')$. For $u = u_1$, (20) reduces to

$$u_1^T W' u_1 - u_1^T W'' u_1 = 1 - u_1^T W'' u_1 \quad (21)$$

where the last inequality follows from the fact that $u_1^T W'' u_1 > 0$ since all entries of the W'' matrix is nonnegative (note that the expectation is taken over nonnegative entry matrices). Thus, $u^T W' u - u^T W'' u < 1$ for all $\{u : u \in \mathbb{R}^N, \|u\|_2 = 1\}$, indicating that $\max_{\|u\|_2=1} u^T W' u - u^T W'' u < 1$. Which, in turn, yields $\lambda_1(\mathbb{E}\{W(t)^T (i - j)W(t)\}) < 1$. ■

C. Almost Sure Convergence to Consensus

Given the results of Sections IV-A and IV-B, we are now in the position of stating our main result.

Theorem 1: The broadcast gossip algorithm converges to a consensus almost surely. That is

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (22)$$

for some random variable $c \in \mathbb{R}$ where

$$\mathbb{E}\{c\} = \frac{1}{N} \mathbf{1}^T x(0). \quad (23)$$

Proof: See Appendix B. ■

The theorem indicates that the broadcasting gossip algorithms achieve consensus with probability one, and the consensus value is, in expectation, equal to the desired value, i.e., average of initial nodes' measurements.

V. OPTIMAL MIXING PARAMETER

The sufficient condition in Lemma 3 depends on $\lambda_1(\mathbb{E}\{W(t)^T (I - J)W(t)\})$, which is the rate of convergence of an upper bound for the mean square deviation. Minimizing this parameter is a meaningful criterion of optimality when trying to approach convergence rapidly. In this section, we derive the optimal mixing parameter defined in (2) by minimizing $\lambda_1(\mathbb{E}\{W(t)^T (I - J)W(t)\})$ with respect to the mixing parameter in the broadcast gossip algorithm. At the same time, this allows us to study the effect of the graph connectivity and network size on the optimal mixing parameter. The following Lemma gives formulae which will be useful in the remaining analysis of the broadcast gossip algorithm.

Lemma 4: The following two formulas hold:

i) let $W' \triangleq \mathbb{E}\{W(t)^T W(t)\}$, then

$$W' = I - \frac{2\gamma(1-\gamma)}{N} (D - \Phi); \quad (24)$$

ii) let $W'' \triangleq \mathbb{E}\{W(t)^T J W(t)\}$, then

$$W'' = \frac{(1-\gamma)^2}{N^2} (D - \Phi)^2 + J \quad (25)$$

where $D = \text{diag}(\Phi\mathbf{1})$ is the diagonal matrix of node degrees.

Proof: See Appendix C. ■

Note that $\lambda_1(\mathbb{E}\{W(t)^T (I - J)W(t)\}) = \lambda_1(W' - W'')$ and recall that $\lambda_1(W' - W'')$ gives the worst-case convergence characteristic of the broadcast gossip algorithms, Lemma 3. Now, consider the matrix $W' - W''$ which, after using Lemma 4, reduces to

$$W' - W'' = I - J - \frac{2\gamma(1-\gamma)}{N} (D - \Phi) - \frac{(1-\gamma)^2}{N^2} (D - \Phi)^2. \quad (26)$$

First note that the vector $\mathbf{1}$ is an eigenvector of $W' - W''$ with eigenvalue 0. The vector $\mathbf{1}$ corresponds to the only nonzero eigenvalue of the matrix J and the only zero eigenvalue for the Laplacian matrix $L = D - \Phi$. Therefore the eigenvectors of $W' - W''$ are exactly the eigenvectors of $D - \Phi$, and the k th eigenvalue of $W' - W''$ for $k = 1, 2, \dots, N - 1$ is

$$\lambda_k(W' - W'') = 1 - \frac{2\gamma(1-\gamma)}{N} \lambda_{N-k}(L) - \frac{(1-\gamma)^2}{N^2} \lambda_{N-k}(L)^2. \quad (27)$$

Thus, the eigenvalue of interest can now be written as

$$\lambda_1(W' - W'') = 1 - \frac{2\gamma(1-\gamma)}{N} \lambda_{N-1}(L) - \frac{(1-\gamma)^2}{N^2} \lambda_{N-1}(L)^2 \quad (28)$$

where $L = D - \Phi$ denotes the Laplacian matrix of the graph. Of note is that $\lambda_{N-1}(L)$ is referred to as the *algebraic connectivity* of the graph [33].

In the following, we investigate the effect of the mixing parameter on the eigenvalue of interest which, as seen by Lemma 3, bounds the rate of convergence of the broadcast gossip algorithms.

Corollary 1: Let us introduce $\lambda_1(W' - W''; \gamma)$ to show the dependency of the eigenvalue of interest to the mixing parameter γ . Then the following statements hold:

- i) $\lambda_1(W' - W''; \gamma)$ is convex in γ .
- ii) The optimal mixing parameter, minimizing a worst-case convergence rate, is given by

$$\gamma^* = \frac{N - \lambda_{N-1}(L)}{2N - \lambda_{N-1}(L)}. \quad (29)$$

Proof: Let us first consider convexity of $\lambda_1(W' - W''; \gamma)$ of w.r.t. to γ . One can show that the first and second derivatives of $\lambda_1(W' - W''; \gamma)$ are given by

$$\begin{aligned} & \frac{\partial \lambda_1(W' - W''; \gamma)}{\partial \gamma} \\ &= \gamma \left(\frac{4}{N} \lambda_{N-1}(L) - \frac{2}{N^2} \lambda_{N-1}(L^2) \right) \\ & \quad - \frac{2}{N} \lambda_{N-1}(L) + \frac{2}{N^2} \lambda_{N-1}(L^2) \end{aligned} \quad (30)$$

and

$$\frac{\partial^2 \lambda_1(W' - W''; \gamma)}{\partial \gamma^2} = \frac{4}{N} \lambda_{N-1}(L) - \frac{2}{N^2} \lambda_{N-1}(L^2) \quad (31)$$

respectively. Recall that $\lambda_1(W' - W''; \gamma)$ is convex in γ if $\partial^2 \lambda_1(W' - W''; \gamma) / \partial \gamma^2 \geq 0$ for $\gamma \in (0, 1)$. Moreover, the eigenvalues of the Laplacian are nonnegative indicating that $\lambda_{N-1}(L) \geq 0$ and $\lambda_{N-1}(L^2) = \lambda_{N-1}^2(L)$. Then, from (31), it is easy to see that, $\partial^2 \lambda_{N-1}(L) / \partial \gamma^2 \geq 0$ if $0 \leq \lambda_{N-1}(L) \leq 2N$. But recall that $\lambda(L) \leq \min\{2 \max_i |\mathcal{N}_i|, N\}$.

Consider next the second claim of the Corollary. The optimal γ is clearly given by

$$\gamma^* = \arg \min_{\gamma} \lambda_1(W' - W''; \gamma). \quad (32)$$

which, since $\lambda_1(W' - W''; \gamma)$ is convex, is simply found by setting (30) to zero and solving it for γ . ■

Interestingly, the above Corollary indicates that the optimal mixing parameter depends on the graph for finite N . For large N we have the following result, whose proof is trivial.

Corollary 2: For graphs such that $\lambda_{N-1}(L) = \Theta(f(N))$ for some function $f(\cdot)$, with $\lim_{N \rightarrow \infty} f(N)/N = 0$, the optimal mixing parameter is given by

$$\lim_{N \rightarrow \infty} \gamma^* = \frac{1}{2}. \quad (33)$$

Hence, for large enough N and standard radius connectivity considerations for random geometric graphs (e.g., $R = \Theta(\sqrt{\log N/N})$ and $\lambda_{N-1}(L) = \Theta(\log N)$), the eigenvalue $\lambda_1(W' - W'')$ increases as $|\gamma - 1/2|$ increases. Therefore the worst-case convergence rate, characterized by $\lambda_1(W' - W'')$, decreases. In words, γ values that are closer to $1/2$ yield a faster worst-case convergence rate compared to the γ values closer to its boundaries, i.e., zero and one.

In the following, we investigate the effect of the graph Laplacian on the optimal mixing parameter.

Corollary 3: Let us introduce $\gamma^*(L) \triangleq \gamma^*$ to denote the dependency of the optimal γ on the graph Laplacian. Then, $\gamma^*(L)$ is monotonically decreasing function of $\lambda_{N-1}(L)$.

Proof: The proof simply follows by analyzing the first derivative of the optimal mixing parameter w.r.t. the parameter of interest, denoted as $\partial \gamma^*(L) / \partial \lambda_{N-1}(L)$, and showing that $\partial \gamma^*(L) / \partial \lambda_{N-1}(L) < 0$ for $\lambda_{N-1}(L) \leq \min\{2 \max_i |\mathcal{N}_i|, N\}$. ■

Thus, the above Corollary indicates that, as the graph connectivity increases, i.e., the eigenvalues of the Laplacian increases, the optimal mixing parameter tends to zero. This result matches the intuition. In fact, in a fully connected graph clearly $\gamma = 0$ would result in a *consensus* at the first iteration.

VI. PERFORMANCE ANALYSIS OF BROADCAST GOSSIP ALGORITHMS

While broadcast gossip algorithms do not preserve the network sum, they do compute a linear combination of the network states. We can define as *error* the deviation of the states from the average of the initial states and use the mean square error as a metric to evaluate the algorithm performance. Even though the average displacement does not give the complete probabilistic picture but lends insight to the average MSE performance of the algorithm. Probabilistic concentration results on a general class of such random consensus algorithms can be found in [28] and some results reported here to make the paper self-contained can be derived as special cases.

This section is dedicated to the derivation of the mean-square error performance of the broadcast gossip algorithm and to studying the mixing parameter effect on the mean-square error performance as well as the convergence. In particular, we prove an upper bound on the discrete time (or equivalently, number of clock ticks) required to get within ϵ of the consensus $c\mathbf{1}$, $c \in \mathbb{R}$. We also derive an upper bound on the limiting mean-square error performance. Finally, we examine the communication complexity of the broadcast gossip algorithms to achieve a certain distance to consensus.

A. Mean Square Error

Since, in general, the broadcast gossip algorithm does not converge to the initial node measurements average $(N)^{-1} \mathbf{1}^T x(0)$, it is of interest to consider the distance of the consensus value to $\bar{x}(0)$. In the remaining, we use

$$\alpha(t) = x(t) - Jx(0). \quad (34)$$

to denote the difference between the state vector at time step t and the average of initial node measurements.

Lemma 5: Let $\mathbb{E}\{\|\alpha(t)\|_2^2\}$ denote the mean square error at time step t . The following two statements hold:

i) the mean square error iteration obeys a recursion given as

$$\begin{aligned} \mathbb{E}\{\|\alpha(t+1)\|_2^2\} &\leq (1 - \lambda_2(W')) \mathbb{E}\{\|J\alpha(t)\|_2^2\} \\ &\quad + \lambda_2(W') \mathbb{E}\{\|\alpha(t)\|_2^2\}; \end{aligned} \quad (35)$$

ii) if $\nexists c \in \mathbb{R}$ such that $x(t) = c\mathbf{1}$ almost surely, then

$$\mathbb{E}\{\|\alpha(t+1)\|_2^2 | \alpha(t)\} < \|\alpha(t)\|_2^2. \quad (36)$$

Proof: See Appendix D. ■

The above Lemma reveals that the mean square error (MSE) conditioned on the current state is a *strictly decreasing* function of time and the strict inequality becomes equality when the nodes converge to consensus. In the following, we consider the limiting MSE behavior of the broadcast gossip algorithms.

Proposition 3: The limiting MSE of the broadcast gossip algorithms is upper bounded by

$$\lim_{t \rightarrow \infty} \mathbb{E}\{\|\alpha(t)\|_2^2\} \leq \|\alpha(0)\|_2^2 \left(1 - \frac{1 - \lambda_2(W')}{1 - \lambda_{N-1}^2(W - J)} \right). \quad (37)$$

Proof: See Appendix E. ■

As in the worst-case convergence-rate case, it is of interest to characterize the effect of the mixing parameter γ on the limiting MSE performance. This is considered in the following Corollary.

Corollary 4: Let $U_\infty(\gamma)$ be the upper-bound on the limiting MSE of the broadcast gossip iterations, given in Proposition 3, as a function of the mixing parameter γ . Then, the following statements hold:

- i) the boundary cases, i.e., $\gamma \rightarrow 0$ and $\gamma \rightarrow 1$, are given by

$$\lim_{\gamma \rightarrow 0} U_\infty(\gamma) = \|\alpha(0)\|_2^2 \quad (38)$$

and

$$\lim_{\gamma \rightarrow 1} U_\infty(\gamma) = \|\alpha(0)\|_2^2 \left(1 - \frac{\lambda_{N-2}(L)}{\lambda_1(L)}\right) \quad (39)$$

respectively;

- ii) $U_\infty(\gamma)$ is a monotonically decreasing function of γ ;
 iii) $U_\infty(\gamma)$, for $\gamma = \gamma^*$, is given by

$$U_\infty(\gamma^*) = \|\alpha(0)\|_2^2 \left(1 - C(L) \frac{\lambda_{N-2}(L)}{\lambda_1(L)}\right) \quad (40)$$

where

$$C(L) = \frac{2N - 2\lambda_{N-1}(L)}{4N - 2\lambda_{N-1}(L) - \lambda_1(L)}. \quad (41)$$

Proof: See Appendix F. ■

The Corollary indicates that the limiting MSE performance of the broadcast gossip algorithm decreases when γ is increasing. This is due to the fact that as γ approaches zero, the broadcasting nodes create a local dominance shifting a multitude of nodes away from the desired mean, whereas, for γ values closer to unity, the nodes receiving the broadcasted value adjust their own state only slightly, thereby changing minimally the network mean.

B. Communication Cost to Achieve Consensus

Generalizing the analysis done for standard sum preserving gossip-based averaging algorithms, we define the ϵ -converging time in the following.

Definition 1: Given $\epsilon > 0$, the ϵ -converging time is the earliest time at which the vector $x(k)$ is ϵ close to the normalized initial deviation with probability greater than $1 - \epsilon$:

$$T(N, \epsilon) = \sup_x \inf \left\{ t : \Pr \left\{ \frac{\|x(t) - Jx(t)\|_2}{\|x(0) - Jx(0)\|_2} \geq \epsilon \right\} \leq \epsilon \right\} \quad (42)$$

where $\|\cdot\|_2$ denotes the l_2 norm of its argument.

Before we move on to the main result of this section, we need the following Lemma giving the order of the eigenvalue of interest.

Lemma 6: For the broadcast gossip algorithm,

$$1 - O\left(\frac{\log^4 N}{N^2}\right) \leq \lambda_1(\mathbb{E}\{W^T(I - J)W\}) \leq 1 - \Omega\left(\frac{\sqrt{\log N}}{N^{5/2}}\right). \quad (43)$$

Proof: See Appendix G. ■

Unfortunately, the upper and lower bounds do not coincide—they differ (ignoring logarithmic terms) by a \sqrt{N} factor. It may be possible to tighten the upper bound by exploiting the fact that for a communication radius slightly larger than the threshold the random geometric graph is regular in an order sense with degree $\Theta(\log N)$ [22]. However, we do not pursue this here.

Given the convergence rate definition, we have the following rate of convergence to a consensus for the broadcast based average consensus.

Proposition 4: The ϵ -converging time of the asynchronous broadcast gossip algorithms is bounded by

$$\Pr \left\{ \frac{\|x(t) - Jx(t)\|_2}{\|x(0) - Jx(0)\|_2} \geq \epsilon \right\} \leq \epsilon \quad (44)$$

where

$$\Omega\left(\frac{N^2 \log \epsilon^{-1}}{\log^4 N}\right) = T(N, \epsilon) = O\left(\frac{N^{5/2} \log \epsilon^{-1}}{\sqrt{\log N}}\right). \quad (45)$$

Proof: At this stage of development, we just have to put the pieces together. Given the Definition 1 and the results of [15], using the Markov inequality and noting that:

$$\begin{aligned} \frac{0.5 \log \epsilon^{-1}}{\log \lambda_1^{-1} \mathbb{E}\{W(t) - J\}} &\leq T(N, \epsilon) \\ &\leq \frac{3 \log \epsilon^{-1}}{\log \lambda_1^{-1} (\mathbb{E}\{W(t)^T (I - J) W(t)\})}. \end{aligned} \quad (46)$$

Now, we have upper and lower bounds on λ_1 of the form $1 - \alpha$, so

$$\Omega\left(\frac{\log \epsilon^{-1}}{\frac{1}{2}(1 - \lambda_1)}\right) = T(N, \epsilon) = O\left(\frac{\log \epsilon^{-1}}{1 - \lambda_1}\right). \quad (47)$$

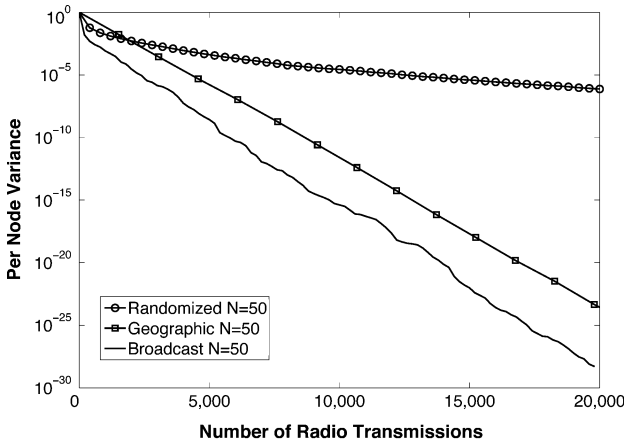
Substituting the bounds in Lemma 6 yields the result. ■

Moreover, note that if we set $\epsilon = 1/N^\alpha$ in the above equation, then we obtain $T(N, 1/N^\alpha) = \Omega(N^2 / \log^3 N)$. Since the number of transmissions per iteration is one in the broadcast gossip algorithms, this result matches also the communication complexity.¹ One can observe that broadcast gossip algorithms improve upon randomized gossip algorithms ($\Theta(N^2 \log N)$), but appears to be worse than the geographic gossip which has communication complexity in the order of $O(N^{3/2} \sqrt{\log N})$. As we will see very shortly through numerical examples, broadcast gossip significantly outperforms both algorithms for practical network sizes. This is interesting, because it illustrates how the constraint of maintaining the sum of the states constant does come with some performance penalty as well.

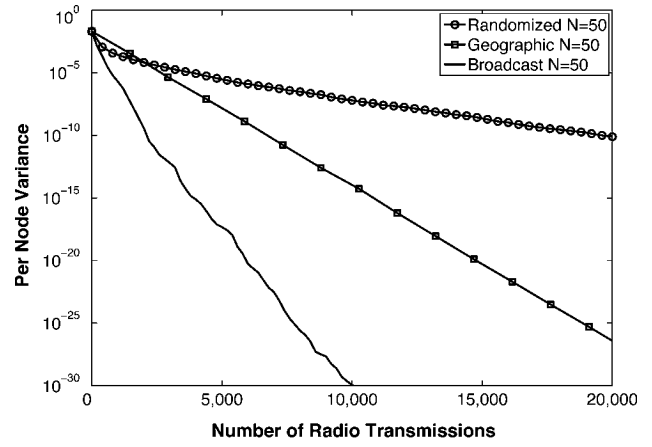
C. Performance Analysis: Numerical Examples

In the following, as in [34], we compare the number of *radio transmissions* to achieve a certain distance from consensus of broadcast gossiping. We choose $\gamma = 1/2$, since this value is the optimal value in terms of convergence speed and provides

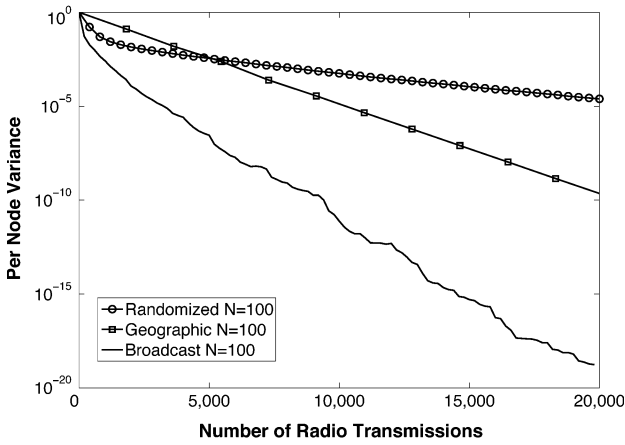
¹Note that larger connectivity radius implies, at the expense of larger broadcasting power, smaller $\lambda_1(\mathbb{E}\{W^T(I - J)W\})$ and, in turn, better convergence rate.



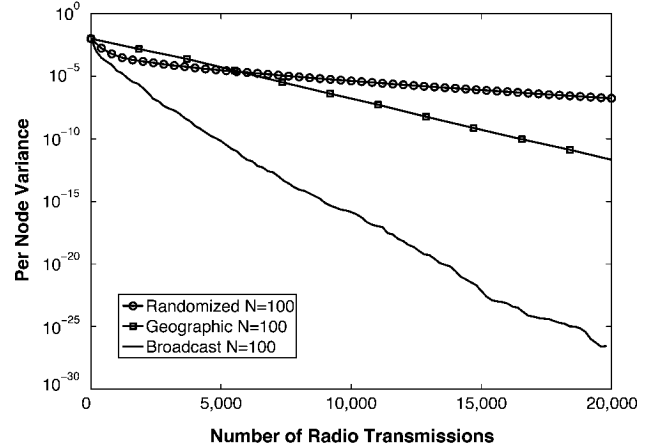
(a)



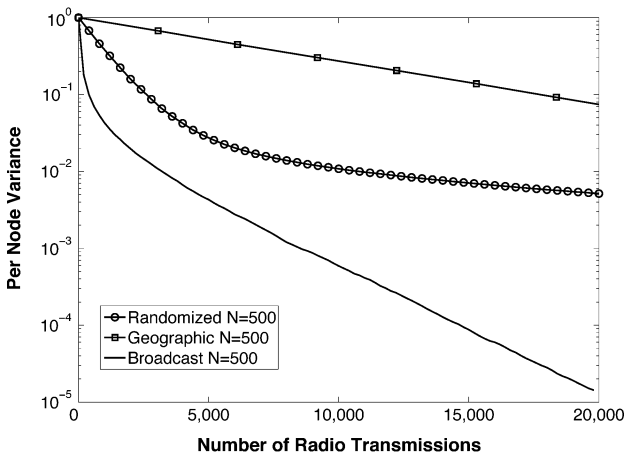
(a)



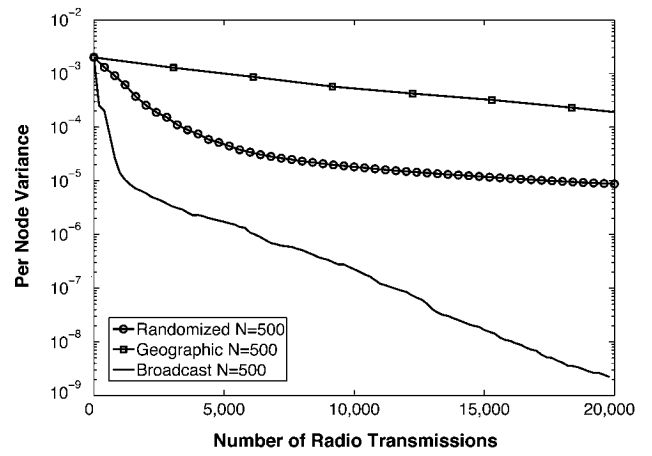
(b)



(b)



(c)



(c)

Fig. 1. Number of radio transmissions required to achieve a given distance (per node variance) from the consensus for $N \in \{50, 100, 500\}$ with initial node values uniformly distributed.

Fig. 2. Number of radio transmissions required to achieve a given distance (per node variance) from the consensus for $N \in \{50, 100, 500\}$ with initial node values zero except one in a single node.

a tradeoff for the MSE. To simulate the random geometric graph, we consider nodes that are uniformly distributed over a unit square. Their initial values are initialized as uniformly distributed random values with unit variance and zero mean. The connectivity radius is chosen as $R = \sqrt{\log(N)/N}$. Of note is that each iteration requires one, two and *the number of hops many* radio transmissions, respectively, for broadcast, standard and geographic gossiping.

The plots show the standard gossip algorithm [15], geographic gossip algorithm [34], and the broadcast gossip algorithm. Fig. 1(a)–1(c) depict per-node variance versus the number of radio transmissions for different network sizes (each data point is an ensemble average of 25 trials). Recall that the transmissions per iterations of randomized, geographic and broadcast gossip algorithms are, respectively equal to two, to the number of hops and to one. The simulation results suggest that broadcast gossiping reaches consensus faster than both

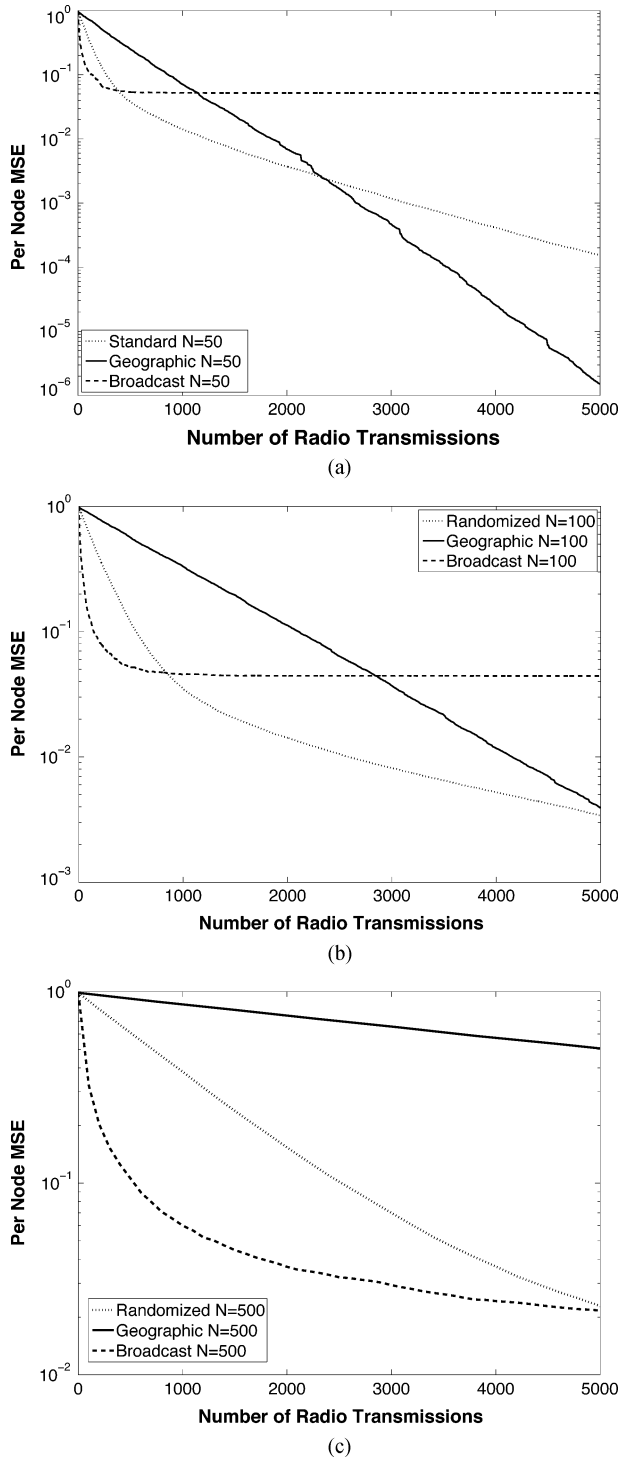


Fig. 3. The MSE performance of the randomized, geographic and broadcast gossip algorithms with respect to the number of radio transmissions for $N \in \{50, 100, 500\}$ with initial node values uniformly distributed.

competing protocols for an equal communication cost (cost that does not account for the extra complexity of routing two ways in geographic gossiping protocol). Indeed, this comparison is not entirely fair since the methods in [15] and [34] do meet an extra constraint; the comparison is useful since it highlights the non negligible penalty in speed that results from the extra constraint of maintaining the sum of states constant.

In Fig. 2(a)–2(c), we initialize the node values with zero except one in a single node following [34]. This is a field where

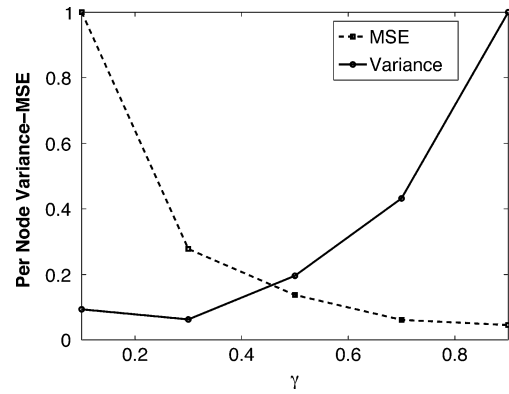


Fig. 4. The MSE and Variance performances of the broadcast gossip algorithms with respect to γ for $N = 500$ with initial node values uniformly distributed.

computing the average over the network is a harder task than the one considered before as in the previous case all the nodes are somewhat closer to the average. However, in this case, the information of the node containing the spike value needs to diffuse over the whole network. The network simulation setup is the same as above. The plot again illustrate the faster diffusion of consensus in broadcast gossiping compared to geographic and randomized gossiping protocols.

Next, we consider the MSE performance of the broadcast gossip algorithm versus the number of iterations and compare them to those of randomized and geographic gossip algorithms. Recall that the MSE of randomized and geographic gossip algorithms is zero in the limit, whereas the MSE of the broadcast algorithm saturates to a non-zero value as the algorithm converges to a consensus. The random geometric graph is simulated exactly as specified for the previous comparisons.

Fig. 3(a)–3(c) depict the MSE performance of the randomized, geographic and broadcast gossip algorithms versus the number of *radio transmissions* (two, number of hops many and one per iteration for randomized, geographic and broadcast gossip) for $N = \{50, 100, 500\}$, respectively. An interesting observation is that, for a reasonable number of radio transmissions, the MSE performance of the broadcast gossip algorithm is better than the randomized and geographic algorithms. However, as the number of radio transmissions increase, the randomized and geographic gossip, outperforms the broadcasting one, as they tend to zero whereas the performance of the broadcast gossip saturates to a nonzero value. Of note is that the crossover point where the randomized and geographic gossip starts to outperform the broadcast gossip increases with increasing number of nodes in the network.

These simulations results corroborate the theoretical analysis since they indicate that the MSE strictly decreases as long as consensus is not achieved. They also show that the faster convergence of broadcast gossip can be exploited to approximate quickly the desired average. In applications that are severely constrained in communication cost broadcast consensus may represent a good practical alternative to the competing methods that preserve the sum of the states.

In Fig. 4, we simulate the MSE and the speed of convergence performances of the algorithm with respect to the parameter γ for $N = 500$. Of note is that the simulation results are normalized by the largest corresponding output. Parallel to our theoretical findings in Corollary 4, the MSE of the algorithm mono-

tonically decreases with increasing γ . Moreover, as Corollary 1 suggests, we observe that per-node variance is convex in γ , i.e., the mixing parameter.² On the other hand, the theory suggests that the optimal mixing parameter γ (although smaller then) is in the neighborhood of 0.5 where the simulation results indicate that the optimal γ is around 0.3. We note that the theoretical results correspond to the optimization of an upper bound, therefore the value of the practical optimal γ may differ from the theoretical one presented here.

VII. CONCLUDING REMARKS

In this work we used the inherent broadcast nature of the wireless medium to present a simple “one-way” protocol that has good performance in simulations of networks with a modest number of nodes. The protocol simplifies the implementation of random gossiping compared to methods that require a pairwise node exchange. We presented the conditions on the weights matrices that would guarantee convergence to a consensus, and showed that the broadcast gossip algorithm achieves consensus with probability one. Moreover, the random consensus value is, in expectation, equal to the average of the initial node states. Although the network sum is not preserved at each broadcasting time-slot, we provided theoretical and simulation results on the mean square error performance in approximating such average. Finally, we presented theoretical and numerical examples evaluating and comparing the communication cost of gossiping algorithms required to achieve a given distance to consensus.

Even though the broadcast gossip algorithm shows promise in terms of convergence rate and MSE performance, the appropriate distributed averaging algorithm, e.g., randomized, geographic or broadcast, with the appropriate tuning parameters, depends ultimately on the application at hand which determines the relative importance of the implementation simplicity, convergence rate, MSE performance, or cap on the number of radio transmissions. Networks operating under adverse conditions may be prone to packet losses, node failures, and other events that may render pairwise exchange protocols and long-haul routing infeasible. Our analysis and method becomes particularly useful in these situations.

APPENDIX A

EXPECTED BROADCAST WEIGHT MATRIX

First we note that

$$W_{jk} = \frac{1}{N} \sum_{i=1}^N W_{jk}^{(i)}. \quad (48)$$

Then, by (5), we have

$$W_{jk} = \begin{cases} 1 - \frac{|\mathcal{N}_j|}{N} + \frac{\gamma|\mathcal{N}_j|}{N}, & k = j \\ 1 - \frac{1-\gamma}{N}, & k \in \mathcal{N}_j \\ 0, & \text{elsewhere.} \end{cases} \quad (49)$$

²It is of interest to note that similar tradeoff between the convergence rate and MSE performance (larger (smaller) convergence rate yields a larger (smaller) MSE) is also observed in the agreement problems with transmission noise case [35].

Therefore (9) follows. Note that (9) is the representation of the weight matrix in terms of graph Laplacian L , i.e., $W = I - \eta L$ where $L = \text{diag}\{\Phi \mathbf{1}\} - \Phi$ and $\eta = (1 - \gamma)/N$. Since $0 < \eta < 1/(N - 1)$ for all $\gamma \in (0, 1)$, W satisfies the conditions given in (10).

APPENDIX B

ALMOST SURE CONVERGENCE OF BROADCAST GOSSIP ALGORITHMS

At this stage of development, we just need to put the pieces together. First, we introduce a Lemma concerning convergence of random sequences that will prove useful to prove the above theorem.

Lemma 7 [36]: Consider a sequence of nonnegative random variables $\{V(t)\}_{t \geq 0}$ with $\mathbb{E}\{V(0)\} < \infty$. Let

$$\mathbb{E}\{V(t+1)|V(t), \dots, V(1), V(0)\} \leq cV(t) \quad (50)$$

where $0 < c < 1$. Then, $V(t)$ almost surely converges to zero, i.e.

$$\Pr \left\{ \lim_{t \rightarrow \infty} V(t) = 0 \right\} = 1. \quad (51)$$

We have almost sure convergence if $\mathbb{E}\{\|\beta(t+1)\|_2^2|\beta(t), \beta(t-1), \dots, \beta(0)\} \leq c\|\beta(t)\|_2^2$, for some $0 < c < 1$ since [7]

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = \Pr \left\{ \lim_{t \rightarrow \infty} \|\beta(t)\|_2^2 = 0 \right\}. \quad (52)$$

However, given $\beta(t)$, we have, as can be seen in the proof of Lemma 3, that

$$\begin{aligned} \mathbb{E}\{\|\beta(t+1)\|_2^2|\beta(t), \beta(t-1), \dots, \beta(0)\} \\ = \mathbb{E}\{\|\beta(t+1)\|_2^2|\beta(t)\}. \end{aligned} \quad (53)$$

From the proof of Lemma 3 and Proposition 2, we know that

$$\begin{aligned} \mathbb{E}\{\|\beta(t+1)\|_2^2|\beta(t)\} \\ \leq \lambda_1(\mathbb{E}\{W(t)(I - J)W(t)\})\|\beta(t)\|_2^2 \end{aligned} \quad (54)$$

where $0 < \lambda_1(\mathbb{E}\{W(t)(I - J)W(t)\}) < 1$. Thus, using the Lemma regarding the convergence of nonnegative random sequences, we have that $\Pr \left\{ \lim_{t \rightarrow \infty} \|\beta(t)\|_2^2 = 0 \right\} = 1$, completing the proof.

APPENDIX C

PROPERTIES OF HIGHER-ORDER MATRICES

Let $d_j = |\mathcal{N}_j|$ be the degree of node j . From the per-node weight matrices, we obtain

$$\{W^{(i)T}W^{(i)}\}_{jk} = \begin{cases} 1 + d_i(1 - \gamma)^2, & k = j = i \\ \gamma(1 - \gamma), & k \in \mathcal{N}_i, j = i \\ \gamma^2, & j \in \mathcal{N}_i, k = j \\ \gamma(1 - \gamma), & j \in \mathcal{N}_i, k = i \\ 1, & j \notin \mathcal{N}_i, k = j \\ 0, & \text{otherwise.} \end{cases} \quad (55)$$

where $W^{(i)}$ denotes the weight matrix corresponding to the case where node i 's clock ticks. Therefore, the average is

$$W'_{jk} = \begin{cases} 1 - \frac{d_j}{N} (1 - (1 - \gamma)^2 - \gamma^2), & k = j \\ \frac{2\gamma(1-\gamma)}{N}, & k \in \mathcal{N}_j \\ \frac{2\gamma(1-\gamma)}{N}, & j \in \mathcal{N}_k \\ 0, & \text{otherwise.} \end{cases} \quad (56)$$

Then, (24) follows. As we noted before, (24) is the representation of W' in terms of graph Laplacian $D - \Phi$. Since $0 < 2\gamma(1 - \gamma)/N < 1/N - 1$ for all γ , W' satisfies the properties given in (10).

Turning now to W'' , we first calculate

$$\left\{ \mathbf{1}^T W^{(i)} \right\}_j = \begin{cases} 1 & j \notin \mathcal{N}_i \cup \{i\} \\ 1 + d_i(1 - \gamma) & j = i \\ \gamma & j \in \mathcal{N}_i. \end{cases} \quad (57)$$

Then we find

$$\left(W^{(i)T} \mathbf{1} \mathbf{1}^T W^{(i)} \right)_{jk} = \begin{cases} 1, & j, k \notin \mathcal{N}_i \cup \{i\} \\ (1 + d_i(1 - \gamma))^2, & j = k = i \\ \gamma^2, & j, k \in \mathcal{N}_i \\ \gamma, & j \in \mathcal{N}_i, k \notin \mathcal{N}_i \text{ or } j \notin \mathcal{N}_i, k \in \mathcal{N}_i \\ \gamma(1 + d_i(1 - \gamma)), & j = i, k \in \mathcal{N}_i \text{ or } k = i, j \in \mathcal{N}_i \\ 1 + d_i(1 - \gamma), & j = i, k \notin \mathcal{N}_i \text{ or } k = i, j \notin \mathcal{N}_i. \end{cases} \quad (58)$$

Now we can take expectations. For each pair (j, k) we can calculate the expectation over i . Note that $(\Phi^2)_{jk}$ is the number of paths of length 2 from j to k or alternatively the number of i which are in the neighborhood of j and k . Consider first the case where $j \neq k$.

- There are $(\Phi^2)_{jk}$ values for i for which $j, k \in \mathcal{N}_i$.
- There are $d_j + d_k - 2(\Phi^2)_{jk}$ values for which $i \neq j, k$ and exactly one of j and k is in \mathcal{N}_i if j and k are not adjacent. There are $d_j + d_k - 2 - 2(\Phi^2)_{jk}$ values if j and k are adjacent.
- Note that $i = j$ once and $i = k$ once. Of the last two alternatives in (58), we have the former if $\Phi_{jk} = 1$ and the latter if $\Phi_{jk} = 0$.
- If j and k are not adjacent, there are $N - 2 - d_j - d_k + (\Phi^2)_{jk}$ values of i for which $j, k \notin \mathcal{N}_i \cup \{i\}$. If they are adjacent, there are $N - d_j - d_k + (\Phi^2)_{jk}$ such values

Then we have for $j \neq k$ that

$$\begin{aligned} N^2 W''_{jk} &= \gamma^2 (\Phi^2)_{jk} + \gamma (d_j + d_k - 2(\Phi^2)_{jk}) (1 - \Phi_{jk}) \\ &\quad + \gamma (d_j + d_k - 2 - 2(\Phi^2)_{jk}) \Phi_{jk} \\ &\quad + \gamma (2 + (1 - \gamma)(d_j + d_k)) \Phi_{jk} \\ &\quad + (2 + (1 - \gamma)(d_j + d_k)) (1 - \Phi_{jk}) \\ &\quad + (N - 2 - (d_j + d_k) + (\Phi^2)_{jk}) (1 - \Phi_{jk}) \\ &\quad + (N - d_j - d_k + (\Phi^2)_{jk}) \Phi_{jk} \\ &= ((1 - \gamma)^2 \Phi^2 + N^2 J \\ &\quad - (1 - \gamma)^2 (D\Phi + \Phi D))_{jk}, \end{aligned} \quad (59)$$

The matrix above has the following values on the diagonal:

$$(1 - \gamma)^2 d_j + N = (NI + (1 - \gamma)^2 D)_{jj}. \quad (60)$$

If $j = k$ then for 1 value of i we have $i = j = k$, for d_j values of i we have $j, k \in \mathcal{N}_i$, and for $N - d_j - 1$ values of i we have $j, k \notin \mathcal{N}_i \cup \{i\}$. Thus on the diagonal we should have

$$N^2 W''_{jj} = (1 + (1 - \gamma)d_j)^2 + \gamma^2 d_j + (N - 1 - d_j) \quad (61)$$

$$= (NI + (1 - \gamma)^2 D^2 + (1 - \gamma)^2 D)_{jj}. \quad (62)$$

So we must add the correction term $(1 - \gamma)^2 D^2$ to (59) to get the correct matrix

$$\begin{aligned} N^2 \mathbb{E} [W^T J W] &= (1 - \gamma)^2 \Phi^2 + N^2 J - (1 - \gamma)^2 \\ &\quad \times (D\Phi + \Phi D) + (1 - \gamma)^2 D^2 \\ &= (1 - \gamma)^2 (D - \Phi)^2 + N^2 J. \end{aligned} \quad (63)$$

APPENDIX D

CONVERGENCE OF THE MEAN SQUARE ERROR

It is easy to see that $\alpha(t + 1) = W(t)\alpha(t)$, yielding the following recursion for the second moment:

$$\begin{aligned} \mathbb{E} \{ \alpha(t + 1)^T \alpha(t + 1) | \alpha(t) \} \\ = \alpha(t)^T \mathbb{E} \{ W(t)^T W(t) \} \alpha(t) \end{aligned} \quad (64)$$

$$= \alpha(t)^T W' \alpha(t) \quad (65)$$

$$= y(t)^T \Lambda y(t) \quad (66)$$

where we utilize the eigendecomposition of $W' = V\Lambda V^T$ and define $y(t) = V^T \alpha(t)$. Given $\alpha(t)$, we can find $y(t)$, so we have the following:

$$\begin{aligned} \mathbb{E} \{ \alpha(t + 1)^T \alpha(t + 1) | \alpha(t) \} \\ = \sum_{i=1}^N \lambda_i(W') |y_i(t)|^2 \end{aligned} \quad (67)$$

$$= |y_1(t)|^2 + \sum_{i=2}^N \lambda_i(W') |y_i(t)|^2 \quad (68)$$

$$\begin{aligned} = (1 - \lambda_2(W')) |y_1(t)|^2 + \lambda_2(W') |y_1(t)|^2 \\ + \sum_{i=2}^N \lambda_i(W') |y_i(t)|^2 \end{aligned} \quad (69)$$

$$\leq (1 - \lambda_2(W')) |y_1(t)|^2 + \lambda_2(W') \sum_{i=1}^N |y_i(t)|^2 \quad (70)$$

$$= (1 - \lambda_2(W')) |y_1(t)|^2 + \lambda_2(W') \|y(t)\|_2^2 \quad (71)$$

$$= (1 - \lambda_2(W')) \|J\alpha(t)\|_2^2 + \lambda_2(W') \|\alpha(t)\|_2^2 \quad (72)$$

where the last line follows from the facts that $\|y(t)\|_2^2 = y(t)^T y(t) = \alpha(t)^T V V^T \alpha(t) = \alpha(t)^T \alpha(t) = \|\alpha(t)\|_2^2$ due to unitary decomposition and $|y_1(t)|^2 = |v_1^T \alpha(t)|^2 = (N^{-1} \alpha(t)^T \mathbf{1} \mathbf{1}^T \alpha(t)) = \alpha(t)^T J \alpha(t) = \|J\alpha(t)\|_2^2$ due to the relation $v_1 = (\sqrt{N})^{-1} \mathbf{1}$. This concludes the proof of the first item.

Let us now consider the second item. Note that J is a paracontracting matrix with respect to ℓ_2 norm since its symmetric and all its eigenvalues are in $(-1, 1]$. Thus, we have

$$Jx \neq x \Leftrightarrow \|Jx\|_2^2 < \|x\|_2^2. \quad (73)$$

Thus, if we can show that $J\alpha(t) = \alpha(t)$ if and only if $x(t) = c\mathbf{1}$ for some $c \in \mathbb{R}$, we are done. If $x(t) = c\mathbf{1}$, then,

$$J\alpha(t) = Jx(t) - Jx(0) = x(t) - Jx(0) = \alpha(t), \quad (74)$$

where we used the facts that $J^2 = J$ and $Jx(t)|_{x(t)=c\mathbf{1}} = x(t)$. Now, if $\alpha(t) = J\alpha(t)$, then $\alpha_i(t) = (N)^{-1} \sum_{i=1}^N \alpha_i(t)$. Thus, $\alpha(t) = \bar{\alpha}(t)\mathbf{1}$. Since

$$\begin{aligned} x(t) &= \alpha(t) + Jx(0) \\ &= \bar{\alpha}(t)\mathbf{1} + \bar{x}(0)\mathbf{1} = (\bar{\alpha}(t) + \bar{x}(0))\mathbf{1} \end{aligned} \quad (75)$$

we are done. Therefore, the proof of the second item is complete.

APPENDIX E LIMITING MEAN SQUARE ERROR

One can check that the following holds:

$$\mathbb{E}\{\|J\alpha(t)\|_2^2\} = \mathbb{E}\{\|\alpha(t)\|_2^2\} - \mathbb{E}\{\|\beta(t)\|_2^2\}. \quad (76)$$

Substituting the above into the claim (i) of Lemma 5, we obtain

$$\begin{aligned} \mathbb{E}\{\|\alpha(t+1)\|_2^2\} &\leq \mathbb{E}\{\|\alpha(t)\|_2^2\} \\ &\quad - (1 - \lambda_2(W'))\mathbb{E}\{\|\beta(t)\|_2^2\}. \end{aligned} \quad (77)$$

To upper bound the above, we need to lower bound $\mathbb{E}\{\|\beta(t)\|_2^2\}$ term. Utilizing the Jensen's inequality, we have

$$\mathbb{E}\{\|\beta(t)\|_2^2\} \geq \|\mathbb{E}\{\beta(t)\}\|_2^2 \quad (78)$$

$$= \|\mathbb{E}\{x(t) - Jx(t)\}\|_2^2 \quad (79)$$

$$= \|\{(W^t - J)x(0)\}\|_2^2 \quad (80)$$

where the last line follows from Lemma 2. Moreover, utilizing the properties of W , the above reduces to

$$\mathbb{E}\{\|\beta(t)\|_2^2\} \geq \|\{(W^t - J)\alpha(0)\}\|_2^2 \quad (81)$$

$$= \alpha(0)^T (W^t - J)^T (W^t - J) \alpha(0) \quad (82)$$

$$= \alpha(0)^T (W^{2t} - J)\alpha(0) \quad (83)$$

$$= \alpha(0)^T (W - J)^{2t} \alpha(0). \quad (84)$$

Note that $(W - J)$ is symmetric since W and J are symmetric. Thus, utilizing the unitary eigendecomposition, $(W - J) = U\Lambda U^T$, and defining $z(0) = U^T \alpha(0)$, we obtain the following:

$$\mathbb{E}\{\|\beta(t)\|_2^2\} \geq \alpha(0)^T U \Lambda^{2t} U^T \alpha(0) \quad (85)$$

$$= z(0)^T \Lambda^{2t} z(0) \quad (86)$$

$$= \sum_{i=1}^N \lambda_i^{2t} (W - J)|_{z_i(0)}|^2. \quad (87)$$

We need the following lemma before we continue the proof.

Lemma 8: All the eigenvalues of $W - J$ except zero are lower bounded by γ , i.e.,

$$\lambda_i(W - J) \geq \gamma \quad (88)$$

for all $i \in \{1, 2, \dots, N - 1\}$.

Proof: Note that the vector $\mathbf{1}$ is an eigenvector of $W - J = I - (1 - \gamma)/NL - J$ with eigenvalue 0. The vector $\mathbf{1}$

corresponds to the only nonzero eigenvalue of the matrix J and the only zero eigenvalue for the Laplacian matrix L . Therefore the eigenvectors of $W - J$ are exactly the eigenvectors of L , and the k th eigenvalue of $W - J$ for $k = 1, 2, \dots, N - 1$ is:

$$\lambda_k(W - J) = 1 - \frac{1 - \gamma}{N} \lambda_{N-k}(L) \geq \gamma \quad (89)$$

where the inequality follows from the fact that $\lambda(L) \leq N$. Thus, the proof is complete. \blacksquare

Since $\lambda(W - J) \geq 0$ by the above lemma, $\lambda_N(W - J) = 0$ and $z_N(0) = \mathbf{1}^T \alpha(0) = 0$ by construction, the above reduces to

$$\mathbb{E}\{\|\beta(t)\|_2^2\} \geq \sum_{i=1}^{N-1} \lambda_i^{2t} (W - J)|_{z_i(0)}|^2 \quad (90)$$

$$= \sum_{i=1}^N \lambda_i^{2t} (W - J)|_{z_i(0)}|^2 \quad (91)$$

$$\geq \lambda_{N-1}^{2t} (W - J) \sum_{i=1}^N |z_i(0)|^2 \quad (92)$$

$$= \lambda_{N-1}^{2t} (W - J) \|\alpha(0)\|_2^2. \quad (93)$$

Substituting (93) into (77) yields

$$\begin{aligned} \mathbb{E}\{\|\alpha(t+1)\|_2^2\} &\leq \mathbb{E}\{\|\alpha(t)\|_2^2\} \\ &\quad - (1 - \lambda_2(W')) \lambda_{N-1}^{2t} (W - J) \|\alpha(0)\|_2^2. \end{aligned} \quad (94)$$

Repeatedly utilizing the above, we obtain

$$\begin{aligned} \mathbb{E}\{\|\alpha(t)\|_2^2\} &\leq \|\alpha(0)\|_2^2 \left[1 - \left((1 - \lambda_2(W')) \sum_{j=0}^t \lambda_{N-1}^{2j} (W - J) \right) \right] \quad (95) \\ &= \|\alpha(0)\|_2^2 \left[1 - \left((1 - \lambda_2(W')) \frac{1 - \lambda_{N-1}^{2t+2} (W - J)}{1 - \lambda_{N-1}^2 (W - J)} \right) \right] \quad (96) \end{aligned}$$

where the second line uses the geometric series. Now, taking the limit of the above as t tends to infinity, we obtain the result of the proposition.

APPENDIX F CHARACTERIZATION OF LIMITING MSE PERFORMANCE

Let us first solve for the boundary cases. We consider the $\gamma \rightarrow 1$ case since $\gamma = 0$ simply follows by replacing $\gamma = 0$ in the $U_\infty(\gamma)$ expression (as is clear in the remainder of the proof). Thus, we have

$$\begin{aligned} \lim_{\gamma \rightarrow 1} U_\infty(\gamma) &= \|\alpha(0)\|_2^2 \lim_{\gamma \rightarrow 1} \left(1 - \frac{1 - \lambda_2(W')}{1 - \lambda_{N-1}^2 (W - J)} \right) \quad (97) \end{aligned}$$

$$= \|\alpha(0)\|_2^2 \left(1 - \lim_{\gamma \rightarrow 1} \frac{\frac{2\gamma(1-\gamma)}{N} \lambda_{N-2}(L)}{1 - \left(1 - \frac{1-\gamma}{N} \lambda_1(L)\right)^2} \right) \quad (98)$$

$$= \|\alpha(0)\|_2^2 \left(1 - \lim_{\gamma \rightarrow 1} \frac{2\gamma\lambda_{N-2}(L)}{2\lambda_1(L) - \frac{1-\gamma}{N}\lambda_1^2(L)} \right) \quad (99)$$

where the second line follows from the facts that $\lambda_2(W') = 1 - 2\gamma(1 - \gamma)/N\lambda_{N-2}(L)$ and $\lambda_{N-1}(W - J) = 1 - (1 - \gamma)/N\lambda_1(L)$ (see Appendix G) and the last line follows from expanding the square. Of note is that $\gamma = 0$ follows straight from (99) by replacing γ with zero. Note that for $\gamma = 1$, the limit of interest reduces to, after substituting $\gamma = 1$ in the limit expression, the expression stated in the Corollary.

Consider next the monotonicity of the upper-bound w.r.t. the mixing parameter. To prove this claim, we simply show that $\partial U_\infty(\gamma)/\partial\gamma < 0$ for $\gamma \in (0, 1)$. Differentiating $U_\infty(\gamma)$ w.r.t. γ and focusing on the numerator of the $\partial U_\infty(\gamma)/\partial\gamma$ expression (since denominator is always positive and does not effect the sign), after tedious algebraic steps, gives

$$\text{sgn} \left\{ \frac{\partial U_\infty(\gamma)}{\partial\gamma} \right\} = \text{sgn} \left\{ \frac{\lambda_1(L)}{N} - 2 \right\}. \quad (100)$$

where $\text{sgn}\{\cdot\}$ denotes the sign operator. Since $\lambda(L) < 2N$ for non-superconnected graphs, it is easy to see that $\partial U_\infty(\gamma)/\partial\gamma < 0$ for $\gamma \in (0, 1)$.

The last claim simply follows by substituting the optimal mixing parameter expression into the upper bound expression.

APPENDIX G

CHARACTERIZATION OF LARGEST EIGENVALUE

We begin by recalling the relation between the eigenvalue of interest and the eigenvalues of the graph Laplacian. We would like to calculate the eigenvalue $\lambda_1(W' - W'')$. The matrix $W' - W''$ is given by

$$W' - W'' = I - J - \frac{2\gamma(1 - \gamma)}{N}(D - \Phi) - \frac{(1 - \gamma)^2}{N^2}(D - \Phi)^2. \quad (101)$$

First note that the vector $\mathbf{1}$ is an eigenvector of $W' - W''$ with eigenvalue 0. The vector $\mathbf{1}$ corresponds to the only nonzero eigenvalue of the matrix J and the only zero eigenvalue for the Laplacian matrix $L = D - \Phi$. Therefore the eigenvectors of $W' - W''$ are exactly the eigenvectors of $D - \Phi$, and the k th eigenvalue $\lambda_k(W' - W'')$ for $k = 1, 2, \dots, N - 1$ is:

$$\lambda_k(W' - W'') = 1 - \frac{2\gamma(1 - \gamma)}{N}\lambda_{N-k}(L) - \frac{(1 - \gamma)^2}{N^2}\lambda_{N-k}(L)^2. \quad (102)$$

Thus to characterize $\lambda_1(W' - W'')$ we must characterize the second-smallest eigenvalue of the Laplacian matrix L . The number $\lambda_{N-1}(L)$ is sometimes called the algebraic connectivity of the graph.

An upper bound on $\lambda_{N-1}(L)$ will yield a lower bound on the largest eigenvalue of $W' - W''$. A result of Alon and Milman [33, Theorem 2.7] shows that $\lambda_{N-1}(L)$ is upper bounded by the following function of the diameter $\text{diam}(G)$ of the graph:

$$\lambda_{N-1}(L) \leq \frac{8d_{\max}}{\text{diam}(G)^2} \log_2^2 N. \quad (103)$$

If the communication radius is chosen large enough, for the random geometric graph with standard connectivity assumptions, $d_{\max} = \Theta(\log N)$ (see [22]). The diameter can be found as the number of hops to get from one corner to the diagonally opposite corner, so it is $\Theta(\sqrt{N/\log N})$. Thus the whole bound is

$$\lambda_{N-1}(L) = O\left(\frac{\log^4 N}{N}\right). \quad (104)$$

This gives the bound

$$\lambda_1(W' - W'') = 1 - O\left(\frac{\log^4 N}{N^2}\right). \quad (105)$$

To upper bound $\lambda_1(W' - W'')$ we need a nontrivial lower bound on $\lambda_{N-1}(L)$. A result of Mohar [37] gives this lower bound in terms of the diameter of the graph. Mohar's lower bound is

$$\lambda_{N-1}(L) \geq \frac{4}{N \cdot \text{diam}(G)}. \quad (106)$$

Therefore

$$\lambda_{N-1}(L) = \Omega\left(\frac{\sqrt{\log N}}{N^{3/2}}\right) \quad (107)$$

and

$$\lambda_1(W' - W'') = 1 - \Omega\left(\frac{\sqrt{\log N}}{N^{5/2}}\right). \quad (108)$$

REFERENCES

- [1] T. C. Aysal, M. E. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. 2008 IEEE Inf. Theory Workshop*, Porto, Portugal, May 2008.
- [2] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms: Design and analysis for consensus," in *Proc. 47th IEEE Conf. Decision Contr. (CDC 2008)*, Cancun, Mexico, Dec. 2008.
- [3] C. C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4753–4766, Nov. 2006.
- [4] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time delays," *IEEE Trans. Autom. Control*, vol. 49, pp. 1520–1533, Sep. 2004.
- [5] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, pp. 655–661, 2005.
- [6] T. Vicsek, A. Czirok, E. B. Jacob, I. Cohen, and O. Schochet, "Novel type of phase transitions in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, pp. 1226–1229, 1995.
- [7] Y. Hatano and M. Mesbahi, "Agreement over random networks," in *Proc. IEEE Conf. Decision Contr.*, Paradise Island, The Bahamas, Dec. 2004.
- [8] A. T. Salehi and A. Jadbabaie, "On consensus in random networks," in *The Allerton Conf. Commun., Contr., Comput.*, Allerton House, IL, Sep. 2007.
- [9] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, pp. 3315–3326, 2008.
- [10] N. Lynch, *Distributed Algorithms*. San Francisco, CA: Morgan Kaufmann, 1996.
- [11] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proce. IEEE/ACM Int. Symp. Inf. Process. Sensor Netw.*, Los Angeles, CA, Apr. 2005.
- [12] Y. Rabani, A. Sinclair, and R. Wanka, "Local divergence of markov chains and the analysis of iterative load-balancing schemes," in *Proc. IEEE Symp. Found. Comput. Sci.*, Palo Alto, CA, Nov. 1998.

- [13] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, pp. 988–1001, 2003.
- [14] J. Tsitsiklis, "Problems in Decentralized Decision Making and Computation," Ph.D. dissertation, Dept. Elec. Eng. Comput. Sci., Massachusetts Inst. Technol. (M.I.T.), Boston, MA, 1984.
- [15] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508–2530, Jun. 2006.
- [16] D. Kempe, A. Dobra, and J. Gehrke, "Computing aggregate information using gossip," in *Proc. Found. Comput. Sci.*, Cambridge, MA, Oct. 2003.
- [17] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," MIT, Cambridge, MA, LIDS Rep. 2274, 2007.
- [18] A. Nedic and A. Ozdaglar, "Convergence rate for consensus with delays," MIT, Cambridge, MA, LIDS Rep. 2774, 2007.
- [19] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Rates of convergence of distributed average consensus with probabilistic quantization," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sep. 2007.
- [20] T. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus using dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, pp. 4905–4918, 2008.
- [21] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, Mar. 2008.
- [22] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Allerton, IL, Sep. 2007.
- [23] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus using probabilistic quantization," in *Proc. IEEE Statist. Signal Process. Workshop*, Madison, WI, Aug. 2007.
- [24] M. E. Yildiz and A. Scaglione, "Differential nested lattice encoding for consensus problems," in *Proc. Inf. Process. Sensor Netw.*, Cambridge, MA, Apr. 2007.
- [25] D. Mosk-Aoyama and D. Shah, "Computing separable functions via gossip," in *Proc. 2006 ACM Symp. Principles of Distrib. Comput. (PODC)*, Denver, CO, Jul. 2006.
- [26] O. Ayaso, D. Shah, and M. Dahleh, "Lower bounds on information rates for distributed computation via noisy channels," in *Proc. 45th Allerton Conf. Commun., Contr., Comput.*, Allerton, IL, Sep. 2007.
- [27] O. Ayaso, D. Shah, and M. Dahleh, "Counting bits for distributed function computation," in *Proc. 2008 IEEE Symp. Inf. Theory (ISIT 2008)*, Toronto, ON, Canada, Jul. 2008.
- [28] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 634–649, 2008.
- [29] A. Giridhar and P. R. Kumar, "Towards a theory of in-network computation in wireless sensor networks," *IEEE Commun. Mag.*, vol. 44, pp. 98–107, Apr. 2006.
- [30] A. Browder, *Mathematical Analysis: An Introduction*. New York: Springer-Verlag, 1996.
- [31] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, pp. 65–78, 2004.
- [32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1987.
- [33] N. Alon and V. Milman, " λ_1 , isoperimetric inequalities for graphs, and superconcentrators," *J. Combin. Theory, Series B*, vol. 38, pp. 73–88, 1985.
- [34] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. Inf. Process. Sensor Netw.*, Nashville, TN, Apr. 2006.
- [35] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with communication channel noise and random link failures," in *Proc. 41st Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007.
- [36] B. Polyak, *Introduction to Optimization*. New York: Optimization Software, Inc., 1987.
- [37] B. Mohar, "The Laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*, Y. Alavi, G. Chartrand, O. Oellermann, and A. Schwenk, Eds. New York: Wiley, 1991, vol. 2, pp. 871–898.



Tuncer Can Aysal (S'05–M'08) received the B.E. degree (high honors) from Istanbul Technical University, Istanbul, Turkey, in May 2003, and the Ph.D. degree from the University of Delaware, Newark, in February 2007, both in electrical and computer engineering.

He held research position at the McGill University and Cornell University, Ithaca, NY, from March 2007 to September 2008. Currently, he is a research scientist with Intellp LLC. His research interests include distributed/decentralized signal processing, sensor networks, consensus algorithms, cognitive radio networks, spectrum sensing, and robust, nonlinear, statistical signal, and image processing.

Dr. Aysal was the recipient of the Competitive Graduate Student Fellowship in 2005, a Signal Processing and Communications Graduate Faculty Award in 2006 (award is presented to an outstanding graduate student in this research area), and a University Dissertation Fellowship in 2007 from the University of Delaware. He was also a Best Student Paper finalist at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2007 and his dissertation was nominated to the prestigious Allan P. Colburn prize.



Mehmet Ercan Yildiz (S'00) received the B.S. degree from the University of Wisconsin, Madison, in 2005, and the M.S. degree from Cornell University, Ithaca, NY, in 2007, both in electrical and computer engineering. He also received an additional major in mathematics and high honors in research from the University of Wisconsin.

He is currently with the Electrical and Computer Engineering Department, Cornell University. His research interests include decentralized signal processing, sensor networks, consensus algorithms, and

statistical signal processing.

Mr. Yildiz is a recipient of a NSF Undergraduate Research Scholarship.



Anand D. Sarwate (S'99–M'09) received S.B. degrees in electrical engineering and computer science and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and the M.S. and Ph.D. degrees in electrical engineering in 2005 and 2008, respectively, from the University of California, Berkeley.

He is currently a postdoctoral researcher at the Information Theory and Applications Center at the University of California, San Diego. His research interests include information theory, distributed signal processing, machine learning, communications, and randomized algorithms for communications and signal processing in sensor networks.

Dr. Sarwate received the Laya and Jerome B. Wiesner Student Art Award from MIT, and the Samuel Silver Memorial Scholarship Award and Demetri Angelakos Memorial Achievement Award from the EECS Department at University of California at Berkeley. He was awarded an NDSEG Fellowship from 2002 to 2005. He is a member of Phi Beta Kappa and Eta Kappa Nu.



Anna Scaglione (SM'09) received the "Laurea" (M.Sc. degree) in 1995 and the Ph.D. degree in 1999 from the University of Rome, "La Sapienza."

She is currently Associate Professor in Electrical and Computer Engineering at University of California at Davis, since July 2008. She was previously at Cornell University, Ithaca, NY, from 2001 and received tenure in 2006; prior to this she was Assistant Professor in the year 2000–2001, at the University of New Mexico. Her expertise is in the broad area of signal processing for communication systems. Her

current research focuses on cooperative networks and sensors' systems.

Dr. Scaglione is first author of the paper that received the 2000 IEEE Signal Processing Transactions Best Paper Award; she has also received the NSF Career Award in 2002 and she is corecipient of the Eilersick Best Paper Award (MILCOM 2005).