

# THE SPEED OF GREED: CHARACTERIZING MYOPIC GOSSIP THROUGH NETWORK VORACITY

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

Department of Electrical and Computer Engineering, McGill University,  
3480 University St, Montréal, Québec, Canada

## ABSTRACT

This paper analyzes the rate of convergence of *greedy gossip with eavesdropping* (GGE). In previous work, we proposed GGE, a fast gossip algorithm based on exploiting the broadcast nature of wireless communications rather than location information. Assuming all transmissions are wireless broadcasts, nodes can keep track of their neighbors' values by eavesdropping on their communications. Then, when it comes time to gossip, a node greedily and myopically gossips with the neighbor whose value is most different from its own, rather than with a randomly chosen neighbor. Previously, we have proved that GGE converges to the average consensus on connected network topologies and demonstrated that GGE outperforms standard *randomized gossip* (RG). In this paper we study the rate of convergence of GGE in terms of network voracity which is a topology-dependent constant analogous to the second-largest eigenvalue characterization for RG. Simulations demonstrate that the convergence rate of GGE is superior to existing average consensus algorithms such as geographic gossip.

**Index Terms**— Distributed signal processing, gossip algorithms, average consensus, wireless sensor networks.

## 1. INTRODUCTION AND BACKGROUND

Distributed consensus and gossip algorithms are a promising tool for performing distributed signal processing tasks. In particular, these methods have received considerable attention, for applications in wireless sensor networks (see, e.g., [1–3] and references therein), because they do not require overhead for forming or maintaining routes, nor do they suffer from any sort of bottleneck near a fusion centre since they are fully decentralized. Existing fast algorithms for average consensus in wireless network-like topologies rely on exploiting geographic location information at each node to speed up gossiping. However, location information may not be available when a network is initially deployed, and existing location-based algorithms are not applicable when nodes are mobile.

In this paper we analyze a fast gossip algorithm that does not use location information. Instead, via wireless broadcast transmissions, we allow a node to track the values at its neighbors, and then perform myopic, greedy updates based on this information. This algorithm was introduced in [4], and is referred to as *greedy gossip with eavesdropping* (GGE). Gossip algorithms have previously been analyzed using techniques from the theory of Markov chains, and their rate of convergence is typically characterized by the second largest eigenvalue of a transition matrix derived from the gossiping mechanism. However, because of the greedy, adaptive update involved in GGE, there is no analogous time-homogeneous Markov structure to study. Instead, we analyze convergence properties of GGE by showing that it solves a convex optimization problem. Similar to the

second largest eigenvalue rate of convergence bounds for randomized gossip, we bound the rate of convergence of GGE in terms of a topology-dependent constant, which we term *network voracity*, that quantifies the amenability of a particular topology to greedy gossiping. We study network voracity for random geometric graphs via simulation.

The prototypical example of a consensus problem is that of computing the *average consensus*: initially, each node in a network has a scalar piece of information, and the goal is to compute the average at every node. The two most widely studied algorithms for solving the average consensus problem are *distributed averaging*, tracing back to the seminal work of Tsitsiklis [5], and *randomized gossip* (RG) [6]. In distributed averaging, every node broadcasts information to its neighbors at every iteration. However, information diffuses slowly across the network in this scheme since the information at each node typically does not change much from iteration to iteration. *Randomized gossip* operates at the opposite extreme, where only two neighboring nodes exchange information at each iteration. At the  $k$ th iteration, a node  $s$  is chosen uniformly at random; it chooses a neighbor,  $t$ , randomly; and this pair performs the update. The fact that nodes only exchange information with their neighbors is attractive from the point of view of simplicity and robustness. However, it also means that in typical wireless network topologies (grids or random geometric graphs [7]), information diffuses slowly. Boyd et al. [6] prove that for random geometric graphs, RG requires  $O(n^2)$  transmissions. This slow convergence, and the consequent poor scaling with network size, is undesirable for large-scale sensor network applications, where each transmission consumes valuable (and scarce) energy resources.

Slow convergence of RG motivated Dimakis et al. to develop geographic gossip. In [3], they show that assuming each node knows its geographic location and the locations of its neighbors, the long-range information exchanges significantly improve the rate of convergence to roughly  $O(n^{3/2})$  transmissions. However, this improvement comes at the cost of increased complexity, since the network must now provide two-way transmission over many hops. Similarly, Jung et al. [8] advocate using the geographic locations of nodes to construct *lifted* Markov chains directing the information exchange to accelerate gossip. These approaches rely on geographic information and thus are not suitable to scenarios where location information is not available.

Aysal et al. propose *broadcast gossip*, which makes use of the broadcast nature of wireless networks [9]. At each iteration, a node is chosen uniformly at random to broadcast its value. The nodes in the broadcast range of this node update their value with the weighted average of their own value and the broadcasted value. In this manner, broadcast gossip achieves rapid convergence, but introduces bias.

The remainder of this paper is organized as follows. In Section 2 we review the formal definition of GGE. In Section 3 we develop a

bound relating the performance of GGE to that of standard RG. In Section 4 we develop a worst-case bound on the rate of convergence of GGE. In Section 5 we investigate the behavior of GGE empirically and find that it performs at least as well as other gossip algorithms. Finally, Section 6 summarizes the contributions of the paper.

## 2. 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  $i$ . Each node has an initial value  $y_i$ , and the goal of the gossip algorithm is to compute  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  at every node while only exchanging information between neighboring nodes. Initially, each node sets its gossip value to  $x_i(0) = y_i$ , and broadcasts it so that neighbors are all aware of each others' values.

At the  $k$ th iteration of GGE, a node  $s_k$  is chosen uniformly at random. This can be accomplished using the asynchronous time model described in [10]. The selected node,  $s_k$  identifies a node  $t_k \in \arg \max_{t \in \mathcal{N}_{s_k}} \{(x_{s_k}(k) - x_t(k))^2 / 2\}$ . That is, since each node is eavesdropping on its neighbors' communications, it is aware of their most recent values, and thus can select  $t_k$  to be a neighbor whose value is most different from its own. Then  $s_k$  and  $t_k$  exchange values and perform the update  $x_{s_k}(k) = x_{t_k}(k) = (x_{s_k}(k-1) + x_{t_k}(k-1)) / 2$ . GGE updates can also be expressed in the matrix form

$$x(k+1) = W^{GGE}(k)x(k)$$

where  $W^{GGE}(k)$  is a stochastic matrix with entries  $W_{s_k, s_k}^{GGE}(k) = W_{s_k, t_k}^{GGE}(k) = W_{t_k, s_k}^{GGE}(k) = W_{t_k, t_k}^{GGE}(k) = \frac{1}{2}$ ,  $W_{i, i}^{GGE}(k) = 1$  for all  $i \notin \{s_k, t_k\}$ , and 0 elsewhere. The standard approach for convergence analysis of gossip algorithms, viewing the algorithm as a Markov chain, cannot be applied here since the matrix  $W^{GGE}(k)$  is data-adaptive. In [4] we show that GGE is a randomized incremental subgradient algorithm [11] for the convex optimization problem,

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

$$\text{subject to} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \quad (2)$$

and the recursive update for GGE results in

$$\|x(k) - \bar{x}\|^2 = \|x(k-1) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2 \quad (3)$$

where the vector  $g(k)$  is a subgradient of the  $s_k$ -th component of the objective function and is defined as:  $g_{s_k}(k) = x_{s_k}(k-1) - x_{t_k}(k-1)$ ,  $g_{t_k}(k) = -(x_{s_k}(k-1) - x_{t_k}(k-1))$  and 0 elsewhere.

In [4], we characterized the convergence behavior by demonstrating that GGE converges almost surely to the consensus value, as a direct consequence of (3). The main idea is that, so long as  $x(k)$  is not a consensus,  $\|g(k)\| > 0$  with positive probability, and thus, the algorithm makes monotonic progress to the solution.

## 3. GGE VS. RANDOMIZED GOSSIP

The following theorem establishes a general expression for the bound on the mean-squared error of GGE after  $k$  iterations and

demonstrates that the upper bound on the MSE of GGE is less than or equal to the upper bound on the MSE of RG. We denote the application of  $k$  successive RG updates by  $W_{1:k}^{RG} = \prod_{j=1}^k W^{RG}(j)$ . Likewise, let  $W_{1:k}^{GGE} = \prod_{j=1}^k W^{GGE}(j)$  denote the successive application of  $k$  GGE updates. Let  $\bar{W} = \mathbb{E}[W^{RG}(k)]$  denote the expected RG matrix with the second largest eigenvalue  $\lambda_2(\bar{W})$ .

**Theorem 1.** *Let the input,  $x(0)$ , be given, and let  $\bar{x}$  denote the corresponding average consensus vector. After  $k$  iterations, the mean squared error of GGE is upper bounded as follows:*

$$\mathbb{E} \left[ \|W_{1:k}^{GGE} x(0) - \bar{x}\|^2 \right] \leq \|x(0) - \bar{x}\|^2 \prod_{i=1}^k (\lambda_2(\bar{W}) - \xi_i) \quad (4)$$

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

$$\xi_k = \frac{\sum_{i=1}^n \max_{t \in \mathcal{N}_i} (x_i(k-1) - x_t(k-1))^2}{2n \mathbb{E}[\|W_{1:k-1}^{GGE} x(0) - \bar{x}\|^2]} - \frac{\sum_{i=1}^n \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (x_i(k-1) - x_j(k-1))^2}{2n \mathbb{E}[\|W_{1:k-1}^{GGE} x(0) - \bar{x}\|^2]} \geq 0, \quad (5)$$

**Remark 1.** *The analogous expression for RG is simply [6]*

$$\mathbb{E}[\|W_{1:k}^{RG} x(0) - \bar{x}\|^2] \leq \lambda_2(\bar{W})^k \|x(0) - \bar{x}\|^2. \quad (6)$$

Since  $\xi_i \geq 0$ , the upper bound on GGE is uniformly upper bounded by the upper bound for RG, for any input  $x(0)$ . The upper bound for random gossip is tight. In practice it means that even for its worst case input, GGE performs at least as well as RG for RG's worst case input.

The form of the terms  $\xi_k$  also provides insight into which scenarios are less favorable for GGE. In particular, when the average neighborhood-wise maximal squared difference is equal to the doubly-averaged squared difference, being greedy does not provide any gain.

*Proof of Theorem 1.* Using the convergence rate bounds for RG, (6), we can identify a related relationship derived from applying  $k-1$  steps of GGE followed by one step of RG:

$$\begin{aligned} \mathbb{E}[\|W^{RG}(k)W_{1:k-1}^{GGE}x(0) - \bar{x}\|^2] &= \mathbb{E}[\|W_{1:k-1}^{GGE}x(0) - \bar{x}\|^2] \\ &\quad - \frac{1}{2} \frac{1}{n} \sum_{t_k=1}^n \frac{1}{|\mathcal{N}_{t_k}|} \sum_{s_k \in \mathcal{N}_{t_k}} (x_{s_k}(k-1) - x_{t_k}(k-1))^2 \\ &\leq \lambda_2(\bar{W}) \mathbb{E}[\|W_{1:k-1}^{GGE}x(0) - \bar{x}\|^2]. \end{aligned} \quad (7)$$

Now we can bound the error of GGE by adding and subtracting the effects of making the  $k$ -th step a RG update:

$$\begin{aligned} \mathbb{E}[\|W_{1:k}^{GGE}x(0) - \bar{x}\|^2] &= \mathbb{E}[\|W_{1:k-1}^{GGE}x(0) - \bar{x}\|^2] \\ &\quad \pm \frac{1}{2} \frac{1}{n} \sum_{t_k=1}^n \frac{1}{|\mathcal{N}_{t_k}|} \sum_{s_k \in \mathcal{N}_{t_k}} (x_{s_k}(k-1) - x_{t_k}(k-1))^2 \\ &\quad - \frac{1}{2} \frac{1}{n} \sum_{t_k=1}^n \max_{s_k \in \mathcal{N}_{t_k}} (x_{s_k}(k-1) - x_{t_k}(k-1))^2 \\ &\leq [\lambda_2(\bar{W}) - \xi_k] \mathbb{E}[\|W_{1:k-1}^{GGE}x(0) - \bar{x}\|^2]. \end{aligned} \quad (8)$$

Repeated application of this inequality yields (4).  $\square$

#### 4. GGE CONVERGENCE RATE: WORST CASE UPPER BOUND AND NETWORK VORACITY

The rate of convergence for gossip algorithms is typically quantified in terms of the  $\epsilon$ -averaging time,

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

Gossip algorithms such as RG and geographic gossip are easily related to a homogeneous Markov chain.  $T_{ave}(\epsilon)$  can be shown to scale as a function of the second largest eigenvalue of the probability transition matrix,  $\bar{W}$ , of this chain [6]. In particular,  $T_{ave}(\epsilon) \leq \frac{3 \log \epsilon^{-1}}{\log \lambda_2(\bar{W})^{-1}}$ . Since the greedy decision made in each iteration of GGE depends on the gossip values at each node,  $x(k)$ , our algorithm cannot be related back to a homogeneous Markov chain. 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 2.** *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

$$A(G) = \max_{x \neq \bar{x}} \frac{1}{|V|} \sum_{v \in V} \left( 1 - \frac{\|g_v(x)\|^2}{4\|x - \bar{x}\|^2} \right),$$

and  $g_v(x)$  refers to a subgradient of the  $v$ -th component of the objective function in (1). 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 2.** *Note that  $A(G)$  is to GGE as  $\lambda_2(\bar{W})$  is to RG, in terms of bounding the averaging time. We refer to  $A(G)$  as network voracity, since it quantifies how well a particular topology is suited to greedy, myopic updates.*

*Proof of Theorem 2.* The proof of the first part of the claim is based on an approach introduced in [12]. We begin by recalling the recursion for the mean squared error of GGE after  $k$  iterations expressed in (3):

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

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

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

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. By applying Chebyshev's inequality, we have

$$\Pr(\|x(k) - \bar{x}\| \geq \epsilon \|x(0) - \bar{x}\|) \leq \frac{\mathbb{E}[\|x(k) - \bar{x}\|^2]}{\epsilon^2 \|x(0) - \bar{x}\|^2} \leq \epsilon^{-2} A(G)^k. \quad (9)$$

To get an upper bound on  $T_{ave}(\epsilon)$ , note that  $\Pr(\|x(k) - \bar{x}\| \geq \epsilon \|x(0) - \bar{x}\|) \leq \epsilon$  provided that  $k \geq \frac{3 \log \epsilon^{-1}}{\log A(G)^{-1}}$ .  $\square$

#### 5. NUMERICAL SIMULATIONS

In this section we compare the performance of GGE with RG [6] and geographic gossip [3]. We also investigate the scaling behavior of the communication complexity using the bound established in Section 4. In our experiments, we focus on random geometric graphs [7], constructed by distributing nodes uniformly at random over the unit square. The transmission radius is set to  $\sqrt{2 \log n/n}$ .

We first compare the convergence rates of the three algorithms by examining the reduction they achieve in relative error which is defined as  $\frac{\|x(k) - \bar{x}\|}{\|x(0) - \bar{x}\|}$ . All figures show averages over 100 realizations of the random geometric graph and 100 runs of the algorithm per graph. We examine performance for a linearly-varying field and a field with the "spike" signal, constructed by setting the value of one random node to 1 and all other node values to 0. Figs. 1(a) and (b) show that GGE converges towards the average at a much faster rate than RG for both initializations. The initial convergence of GGE is faster than geographic gossip for the field with a spike but is very similar for the linearly-varying field. On the other hand, asymptotically the algorithms achieve similar rate. For GGE, the linearly-varying field is the worst case, as was anticipated from the convergence analysis conducted in Section 3.

Finally, we examine how the communication complexity scales with respect to the number of nodes in the network. Since there is no closed-form solution for  $A(G)$ , we solve the optimization problem identified in Theorem 2 numerically, using an incremental subgradient algorithm. The network voracity can be expressed as

$$A(G) = 1 - \frac{1}{2n} \min_{x \neq \bar{x}} \sum_{i=1}^n \max_{j \in \mathcal{N}_i} (x_i - x_j)^2$$

and we can further focus on a simplified problem

$$A'(G) = \min_{x \in \mathcal{X}} \sum_{i=1}^n \max_{j \in \mathcal{N}_i} (x_i - x_j)^2$$

where  $\mathcal{X} = \{x \mid \bar{x} = 0, \|x(k)\|^2 = 1\}$ . Then  $A(G) = 1 - A'(G)/2n$ . We approximate the solution to this minimization using a projected incremental subgradient method. To avoid the problem of local minima (since the constraint set is non-convex) we rerun the optimization algorithm from multiple initial conditions. Figure 2 displays how  $A(G)$  and the theoretical bound on the averaging time change as the number of nodes  $n$  is increased. To obtain these data-points, we generated 50 random geometric graphs for each value of  $n$ , and using the described procedure evaluated  $A(G)$  numerically. Note that the averaging time is plotted in terms of the number of iterations per node. For comparison purposes, the dotted line depicts  $7\sqrt{n}$ . This provides some experimental support for a conclusion that the averaging time is  $O(n^{3/2})$ , which implies a communication complexity similar to geographic gossip.

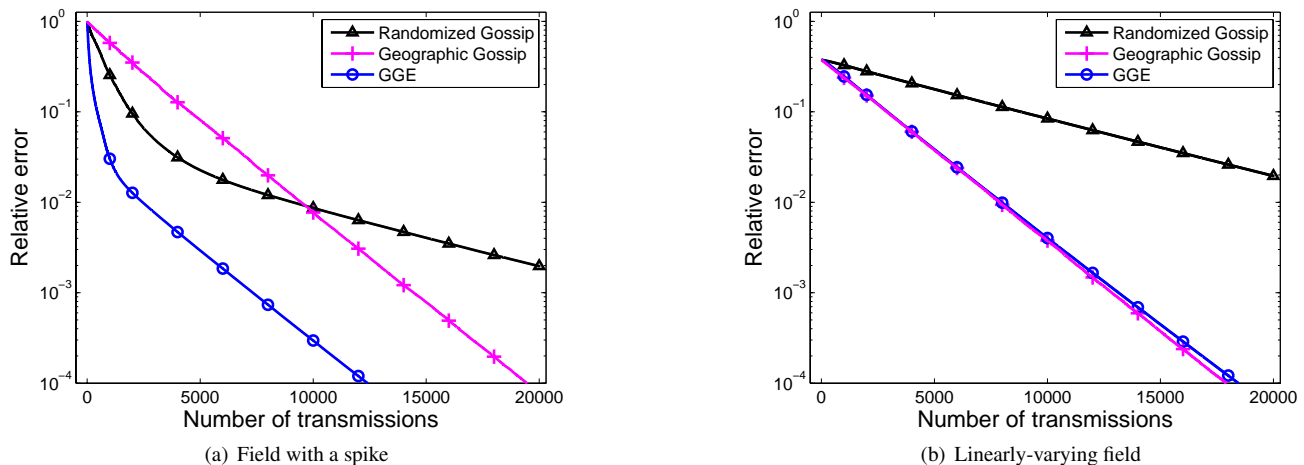


Fig. 1. A comparison of the performance of RG, GGE, and geographic gossip for two different initial conditions.

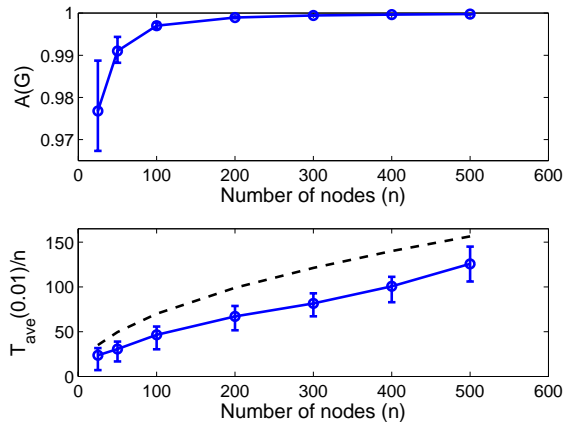


Fig. 2. The scaling behaviour of  $A(G)$  and the bound on the averaging time  $T_{avg}(\epsilon)$  for  $\epsilon = 0.01$  as  $n$  grows. The dashed line which is  $7\sqrt{n}$  is plotted for comparison purposes.

## 6. SUMMARY

In this paper we analyzed the convergence behavior of greedy gossip with eavesdropping (GGE), an algorithm we proposed in [4]. The theoretical contributions of this paper are (i) a bound on the mean-squared error of GGE; (ii) a bound on the  $\epsilon$ -averaging time of GGE; and (iii) a proof that GGE always converges faster than RG and a characterization of how the convergence rate differs. Simulations compare the performance of GGE, RG [6], and geographic gossip [3]. The simulations also investigated the scaling behavior of the communication complexity of GGE, and provided some evidence that it is  $O(n^{3/2})$ , similar to geographic gossip. A theoretical characterization of the scaling of this communication complexity is the focus of our current research.

## 7. REFERENCES

- [1] M. Rabbat, R. Nowak, and J. Bucklew, "Robust decentralized source localization via averaging," in *Proc. IEEE ICASSP*, Phil., PA, Mar. 2005.
- [2] S. Sundaram and C.N. Hadjicostis, "Distributed function calculation and consensus using linear iterative strategies," *IEEE J. Selected Areas in Comm.*, vol. 26, no. 4, pp. 650–660, May 2008.
- [3] A. Dimakis, A. Sarwate, and M. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. Info. Proc. in Sensor Nets.*, Nashville, TN, Apr. 2006.
- [4] D. Üstebay, M. Coates, and M. Rabbat, "Greedy gossip with eavesdropping," in *Proc. Int. Symp. on IEEE Wireless Pervasive Computing*, Santorini, Greece, May 2008.
- [5] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, MIT, 1984.
- [6] 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.
- [7] P. Gupta and P.R. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Th.*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [8] 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.
- [9] T.C. Aysal, M.E. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. IEEE Info. Theory Workshop*, Porto, Portugal, May 2008.
- [10] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Belmont, MA, 1997.
- [11] A. Nedić and D.P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. on Opt.*, vol. 12, no. 1, pp. 109–138, 2001.
- [12] M.V. Burnashev and K.S. Zigangirov, "An interval estimation problem for controlled observations," *Problems in Information Transmission*, vol. 10, pp. 223–231, 1974.