Selective Gossip

Deniz Üstebay
Electrical & Computer Engineering
McGill University
Montréal, QC, Canada
deniz.ustebay@mail.mcgill.ca

Rui Castro
Electrical Engineering
Columbia University
New York, NY, U.S.A.
rmcastro@ee.columbia.edu

Michael Rabbat
Electrical & Computer Engineering
McGill University
Montréal, QC, Canada
michael.rabbat@mcgill.ca

Abstract—**Motivated by applications in compression and distributed transform coding, we propose a new gossip algorithm called Selective Gossip to efficiently compute sparse approximations of network data. We consider running parallel gossip algorithms on the elements of a vector of transform coefficients. Unlike classical randomized gossip, communication between adjacent nodes is data driven and only performed if deemed to significantly improve the estimate of the signal vector. In particular nodes adaptively estimate and focus on using communication resources to compute significant coefficients (above a pre-defined threshold in magnitude). Consequently, energy and bandwidth are conserved by not gossiping on insignificant coefficients. The proposed procedure guarantees that all nodes will reach consensus on (i) the values of significant coefficients and (ii) the indices of insignificant coefficients. Insignificant values are not computed. We illustrate the significant communication savings over global randomized gossiping in a distributed transform coding application.**

I. INTRODUCTION

Decentralized signal processing algorithms are needed for wireless sensor networks and cyber-physical systems applications, where battery-powered devices autonomously form an ad hoc network and operate as a collective system. In this setting, collecting and processing data at a fusion center causes a bottleneck, and previous studies have shown that in-network processing can lead to significant energy savings. Gossip algorithms are emerging as an attractive mechanism for in-network signal processing. Gossiping refers to an iterative decentralized framework for computation where each node maintains a local estimate of the quantity being computed, and the goal is to reach a consensus where all nodes agree on the same estimate. At each iteration, neighboring nodes (that communicate directly) exchange information and then update their local estimates. Because all information exchange is local and asynchronous, there is no need for network-wide coordination, and consequently, no communication or security bottlenecks arise, e.g., around a fusion center.

This paper describes *selective gossip*, a decentralized algorithm designed specifically for computing approximations to large vectors of data over a network. To reach consensus on a vector of data, a direct approach would be to gossip on each component of the vector in parallel. However, in many applications, one may only be interested in the components which contain significant energy (i.e., they exceed a pre-specified threshold in absolute value); and components that are below the threshold are simply ignored or forced to zero.

In such applications, there is always a dilemma: a priori we do not know which components are the significant ones. If we knew which were significant, we would save energy by disregarding the insignificant ones from the outset. Without knowing, one must somehow estimate which are significant, or compute all components in advance and then discard the insignificant ones. Of course, this latter approach is wasteful and we would prefer to conserve energy and bandwidth resources by only gossiping about significant values.

Selective gossip is an asynchronous decentralized algorithm that adaptively determines which components are significant and insignificant while gossiping. In each gossip round, two neighboring nodes exchange information for components of the vector that at least one believes to be significant, based on their current estimates. By doing so, few transmissions are spent gossiping on insignificant coefficients. We prove that selective gossip converges asymptotically, in the following sense. Significant components (those with energy above the pre-defined threshold) converge asymptotically to their true value at every node in the network. Insignificant components eventually reach a state where the local estimates at every node are below the threshold, and in this manner all nodes consent to disregard the component. We demonstrate the utility of selective gossip for sparse approximation in a field estimation application and find that selective gossip obtains a network-wide estimate while transmitting significantly fewer values than an approach that gossips in parallel on all coefficients.

A. Background and Related Work

Distributed consensus, which has its roots in the seminal work of Tsitsiklis [1], has recently been identified as a canonical problem in distributed signal processing and control (see, e.g., [2] for a survey). In the context of average consensus, for a network of n nodes where each node has an associated scalar value y_i, the goal is to compute the average $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ at every node in a decentralized fashion. Although simple to formulate, if one can solve the average consensus problem then one can compute any linear function of the network data in a decentralized fashion. Consequently, consensus algorithms have proven useful in many applications ranging from decentralized compression [3] to localization in sensor networks [4].

Gossip asymptotically solves the average consensus problem through iterative information exchange between neighboring nodes. Since information exchange is local, gossip algo-

rithms are simple, scalable, and robust to changes in network topology or unreliable communications. Randomized gossip is one such algorithm and is analyzed in [5]. At every iteration of randomized gossip one of the nodes wake up uniformly at random and randomly selects one of its neighbors. These two nodes take the average of their values and all other nodes remain unchanged. Under very mild conditions on how the neighbor is selected, it can be shown that the values at every node converge to the initial average. Recently, several variants of randomized gossip have been proposed with the purpose of accelerating the speed of convergence [6]–[9].

In this paper we propose a new variant of randomized gossip called *selective gossip*. Unlike randomized gossip, which computes the average for a single scalar value, selective gossip is designed for efficiently reaching consensus on a vector of values where some components may be insignificant relative to the others. We illustrate the utility of selective gossip in a distributed field estimation application. In [3] and [10] the authors describe alternative field estimation methods using gossip algorithms to compute random linear transformations of the network data, and then recover the field using techniques from compressed sensing. This approach is useful for exploratory data analysis but is inefficient when one has available a linear transformation that sparsifies the data. In [11] and [12] wavelet transform methods are proposed for wireless sensor networks. However both of these methods aggregate data along trees and therefore require coordination to form and maintain specialized routes.

II. ALGORITHM

We consider a network of n nodes and represent network connectivity with a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The vertices, $\mathcal{V} = \{1, \ldots, n\}$, are the nodes and edges are the communication links between the nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. We assume that links are symmetric and the network is connected. Each node i in the network has an initial vector, $\mathbf{y}_i \in \mathbb{R}^m$. The goal is to compute $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i$. Let $y_{i,j}$ denote the jth component of the vector at node i, and let $\bar{y}_j$ denote the jth component of $\bar{\mathbf{y}}$. We are given a threshold $\tau > 0$ which indicates the significance for components of $\bar{\mathbf{y}}$. A component $\bar{y}_j$ is considered significant when $|\bar{y}_j| \geq \tau$; otherwise $\bar{y}_j$ is insignificant.

Selective gossip asymptotically computes the values of significant components, and only invests enough transmissions on insignificant components to have every node consent on their insignificance. At the kth iteration each node maintains a local estimate $\mathbf{x}_i(k)$ of $\bar{\mathbf{y}}$, and the estimates are updated in an iterative fashion. Node i initializes its jth component to $x_{ij}(0) = y_{i,j}$. At the kth iteration, a node s is chosen uniformly at random from $\{1, \ldots, n\}$ (this can be implemented using the asynchronous time model described in [13]), and s randomly selects a neighboring node t uniformly at random. Then s and t gossip only on the components of $\mathbf{y}$ which are significant; i.e., they update components j for which either $|x_{s,j}(k-1)| \geq \tau$ or $|x_{t,j}(k-1)| \geq \tau$ by setting

$$x_{s,j}(k) = x_{t,j}(k) = \frac{1}{2}\big(x_{s,j}(k-1) + x_{t,j}(k-1)\big). \quad (1)$$

No change is made to component j when $|x_{s,j}| < \tau$ and $|x_{t,j}| < \tau$, and these values are not transmitted.

III. CONVERGENCE OF SELECTIVE GOSSIP

In this section we prove that selective gossip converges to the correct values. Since there is no coupling between the different components of the vector, we treat each individually. Without loss of generality, let $x_i(0)$ denote the initial value for this component at node i, let $\bar{x}$ denote the average, and let $\tau > 0$ be the given threshold. Below, we prove convergence in the sense that when $|\bar{x}| \geq \tau$, for all i, $x_i(k) \to \bar{x}$ as $k \to \infty$, and when $|\bar{x}| < \tau$, there exists a K such that, for every i, $|x_i(k)| < \tau$ for all $k \geq K$. It is well known that, under the assumptions stated above, randomized gossip converges asymptotically to the average consensus [5]. Selective gossip only differs from randomized gossip in that, at some iterations, two nodes may choose not to gossip about a particular component, so it will not be updated. Thus, intuitively, to show convergence when $\bar{x} \geq \tau$ (resp., $-\bar{x} \leq \tau$) we just need to show that nodes gossip sufficiently often so that eventually they all have $x_i(k) \geq \tau$ (resp., $-x_i(k) \leq \tau$); at that point the algorithm becomes just randomized gossip. Similarly, when $|\bar{x}| < \tau$, we just need to show that nodes gossip enough so that eventually they reach a state where $|x_i(k)| < \tau$ for all i, at which point the entire network will cease to gossip about the insignificant coefficient.

To make this argument rigorous we will define a cost function measuring the average distance to $\bar{x}$, and demonstrate that it is strictly decreasing in expectation. In doing so, we will make use of the following lemma.

Lemma 1: Let $S(k) = \sum_{i=1}^{n}(x_i(k) - \bar{x})^2$, and let s and t be the nodes that gossip at iteration k. Then $S(k) = S(k-1) - \frac{1}{2}(x_s(k-1) - x_t(k-1))^2$.

This follows directly from the definition of $S(k)$ and the update rule (1).

Theorem 1: Let $S(k)$ be defined as above and suppose $|\bar{x}| > \tau$. Then $\mathbb{E}[S(k)] \leq (1 - \frac{1}{2n^7})^k S(0)$.

Proof: We will show that there exist a pair of neighboring nodes s and t for which $(x_s(k) - x_t(k))^2 > S(k)/n^5$, and this pair has at least probability of $1/n^2$ of gossiping at the $(k+1)$th iteration. Then we will apply Lemma 1 and find that at every iteration we decrease the cost function by at least $S(k)/(2n^7)$ in expectation. In the rest of the proof, we consider the case $\bar{x} \geq \tau$ (the case $\bar{x} \leq -\tau$ is analogous).

The first step is to show that there exists a node a with $x_a(k) \geq \bar{x} + \frac{1}{n}\sqrt{\frac{S(k)}{n}}$. Without loss of generality assume that $S(k) > 0$ (otherwise consensus has been attained). Note that there exists a node i for which $(x_i(k) - \bar{x})^2 \geq S(k)/n$, otherwise we get a contradiction $\sum_{i=1}^{n}(x_i(k) - \bar{x})^2 < S(k)$. Let i denote such a node. If $x_i(k) > \bar{x}$ then $x_i(k) \geq \bar{x} + \sqrt{S(k)/n}$ and we take i to be the node a we are looking for. If $x_i(k) < \bar{x}$, then $x_i(k) \leq \bar{x} - \sqrt{S(k)/n}$. Note that $\sum_{j \neq i} x_j(k) = n\bar{x} - x_i(k) = (n-1)\bar{x} + \bar{x} - x_i(k)$, and therefore $\sum_{j \neq i} x_j(k) \geq (n-1)\bar{x} + \sqrt{S(k)/n}$. This implies that there is at least a node a such that $x_a(k) \geq \frac{1}{n-1}\left((n-1)\bar{x} + \sqrt{\frac{S(k)}{n}}\right) \geq \bar{x} + \frac{1}{n}\sqrt{\frac{S(k)}{n}}$.

Now, define $\mathcal{H} = \{h \in \mathcal{V} : x_h(k) < \bar{x}\}$ and note that $\mathcal{H}$ is non-empty since $S(k) > 0$. Recall that $x_a(k) > \bar{x}$, and let $a = a_1, a_2, \ldots, a_m = h$ be a shortest path in $\mathcal{G}$ from a to $\mathcal{H}$ such that $h \in \mathcal{H}$ and $a_\ell \notin \mathcal{H}$ for all $\ell < m$. Because $x_a(k) \geq \bar{x} + \frac{1}{n}\sqrt{\frac{S(k)}{n}}$ and $x_h(k) < \bar{x}$ it follows that there is at least one step $(a_\ell, a_{\ell+1})$ along this path for which $|x_{a_\ell}(k) - x_{a_{\ell+1}}(k)| \geq \frac{1}{mn}\sqrt{\frac{S(k)}{n}}$. Since $m < \mathrm{diam}(\mathcal{G}) \leq n$ and a_ℓ is not in $\mathcal{H}$ the probability that these nodes gossip at iteration $k + 1$ is at least $1/n^2$ (a_ℓ ticks with probability $1/n$, and it has at most $n - 1$ neighbors). From Lemma 1 it then follows that $\mathbb{E}[S(k+1)|S(k)] \leq S(k) - \frac{S(k)}{2n^7} = (1 - \frac{1}{2n^7})S(k)$ and recursing back to 0 from iteration k leads to the claim. ∎

When a component is significant, selective gossip will always compute the correct value in expectation. Standard arguments [5] based on Markov's inequality can then be applied to show convergence in probability. To prove that the algorithm converges for insignificant components ($|\bar{x}| < \tau$) a similar methodology can be used, but using a modified cost function instead, $\tilde{S}(k) = \sum_{i:|x_i(k)|\geq\tau}(x_i(k) - \bar{x})^2$. We omit the full proof here due to lack of space.

It is also worth noting that the bounding arguments used in the proof above are extremely loose, and should not be taken as an indicator of the rate of convergence. In fact, it is easy to see that once all nodes agree the component is significant, selective gossip behaves identically to randomized gossip, and so asymptotically the rates of convergence are the same. As illustrated in the simulations presented below, the error decay rate of selective gossip, as a function of the number of values transmitted per iteration, is in fact substantially faster than randomized gossip when some components are insignificant.

IV. Distributed Transform Coding

Next, we show how selective gossip can be used to carry out decentralized sparse approximation of transform coefficients. In general, an orthonormal linear transformation can be expressed as matrix $\Phi \in \mathbb{R}^{n \times n}$ which takes a signal $\mathbf{f} \in \mathbb{R}^n$ to a set of coefficients $\boldsymbol{\beta} = \Phi^T \mathbf{f}$. Informally, if the signal $\mathbf{f}$ is compressible in Φ then $\boldsymbol{\beta}$ will have many components which are nearly zero (e.g., $|\beta_j| < \tau$), and one can obtain a sparse approximation $\hat{\mathbf{f}}$ to $\mathbf{f}$ by forcing these insignificant coefficients to zero to obtain $\hat{\boldsymbol{\beta}}$, and then inverting the transformation to get $\hat{\mathbf{f}} = \Phi\hat{\boldsymbol{\beta}}$. The main challenge for efficiently computing the sparse approximation $\hat{\boldsymbol{\beta}}$ in a decentralized fashion is that we generally do not know in advance which components are insignificant.

To apply transform coding in the context of sensor network field estimation, we view the data value at each node as one element of the signal $\mathbf{f}$. The goal is to compute $\hat{\boldsymbol{\beta}}$ at every node, so that an approximation to the network data is available everywhere. Each transform coefficient is a linear function of the network data, and thus can be computed using gossip. We assume each node i knows its corresponding row of the transformation matrix, Φ, and node i takes $y_{i,j} = n\Phi_{i,j}f_i$. Then $\bar{\mathbf{y}} = \boldsymbol{\beta}$. By using selective gossip with threshold τ, we

(a) Original Field (b) Sensed Field

(c) Centralized approximation (d) Gossip approximation, $\tau=0.25$
18% of coeffs used 19% of coeffs gossiped
MSE=1.062 MSE=1.062

(e) Centralized approximation (f) Gossip approximation, $\tau=0.6$
6% of coeffs used 7% of coeffs gossiped
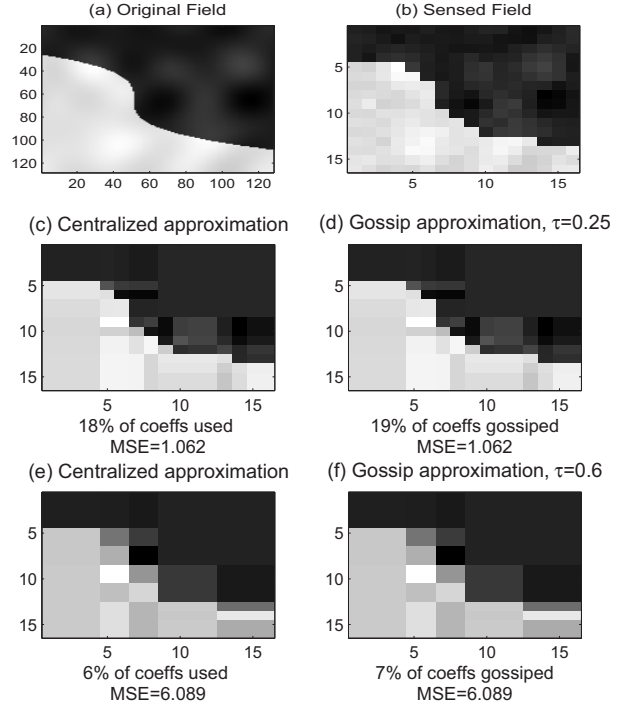MSE=6.089 MSE=6.089

Fig. 1. Original, sensed and approximated field values as gray scaled image.

compute an approximation $\hat{\boldsymbol{\beta}}$ at every node, where coefficients with $|\beta_j| < \tau$ are forced to zero.

V. Simulation Results

In this section we illustrate the performance of selective gossip algorithm via simulations. The field to be estimated is a 128×128 discrete sampling of a piecewise smooth field with additive Gaussian noise, $\mathcal{N}(0, \sigma^2)$. Figure 1(a) shows this field as a gray scaled image. 256 sensor nodes are arranged with network connectivity forming a 16×16 grid. Figure 1(b) is an image generated from the noisy sensor measurements. We use a three-level Haar wavelet basis as the linear transform. Selective gossip is run 20 times and the results presented here illustrate the average performance.

Figure 1(c-f) illustrates the results of approximation. For Figures 1(c) and (e) the approximations are obtained using centralized wavelet transform (assuming all the data was gathered to a single location) to compute coefficients and then insignificant coefficients are discarded. Figures 1(d) and (f) show the results of approximation using selective gossip to estimate significant wavelet coefficients. The approximation error is the mean square error, $\frac{1}{n}\sum_{i=1}^n \|\hat{\mathbf{f}}^i - \mathbf{f}\|^2$ where $\mathbf{f}$ is the vector of sensor measurements and $\hat{\mathbf{f}}^i$ is the reconstructed field using only significant coefficients estimated by node i. We conclude that selective gossip estimates of significant coefficients are similar to the percentage of significant coefficients found by centralized wavelet transform and gossip reaches similar approximation error values. Varying τ changes the approximation quality.

Figure 2 plots mean squared error versus number of gossiped coefficients for different values of the threshold, τ. First,
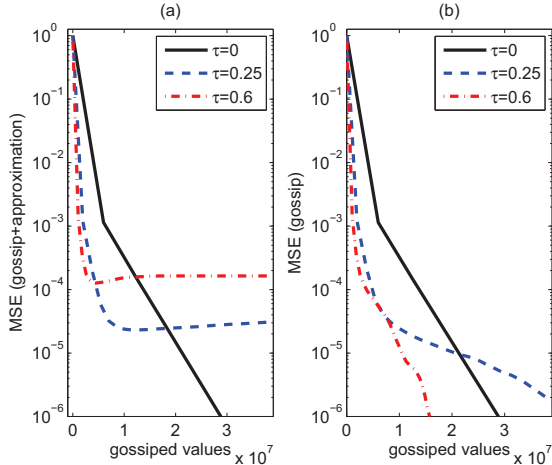
Fig. 2. Mean square error in coefficient estimates vs total number of gossiped values: (a) error due to approximation and gossip, (b) error due to gossip.
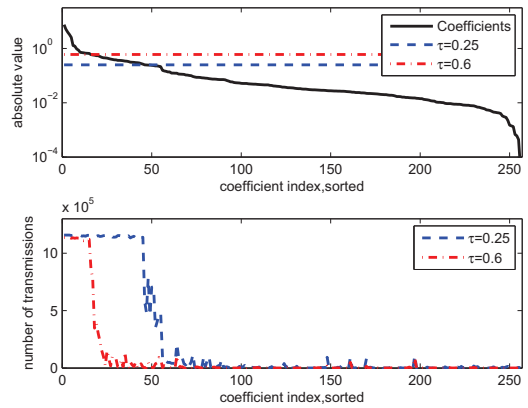

Fig. 3. Coefficients and corresponding convergence requirements. Top: original wavelet coefficients and threshold levels, Bottom: Number of transmissions per coefficient for convergence to error value 10^{-4}.

Figure 2(a) accounts errors due to both approximation (thresholding coefficients) and gossip, $\frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}\|^2$, where $\boldsymbol{\beta} = \Phi^T \mathbf{f}$ is the vector of true wavelet coefficients and $\hat{\boldsymbol{\beta}}_i$ is the vector of estimated coefficients at node i. The selective gossip curves level off when gossip has effectively converged, and all remaining error is only due to approximation. Figure 2(b) shows the error due to gossip only, $\frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}\|^2$ where $\hat{\boldsymbol{\beta}}$ is the thresholded version of $\boldsymbol{\beta}$, as if computed in a centralized fashion, and $\hat{\boldsymbol{\beta}}_i$ is the gossip estimate of $\hat{\boldsymbol{\beta}}$ at node i. The error is calculated using thresholded coefficients instead of true coefficients hence error due to approximation is ignored. As expected, higher values of τ result in higher approximation error, but with fewer transmitted values.

Figure 3 illustrates the effect of two different values of τ on number of transmissions required to approximate coefficients well. The top panel shows the original wavelet coefficients in absolute value, sorted in descending order. The panel below shows the number of gossip transmissions for each coefficient, where the order of indexing is the same as the sorting above. The two curves shown on the bottom panel correspond to two different thresholds, $\tau = 0.25$ and 0.6, as illustrated in the panel above. The total number of transmissions shown is that required to reach a relative mean squared error of 10^{-4}. Observe that selective gossip automatically determines which coefficients are insignificant, and spends a minimal number of transmissions on these coefficients.

VI. DISCUSSION

In this paper we propose a new gossip algorithm for decentralized sparse approximation. Selective gossip is aimed at computing a vector of data, where some components of the vector are insignificant and do not need to be computed exactly, but the locations of these components are not known in advance. Instead, selective gossip adaptively determines which components are significant while the computation is being carried out, and automatically adjusts where transmissions are invested in order to efficiently obtain an accurate estimate. We prove that the algorithm converges. Investigating the

rate of this convergence is the next step and the topic of ongoing research. In the future, we will study sparsifying transforms for general network topologies which are amenable to decentralized implementation and/or construction. Another interesting question is how to design a schedule for controlling τ to automatically compute the best (in terms of MSE) approximation for a given budget of transmissions.

REFERENCES

[1] J. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, 1984.
[2] R. Olfati-Saber, J. A. Fax, and R.M.Murray, "Consensus and cooperationin networked multi-agent systems," in *Proc. IEEE Information Theory Workshop*, Porto, Portugal, May 2007.
[3] 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.
[4] M. Yildiz, F. Ciaramello, and A. Scaglione, "Distributed distance estimation for manifold learning and dimensionality reduction," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] D. Üstebay, B. Oreshkin, M. Coates, and M. Rabbat, "The speed of greed: Characterizing myopic gossip through network voracity," Taipei, Taiwan, April 2009.
[10] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proc. Information Processing in Sensor Networks*, Cambridge, MA, Apr. 2007.
[11] G. Shen, S. Pattem, and A. Ortega, "Energy-efficient graph-based wavelets for distributed coding in wireless sensor networks," Taipei, Taiwan, Apr. 2009.
[12] R. Wagner, R. Baraniuk, S. Du, D. Johnson, and A. Cohen, "An architecture for distributed wavelet analysis and processing in sensor networks," in *Proc. ACM/IEEE Int. Conf. Inf. Proc. in Sensor Networks*, Nashville, TN, Apr. 2006.
[13] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA: Athena Scientific, 1997.