Queue-based Broadcast Gossip Algorithm for Consensus

Soummya Kar, Rohit Negi, Majid Mahzoon and Anit Kumar Sahu
Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh PA 15213
Email: {soummyak, negi, mmahzoon, anits}@ece.cmu.edu

***Abstract*— We study broadcast gossip algorithms to compute the average of given initial sensor measurements. In the context of wireless networks, an algorithm was proposed by Scaglione et al. that allows a single node to broadcast at each time with geometrically fast convergence to consensus. To improve the rate of convergence to consensus, we go beyond this single node broadcast approach and propose a queue-based broadcast gossip algorithm, in which simultaneous node broadcasts are allowed. Since packet collisions may happen, we choose time-varying update weights. Using a novel interval-based consensus error analysis to handle the time-dependent update weights, we show that with appropriate choice of parameters, the proposed algorithm converges to consensus in the mean-squared sense geometrically fast. We prove that for the class of networks modeled as non-bipartite Ramanujan graphs, the exponent of convergence of the proposed algorithm is independent of the number of nodes, unlike the single node broadcast case which converges slowly in large networks. We also demonstrate through simulations that our proposed algorithm improves the rate of convergence for some other example networks.**

1. Introduction

Distributed computation in multi-agent networks has been instrumental in the evolution of a large body of literature pertaining to distributed inference algorithms, e.g., distributed estimation, detection, optimization [1]–[6]. Distributed average consensus is one of the most simple yet elegant distributed computation schemes which involves computing the global average of distributed data using iterative local communications. Various extensions and variants of the basic consensus scheme can be found in [7]–[11].

Gossip-based algorithms have been proposed to achieve consensus, including randomized gossip algorithm [12], [13] and geographic gossip algorithm [14]. By a gossip algorithm, we mean specifically an algorithm in which each node receives a packet from no more than one neighbor at any time instant. For wireless sensor networks, which is the scenario we have in mind in this paper, a natural gossip method would be for a node to *broadcast* its value, so that all its neighbors (nodes within some distance of the broadcasting node) can receive this value correctly. This is called the broadcast gossip method in [15], and will be the primary comparison point of our paper. In that paper, a *single node* chosen uniformly at random broadcasts its state value successfully to all its neighbors, where the single broadcast ensures that no collision of packets occurs. The neighbors then update their own state value using *fixed weights* and the remaining nodes keep their value unchanged.

In this paper, we are inspired by the intuition that allowing *simultaneous broadcasts* should improve the rate of convergence to consensus. However, this approach also increases the probability of packet collision. In fact, in a large network, the probability will be high that a collision occurs somewhere in the network, which will cause the packet reception rates at different nodes to be unequal. This will result in significant deviation of the converged consensus value (assuming that consensus is indeed reached) from the initial (unweighted) consensus value. Thus, there are two types of error we focus on. The first is the *consensus error*, which measures the deviation of the nodes' individual state values from the average (sample mean) of these values, to check whether consensus is occurring. The second is the *error-in-average*, which measures the deviation of the unweighted average of the nodes' state values from the *initial unweighted average* of these values, since the latter is the desired consensus value to which the states must be driven. These errors will be formally defined in Section 2. Many traditional developments in the consensus literature assume that some underlying wireless scheduling mechanism is in place to equalize rates; in contrast, in this paper, we use a realistic transmission scheduling and deal with the impact of the resulting packet collisions. The authors in [15] equalize the rates by choosing equal probability of broadcast for all nodes, and avoiding collisions by only allowing one node to transmit at any time. However, the price paid is that N nodes will require time proportional to N to obtain their turn at broadcasting, resulting in slow convergence to consensus in large networks. Since we allow simultaneous broadcasts in this paper, packet collisions are inevitable. (An alternative strategy may be to explore a type of 'broadcast collision avoidance' protocol, which we do not pursue in this paper.) Therefore, we modify the standard gossip algorithm for state update to use *time-varying* update weights. Intuitively, the weights must be chosen so that the expected weight that node i uses for transmissions from j is the same as the one that j uses for transmissions from i. However, in this paper the weights are chosen *adaptively* using *local information*, since nodes are not assumed to know the network structure. (Knowing the network structure would need a burdensome preliminary step where global information would need to be gathered and appropriate parameters calculated, and burdensome maintenance where the global information would need to be updated and disseminated whenever the network changes.) However, the adaptive choice of weights results in time dependencies, which makes the algorithm significantly more complex to analyze. A key result in the paper is, therefore, to prove formally, using a novel interval-based analysis, that this adaptive algorithm indeed results in consensus (with geometric decay of the mean-squared consensus error). Intuitively, since the algorithm allows simultaneous broadcasts, it should converge faster to consensus, compared to [15]. Unfortunately, in this preliminary work, the analytical bounds obtained by us, although do indeed prove geometrically fast convergence to

The work of R. Negi and M. Mahzoon was supported in part by NSF under grants CNS-1218823 and CCF-1422193. The work of S. Kar and A. K. Sahu was supported in part by NSF under grants CCF-1513936 and ECCS-1408222.

consensus, are not at present sufficiently tight to prove faster convergence in all networks, except in certain (important) networks, such as non-bipartite Ramanujan graphs. Simulation results, however, clearly show that our claim that the new algorithm converges faster in general is indeed plausible.

The rest of the paper is organized as follows. In Section 2, we provide the model and notation used. The queue-based broadcast gossip algorithm is introduced in Section 3. Section 4 shows a Lyapunov analysis of the proposed algorithm while Section 5 provides the main result proving the convergence to consensus of the proposed algorithm. Simulation experiments of the proposed algorithm are discussed in Section 6 and Section 7 concludes the paper.

2. Network Model

Consider a network of N nodes (or sensors or agents). The inter-sensor communication network is modeled as a simple[1] *undirected* connected graph $G = (V, E)$, with $V = \{1, \cdots, N\}$ and E denoting the set of nodes and communication links, respectively. The edge set E includes both directions of a link so that (i, j) and (j, i) are considered separate edges in E. The neighborhood of node i is

$$\mathcal{N}_i = \{j \in V \,|\, (i, j) \in E\}\,. \tag{1}$$

We assume that if node i transmits, all the neighbors $\mathcal{N}_i$ can hear the transmission. However, if another node k simultaneously transmits, some of these neighbors may experience a packet collision, so that node i's packet is lost. For the purposes of this paper, the precise collision model is not important since the algorithm we propose will adapt to the physical communication model.

Node i has degree $d_i = |\mathcal{N}_i|$. The structure of the graph is described by the $N \times N$ adjacency matrix, $\mathbf{A} = \mathbf{A}^\top = [\mathbf{A}_{nl}]$, where $\mathbf{A}_{nl} = 1$, if $(n, l) \in E$, and $\mathbf{A}_{nl} = 0$, otherwise. Let $\mathbf{D} = \text{diag}\,(d_1, \cdots, d_N)$. The graph Laplacian $\overline{\mathbf{L}} = \mathbf{D} - \mathbf{A}$ is positive semi-definite, with eigenvalues ordered as $0 = \lambda_1(\overline{\mathbf{L}}) \le \lambda_2(\overline{\mathbf{L}}) \le \cdots \le \lambda_N(\overline{\mathbf{L}})$. The eigenvector of $\overline{\mathbf{L}}$ corresponding to $\lambda_1(\overline{\mathbf{L}})$ is $(1/\sqrt{N})\mathbf{1}$, where $\mathbf{1}$ is the vector of all 1's. Since we assume a connected graph, $\lambda_2(\overline{\mathbf{L}}) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network.

We consider a time-slotted system. At time slot $t \ge 0$, each sensor $i \in V$ has its local estimate $x_i(t)$ of the global average which is called its state value. We denote $\mathbf{x}^\top(t) = [x_1(t), \cdots, x_N(t)]^\top$ as the vector of estimates at all the nodes at the end of time slot t. The ultimate goal of an average consensus scheme is to drive the estimate $\mathbf{x}(t)$ as close as possible to the initial average vector $\overline{x}(0)\mathbf{1}$, which we call the *unweighted consensus value*, where

$$\overline{x}(0) = \frac{1}{N} \sum_{i=1}^{N} x_i(0). \tag{2}$$

Similarly, $\overline{x}(t)$ is the average of the nodes' state values at time slot t.

As mentioned in Section 1, there are two types of error we focus on: 1. *Consensus error*, defined as $\mathbf{x}(t) - \overline{x}(t)\mathbf{1}$, which measures the deviation of the nodes' individual state values from the average of these values. 2. *Error-in-average*, defined as $\overline{x}(t) - \overline{x}(0)$, which measures the deviation of the unweighted average of the nodes' state values from the *unweighted consensus value* .

[1] A graph is said to be simple if it is devoid of self loops and multiple edges.

3. Queue-based Broadcast Gossip Algorithm

As in standard gossip algorithms, each node must transmit its current state to other nodes. In the proposed algorithm, node i can only broadcast $x_i(t)$ in time slot t to its neighborhood if it is included in the schedule of the network at that time. Formally, every node i at any time slot t is assigned a schedule $\sigma_i(t)$. The schedule $\sigma_i(t)$ takes the value 1 with probability $p_i \in (0, 1)$, with $\sigma_i(t)$ being independent for each i, t. A schedule of the network can then be represented as a vector $\boldsymbol{\sigma}(t)$, where $\boldsymbol{\sigma}(t) \in \mathbb{R}^N$ and $\sigma_i(t) \in \{0, 1\}$ for all $i \in V$. Node j experiences a collision if it receives more than one packet in a given time slot. Let p_{ji} be the probability that node j successfully receives a packet from node i in a given time slot. In general, $p_{ji} \neq p_{ij}$, so that the reception rate of packets from i to j may not be the same in the other direction. We assume that a lower bound on this probability p_0 is known, so that $p_{ji} > p_0,\ \forall (j, i) \in E$. *Note that the bound p_0 is the only global knowledge needed at each node in our algorithm. In particular, the graph G or the values p_i of other nodes are not required to be known.* The expected number of correct packet receptions in each time slot is $\sum_{(j,i)\in E} p_{ji} > p_0|E| = p_0\bar{d}N$, where $\bar{d}$ is the average degree of G. Thus, we can expect the number of successful packet receptions to be proportional to the number of nodes in our algorithm, unlike the single-node broadcast algorithm in [15], which should allow faster convergence to consensus. We describe one plausible scenario where this bound p_0 can be calculated. Assume that a collision occurs at j if and only if two or more nodes in $j \cup \mathcal{N}_j$ transmit in that time slot. Thus, there is a successful reception of node i's packet at node $j \in \mathcal{N}_i$, denoted by $\sigma_{ji} = 1$, if node i is included in the schedule but neither node j nor any of j's neighbors is included, i.e.,

$$\sigma_{ji} = \sigma_i \overline{\sigma}_j \prod_{k \in \mathcal{N}_j \setminus i} \overline{\sigma}_k, \tag{3}$$

where $\overline{\sigma} = 1 - \sigma$. Then, at each time slot, node $j \in \mathcal{N}_i$ receives node i's state value without collision, i.e. $\sigma_{ji} = 1$, with probability

$$p_{ji} = \mathbb{E}[\sigma_{ji}] = p_i(1 - p_j) \prod_{k \in \mathcal{N}_j \setminus i} (1 - p_k) \ge p_0, \tag{4}$$

where p_0 is a lower bound on the probability of successful reception for all links that must be determined. Now if we choose $p_i = p$ for all nodes $i \in V$, then a lower bound is $p_0 \doteq p(1 - p)^\Delta$, where $\Delta = \max_{i=1,\cdots,N} d_i$ is the maximum vertex degree of G. Note that $\Delta \ge 1$ since G is assumed connected. However, we stress again that the scenario outlined above is not necessary, so long as a bound p_0 can be obtained by some means.

For each node $j \in V$, we consider $|\mathcal{N}_j|$ queues, $q_{jk},\ k \in \mathcal{N}_j$, each corresponding to one of its neighbors. The length (occupancy) of q_{jk} at time t is $Q_{jk}(t)$. At the beginning of each time slot, $\alpha \in (0, 1)$ (fractional) tokens enter each queue. (The tokens and the queues are simply a means for book-keeping and do not involve physical buffers.) The maximum queue length is chosen to be $\frac{\alpha\gamma}{p_0} < 1$ for suitably chosen $\alpha > 0$ and $\gamma > p_0$. When a queue is at the maximum allowed length, all tokens entering that queue are rejected. Furthermore, once node j receives a packet from node k successfully, *it empties the corresponding queue q_{jk} completely*, i.e., $Q_{jk}(t)$ tokens depart the queue (rather than a constant number of tokens departing, as in typical queuing literature).

Once the broadcast packet of node i is received, the neighboring nodes which received it successfully, denoted by $\mathcal{N}_i^0$, set their state values equal to the weighted average of their current state value and the state value broadcast by node i, i.e.,

$$x_j(t+1) = (1 - Q_{ji}(t))\, x_j(t) + Q_{ji}(t) x_i(t), \quad \forall j \in \mathcal{N}_i^0. \quad (5)$$

Note that this consensus step differs from standard consensus algorithms, since the queues are used as weights here, instead of using a fixed weight as in [15].

Due to the queuing model, where the queues are emptied completely when a packet is successfully received, the queue length at the end of time slot t can be written as $Q_{ji}(t) = \min\left\{\alpha T_{ji}(t), \frac{\alpha\gamma}{p_0}\right\}$, where $T_{ji}(t) = t - \max\{k : k < t, \sigma_{ji}(k) = 1\}$ is the number of time slots that node j waited to receive the state value of node i without collision. Therefore, with $T^{\max} \doteq \frac{\gamma}{p_0}$, (5) can be written as

$$\begin{aligned} x_j(t+1) &= \left(1 - \alpha \min\left\{T_{ji}(t), T^{\max}\right\}\right) x_j(t) \\ &+ \alpha \min\left\{T_{ji}(t), T^{\max}\right\} x_i(t), \quad \forall j \in \mathcal{N}_i^0. \end{aligned} \quad (6)$$

There will typically be more than one node broadcasting in the same time slot. The above scheme is replicated for the other successfully received packets as well. The nodes that do not successfully receive any packet from their neighbors do not update their state values. This procedure is repeated in every time slot.

The update in (6) can be written in a compact manner as follows:

$$\mathbf{x}(t+1) = \left(\mathbf{I} - \alpha \mathbf{L}(t)\right)\mathbf{x}(t), \quad (7)$$

where $\mathbf{I}$ denotes the identity matrix and the random non-symmetric 'Laplacian-like' matrix $\mathbf{L}(t)$ has the j,i entry

$$\mathbf{L}_{ji}(t) = \begin{cases} \sum_{k \in \mathcal{N}_j} \min\left\{T_{jk}(t), T^{\max}\right\} \sigma_{jk}(t) & \text{if } j = i \\ -\min\left\{T_{ji}(t), T^{\max}\right\} \sigma_{ji}(t) & \text{if } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The main result in this paper is to show that the broadcast gossip algorithm (7) results in consensus asymptotically in time. However, there are two difficulties that we face in analyzing (7) - firstly, the state updates depend on time-varying weights $Q_{ji}(t)$, and secondly, these weights are random variables that are dependent in time (since they are queue lengths). In contrast, in [15] the weights are i.i.d., which allows a contraction-type analysis for every slot. Our solution to this problem is to do a Lyapunov analysis over a suitable large interval δ, which allows us to carefully consider the impact of Markov $Q_{ji}(t)$, as shown in Section 4-A.

4. Consensus Error Analysis

In this section we formally state and prove results concerning the convergence of the average of the Laplacian-like matrix sequence $\{\mathbf{L}(t)\}$ (see (8)) and the Lyapunov analysis of the sequence $\{\mathbf{x}(t)\}$, which will be crucial in establishing consensus using the proposed algorithm in Section 5.

A. Convergence of Laplacian Average

For each $t \geq 0$, we denote by $\mathcal{F}_t$ the natural filtration generated by the Laplacian-like matrices $\{\mathbf{L}(s)\}_{s=0}^t$ and the queue lengths $\{Q_{ji}(s)\}_{s=0}^t$, $\forall j \in \{1, \cdots, N\}, \forall i \in \mathcal{N}_j$, i.e.,

$$\mathcal{F}_t = \left\{\{\mathbf{L}(s)\}_{s=0}^t, \{Q_{ji}(s)\}_{s=0}^t, \forall j \in \{1, \cdots, N\}, \forall i \in \mathcal{N}_j\right\}, \quad (9)$$

which is the σ-algebra induced by the Laplacian-like matrices and the queue lengths. The following proposition proves the convergence of the average of non-symmetric (i.e., directed) Laplacian-like matrices to a *symmetric* connected Laplacian, which will be instrumental in establishing the geometric convergence of the process $\{\mathbf{x}(t)\}$.

Proposition 4.1. *Assuming G is connected, the sequence $\{\mathbf{L}(t)\}$ satisfies the following convergence condition for all $t \geq 0$:*

$$\lim_{\gamma, \frac{\delta}{\gamma} \to \infty} \mathbb{E}\left[(\delta)^{-1}.\left(\mathbf{L}(t+\delta) + \cdots + \mathbf{L}(t+1)\right) \mid \mathcal{F}_t\right] = \overline{\mathbf{L}}, \quad (10)$$

where $\overline{\mathbf{L}}$ is the symmetric (undirected) connected Laplacian of graph G, i.e., with $\lambda_2\left(\overline{\mathbf{L}}\right) > 0$. More precisely,

$$\overline{\mathbf{L}}_{ji} = \begin{cases} |\mathcal{N}_j| & \text{if } j = i \\ -1 & \text{if } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $|\mathcal{N}_j|$ is the degree of node j in graph G.

For convenience, we introduce the following notation:

$$\overline{\mathbf{L}}_{t,\delta} \doteq (\delta)^{-1}.\left(\mathbf{L}(t+\delta) + \cdots + \mathbf{L}(t+1)\right), \quad (12)$$

$$\widetilde{\mathbf{L}}_{t,\delta} = \mathbb{E}\left[\overline{\mathbf{L}}_{t,\delta} \mid \mathcal{F}_t\right] - \overline{\mathbf{L}}. \quad (13)$$

Also, for a vector $\mathbf{z} \in \mathbb{R}^N$, we denote by $\mathbf{z}_{\mathcal{C}}$ its projection onto the consensus subspace, i.e.,

$$\mathbf{z}_{\mathcal{C}} = \frac{1}{N}\mathbf{1}\mathbf{1}^T\mathbf{z}, \quad (14)$$

and, by $\mathbf{z}_{\mathcal{C}^\perp}$ the projection onto the orthogonal subspace $\mathcal{C}^\perp$.

Proof of Proposition 4.1. Considering $\delta \geq T^{\max} + 1 = \frac{\gamma}{p_0} + 1$ and defining $T_{ji}^{\text{ini}} = t - \max\{k : k < t, \sigma_{ji}(k) = 1\}$ (which is known conditioned on $\mathcal{F}_t$), for $j \in \mathcal{N}_i$, we have

$$\begin{aligned} &\sum_{k=t+1}^{t+\delta} \mathbb{E}\left[\mathbf{L}_{ji}(k) | \mathcal{F}_t\right] \\ &= -\sum_{k=t+1}^{t+\delta} \mathbb{E}\left[\min\left\{T_{ji}(k), T^{\max}\right\} \sigma_{ji}(k) | \mathcal{F}_t\right] \quad (15) \\ &= -\sum_{k=t+1}^{t+\delta} \mathbb{E}\left[\min\left\{T_{ji}(k), T^{\max}\right\} | \mathcal{F}_t\right] \mathbb{E}\left[\sigma_{ji}(k) | \mathcal{F}_t\right] \quad (16) \\ &= -\sum_{k=t+1}^{t+\delta} \sum_{l=1}^{\min\left\{k-t-1, T^{\max}\right\}} l p_{ji}^2 (1 - p_{ji})^{l-1} \\ &- \sum_{k=t+1}^{t+\delta} \min\left\{k - t + T_{ji}^{\text{ini}}, T^{\max}\right\} \\ &\times p_{ji}(1 - p_{ji})^{\min\left\{k-t-1, T^{\max}\right\}} \quad (17) \\ &= -\sum_{l=1}^{T^{\max}} \sum_{k=\min\left(t+l+1, t+T^{\max}+1\right)}^{t+\delta} l p_{ji}^2 (1 - p_{ji})^{l-1} \\ &- \sum_{k=t+1}^{t+\delta} \min\left\{k - t + T_{ji}^{\text{ini}}, T^{\max}\right\} \\ &\times p_{ji}(1 - p_{ji})^{\min\left\{k-t-1, T^{\max}\right\}} \quad (18) \\ &= -\sum_{l=1}^{T^{\max}} \left(\delta - \min\left(T^{\max}, l\right)\right) l p_{ji}^2 (1 - p_{ji})^{l-1} \\ &- \sum_{l=1}^{\delta} \min\left\{l + T_{ji}^{\text{ini}}, T^{\max}\right\} p_{ji}(1 - p_{ji})^{\min\left(l-1, T^{\max}\right)} \quad (19) \end{aligned}$$

$$
\begin{aligned}
&= -\delta + \delta \sum_{l=T^{\max}+1}^{\infty} l p_{ji}^2 (1-p_{ji})^{l-1} \\
&+ \sum_{l=1}^{T^{\max}} l^2 p_{ji}^2 (1-p_{ji})^{l-1} \\
&- \sum_{l=1}^{T^{\max}} \min\left\{l + T_{ji}^{\text{ini}}, T^{\max}\right\} p_{ji}(1-p_{ji})^{l-1} \\
&- \sum_{l=T^{\max}+1}^{\delta} T^{\max} p_{ji}(1-p_{ji})^{T^{\max}} \qquad (20)\\
&= -\delta + \delta \left(T^{\max} p_{ji}(1-p_{ji})^{T^{\max}} + (1-p_{ji})^{T^{\max}}\right) \\
&+ \sum_{l=1}^{T^{\max}} l^2 p_{ji}^2 (1-p_{ji})^{l-1} \\
&- \sum_{l=1}^{T^{\max}} \min\left\{l + T_{ji}^{\text{ini}}, T^{\max}\right\} p_{ji}(1-p_{ji})^{l-1} \\
&- \sum_{l=T^{\max}+1}^{\delta} T^{\max} p_{ji}(1-p_{ji})^{T^{\max}}, \qquad (21)
\end{aligned}
$$

where (17) follows from the facts that the probability distribution of $T_{ji}(k)$ conditioned on $\mathcal{F}_t$ for all $k > t$ is

$$
\Pr(T_{ji}(k) = l|\mathcal{F}_t) = \begin{cases} p_{ji}(1-p_{ji})^{l-1} & \text{if } l = 1, 2, \cdots, k-t-1 \\ (1-p_{ji})^{k-t-1} & \text{if } l = k-t+T_{ji}^{\text{ini}} \\ 0 & \text{otherwise} \end{cases} \tag{22}
$$

and $\mathbb{E}\left[\sigma_{ji}(k)|\mathcal{F}_t\right] = p_{ji}$ for all k, and (18) follows from the assumption that $\delta \geq T^{\max} + 1$. Therefore, for $j \in \mathcal{N}_i$,

$$
\begin{aligned}
&\left|(\delta)^{-1}. \sum_{k=t+1}^{t+\delta} \mathbb{E}\left[\mathbf{L}_{ji}(k)|\mathcal{F}_t\right] - \overline{\mathbf{L}}_{ji}\right| \leq T^{\max} p_{ji}(1-p_{ji})^{T^{\max}} \\
&+ (1-p_{ji})^{T^{\max}} + \frac{2-p_{ji}}{\delta p_{ji}} \\
&+ \frac{1}{\delta}\sum_{l=1}^{T^{\max}} \min\left\{l + T_{ji}^{\text{ini}}, T^{\max}\right\} p_{ji}(1-p_{ji})^{l-1} \\
&+ \frac{1}{\delta}\sum_{l=T^{\max}+1}^{\delta} T^{\max} p_{ji}(1-p_{ji})^{T^{\max}} \qquad (23)\\
&\leq T^{\max}(1-p_{ji})^{T^{\max}} + (1-p_{ji})^{T^{\max}} + \frac{2}{\delta p_{ji}} \\
&+ T^{\max}\frac{1-(1-p_{ji})^{T^{\max}}}{\delta} + T^{\max}(1-p_{ji})^{T^{\max}} \qquad (24)\\
&\leq 3T^{\max}(1-p_0)^{T^{\max}} + \frac{2}{\delta p_0} + \frac{T^{\max}}{\delta} \qquad (25)\\
&= \frac{3\gamma}{p_0}(1-p_0)^{\frac{\gamma}{p_0}} + \frac{2}{\delta p_0} + \frac{\gamma}{\delta p_0} \qquad (26)\\
&\leq \frac{3\gamma}{p_0}e^{-\gamma} + \frac{\gamma+2}{\delta p_0}, \qquad (27)
\end{aligned}
$$

where (25) and (27) follow from the facts that $p_{ji} \geq p_0$ and $1 - p_0 \leq e^{-p_0}$ respectively. Since in each time slot k, by the property of Laplacian-like matrices (8), $\mathbf{L}(k)\mathbf{1} = \mathbf{0}$, we have

$$
\left|(\delta)^{-1}. \sum_{k=t+1}^{t+\delta} \mathbb{E}\left[L_{ii}(k)|\mathcal{F}_t\right] - \overline{\mathbf{L}}_{ii}\right| \leq |\mathcal{N}_j|. \left(\frac{3\gamma}{p_0}e^{-\gamma} + \frac{\gamma+2}{\delta p_0}\right). \tag{28}
$$

Thus, we have

$$
\left\|\widetilde{\mathbf{L}}_{t,\delta}\right\|_1 \leq 2\Delta\left(\frac{3\gamma}{p_0}e^{-\gamma} + \frac{\gamma+2}{\delta p_0}\right), \tag{29}
$$

$$
\left\|\widetilde{\mathbf{L}}_{t,\delta}\right\|_\infty \leq 2\Delta\left(\frac{3\gamma}{p_0}e^{-\gamma} + \frac{\gamma+2}{\delta p_0}\right), \tag{30}
$$

where $\widetilde{\mathbf{L}}_{t,\delta}$ is as defined in (13). From (29), (30) and Holder's inequality for matrices, $\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1 \ \|\mathbf{A}\|_\infty}$, we conclude that

$$
\left\|\widetilde{\mathbf{L}}_{t,\delta}\right\|_2 \leq 2\Delta\left(\frac{3\gamma}{p_0}e^{-\gamma} + \frac{\gamma+2}{\delta p_0}\right). \tag{31}
$$

Therefore, $\left\|\widetilde{\mathbf{L}}_{t,\delta}\right\|_2 \to 0$ as $\gamma, \frac{\delta}{\gamma} \to \infty$. ■

Propositon 4.1 shows that with suitable choice of parameters, the sample average of non-symmetric (directed) Laplacian-like matrices converges to a *symmetric (undirected)* Laplacian in expectation. The proposition below shows an appropriate choice of parameters.

Proposition 4.2. *For each $\varepsilon > 0$, there exist a $\gamma \doteq \gamma(\varepsilon) > p_0$ and a positive integer $\delta \doteq \delta(\varepsilon) \geq \frac{\gamma}{p_0}$ such that*

$$
\sup_{t\geq 0}\left\|\mathbb{E}\left[(\delta)^{-1}.\left(\mathbf{L}(t+\delta) + \cdots + \mathbf{L}(t+1)\right) \mid \mathcal{F}_t\right] - \overline{\mathbf{L}}\right\|_2 \leq \varepsilon \text{ a.s.} \tag{32}
$$

Proof. The proof follows from Proposition 4.1. One such choice is $\gamma = \frac{12\Delta}{\varepsilon p_0} =$ and $\delta = \frac{4\Delta(\gamma+2)}{\varepsilon} = \frac{48\Delta^2}{\varepsilon^2 p_0} + \frac{8\Delta}{\varepsilon}$. ■

B. Lyapunov Analysis of Consensus Error

In this section we will fix a $(\varepsilon, \gamma, \delta)$ triple and analyze the evolution of the periodically sampled subsequence $\{\mathbf{x}(k\delta)\}_{k\geq 0}$. This will allow us to analyze the consensus error over the (large) interval δ, which will enable us to carefully consider the Markov property of the queue weights. Specifically, as a measure of squared distance from the consensus subspace, we will study the evolution of the non-negative functional

$$
V(k\delta) = \mathbf{x}^\top(k\delta)\overline{\mathbf{L}}\mathbf{x}(k\delta) \tag{33}
$$

along the $\{\mathbf{x}(k\delta)\}$ trajectory. Note that, since $\overline{\mathbf{L}}$ is a symmetric connected Laplacian, it follows that

$$
V(k\delta) = \mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}\mathbf{x}_{\mathcal{C}^\perp}(k\delta) \geq \lambda_2(\overline{\mathbf{L}})\left\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right\|_2^2, \tag{34}
$$

and hence, $V(k\delta) \to 0$ implies consensus of the subsequence, i.e., $\left\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right\|_2 \to 0$.
Later, in Section 5, we will obtain conditions on the weight parameter α that ensure geometric decay of the sequence $\{V(k\delta)\}$ to 0 and subsequently extend the (geometric) consensus property of the subsequence $\{\mathbf{x}(k\delta)\}$ to the entire sequence $\{\mathbf{x}(t)\}$.
Note that,

$$
\begin{aligned}
&\mathbf{x}\left((k+1)\delta\right) = \left(\mathbf{I} - \alpha\mathbf{L}\left((k+1)\,\delta\right)\right)\left(\mathbf{I} - \alpha\mathbf{L}\left(k\delta+\delta-1\right)\right) \\
&\cdots\left(\mathbf{I} - \alpha\mathbf{L}\left(k\delta+1\right)\right)\mathbf{x}(k\delta). \qquad (35)
\end{aligned}
$$

Hence, the product $\left(\mathbf{I} - \alpha\mathbf{L}\left((k+1)\,\delta\right)\right)\cdots\left(\mathbf{I} - \alpha\mathbf{L}(k\delta+1)\right)$ may be written as

$$
\begin{aligned}
&\left(\mathbf{I} - \alpha\mathbf{L}\left((k+1)\delta\right)\right)\left(\mathbf{I} - \alpha\mathbf{L}\left(k\delta+\delta-1\right)\right) \\
&\cdots\left(\mathbf{I} - \alpha\mathbf{L}\left(k\delta+1\right)\right) \qquad (36)\\
&= \mathbf{I} - \alpha\left(\mathbf{L}\left((k+1)\delta\right) + \cdots + \mathbf{L}\left(k\delta+1\right)\right) + R_{(k+1)\delta} \qquad (37)\\
&= \mathbf{I} - \alpha\delta\overline{\mathbf{L}}_{k\delta,\delta} + R_{(k+1)\delta}. \qquad (38)
\end{aligned}
$$

To bound the error matrix $R_{(k+1)\delta}$ in (38), we first note the following:

Proposition 4.3. *The norm of the Laplacian-like matrix $\mathbf{L}(t)$ in each time slot t is bounded as*

$$
\|\mathbf{L}(t)\|_2 \leq \sqrt{2\Delta}\frac{\gamma}{p_0}. \tag{39}
$$

Proof. It can be easily seen that in each time slot t

$$\|\mathbf{L}(t)\|_1 \le 2\frac{\gamma}{p_0}, \|\mathbf{L}(t)\|_\infty \le \Delta\frac{\gamma}{p_0}. \tag{40}$$

From (40) and Holder's inequality for matrices, we conclude that

$$\|\mathbf{L}(t)\|_2 \le \sqrt{2\Delta}\frac{\gamma}{p_0}. \tag{41}$$

∎

Now, we choose $\alpha < \frac{p_0}{4\Delta\delta\gamma}$ (which is tighter than our earlier stated assumption that $\frac{\alpha\gamma}{p_0} < 1$) so that $\alpha\delta\|\mathbf{L}\|_2 < \frac{1}{2}$. Then, from definition of $R_{(k+1)\delta}$ and Proposition 4.3, we have

$$\|R_{(k+1)\delta}\|_2 \le \frac{\alpha^2\delta^2\|\mathbf{L}\|_2^2}{1-\alpha\delta\|\mathbf{L}\|_2} \le 2\alpha^2\delta^2\|\mathbf{L}\|_2^2 \le \frac{4\Delta\alpha^2\delta^2\gamma^2}{p_0^2}, \tag{42}$$

where $\mathbf{L}$ is the Laplacian-like matrix with the largest norm. We now characterize the evolution of the $\{V(k\delta)\}$ sequence. Note that, by (33) and (38),

$$\begin{aligned} V((k+1)\delta) &= \mathbf{x}^T(k\delta)\left(\mathbf{I} - \alpha\delta\overline{\mathbf{L}}_{k\delta,\delta} + R_{(k+1)\delta}\right)^\top \overline{\mathbf{L}} \\ &\times \left(I - \alpha\delta\overline{\mathbf{L}}_{k\delta,\delta} + R_{(k+1)\delta}\right)\mathbf{x}(k\delta) \qquad (43)\\ &= V(k\delta) - \alpha\delta\mathbf{x}^\top(k\delta)\left(\overline{\mathbf{L}}\overline{\mathbf{L}}_{k\delta,\delta} + \overline{\mathbf{L}}_{k\delta,\delta}\overline{\mathbf{L}}\right)\mathbf{x}(k\delta) \\ &+ \mathbf{x}^\top(k\delta)J_{k+1}\mathbf{x}(k\delta), \qquad (44) \end{aligned}$$

where

$$\begin{aligned} J_{k+1} &= \overline{\mathbf{L}}R_{(k+1)\delta} + \alpha^2\delta^2\overline{\mathbf{L}}_{k\delta,\delta}^\top\overline{\mathbf{L}}\overline{\mathbf{L}}_{k\delta,\delta} - \alpha\delta\overline{\mathbf{L}}_{k\delta,\delta}^\top\overline{\mathbf{L}}R_{(k+1)\delta} \\ &+ R_{(k+1)\delta}^\top\overline{\mathbf{L}} - \alpha\delta R_{(k+1)\delta}^\top\overline{\mathbf{L}}\overline{\mathbf{L}}_{k\delta,\delta} + R_{(k+1)\delta}^\top\overline{\mathbf{L}}R_{(k+1)\delta}. \end{aligned} \tag{45}$$

From the properties of undirected Laplacians and directed Laplacian-like matrices respectively ($\overline{\mathbf{L}}$ is an undirected Laplacian, whereas $\overline{\mathbf{L}}_{k\delta,\delta}$ is a directed Laplacian-like matrix), we have,

$$\mathbf{x}^\top(k\delta)\overline{\mathbf{L}} = \mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}, \overline{\mathbf{L}}\mathbf{x}(k\delta) = \overline{\mathbf{L}}\mathbf{x}_{\mathcal{C}^\perp}(k\delta), \tag{46}$$

$$\overline{\mathbf{L}}_{k\delta,\delta}\mathbf{x}(k\delta) = \overline{\mathbf{L}}_{k\delta,\delta}\mathbf{x}_{\mathcal{C}^\perp}(k\delta), \mathbf{x}^\top(k\delta)\overline{\mathbf{L}}_{k\delta,\delta}^\top = \mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}_{k\delta,\delta}^\top. \tag{47}$$

Also, since $R_{(k+1)\delta}$ corresponds to a sum of products of matrices, the right-most entry of each product being a directed Laplacian-like matrix, we have

$$R_{(k+1)\delta}\mathbf{x}(k\delta) = R_{(k+1)\delta}\mathbf{x}_{\mathcal{C}^\perp}(k\delta), \tag{48}$$

$$\mathbf{x}^\top(k\delta)R_{(k+1)\delta}^\top = \mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)R_{(k+1)\delta}^\top. \tag{49}$$

From (42) and noting that $\left\|\overline{\mathbf{L}}\right\|_2 \le 2\Delta$, $\alpha\delta\|\mathbf{L}\|_2 < \frac{1}{2}$ and $\alpha < 1$, it can be concluded that

$$\|J_{k+1}\|_2 \le \alpha^2 32\left(\frac{\Delta\delta\gamma}{p_0}\right)^2 \doteq \alpha^2 K(\gamma,\delta), \tag{50}$$

where

$$K(\gamma,\delta) = 32\left(\frac{\Delta\delta\gamma}{p_0}\right)^2. \tag{51}$$

From (46)-(49), we obtain

$$\begin{aligned} &\left|\mathbf{x}^\top(k\delta)J_{k+1}\mathbf{x}(k\delta)\right| = \left|\mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)J_{k+1}\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right| \\ &\le \alpha^2 K(\delta,\alpha)\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\|_2^2 \le \frac{\alpha^2K(\gamma,\delta)}{\lambda_2\left(\overline{L}\right)}V(k\delta). \end{aligned} \tag{52}$$

We also have that,

$$\begin{aligned} &\mathbb{E}\left[\mathbf{x}^\top(k\delta)\left(\overline{\mathbf{L}}\overline{\mathbf{L}}_{k\delta,\delta} + \overline{\mathbf{L}}_{k\delta,\delta}\overline{\mathbf{L}}\right)\mathbf{x}(k\delta) \mid \mathcal{F}_{k\delta}\right] \\ &= 2\mathbf{x}^\top(k\delta)\overline{\mathbf{L}}^2\mathbf{x}(k\delta) + \mathbf{x}^\top(k\delta)\left(\overline{\mathbf{L}}\widetilde{\mathbf{L}}_{k\delta,\delta} + \widetilde{\mathbf{L}}_{k\delta,\delta}^T\overline{\mathbf{L}}\right)\mathbf{x}(k\delta) \\ &= 2\mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}^2\mathbf{x}_{\mathcal{C}^\perp}(k\delta) \\ &+ \mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\left(\overline{\mathbf{L}}\widetilde{\mathbf{L}}_{k\delta,\delta} + \widetilde{\mathbf{L}}_{k\delta,\delta}^\top\overline{\mathbf{L}}\right)\mathbf{x}_{\mathcal{C}^\perp}(k\delta). \end{aligned} \tag{53}$$

Moreover, the symmetry and connectedness of $\overline{\mathbf{L}}$ imply

$$\begin{aligned} &\mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}^2\mathbf{x}_{\mathcal{C}^\perp}(k\delta) \ge \lambda_2^2(\overline{\mathbf{L}})\left\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right\|_2^2 \\ &\ge \frac{\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})}\mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\overline{\mathbf{L}}\mathbf{x}_{\mathcal{C}^\perp}(k\delta) = \frac{\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})}V(k\delta). \end{aligned} \tag{54}$$

Now, if δ and γ are chosen so that (32) holds, we have

$$\begin{aligned} &\left|\mathbf{x}_{\mathcal{C}^\perp}^\top(k\delta)\left(\overline{\mathbf{L}}\widetilde{\mathbf{L}}_{k\delta,\delta} + \widetilde{\mathbf{L}}_{k\delta,\delta}^\top\overline{\mathbf{L}}\right)\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right| \\ &\le 2\varepsilon\lambda_N(\overline{\mathbf{L}})\left\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\right\|_2^2 \le \frac{2\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})}V(k\delta). \end{aligned} \tag{55}$$

By (43) and (52)-(55), we obtain

$$\begin{aligned} &\mathbb{E}\left[V((k+1)\delta) \mid \mathcal{F}_{k\delta}\right] \le V(k\delta) - \frac{2\alpha\delta\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})}V(k\delta) \\ &+ \frac{2\alpha\delta\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})}V(k\delta) + \frac{\alpha^2K(\gamma,\delta)}{\lambda_2(\overline{\mathbf{L}})}V(k\delta) \\ &= \left(1 - \frac{2\alpha\delta\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})} + \frac{2\alpha\delta\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})} + \frac{\alpha^2K(\gamma,\delta)}{\lambda_2(\overline{\mathbf{L}})}\right)V(k\delta). \end{aligned} \tag{56}$$

Finally, using the law of iterated expectations, we obtain

$$\begin{aligned} &\mathbb{E}\left[V((k+1)\delta)\right] \\ &\le \left(1 - \frac{2\alpha\delta\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})} + \frac{2\alpha\delta\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})} + \frac{\alpha^2K(\gamma,\delta)}{\lambda_2(\overline{\mathbf{L}})}\right)\mathbb{E}\left[V(k\delta)\right]. \end{aligned} \tag{57}$$

With the above development in place, we formally state and prove the main result of this paper.

5. Main Result On Consensus Convergence

In this section we formally state the main result of this paper concerning the convergence to consensus of the proposed algorithm.

Theorem 5.1. *In a connected graph, for the queue-based broadcast gossip algorithm in* (7)-(8)*, there exist small enough $\alpha > 0$ and appropriately large δ and γ such that the sequence $\{\mathbf{x}(t)\}$ achieves consensus in mean-squared sense geometrically fast. Formally, the convergence exponent is*

$$\limsup_{t\to\infty}\frac{1}{t}\log\left(\mathbb{E}\left\|\mathbf{x}_{\mathcal{C}^\perp}(t)\right\|\right) \le \frac{1}{2\delta}\log\left(1 - f(\alpha,\epsilon,\gamma,\delta)\right) < 0, \tag{58}$$

Here, $f(\alpha,\epsilon,\gamma,\delta)$ is given by

$$f(\alpha,\varepsilon,\gamma,\delta) = 2\alpha\delta\left(\frac{\lambda_2^2\left(\overline{\mathbf{L}}\right)}{\lambda_N\left(\overline{\mathbf{L}}\right)} - \frac{\varepsilon\lambda_N\left(\overline{\mathbf{L}}\right)}{\lambda_2\left(\overline{\mathbf{L}}\right)} - \frac{\alpha K(\gamma,\delta)}{2\delta\lambda_2\left(\overline{\mathbf{L}}\right)}\right), \tag{59}$$

$\overline{\mathbf{L}}$ is as given in (11) *and $K(\gamma,\delta)$ is as defined in* (51)*.*

Proof of Theorem 5.1. We establish geometrically fast consensus in the mean-squared sense in two steps. The first step deals with establishing consensus of a subsequence and the choice of parameters needed for this, while the second step extends the geometrically fast mean-squared consensus to the entire sequence.

Lemma 5.1. *Let graph G be connected. Given the queue-based broadcast gossip algorithm in* (7)-(8)*, there exist α and appropriately large δ and γ such that the subsequence $\{\mathbf{x}(k\delta)\}$ achieves consensus in mean-squared sense geometrically fast.*

Proof. In what follows, we provide a constructive proof that identifies such α, γ and δ and a characterization of the associated consensus exponent. To this end, it suffices to show that (see (57)) $(\alpha, \varepsilon, \gamma, \delta)$ can be chosen such that

$$0 < f(\alpha, \varepsilon, \gamma, \delta) \doteq \left(\frac{2\alpha\delta\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})} - \frac{2\alpha\delta\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})} - \frac{\alpha^2 K(\gamma,\delta)}{\lambda_2(\overline{\mathbf{L}})}\right) < 1. \quad (60)$$

First, choose $\varepsilon > 0$ small enough such that

$$\frac{\varepsilon\lambda_N(\overline{\mathbf{L}})}{\lambda_2(\overline{\mathbf{L}})} \leq \frac{1}{4}\frac{\lambda_2^2(\overline{\mathbf{L}})}{\lambda_N(\overline{\mathbf{L}})}. \quad (61)$$

Now, fixing ε as above, by Proposition 4.2, choose δ and α such that the assertion in Proposition 4.2, i.e.,

$$\sup_{t\geq 0}\left\|\widetilde{\mathbf{L}}_{t,\delta}\right\|_2 \leq \varepsilon \text{ a.s.} \quad (62)$$

holds. (The proof of that proposition suggests one such choice.) Now, define

$$\alpha^{\max} = \min\left(\frac{p_0}{4\Delta\delta\gamma}, \frac{\delta\lambda_2^3\left(\overline{\mathbf{L}}\right)}{2K(\gamma,\delta)\lambda_N\left(\overline{\mathbf{L}}\right)}\right). \quad (63)$$

Note that, by (61)-(63), for any $\alpha \in (0, \alpha^{\max})$ we have that

$$f(\alpha,\varepsilon,\gamma,\delta) = 2\alpha\delta\left(\frac{\lambda_2^2\left(\overline{\mathbf{L}}\right)}{\lambda_N\left(\overline{\mathbf{L}}\right)} - \frac{\varepsilon\lambda_N\left(\overline{\mathbf{L}}\right)}{\lambda_2\left(\overline{\mathbf{L}}\right)} - \frac{\alpha K(\gamma,\delta)}{2\delta\lambda_2\left(\overline{\mathbf{L}}\right)}\right) \geq \frac{\alpha\delta\lambda_2^2\left(\overline{\mathbf{L}}\right)}{\lambda_N\left(\overline{\mathbf{L}}\right)} > 0. \quad (64)$$

Also, by (60), we have that,

$$f(\alpha,\varepsilon,\gamma,\delta) \leq \frac{2\alpha\delta\lambda_2^2\left(\overline{\mathbf{L}}\right)}{\lambda_N\left(\overline{\mathbf{L}}\right)} \leq 2\alpha\delta\lambda_N\left(\overline{\mathbf{L}}\right) = 4\alpha\delta\left\|\overline{\mathbf{L}}\right\|_2 = 4\alpha\Delta\delta < 1, \quad (65)$$

since $\alpha < \alpha^{\max} \leq \frac{p_0}{4\Delta\gamma}$ and $\gamma > p_0$.
Finally, by (57) and the above, we note that, as long as $\alpha \in (0, \alpha^{\max})$, the subsequence $\{\mathbf{x}(k\delta)\}$ satisfies

$$\mathbb{E}\left[V((k+1)\delta)\right] \leq (1 - f(\alpha,\varepsilon,\gamma,\delta))\,\mathbb{E}\left[V(k\delta)\right], \quad (66)$$

with $f(\alpha,\varepsilon,\gamma,\delta) \in (0,1)$. The above implies geometric decay of $\{V(k\delta)\}$ as $k \to \infty$, i.e.,

$$\limsup_{k\to\infty}\frac{1}{k}\log\left(\mathbb{E}\left[V(k\delta)\right]\right) \leq \log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right) < 0. \quad (67)$$

Since by (34), $V(k\delta) \geq \lambda_2\left(\overline{\mathbf{L}}\right)\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\|_2^2$ and $\lambda_2\left(\overline{\mathbf{L}}\right) > 0$, the geometric decay assertion in (67) also holds for the process $\{\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\}$. ■

In fact, we have shown the following constructive version of Lemma 5.1.

Lemma 5.2. *Let the hypotheses of Theorem 5.1 hold. Let ε be chosen to satisfy* (61)*. Further, using Proposition 4.2, choose δ and γ to satisfy* (62)*, and let $\alpha^{\max}$ be defined as in* (63)*. Then, as long as the weight parameter α is chosen to satisfy the condition $\alpha \in (0, \alpha^{\max})$, we have*

$$\limsup_{k\to\infty}\frac{1}{k}\log\left(\mathbb{E}\left[V(k\delta)\right]\right) \leq \log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right) < 0 \quad (68)$$

and

$$\limsup_{k\to\infty}\frac{1}{k}\log\left(\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\|_2^2\right]\right) \leq \log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right) < 0. \quad (69)$$

Remark 5.1. *Note that with the above choice of parameters,* (56) *implies that the sequence $\{V(k\delta)\}$ is a (non-negative) supermartingale. Using this property, we can further derive pathwise (in the a.s. sense) convergence and in $\mathcal{L}_1$ for the subsequences $\{V(k\delta)\}$ and $\{\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\}$. In fact, we have the following:*

$$\mathbb{P}\left(\limsup_{k\to\infty}\frac{1}{k}\log\left(V(k\delta)\right) \leq \log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right)\right) = 1, \quad (70)$$

$$\mathbb{P}\left(\limsup_{k\to\infty}\frac{1}{k}\log\left(\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\|_2\right) \leq \frac{1}{2}\log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right)\right) = 1, \quad (71)$$

and

$$\limsup_{k\to\infty}\frac{1}{k}\log\left(\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta)\|_2\right]\right) \leq \frac{1}{2}\log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right) < 0. \quad (72)$$

To extend the consensus results to the entire sequence $\{\mathbf{x}(t)\}$, first note that the assertions of Lemma 5.2 and Remark 5.1 apply to other subsequences of the form $\{V_{k\delta+c}\}_{k\geq 0}$ and $\{\mathbf{x}(k\delta+c)\}_{k\geq 0}$ where c is an integer in $[0, \delta-1]$. Now for instance, following (72), we have

$$\limsup_{t\to\infty}\frac{1}{t}\log\left(\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(t)\|_2\right]\right) = \limsup_{k\to\infty}\ \max_{c}\frac{1}{k\delta+c}\log\left(\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(k\delta+c)\|_2\right]\right) \leq \frac{1}{2\delta}\log\left(1 - f(\alpha,\varepsilon,\gamma,\delta)\right) < 0. \quad (73)$$

and similarly for the other assertions in Lemma 5.2 and Remark 5.1, which continue to hold, but with the corresponding exponents scaled by $1/\delta$ as seen in (73). ■

Remark 5.2. *The exponent of geometric consensus derived above is somewhat loose due to the complex bounds used in the analysis, necessitated by the Markov queues used as state update weights. This prevents us from comparing our exponent directly with the one for the single-node broadcast algorithm in [15] for general networks. However, for networks where the graph Laplacian $\overline{\mathbf{L}}$ has eigenvalue ratio $\lambda_N\left(\overline{\mathbf{L}}\right)/\lambda_2\left(\overline{\mathbf{L}}\right)$ and maximum degree bounded for all N, such as non-bipartite Ramanujan graphs, we can show that while the convergence exponent of the broadcast gossip algorithm in [15] goes to 0 as $N \to \infty$, our exponent is strictly negative, independent of the number of nodes in the network. Thus, for such networks, our algorithm converges quickly, since it makes use of simultaneous broadcasts.*

We show the above property of fast convergence specifically for a k-regular non-bipartite Ramanujan graph with fixed k (see [16]). There,

$$\lambda_2\left(\overline{\mathbf{L}}\right) \geq k - 2\sqrt{k-1}, \quad \lambda_N\left(\overline{\mathbf{L}}\right) \leq k + 2\sqrt{k-1}. \quad (74)$$

From the above bounds and (64), we conclude that

$$f(\alpha,\varepsilon,\gamma,\delta) \geq \frac{\alpha\delta\left(k - 2\sqrt{k-1}\right)^2}{k + 2\sqrt{k-1}}. \quad (75)$$

Choosing ε such that (61) holds with equality, we have

$$\varepsilon = \frac{1}{4}\frac{\lambda_2^3\left(\overline{\mathbf{L}}\right)}{\lambda_N^2\left(\overline{\mathbf{L}}\right)} \geq \frac{1}{4}\frac{\left(k - 2\sqrt{k-1}\right)^3}{\left(k + 2\sqrt{k-1}\right)^2}. \quad (76)$$

Therefore, from the choice of parameters indicated in the proof of Proposition 4.2 and noting that $\Delta = k$, we have

$$\gamma \leq \frac{48k\left(k + 2\sqrt{k-1}\right)^2}{p_0\left(k - 2\sqrt{k-1}\right)^3}. \quad (77)$$

Also, from (63), choosing $\alpha = \frac{c\delta\lambda_2^3(\overline{\mathbf{L}})}{2K(\gamma,\delta)\lambda_N(\overline{\mathbf{L}})}$, where c is a constant chosen such that $\alpha < \frac{p_0}{4\Delta\delta\gamma}$, and from (51), (74) and (77), we have

$$\alpha\delta \geq \frac{c' p_0^4 \left(k - 2\sqrt{k-1}\right)^9}{k^4 \left(k + 2\sqrt{k-1}\right)^5}, \tag{78}$$

where $c' = \frac{c}{2^{14} \times 9}$. Now, from (75) and (78), it can be concluded that

$$f(\alpha, \varepsilon, \gamma, \delta) \geq \frac{c' p_0^4 \left(k - 2\sqrt{k-1}\right)^{11}}{k^4 \left(k + 2\sqrt{k-1}\right)^6}, \tag{79}$$

where the last term is a positive constant. Thus, the exponent, $\frac{1}{2\delta}\log\left(1 - f(\alpha, \varepsilon, \gamma, \delta)\right)$, is less than a negative constant. For comparison, the exponent of the broadcast gossip algorithm in [15] is given by $\log(1-g)$ where

$$g = \frac{2\beta(1-\beta)}{N}\lambda_2\left(\overline{\mathbf{L}}\right) + \frac{(1-\beta)^2}{N^2}\lambda_2^2\left(\overline{\mathbf{L}}\right), \tag{80}$$

and $1-\beta$ is the update weight. The factor $\frac{1}{N}$ in (80) is due to the fact that only one node can transmit at each time. Note that $\log(1-g) \to 0$ as $N \to \infty$

While we are unable to formally show that our algorithm's exponent of convergence is better than in the single-broadcast case in general graphs, we compare the two through simulations in the next section.

A. Analysis of Error-in-average

In this section we characterize the error-in-average. Note that

$$\mathbf{x}(t+1) = \left(\mathbf{I} - \alpha\mathbf{L}(t+1)\right)\mathbf{x}(t) = \mathbf{x}(t) - \alpha\mathbf{L}(t+1)\mathbf{x}_{\mathcal{C}^\perp}(t). \tag{81}$$

Thus,

$$\overline{x}(t+1) - \overline{x}(t) = -\frac{1}{N}\alpha(\mathbf{1}^T\mathbf{L}(t+1))\mathbf{x}_{\mathcal{C}^\perp}(t). \tag{82}$$

Using telescoping and the triangle inequality, we obtain

$$\limsup_{t\to\infty} \mathbb{E}\left[|\overline{x}(t+1) - \overline{x}(0)|\right]$$
$$\leq \sum_{s=0}^{\infty} \mathbb{E}\left[|\overline{x}(s+1) - \overline{x}(s)|\right] \tag{83}$$
$$\leq \frac{1}{N}\alpha\left(\sqrt{N}.\sqrt{2\Delta}\frac{\gamma}{p_0}\right)\sum_{s=0}^{\infty}\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(s)\|_2\right], \tag{84}$$

where (84) follows from (82) and Proposition 4.3. The summation in (84) is clearly bounded by the geometric decay in (73) (a geometric bound for each t needs to be obtained using (34), (63), (66) and (81), the details of which are omitted). Therefore, the asymptotic modulus of error-in-average can be expressed as a function of the exponent in (73) as follows:

$$\limsup_{t\to\infty} \mathbb{E}\left[|\overline{x}(t) - \overline{x}(0)|\right]$$
$$\leq \frac{1}{\sqrt{N}}\alpha\sqrt{2\Delta}\frac{\gamma}{p_0}\frac{1}{1 - \left(1 - f(\alpha, \varepsilon, \gamma, \delta)\right)^{\frac{1}{2\delta}}}2\sqrt{\frac{\lambda_N\left(\overline{\mathbf{L}}\right)}{\lambda_2\left(\overline{\mathbf{L}}\right)}}\|\mathbf{x}_{\mathcal{C}^\perp}(0)\|_2 \tag{85}$$
$$\leq \frac{1}{\sqrt{N}}\sqrt{2\Delta}\frac{\alpha\gamma}{p_0}\frac{2\delta}{f(\alpha, \varepsilon, \gamma, \delta)}2\sqrt{\frac{\lambda_N\left(\overline{\mathbf{L}}\right)}{\lambda_2\left(\overline{\mathbf{L}}\right)}}\|\mathbf{x}_{\mathcal{C}^\perp}(0)\|_2 \tag{86}$$
$$\leq \frac{1}{\sqrt{N}}4\sqrt{2\Delta}\frac{\gamma}{p_0}\frac{\lambda_N^{3/2}\left(\overline{\mathbf{L}}\right)}{\lambda_2^{5/2}\left(\overline{\mathbf{L}}\right)}\|\mathbf{x}_{\mathcal{C}^\perp}(0)\|_2, \tag{87}$$

where (87) follows from (64).

Remark 5.3. *The upper bound in* (87) *shows that for a k-regular non-bipartite Ramanujan graph with fixed k, using the bounds in* (74)*, the asymptotic error-in-average is $O(\frac{1}{\sqrt{N}})$. Thus this error tends to zero for large graphs, so that the consensus value is indeed accurate, similar to [15], although the latter converges slowly compared to our algorithm for this graph, as noted in Remark 5.2.*

6. Simulation Results

The performance of the queue-based broadcast gossip algorithm proposed in this paper is illustrated through simulations on the following networks of 100 nodes: networks modeled as ring-structured and grid-structured graphs. These are compared against the performance of the single-node broadcast algorithm in [15]. In each case, the initial value of each node is chosen uniformly at random from the interval $(0, 100)$. The single-node broadcast gossip algorithm uses $\beta = 0.5$, where $1-\beta$ is the fixed update weight, since Corollary 2 in [15] shows this to be the optimal value for such graphs as $N \to \infty$. For the queue-based algorithm, the transmission probability of each node is chosen to be $p = 0.33$ in ring-structured network and $p = 0.2$ in grid-structured network. $\gamma = 0.15$ for ring-structured network and $\gamma = 0.1$ for grid-structured network; both use $\alpha = 0.5$.

Figures 1 and 2 plot the expected norm of the consensus error $\mathbb{E}\left[\|\mathbf{x}_{\mathcal{C}^\perp}(t)\|_2\right]$ vs. the number of iterations for the two networks. In each case, by choosing appropriate parameters, our proposed algorithm converges much faster compared to the single-node broadcast gossip algorithm, essentially because we allow simultaneous broadcasts.

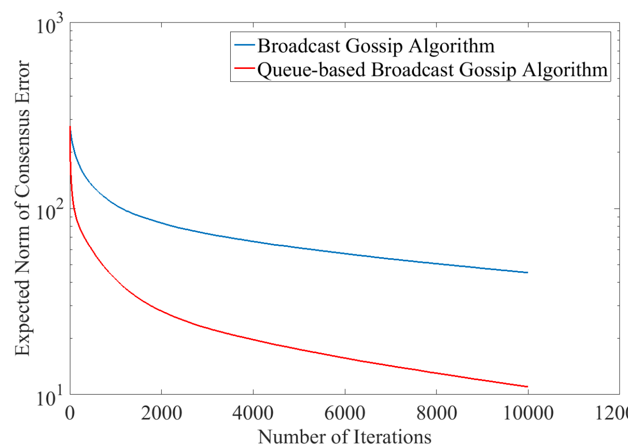

Fig. 1: Expected norm of the consensus error vs. number of iterations achieved by queue-based and single-node broadcast gossip algorithms for a network of 100 nodes modeled as a ring-structured graph.

Now we demonstrate that the requirement of the broadcast gossip algorithm in [15], that only one node transmits at each time, cannot be easily dispensed with in that scheme. By increasing the probability of transmission in their scheme, we can force simultaneous transmissions there, so that their algorithm may also converge faster. However, due to the resulting collisions, the consensus result obtained in that fixed-weight algorithm will be unsatisfactory, as seen below. Consider an unbalanced network consisting of two cliques of 20 and 10 nodes, respectively, connected by one common node. Assume that the initial value is -1 in all nodes of the larger clique and +1 in all other nodes. It is assumed that the nodes have reached a consensus when the difference between the maximum and minimum values of the nodes becomes less than 0.01 at which point the iterations are stopped. Figure 3 plots modulus of the error-in-average vs. the number of iterations before being stopped, for our

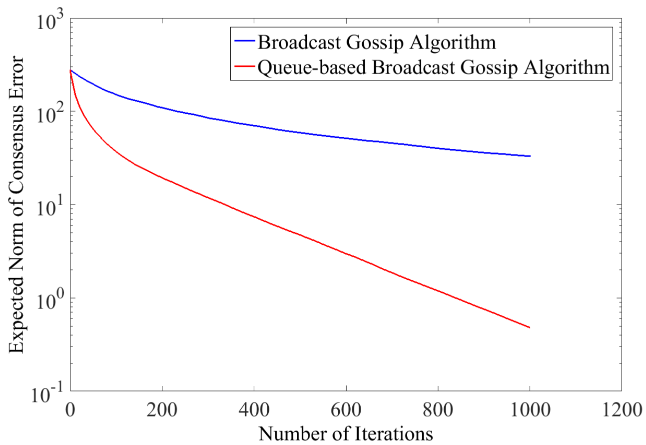

Fig. 2: Expected norm of the consensus error vs. number of iterations achieved by queue-based and single-node broadcast gossip algorithms for a network of 100 nodes modeled as a grid-structured graph.

proposed algorithm and the one with fixed weights. The transmission probability of each node in both algorithms is chosen to be $p = 0.1$ so that collisions occur frequently. As illustrated in Figure 3, for any given number of iterations, the consensus value of the algorithm in [15] deviates significantly from unweighted consensus value, compared to the error in our queue-weighted algorithm. Thus, the trade-off between the time needed for convergence and the error-in-average after convergence is significantly better in our algorithm. This happens because the rate of correct packet reception in the link (i, j) is different from that in link (j, i) even in expectation in their case. i.e., the expected update weight matrix is non-symmetric. While this will happen in all graphs, the effect is particularly strong in our chosen example because the node connecting the two cliques will transmit packets successfully at a much higher rate to the smaller clique than to the larger clique, resulting in a highly non-symmetric expected weight matrix. This illustrates the importance of using queue weights to automatically adapt to these rate differences, as in our algorithm.

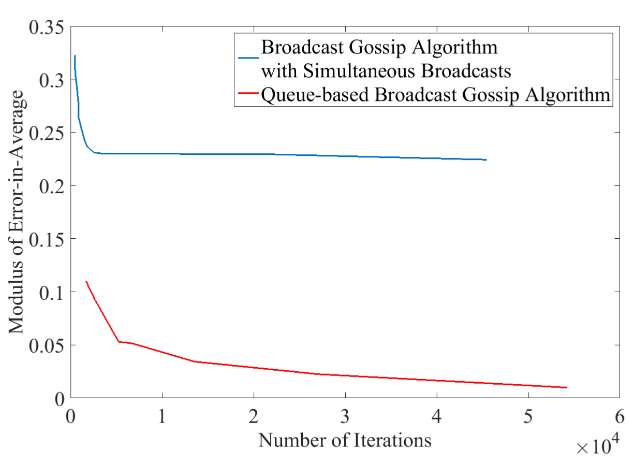

Fig. 3: Consensus value deviates significantly from unweighted consensus value if fixed weights are used to update state.

7. Conclusion

This paper proposed a queue-based broadcast gossip algorithm for distributed average consensus in an arbitrarily connected network of sensors. Our algorithm allows simultaneous broadcasts to speed up convergence to consensus. Since these broadcasts naturally cause collisions, we specify a novel state update algorithm that uses certain queue lengths as update weights. We proved the geometric convergence of this algorithm to consensus, assuming an appropriate choice of parameters. The algorithm does not require the nodes to have global information, such as the structure of the network. (It does need a single global parameter - a lower bound on correct packet reception probability.) For the particular case of networks modeled as non-bipartite Ramanujan graphs, we have shown formally that the convergence exponent of the proposed algorithm is independent of the number of nodes in the network, which is important in large-scale sensor networks. Tightening the bounds used to derive the convergence exponent remains to be addressed in the future. This could help us compare our algorithm formally with ones that only allow a single node broadcast.

References

[1] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575 – 3605, June 2012.

[2] J. Chen and A. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, 2012.

[3] I. Schizas, G. Mateos, and G. Giannakis, "Stability analysis of the consensus-based distributed LMS algorithm," in *Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, Nevada, USA, April 1-4 2008, pp. 3289–3292.

[4] S. Ram, A. Nedić, and V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of optimization theory and applications*, vol. 147, no. 3, pp. 516–545, 2010.

[5] A. K. Sahu and S. Kar, "Recursive distributed detection for composite hypothesis testing : Algorithms and Asymptotics," *arXiv preprint arXiv:1601.04779*, 2016.

[6] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations analysis of consensus+innovations detection in random networks," in *The 49th Annual Allerton Conference on Control, Communication, and Computing*, Monticello, IL, Sept. 28 - 30 2011, pp. 151–155.

[7] J. Tsitsiklis, "Problems in decentralized decision making and computation," PHD, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[8] R. O. Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *42nd IEEE Conference on Decision and Control*, 2003.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, pp. 65–78, 2004.

[10] A. T. Salehi and A. Jadbabaie, "On consensus in random networks," in *The Allerton Conference on Communication, Control, and Computing*, Allerton House, IL, September 2006.

[11] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[13] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.

[14] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, March 2008.

[15] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, July 2009.

[16] S. Kar, S. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, 2008.