Broadcast Gossip Algorithms

Xiaole Zou

August 22, 2025

Abstract

From the 30th of June to the 22nd of August 2025, I studied at the institute of Science Tokyo under the supervision of Prof. Naoto Miyoshi, cosupervised by Keigo Misawa. During this period, broadcast gossip algorithms were investigated. Through studying existing literature and techniques, this project aims to generalise the theory of convergence for broadcast gossip algorithms onto weighted graphs with stochastic mixing parameters. The following is a complete set of notes on the project.

Contents

1 Introduction

Networks are often used as models to many applications, ranging from physics to biology and social sciences. Often it is the case that nodes will have certain values associated with them. For example, in social contexts, the nodes could represent people whilst the values represents their opinion on a certain event. In such models, algorithms are performed in order to model change. The change of opinion perhaps in the context of social networks. Hence, for many algorithms it is essential to study the achievement of consensus. Moreover, in an ideal case, you may expect the value reached at consensus to be related to the initial average in some way. Due to this possible social application of such models, a key component of human societies is the ability to gossip. Hence, inspired by this, gossip algorithms are performed in order to drive systems towards consensus.

Gossip Algorithms were first introduced as a gossip inspired type of algorithm aiming for nodes in a network to achieve global consensus through local communications [4]. The classical approach studies a peer to peer gossip algorithm. However, considering that in addition to achieving consensus it is also ideal for communication cost (in this case the number of steps required) to be minimised. In order to achieve this, a pairwise communication method is replaced with a broadcasting method, allowing one-to-many information transfer which may lead to a faster convergence rate.

Previous studies have shown that the standard broadcast gossip algorithm not only achieves consensus almost surely, the communication cost is also lower than other existing approaches [1]. Hence, the aim of this project is to understand the mathematical techniques used in proving convergence for the standard model, and extend it to more general settings.

However, also inspired by gossiping, it makes intuitive sense that the importance of characters may vary and more influential nodes may have a more significant impact on the neighbouring nodes whilst others have a lesser impact. Hence we study a stochastic mixing parameter which is dependent on the broadcasting node but independent of time and aim to show convergence in this context as well.

This set of notes studies the existing paper, filling in the gaps of proofs as well as assumptions made which were not explicitly mentioned. Then, we aim to generalise to a weighted setting, and explore the impacts of adding stochasticity to the mixing parameter.

Section 2 aims to complete the original paper's proofs for convergence. Section 3 discusses how the algorithm can be extended to a weighted graph. Section 4 introduces a stochastic mixing parameter. Sections 3 and 4 will also be accompanied with numerical results to demonstrate the results derived.

2 Initial Model

This section is based on the existing paper [1]. The model produced throughout the project is an adjustment to the model discussed in the above literature.

In summary, the original paper proposes a broadcast gossip algorithm on a random graph. Then, using matrices and matrix properties, it proves convergence in expectation, second moment to a consensus as hence deducing convergence almost surely. It also studies mean square error theoretically, and the per node variance using numerical methods. Finally, it touches on communication cost being lower than traditional models.

2.1 Graph and Time Models

We first discuss the graph model. The graph model considered is a random geometric graph denoted $G(N, R)$. N the number of nodes which are generated uniformly in the euclidean unit square. Two nodes are then connected by a link if their euclidean distance is less than R. Previous research has shown that to have good connectivity and minimise interference, R has to scale like $\Theta(\sqrt{\log(N)/N})$.

For the purpose of this paper, we assume the resultant graph to be strongly connected, and denote its adjacency matrix by Φ. Furthermore, we define $\mathcal{N}_i = \{j \in \{1, 2, ..., N\} : \Phi_{ij} \neq 0\}$ the neighbours of node i.

We consider an asynchronous time mode. For this model, each node has a clock which ticks independently at a rate μ Poisson process. It can be shown that this is equivalent to a single clock ticking with rate $N\mu$ Poisson process. For purposes of studying convergence, we discretize time by counting the number of ticks rather than the absolute time spanned.

At time t, each node has $x_i(t)$ an estimate of the global average. Let $x(t)$ denote the N-vector of such estimates. We aim to drive $x(k)$ as close as possible to $\overline{x}(0)\mathbf{1}$ where $\overline{x}(0) = 1/N \sum_{i=1}^{N} x_i(0)$.

2.2 The Algorithm

The algorithm is conducted as follows:

1. At time t, a node i is chosen uniformly and broadcasts its current value, $x_i(t)$ over the network.

2. This is received by every node within a radius R.

3. Neighbouring nodes update their value as follows:

$$x_k(t+1) = \gamma x_k(t) + (1-\gamma)x_i(t) \quad \forall k \in \mathcal{N}_i \tag{1}$$

   γ is the mixing parameter.

4. The remaining nodes leave their values unchanged.

This can be written as the following:

$$x(t+1) = W(t)x(t) \tag{2}$$

where random matrix $W(t)$ with probability $1/N$ is given by

$$W_{jk}^{(i)} = \begin{cases} 1 & j \notin \mathcal{N}_i, \quad k = j \\ \gamma & j \in \mathcal{N}_i, \quad k = j \\ 1-\gamma & j \in \mathcal{N}_i, \quad k = i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Note since $i \notin \mathcal{N}_i$ by definition of adjacency matrix, we have that i, j distinct hence the matrix is well defined.

Lemma 2.1. *The weight matrices $\{W^{(i)} : i = 1, 2, ..., N\}$ satisfy the followings:*

1. *$\mathbf{1}$ is a right eigenvector of $W^{(i)}$ with eigenvalue $1 \,\forall i$*

2. *$\mathbf{1}$ is not a left eigenvector for any $W^{(i)}$.*

Proof. It is easy to show the rows sum to 1 by definition. Hence, we only show the second item of the lemma. Fix i and let $k = i$. Then, we have:

$$\sum_{k=1}^{N} W_{jk}^{(i)} = 1 + (1-\gamma)|\mathcal{N}_i| > 1 \tag{4}$$

where in the last inequality, we used that $\gamma \in (0,1)$ and strong-conectedness wich gives $|\mathcal{N}_i| \geq 1$. Hence for every i, there exists a column which doesn't sum to 1 implying that $\mathbf{1}$ can not be a left eigenvector. $\square$

The above lemma concludes that $c\mathbf{1}$ is a fixed point and hence if the algorithm reaches a consensus, it does not leave. This also shows that the sum of the states is not preserved at each step, meaning that the convergent state may not be the desired one stated at the start. In fact, we can bound the change in sum of global estimates by spotting the following:

$$\left|\sum_{j=1}^{N}(x_j(t+1) - x_j(t))\right| \leq (1-\gamma)\,|\mathcal{N}_i|\,K \tag{5}$$

where K is the maximal difference between $x_i(t)$ and $x_j(t)$, $j \in \mathcal{N}_i$. This relation is obtained by applying equation (1) and bounding each difference from above as global estimations are finite.

The next lemma considers the mean of the i.i.d. $W(t)$ as $\mathbb{E}(W(t)) = W$.

Lemma 2.2. *The average weight matrix W is given by*

$$W = I - \frac{1-\gamma}{N}\operatorname{diag}\{\Phi\mathbf{1}\} + \frac{1-\gamma}{N}\Phi \tag{6}$$

and we have the following relation:

$$W\mathbf{1} = \mathbf{1},\ \mathbf{1}^T W = \mathbf{1}^T,\ \rho(W-J) < 1 \tag{7}$$

where $\rho(\cdot)$ is the spectral radius of its arguement and $J = (N)^{-1}\mathbf{1}\mathbf{1}^T$

Proof. Note the current proof is different to that of the literature. To be confirmed with supervisor.

By definition of the average weight matrix, we find

$$W_{jk} = \frac{1}{N}\sum_{i=1}^{N} W_{jk}^{(i)} \tag{8}$$

Then, by (3) we obtain

$$W_{jk} = \begin{cases} 1 - \frac{|\mathcal{N}_j|}{N} + \frac{\gamma|\mathcal{N}_j|}{N} & k = j \\ \frac{1-\gamma}{N} & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

This coincides with the expression given in (6) hence the first half of the lemma is complete.

We note that further $W = I - \eta L$ for $L = \operatorname{diag}\{\Phi\mathbf{1}\} - \Phi$ the graph laplacian and $\eta = (1-\gamma)/N$. Since we have $0 < \eta < 1 - 1/N$ we apply this to prove the remaining lemma.

We know that both I and L are symmetric so if $\mathbf{1}$ is a right eigenvector then it is also a left eigenvector. Right eigenvector is readily checked using the previous expression we've found hence the first two claims are done.

First, we note that the graph laplacian is positive-semidefinite which means for λ_i eigenvalues of the graph laplacian, we have $\lambda_i \geq 0$. Assuming connectivity, we have $\lambda_N = 0$ and all other eigenvalues strictly positive.

Let $\mathbf{v}_i$ be the corresponding eigenvectors of L, then we have $\mathbf{v}_N = \mathbf{1}$. Hence, by orthogonality of eigenspaces, we have $\mathbf{1}^T\mathbf{v}_i = 0, i \neq N$. Hence we have for all other eigenvectors:

$$\begin{aligned}(I - \eta L - J)\mathbf{v}_i &= \mathbf{v}_i - \eta\lambda_i\mathbf{v}_i - J\mathbf{v}_i \\ &= (1-\eta\lambda_i)\mathbf{v}_i\end{aligned}$$

which shows all other eigenvalues of $W - J$ must be of form $1 - \eta\lambda_i$. I now claim these values are of modulus less than 1.

Note by strictly positive assumption justified before, we have $1 - \eta\lambda_i < 1$. Also, note that by the properties of laplacian matrix, we have $\lambda_i \leq 2 * d$ for d the maximum degree of graph (this is shown in B). Hence we have $\lambda_i \leq 2N < \frac{2N}{1-\gamma}$. Rearrangeing we get that the eigenvalues of $W - J$ must be of absolute value < 1 which concludes the proof. □

This shows that the average weight matrix is unlike the individual weight matrix with $\mathbf{1}$ and both a left and right eigenvector. Furthermore, the result on spectral radius is important and will be used later.

2.3 Convergence

The theorems provided here hold true for any iterative model $x(t+1) = W(t)x(t)$ with $W(t)$ satisfying some of the conditions provided in section 2.2. In summary, to obtain convergence in expectation, we need $x(0)$ is bounded and $W = \mathbb{E}(W(t))$ to satisfy 2.2. In order to obtain convergence in second moment, we require the leading eigenvalue of $\mathbb{E}\{W(t)^T(I - J)W(t)\}$ to be strictly less than 1. In order to obtain convergence almost surely, we require $x(0)$ bounded and we require convergence in second moment. To show that the value it converges to is the desired value, we need convergence in expectation.

We first show **convergence in expectation**:

Proposition 2.3. *Given the matrix W satisfies lemma 2.2, and the values $x_i(0)$ are bounded, we have the following:*

$$\mathbb{E}(\lim_{t\to\infty} x(t)) = \frac{1}{N}\boldsymbol{1}\boldsymbol{1}^T x(0) \tag{10}$$

Proof. Later we will show that the mean square error of $x(t)$ is bounded above and bellow. Hence, we have that if $x(0)$ is bounded, so is $x(t)$. We are then able to apply dominating convergence theorem and get the following:

$$\begin{aligned}
\mathbb{E}(\lim_{t\to\infty} x(t)) &= \lim_{t\to\infty} \mathbb{E}(x(t)) \\
&= \lim_{t\to\infty} \mathbb{E}(W(t)x(t-1)) \\
&= \lim_{t\to\infty} W^t x(0)
\end{aligned}$$

where the final equality follows from the fact that the $W^{(i)}$'s are i.i.d. Hence, we require that $\lim_{t\to\infty} W^t = 1/N\ \mathbf{1}\mathbf{1}^T$ to complete the proof.

The requirement for the above to be true is given in the literature [2], which are exactly the second statement of 2.2. Hence, the proof is complete. □

We now move on to show **convergence in second moment**. To study this, we define the vector of deviations, $\beta(t)$, by defining the entries $\beta_i(t) = x_i(t) - \overline{x}(t)$. Equivalently, we write:

Definition 2.4.

$$\beta(t) = x(t) - J\ x(t) = (I - J)\ x(t) \tag{11}$$

for J defined as in lemma 2.2.

We first study consensus being reached in finite time:

Lemma 2.5. *There is a consensus at time slot $T \in \mathbb{N}$ if and only if* $\mathbb{E}(\|\beta(T)\|_2^2) = 0$

Proof. ($\Rightarrow$) Suppose we have a consensus at time T. Then, we have $x(t) = c\mathbf{1}$ for some c:

$$\beta(t) = (I - J)\ c\mathbf{1} = c\mathbf{1} - \frac{c}{N}\mathbf{1}\mathbf{1}^T\mathbf{1} = 0$$

Hence, taking the norm and the expectation we get the right-hand side.

($\Leftarrow$) Suppose at them T we have that $\mathbb{E}(\|\beta(T)\|_2^2) = 0$. Then, since $\|\beta(T)\|_2^2 \geq 0$ we must also have $\|\beta(T)\|_2^2 = 0 \Rightarrow \beta_i(T) = 0\, \forall i$. Hence, we can conclude there is no deviation from the average value indicating $x(t) = c\mathbf{1}$ and hence a consensus is reached. □

We now provide sufficient conditions for convergence in second moment. Let λ_i denote the i^{th} ranked eigenvalue.

Lemma 2.6. *$\mathbb{E}(\|\beta(t)\|_2^2)$ converges to 0 if*

$$\lambda_1(\mathbb{E}(W(t)^T(I-J)W(t))) < 1 \tag{12}$$

Proof. First we note the following:

$$\begin{aligned}
(W(t) - JW(t))\beta(t) &= (W(t) - JW(t))(x(t) - Jx(t)) \\
&= W(t)x(t) - JW(t)x(t) - W(t)Jx(t) + JW(t)Jx(t) \\
&= x(t+1) - Jx(t+1) - Jx(t) + Jx(t) \\
&= \beta(t+1)
\end{aligned}$$

Noting that we used the definition $J = N^{-1}\mathbf{1}\mathbf{1}^T$ and the fact that $\mathbf{1}$ is a right eigenvector of $W(t)$ to conclude $W(t)J = J$ and $J^2 = J$ to complete the above recurrence relation. Hence let us denote $Y(t) = W(t) - JW(t)$ then we have $\beta(t+1) = Y(t)\beta(t)$.

We now note that since $Y(t)^TY(t)$ is symmetric, so is $\mathbb{E}(Y(t)^TY(t))$. Therefore, there exists an orthogonal P such that $\mathbb{E}(Y(t)^TY(t)) = P^T\Lambda P$ with Λ diagonal due to spectral theorem. Denoting $y(t) = P\beta(t)$ we have:

$$\begin{aligned}
\mathbb{E}(\|\beta(t+1)\|_2^2 \mid \beta(t)) &= \beta(t)^T\mathbb{E}(Y(t)^TY(t))\beta(t) \\
&= y(t)^T\Lambda y(t) \\
&= \sum_{i=1}^{N} \lambda_i(\mathbb{E}(Y(t)^TY(t)))|y(t)_i|^2 \\
&\leq \lambda_1(\mathbb{E}(Y(t)^TY(t)))y(t)^Ty(t) \\
&= \lambda_1(\mathbb{E}(Y(t)^TY(t)))\beta(t)^T\beta(t)
\end{aligned}$$

Where in the last equality we use the orthogonality of P. We can repeat the above process recursively to obtain the following relation:

$$\mathbb{E}(\|\beta(t)\|_2^2) \leq \lambda_1^t(\mathbb{E}(Y(t)^TY(t)))\|\beta(0)\|_2^2 \tag{13}$$

We notice that in order to conclude the above result, we need to ensure $|\lambda_1| < 1$. However, we note that $\mathbf{1}$ is a right eigenvector of $W(t)$ and hence of $Y(t)$ with eigenvalue 0. Hence we know $\lambda_1 \geq 0$ and hence the result follows as $t \to \infty$. $\square$

We now aim to show that the broadcast gossip algorithm actually satisfies this property. In order to do this, we first show an intermediate lemma:

Lemma 2.7. *The second moment of the weighted matrices i.e. $W' = \mathbb{E}(W(t)^TW(t))$ is given by:*

$$W' = I - \frac{2\gamma(1-\gamma)}{N}\operatorname{diag}(\Phi\mathbf{1}) + \frac{2\gamma(1-\gamma)}{N}\Phi \tag{14}$$

Proof. Let node i be updated at time t then we have the following:

$$W^{(i)^T}W^{(i)} = \begin{cases}
1 + |\mathcal{N}_i|(1-\gamma)^2 & k = j = i \\
\gamma(1-\gamma) & k \in \mathcal{N}_i, j = i \\
\gamma(1-\gamma) & j \in \mathcal{N}_i, k = i \\
\gamma^2 & j \in \mathcal{N}_i, k = j \\
1 & j \notin \mathcal{N}_i, k = j \neq i \\
0 & \text{otherwise}
\end{cases} \tag{15}$$

Hence, taking the expectation, we find that:

$$W'_{jk} = \begin{cases} 1 - \frac{|\mathcal{N}_j|}{N}(1-(1-\gamma)^2 - \gamma^2) & j = k \\ \frac{2\gamma(1-\gamma)}{N} & k \in \mathcal{N}_j \\ 0 & \text{elsewhere} \end{cases} \tag{16}$$

And the lemma is complete. We note that again we can use the graph Laplacian matrix to represent W' and since $0 < 2\gamma(1-\gamma) < 1$ for $0 < \gamma < 1$ I can apply the same arguement as before and show matrix W' satisfies (7). □

Proposition 2.8. *The broadcast gossip algorithm satisfies* $\lambda_1(\mathbb{E}(W(t)^T(I-J)W(t))) < 1$.

Proof. Firstly note that by the Rayleigh-Ritz Theorem we have $\lambda_1 = \max_{\|u\|_2^2=1} u^T W' u - u^T W'_J u$ where we define $W'_J = \mathbb{E}(W(t)^T J W(t))$.

By previous lemma, we have W' has maximum eigenvalue 1 with eigenvector $1/\sqrt{N}\ \mathbf{1}$ and for any other u_i eigenvectors, we have that $u_i^T W' u_i < 1$. Also, as $W(t)$ and J have non-negative entries, $W(t)^T J W(t)$ also has non-negative entries meaning all of its eigenvalues must be positive. We are now left to check cases.

For u_i eigenvectors of W' with $i \neq 1$ we have $u_i^T W' u_i - u_i^T W'_J u_i < 1 - u_i^T W'_J u_i < 1$. For u_1 we have $u_1^T W' u_1 - u_1^T W'_J u_1 = 1 - u_i^T W'_J u_i < 1$. Hence, concluding the proof. □

Finally we study **almost sure convergence**, which is summarised in the following theorem.

Theorem 2.9. *The broadcast gossip algorithm converges to consensus almost surely. i.e.*

$$\mathbb{P}(\ \lim_{t\to\infty} x(t) = c\mathbf{1}\ ) = 1 \tag{17}$$

for some $c \in \mathbb{R}$ *where* $\mathbb{E}(c) = \frac{1}{N}\mathbf{1}^T x(0)$

Proof. Note that for a fixed $\epsilon > 0$ we have

$$1 - \mathbb{P}(\ \lim_{t\to\infty} \|\beta(t)\|_2^2 \geq \epsilon\ ) \geq\ 1 - \frac{\mathbb{E}(\lim_{t\to\infty} \|\beta(t)\|_2^2)}{\epsilon} \tag{18}$$

$$= 1 - \frac{\lim_{t\to\infty} \mathbb{E}(\|\beta(t)\|_2^2)}{\epsilon} \tag{19}$$

$$= 1 \tag{20}$$

For the above, we first applied Markov's inequality and then Lebesgue Dominated Convergence Theorem to swap the order of limits and expectations.

Since the above is true $\forall\epsilon > 0$, we have that taking limits:

$$\mathbb{P}(\ \lim_{t\to\infty} x(t) = c\mathbf{1}\ ) = \mathbb{P}(\ \lim_{t\to\infty} \|\beta(t)\|_2^2 = 0\ ) = 1 - \mathbb{P}(\ \lim_{t\to\infty} \|\beta(t)\|_2^2 > 0\ ) = 1 \tag{21}$$

The final requirement follows from convergence in expectation 2.3. □

2.4 Performance Analysis

We first study **Mean Square Error**. To do this, we define $\alpha(t) = x(t) - Jx(0)$ with J defined as before 2.2 to denote the difference between the state vector and the average of the initial node measurements. Hence $\mathbb{E}(\|\alpha(t)\|_2^2)$ is the mean square error.

Lemma 2.10. *1. The mean square error obeys the following recursive rule:*

$$\mathbb{E}(\|\alpha(t+1)\|_2^2) \leq (1 - \lambda_2(W'))\mathbb{E}(\|J\alpha(t)\|_2^2) + \lambda_2(W')\mathbb{E}(\|\alpha(t)\|_2^2) \tag{22}$$

2.

$$x(t) \neq c\mathbf{1} \Leftrightarrow \mathbb{E}(\|\alpha(t+1)\|_2^2) < \mathbb{E}(\|\alpha(t)\|_2^2) \tag{23}$$

for some $c \in \mathbb{R}$

The proof is omitted due to length however this shows that the mean square error is strictly decreasing if a consensus is not reached.

Then, they conducted numerical studies on the per-node variance defined as $(N)^{-1}\|x(t) - Jx(t)\|_2^2$ with varying γ It was found that as γ increases, per-node variance increases but MSE decreases. This is suggesting a trade-off between time to reach consensus and MSE.

3 Broadcasting on Weighted Graphs

This is a generalisation of the broadcasting model based on existing literature [1]. We generalise the topology to be finite weighted connected graphs.

We employ the same time model as the original model, described in section 2.1. However, we adjust the graph model and the update algorithm in order to obtain the same aim which is also discussed in section 2.1.

3.1 Time Model

We consider an asynchronous time model as previously explored by [2] and [1]. Each node has a clock with ticks at the times of a rate 1 Poisson process. Therefore, the times between the ticks are exponentially distributed with rate 1 and are independent of each other. Moreover, the above is equivalent to a single clock ticking at rate N Poisson process, where at each tick a node is chosen uniformly to tick. Let Z_k be the ticking times for $k \geq 1$ then $Z_{k+1} - Z_k$ follows an exponential distribution with rate N. Let $I_k \in \{1, \dots, N\}$ be the node which ticks at time Z_k. We have that I_k are independent and identically distributed according to a uniform distribution on $\{1, \dots, N\}$.

Since $x(t)$ remains constant in-between ticks, we discretise time as done in [2]. From now, we consider $x(k)$ to be the vector of values at the end of time slot k.

3.2 Graph Model

As gossip algorithms were first introduced in the context of decentralised group decision making [4], the closeness between agents should play a role in how impactful somebody's opinion is. Intuitively it makes sense that if you're closer with somebody, their opinion may have a larger impact on your opinion and hence we introduce a weighted graph.

We consider a connected undirected weighted graph $G = (V, E, w)$ where $V = 1, 2, \dots, N$ the set of nodes, $E \subseteq \{\{u, v\} \in V^2, u \neq v\}$ the set of edges and $w : E \to \mathbb{R}^+$ the weight function of the graph. We then define the adjacency matrix $\Phi \in \mathbb{R}^{N \times N}$ with $\Phi_{uv} = w(\{u, v\})$ if $\{u, v\} \in E$ and 0 otherwise. We also denoted $\mathcal{N}_u = \{v \in V : \{u, v\} \in E\}$ the set of neighbours of the node u.

We define $x(0) = [x_1(0), \dots, x_N(0)]^T$ to be the initial value of the nodes. Defining $x_{avg} = \mathbf{1}^T x(0)/N$ the average of the initial node, our goal is to drive $x(t)$ towards $x_{avg}\mathbf{1}$.

3.3 Algorithm Generalisation

Suppose node $u \in \{1, \dots, N\}$ ticks at time $t \in \mathbb{Z}_+$. Then, node u broadcasts its value to its neighbours, causing the vector $x(t+1)$ to update as follows. If $v \notin \mathcal{N}_u$ then $x_v(t+1) = x_v(t)$. If $v \in \mathcal{N}_u$ then $x_v(t+1) = \alpha_{uv}\, x_u(t) + (1 - \alpha_{uv})x_v(t)$ for $\alpha_{uv} = \gamma \cdot f \circ w(\{u, v\})$, $\gamma \in (0, 1)$ and $f : \mathbb{R} \to (0, 1]$ some arbitrary function which scales the weights.

Note we perform the scaling as if such a scaling doesn't exist, we may have $1 - \alpha_{uv} < 0$ indicating the original opinion of node v at time t having negative impact on the node's opinion at time $t + 1$. This may in general lead to lack of consensus as you may fluctuate away from the desired average. Hence we consider a normalised case

This is also equivalent to the recursive formula $x(t+1) = W(t)x(t)$ where $W(t) = W^{(u)}$ if the node u ticks at time t, where the matrix entries are

$$W_{jk}^{(u)} = \begin{cases} 1 & j \notin \mathcal{N}_u,\ k = j \\ \alpha_{uj} & j \in \mathcal{N}_u,\ k = u \\ 1 - \alpha_{uj} & j \in \mathcal{N}_u,\ k = j \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Note hence that $W(t)$ is a random matrix distributed uniformly on the set $\{W^{(u)} : u \in V\}$. We note that this setup is equivalent to considering a graph $G' = (V, E, w')$ where $w' = f \circ w$ where now the weights of the graph are normalised to be bounded by 1 and we redefine $\alpha_{uv} = \gamma \cdot w'(\{u, v\})$. Hence, we will assume the weights of the edges are bounded above by 1 for the remainder of this article.

We now begin outlining some important characteristics of the matrices which will prove useful for proving convergence.

Lemma 3.1. *For* $\mathbf{1} = [1, \ldots, 1]^T$*, we have that* $W^{(u)}\mathbf{1} = \mathbf{1}\ \forall u \in V$

Proof. We aim to show that the rows of $W^{(u)}$ sum to unity. By applying (24), we obtain:

$$\sum_{k=1}^{N} W_{jk}^{(u)} = \underline{\mathbf{1}}\{j \notin \mathcal{N}_u\} + \underline{\mathbf{1}}\{j \in \mathcal{N}_u\}(1 - \alpha_{uv} + \alpha_{uv}) = 1 \tag{25}$$

where $\underline{\mathbf{1}}\{\cdot\}$ is the indicator function. This concludes the proof. □

Lemma 3.2. *For* $W = \mathbb{E}(W(t))$ *we have* $W = I - \eta L$ *where* $\eta = \gamma/N$*,* $L = \operatorname{diag}\{\Phi\mathbf{1}\} - \Phi$ *the weighted graph Laplacian matrix. Moreover,* $\forall\gamma \in (0, 1)$ *we have:*

$$W\mathbf{1} = \mathbf{1},\ \mathbf{1}^T W = \mathbf{1}^T,\ \rho(W - J) < 1 \tag{26}$$

for $\rho(\cdot)$ *the spectral radius of a matrix,* $J = N^{-1}\mathbf{1}\mathbf{1}^T$

Proof. Due to a uniform distribution, we have $W = 1/N \sum_{u=1}^{N} W^{(u)}$ Using (24), we get:

$$W_{jk} = \begin{cases} 1 - \frac{|\mathcal{N}_j|}{N} + \frac{1}{N}\sum_{u=1}^{N}(1 - \alpha_{uj})\underline{\mathbf{1}}\{u \in \mathcal{N}_j\} & k = j \\ \frac{\alpha_{jk}}{N} & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

Applying the definition of Φ, we see that the relation $W = I - \eta L$ holds.

We now aim to prove the second part of the theorem. The fact that $\mathbf{1}$ is both a right and left eigenvector is done similarly to the proof of Lemma 3.1. We now aim to prove that the spectral radius is less than 1.

Let $\lambda_1 \leq \cdots \leq \lambda_N$ be the ranked eigenvalues of the graph Laplacian with corresponding eigenvectors $\mathbf{v}_i$. By properties of the graph Laplacian, we have $0 \leq \lambda_i \leq 2d$ for $d = \max_{u \in V}\{|\mathcal{N}_j|\}$ the maximum degree of a node. In particular, we have $\lambda_1 = 0$ with $\mathbf{v}_1 = \mathbf{1}$ by properties of the graph Laplacian. We also note that by connectivity, $0 < \lambda_2$ and hence by orthogonality of eigenspaces we have $\mathbf{v}_i^T\mathbf{1} = 0\ \forall 2 \leq i \leq N$.

Now for arbitrary i:

$$(W - J)\mathbf{v}_i = (I - \eta L - J)\mathbf{v}_i = \mathbf{v}_i - \eta\lambda_i\mathbf{v}_i - J\mathbf{v_i} \tag{28}$$

By applying the above, we hence conclude that the eigenvalues of $W - J$ are given by $\{0\} \cup \{1 - \eta\lambda_i : 2 \leq i \leq N\}$. The 0 case is trivial, and by applying bounds on λ_i we obtain the full inequality. □

3.4 Convergence of Algorithm

3.4.1 Theoretical Results

We note that by the proof of 2.3, we can immediately conclude the above algorithm converges in expectation. Hence, we immediately have the following.

Lemma 3.3.

$$\mathbb{E}\{\lim_{t \to \infty} x(t)\} = x_{avg}\mathbf{1} \tag{29}$$

In order to show convergence in second moment, we again aim to show the generalised broadcast gossip algorithm satisfies 2.6 using a similar technique, but using a secondary graph.

Lemma 3.4. *The eigenvalues of* $\mathbb{E}\{W(t)^T W(t)\}$ *are bounded above by 1 and bellow by 0.*

Proof. First assume that node u is broadcasting. Then we have:

$$[W^{(u)^T}W^{(u)}]_{jk} = \begin{cases} 1 + \sum_{l=1}^{N} \alpha_{ul}^2 \mathbf{1}\{l \in \mathcal{N}_u\} & k = j = u \\ \alpha_{uk}(1-\alpha_{uk}) & k \in \mathcal{N}_u;\ j = u \\ \alpha_{uj}(1-\alpha_{uj}) & k = u;\ j \in \mathcal{N}_u \\ (1-\alpha_{uj})^2 & k \in \mathcal{N}_u;\ j = k \\ 1 & j \notin \mathcal{N}_u;\ j = k \neq u \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

Then, taking the expectation with respect to u we obtain:

$$[\mathbb{E}\{W(t)^T W(t)\}]_{jk} = \begin{cases} 1 - \frac{2}{N}\sum_{l=1}^{N} \alpha_{jl}(1-\alpha_{jl})\mathbf{1}\{l \in \mathcal{N}_j\} & k = j \\ \frac{2}{N}\alpha_{jk}(1-\alpha_{jk}) & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

By defining a second graph $G_2 = (V, E, w_2)$ where for $\{i, j\} \in E$ we define $w_2(\{i, j\}) = \alpha_{ij}(1-\alpha_{ij})$ we note that the above matrix can be written as $I - \frac{2}{N}L_2$ where L_2 is the weighted Laplacian matrix of the graph G_2.

We now apply Gershgorin Circle Theorem to bound the eigenvalues. First we note that $R_i = \frac{2}{N}\sum_{l=1}^{N} \alpha_{jl}(1-\alpha_{jl})\mathbf{1}\{l \in \mathcal{N}_j\}$ and hence we conclude the following:

$$\lambda_i \in \bigcup_{j \in V} \overline{B([\mathbb{E}\{W(t)^T W(t)\}]_{jj}, R_j)} \subseteq [0, 1] \tag{32}$$

Noting that both the bound above and bellow can be obtained by considering the discs. We denote $B(a, r)$ as the open disc centred at a with radius r. $\square$

Theorem 3.5. *The general broadcast gossip algorithm on weighted graphs converges in second moment.*

Proof. In order to show the above we show condition 2.6. By the Rayleigh-Ritz Theorem we have by denoting $W' = \mathbb{E}\{W(t)^T W(t)\}$ and $W'_J = \mathbb{E}\{W(t)^T J W(t)\}$:

$$\lambda_1 = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T W' \mathbf{x} - \mathbf{x}^T W'_J \mathbf{x} \tag{33}$$

and hence we aim to apply the previous lemma in order to show that the maximum value is strictly less than 1.

We note that

$$\lambda_1 \leq \max_{\|\mathbf{x}\|=1} \mathbf{x}^T W' \mathbf{x} - \min_{\|\mathbf{x}\|=1} \mathbf{x}^T W'_J \mathbf{x} \tag{34}$$

The first term is bounded by the previous lemma by 1 and the second term is bounded bellow by 0 and hence we obtain a bound of 1. We aim to show this bound is not obtained.

Firstly note that by properties of the weighted Laplacian and connectivity, we have hat W' is a non-negative irreducible matrix. We can then apply Perron-Frobenious and say that 1 is an eigenvalue of W' (multiplicity 1) with eigenvector $1/\sqrt{N}\,\mathbf{1}$ and any other eigenvector $\mathbf{x}_i$ has eigenvalues $0 \leq \mu_i < 1$. We also note that W'_J is positive semi-definite and hence $\mathbf{x}^T W'_J \mathbf{x} \geq 0$.

Hence, we consider the eigenvector $1/\sqrt{N}\,\mathbf{1}$ we have:

$$\mathbf{x}^T W' \mathbf{x} - \mathbf{x}^T W'_J \mathbf{x} = 1 - \frac{1}{N} < 1 \tag{35}$$

where we also used the fact that $\mathbf{1}$ is a right eigenvector of $W(t)$.

This shows that the maximum bound of 1 is never obtained as maximising the first term turns the inequality in (34) into a strictly less than. For any other $\mathbf{x}$, we have a value < 1 subtracting a value ≥ 0 which results also in a strict bound. This concludes the proof. $\square$

Note that since we have convergence in expectation as well as convergence in second moment, we can conlude convergence almost surely.

Theorem 3.6. *The generalised broadcast gossip algorithm on weighted graphs converges to consensus almost surely i.e.*

$$\mathbb{P}(\lim_{t\to\infty} x(t) = c\mathbf{1}) = 1 \tag{36}$$

for some $c \in \mathbb{R}$ *where* $\mathbb{E}(c) = x_{avg}$

3.4.2 Numerical Results

We can confirm our theoretical derivations with some numerical studies. In this case, we consider an explicit example for the weight of graphs.

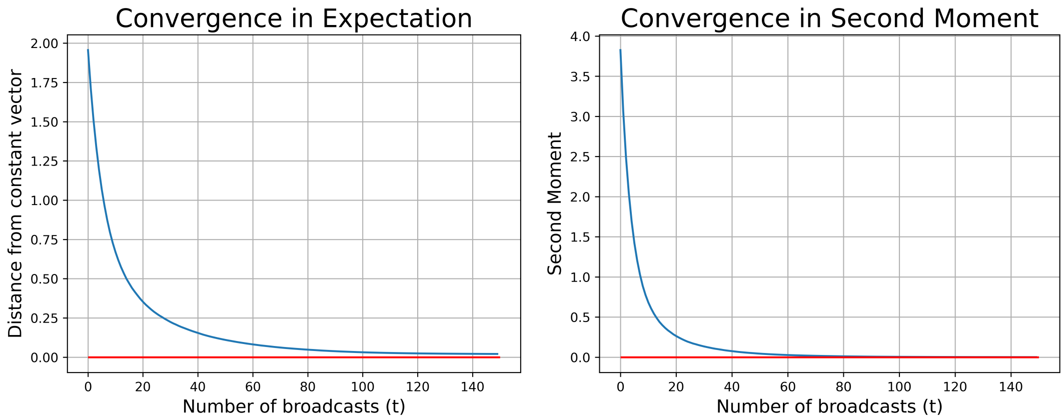

Figure 1: Figures above shows convergence in expectation and second moment for broadcast gossip algorithms on weighted graphs. 1000 simulations were performed when computing the expectation is required. The graph is generated using 2.1 with $N = 50$ and $R = 2\sqrt{\log(N)/N}$. The mixing parameter is $\gamma = 50$ and the weights as well as the initial values of nodes are outlined in 3.4.2. The red lines show $y = 0$ which is the value we expect to be obtaining in the long run.

We generate an original unweighted graph like in 2.1, however we define the weight of the edges as the following:

Definition 3.7. *For $\{i, j\} \in E$ let $d < R$ be the euclidean distance between the nodes in the unit square. Then, we define $w(\{i, j\}) = 1 - \frac{d}{R}$*

Physically speaking, this is a linear decrease based on euclidean distance up to the radius R. Hence, nodes which are closer together would have a stronger impact on each other whilst nodes which are further away will have smaller impacts. We also generate $x_i(0)$ for all $i \in V$ as independent uniform distribution on $(0, 1)$. We compute the distance away from constant vector as $\|\mathbb{E}\{x(t)\} - x_{avg}\mathbf{1}\|$ at each time t. We compute the second moment as $\mathbb{E}\{\|\beta(t)\|_2^2\}$. This, along with other parameters described in 1, produced the figures above.

The above figures demonstrate convergence in expectation as well as convergence in second moment from a numerical perspective, coinciding with the theoretical results.

We also later study the communication cost of the above model, comparing it to both the original model as well as the later models. This is done in section 5.

4 Weighted Graph Broadcasting with Stochastic Mixing Parameter

Based on our work in section 3, we further generalise by generalising the mixing parameter. Previously, the mixing parameter is a constant to be determined prior to the algorithm being performed. However, we aim to generalise this mixing parameter to follow some distribution and study if convergence almost surely still holds in this context.

We employ graph model 3.2 and time model 3.1 and update the algorithm.

4.1 Independent Mixing Parameter

Suppose node $u \in V$ ticks at time $t \in \mathbb{Z}_+$. Then, we compute the vector $x(t+1) = W(t)x(t)$ using the following random matrix:

$$W_{jk}^{(u)} = \begin{cases} 1 & j \notin \mathcal{N}_u,\, k = j \\ \alpha_{uj} & j \in \mathcal{N}_u,\, k = u \\ 1 - \alpha_{uj} & j \in \mathcal{N}_u,\, k = j \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

where $\alpha_{uv} = \Gamma \cdot f \circ w(\{u, v\})$ for $f : \mathbb{R}_+ \to (0, 1]$ and Γ is some random variable s.t. $\mathbb{P}(\Gamma \in (0, 1)) = 1$ i.e. Γ is bounded above by 1 and bellow by 0 almost surely. We will also assume that Γ is independent to the broadcasting node u.

By the same method as 3, we note that we can assume without loss of generality we have that the weights of the graph are bounded above by 1, and hence remove the requirement of the normalising function f. Hence, we will be working with the definition $\alpha_{uv} = \Gamma \cdot w(\{u, v\})$ instead and note that any other case can be reduced to this one due to the normalising function.

Following this framework we see again that by definition, $\mathbf{1}$ is still a right eigenvector of $W^{(u)}$ for all u as the rows of the matrix sum to 1. We can now study convergence in the case of independence.

Firstly we note that $\mathbb{E}\{\Gamma\} = \gamma$ is well defined as Γ has bounded support. We define $\mathbb{E}\{W(t)\}$ to be the expectation with respect to the product probability measure from the discrete uniform distribution I_k and random variable Γ. Then we claim the following:

Lemma 4.1. *$W = \mathbb{E}\{W(t)\}$ satisfies the following:*

$$W\mathbf{1} = \mathbf{1},\, \mathbf{1}^T W = \mathbf{1}^T,\, \rho(W - J) < 1 \tag{38}$$

for $\rho(\cdot)$ the spectral radius of a matrix, $J = N^{-1}\mathbf{1}\mathbf{1}^T$

Proof. Firstly note that since $W(t)$ has non-negative entries, by Fubini's theorem, we can compute W by first computing the expectation with respect to u and then with respect to Γ.

By the same procedure as in the previous sections (3.2), we find that the relation $\mathbb{E}\{W(t) \,|\, \Gamma = a\} = I - \eta L$ for $\eta = \frac{a}{N}$ to still be true. Hence, applying linearity of expectation we obtain that $W = I - \eta L$ for $\eta = \frac{\gamma}{N}$ to be true.

By noting that since Γ is almost surely in $(0, 1)$ we also have that $\gamma \in (0, 1)$ and hence by the same arguments as those outlined in 3.2, we conclude the proof. □

From this we can immediately conclude convergence in expectation similar to previous sections.

We now again aim to show convergence in second moment, as these are the two parts required for almost sure convergence. To do so, we again show the following intermediate lemma which will prove useful for the final result.

Lemma 4.2. *The eigenvalues of $\mathbb{E}\{W(t)\}$ are bounded above by 1 and bellow by 0.*

Proof. We employ the same technique as the previous proof, applying Fubini's theorem to first take the expectation with respect to Γ and then with respect to u. By taking this approach, we find that:

$$[\mathbb{E}\{W(t)^T W(t)\}]_{jk} = \begin{cases} 1 - \frac{2}{N} \sum_{l=1}^{N} \gamma w_{jl}(1 - \gamma w_{jl}) \mathbf{1}\{l \in \mathcal{N}_j\} & k = j \\ \frac{2}{N} \gamma w_{jk}(1 - \gamma w_{jk}) & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

By defining a second graph $G_2 = (V, E, w_2)$ where for $\{i, j\} \in E$ we define $w_2(\{i, j\}) = \gamma w_{jk}(1 - \gamma w_{jk})$ then the above matrix can be written as $I - \frac{2}{N} L_2$ for L_2 the weighted Laplacian matrix of the graph G_2.

Now by the same methods as the proof of lemma 3.4, we can apply Gershgorin Circle Theorem to compute the rest of the proof. □

Now with an identical proof to the theorem 3.5, we can show the following theorem.

Theorem 4.3. *The general broadcast gossip algorithm on weighted graphs with an independent stochastic mixing parameter converges in second moment.*

Corollary 4.3.1. *The general broadcast gossip algorithm on weighted graphs with an independent stochastic mixing parameter converges almost surely i.e.*

$$\mathbb{P}(\lim_{t \to \infty} x(t) = c\mathbf{1}) = 1 \tag{40}$$

for some $c \in \mathbb{R}$ *where* $\mathbb{E}\{c\} = x_{avg}$.

4.2 Dependent Mixing Parameter

We consider the case where the mixing parameter is dependent on the broadcasting node. This is a relatively simple case as its methods are almost identical to the independent case. We first formalise how this dependence is defined and then briefly explain why convergence still holds in this context.

For any $u \in V$ we consider a random variable Γ_u such that $\mathbb{P}(\Gamma_u \in (0, 1)) = 1$. We denote $\gamma_u = \mathbb{E}\{\Gamma_u\} \in (0, 1)$. Then suppose node u is broadcasting at time t. Then we have $x(t+1) = W^{(u)} x(t)$ with

$$W_{jk}^{(u)} = \begin{cases} 1 & j \notin \mathcal{N}_u,\ k = j \\ \alpha_{uj} & j \in \mathcal{N}_u,\ k = u \\ 1 - \alpha_{uj} & j \in \mathcal{N}_u,\ k = j \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

where $\alpha_{uj} = \Gamma_u \cdot w(\{u, j\})$ where we've normalised such that $w(\{u, v\}) \in (0, 1]$ for any node $u, v \in V$.

The dependence with respect to u may represent how persuasive an individual is when applied in the context of gossiping. More generally, it can be interpreted as how influential or important a node is compared to the other nodes. Note that the deterministic case is a special case of the above setup.

Again we note that by definition, $\mathbf{1}$ is a right eigenvector of $W(t)$ with probability 1. Now we aim to derive conditions for convergence in expectation. We note that this is not generally true because by [5], convergence in expectation is true if and only if $W = \mathbb{E}\{W(t)\}$ has left and right eigenvector $\mathbf{1}$ and the spectral radius $\rho(W - J) < 1$. However this is not generally true.

Example 4.1. *For the case* $N = 2$ *consider the graph* $G = (V, E, w)$ *where* $V = \{1, 2\}$, $E = \{\{1, 2\}\}$ *and* $w(\{1, 2\}) = 1$ *and* 0 *otherwise. This is just a graph of two nodes with a single edge connecting them of weight 1. Now define* Γ *such that* $\Gamma | \{\textit{node } 1 \textit{ is broadcasting}\} \sim U(0, 2/3)$ *and* $\Gamma | \{\textit{node } 2 \textit{ is broadcasting}\} \sim U(0, 1)$ *then* Γ *satisfies the required properties.*

Following this setup, we can apply Fubini's theorem to compute $\mathbb{E}\{W(t)\}$:

$$\mathbb{E}\{W(t)\} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{6} & \frac{5}{6} \end{pmatrix} \tag{42}$$

and the columns clearly do not sum to 1, showing that in this case, the broadcast gossip algorithm does not converge in expectation.

This behaviour can be physically interpreted as follows. As some nodes become more important/influential than others, there is no reason to expect the entire group of people converges in expectation to the initial average. The more influential nodes will have a weighted impact on the result and hence convergence in expectation to the desired value occurs only when a careful balance is achieved.

Since we already have the if and only if requirement for convergence in expectation, we can then derive the conditions such that convergence in expectation is satisfied.

Proposition 4.4. *Defining $W = \mathbb{E}\{W(t)\}\}$, we claim the broadcast gossip algorithm with stochastic mixing parameter converges in expectation if and only if the following set linear equation (in γ_u) is satisfied:*

$$\sum_{u=1}^{N} \gamma_u [W]_{uk} \mathbf{1}\{u \in \mathcal{N}_k\} = \gamma_k \sum_{u=1}^{N} [W]_{uk} \mathbf{1}\{u \in \mathcal{N}_k\} \quad \forall k \in V \tag{43}$$

Proof. Convergence in expectation is true if and only if $\mathbf{1}$ is both a left and a right eigenvector of W and the spectral radius of $W - J$ is strictly less than 1. We show the above properties by first computing W. By applying Fubini's theorem, we first take the expectation fixing the broadcasting node u and then compute the final matrix. From this, we obtain:

$$[W]_{jk} = \begin{cases} 1 - N^{-1} \sum_{u=1}^{N} \gamma_u w_{uj} \mathbf{1}\{u \in \mathcal{N}_J\} & k = j \\ N^{-1} \gamma_k w_{jk} & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

We first note that the above matrix has right eigenvector $\mathbf{1}$ with eigenvalue 1. Hence the first condition is satisfied.

We now aim to show that $\mathbf{1}$ to be a right eigenvector of W if and only if the above condition is true. To do this, we apply a sequence of equalities:

$$[\mathbf{1}^T W]_k = \sum_{j=1}^{N} [W]_{jk} = 1 - \frac{1}{N} \sum_{u=1}^{N} [\gamma_u w_{uk} \mathbf{1}\{u \in \mathcal{N}_k\}] + \frac{1}{N} \sum_{u=1}^{N} [\gamma_k w_{uk} \mathbf{1}\{u \in \mathcal{N}_k\}] \tag{45}$$

which is equal to 1 if and only if

$$\sum_{u=1}^{N} \gamma_u [W]_{uk} \mathbf{1}\{u \in \mathcal{N}_k\} = \gamma_k \sum_{u=1}^{N} [W]_{uk} \mathbf{1}\{u \in \mathcal{N}_k\} \tag{46}$$

In the following we aim to show that the spectral radius is also strictly bounded above by 1. We now assume the above property is true and $\mathbf{1}$ is a left eigenvector of W.

To do this, we first note that W is irreducible as its associated directed graph $G_A = (V, E_A)$ where $E_A = \{(i,j) : \{i,j\} \in E\}$ is strongly connected as we have G is connected. Furthermore, we note that the periodicity of W is 1 as the diagonal entries of W are already positive. Hence, we can apply Perron-Frobenius theorem to W.

Since we've shown that $\mathbf{1}$ is a right eigenvector, we know that the spectral radius of W is 1 and by periodicity, -1 is not an eigenvalue of W and 1 has geometric multiplicity 1. Say W has l linearly independent eigenvectors, we can denote them v_i, $i \in \{1, ..., l\}$. Then we let $v_1 = \mathbf{1}$ without loss of generality and obtain:

$$(W - J)v_i = Wv_i - Jv_i = \begin{cases} 0 & i = 1 \\ \lambda_i v_i & \text{otherwise} \end{cases} \tag{47}$$

hence we know that $\{v_1, ...v_l\}$ is a linearly independent family of eigenvectors of $W - J$.

Now we aim to show that W and $W - J$ has the same number of linearly independent eigenvectors. In order to do this, we write $\mathbb{R}^N = \text{span}\{\mathbf{1}\} \oplus U$ a direct sum of two vector spaces with $U = \{x \in \mathbb{R}^N : \mathbf{1}^T x = 0\}$. This subspace is W-invariant due to the fact that $\mathbf{1}$ is a left eigenvector of W.

Let v be an eigenvector of $W - J$ with eignevalue λ. Then $v = \alpha\mathbf{1} + w$ for $w \in U$. Then:

$$\begin{aligned} (W - J)v &= Wv - Jv \\ &= W(\alpha\mathbf{1} + w) - J(\alpha\mathbf{1} + w) \\ &= \alpha\mathbf{1} + Ww - \alpha\mathbf{1} \\ &= Ww = \lambda v = \lambda\alpha\mathbf{1} + \lambda w \end{aligned}$$

where we used orthogonality of w and $\mathbf{1}$ to eliminate the term Jw.

We now note that due to W-invariance, we have that either $\alpha = 0$ or $w = 0$ the zero vector. In the case that $\alpha = 0$ we have that $v = w$ is also an eigenvector of W with eigenvalue λ. If $w = 0$ then the eigenvector is some constant multiple of $\mathbf{1}$ which we know is also an eigenvector of W.

Hence, we've shown that any eigenvector of $W - J$ is also an eigenvector of W. We have also shown the converse when bounding the eigenvalues. This then shows that λ is an eigenvalue of W if and only if it is an eigenvalue of $W - J$. Hence, we've shown that the spectral radius of $W - J$ must be bounded by 1. $\square$

We now move on to proving convergence in second moment. From physical intuition, it makes intuitive sense that convergence in second moment still holds due to the fact that introducing dependence may only have an impact on the rate of convergence in second moment. We prove this in the following:

Theorem 4.5. *The general broadcast gossip algorithm on weighted graphs with a dependent stochastic mixing parameter converges in second moment to a consensus.*

Proof. Firstly, let $\gamma = \mathbb{E}\{\Gamma\}$ and $K = \mathbb{E}\{\Gamma^2\}$. We note that both exists due to Γ being bounded almost surely. Then, we obtain the following by applying Fubini's theorem to $W(t)^T W(t)$ noting that the entries are positive:

$$[\mathbb{E}\{W(t)^T W(t)\}]_{jk} = \begin{cases} 1 - \frac{2}{N}\sum_{l=1}^{N}\left(\gamma w_{jl} - K w_{jl}^2\right)\mathbf{1}\{l \in \mathcal{N}_j\} & k = j \\ \frac{2}{N}\left(\gamma w_{jl} - K w_{jl}^2\right) & k \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \tag{48}$$

We note that since $\Gamma \in (0,1)$ almost surely, we have that $\Gamma^2 < \Gamma$ almost surely which implied $K < \gamma$. This along with $w_{jk} \leq w_{jk}^2$ means that we can again apply Gershgorin's bound for eigenvalues finding that $\mathbb{E}\{W(t)^T W(t)\}$ has eigenvalues bounded by 1 and 0.

The remainder of the proof follows like that in theorem 3.5 and hence we obtain convergence in second moment. $\square$

This, in combination with convergence in expectation give the following result:

Theorem 4.6. *The general broadcast gossip algorithm on weighted graphs with stochastic mixing parameters dependent on the broadcasting node converges almost surely to a consensus if the following holds:*

$$\sum_{u=1}^{N}\gamma_u [W]_{uk}\mathbf{1}\{u \in \mathcal{N}_k\} = \gamma_k \sum_{u=1}^{N}[W]_{uk}\mathbf{1}\{u \in \mathcal{N}_k\} \quad \forall k \in V \tag{49}$$

4.3 Numerical Results

We now aim to numerically study the convergence of the models discussed in the current section. To do so, we work with two explicit examples. We employ the same method for generating weights to the graph as in section 3.4.2.

In the independent case, we let $\Gamma \sim \text{Unif}\,(0,1)$ a uniform distribution. We do this such that $\mathbb{E}\{\Gamma\} = 1/2$ which aligns with the constant case studied previously.

In the dependent case, we again fall back onto physical intuition. We want more important nodes to have a larger influence whereas less important nodes to have a smaller influence. Hence, we require a method for measuring importance in a graph. We do this by using a normalised version of the pagerank centrality. Let p_u be the pagerank centrality of node $u \in V$ then we normalise this by defining $\hat{p}_u = p_u / \max_v\{p_v\}$ the normalised pagerank centrality. We then define Γ such that $\Gamma|\{\text{node } u \text{ is broadcasting}\} \sim \text{Unif}(0, \hat{p}_u)$ to be the mixing parameter. This, along with the other characteristics outlined in the captions of the figures produced figures 2 and 3.

We note that the y-axis are outlined in 3.4.2.

The figures demonstrate what we expect from theory, with convergence in expectation only being true for the independent case whereas convergence in second moment is achieved in both cases.

It is important to note, however, that convergence in second moment does imply that the broadcast gossip algorithm converges almost surely to a consensus. However, the value of consensus obtained is not in expectation equal to the average of the initial opinions. Physically we can interpret this as under broadcasting, consensus is obtained throughout the population almost surely. However, if there exists more "influential" people who have a stronger impact than others, then it may lead to a drive away from the true average of the population before any broadcasting has happened.

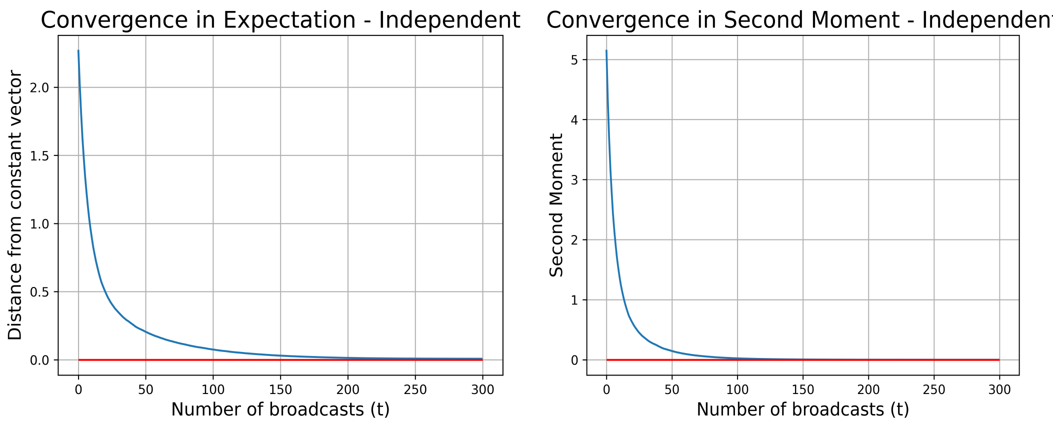

Figure 2: Figures above shows convergence in expectation and second moment for broadcast gossip algorithms on weighted graphs with independent stochastic mixing parameters. 1000 simulations were performed when computing the expectation is required. The graph is generated using 2.1 with $N = 50$ and $R = 2\sqrt{\log(N)/N}$. Weights and mixing parameters are outlined in 4.3. The red lines show $y = 0$.

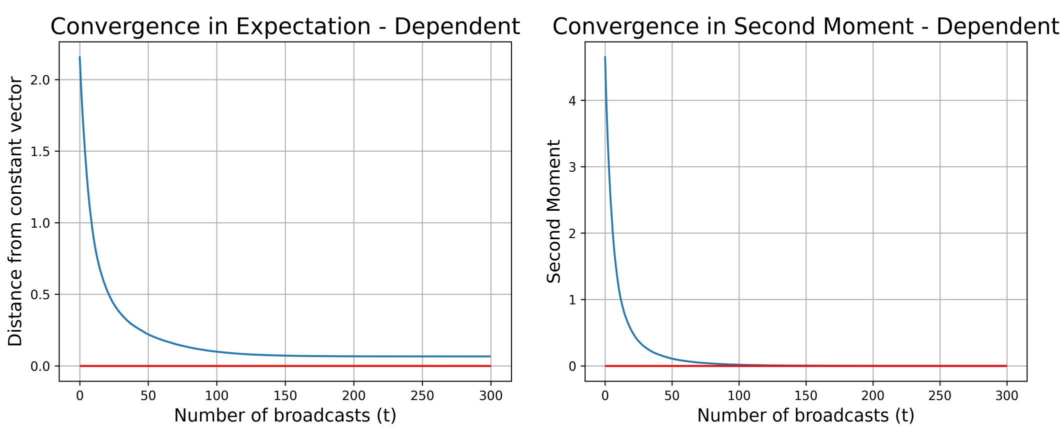

Figure 3: Figures above shows convergence in expectation and second moment for broadcast gossip algorithms on weighted graphs with dependent stochastic mixing parameters. 1000 simulations were performed when computing the expectation is required. The graph is generated using 2.1 with $N = 50$ and $R = 2\sqrt{\log(N)/N}$. Weights and mixing parameters are outlined in 4.3. The red lines show $y = 0$.

5 Communication Cost

The original reasoning or introducing broadcast gossip algorithms was that not only did they demonstrate almost sure convergence, they also converged quicker than other existing models [1]. However, we note that in the generalisation considered here there is no reason to expect a lower communication cost compared to the original model. This is due to the fact that in the generalisation, we enforced that the weights of the graph are bounded by 1. Hence, with the same mixing parameters, nodes will have less influence then that in the original model, meaning intuitively that information is shared lower hence a larger communication cost.

We explore this numerically, using per-node variance as second moment becomes too small too quickly causing floating-point error to occur too early before any patterns are spotted. We define the per-node variance as $N^{-1}\|x(t) - Jx(t)\|$ and the figure bellow is produced as an average of 100 simulations.

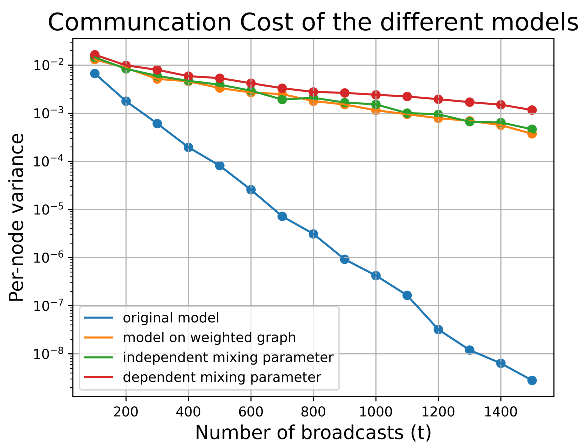

Figure 4: The figure above shows the per-node variance at the end of a k step broadcast. The graph used is outlined in 2.1 with $N = 50$ and $R = \sqrt{\log(N)/N}$. The original model has $\gamma = 1/2$. The model for weighted graphs is identical to that in 3.4.2. The independent and dependent case are exactly as outlined in 4.3.

As shown in figure 4, the original model has much lower communication cost compared to the remaining models. This aligns with previous intuition of more "efficient" communications when the weights of the graph are uniformly 1. We also note that the dependent mixing parameter model is slightly more expensive than the other two models again due to less effective communication, as for the constant and the independent case, we have that $\mathbb{E}\{\Gamma\} = \gamma = 1/2$ whereas in the dependent case, $\mathbb{E}\{\Gamma\} < 1/2$.

A Program Used for Numerical Studies

The program used to generate numerical results can be found here: https://github.com/John-Zou-IC/IROP_Broadcast

B Bound for Eigenvalues of Graph Laplacian

The following proof is partially taken from the stack exchange conversation linked here accessed 17th of July 2025.

Lemma B.1. *Given a graph G with weights $w_{ij} \in [0, 1]$ we have for λ_i eigenvalues of the graph laplacian L: $0 \leq \lambda_i \leq 2 * d_{max}$ for d_{max} the maximum degree of a node in the graph.*

Proof. We split the proof into two parts. We first show that $\lambda_i \geq 0$. Note this follows from L being positive semi-definite. This is reliant on the fact that $L = \sum_{(i,j)\in E} w_{i,j}(e_i - e_j)(e_i - e_j)^T$. From this we can conclude for any $x \in \mathbb{R}^N$:

$$\begin{aligned} x^T L x &= x^T \left(\sum_{(i,j)\in E} w_{i,j}(e_i - e_j)(e_i - e_j)^T \right) x \\ &= \sum_{(i,j)\in E} w_{i,j}\, x^T(e_i - e_j)(e_i - e_j)^T x \\ &= \sum_{(i,j)\in E} w_{i,j}\, (x_i - x_j)(x_i - x_j) \geq 0 \end{aligned}$$

Hence, we've shown L is positive semi-definite and hence the first inequality follows.

We now begin to show that the eigenvalues are also bounded above. In order to do this, we use Gershgorin Circle Theorem. To apply the theorem, first we compute the R_i's:

$$R_i = \sum_{j\neq i} |[L]_{i,j}| = \sum_{j\neq i} w_{i,j} = \alpha_i \tag{50}$$

where we denote α_i as the sum of the weights. From here, applying the Gershgorin Circle Theorem, we have $\forall i$:

$$\lambda_i \in \bigcup_j \overline{B}(\alpha_j, R_j) = \bigcup_j \overline{B}(\alpha_j, \alpha_j) \tag{51}$$

We note that $B(r, r) \subset B(t, t)$ whenever $r \leq t$. Hence we can conclude for $\alpha = \max_j\{\alpha_j\}$ we have $\lambda_i \in \overline{B}(\alpha, \alpha)$ and therefore $\lambda_i \leq 2\alpha \leq 2 * d_{max}$ noting that α is bounded above by maximum degree due to weights being bounded above by 1. □

C Stochastic Ordering

During my studies I also attended a reading seminar. The reading done was [3]. I found it fascinating that under certain orders, the set of random variables forms a type of partially ordered set, specifically a lattice. More importantly, I found the existence of such an ordering to be non-trivial as it was something I had not considered before. Historically it was often the case that when I was presented with multiple random variables, they represented distinct events which may be dependent on each other. However, stochastic ordering has wide applications in queueing theory and finance comparing response times or risks. Hence, I spend this section of the appendix defining the two orders encountered and describing briefly how they lead to a lattice structure.

Before we begin, I should note that if you intend on reading the original literature, it has some rather weird notation on the naturals and integers which is covered in the chapter "Note". It is just before the first chapter and worth reading before skipping over.

We employ the idea of a survival function later on when stating the final result. This is defined as follows: For X a random variable with distribution F, its survival function denoted $\bar{F}$ is given by $1 - F$.

Definition C.1. *Let X, Y be two random variable such that*

$$\mathbb{P}(X > x) \leq \mathbb{P}(Y > x) \quad \forall x \in (-\infty, \infty) \tag{52}$$

then we say X is smaller than Y in the usual stochastic order denoted $X \leq_{st} Y$.

Note that since every closed interval is a countable intersection of open intervals, we can alter the above definition to become $\geq$ rather than $>$ which sometimes may be easier to check.

Denoting $=_{\text{st}}$ as equal in distribution, we show the following lemma.

Lemma C.2. *Two random variables X and Y satisfy $X \leq_{st} Y$, if and only if, there exists two random variables $\hat{X}$ and $\hat{Y}$ defined on the same probability space such that:*

$$\begin{aligned} \hat{X} &=_{st} X, \\ \hat{Y} &=_{st} Y, \\ \mathbb{P}(\hat{X} \leq \hat{Y}) &= 1 \end{aligned}$$

Proof. ($\Rightarrow$) Suppose $X \leq_{\text{st}} Y$. Let F and G be the distribution functions of X and Y respectively, and we define F^{-1} and G^{-1} to be the right-continuous inverse of said functions i.e. $F(u)^{-1} = \sup\{x : F(x) \leq u\}$.

Then, by the usual stochastic ordering, we have $1 - F(x) \leq 1 - G(x)$ for any x. Hence, we conclude that $F^{-1}(u) \leq G^{-1}(u)$ for any u. Then, if we let $U \sim \text{Unif}(0, 1)$ then we can define $\hat{X} = F^{-1}(U)$ and $\hat{Y} = G^{-1}(U)$ which completes the proof.

($\Leftarrow$) This follows immediately from the conditions given. □

As a direct result of the above lemma, we have the following theorem and obtains the desired result:

Theorem C.3. *Let X be a random variable and let ϕ_1 and ϕ_2 be two functions such that $\phi_1(x) \leq \phi_2(x)$ for all $x \in \mathbb{R}$. Then we have $\phi_1(X) \leq_{st} \phi_2(X)$.*

In particular, if ϕ is a function such that $x \leq [\geq]\phi(x)$ for all $x \in \mathbb{R}$ then $X \leq_{st} [\geq_{st}]\phi(X)$.

Corollary C.3.1. *The set of all distribution functions on $\mathbb{R}$ is a lattice with respect to the usual stochastic order. That is, if X and Y are random variables with distributions F and G, then there exist random variables Z and W such that $Z \leq_{st} X$, $Z \leq_{st} Y$, $X \leq_{st} W$ and $Y \leq_{st} W$. Explicitly, Z has survival function $\min\{\bar{F}, \bar{G}\}$ and W has survival function $\max\{\bar{F}, \bar{G}\}$.*

The same conclusion can be obtained on the hazard rate order which is discussed later in the literature.

References

[1] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *Signal Processing, IEEE Transactions on*, 57:2748 – 2761, 08 2009.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[3] M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer New York, New York, NY, 1st ed. 2007. edition, 2007.

[4] J. N. Tsitsiklis. *Problems in decentralized decision making and computation*. PhD thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., Boston, MA, 11 1984.

[5] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. In *42nd IEEE International Conference on Decision and Control*, volume 5, pages 4997–5002 Vol.5, Dec. 2003.