

Paving the Way for Consensus: Convergence of Block Gossip Algorithms

Jamie Haddock, *Member, IEEE*, Benjamin Jarman, and Chen Yap

Abstract

Gossip protocols are popular methods for average consensus problems in distributed computing. We prove new convergence guarantees for a variety of such protocols, including path, clique, and synchronous pairwise gossip. These arise by exploiting the connection between these protocols and the block randomized Kaczmarz method for solving linear systems. Moreover, we extend existing convergence results for block randomized Kaczmarz to allow for a more general choice of blocks, rank-deficient systems, and provide a tighter convergence rate guarantee. We furthermore apply this analysis to inconsistent consensus models and obtain similar guarantees. An extensive empirical analysis of these methods is provided for a variety of synthetic networks.

Index Terms

average consensus, randomized block gossip algorithms, linear systems, stochastic iterative methods, randomized block Kaczmarz

I. INTRODUCTION

Consider a network in which every node has a secret (unknown by the other nodes) value, and the goal is for all nodes to learn the average of these values. This problem is a classical and fundamental problem in distributed computing and multi-agent systems known as *average consensus* [1]; it has additional real-world applications in clock synchronization [2], PageRank, localization without GPS [3], opinion formation, distributed data fusion in sensor networks [4], blockchain technology, and load balancing [5]. See Figure 1 for a visualization of an average consensus problem.

Initial approaches for this problem may be to allow all nodes to pass their secret value to a single *hub* node which would then perform the averaging and pass this value back to the others, or for every node to share its stored knowledge of all other nodes' secret values with its neighbours until all nodes have learned all stored values (a process known as *flooding* [6]). However, such methods are problematic. The first requires communication that may be infeasible as it may not respect the topology of the underlying network; in particular, there may be no hub node (see Figure 1). The second may require many instances of communication between nodes and thus struggle to scale to modern large-scale networks. An attractive class of methods that go some way towards rectifying these issues are *gossip* protocols, where at each time-step some subset of nodes are 'activated' to share their stored information with each other across network edges. As is common in large-scale settings, such protocols are *randomized* in the sense that activated nodes are often chosen at random, with distribution depending on the particular gossip protocol. Over many iterations, the values held at each node converge to the average over the whole network (with mild assumptions including connectivity of the underlying network) and the problem is solved in a distributed manner. Such methods have been a topic of popular study this side of the millenium, with the seminal 2006 paper of Boyd et al. [7] sparking a flurry of research on the topic [8]–[15].

J. Haddock is with the Department of Mathematics, Harvey Mudd College, Claremont, CA, 91711 USA e-mail: jhaddock@g.hmc.edu.

B. Jarman and C. Yap are with University of California, Los Angeles.

Manuscript received ; revised .

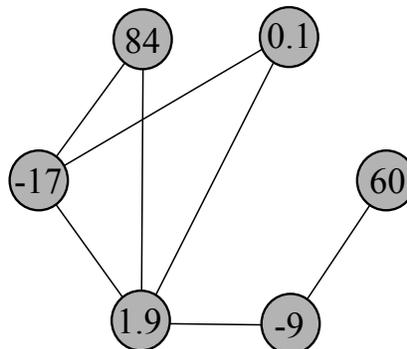


Fig. 1: Average consensus problem with initial (secret) values listed. The consensus value for this problem is $\bar{c} = 20$.

In [16], Loizou and Richtárik united the study of randomized gossip protocols with randomized numerical linear algebra. They showed that a wide class of gossip protocols can be interpreted, under certain assumptions, as randomized iterative linear system solvers applied to a linear system derived from the network at hand. This remarkable connection prompted a variety of convergence results, new accelerated and weighted gossip variants, and dual edge-based gossip protocols, opening up the possibility for further links to be developed. In this paper, we hone in on the class of *block* gossip algorithms discussed in their work: we show that several popular gossip protocols are members of this class, give strong convergence results for said methods, and generalize the theory on both the linear algebra and gossip protocol sides of the problem.

A. Contributions

Our main contributions in this work are threefold:

- We generalize previous results on randomized block iterative methods for solving linear systems (see [17]–[19]) to include the case where the system is less than full rank (which is vital for the average consensus case, as we will see), to include a wider class of block sampling protocols, and to sharpen the resulting convergence rate guarantee.
- We derive new convergence rates for popular gossip protocols such as *path* [20], *clique* [21], and *edge-independent set* [7] gossiping, by showing they can be interpreted as block gossip methods. We furthermore generalize to include the case where multiple node sets can be activated simultaneously, and analyze the dependence of a protocol’s performance on the spanning trees of the activated subgraphs.
- We give new analyses of gossip for *inconsistent* consensus models. We link edge communication errors in a network to inconsistent, or noisy, linear systems, and connect the performance of iterative solvers on said systems to the performance of gossip protocols on said networks.

Furthermore, we provide a wide range of experiments to demonstrate the comparative performance of the discussed algorithms, both previously existing and new, on a variety of network structures.

We remark that we do not go into detail regarding the specific communication protocols that networks may have. For example, certain methods (e.g., path gossiping [20]) require multiple instances of information sharing at each iteration, which may not be implementable in all networks, and others (e.g., edge independent set gossiping [7]) require communication across each edge only once at each time step. We point the reader to [7], [22] for more details on network communication protocols and their use in gossip algorithms.

B. Organization

The rest of the paper is organized as follows. In Section I-C, we introduce notation that will be used throughout the work. In Section I-D, we describe in detail the methods considered in this paper and present our main theoretical results. In Section I-E, we provide detailed background on the average consensus problem, and recent work on block iterative methods for solving linear systems. In Section II, we prove our generalization of the block iterative method theory, and go into detail on the connection between this and block gossip algorithms for average consensus. In Section III, we connect particularly popular block gossip algorithms with our theory and produce explicit convergence rates for them. In Section IV, we explore the case of average consensus models with edge miscommunication and provide convergence results for the gossip method on this type of faulty model. Lastly, in Section V, we provide numerous experiments and compare our theoretical results with the empirical behavior of the considered gossip methods.

C. Notation and Definitions

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected network where \mathcal{V} denotes the n nodes of the network and \mathcal{E} denotes the m edges. Let $\mathbf{Q} \in \mathbb{R}^{m \times n}$ denote the incidence matrix of the network. Each row of \mathbf{Q} corresponds to an edge of the network and each column corresponds to a node. If row l of \mathbf{Q} , denoted \mathbf{q}_l^T , corresponds to the edge $e_{ij} \in \mathcal{E}$ connecting node i and j , then all but the i th and j th entries of \mathbf{q}_l^T are zero with these entries containing a one and negative one (the order of the positive and negative entries does not matter).

We recall the notion of several special subgraph structures. We remind the reader of the definition of an *independent edge set*, that is a subset of edges of the graph in which no two edges are incident to the same node. Additionally, we remind the reader that a *clique subgraph*, or *complete subgraph*, is a subset of edges of the graph which together form a subgraph in which every pair of nodes is connected; that is, the edge-induced subgraph is a clique. Finally, we remind the reader of the definition of a *path subgraph*, a subset of edges which together form a path graph; that is, the edge-induced subgraph is a path.

Throughout, we let $[m]$ denote the set of integers from one to m ; $[m] := \{1, 2, \dots, m\}$. We additionally let $\mathbf{0}$ denote the vector of all zeros and let $\mathbf{1}$ denote the vector of all ones (the dimensions of these vectors will be given or obvious from context).

In what follows, we let A_τ denote the subset of matrix A with rows indexed by the set τ . We let $\lambda_{\min}(A)$, $\lambda_{\min+}(A)$, and $\lambda_{\max}(A)$ denote the minimum, minimum non-zero, and maximum eigenvalues of the matrix A , respectively. Additionally, we denote by A^\dagger the Moore-Penrose pseudoinverse of the matrix A .

We additionally recall the notion of a *row paving* of a matrix \mathbf{A} , which controls the conditioning of a set of submatrices that partition the matrix \mathbf{A} . We include here the definition provided by Needell and Tropp in [18], and note that earlier work on the construction of pavings for block projection methods is due to Popa [23]. In the wider operator theory context, pavings have long been a topic of interest; see [24] for a review.

Definition I.1. A (d, α, β) row paving of a matrix \mathbf{A} is a partition $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ of the row indices that satisfies

$$\alpha \leq \lambda_{\min}(\mathbf{A}_\tau \mathbf{A}_\tau^\top) \text{ and } \lambda_{\max}(\mathbf{A}_\tau \mathbf{A}_\tau^\top) \leq \beta \text{ for each } \tau \in T.$$

We will additionally be interested in subsets of the row indices that do not necessarily partition the rows. We define a *row covering* to be the natural relaxation of a row paving in which the requirement that the subsets partition the row indices be relaxed and where the parameter α provides a lower-bound for the minimum *non-zero* eigenvalue rather than the minimum eigenvalue. Each of these are important generalizations for our application of interest, gossip algorithms for average consensus.

Definition I.2. A (d, α, β, r, R) row covering of a matrix \mathbf{A} is a collection of subsets $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ of the row indices, $\tau_i \subset [m]$ for all $i = 1, \dots, d$, that covers the row indices, for each $i \in [m]$ we have $i \in \tau_l$ for some $l = 1, \dots, d$, and that satisfies

$$\alpha \leq \lambda_{\min+}(\mathbf{A}_\tau \mathbf{A}_\tau^\top) \text{ and } \lambda_{\max}(\mathbf{A}_\tau \mathbf{A}_\tau^\top) \leq \beta \text{ for each } \tau \in T,$$

where r and R are the minimum and maximum, respectively, number of blocks in which a single row appears, i.e., $r = \min_{i \in [m]} |\{\tau_l \in T : i \in \tau_l\}|$ and $R = \max_{i \in [m]} |\{\tau_l \in T : i \in \tau_l\}|$.

First, note that the minimum and maximum number of repeated occurrences of a row in the blocks, r and R , satisfy

$$1 \leq r \leq R \leq d.$$

Further, note that a (d, α, β) row paving is by definition a $(d, \alpha, \beta, 1, 1)$ row covering. Next, we remind the readers that every sufficiently tall row-normalized matrix \mathbf{A} admits a row paving with $\alpha > 0$ and $\beta < 2$ [18]; the authors illustrate that additional mild assumptions on the matrix allows one to produce such a row paving via random partitioning. Additionally, we note that if the matrix in question is an incidence matrix \mathbf{Q} , where rows correspond to edges in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, then the row covering $T = \{\tau_1, \dots, \tau_d\}$ corresponds to a collection of sets of edges such that every edge appears in at least one set; that is $\tau_i \subset \mathcal{E}$ for all $i \in [d]$ and for all $e \in \mathcal{E}$, we have $e \in \tau_l$ for some $l \in [d]$.

D. Main Results

The average consensus problem is defined over an undirected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We let $\mathbf{c} = (c^{(1)}, c^{(2)}, \dots, c^{(n)})^T$ denote the vector of secret values (i.e., $c^{(i)}$ is the secret value of the i th node) initially held by the nodes of the network. The average consensus problem is then to ensure, after some communication protocol is applied, that each node stores the averaged value $\bar{c} := \text{mean}(\mathbf{c})$; that is the final vector of updated node values is $\mathbf{c}^* = \bar{c}\mathbf{1}$ where $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones. We can formulate this problem as:

$$\text{Find } \mathbf{c}^* = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \|\mathbf{x} - \mathbf{c}\|^2 \text{ s.t. } \mathbf{Q}\mathbf{x} = \mathbf{0}. \quad (1)$$

Here $\mathbf{0} \in \mathbb{R}^{|\mathcal{E}|}$ denotes the all zeros vector. The average consensus problem may be formulated in this way using either the incidence matrix or the Laplacian matrix, $L = D - A$ where D is the diagonal matrix of node degrees and A is the adjacency matrix, or more generally as an *average consensus system* as defined in [25]. In this work, we will focus upon the *block gossip* methods for this problem, in which groups of edges are sampled and nodes connected by these edges update their value to their average (or the average of a subset of these nodes); pseudocode for this method is provided in Algorithm 1. See Figure 2 for a visualization of one step of Algorithm 1 on the average consensus problem of Figure 1 with independent edge set, path, and clique block sampling.

We next present our main result which illustrates that the block gossip method converges at least linearly in expectation to consensus if the underlying graph is connected. Moreover, we specialize this result to three special cases: the case when the blocks of edges sampled in each iteration are independent edge sets, when they are clique subgraphs, and when they are path subgraphs, and give refined convergence rates for these cases.

Corollary I.3. Suppose graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is connected, $\mathbf{Q} \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{V}|}$ is the incidence matrix for \mathcal{G} , and $T = \{\tau_1, \dots, \tau_d\}$ is a (d, α, β, r, R) row covering for \mathbf{Q} with $M = \max_{i \in [d]} |\tau_i|$. Then the block gossip method with blocks determined by T converges at least linearly in expectation with the guarantee

$$\mathbb{E} \|\mathbf{c}_k - \mathbf{c}^*\|^2 \leq \left(1 - \frac{r\alpha(\mathcal{G})}{\beta d}\right)^k \|\mathbf{c} - \mathbf{c}^*\|^2,$$

where $\alpha(\mathcal{G})$ is the algebraic connectivity of graph \mathcal{G} .

1) If T consists of independent edge sets, the rate constant can be bounded by

$$\left(1 - \frac{r\alpha(\mathcal{G})}{2d}\right).$$

2) If T consists of path subgraphs, the rate constant can be bounded by

$$\left(1 - \frac{r\alpha(G)}{(2 - 2 \cos \frac{M\pi}{M+1})d}\right) \leq \left(1 - \frac{r\alpha(\mathcal{G})}{4d}\right).$$

3) If T consists of clique subgraphs, the rate constant can be bounded by

$$\left(1 - \frac{r\alpha(G)}{(2 - 2 \cos \frac{M\pi}{M+1})d}\right) \leq \left(1 - \frac{r\alpha(\mathcal{G})}{4d}\right).$$

4) If T consists of arbitrary connected subgraphs, the rate constant can be bounded by

$$\left(1 - \frac{r\alpha(\mathcal{G})}{Md}\right).$$

Algorithm 1 Block Gossip Method

```

1: procedure BG( $\mathcal{G}, \mathbf{c}_0 = \mathbf{c}, T = \{\tau_1, \dots, \tau_d\}$ )
2:    $k = 0$ 
3:   repeat
4:      $k \leftarrow k + 1$ 
5:     Choose edge subset  $\tau$  uniformly at random from elements of  $T$ .
6:     Form  $\mathcal{G}_\tau = (\mathcal{V}_\tau, \mathcal{E}_\tau)$ , the edge-induced subgraph of  $\mathcal{G}$  defined by edges in  $\tau$ .
7:     Nodes outside of  $\mathcal{V}_\tau$  do not update secret values,  $(\mathbf{c}_k)_{\mathcal{V}_\tau^c} \leftarrow (\mathbf{c}_{k-1})_{\mathcal{V}_\tau^c}$ .
8:     for all connected components  $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$  of  $\mathcal{G}_\tau$  do
9:       Nodes in  $\mathcal{V}'$  solve average consensus on  $\mathcal{G}'$ ,  $(\mathbf{c}_k)_{\mathcal{V}' } \leftarrow \left[ \frac{1}{|\mathcal{V}'|} \sum_{i \in \mathcal{V}'} (\mathbf{c}_{k-1})_i \right] \mathbf{1}_{|\mathcal{V}'|}$ .
10:    end for
11:  until stopping criterion reached
12:  return  $\mathbf{c}_k$ 
13: end procedure

```

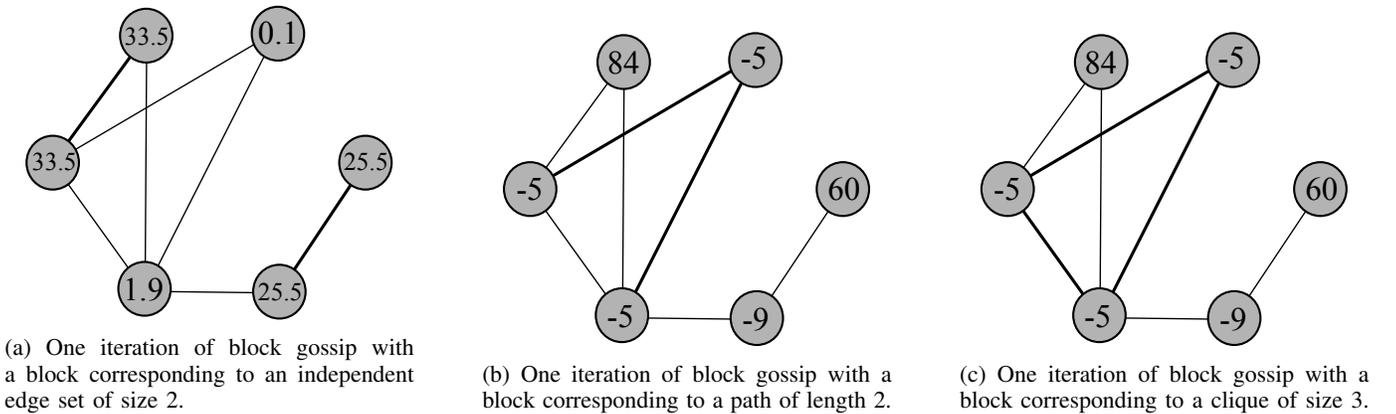


Fig. 2: The average consensus problem of Figure 1 after an iteration of the block gossip method (Algorithm 1) with various types of block structures. The edges defining the sampled block are represented by bold lines.

This result follows from a new bound on the convergence rate of the *block Kaczmarz method* on a potentially rank-deficient least-squares problem. The block Kaczmarz method samples blocks of rows of the matrix \mathbf{A} in each iteration and performs an update which projects the previous iterate onto the solution space of the subset of sampled equations; note that the standard single-row Kaczmarz updates are a special case of block Kaczmarz with block size one. The details of this method are provided in Algorithm 2. The block gossip method with blocks T produces the same iterates as the block Kaczmarz method performed with $\mathbf{A} = \mathbf{Q}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{x}_0 = \mathbf{c}$ with blocks T .

Our main result regarding the block Kaczmarz method generalizes the main result of [18] in several ways:

Algorithm 2 Block Kaczmarz Method

```

1: procedure BK( $\mathbf{A}, \mathbf{b}, \mathbf{x}_0, T = \{\tau_1, \dots, \tau_d\}$ )
2:    $k = 0$ 
3:   repeat
4:      $k \leftarrow k + 1$ 
5:     Choose row block  $\tau$  uniformly at random from  $T$ .
6:      $\mathbf{x}_k \leftarrow \mathbf{x}_{k-1} + \mathbf{A}_\tau^\dagger(\mathbf{b}_\tau - \mathbf{A}_\tau \mathbf{x}_{k-1})$ 
7:   until stopping criterion reached
8:   return  $\mathbf{x}_k$ 
9: end procedure

```

- Generalizes to the case when the least-squares problem is rank-deficient.
- Relaxes the requirement that the row blocks be sampled from a matrix paving.
- Demonstrates that the convergence horizon depends upon the minimum *nonzero* singular value of the blocks \mathbf{A}_τ rather than the absolute minimum singular value (often 0).

These generalizations are important in our main application to the average consensus problem and block gossip methods, but are likely of independent interest in other applications.

Theorem I.4. Consider the least-squares problem

$$\min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is not necessarily full-rank and $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{e} = \mathbf{A}\mathbf{x}_e - \mathbf{b}$ for some \mathbf{x}_e and let $T = \{\tau_1, \dots, \tau_d\}$ be a (d, α, β, r, R) covering (not necessarily a paving) of the rows of \mathbf{A} . Let \mathbf{x}_j denote the j th iterate produced by Block RK on the system defined by \mathbf{A} and \mathbf{b} with initial iterate \mathbf{x}_0 , and let $\mathbf{x}^* := (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}_0 + \mathbf{A}^\dagger(\mathbf{b} + \mathbf{e}) = (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}_0 + \mathbf{A}^\dagger \mathbf{A} \mathbf{x}_e$. Then we have

$$\mathbb{E} \left(\|\mathbf{x}_j - \mathbf{x}^*\|^2 \right) \leq \rho(\mathbf{A}, T)^j \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{\beta R}{\alpha r \sigma_{\min+}^2(\mathbf{A})} \|\mathbf{e}\|^2, \quad (2)$$

where $\rho(\mathbf{A}, T) = 1 - \frac{r \sigma_{\min+}^2(\mathbf{A})}{\beta d}$ and $\sigma_{\min+}(\mathbf{A})$ is the smallest nonzero singular value of \mathbf{A} .

Remark I.5. The convergence horizon term, $\frac{\beta R}{\alpha r \sigma_{\min+}^2(\mathbf{A})} \|\mathbf{e}\|^2$, is minimized in the case that $\mathbf{e} = \mathbf{e}^*$ where \mathbf{e}^* is the minimum norm residual, i.e.,

$$\mathbf{e}^* = \operatorname{argmin} \|\mathbf{e}\|^2 \text{ s.t. } \mathbf{e} = \mathbf{A}\mathbf{x} - \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^n.$$

In this case, the iterates converge to $\mathbf{x}^* := (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}_0 + \mathbf{A}^\dagger(\mathbf{b} + \mathbf{e}^*)$.

Remark I.6. In other Kaczmarz literature, including Needell's original result for Randomized Kaczmarz applied to inconsistent linear systems [26, Theorem 2.1], the problem is formulated by adding some vector of measurement noise \mathbf{r} to a consistent system $\mathbf{A}\mathbf{x} = \mathbf{b}$, leading to a convergence horizon term proportional to $\|\mathbf{r}\|$. We instead work with $\mathbf{A}\mathbf{x} = \mathbf{b}$ being inconsistent and give a horizon proportional to a residual, but we note that the two formulations are equivalent.

Remark I.7. We include a generalization of Theorem I.4 to the case when the right hand side, \mathbf{b} , is varying in each iteration according to mean-zero randomly distributed additive noise in Proposition IV.2.

E. Related Work

In this subsection we offer some related reading in the fields of average consensus, gossip protocols, and block iterative methods, and draw connections between them and our work.

Average consensus and gossip protocols. As mentioned, average consensus has been a fundamental topic in distributed computing since the inception of the field. We refer the reader to the classical work of DeGroot [27] for an inception of the consensus problem, and to the work of Tsitsiklis, Bertsekas, and Athans [28] for a first look into stochastic protocols for distributed computing. As networks have grown larger in size and have appeared in more applications, the need for more efficient average consensus solvers motivated the development of *gossip algorithms*: protocols that, in general, select some subset of nodes and allow them to ‘gossip’, i.e., share and average their stored values. Boyd et al.’s 2006 paper [7] provided a fundamental exposition of said protocols, particularly on the connection between their convergence rate and the underlying network topology, and has motivated research in the topic ever since [9].

There have been a variety of works analyzing different node-selection protocols for gossiping. In [20], at each epoch a path of nodes in the network is formed and the values along said path are averaged. In [21], the network is decomposed beforehand into cliques (connected subgraphs), and at each epoch one such clique is activated. Lastly in [7], at each epoch a selection

of pairs are chosen and each pair computes its own average (we call this *edge independent set* gossiping throughout). These analyses are, however, somewhat disjoint, so we believe the unified convergence analysis presented in this paper (which covers all of the aforementioned methods) to be novel.

Block randomized Kaczmarz. The Kaczmarz method [29] is an iterative linear system solver whose popularity boomed after its randomized variant was proven to have exponential convergence by Strohmer and Vershynin [30]. After this work proved convergence for full-rank, consistent systems, further work was done to generalize to the case of inconsistent [26] and rank-deficient [31] systems.

A well-studied family of variants are *block Kaczmarz methods*, in which iterates are projected onto subspaces corresponding to blocks of rows rather than single equations. Early references include [23], [32], [33], but we focus on the block randomized Kaczmarz method introduced by Needell and Tropp [18]. The authors proved that under certain restrictions on the choice of blocks, the method achieves exponential convergence, and converges up to a threshold if the system is inconsistent. In our work we significantly relax these conditions and achieve a similar convergence guarantee.

The connection between gossip algorithms and Kaczmarz methods for linear systems was analyzed in depth by Loizou and Richtárik in their 2019 paper [16]; others considering this connection include [15], [34]. This connection was exploited in [16] to build a framework giving new convergence guarantees for gossip protocols, accelerations via momentum, and other interesting discussions such as dual gossip algorithms. In our work we hone in on their general exposition of block gossip algorithms, and give more explicit links and convergence guarantees for previously mentioned existing gossip protocols.

II. CONVERGENCE OF BLOCK RANDOMIZED KACZMARZ

In this section, we prove our main result, Theorem I.4, which illustrates that the block randomized Kaczmarz method converges at least linearly in expectation on least-squares problems even in the case that the matrix \mathbf{A} is rank-deficient and the blocks are not sampled from a matrix paving. We will then illustrate how this result specializes to prove Corollary I.3.

Proof of Theorem I.4. First, note that by definition, $\mathbf{e} = \mathbf{A}\mathbf{x}_e - \mathbf{b}$ and so

$$\begin{aligned} \mathbf{A}\mathbf{x}^* - \mathbf{b} &= (\mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{x}_0) + \mathbf{A}\mathbf{A}^\dagger\mathbf{b} + \mathbf{A}\mathbf{A}^\dagger\mathbf{e} - \mathbf{b} \\ &= \mathbf{A}\mathbf{A}^\dagger\mathbf{b} + \mathbf{A}\mathbf{A}^\dagger(\mathbf{A}\mathbf{x}_e - \mathbf{b}) - \mathbf{b} \\ &= \mathbf{A}\mathbf{x}_e - \mathbf{b} = \mathbf{e}, \end{aligned} \quad (3)$$

where the second and third equation used the fact that $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.

Recall that our updates take the form $\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{A}_\tau^\dagger(\mathbf{b}_\tau - \mathbf{A}_\tau\mathbf{x}_j)$, where τ is chosen uniformly at random from our set of blocks. We then have

$$\begin{aligned} \|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}_j + \mathbf{A}_\tau^\dagger(\mathbf{b}_\tau - \mathbf{A}_\tau\mathbf{x}_j) - \mathbf{x}^*\|^2 \\ &= \|(\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau)\mathbf{x}_j + \mathbf{A}_\tau^\dagger(\mathbf{A}_\tau\mathbf{x}^* - \mathbf{e}_\tau) - \mathbf{x}^*\|^2 \\ &= \|(\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau)(\mathbf{x}_j - \mathbf{x}^*) - \mathbf{A}_\tau^\dagger\mathbf{e}_\tau\|^2 \\ &= \|(\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau)(\mathbf{x}_j - \mathbf{x}^*)\|^2 + \|\mathbf{A}_\tau^\dagger\mathbf{e}_\tau\|^2, \end{aligned} \quad (4)$$

where we used that $\mathbf{A}_\tau^\dagger(\mathbf{b}_\tau + \mathbf{e}_\tau) = \mathbf{A}_\tau^\dagger\mathbf{A}_\tau\mathbf{x}^*$ and that $\text{Im}(\mathbf{A}_\tau^\dagger)$ and $\text{Im}(\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau)$ are orthogonal. Then since $\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau$ is an orthogonal projector, we have

$$\|(\mathbf{I} - \mathbf{A}_\tau^\dagger\mathbf{A}_\tau)(\mathbf{x}_j - \mathbf{x}^*)\|^2 = \|\mathbf{x}_j - \mathbf{x}^*\|^2 - \|\mathbf{A}_\tau^\dagger\mathbf{A}_\tau(\mathbf{x}_j - \mathbf{x}^*)\|^2. \quad (5)$$

Note that $\|\mathbf{A}_\tau^\dagger\mathbf{e}_\tau\|^2 \leq \sigma_{\max}^2(\mathbf{A}_\tau^\dagger) \|\mathbf{e}_\tau\|^2 = \frac{1}{\sigma_{\min}^2(\mathbf{A}_\tau)} \|\mathbf{e}_\tau\|^2 \leq \frac{1}{\alpha} \|\mathbf{e}_\tau\|^2$. Using this fact and taking expectations, we obtain

$$\mathbb{E}_j \left(\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 \right) \quad (6)$$

$$\begin{aligned} &\leq \|\mathbf{x}_j - \mathbf{x}^*\|^2 - \mathbb{E}_j \left(\|\mathbf{A}_\tau^\dagger\mathbf{A}_\tau(\mathbf{x}_j - \mathbf{x}^*)\|^2 \right) + \frac{1}{\alpha} \mathbb{E}_j \left(\|\mathbf{e}_\tau\|^2 \right) \\ &\leq \|\mathbf{x}_j - \mathbf{x}^*\|^2 - \frac{1}{\beta} \mathbb{E}_j \left(\|\mathbf{A}_\tau(\mathbf{x}_j - \mathbf{x}^*)\|^2 \right) + \frac{1}{\alpha} \mathbb{E}_j \left(\|\mathbf{e}_\tau\|^2 \right) \\ &\leq \|\mathbf{x}_j - \mathbf{x}^*\|^2 - \frac{r}{\beta d} \|\mathbf{A}(\mathbf{x}_j - \mathbf{x}^*)\|^2 + \frac{R}{\alpha d} \|\mathbf{e}\|^2, \end{aligned} \quad (7)$$

where the last inequality follows from the fact that

$$\begin{aligned} \mathbb{E}_j(\|\mathbf{v}_\tau\|^2) &= \frac{1}{d} \sum_{l=1}^d \|\mathbf{v}_{\tau_l}\|^2 = \frac{1}{d} \sum_{l=1}^d \sum_{i=1}^m \mathbf{1}(i \in \tau_l) v_i^2 \\ &= \frac{1}{d} \sum_{i=1}^m \left[\sum_{l=1}^d \mathbf{1}(i \in \tau_l) \right] v_i^2 \end{aligned}$$

so we have $\frac{r}{d} \|\mathbf{v}\|^2 \leq \mathbb{E}_j(\|\mathbf{v}_\tau\|^2) \leq \frac{R}{d} \|\mathbf{v}\|^2$.

We now claim that for all j , $\mathbf{x}_j - \mathbf{x}^* \in \text{Im}(\mathbf{A}^T)$, i.e., the row space of \mathbf{A} . We do so by induction; firstly for $j = 0$ we have

$$\mathbf{x}_0 - \mathbf{x}^* = \mathbf{x}_0 - (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}_0 - \mathbf{A}^\dagger(\mathbf{b} + \mathbf{e}) = \mathbf{A}^\dagger(\mathbf{A}\mathbf{x}_0 - (\mathbf{b} + \mathbf{e})),$$

and since $\text{Im}(\mathbf{A}^\dagger) = \text{Im}(\mathbf{A}^T)$, we are done.

Now assume $\mathbf{x}_l - \mathbf{x}^* \in \text{Im}(\mathbf{A}^T)$. Then we have, for some τ ,

$$\mathbf{x}_{l+1} - \mathbf{x}^* = \mathbf{x}_l + \mathbf{A}_\tau^\dagger(\mathbf{b}_\tau - \mathbf{A}_\tau \mathbf{x}_l) - \mathbf{x}^*.$$

By assumption $\mathbf{x}_l - \mathbf{x}^* \in \text{Im}(\mathbf{A}^T)$, and furthermore since \mathbf{A}_τ is a row submatrix of \mathbf{A} , we have

$$\text{Im}(\mathbf{A}_\tau^\dagger) = \text{Im}(\mathbf{A}_\tau^T) \subseteq \text{Im}(\mathbf{A}^T).$$

Thus $\mathbf{x}_{l+1} - \mathbf{x}^* \in \text{Im}(\mathbf{A}^T)$, and we are done by induction.

Now, returning to (6), since $\mathbf{x}_j - \mathbf{x}^* \in \text{Im}(\mathbf{A}^T) = \text{Ker}(\mathbf{A})^\perp$, we have

$$\|\mathbf{A}(\mathbf{x}_j - \mathbf{x}^*)\|^2 \geq \sigma_{\min+}^2(\mathbf{A}) \|\mathbf{x}_j - \mathbf{x}^*\|^2.$$

This yields

$$\mathbb{E}_j \left(\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 \right) \leq \rho(\mathbf{A}, T) \|\mathbf{x}_j - \mathbf{x}^*\|^2 + \frac{R}{\alpha d} \|\mathbf{e}\|^2,$$

where $\rho(\mathbf{A}, T) = 1 - \frac{r\sigma_{\min+}^2(\mathbf{A})}{\beta d}$ and by induction we obtain

$$\mathbb{E} \left(\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 \right) \tag{8}$$

$$\leq \rho(\mathbf{A}, T)^{j+1} \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \left[\sum_{i=0}^j \rho(\mathbf{A}, T)^i \right] \frac{R}{\alpha d} \|\mathbf{e}\|^2$$

$$\leq \rho(\mathbf{A}, T)^{j+1} \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \left[\sum_{i=0}^{\infty} \rho(\mathbf{A}, T)^i \right] \frac{R}{\alpha d} \|\mathbf{e}\|^2 \tag{9}$$

$$= \rho(\mathbf{A}, T)^{j+1} \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{\beta R}{\alpha r \sigma_{\min+}^2(\mathbf{A})} \|\mathbf{e}\|^2.$$

□

Remark II.1. We note that in most applications, including our application of average consensus, $\sigma_{\min+}(\mathbf{A})$ is fixed, and so it is natural to seek to maximize $\frac{r}{\beta d}$. For the average consensus problem, this equates to careful selection of the block set T .

We next include a proof of Corollary I.3 which follows from Theorem I.4 due to the fact that block gossip on the network \mathcal{G} with initial secret node values \mathbf{c} coincides with block randomized Kaczmarz on the problem (1).

Proof of Corollary I.3. Consider running block RK with $\mathbf{A} = \mathbf{Q}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{x}_0 = \mathbf{c}$ with samples $\tau \in T$ determined by the run of block gossip on \mathcal{G} with $\mathbf{c}_0 = \mathbf{c}$. Note that $\text{ker}(\mathbf{Q})$ is nonempty since $\text{rank}(\mathbf{Q}) \leq n - 1$, so $\mathbf{e} = \mathbf{0}$. It follows from [16, Theorem 5] that block gossip iterate \mathbf{c}_k and block RK iterate \mathbf{x}_k coincide for all k .

If we take $\mathbf{x}_0 = \mathbf{c}$, then by Theorem I.4 we know that block RK will converge to $(\mathbf{I} - \mathbf{Q}^\dagger \mathbf{Q})\mathbf{c}$ (since $\mathbf{b} = \mathbf{e} = \mathbf{0}$ in this application). This is exactly the orthogonal projection of \mathbf{c} onto $\text{ker}(\mathbf{Q}) = \text{span}\{\mathbf{1}\}$, where $\mathbf{1}$ is the length- $|\mathcal{E}|$ vector of all ones. Then, since $\left\{ \frac{1}{\sqrt{|\mathcal{E}|}} \mathbf{1} \right\}$ is an orthonormal basis for $\text{ker}(\mathbf{Q})$, we can compute this projection as

$$(\mathbf{I} - \mathbf{Q}^\dagger \mathbf{Q})\mathbf{c} = \left\langle \mathbf{c}, \frac{1}{\sqrt{|\mathcal{E}|}} \mathbf{1} \right\rangle \frac{1}{\sqrt{|\mathcal{E}|}} \mathbf{1} \tag{10}$$

$$= \frac{1}{|\mathcal{E}|} \left(\sum_{i=1}^{|\mathcal{E}|} c_i \right) \mathbf{1} \tag{11}$$

$$= \mathbf{c}^*. \tag{12}$$

Finally, note that $\sigma_{\min+}^2(\mathbf{Q}) = \lambda_{\min+}(\mathbf{L}) = \alpha(\mathcal{G})$ where \mathbf{L} is the Laplacian matrix of \mathcal{G} . The specific results enumerated in Corollary I.3 follow from the singular value upper bounds presented in Section III. □

III. BLOCK GOSSIP SAMPLING

In this section, we consider particular cases when the blocks used in the block gossip method correspond to special subgraph structures, namely independent edge sets, clique subgraphs, path subgraphs, and arbitrary connected subgraphs.

We begin with a lemma that will be used to strengthen our convergence results for these cases.

Lemma III.1. *Let τ be a subset of edges of \mathcal{G} , and let $\tau' \subseteq \tau$ be the edge set of a spanning tree of \mathcal{G}_τ . Then the block gossip updates produced by choosing blocks τ and τ' are identical.*

Proof. This follows immediately from the fact that \mathcal{G}_τ and $\mathcal{G}_{\tau'}$ have the same vertex set, and block gossip simply averages the stored values of said vertex set at each iteration. \square

The value of this lemma comes from the fact that our convergence rate depends (in part) on the maximum singular values of our blocks: removing rows from a matrix (i.e., using a subset of edges in our block) decreases said singular values, improving the convergence rate.

Throughout this section, we make use of the fact that for any collection of row indices τ , the spectra of $\mathbf{Q}_\tau \mathbf{Q}_\tau^\top$ and \mathbf{L}_τ are the same, up to zeros. In particular, we can identify our covering constants α and β by analyzing the spectrum of \mathbf{L}_τ , which is well understood for many of the graph structures we will consider.

A. Independent edge set blocks

In the case that $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ is a row covering of \mathbf{Q} where each \mathcal{G}_{τ_i} is an independent edge set, we have that

$$\mathbf{Q}_{\tau_i} \mathbf{Q}_{\tau_i}^\top = 2\mathbf{I}$$

for each $i \in [d]$. Thus, T is a $(d, 2, 2, r, R)$ row covering of \mathbf{Q} and we have that

$$\left(1 - \frac{r\alpha(\mathcal{G})}{\beta d}\right) = \left(1 - \frac{r\alpha(\mathcal{G})}{2d}\right).$$

Note that each independent edge set is its own spanning tree, so there is no improvement to be made here via Lemma III.1.

B. Path blocks

In the case that $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ is a row covering of \mathbf{Q} where each \mathcal{G}_{τ_i} is a path, we make use of the following fact about the eigenvalues of the Laplacian of a path subgraph (see e.g., [35]).

Fact III.2. *Let P_n be a path subgraph of G of length n . Then the eigenvalues of L_{P_n} are*

$$2 - 2 \cos \frac{\pi k}{n} \quad \text{for } k = 0, 1, \dots, n-1. \quad (13)$$

We then have for each $i \in [d]$ that

$$\lambda_{\max}(\mathbf{Q}_{\tau_i} \mathbf{Q}_{\tau_i}^\top) = \lambda_{\max}(\mathbf{Q}_{\tau_i}^\top \mathbf{Q}_{\tau_i}) = 2 - 2 \cos \frac{|\tau_i| \pi}{|\tau_i| + 1} \quad (14)$$

$$\leq 2 - 2 \cos \frac{M\pi}{M+1}, \quad (15)$$

so T is a $(d, \alpha, 2 - 2 \cos \frac{M\pi}{M+1}, r, R)$ row covering. Thus we have that

$$\left(1 - \frac{r\alpha(\mathcal{G})}{\beta d}\right) = \left(1 - \frac{r\alpha(\mathcal{G})}{(2 - 2 \cos \frac{M\pi}{M+1})d}\right) \leq \left(1 - \frac{r\alpha(\mathcal{G})}{4d}\right).$$

Again, each path is its own spanning tree, so this rate cannot be improved via Lemma III.1.

C. Clique blocks

In the case that $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ is a row covering of \mathbf{Q} where each \mathcal{G}_{τ_i} is a complete subgraph of \mathcal{G} , we make use of Lemma III.1; in particular, for each $\tau \in T$ there exists $\tau' \subset \tau$ such that $\mathcal{G}_{\tau'}$ is a spanning path of \mathcal{G}_τ . See Figures 2b and 2c for an example of how a spanning path block update coincides with the update produced by a clique block update.

In this way, we see that the bound on the convergence rate constant for complete subgraphs must be no larger than that of path subgraphs and we recover the constant

$$\left(1 - \frac{r\alpha(\mathcal{G})}{(2 - 2 \cos \frac{M\pi}{M+1})d}\right) \leq \left(1 - \frac{r\alpha(\mathcal{G})}{4d}\right).$$

D. Arbitrary connected subgraph blocks

In the case that $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ is a row covering of \mathbf{Q} where each \mathcal{G}_{τ_i} is an arbitrary connected subgraph of \mathcal{G} , we again make use of the fact that we may replace every block $\tau \in T$ with a block $\tau' \subset \tau$ such that $\mathcal{G}_{\tau'}$ is a spanning tree of \mathcal{G}_τ to form $T' = \{\tau'_1, \tau'_2, \dots, \tau'_d\}$. We note that the block gossip method with blocks sampled from T will produce the same set of iterates as those sampled identically from T' .

Now, we use that fact that the eigenvalues of the Laplacian of a tree, $L_{\mathcal{T}}$, are bounded above by $|\mathcal{E}(\mathcal{T})|$ (and that in fact this bound is tight for the star graph), see [35]. Thus, we have

$$\lambda_{\max}(\mathbf{Q}_{\tau'_i} \mathbf{Q}_{\tau'_i}^\top) = \lambda_{\max}(\mathbf{Q}_{\tau'_i}^\top \mathbf{Q}_{\tau'_i}) \leq |\tau_i| \leq M.$$

We use this bound to recover the constant upper bound

$$\left(1 - \frac{r\alpha(\mathcal{G})}{Md}\right).$$

E. Multiple subgraph blocks

If the network allows for multiple disjoint components to be activated at a single instance, we may form blocks consisting of multiple disjoint subgraphs of \mathcal{G} . In this case, to compute β we may use the following Lemma.

Lemma III.3. *Let \mathcal{G}_τ be a subgraph of \mathcal{G} consisting of disjoint connected subgraphs $\mathcal{G}_{\tau_1}, \dots, \mathcal{G}_{\tau_k}$. Then*

$$\lambda_{\max}(\mathbf{Q}_\tau \mathbf{Q}_\tau^\top) = \max_i \lambda_{\max}(\mathbf{Q}_{\tau_i} \mathbf{Q}_{\tau_i}^\top). \quad (16)$$

Proof. This follows immediately from the fact that since said subgraphs are edge-disjoint, we have that the Laplacian of their union is the direct sum of their individual Laplacians:

$$\mathbf{L}_\tau = \mathbf{L}_{\tau_1} \oplus \dots \oplus \mathbf{L}_{\tau_k}, \quad (17)$$

and so the spectrum of \mathbf{L}_τ is exactly the union of the spectra of $\mathbf{L}_{\tau_1}, \dots, \mathbf{L}_{\tau_k}$. \square

Suppose then that we take $T = \{\tau_1, \tau_2, \dots, \tau_d\}$, where each τ_i is the union of a collection of rows corresponding to disjoint connected subgraphs, say $\tau_i = \bigcup_j \tau_i^j$. Then by Lemma III.3 we have

$$\lambda_{\max}(\mathbf{Q}_{\tau_i} \mathbf{Q}_{\tau_i}^\top) = \max_j \lambda_{\max}(\mathbf{Q}_{\tau_i^j} \mathbf{Q}_{\tau_i^j}^\top).$$

One may then apply the relevant previous results of Section III to compute an upper bound on this quantity, yielding β .

IV. INCONSISTENT CONSENSUS MODELS

In this section, we consider two models of inconsistent average consensus where communication across edges is noisy and provide analyses of the natural block gossip method in these cases.

A. Constant edge communication error

Consider the average consensus problem in the presence of constant edge communication error; that is, some blocks of nodes do not update to local consensus during iterations of the block gossip method, but instead update according to an attempt to satisfy constant edge miscommunication values. We denote $\mathbf{m} \in \mathbb{R}^{|\mathcal{E}|}$ as the edge miscommunication values and consider the block gossip updates under this edge miscommunication to be

$$\mathbf{c}_k = \mathbf{c}_{k-1} + \mathbf{Q}_\tau^\dagger (\mathbf{m}_\tau - \mathbf{Q}_\tau \mathbf{c}_{k-1}). \quad (18)$$

We can apply Theorem I.4 to prove the following corollary. This result yields a guarantee of convergence to a convergence horizon that depends upon the edge miscommunication vector \mathbf{m} .

Corollary IV.1. *Suppose graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is connected, $\mathbf{Q} \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{V}|}$ is the incidence matrix for \mathcal{G} , and $T = \{\tau_1, \dots, \tau_d\}$ is a (d, α, β, r, R) row covering for \mathbf{Q} . Then the block gossip method under edge miscommunication \mathbf{m} as defined in (18) with blocks determined by T converges at least linearly in expectation to a horizon determined by \mathbf{m} with the guarantee*

$$\mathbb{E} \|\mathbf{c}_k - \mathbf{c}^*\|^2 \leq \left(1 - \frac{r\alpha(\mathcal{G})}{\beta d}\right)^k \|\mathbf{c} - \mathbf{c}^*\|^2 + \frac{\beta R}{\alpha r \alpha(\mathcal{G})} \|\mathbf{m}\|^2,$$

where $\alpha(\mathcal{G})$ is the algebraic connectivity of graph \mathcal{G} .

Proof. This result follows from Theorem I.4 where $\mathbf{b} = \mathbf{m}$, $\mathbf{A} = \mathbf{Q}$, and $\mathbf{e} = -\mathbf{m}$. Note that $\mathbf{x}^* = (\mathbf{I} - \mathbf{Q}^\dagger \mathbf{Q})\mathbf{c} + \mathbf{Q}^\dagger (\mathbf{m} - \mathbf{m}) = (\mathbf{I} - \mathbf{Q}^\dagger \mathbf{Q})\mathbf{c} = \mathbf{c}^*$. \square

B. Randomly varying edge communication error

We now consider the average consensus problem in the presence of randomly varying edge communication error. During the block gossip method, blocks do not update to local consensus but instead update to attempt to satisfy the iteration dependent edge miscommunication values. We denote $\mathbf{m}_k \in \mathbb{R}^{|\mathcal{E}|}$ as the edge miscommunication values during the k th iteration and consider the block gossip updates under edge miscommunication to be

$$\mathbf{c}_k = \mathbf{c}_{k-1} + \mathbf{Q}_\tau^\dagger((\mathbf{m}_k)_\tau - \mathbf{Q}_\tau \mathbf{c}_{k-1}). \quad (19)$$

We prove a generalization of Theorem I.4 and use it to prove a guarantee of convergence to a convergence horizon that depends upon the distribution of the edge miscommunication values.

Proposition IV.2. *Let $\mathbf{b} \in r(\mathbf{A})$ with \mathbf{A} not necessarily full rank. Consider running the block Kaczmarz method with matrix \mathbf{A} and vector $\mathbf{b}_k = \mathbf{b} + \mathbf{e}_k$ in the k th iteration; that is*

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{A}_\tau^\dagger((\mathbf{b}_k)_\tau - \mathbf{A}_\tau \mathbf{x}_{k-1}).$$

Assume that $\{\mathbf{e}_k\}$ is sampled i.i.d. according to distribution \mathcal{D} , $\mathbf{e}_k \sim \mathcal{D}$, with $\mathbb{E}_{\mathcal{D}}[\mathbf{e}_k] = \mathbf{0}$ and $\text{cov}(\mathbf{e}_k) = \mathbb{E}_{\mathcal{D}}[\mathbf{e}_k \mathbf{e}_k^\top] = \mathbf{\Sigma}$. Let $T = \{\tau_1, \tau_2, \dots, \tau_d\}$ be a (d, α, β, r, R) row covering of \mathbf{A} . Let $\mathbf{x}^* = (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}_0 + \mathbf{A}^\dagger \mathbf{b}$. Then we have

$$\mathbb{E}\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 \leq \rho(\mathbf{A}, T)^{j+1} \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{\beta R}{\alpha r \sigma_{\min}^2(\mathbf{A})} \text{tr}(\mathbf{\Sigma}),$$

where $\rho(\mathbf{A}, T) = 1 - \frac{r \sigma_{\min}^2(\mathbf{A})}{\beta d}$.

Proof. This proof proceeds in a manner highly similar to that of Theorem I.4. First, note that in a calculation similar to (3), we have that $\mathbf{A}\mathbf{x}^* - \mathbf{b}_k = \mathbf{e}_k$. Then, recalling that the updates take the form $\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{A}_\tau^\dagger((\mathbf{b}_k)_\tau - \mathbf{A}_\tau \mathbf{x}_j)$ where τ is chosen uniformly from T , we compute

$$\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 = \|(\mathbf{I} - \mathbf{A}_\tau^\dagger \mathbf{A}_\tau)(\mathbf{x}_j - \mathbf{x}^*)\|^2 + \|\mathbf{A}_\tau^\dagger(\mathbf{e}_j)_\tau\|^2$$

in a manner similar to that of (4). Using (5) and taking expectation with respect to the sampled block in iteration j , τ_j , conditioned on all previously sampled blocks, we have

$$\begin{aligned} \mathbb{E}_{\tau_j} \|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}_j - \mathbf{x}^*\|^2 - \mathbb{E}_j \|\mathbf{A}_\tau^\dagger \mathbf{A}_\tau (\mathbf{x}_j - \mathbf{x}^*)\|^2 + \frac{1}{\alpha} \mathbb{E}_j \|(\mathbf{e}_j)_\tau\|^2 \\ &\leq \left(1 - \frac{r \sigma_{\min}^2(\mathbf{A})}{\beta d}\right) \|\mathbf{x}_j - \mathbf{x}^*\|^2 + \frac{R}{\alpha d} \|\mathbf{e}_j\|^2. \end{aligned}$$

Now, taking expectation with respect to the sampled error $\mathbf{e}_j \sim \mathcal{D}$ conditioned upon all previously sampled errors, we arrive at

$$\begin{aligned} \mathbb{E}_{\mathbf{e}_j} [\mathbb{E}_{\tau_j} \|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2] &\leq \left(1 - \frac{r \sigma_{\min}^2(\mathbf{A})}{\beta d}\right) \|\mathbf{x}_j - \mathbf{x}^*\|^2 + \frac{R}{\alpha d} \mathbb{E}_{\mathbf{e}_j} \|\mathbf{e}_j\|^2 \\ &= \left(1 - \frac{r \sigma_{\min}^2(\mathbf{A})}{\beta d}\right) \|\mathbf{x}_j - \mathbf{x}^*\|^2 + \frac{R}{\alpha d} \text{tr}(\mathbf{\Sigma}). \end{aligned}$$

Iterating this expectation and proceeding inductively as in (8), we arrive at the desired result. \square

We may now use this result to prove the following guarantee for convergence of block gossip methods in the presence of randomly varying edge miscommunication error.

Corollary IV.3. *Suppose graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is connected, $\mathbf{Q} \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{V}|}$ is the incidence matrix for \mathcal{G} , and $T = \{\tau_1, \dots, \tau_d\}$ is a (d, α, β, r, R) row covering for \mathbf{Q} . Assume that $\{\mathbf{m}_k\}$ is sampled i.i.d. according to distribution \mathcal{D} , $\mathbf{m}_k \sim \mathcal{D}$, with $\mathbb{E}_{\mathcal{D}}[\mathbf{m}_k] = \mathbf{0}$ and $\text{cov}(\mathbf{m}_k) = \mathbb{E}_{\mathcal{D}}[\mathbf{m}_k \mathbf{m}_k^\top] = \mathbf{\Sigma}$. Then the block gossip method under randomly varying edge miscommunication \mathbf{m}_k as defined in (19) with blocks determined by T converges at least linearly in expectation to a horizon determined by $\text{tr}(\mathbf{\Sigma})$ with the guarantee*

$$\mathbb{E}\|\mathbf{c}_k - \mathbf{c}^*\|^2 \leq \left(1 - \frac{r \alpha(\mathcal{G})}{\beta d}\right)^k \|\mathbf{c} - \mathbf{c}^*\|^2 + \frac{\beta R}{\alpha r \alpha(\mathcal{G})} \text{tr}(\mathbf{\Sigma}),$$

where $\alpha(\mathcal{G})$ is the algebraic connectivity of graph \mathcal{G} .

Proof. This result follows from Proposition IV.2 where $\mathbf{b} = \mathbf{0}$, $\mathbf{e}_k = \mathbf{m}_k$, and $\mathbf{A} = \mathbf{Q}$. Note that $\mathbf{x}^* = (\mathbf{I} - \mathbf{Q}^\dagger \mathbf{Q})\mathbf{c} = \mathbf{c}^*$. \square

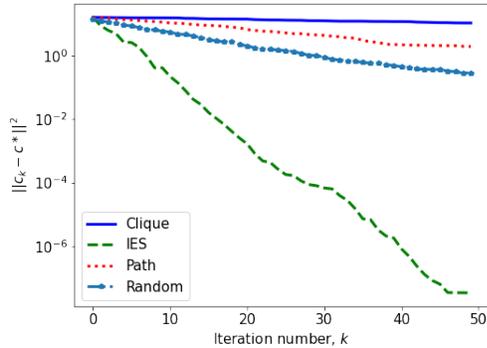


Fig. 3: Errors for all protocols applied to ER(200, 0.2).

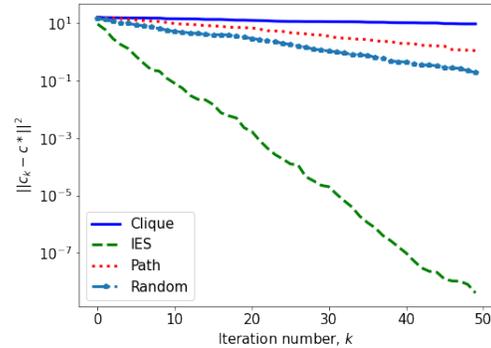


Fig. 4: Errors for all protocols applied to ER(200, 0.4).

Remark IV.4. *In the case that the blocks consist of single rows, then updates (18) and (19) correspond to nodes updating to satisfy a misspecified (nonhomogenous) equation in the AC system, which could model link communication failure. However, for larger blocks, the interpretation of these updates break down and it is less clear that they model a natural gossip process, as the individual node value updates are produced by the collection of edge miscommunications. We note, however, that such nonhomogenous systems of equations arise in practice elsewhere, e.g., in rank aggregation from pairwise comparisons via Massey’s method [36] or Hodgerank [37].*

V. EXPERIMENTS

In this section we present empirical results from applying block gossip with various choices of block structure to AC problems on multiple graph structures, including Erdős-Rényi graphs of varying connectivity and square lattice graphs. All experiments were conducted in Python 3.8 with the NetworkX [38] package used to generate and work with graph structures. We also present results from applying said protocols to inconsistent consensus models as detailed in Section IV.

A. Preliminaries

Recall that an Erdős-Rényi graph $ER(n, p)$ on n vertices is formed by randomly including edges between each pair of nodes independently with probability p . We choose to experiment on such graphs as they are popular models for real-life networks and highlight the effects of varying connectivity (by varying p) on the convergence of the considered gossip protocols. We also run experiments on $n \times n$ square lattice graphs, another widely-studied network structure; see e.g., [20].

For all graphs we perform experiments with four block sampling protocols: independent edge sets (IES), cliques, paths, and randomly selected blocks of fixed size. These protocols and the graph structures underlying them are detailed in Subsection I-D and Section III.

To produce an IES cover we use a greedy algorithm that repeatedly finds the largest independent edge set, then removes it from the graph until there are no remaining edges. Similarly, a clique edge cover of the graph is generated by a greedy algorithm that repeatedly finds the largest clique, then removes it from the graph until there are no remaining edges. For path gossip, paths are formed by selecting a node uniformly at random, adding a randomly selected neighbour to it, and continuing to sequentially add neighbours until we have a path of the desired length l . Randomly selected blocks are sampled by selecting edges uniformly at random to form a block of a specified size. The blocks generated by these algorithms are then passed into our block gossip algorithm, which randomly samples a block from the list of blocks at each iteration.

We produce two types of plot: *collapse plots*, which are a visualization of individual node values by iteration (as in [39]), and *error plots*, which show the error at each iteration, $\|c_k - c^*\|$, and for some examples also display the predicted upper bound on convergence given by Corollary I.3.

B. Erdős-Rényi Graphs

We apply each of our block sampling protocols to Erdős-Rényi graphs $ER(n, p)$ with $p = 0.2, 0.4, 0.6, 0.8, 1$ and $n = 200$.

In Figures 3 to 6 we compare the performance of each protocol (with path length $l = 10$ for path gossip) across a range of p . It appears that, in the ER case, connectivity does not have a substantial effect on the relative performance of our protocols, with IES gossip consistently being the strongest by some margin. This aligns with the fact that independent edge sets will have the greatest *node* overlap, particularly when compared to cliques – in the sense that a single node is likely to be in many more independent edge sets than cliques – and so a larger amount of information is transferred per iteration.

In Figures 7 to 10, we present a closer look at best (IES) and worst (clique) performing protocols from the previous experiments. The dramatic difference in convergence rate can be partially explained by the collapse plots: we see that during

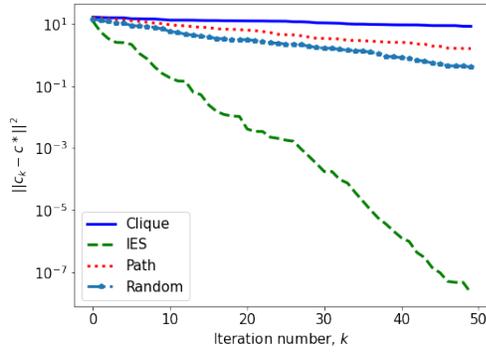


Fig. 5: Errors for all protocols applied to ER(200, 0.6).

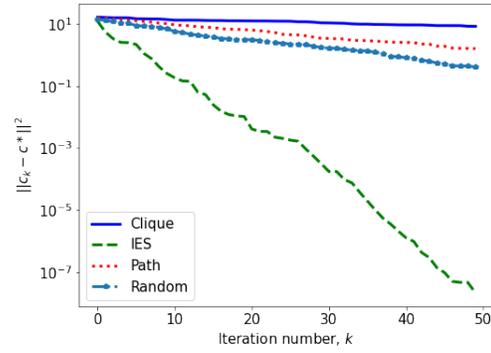


Fig. 6: Errors for all protocols applied to ER(200, 1).

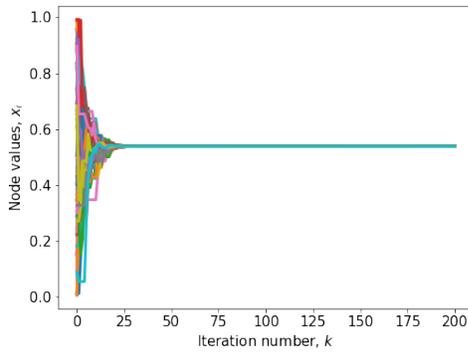


Fig. 7: Collapse plot for ER(100, 0.6) under IES gossip.

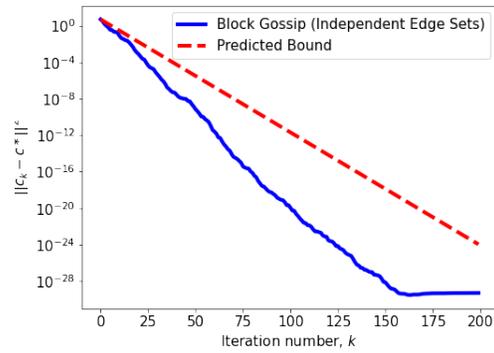


Fig. 8: Error plot for ER(100, 0.6) under IES gossip.

clique gossip certain nodes will hold their value for many iterations before updating, leading to dramatically slower convergence than IES gossip seen in the error plots. This is again connected to the greater amount of node overlap that persists in IES blocks, compared to cliques. We see from the error plots that both protocols respect the upper bound on convergence given by Corollary I.3, and that said bound predicts that IES gossip should outperform clique gossip.

We analyze the effect of increasing p (and thus the connectivity of the graph) in Figures 11 to 14. It can be seen that IES gossip is both fast and robust to variations in connectivity compared to other protocols. This can be attributed heuristically to the fact that independent edge sets are formed by selecting edges which separate nodes well, rather than selecting edges which join nodes well (as in forming cliques). The performance of clique gossip improves significantly with p , corresponding to the fact that ER(n, p) is likely to have a greater number of larger cliques as p , and thus its average degree, increases. Note that we exclude clique gossip for $p = 1$, as the entire graph will be selected as a clique and consensus will be reached in a single

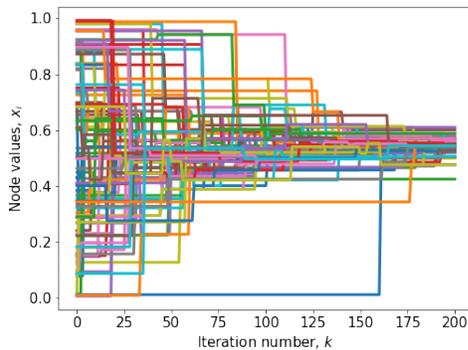


Fig. 9: Collapse plot for ER(100, 0.6) under clique gossip.

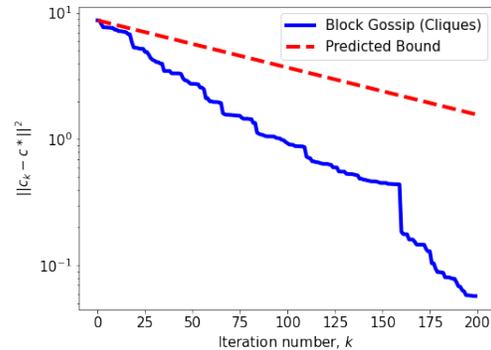


Fig. 10: Error plot for ER(100, 0.6) under clique gossip.

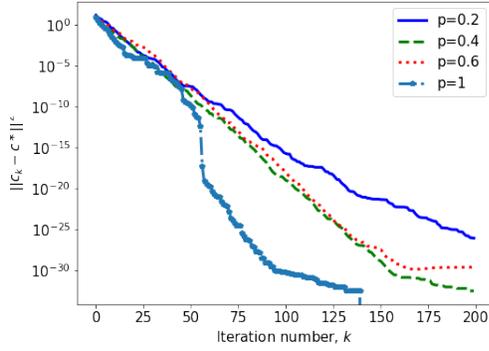


Fig. 11: IES gossip errors on $ER(200, p)$ for various p .

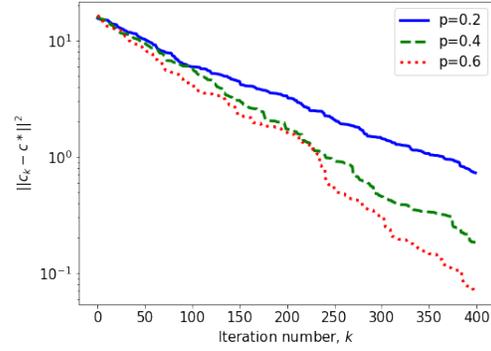


Fig. 12: Clique gossip errors on $ER(200, p)$ for various p .

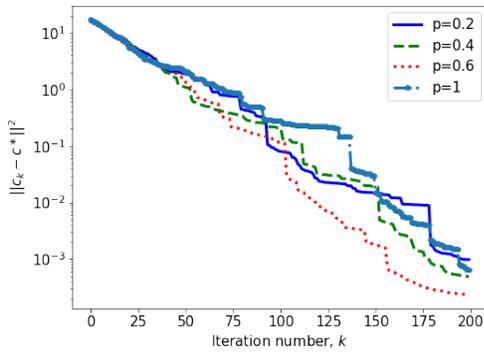


Fig. 13: Path gossip errors on $ER(200, p)$ for various p .

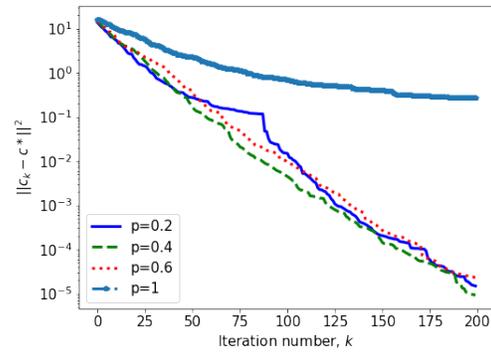


Fig. 14: Random gossip errors on $ER(200, p)$ for various p .

iteration.

C. Square Lattice

In Figure 15 we show the error plots from each of our sampling protocols applied to the 10×10 square lattice. The graph structure restricts the size of cliques to only single edges, leading to poor performance versus other protocols. Moreover, path gossip struggles as nodes on opposite sides of the grid are a large graph distance apart. In this scenario, with our ‘special’ structures being limited, it is sensible to choose blocks at random to attempt to maximise the dispersement of information, and this strategy indeed yields the best performance.

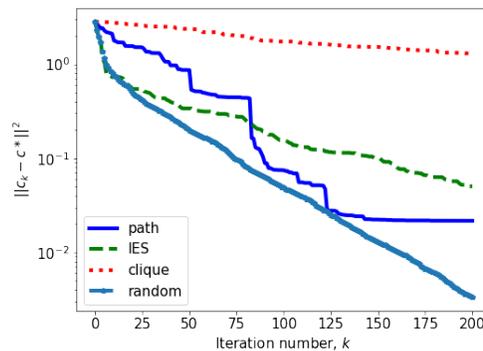


Fig. 15: Comparison of sampling protocols applied to a 10×10 square lattice.

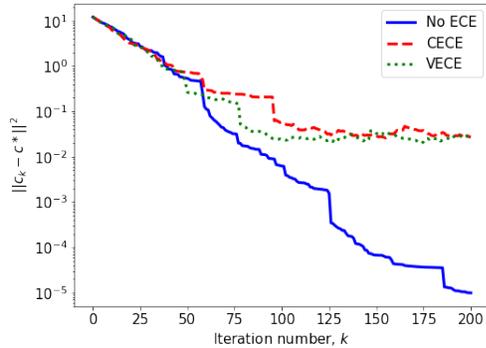


Fig. 16: Comparison of effect of different communication errors on path gossip.

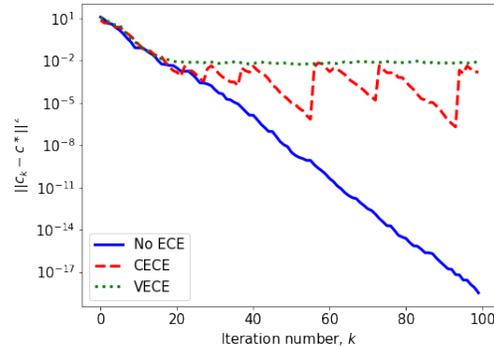


Fig. 17: Comparison of effect of different communication errors on IES gossip.

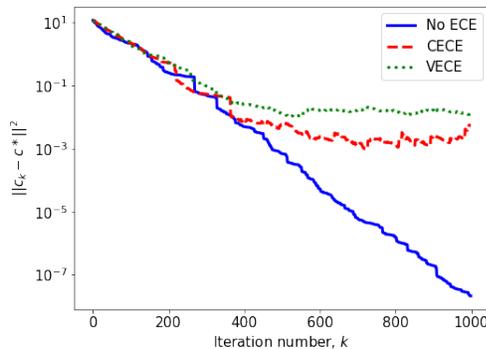


Fig. 18: Comparison of effect of different communication errors on clique gossip.

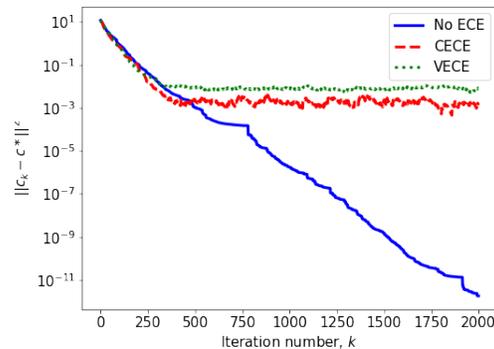


Fig. 19: Comparison of effect of different communication errors on random gossip.

D. Inconsistent Consensus Models

We experimented with the two types of inconsistent average consensus systems described in Section IV, namely systems with a constant edge communication error (CECE) and systems with a randomly varying edge communication error (VECE). In CECE, all iterations are affected by the same edge miscommunication values $\mathbf{m} \in \mathbb{R}^{|\mathcal{E}|}$ which is generated (in advance) with entries sampled independently from $\mathcal{N}(0, 0.01)$, the mean-zero Gaussian distribution with variance 0.01. In VECE, the edge miscommunication for the k th iteration, $\mathbf{m}_k \in \mathbb{R}^{|\mathcal{E}|}$, has entries sampled independently from $\mathcal{N}(0, 0.01)$.

In Figures 16 and 17, we display the errors for consistent block gossip updates ($\mathbf{m} = \mathbf{0}$), CECE updates (18) with constant edge communication values \mathbf{m} sampled as described above, and VECE updates (19) with varying edge communication values \mathbf{m}_k sampled as described above, on an $ER(n, p)$ graph with $n = 150$ and $p = 0.6$ with path block sampling and IES block sampling, respectively. Note that IES block sampling gossip converges far more quickly than path block sampling, and so we are able to see more clearly the effect of varying error in this plot. Note that VECE has a fairly smooth convergence horizon, whereas the convergence horizon for CECE varies widely as the updates sample the (fixed) values of edge miscommunication of differing magnitude.

In Figures 18 and 19, we display the errors for consistent block gossip updates ($\mathbf{m} = \mathbf{0}$), CECE updates (18) with constant edge communication values \mathbf{m} sampled as described above, and VECE updates (19) with varying edge communication values \mathbf{m}_k sampled as described above, on an $ER(n, p)$ graph with $n = 150$ and $p = 0.6$ with clique block sampling and random block sampling, respectively. Note that clique block gossip and random block gossip converge so slowly that it takes far longer to see clearly the convergence horizon for the CECE and VECE updates.

VI. CONCLUSIONS

In this work, we prove a new convergence result for the block gossip method for the average consensus problem and specialize this theoretical result to *path* [20], *clique* [21], and *edge-independent set* [7] gossip protocols. We prove this result by exploiting the fact that these methods are generalized by the block randomized Kaczmarz method for solving linear systems.

We prove a generalized convergence result for the block randomized Kaczmarz method which generalizes the main result of [18] to the case of rank-deficient systems and relaxes requirements on the set of blocks to be sampled. While these generalizations are highly important for the average consensus problem, we expect that they will be of interest in other applications as well.

We additionally prove convergence results for the block gossip method on inconsistent consensus models, and perform a broad set of experiments to compare our theoretical results to the empirical behavior of the block gossip methods on various network structures.

Future directions include further exploration of inconsistent average consensus models, including random link failure and adversarial nodes, bounded confidence models, which generalize the average consensus model, and block Kaczmarz variants for randomly varying noise.

ACKNOWLEDGMENT

The authors thank Jesús A. De Loera, Deanna Needell, Braxton Osting, and Mason Porter for useful conversations and suggestions.

REFERENCES

- [1] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Ann. IEEE Symp. Found.* IEEE, 2003, pp. 482–491.
- [2] N. M. Freris and A. Zouzias, "Fast distributed smoothing of relative measurements," in *Proc. 51st IEEE Conf. on Decision and Control (CDC)*. IEEE, 2012, pp. 1411–1416.
- [3] J. Zhang, J. Cui, Z. Wang, Y. Ding, and Y. Xia, "Distributed joint cooperative self-localization and target tracking algorithm for mobile networks," *Sensors*, vol. 19, no. 18, p. 3829, 2019.
- [4] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Symp. on Information Processing in Sensor Networks*. IEEE, 2005, pp. 63–70.
- [5] G. Cybenko, "Dynamic load balancing for distributed memory multiprocessors," *J. Parallel Distr. Com.*, vol. 7, no. 2, pp. 279–301, 1989.
- [6] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distr. Com.*, vol. 67, no. 1, pp. 33–46, 2007.
- [7] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE T. Inform. Theory*, vol. 52, pp. 2508–2530, 2006.
- [8] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE T. Signal Proces.*, vol. 57, no. 7, pp. 2748–2761, 2009.
- [9] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *P. IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.
- [10] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE T. Signal Proces.*, vol. 56, no. 3, pp. 1205–1216, 2008.
- [11] F. Hanzely, J. Konečný, N. Loizou, P. Richtárik, and D. Grishchenko, "Privacy preserving randomized gossip algorithms," *arXiv preprint arXiv:1706.07636*, 2017.
- [12] J. Liu, B. D. Anderson, M. Cao, and A. S. Morse, "Analysis of accelerated gossip algorithms," *Automatika*, vol. 49, no. 4, pp. 873–883, 2013.
- [13] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *P. IEEE*, vol. 106, no. 5, pp. 953–976, 2018.
- [14] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, 2009.
- [15] A. Zouzias and N. M. Freris, "Randomized gossip algorithms for solving Laplacian systems," in *Proc. European Control Conference (ECC)*. IEEE, 2015, pp. 1920–1925.
- [16] N. Loizou and P. Richtárik, "Revisiting randomized gossip algorithms: General framework, convergence rates and novel block and accelerated protocols," *arXiv preprint arXiv:1905.08645*, 2019.
- [17] I. Necoara, "Faster randomized block Kaczmarz algorithms," *SIAM J. Matrix Anal. A.*, vol. 40, no. 4, pp. 1425–1452, 2019.
- [18] D. Needell and J. A. Tropp, "Paved with good intentions: Analysis of a randomized block Kaczmarz method," *Linear Algebra Appl.*, 2013.
- [19] D. Needell, R. Zhao, and A. Zouzias, "Randomized block Kaczmarz method with projection for solving least squares," *Linear Algebra Appl.*, vol. 484, pp. 322–343, 2015.
- [20] F. Bénézit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Order-optimal consensus through randomized path averaging," *IEEE T. Inform. Theory*, vol. 56, no. 10, pp. 5150–5167, 2010.
- [21] Y. Liu, B. Li, B. O. Anderson, and G. Shi, "Clique gossiping," *IEEE/ACM Transactions on Networking*, vol. 27, no. 06, pp. 2418–2431, nov 2019.
- [22] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *P. IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [23] C. Popa, "Block-projections algorithms with blocks containing mutually orthogonal rows and columns," *BIT*, vol. 39, no. 2, pp. 323–338, 1999. [Online]. Available: <http://dx.doi.org/10.1023/A:1022398014630>
- [24] E. Weber, "Algebraic aspects of the paving and Feichtinger conjectures," in *Topics in Operator Theory*. Birkhäuser Basel, 2010, pp. 569–578.
- [25] N. Loizou and P. Richtárik, "A new perspective on randomized gossip algorithms," in *IEEE Glob. Conf. Sig.* IEEE, 2016, pp. 440–444.
- [26] D. Needell, "Randomized Kaczmarz solver for noisy linear systems," *BIT*, vol. 50, no. 2, pp. 395–403, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s10543-010-0265-5>
- [27] M. H. DeGroot, "Reaching a consensus," *J. Am. Stat. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974. [Online]. Available: <http://www.jstor.org/stable/2285509>
- [28] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE T. Automat. Contr.*, vol. 31, no. 9, pp. 803–812, 1986.
- [29] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bull. Internat. Acad. Polon. Sci. Lettres A*, p. 335–357, 1937.
- [30] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *J. Fourier Anal. Appl.*, vol. 15, pp. 262–278, 2007.
- [31] A. Zouzias and N. M. Freris, "Randomized extended Kaczmarz for solving least squares," *SIAM J. Matrix Anal. A.*, vol. 34, no. 2, p. 773–793, 2013.
- [32] P. Eggermont, G. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems, with applications to image reconstruction," *Linear Algebra Appl.*, vol. 40, pp. 37–67, 1981.
- [33] T. Elfving, "Block-iterative methods for consistent and inconsistent linear equations," *Numer. Math.*, vol. 35, no. 1, pp. 1–12, 1980. [Online]. Available: <http://dx.doi.org/10.1007/BF01396365>
- [34] J. Haddock and A. Ma, "Greed works: An improved analysis of sampling Kaczmarz-Motzkina," *SIAM J. Math. Data Sci.*, 2020, to appear.
- [35] C. Godsil and G. Royle, *Algebraic Graph Theory*, 01 2001, vol. 207.

- [36] K. Massey, “Statistical models applied to the rating of sports teams,” *Bluefield College*, 1997.
- [37] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, “Statistical ranking and combinatorial Hodge theory,” *Math. Program.*, vol. 127, no. 1, pp. 203–244, 2011.
- [38] A. Hagberg, P. Swart, and D. S. Chult, “Exploring network structure, dynamics, and function using NetworkX,” Los Alamos National Lab. (LANL), Los Alamos, NM (United States), Tech. Rep., 2008.
- [39] H. Z. Brooks and M. A. Porter, “A model for the influence of media on the ideology of content in online social networks,” *Phys. Rev. Res.*, vol. 2, no. 2, p. 023041, 2020.



Jamie Haddock received her B.S. in Mathematics (2013) from Gonzaga University (Spokane, WA, USA) and Ph.D. in Applied Mathematics (2018) from University of California, Davis (Davis, CA, USA). She is now Assistant Professor of Mathematics at Harvey Mudd College (Claremont, CA, USA). Her research interests include mathematical data science, optimization, and applied convex geometry, with work in randomized method in numerical linear algebra, combinatorial and convex optimization, topic modeling and tensor decomposition, distributed computing, compressive sensing, community detection, combinatorial linear algebra, and medical applications of data science and statistics.



Benjamin Jarman received his B.A. (2017) and M.Math. (2018) degrees from Trinity College, University of Cambridge (Cambridge, England). He is now a doctoral student at University of California, Los Angeles (Los Angeles, CA, USA), under the supervision of Professor Deanna Needell. His research interests include the mathematics of data theory and machine learning, with work in randomized linear system solvers, distributed computing, topological data analysis, and self-supervised learning.

Chen Yap received her B.S. in Mathematics (2021) from University of California, Los Angeles (Los Angeles, CA, USA).