

# Greedy Gossip with Eavesdropping

Deniz Üstebay\*, Boris Oreshkin, Mark Coates, and Michael Rabbat

Department of Electrical and Computer Engineering

McGill University

3480 University St, Montréal, Québec, Canada

Email: {deniz.ustebay,boris.oreshkin}@mail.mcgill.ca,

{mark.coates,michael.rabbat}@mcgill.ca

## Abstract

This paper presents *greedy gossip with eavesdropping* (GGE), a novel randomized gossip algorithm for distributed computation of the average consensus problem. In gossip algorithms, nodes in the network randomly communicate with their neighbors and exchange information iteratively. The algorithms are simple and decentralized, making them attractive for wireless network applications. In general, gossip algorithms are robust to unreliable wireless conditions and time varying network topologies. In this paper we introduce GGE and demonstrate that greedy updates lead to rapid convergence. We do not require nodes to have any location information. Instead, greedy updates are made possible by exploiting the broadcast nature of wireless communications. During the operation of GGE, when a node decides to gossip, instead of choosing one of its neighbors at random, it makes a greedy selection, choosing the node which has the value most different from its own. In order to make this selection, nodes need to know their neighbors' values. Therefore, we assume that all transmissions are wireless broadcasts and nodes keep track of their neighbors' values by eavesdropping on their communications. We show that the convergence of GGE is guaranteed for connected network topologies. We also study the rates of convergence and illustrate, through theoretical bounds and numerical simulations, that GGE consistently outperforms randomized gossip and performs comparably to geographic gossip on moderate-sized random geometric graph topologies.

## Index Terms

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

## I. INTRODUCTION AND BACKGROUND <sup>1</sup>

Distributed consensus is recognized as a fundamental problem of distributed control and signal processing applications (see, e.g., [4]–[10] and references therein). The prototypical example of a consensus problem is computation of the *average consensus*: for a network of  $n$  nodes, initially each node has a scalar data value,  $y_i$ , and the goal is to find a distributed algorithm that asymptotically computes the average,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  at every node  $i$ . Such an algorithm can further be used for computing linear functions of the data and can be generalized for averaging vectorial data.

One of the algorithms proposed for solving the average consensus problem is distributed averaging [11]. In distributed averaging, every node in the network broadcasts its information at each iteration so that neighboring nodes can receive and use this information for their updates. Recently, gossip algorithms have gained attention for the computation of average consensus [10], [12]. In contrast to distributed averaging, gossip algorithms allow only two neighboring nodes to communicate and exchange information at each iteration. Restricting all information exchange to be local in this fashion is attractive from the point of view of simplicity and robustness (e.g., to changing topology and unreliable network conditions). However, with both distributed averaging and gossip, the speed of information diffusion across the network is slow for topologies used to model wireless mesh and sensor networks. The convergence analysis leads to similar results for both algorithms; for detailed analysis see [10]. The information at each node typically does not change much from iteration to iteration. Hence, the broadcast medium is not efficiently used.

In this paper we propose a new randomized gossip algorithm, *greedy gossip with eavesdropping* (GGE), for average consensus computation. Unlike previous randomized gossip algorithms, which perform updates completely at random, GGE takes advantage of the broadcast nature of wireless communications and implements a greedy neighbor selection procedure. We assume a broadcast transmission model such that all neighbors within range of a transmitting node receive the message. Thereby, in addition to keeping track of its own value, each node tracks its neighbors values by eavesdropping on their transmissions. At each iteration, one of the nodes is activated according to an asynchronous time model [13]. Then, the activated node uses this information to greedily choose the neighbor with which it will gossip, selecting the neighbor whose value is most different from its own. Accelerating convergence in this myopic way

<sup>1</sup>Portions of this work were presented in [1], [2], [3].

does not bias computation and does not rely on geographic location information, which may change in networks of mobile nodes.

Although GGE is a powerful yet simple variation on gossip-style algorithms, analyzing its convergence behavior is non-trivial. The main reason is that each GGE update depends explicitly on the values at each node (via the greedy decision of with which neighbor to gossip). Thus, the standard approach to proving convergence to the average consensus solution (i.e., expressing updates in terms of a linear recursion and then imposing properties on this recursion) cannot be applied to guarantee convergence of GGE. To prove convergence, we demonstrate that GGE updates correspond to iterations of a distributed randomized incremental subgradient optimization algorithm. Similarly, analysis of the convergence rate of GGE requires a different approach than the standard approach of examining the mixing time of a related Markov chain. We develop a bound relating the rate of convergence of GGE to the rate of standard randomized gossip. The bound indicates that GGE always converges faster than randomized gossip, a finding supported by simulation results. We also provide a worst-case bound on the rate of convergence of GGE. For other gossip algorithms the rate of convergence is generally characterized as a function of the second largest eigenvalue of a related stochastic matrix. In the case of GGE, our worst-case bound characterizes the rate of convergence in terms of a constant that is completely determined by the network topology. We investigate the behavior of this constant empirically for random geometric graph topologies and derive lower bounds that provide some characterization of its scaling properties.

### A. Background and Related Work

Randomized gossip was proposed in [12] as a decentralized asynchronous scheme for solving the average consensus problem. At the  $k$ th iteration of randomized gossip, a node  $s_k$  is chosen uniformly at random. It chooses a neighbor,  $t_k$ , randomly, and this pair of nodes “gossips”:  $s_k$  and  $t_k$  exchange values and perform the update  $x_{s_k}(k) = x_{t_k}(k) = (x_{s_k}(k-1) + x_{t_k}(k-1))/2$ , and all other nodes remain unchanged. One can show that under very mild conditions on the way a random neighbor,  $t_k$ , is drawn, the values  $x_i(k)$  converge to  $\bar{y}$  at every node  $i$  as  $k \rightarrow \infty$  [11]. Because of the broadcast nature of wireless transmission, other neighbors overhear the messages exchanged between the active pair of nodes, but they do not make use of this information in existing randomized gossip algorithms.

The convergence rate of randomized gossip is characterized by relating the algorithm to a Markov chain [12]. The mixing time of this Markov chain is closely related to the averaging time of the gossip algorithm, and therefore defines the rate of convergence. For certain types of graph topologies, the mixing times are small and convergence of the gossip algorithm is fast. For example, in the case of a complete

graph, the algorithm requires  $O(n)$  iterations to converge. However topologies such as random geometric graphs [14] or grids are more realistic for wireless applications. Boyd et al. [12] prove that for random geometric graphs, randomized gossip requires  $O(n^2/\log n)$  transmissions to approximate the average consensus well<sup>2</sup>.

Motivated by the slow convergence of randomized gossip, Dimakis et al. introduced *geographic gossip* in [10]. Geographic gossip enables information exchange over multiple hops with the assumption that nodes have the knowledge of their geographic locations and the locations of their neighbors. It has been shown that long-range information exchange improves the rate of convergence to  $O(n\sqrt{n/\log n})$  for random geometric graphs. However, geographic gossip involves overhead due to localization and geographic routing. Furthermore, the network needs to provide reliable two-way transmission over many hops. Otherwise, messages which are lost in transit will result in biasing the average consensus computation.

Recently, other fast gossiping algorithms have also been proposed. Most related is the work of Li and Dai [15], and Jung et al. [16]. Both approaches are based on directing the exchange of information across the network by constructing *lifted* Markov chains using knowledge of the geographic locations of nodes. As an extension to geographic gossip, Benezit et al. [17] have recently proposed averaging along paths, an algorithm which converges in  $O(n)$  transmissions. All of these approaches rely on geographic information and thus are not suitable to scenarios where nodes are mobile or location information is not available. The focus of our work is on providing fast and communication-efficient computation that exploits broadcast transmissions rather than geo-location information to gossip quickly.

Aysal et al. have proposed *broadcast gossip*, a consensus algorithm that also makes use of the broadcast nature of wireless networks [18], [19]. At each iteration, a node is activated uniformly at random to broadcast its value. All nodes within transmission range of the broadcasting node calculate a weighted average of their own value and the broadcasted value, and they update their local value with this weighted average. Broadcast gossip does not preserve the network average at each iteration. It achieves a low variance (i.e., rapid convergence), but introduces bias: the value to which broadcast gossip converges can be significantly different from the true average.

Sundhar Ram et al. introduced a general class of incremental subgradient algorithms for distributed optimization in [20]. In this study, the effects of stochastic errors (e.g., due to quantization) on the

<sup>2</sup>Throughout this paper, when we refer to randomized gossip we specifically mean the natural random walk version of the algorithm, where the node  $t_k$  is chosen uniformly from the set of neighbors of  $s_k$  at each iteration. For random geometric graph topologies, which are of most interest to us, Boyd et al. [12] prove that the performance of the natural random walk algorithm scales order-wise identically to that of the optimal choice of transition probabilities, so there is no loss of generality.

convergence of consensus-like distributed optimization algorithms are investigated. Convergence of their algorithm is guaranteed under certain conditions on the errors, but the convergence rates are not characterized.

Nedić and Ozdaglar have also developed a distributed form of incremental subgradient optimization that generalizes the consensus framework [21]. Our problem formulation is not as general as theirs, but with the specific formulation addressed in this paper we achieve stronger results. In particular, our cost function has a specific form and, by exploiting it, we are able to guarantee convergence to an optimal solution and obtain tight bounds on the rate of convergence as a function of the network topology.

### B. Paper Organization

The remainder of this paper is organized as follows. Section II introduces the formal definition of the algorithm. Section III presents two bounds; the first relates the performance of GGE to randomized gossip and indicates that GGE always outperforms randomized gossip, and the second is a worst-case upper bound on the rate of convergence of GGE in terms of a topology-dependent constant. Results from numerical simulations are presented in Section IV. Motivated by these results we provide a multi-hop extension to our algorithm in Section V. Section VI summarizes the contributions of the paper and discusses future work.

## II. GREEDY GOSSIP WITH EAVESDROPPING (GGE)

We consider a network of  $n$  nodes and represent network connectivity as a graph,  $G = (V, E)$ , with vertices  $V = \{1, \dots, n\}$ , and edge set  $E \subset V \times V$  such that  $(i, j) \in E$  if and only if nodes  $i$  and  $j$  directly communicate. We assume that communication relationships are symmetric and that the graph is connected. Let  $\mathcal{N}_i = \{j : (i, j) \in E\}$  denote the set of neighbors of node  $i$  (not including  $i$  itself). Each node in the network has an initial value  $y_i$ , and the goal of the gossip algorithm is to use only local broadcast exchanges to arrive at a state where every node knows the average  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . To initialize the algorithm, each node sets its gossip value to  $x_i(0) = y_i$ .

At the  $k$ th iteration of GGE, a node  $s_k$  is chosen uniformly at random from  $\{1, \dots, n\}$ . (This can be accomplished using the asynchronous time model described in [13], where each node “ticks” according to a Poisson clock with rate 1.) Then,  $s_k$  identifies a neighboring node  $t_k$  satisfying

$$t_k \in \arg \max_{t \in \mathcal{N}_{s_k}} \left\{ \frac{1}{2} (x_{s_k}(k-1) - x_t(k-1))^2 \right\}, \quad (1)$$

which is to say,  $s_k$  identifies a neighbor that currently has the most different value from its own. This choice is possible because each node  $i$  maintains not only its own local variable,  $x_i(k-1)$ , but also a

copy of the current values at its neighbors,  $x_j(k-1)$ , for  $j \in \mathcal{N}_i$ . When  $s_k$  has multiple neighbors whose values are all equally (and maximally) different from  $s_k$ 's, it chooses one of these neighbors at random. Then  $s_k$  and  $t_k$  perform the update

$$x_{s_k}(k) = x_{t_k}(k) = \frac{1}{2}(x_{s_k}(k-1) + x_{t_k}(k-1)), \quad (2)$$

while all other nodes  $i \notin \{s_k, t_k\}$  hold their values at  $x_i(k) = x_i(k-1)$ . Finally, the two nodes,  $s_k$  and  $t_k$ , broadcast these new values so that their neighbors have up-to-date information.

Note that one GGE iteration could be accomplished with just two transmissions since  $s_k$  already knows the values at its neighbors: one broadcast from  $s_k$  to  $t_k$  notifying it of the change and simultaneously announcing the new value to all of  $s_k$ 's neighbors, and one broadcast from  $t_k$  to its neighbors to echo the new value to them. However, in networks with unreliable transmission or systems where nodes periodically shut off their radios to conserve energy,  $s_k$  may miss some transmissions from its neighbors, and thus may not always have accurate information about their values. In this case, mistakes in calculation at  $s_k$  in the two-transmission scheme just described would introduce errors, biasing the consensus computation. To make our algorithm more robust and address the case where  $s_k$  does not precisely know the values at all neighbors, we assume a three-transmission version of our scheme throughout the rest of this paper: one transmission from  $s_k$  to  $t_k$  to initiate gossiping, one from  $t_k$  to its neighbors to inform them of the new value, and one from  $s_k$  to its neighbors to inform them of its new value. We comment further on this issue in Section IV-C.

#### A. Initialization

Calculating the greedy update in (1) requires nodes to know their neighbors' values. Similar to other randomized gossip algorithms, we assume that at the outset of gossip computation each node  $i$  has already discovered its neighbor set,  $\mathcal{N}_i$ , but it does not know its neighbors' values. Instead, these values are learned during an initialization phase. During the initialization phase, the node  $s_k$  that is activated at iteration  $k$  chooses  $t_k$  randomly from the subset of its neighbors whose values it does not know, rather than performing a GGE update. Since  $s_k$  and  $t_k$  broadcast their new values after averaging, the nodes in their neighborhoods overhear and acquire information accordingly. Once  $s_k$  has heard from all of its neighbors, the initialization process is complete for that particular node and it chooses  $t_k$  greedily via (1) for all subsequent iterations.

Figure 1 illustrates the effects of this initialization scheme relative to an idealized initialization scheme and a naïve initialization scheme. In the idealized scheme, each node clairvoyantly knows all of its

neighbors' initial values, and all nodes immediately commence with greedy updates. In the naïve scheme, before any node begins gossiping, all nodes broadcast once (without performing an update) to inform their neighbors of their starting value. The results shown correspond to a network of 200 nodes<sup>3</sup>. Observe that the proposed scheme and the naïve scheme attain similar performance, and both involve an overhead of roughly  $n = 200$  transmissions, which is not substantial relative to the total number of transmission required. We prefer the proposed initialization scheme to the naïve broadcast approach because it does not require any scheduling mechanism.

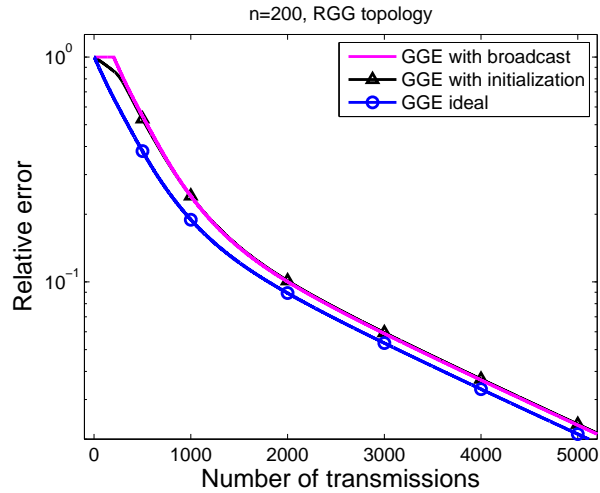


Fig. 1. A comparison of GGE performance for three different initializations of the algorithm. GGE with broadcast corresponds to the case where each node broadcasts its value once (without performing any updates) to initialize the values at all neighbors, before beginning to gossip. GGE with initialization corresponds to the proposed scheme, where, before a node has heard from all its neighbors, it gossips randomly with one it has not heard from, and after it has heard from all neighbors its performs greedy updates. Thus, the proposed GGE initialization scheme gives priority to learning neighbors' values during the initialization phase. GGE ideal assumes that each node clairvoyantly knows its neighbors' values at the outset. Note that GGE with broadcast and GGE with initialization give similar asymptotic performance (accruing overhead of roughly  $n = 200$  transmissions), although the proposed scheme does not require any scheduling mechanism and during early iterations the proposed scheme actually uses each transmission to perform an update.

<sup>3</sup>The topology is generated randomly from the family of random geometric graphs on 200 nodes with the standard connectivity radius. See Section IV for further details on simulations, and note that the setup used here is the same as that used to generate Figure 2(a). Although not reported here due to space constraints, we observe similar behavior on other network topologies.

### III. CONVERGENCE ANALYSIS

#### A. Convergence of GGE

To derive convergence results, we interpret GGE as a randomized incremental subgradient method [22]. Consider the constrained optimization problem,

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^n f_i(x) \\ \text{subject to} \quad & x \in X, \end{aligned}$$

where we assume that each  $f_i(x)$  is a convex function, not necessarily differentiable, and  $X$  is a non-empty convex subset of  $\mathbb{R}^n$ . An incremental subgradient algorithm for solving this optimization problem is an iterative algorithm of the form:

$$x(k) = \mathcal{P}_X[x(k-1) - \alpha_k g(s_k, x(k-1))], \quad (3)$$

where  $\alpha_k > 0$  is the step-size,  $g(s_k, x(k-1))$  is a subgradient<sup>4</sup> of  $f_{s_k}$  at  $x(k-1)$ , and  $\mathcal{P}_X[\cdot]$  projects its argument onto the set  $X$ . The algorithm is randomized when the component updated at each iteration,  $s_k$ , is drawn uniformly at random from the set  $\{1, \dots, n\}$ , and is independent of  $x(k-1)$ . Intuitively, the algorithm resembles gradient descent, except that instead of taking a descent step in the direction of the gradient of the cost function,  $f(x) = \sum_{i=1}^n f_i(x)$ , at each iteration we focus on a single component,  $f_i(x)$ . The projection,  $\mathcal{P}_X[\cdot]$ , ensures that each new iterate  $x(k)$  is feasible. Under mild conditions on the sequence of step sizes,  $\alpha_k$ , and on the regularity of each component function  $f_i(x)$ , Nedić and Bertsekas have shown that the randomized incremental subgradient method described above converges to a neighborhood of the global minimizer [22].

GGE is a randomized incremental subgradient algorithm for the problem

$$\min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^n \max_{j \in \mathcal{N}_i} \left\{ \frac{1}{2} (x_i - x_j)^2 \right\} \quad (4)$$

$$\text{subject to} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \quad (5)$$

<sup>4</sup>Subgradients generalize the notion of a gradient for non-smooth functions. The subgradient of a convex function  $f_i$  at  $x$  is any vector  $g$  that satisfies  $f_i(y) \geq f_i(x) + g^T(y-x)$ . The set of subgradients of  $f_i$  at  $x$  is referred to as the *subdifferential* and is denoted by  $\partial f_i(x)$ . If  $f_i$  is continuous at  $x$ , then  $\partial f_i(x) = \{\nabla f_i(x)\}$ ; i.e., the only subgradient of  $f_i$  at  $x$  is the gradient. A sufficient and necessary condition for  $x^*$  to be a minimizer of the convex function  $f_i$  is that  $0 \in \partial f_i(x^*)$ . See [22] and references therein.



where  $y_i$  is the initial value at node  $i$ . The objective function in (4) has a minimum value of 0 which is attained when  $x_i = x_j$  for all  $i, j$ . Thus, any minimizer is a consensus solution. Moreover, the constraint  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$  ensures that the unique global minimizer is the average consensus.

To connect the GGE update, (2), and the incremental subgradient update, (3), observe that  $g(k)$ , the subgradient of  $f_{s_k}(x(k-1)) = \max_{j \in \mathcal{N}_{s_k}} \{\frac{1}{2}(x_{s_k}(k-1) - x_j(k-1))^2\}$ , is defined by

$$g_i(k) = \begin{cases} x_{s_k}(k-1) - x_{t_k}(k-1) & \text{for } i = s_k, \\ -(x_{s_k}(k-1) - x_{t_k}(k-1)) & \text{for } i = t_k, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Here the subscript  $i$  indexes the components of the vector  $g$ . Fixing a constant step size  $\alpha_k = \frac{1}{2}$ , the update (3) is identical to (2). The recursive update for GGE thus has the form

$$x(k) = x(k-1) - \frac{1}{2}g(k), \quad (7)$$

Note that the projection is unnecessary since the choice  $\alpha_k = \frac{1}{2}$  ensures that the constraint  $\sum_{i=1}^n x_i(k) = \sum_{i=1}^n y_i$  is satisfied at each iteration.

Nedić and Bertsekas study the convergence behavior of randomized incremental subgradient algorithms in [22]. For a constant step size, their analysis only guarantees the iterates  $x(k)$  will reach a neighborhood of the optimal solution: with probability 1,  $\min_k f(x(k)) \leq \alpha n C^2 / 2$ , where  $C \geq \|g(s_k, x(k))\|$  is an upper bound on the norm of the subgradient [22]. We wish to show that  $x(k)$  actually converges to the average consensus,  $\bar{x}$ , the global minimizer of our optimization problem, and not just to a neighborhood of  $\bar{x}$ . By exploiting the specific form of the GGE cost function, we are able to prove the following stronger result.

*Theorem 1:* Let  $x(k)$  denote the sequence of iterates produced by GGE. Then  $x(k) \rightarrow \bar{x}$  almost surely.

*Proof:* To begin, we examine the improvement in squared error after one GGE iteration. Expanding  $x(k+1)$  via the expression (7), we have

$$\begin{aligned} \|x(k) - \bar{x}\|^2 &= \|x(k-1) - \frac{1}{2}g(k) - \bar{x}\|^2 \\ &= \|x(k-1) - \bar{x}\|^2 - \langle x(k-1) - \bar{x}, g(k) \rangle + \frac{1}{4}\|g(k)\|^2. \end{aligned}$$

Based on the definition of  $g(k)$  in (6), given  $s_k$  and  $t_k$ , we have

$$\|g(k)\|^2 = 2(x_{s_k}(k-1) - x_{t_k}(k-1))^2,$$

and,

$$\begin{aligned}\langle x(k-1) - \bar{x}, g(k) \rangle &= \sum_{i=1}^n (x_i(k-1) - \bar{x}_i) g_i(k) \\ &= (x_{s_k}(k-1) - x_{t_k}(k-1))^2.\end{aligned}$$

Therefore, we have

$$\|x(k) - \bar{x}\|^2 = \|x(k-1) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2 \quad (8)$$

with probability 1, since the expression holds independent of the value of  $s_k$  and  $t_k$ . Recursively applying our update expression, we find that w.p. 1,

$$\begin{aligned}\|x(k) - \bar{x}\|^2 &= \|x(k-1) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2 \\ &= \|x(k-2) - \bar{x}\|^2 - \frac{1}{4} \sum_{j=k-1}^k \|g(j)\|^2 \\ &\vdots \\ &= \|x(0) - \bar{x}\|^2 - \frac{1}{4} \sum_{j=1}^k \|g(j)\|^2.\end{aligned}$$

Since  $\|x(k) - \bar{x}\|^2 \geq 0$ , we must have

$$\sum_{j=1}^k \|g(j)\|^2 \leq 4 \|x(0) - \bar{x}\|^2$$

w.p. 1, and, consequently, the series  $\sum_{j=1}^k \|g(j)\|^2$  converges a.s. as  $k \rightarrow \infty$ . Since each term  $\|g(j)\|^2 \geq 0$ , this also implies that  $\|g(k)\|^2 \rightarrow 0$  a.s. as  $k \rightarrow \infty$ . However, by definition,  $g(k)$  is the subgradient of a convex function, and  $g(k) = 0$  is both a sufficient and necessary condition for  $x(k)$  to be a global minimizer. Thus,  $g(k) \rightarrow 0$  a.s. implies that  $x(k) \rightarrow \bar{x}$  a.s., since  $\bar{x}$  is the unique minimizer of (4)-(5). ■

### B. Convergence Rate: GGE vs. Randomized Gossip

The following theorem establishes a general expression for the bound on the mean-squared error of GGE after  $k$  iterations. Moreover, it demonstrates that the upper bound on the MSE of GGE is less than or equal to the upper bound on the MSE of randomized gossip.

The GGE updates can also be expressed in the form  $x(k) = W^{GGE}(k)x(k-1)$  where  $W^{GGE}(k)$  is a stochastic matrix with  $W_{s_k, s_k}^{GGE}(k) = W_{s_k, t_k}^{GGE}(k) = W_{t_k, s_k}^{GGE}(k) = W_{t_k, t_k}^{GGE}(k) = \frac{1}{2}$ ,  $W_{i,i}^{GGE}(k) = 1$  for all  $i \notin \{s_k, t_k\}$ , and 0 elsewhere. We denote the application of  $k$  successive GGE updates by

$W^{GGE}(1 : k) = \prod_{j=1}^k W^{GGE}(j)$ . Likewise, let  $W^{RG}(1 : k) = \prod_{j=1}^k W^{RG}(j)$  denote the successive application of  $k$  randomized gossip updates. Let  $\bar{W} = \mathbb{E}[W^{RG}(k)]$  denote the expected value of the randomized gossip matrix, and let  $\lambda_2(\bar{W})$  denote the second largest eigenvalue of  $\bar{W}$ .

*Theorem 2:* Let the algorithm input,  $x(0)$ , be given, let  $\bar{x}$  denote the corresponding average consensus vector. After  $k$  iterations, the expected mean squared error of GGE is upper bounded as follows:

$$\mathbb{E} [\|W^{GGE}(1 : k)x(0) - \bar{x}\|^2] \leq \|x(0) - \bar{x}\|^2 \prod_{i=1}^k (\lambda_2(\bar{W}) - \xi_i) \quad (9)$$

where  $\xi_i = 0$  if  $\mathbb{E} [\|W^{GGE}(1 : i - 1)x(0) - \bar{x}\|^2] = 0$ , and otherwise,

$$\xi_i = \frac{\mathbb{E} \left[ \sum_{s=1}^n \left( \max_{t \in \mathcal{N}_s} (x_s(i-1) - x_t(i-1))^2 \right) - \sum_{s=1}^n \left( \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s(i-1) - x_t(i-1))^2 \right) \right]}{2n \mathbb{E} [\|W^{GGE}(1 : i - 1)x(0) - \bar{x}\|^2]} \geq 0. \quad (10)$$

Note that  $x(i-1)$  is a random quantity determined by the choice of  $[s_1 \ s_2 \ \dots \ s_{i-1}]$ , and the expectation in (10) is over these variables.

*Remark 1:* The analogous expression for randomized gossip is simply [12]

$$\mathbb{E} [\|W^{RG}(1 : k)x(0) - \bar{x}\|^2] \leq \|x(0) - \bar{x}\|^2 \lambda_2(\bar{W})^k.$$

(Note that here, the expectation is taken with respect to both random nodes chosen at each iteration,  $[s_1 \ s_2 \ \dots \ s_k]$  and  $[t_1 \ t_2 \ \dots \ t_k]$ , whereas in the expressions in the theorem, the only randomness is in  $[s_1 \ s_2 \ \dots \ s_k]$ .) Since  $\xi_i \geq 0$  for all  $i = 1, \dots, k$ , this implies that the upper bound on GGE is uniformly upper bounded by the upper bound for randomized gossip, for any  $k \geq 0$  and any input  $x(0)$ . The upper bound for random gossip is tight; if  $v_2$  denotes the eigenvector corresponding to the second-largest eigenvalue of  $\bar{W}$ , then if  $x(0) = cv_2$  for some constant  $c$ , the upper bound holds with equality (in expectation).

*Remark 2:* The form of the terms  $\xi_i$  also provides insight into which scenarios are less favorable for GGE. In general, we know that randomized gossip is slow to converge on random geometric graphs [12], and so we hope that  $\xi_i > 0$  so that GGE achieves some improvement. Note that the numerator of  $\xi_i$  measures how much larger (on average) a GGE step from  $x(i-1)$  is in comparison to the step taken by randomized gossip from the same location. There are two scenarios where the expression for  $\xi_i$  in (10) evaluates to 0. The first is when  $x(i-1) = \bar{x}$ , in which case consensus has already been achieved. The second is when the difference between any two neighbors is constant across the network; i.e.,  $(x_s - x_t)^2 = c$  for all  $t \in \mathcal{N}_s$  and all  $s = 1, \dots, n$ . In this setting, being greedy does not provide any gain, since gossiping with any neighbor provides the same amount of immediate improvement.

*Proof of Theorem 2:* We recall the known convergence rate bounds for randomized gossip [12],

$$\mathbb{E} [\|W^{RG}(1:k)x(0) - \bar{x}\|^2] \leq \lambda_2 (\overline{W})^k \|x(0) - \bar{x}\|^2 \quad (11)$$

and the related recursive relationship,

$$\begin{aligned} & \mathbb{E} [\|W^{RG}(1:k)x(0) - \bar{x}\|^2] \\ &= \mathbb{E} \left[ \|W^{RG}(1:k-1)x(0) - \bar{x}\|^2 - \frac{1}{2n} \sum_{s=1}^n \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s(k-1) - x_t(k-1))^2 \right] \\ &\leq \lambda_2 (\overline{W}) \mathbb{E} [\|W^{RG}(1:k-1)x(0) - \bar{x}\|^2]. \end{aligned} \quad (12)$$

We can identify an equivalent relationship derived from applying  $k-1$  steps of GGE followed by one step of random gossip:

$$\begin{aligned} & \mathbb{E} [\|W^{RG}(k)W^{GGE}(1:k-1)x(0) - \bar{x}\|^2] \\ &= \mathbb{E} \left[ \|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2 - \frac{1}{2n} \sum_{s=1}^n \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s(k-1) - x_t(k-1))^2 \right] \\ &\leq \lambda_2 (\overline{W}) \mathbb{E} [\|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2]. \end{aligned} \quad (13)$$

With this relationship in hand, we can bound the error of the GGE algorithm by adding and subtracting the effects of making the  $k$ -th step a randomized gossip update:

$$\begin{aligned} & \mathbb{E} [\|W^{GGE}(1:k)x(0) - \bar{x}\|^2] \\ &= \mathbb{E} \left[ \|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2 - \frac{1}{2n} \sum_{s=1}^n \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s(k-1) - x_t(k-1))^2 \right] \\ &\quad - \mathbb{E} \left[ \frac{1}{2n} \sum_{s=1}^n \max_{t \in \mathcal{N}_s} (x_s(k-1) - x_t(k-1))^2 + \frac{1}{2n} \sum_{s=1}^n \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s(k-1) - x_t(k-1))^2 \right] \\ &\leq [\lambda_2 (\overline{W}) - \xi_k] \mathbb{E} [\|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2]. \end{aligned} \quad (14)$$

Repeated application of this inequality from  $i = 1, \dots, k$  yields the bound (9). ■

### C. GGE Convergence Rate: Worst Case Bounds

The previous subsection related the performance of GGE to that of standard randomized gossip. Next, we seek a more direct characterization of the GGE rate of convergence in terms of properties of the underlying communication topology. We then revisit our comparison to randomized gossip. The rate of convergence for gossip algorithms is typically quantified in terms of the  $\epsilon$ -averaging time,

$$T_{ave}(\epsilon) = \sup_{x(0) \neq 0} \inf \left\{ k : \Pr \left( \frac{\|x(k) - \bar{x}\|}{\|x(0) - \bar{x}\|} \geq \epsilon \right) \leq \epsilon \right\}.$$

Other gossip algorithms such as randomized gossip and geographic gossip are easily related to a homogeneous Markov chain, and  $T_{ave}(\epsilon)$  can be shown to scale as a function of the second largest eigenvalue of  $\overline{W}$ . In particular (see Theorem 3 in [12]),  $T_{ave}(\epsilon) \leq \frac{3 \log \epsilon^{-1}}{\log \lambda_2(\overline{W})^{-1}}$ . For randomized gossip, the matrix  $\overline{W}$  depends on the choice of probabilities assigned to each edge in the network and hence depends on the network topology.

Since, in each iteration of GGE, the greedy decision depends on the gossip values at each node,  $x(k)$ , our algorithm cannot be related to a homogeneous Markov chain ( $t_k$  depends on  $x(k)$ ). Consequently, the same machinery cannot be used to characterize the rate of convergence for GGE. The goal of this section is to bound the rate of convergence of GGE through alternative means. To this end, our main result is the following.

*Theorem 3:* Let  $G = (V, E)$  denote the graph on which we are gossiping, let  $x(k)$  denote the vector of GGE values after  $k$  iterations, and let  $\bar{x}$  denote the average vector. Then

$$\mathbb{E} [\|x(k) - \bar{x}\|^2] \leq A(G)^k \|x(0) - \bar{x}\|^2,$$

where  $A(G)$  is the graph-dependent constant defined as

$$A(G) = \max_{x \neq \bar{x}} \frac{1}{n} \sum_{s=1}^n \left( 1 - \frac{\|g_s(x)\|^2}{4\|x - \bar{x}\|^2} \right),$$

where  $g_s(x)$  refers to a subgradient of  $f_s(x)$ , when viewing GGE as an incremental subgradient algorithm<sup>5</sup>. Moreover, the  $\epsilon$ -averaging time for GGE is bounded above by

$$T_{ave}(\epsilon) \leq \frac{3 \log \epsilon^{-1}}{\log A(G)^{-1}}.$$

*Remark 3:* Note that the constant  $A(G)$  only depends on the topology of the graph. This constant plays a role for GGE similar to that played by the second-largest eigenvalue of  $\overline{W}$  for regular gossip algorithms.

*Proof of Theorem 3:* The proof of the first part of Theorem 3 is based on an approach introduced in [23] and developed in [24] for analyzing data-adaptive algorithms. We begin by recalling the recursion for the mean squared error of GGE after  $k$  iterations expressed in (8):

$$\begin{aligned} \|x(k) - \bar{x}\|^2 &= \|x(k-1) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2 \\ &= \left( 1 - \frac{\|g(k)\|^2}{4\|x(k-1) - \bar{x}\|^2} \right) \|x(k-1) - \bar{x}\|^2, \end{aligned}$$

<sup>5</sup>We explicitly note that this constant is a function of the underlying topology by writing  $A(G)$ . The constant  $A(G)$  is completely determined by the neighborhood structure of the network because the maximization is over all  $x$ . For a fixed  $x$ , the subgradients are determined by the neighborhood structure.

where  $g(k)$  denotes the subgradient at iteration  $k$  (when viewing GGE as a randomized incremental subgradient algorithm), and is a random quantity, depending on which node  $s(k)$  is activated at iteration  $k$ . Let  $M(k) = \|x(k) - \bar{x}\|^2$  denote the error after  $k$  iterations, and let  $N(k) = 1 - \frac{\|g(k)\|^2}{4\|x(k-1) - \bar{x}\|^2}$  denote the amount of contraction at iteration  $k$ . Using these definitions and some successive conditioning, we get

$$\begin{aligned}
\mathbb{E}[M(k)] &= \mathbb{E}[N(k)M(k-1)] \\
&= \mathbb{E}[\mathbb{E}[N(k)M(k-1)|x(k-1)]] \\
&= \mathbb{E}[M(k-1)\mathbb{E}[N(k)|x(k-1)]] \\
&\quad \vdots \\
&= M(0)\mathbb{E}[\mathbb{E}[N(1)|x(0)] \cdots \mathbb{E}[N(k)|x(k-1)]].
\end{aligned}$$

Note that  $A(G)$  is defined in such a way that  $\mathbb{E}[N(k)|x(k-1)] \leq A(G)$  for all  $k$ . Therefore, it follows that

$$\mathbb{E}[\|x(k) - \bar{x}\|^2] \leq A(G)^k \|x(0) - \bar{x}\|^2.$$

Next, we prove the second part of the claim: the bound on  $\epsilon$ -averaging time. To do this, we will use the bound we have just derived to develop an upper bound on  $\Pr(\|x(k) - \bar{x}\| \geq \epsilon \|x(0) - \bar{x}\|)$ , the probability that after  $k$  iterations we are still more than a factor of  $\epsilon$  away from the initial error. Applying Markov's inequality and the bound we just derived for  $\mathbb{E}[\|x(k) - \bar{x}\|^2]$ , we have

$$\begin{aligned}
\Pr(\|x(k) - \bar{x}\| \geq \epsilon \|x(0) - \bar{x}\|) &= \Pr(\|x(k) - \bar{x}\|^2 \geq \epsilon^2 \|x(0) - \bar{x}\|^2) \\
&\leq \frac{\mathbb{E}[\|x(k) - \bar{x}\|^2]}{\epsilon^2 \|x(0) - \bar{x}\|^2} \\
&\leq \epsilon^{-2} A(G)^k.
\end{aligned}$$

To get an upper bound on  $T_{ave}(\epsilon)$ , first note that  $\Pr(\|x(k) - \bar{x}\| \geq \epsilon \|x(0) - \bar{x}\|) \leq \epsilon$  provided that  $k \geq \frac{3 \log \epsilon^{-1}}{\log A(G)^{-1}}$ . Since in the first part of our proposition, the bound on  $\mathbb{E}[\|x(k) - \bar{x}\|^2]$  is based on a worst-case one-step analysis, it is an upper bound on the mean squared error at iteration  $k$ , effectively a lower bound on the rate of convergence. Therefore, we have an upper bound on the  $\epsilon$ -averaging time for GGE; that is  $T_{ave}(\epsilon) \leq \frac{3 \log \epsilon^{-1}}{\log A(G)^{-1}}$ . ■

Theorem 3 provides a direct link between the rate of convergence of GGE and the underlying network topology through the constant,  $A(G)$ . This motivates further study of how  $A(G)$  behaves for different classes and sizes of network topologies. Next, we derive a lower-bound on  $A(G)$  as a function of the

maximum degree of the network,  $d_{\max} = \max_i |\mathcal{N}_i|$ . Then we apply this result to characterize  $A(G)$  for two-dimensional grid and random geometric graph topologies.

*Theorem 4:* Let  $G$  be a graph with  $n$  nodes and maximum degree  $d_{\max}$ . As above, let  $\overline{W}$  denote the expected update for one step of randomized gossip on  $G$ . There exists a vector  $x \in \mathbb{R}^n$  with corresponding average consensus vector  $\bar{x}$  such that

$$\frac{\mathbb{E}\|W^{GGE}x - \bar{x}\|^2}{\|x - \bar{x}\|^2} \geq (1 - d_{\max}(1 - \lambda_2(\overline{W}))), \quad (15)$$

and this implies a lower-bound for  $A(G)$ ,

$$A(G) \geq 1 - d_{\max}(1 - \lambda_2(\overline{W})). \quad (16)$$

The proof appears below. We can use this result to relate the upper bounds on averaging time for GGE and standard randomized gossip. Let  $\mathcal{U}^{GGE}(G, \epsilon) = \frac{3 \log \epsilon^{-1}}{\log A(G)^{-1}}$  denote the upper bound on the averaging time of GGE obtained in Theorem 3, and let  $\mathcal{U}^{RG}(G, \epsilon) = \frac{3 \log \epsilon^{-1}}{\log \lambda_2(\overline{W})^{-1}}$  denote the corresponding upper bound on the averaging time of randomized gossip [12]. Using two inequalities for the logarithm — namely,  $a < -\log(1 - a)$  for  $a \in (0, 1)$ , and  $\log \lambda \leq \lambda - 1$  for  $\lambda > 0$  — we obtain

$$\mathcal{U}^{GGE}(G, \epsilon) \geq \frac{3 \log \epsilon^{-1}}{d_{\max}(1 - \lambda_2(\overline{W}))} \geq \mathcal{U}^{RG}(G, \epsilon)/d_{\max}. \quad (17)$$

In words, the upper bound on the averaging time of GGE is at most a factor of  $d_{\max}$  better than the upper bound for randomized gossip. Of course, this only links the upper bounds of the two algorithms and does not directly relate their actual performance. However, simulation results presented in the next section indicate that this relationship indeed captures the improvements seen for GGE over randomized gossip. Moreover, recall that the bounds on the expected improvement after a single gossip iteration are tight for both GGE and randomized gossip. The bound for GGE in Theorem 3 becomes an equality for  $k = 1$  when  $x(0)$  is taken to be the  $x$  that solves the optimization problem defining  $A(G)$ . Similarly, the bound for randomized gossip becomes an equality when  $x(0)$  is taken to be the eigenvector corresponding to the second largest eigenvalue of  $\overline{W}$ .

We are interested in understanding the performance of GGE for applications primarily in wireless networks. Random geometric graphs, first introduced in [14], are commonly used to model connectivity in wireless networks for the purpose of analyzing scaling behaviour of algorithms. A random geometric graph on  $n$  nodes is obtained by uniformly assigning each node i.i.d. coordinates in the unit square and then connecting nodes whose distance is less than connectivity radius  $r(n)$ . In this paper we adopt the common scaling  $r(n) = \sqrt{\frac{2 \log n}{n}}$ , which guarantees the network is connected with high probability [14].

For a random geometric graph with  $n$  nodes, it is known [12] that, for  $r(n)$  as given above, every node has  $2\pi \log(n)(1 - o(1))$  neighbors with high probability. Thus, with  $d_{\max} = 2\pi \log(n)(1 - o(1))$ , we see that GGE gives essentially a factor of  $\log n$  improvement in averaging time over randomized gossip. For a two-dimensional grid  $d_{\max} = 4$ , and GGE gives only a constant improvement in averaging time. These results are illustrated via simulation in the next section.

*Proof of Theorem 4:* Our starting point is (10) from Theorem 2. Since we are focusing on the effect of applying a single gossip iteration, we drop the time index  $i$  in (10) to simplify the notation:

$$\xi = \frac{\mathbb{E} \left[ \sum_{s=1}^n \left( \max_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right) - \sum_{s=1}^n \left( \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right) \right]}{2n \mathbb{E} [\|x - \bar{x}\|^2]} \geq 0, \quad (18)$$

Using the fact that the maximum of a set of non-negative values is always less than or equal to the sum of those values, we can write

$$\mathbb{E} \left[ \max_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right] \leq \frac{|\mathcal{N}_s|}{|\mathcal{N}_s|} \mathbb{E} \left[ \sum_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right] \quad (19)$$

$$\leq \frac{d_{\max}}{|\mathcal{N}_s|} \mathbb{E} \left[ \sum_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right]. \quad (20)$$

Therefore we can upper bound  $\xi$  by

$$\xi \leq \frac{(d_{\max} - 1) \mathbb{E} \left[ \sum_{s=1}^n \left( \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right) \right]}{2n \mathbb{E} [\|x - \bar{x}\|^2]}. \quad (21)$$

Next, take  $x$  to be the eigenvector corresponding to the second largest eigenvalue of  $\bar{W}$  and equate  $W^{RG}(1 : k - 1)x(0) = x$  in (12) to get

$$\mathbb{E} \left[ -\frac{1}{2n} \sum_{s=1}^n \frac{1}{|\mathcal{N}_s|} \sum_{t \in \mathcal{N}_s} (x_s - x_t)^2 \right] \leq -(1 - \lambda_2(\bar{W})) \mathbb{E} [\|x - \bar{x}\|^2]. \quad (22)$$

Applying this inequality in (21) gives

$$\xi \leq (d_{\max} - 1)(1 - \lambda_2(\bar{W})). \quad (23)$$

Observe that the one-step bound obtained by taking  $k = 1$  in (9) is tight. In particular, for our choice of  $x$  as the eigenvector corresponding to the second largest eigenvalue of  $\bar{W}$ , we have equality in (9):

$$\mathbb{E} [\|W^{GGE}x - \bar{x}\|^2] = \|x - \bar{x}\|^2 (\lambda_2(\bar{W}) - \xi). \quad (24)$$

Inserting (23) into (24) leads to the first claim in the proposition. Then, combining this bound with the first inequality in Theorem 3 for the case  $k = 1$  and  $x(0) = x$  yields the desired lower bound on  $A(G)$ . ■



#### IV. NUMERICAL SIMULATIONS

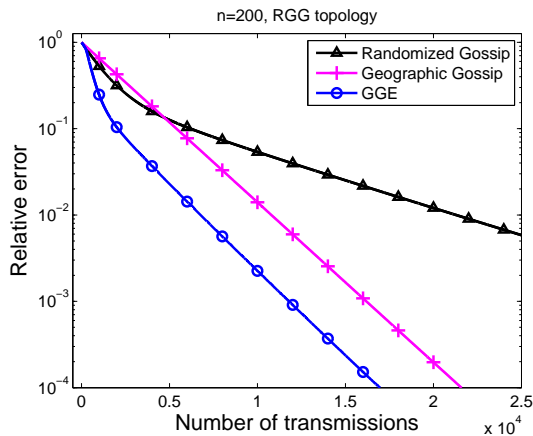
In this section we report the results of simulations conducted to compare the performance of GGE with randomized gossip [12] and geographic gossip [10] for a variety of initial conditions. We also compare the empirically achieved convergence rates to the bound established in Section III-C and investigate how this bound behaves as the number of nodes in the network grows.

##### A. Comparison of Convergence Rates

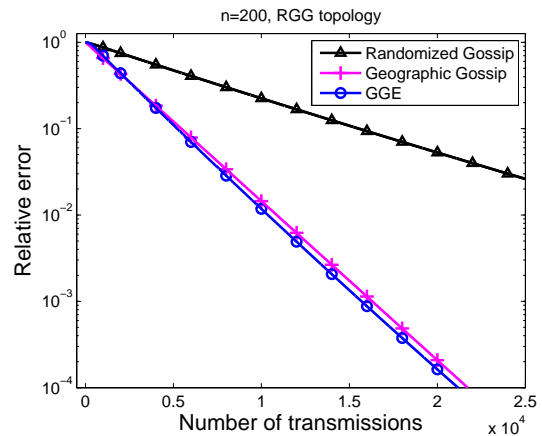
We first compare the convergence rates of GGE with randomized gossip and geographic gossip by examining the reduction they achieve in relative error,  $\frac{\|x(k) - \bar{x}\|}{\|x(0) - \bar{x}\|}$ , as a function of the number of transmissions (communication complexity). Since the number of transmissions per iteration is different for each algorithm, this is a fairer comparison than examining convergence rate relative to the number of iterations. Randomized gossip requires two wireless transmissions per iteration, GGE requires three transmissions (see the discussion in Section II), and geographic gossip has a variable number of transmissions per iteration, depending on the number of hops between the gossiping nodes. We simulate networks with random geometric graph topologies, and all figures show averages over 100 realizations of the random geometric graph. We examine performance for four different initial conditions,  $x(0)$ , in order to explore the impact of the initial values on performance. The first two of these cases are a Gaussian bumps field, and a linearly-varying field. For these two cases, the initial value  $x(0)$  is determined by sampling these fields at the locations of the nodes. The remaining two initializations consist of the ‘‘spike’’ signal, constructed by setting the value of one random node to 1 and all other node values to 0, and a random initialization where each value is i.i.d. drawn from a Gaussian distribution  $\mathcal{N}(0,1)$  of zero mean and unit variance. The first three of these signals were also used to examine the performance of geographic gossip in [10].

Figs. 2(a)-(d) show that GGE converges to the average at a much faster rate (both initially and asymptotically) than randomized gossip for all initial conditions. The initial rate of convergence of GGE is faster than geographic gossip for all but the linearly-varying field, and for the simulated network size ( $n = 200$ ), the asymptotic rates of reduction in relative error are similar for the two algorithms. Out of these candidate initializations, the linearly-varying field is the worst case. This is not surprising since the convergence analysis conducted in Section III suggests that constant differences between neighbors causes GGE to provide minimal gain.

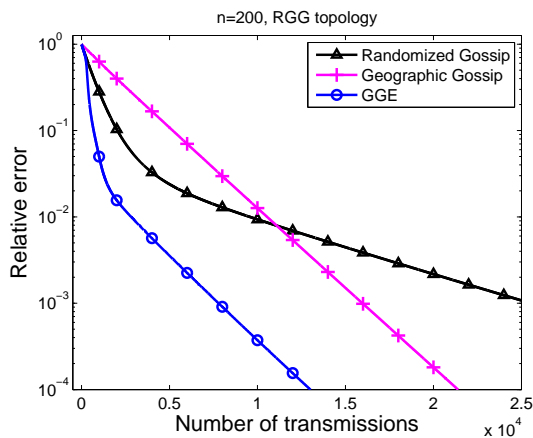
We also compare the performances of the three gossip algorithms for the grid topology. Figure 3 shows that in grid-structured networks the performance of GGE is close to the performance of randomized gossip



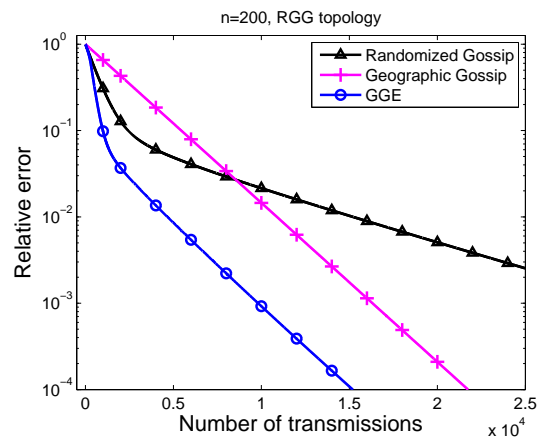
(a) Gaussian bumps convergence rate comparison



(b) Linearly-varying field convergence rate comparison



(c) Spike convergence rate comparison



(d) Uniform random field convergence rate comparison

Fig. 2. A comparison of the performance of randomized gossip, GGE, and geographic gossip for four initializations of  $x(0)$ . Results are averaged over 100 realizations of the random geometric graph and 100 runs of the algorithm per graph.

(constant improvement). Clearly geographic gossip has the best performance in this case. As discussed in Section III-C, the small number of neighbors in the grid topology restricts the improvement that GGE can achieve relative to randomized gossip.

For random geometric graph topologies, the expected node degree,  $\mathbb{E}[|\mathcal{N}_s|]$ , scales as  $\log n$  in contrast to constant average node degree in the case of grid topology. Therefore, GGE is able to provide better performance results compared to randomized gossip for graph topologies where average node degree increases with the number of nodes. To improve GGE performance for topologies with low expected node degree values, we propose an extension to the algorithm. Details of this extension, which we call Multi-hop GGE, are provided in Section V.

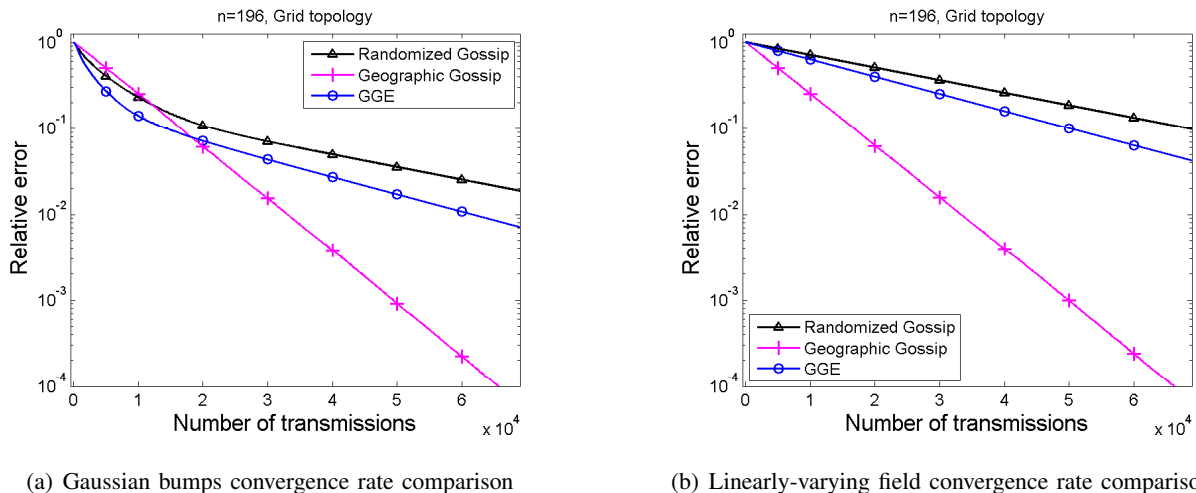


Fig. 3. A comparison of the performance of randomized gossip, GGE, and geographic gossip for two initializations of  $x(0)$  in grid topology. Results are averaged over 100 runs of the algorithm.

### B. Comparison with the Theoretical Upper Bound

We now compare the empirical average relative error for the random geometric graph with the bound developed in Theorem 3. There is no closed-form solution for  $A(G)$ , so we solve the optimization problem identified in Theorem 3 numerically. Since the cost function can be expressed as a function of  $(x_i - \bar{x})/\|x - \bar{x}\|$  where  $i$  is one of the nodes, without loss of generality we can maximize over the set  $\mathcal{X} = \{x \in \mathbb{R}^n : \bar{x} = 0, \|x\| = 1\}$ . Furthermore, we can simplify the objective function and calculate  $A(G) = 1 - \frac{1}{2n} \min_{x \in \mathcal{X}} \sum_{s=1}^n \max_{t \in \mathcal{N}_s} (x_s - x_t)^2$ . To solve this minimization problem, we use a randomized projected incremental subgradient algorithm with updates  $x^{(k+1)} = \mathcal{P}_{\mathcal{X}}[x^{(k)} - \mu \tilde{g}_{i(k)}(x)]$ , where  $\tilde{g}_i(x)$  is a subgradient of  $\max_{j \in \mathcal{N}_i} (x_i - x_j)^2$ ,  $\mu > 0$  is the step-size, and the random variables  $i(k)$  are i.i.d., drawn uniformly from the integers  $1, \dots, n$ . The projection,  $x = \mathcal{P}_{\mathcal{X}}[y]$  is implemented by setting  $x = (y - \bar{y})/\|y - \bar{y}\|$ , where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . This projection is unique and well-defined unless  $y \equiv 0$ . We avoid this situation by initializing  $x^{(0)} \in \mathcal{X}$ , and taking the step size  $\mu$  to be sufficiently small. To avoid the problem of local minima (since the constraint set is non-convex) we rerun the optimization algorithm from multiple initial conditions. Figure 4 shows the relative error achieved by GGE as a function of the number of iterations for different initial conditions of  $x(0)$ , averaged over 100 realizations of the algorithm. Also plotted is the bound identified in Theorem 3, substituting in  $A(G)$  calculated numerically. For all but the linearly-varying field, GGE achieves a much more rapid initial decrease in error than indicated by the bound. After approximately 1000 iterations, the bound provides a

good indication of the rate of decrease in error. We again observe that the linearly-varying field is close to a worst-case scenario for GGE.

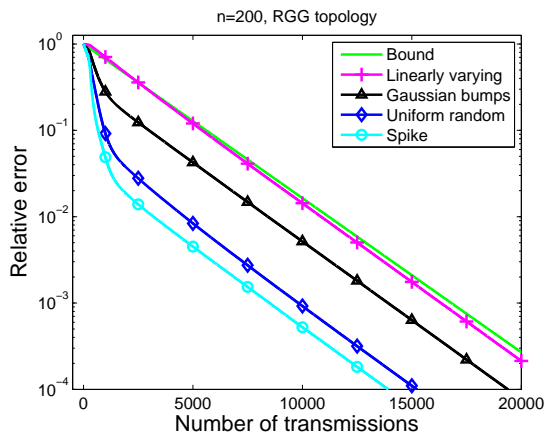


Fig. 4. A comparison of the theoretical bound on relative error and the experimental performance of GGE for four initializations. Results are for 100 realizations of the random geometric graph, averaged over 100 runs of the algorithm.

Next, we examine how the communication complexity scales with respect to the number of nodes in the network. Figs. 5(a) and 5(b) display how  $A(G)$  and the averaging time scale as a function of the number of nodes  $n$ , for random geometric graph and grid topologies, respectively. To obtain the random geometric graph curve, we generate 50 random graphs for each value of  $n$ , and numerically evaluate  $A(G)$  for each using the procedure detailed above. The top panel shows how the values of  $A(G)$  change as the number of nodes increases. The bottom panel shows the  $\epsilon$ -averaging time,  $T_{ave}(\epsilon)$ , evaluated via simulation, for  $\epsilon = 0.01$  versus the number of nodes. Note that Figs. 5(a) and 5(b) show the averaging time in terms of the number of iterations per node. The errorbars depict the minimum, mean and maximum values obtained for the 50 simulated graphs for each  $n$ . For reference, the dotted lines depict  $1.5n/\log n$  for the random geometric graph and  $2.5n$  for the grid topology.

### C. Stale Information

In wireless networks, links can be unreliable, and for GGE, it is possible for nodes to miss some updates from their neighbors. Consequently, nodes will have stale information about their neighbors' values and therefore the greedy selection in GGE may be affected. Here we investigate the effect of stale information on the performance of GGE through a simulation study.

We consider random graph topologies with 200 nodes. The initial measurements  $x(0)$  correspond to sampling the Gaussian bumps field, similar to Figure 2(a). As described in Section II, at the  $k$ th

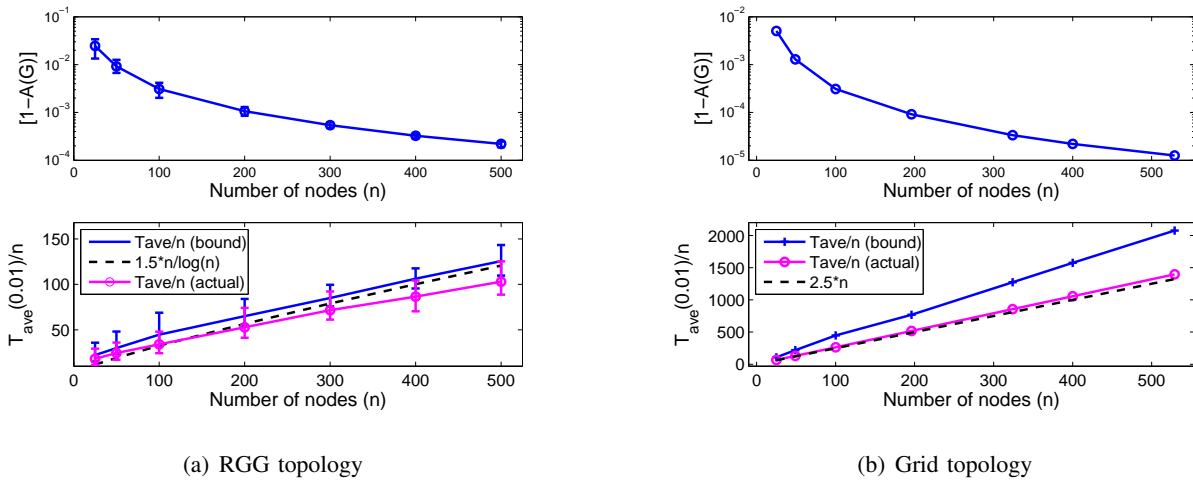


Fig. 5. The scaling behavior of  $A(G)$  and the averaging time  $T_{ave}(\epsilon)$  for  $\epsilon = 0.01$  as a function of the number of nodes  $n$  in the network. The top panels show  $A(G)$  evaluated numerically, and the bottom panels compare  $T_{ave}(0.01)$  computed via simulation and via the corresponding bound from Theorem 3. (a) 50 random geometric graphs are simulated for each value of  $n$ . The error bars depict the minimum, mean, and maximum values obtained over these 50 realizations. The curve  $1.5n/\log n$  is also shown for reference. (b) For the grid topology, error bars are not used since there is only one realization of the grid for a given network size. The curve  $2.5n$  is also shown for reference.

GGE iteration, two nodes  $s_k$  and  $t_k$  perform averaging. To provide up-to-date information to their neighbors,  $s_k$  and  $t_k$  broadcast their new values. We simulate the case when nodes randomly miss the broadcasted messages. We assume that the gossiping nodes  $s_k$  and  $t_k$  communicate reliably, but sometimes eavesdropping nodes miss an update from their neighbor. Specifically, each eavesdropping node independently misses the transmission from its neighbor with probability  $p$ . Figure 6 illustrates the performance degradation in GGE. Curves are shown for four different values of  $p$  between 0 and 0.5, and standard randomized gossip is also shown for comparison. We conclude that GGE provides significantly better performance than randomized gossip even when 50 percent of the broadcast messages are missed.

## V. MULTI-HOP GREEDY GOSSIP WITH EAVESDROPPING

As Figs. 2 and 3 indicate, the improvement of GGE over randomized gossip is less for the grid topology compared to the RGG topology. The decrease in improvement is due to the fact that the node degree in a two-dimensional grid is bounded at 4 and does not increase with the network size. Here we propose an extension to our algorithm that improves the performance of GGE such that it can be employed for topologies with low average node degree. Essentially, this extension allows nodes to perform greedy gossip updates with nodes beyond their immediate one-hop neighborhood.

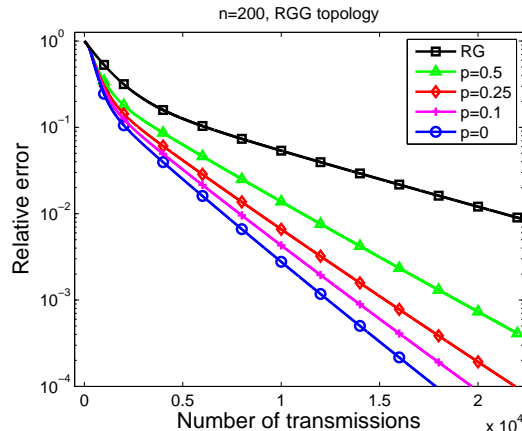


Fig. 6. A comparison of the performance of randomized gossip with GGE in the case of stale information due to independent link failures with probability  $p$ . Results are for 100 realizations of the 200-node random geometric graph, averaged over 100 runs of the algorithm.

In one-hop GGE, at the  $k$ th iteration,  $s_k$  determines which neighbor,  $t_k$ , has a value most different from its own. In two-hop GGE, instead of completing the update,  $t_k$  checks if any of its neighbors  $u_k \in \mathcal{N}_{t_k}$  has a value even more different from  $s_k$  than its own; i.e.,  $\|x_{s_k} - x_{u_k}\| > \|x_{s_k} - x_{t_k}\|$ . If so, nodes  $s_k$  and  $u_k$  gossip; otherwise  $s_k$  and  $t_k$  gossip. Multi-hop gossip generalizes this idea to even larger neighborhoods. For example,  $u_k$  can search its neighborhood, and so on.

To observe the effect of performing greedy updates over multiple hops, we conduct an experimental comparison between 1-hop, 2-hop, and 3-hop GGE. As a point of comparison, we also include curves for randomized gossip and geographic gossip. Figure 7 illustrates the results for grid and random geometric graph topologies. In the grid, 3-hop GGE achieves an asymptotic rate of reduction in relative error that is comparable to geographic gossip. In the random geometric graph topology, the asymptotic performance of 1-hop GGE is already similar to that of geographic gossip (for a network of 200 nodes) and the multi-hop versions lead to significant improvements while still limiting all gossip exchanges to be between nodes separated by at most three hops.

## VI. CONCLUSION

In this paper we propose a new average consensus algorithm for wireless sensor networks. Greedy gossip with eavesdropping (GGE) makes use of the broadcast nature of wireless communications and provides fast and reliable computation of average consensus. We provide (i) a proof that GGE converges to the average consensus; (ii) a bound on the mean-squared error after  $k$  iterations of GGE; (iii) a

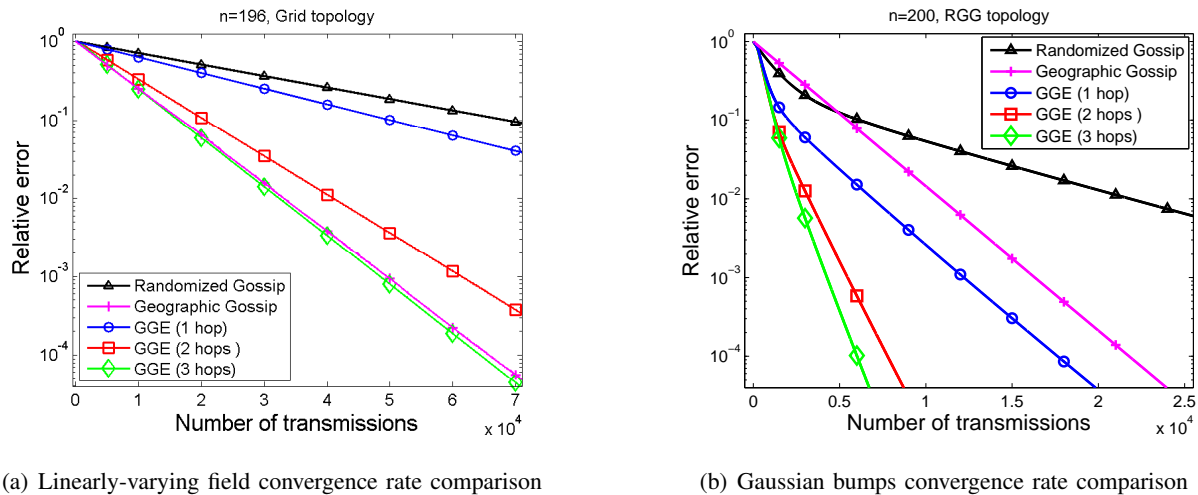


Fig. 7. A comparison of the performance of randomized gossip, GGE (1-hop), multi hop GGE (2 and 3-hops) and geographic gossip for linearly-varying field initialization of  $x(0)$  in grid topology and Gaussian bumps initialization of  $x(0)$  in RGG topology. Results are averaged over 100 runs of the algorithm.

bound on the  $\epsilon$ -averaging time of GGE; and (iv) theoretical bounds suggesting that GGE converges faster than randomized gossip, and (v) a characterization of the improvement in convergence rate achieved by GGE over randomized gossip as a function of the maximum degree. Simulation experiments compare the performance of GGE, randomized gossip [12], and geographic gossip [10] and demonstrate that the theoretical bound on mean-squared error provides a good characterization of the algorithm performance. The simulation experiments also investigate the scaling behavior of the communication complexity of GGE.

GGE retains the robustness and simplicity of randomized gossip; it does not require nodes to acquire location information and it does not introduce the overhead of geographic routing. There is an additional memory overhead (nodes store their neighbors' values), but this storage requirement is small. Nodes do need to learn their neighbors' values, and we propose an initialization process that introduces a minor performance penalty with negligible added complexity. Since nodes eavesdrop on their neighbors' broadcasts, they must remain in "Receive" mode throughout the entire operation of the GGE algorithm. In randomized gossip, nodes can enter "Idle" mode and only switch to "Receive" mode when they detect that a neighbor is requesting a data exchange. In a wireless sensor network implementation, this difference could lead to concerns that GGE would consume more energy than randomized gossip. However, empirical studies have shown that energy consumption in "Idle" and "Receive" modes is very similar for most existing wireless sensor network architectures [25].

Our future work will investigate the benefits of GGE in networks of mobile nodes. When nodes are mobile, other fast consensus approaches which exploit knowledge of geographic location are no longer applicable. However, because GGE is purely local and adaptive, we believe it is a promising candidate for accelerating gossip algorithms in time-varying networks. We are also interested in constructing a synchronous version of GGE which will effectively be implementing a projected subgradient algorithm. Moreover, we plan to investigate further connections between consensus algorithms and incremental subgradient optimization algorithms, towards computing more general functions than the average.

## REFERENCES

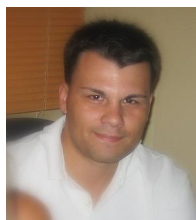
- [1] D. Üstebay, M. Coates, and M. Rabbat, “Greedy gossip with eavesdropping,” in *Proc. IEEE Int. Symp. on Wireless Pervasive Computing*, Santorini, Greece, May 2008.
- [2] D. Üstebay, B. Oreshkin, M. Coates, and M. Rabbat, “Rates of convergence for greedy gossip with eavesdropping,” in *Proc. Allerton Conf. on Comm., Control, and Computing*, IL, USA, September 2008.
- [3] —, “The speed of greed: Characterizing myopic gossip through network voracity,” to appear in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009.
- [4] J. Tsitsiklis, “Problems in decentralized decision making and computation,” Ph.D. dissertation, Massachusetts Institute of Technology, 1984.
- [5] S. Sundhar Ram, V. Veeravalli, and A. Nedić, “Distributed and recursive parameter estimation in parametrized linear state-space models,” Submitted, Apr. 2008.
- [6] M. Rabbat, R. Nowak, and J. Bucklew, “Robust decentralized source localization via averaging,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Phil., PA, Mar. 2005.
- [7] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, “Decentralized compression and predistribution via randomized gossiping,” in *Proc. Information Processing in Sensor Networks*, Nashville, TN, Apr. 2006.
- [8] A. Kashyap, T. Basar, and R. Srikant, “Quantized consensus,” *Automatica*, vol. 43, pp. 1192–1203, Jul. 2007.
- [9] S. Sundaram and C. Hadjicostis, “Distributed function calculation and consensus using linear iterative strategies,” *IEEE J. Selected Areas in Communications*, vol. 26, no. 4, pp. 650–660, May 2008.
- [10] A. Dimakis, A. Sarwate, and M. Wainwright, “Geographic gossip: Efficient aggregation for sensor networks,” in *Proc. Int. Conf. Inf. Proc. in Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006.
- [11] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [13] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA: Athena Scientific, 1997.
- [14] P. Gupta and P. Kumar, “The capacity of wireless networks,” *IEEE Trans. Info. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [15] W. Li and H. Dai, “Location-aided distributed averaging algorithms: Performance lower bounds and cluster-based variant,” in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.



- [16] K. Jung, D. Shah, and J. Shin, "Fast gossip through lifted Markov chains," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.
- [17] F. Benezit, A. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.
- [18] T. Aysal, M. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. IEEE Information Theory Workshop*, Porto, Portugal, May 2008.
- [19] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms: Design and analysis for consensus," in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008.
- [20] S. Sundhar Ram, A. Nedić, and V. Veeravalli, "Incremental stochastic subgradient algorithms for convex optimization," to appear in *SIAM J. on Optimization*.
- [21] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Automatic Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [22] A. Nedić and D. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.
- [23] M. Burnashev and K. Zigangirov, "An interval estimation problem for controlled observations," *Problems in Information Transmission*, vol. 10, pp. 223–231, 1974.
- [24] R. Castro and R. Nowak, "Active learning and sampling," in *Foundations and Applications of Sensor Management*, A. Hero, D. Castanon, D. Cochran, and K. Kastella, Eds. Springer-Verlag, 2007, pp. 177–200.
- [25] S. P. V. Raghunathan, C. Schurgers and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, March 2002.



**Deniz Üstebay** received the B.Sc. degree with honours in electrical and electronics engineering from Middle East Technical University (Ankara, Turkey) in 2004; and M.Sc. degree in electrical and electronics engineering from Bilkent University (Ankara, Turkey) in 2007. She is currently pursuing a Ph.D. degree in electrical engineering at McGill University (Montréal, Canada). Her research interests include sensor networks and distributed signal processing. Her email address is [deniz.ustebay@mail.mcgill.ca](mailto:deniz.ustebay@mail.mcgill.ca).



**Boris N. Oreshkin** received the B.E. degree with highest honours in electrical engineering from Ryazan State University of Radio Engineering, Russia, in 2004; and a Ph.D. degree in electrical engineering from McGill University, Canada in 2009. From 2004 to 2006 he was involved in hardware and software design and digital signal processing algorithms development and implementation in industry. In 2006 he was awarded Dean's Doctoral Student Research Recruitment Award by the Faculty of Engineering at McGill University. He is currently a post doctoral fellow with computer networks lab, McGill University. His research interests include statistical signal processing, Bayesian and Monte Carlo inference, and distributed information fusion. His email address is [boris.oreshkin@gmail.com](mailto:boris.oreshkin@gmail.com).



**Mark J. Coates** (M'99,SM'07) received the B.E. degree (first class honours) in computer systems engineering from the University of Adelaide, Australia, in 1995, and a Ph.D. degree in information engineering from the University of Cambridge, U.K., in 1999. He joined McGill University (Montreal, Canada) in 2002, where he is currently an Associate Professor. He was awarded the Texas Instruments Postdoctoral Fellowship in 1999 and was a research associate and lecturer at Rice University, Texas, from 1999-2001. His research interests include communication and sensor networks, statistical signal processing, causal analysis, and Bayesian and Monte Carlo inference. His email address is [coates@ece.mcgill.ca](mailto:coates@ece.mcgill.ca).



**Michael G. Rabbat** earned the B.Sc. from the University of Illinois at Urbana-Champaign (2001), the M.Sc. degree from Rice University (2003), and the Ph.D. from the University of Wisconsin-Madison (2006), all in electrical engineering. He is currently an Assistant Professor at McGill University. He was a visiting researcher at Applied Signal Technology, Inc., during the summer of 2003. He received the Best Student Paper award at the 2004 ACM/IEEE Conference on Information Processing in Sensor Networks, Outstanding Student Paper Honorable Mention at the 2006 Conference on Neural Information Processing Systems, and the Harold A. Peterson Thesis Prize. His current research focuses on distributed information processing in sensor networks, network monitoring, and network inference.