Stochastic Activation based Broadcast Push-Sum for Distributed Estimation

Yue Bi[*†], Philippe Ciblat[†], Yue Wu[*], Cunqing Hua[*]
[*]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China
{wuyue,cqhua}@sjtu.edu.cn
[†] LTCI, Telecom Paris, IP Paris, 91120 Palaiseau, France, {bi,ciblat}@telecom-paris.fr

***Abstract*—Distributed estimation systems enable nodes to estimate a target parameter in a collaborative manner. These systems are useful in sensor networks or distributed machine learning. Here, we explore distributed estimation in graph-connected networks without a fusion center, where nodes exchange information with neighbors to estimate this target parameter synchronously. Due to packet collision, there is a tradeoff between the number of exchanges and the quality of these exchanges. To fix this issue, we propose to activate the nodes randomly. The main contribution of the paper is to determine an activation rate offering a good target estimation quality as fast as possible.**

***Index Terms*—Distributed estimation, stochastic activation, consensus.**

I. Introduction

We focus on systems without a fusion center, where sensors are connected by an undirected, self-loop-free graph. Each node observes a noisy version of the target parameter θ and exchanges data with its neighbors to improve its estimation performance. The objective of each node is to collect sufficient information on θ spread through the other nodes. If the samples depend linearly on θ, the objective for all nodes is to achieve an average consensus, corresponding to the mean of the collected samples.

To reach this average consensus, two node activation strategies exist: *i)* the asynchronous strategy, where the activation is random and in practice restricted to a single sensor or, at most, a pair of sensors at a given time. With this strategy, there is no packet collision but the number of exchanges per unit of time is inherently small which leads to a slow convergence rate. The main advantage is the absence of coordination between nodes for synchronization. In that setting, algorithms like *Random Gossip* [1], [2] and *Broadcast Gossip* [3] have been developed and extensively studied over some imperfect configurations, such as directed graphs [4], [5], link failures [6], and unstable sensors [7]. The broadcast gossip leverages the broadcast nature of wireless channels but even if it converges to a consensus, the consensus does not correspond to the sample average. To overcome this issue, a *Push-Sum* approach has been developed in that wireless context by [8] inspired by [9]. Extensions to scenarios with link failures have been explored in [10]. Analyses of the convergence rate of this Push-Sum approach have been conducted in [11], [12]. *ii)* the synchronous strategy where the activation is synchronous and may be coordinated, for instance, through Global Positioning System (GPS). In that case, we assume nodes operating in full-duplex mode. During each time slot, all nodes simultaneously transmit and receive, which may lead to packet collisions at the receiver side even with a multi-user detector. Therefore, an activation policy is required to balance data transmission and collision risk.

This work has been supported by National Key R&D Program of China under Grant No 2023YFB2704903 and No 2020YFB1807504, and ERC Grant "CTOCom".

The goal of the paper is thus to provide a stochastic node activation mode in a synchronous strategy that minimizes the Mean Square Error (MSE) of the target parameter as quickly as possible. To achieve this, we propose applying Push-Sum approach along with an independent identically-distributed Bernoulli activation distribution process with an common activation rate of $(1-\gamma)$. We also propose a simple way to design γ by establishing a link between the theoretical convergence rate (given in [11]) and γ. Numerical results show the effectiveness of the proposed stochastic activation principle and the designed γ. They also show that using an averaged optimized γ, which depends only on the stochastic properties of the graph rather than on individual graph realizations, performs comparably to a graph-dependent γ.

The rest of this paper is organized as follows: the system model is introduced on Section II. Push-Sum approach with stochastic activation is proposed in Section III. The design of γ based on the convergence rate of the Push-Sum is given in Section IV. Numerical results are in Section V. Concluding remarks are drawn in Section VI.

II. System model

We consider a network of K full-duplex sensors modeled as an undirected connected simple graph $\mathcal{G}=(\mathcal{K},\mathcal{E})$ where $\mathcal{K}$ is the set of sensors, and $\mathcal{E}$ is the set of links between sensors. Let $\mathcal{K}_k$ denote the set of neighbors of sensor k. We also define $\mathbf{A}$ as the *adjacency matrix* of the graph $\mathcal{G}$.

a) Node activation: the wireless network operates synchronously with a subset of nodes active at each time slot, communicating only with neighbors. We define the *activation vector* $\mathbf{c}_t \in \{0,1\}^K$ at time t as

$$\mathbf{c}_t \triangleq (c_{1,t}, c_{2,t}, \dots, c_{K,t})^\intercal, \tag{1}$$

where $c_{k,t}=1$ if sensor k is activated at time t, and 0 otherwise. Superscript $(.)^\intercal$ stands for matrix transposition. We denote by $\tilde{\mathcal{K}}_{k,t} \subseteq \mathcal{K}_k$ the set of the activated neighbors of sensor k at time t, and $\tilde{K}_{k,t} \triangleq |\tilde{\mathcal{K}}_{k,t}|$ its cardinality.

b) Communication failure: receiving nodes may fail to decode the packets sent by its neighbors. It means that the multi-packets detector did not succeed sometimes to recover each packet. We define the *collision-free vector* $\mathbf{s}_t \in \{0,1\}^K$ at time t as

$$\mathbf{s}_t \triangleq (s_{1,t}, s_{2,t}, \ldots, s_{K,t})^\intercal, \tag{2}$$

where $s_{k,t} = 1$ if sensor k has correctly decoded all the packets, i.e., is collision-free, and 0 otherwise. We assume that $s_{k,t}$ follows a independent (over k and t) Bernoulli distribution with success probability $(1 - p_{k,t})$. We model the failure probability $p_{k,t}$ as an increasing function of the number of active neighbors of sensor k at time t. Therefore we get

$$p_{k,t} = 1 - f(\tilde{K}_{k,t}), \tag{3}$$

where $f : \mathbb{N} \mapsto [0,1]$ is a monotonically decreasing function. The exact form of f depends on the physical layer protocol and is beyond the scope of this work.

We also assume that a sensor not succeeding to decode its received packets broadcasts a negative-acknowledgment (NACK) message to all its neighboring nodes which become aware of the communication failure. We denote $\mathcal{K}'_{k,t} \subseteq \mathcal{K}_k$ the set of collision-free nodes neighboring to node k at time t, and $K'_{k,t} = |\mathcal{K}'_{k,t}|$.

c) Sensor observations and estimations: we consider each sensor gets one noisy sample linearly dependent of the target parameter θ as follows

$$y_k = \theta + \varepsilon_k \tag{4}$$

where ε_k is i.i.d. and follows a zero-mean Gaussian distribution with variance σ^2.

In centralized setting, according to Eq. (4), the minimum variance unbiased estimate is

$$x_{avg} = \frac{1}{K}\sum_{k=1}^{K} y_k.$$

The objective of the distributed estimation algorithm is to share x_{avg} at each node after convergence. This algorithm is iterative and we denote by $x_{k,t}$ the shared information (equivalently, the estimate of θ) at sensor k and time/iteration t. Obviously, we have $x_{k,0} = y_k$. Consequently, we force the distributed algorithm to satisfy both properties:

- (**P1**) *Average conservation*: the average $x_{avg,t} = (1/K)\sum_{k=1}^{K} x_{k,t}$ at each time is equal to $x_{avg,0} = x_{avg}$.
- (**P2**) *Consensus*: after convergence, each sensor shares the same value x_c. Due to **P1**, it can be proven that $x_c = x_{avg}$. Consequently, after convergence, each sensor knows the best centralized estimate.

We define the MSE at time t as

$$\mathrm{MSE}_t \triangleq \mathbb{E}\left[\frac{1}{K}\left\|\mathbf{x}_t - \theta\cdot\mathbf{1}\right\|^2\right], \tag{5}$$

where $\mathbf{x}_t = (x_{1,t}, \ldots, x_{K,t})^\intercal$ and $\mathbf{1} = (1, \ldots, 1)^\intercal$.

According to **P1** and **P2** as well as the model given by Eq. (4), one can easily prove that

$$\mathrm{MSE}_t = \frac{1}{K}\mathrm{MSE}'_t + \frac{\sigma^2}{K} \tag{6}$$

where $\mathrm{MSE}'_t = \mathbb{E}[\|\mathbf{x}_t - x_{avg}.\mathbf{1}\|^2]$, (σ^2/K) comes from $\mathbb{E}[|x_{avg} - \theta|^2]$ and the cross-term vanishes since $\sum_{k=1}^{K}(x_{k,t} - x_{avg}) = 0$.

III. Broadcast Push-Sum approach with stochastic activation

The goal of this section is to exhibit a distributed algorithm satisfying **P1** and **P2** even in presence of packet collision.

One approach is to combine linearly the received samples at each sensor, i.e., $\mathbf{x}_{t+1} = \mathbf{G}_t \cdot \mathbf{x}_t$ where $\mathbf{G}_t$ depends on the chosen algorithm. In that case, **P1** and **P2** are satisfied if and only if $\mathbf{G}_t$ is doubly-stochastic at any time t. Due to the collision, $\mathbf{G}_t$ may loose its initial double-stochastic property. To overcome this issue, the nodes involved in the collision should be removed in advance to work on a subgraph during one iteration. But it is impossible since the collision is not known in advance. Retransmission may be done but during a specific iteration and this additional iteration requires a global coordination which is not advocated. Therefore algorithms based on $\mathbf{x}_{t+1} = \mathbf{G}_t \cdot \mathbf{x}_t$ are not appropriate.

Actually, the appropriate approach is the so-called *Push-Sum* introduced in [8], [9]. These algorithms update two variables per node: $v_{k,t}$ (initialized by $x_{k,0}$) and $w_{k,t}$ (initialized by 1). The variable related to the average $x_{k,t}$ is obtained by dividing $v_{k,t}$ by $w_{k,t}$, assuming $w_{k,t} \neq 0$. Therefore, the update process can be expressed as

$$\mathbf{v}_{t+1} = \mathbf{G}_t \cdot \mathbf{v}_t, \quad \mathbf{w}_{t+1} = \mathbf{G}_t \cdot \mathbf{w}_t, \quad \mathbf{x}_{t+1} = \mathbf{v}_{t+1}/\mathbf{w}_{t+1},$$

where $\mathbf{G}_t$ is the update matrix, $\mathbf{v}_t = (v_{1,t}, v_{2,t}, \cdots, v_{K,t})^\intercal$, $\mathbf{w}_t = (w_{1,t}, w_{2,t}, \cdots, w_{K,t})^\intercal$, and division is elementwise.

To ensure **P1** and **P2**, the update matrix $\mathbf{G}_t$ in the Push-Sum approach have to meet the following conditions [13]: (**C1**) they are column-stochastic, and have strictly positive diagonal entries, (**C2**) they are chosen through an i.i.d. process, and (**C3**) the average update matrix $\mathbb{E}[\mathbf{G}_t]$ is primitive.

Now, we will design $\mathbf{G}_t$ adapted to the synchronous strategy with random collisions and stochastic activation. For the stochastic activation scheme, we consider an i.i.d. (over k and t) Bernoulli process with activation probability $(1-\gamma)$. Notice that γ is the same for any node for the sake of simplicity. At time 0, we set $v_{k,0} = y_k = x_{k,0}$ and $w_{k,0} = 1$. At time t,

- with probability $(1 - p_{k,t})$, node k updates its data as

$$v_{k,t+1} = \sum_{\ell \in \tilde{\mathcal{K}}_{k,t} \cup \{k\}} h_{k,\ell,t} \cdot v_{\ell,t}, \tag{7}$$

$$w_{k,t+1} = \sum_{\ell \in \tilde{\mathcal{K}}_{k,t} \cup \{k\}} h_{k,\ell,t} \cdot w_{\ell,t}. \tag{8}$$

- with probability $p_{k,t}$ (which corresponds to collision event), node k updates its data as

$$v_{k,t+1} = h_{k,k,t} \cdot v_{k,t}, \quad w_{k,t+1} = h_{k,k,t} \cdot w_{k,t}. \tag{9}$$

where, for a given $\ell \in [K]$ and $k \in \mathcal{K}'_{\ell,t}$, $h_{k,\ell,t}$ is defined as

$$h_{k,\ell,t} = \begin{cases} \dfrac{1}{K_\ell + 1} & \text{if } k \neq \ell, \\ \dfrac{K_k - K'_{k,t} + 1}{K_k + 1} & \text{if } k = \ell, \text{node } k \text{ activated}, \\ 1 & \text{if } k = \ell, \text{node } k \text{ inactivated}. \end{cases} \tag{10}$$

Then, the entries of $\mathbf{G}_t$, denoted by $g_{k,\ell,t}$, are equal to $h_{k,\ell,t}$ if this last term is involved in Eqs (7)-(9). It is easy to check that **C1** holds, that **C2** is satisfied since activation and collision processes are i.i.d., and that **C3** is satisfied since the support of $\mathbb{E}[\mathbf{G}]$ is those of $(\mathbf{I} + \mathbf{A})$ where $\mathbf{I}$ is the identity matrix.

IV. Activation Rate Design

The goal of this section is to characterize γ with respect of the properties of the graph and the proposed algorithm in Section III. Based on Eq. (6), the MSE_t is directly connected to MSE'_t. Therefore, hereafter, we exhibit a mathematical relationship between γ and MSE'_t.

In [11], it is proven that, for t large enough, $\log \mathrm{MSE}'_t$ is well approximated by an upper bound described by a decreasing affine function whose the absolute value of the slope (in the rest of the paper, called "slope") is given by

$$\omega \triangleq -\log\left(\rho\left(\mathbb{E}\left[\mathbf{G}_t \otimes \mathbf{G}_t\right] \cdot (\mathbf{J}^\perp \otimes \mathbf{J}^\perp)\right)\right), \tag{11}$$

where $\rho(\cdot)$ is the spectral radius, and $\mathbf{J}^\perp \triangleq \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal$. Consequently, we would like to design γ as the maximum of ω. For this purpose, we derive ω in closed-form partially by establishing an expression for $\mathbb{E}\left[\mathbf{G}_t \otimes \mathbf{G}_t\right]$.

For the sake of simplicity, we now omit the index t as all matrices since they are i.i.d.. The entries of $\mathbb{E}\left[\mathbf{G}_t \otimes \mathbf{G}_t\right]$ are $G_{k,\ell,k',\ell'} \triangleq \mathbb{E}\left[g_{k,\ell} \cdot g_{k',\ell'}\right]$. We have

$$g_{k,\ell} = \begin{cases} \dfrac{\tilde{a}_{k,\ell}}{K_\ell + 1} & \text{if } k \neq \ell, \\ 1 - \dfrac{\sum_{k_1=1}^K \tilde{a}_{k_1,\ell}}{K_\ell + 1} & \text{if } k = \ell, \end{cases} \tag{12}$$

with $\tilde{a}_{k,\ell} = s_k a_{k,\ell} c_\ell$. For the expression of $G_{k,\ell,k,\ell'}$ (shortened in G), we have three subcases:

- Case 1: $k \neq \ell$ and $k' \neq \ell'$.

$$G = \frac{1}{(K_\ell + 1)(K_{\ell'} + 1)} \mathbb{E}\left[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}\right]. \tag{13}$$

- Case 2: $k = \ell$ and $k' \neq \ell'$ (or $k \neq \ell$ and $k' = \ell'$).

$$G = \frac{1}{K_{\ell'} + 1} \mathbb{E}\left[\tilde{a}_{k',\ell'}\right] - \frac{1}{(K_\ell + 1)(K_{\ell'} + 1)} \sum_{k_1=1}^K \mathbb{E}\left[\tilde{a}_{k_1,\ell} \cdot \tilde{a}_{k',\ell'}\right]. \tag{14}$$

- Case 3: $k = \ell$ and $k' = \ell'$.

$$G = 1 - \frac{1}{K_\ell + 1} \sum_{k_1=1}^K \mathbb{E}\left[\tilde{a}_{k_1,\ell}\right] - \frac{1}{K_{\ell'} + 1} \sum_{k'_1=1}^K \mathbb{E}\left[\tilde{a}_{k'_1,\ell'}\right] + \frac{1}{(K_\ell + 1)(K_{\ell'} + 1)} \sum_{k_1=1}^K \sum_{k'_1=1}^K \mathbb{E}\left[\tilde{a}_{k_1,\ell} \cdot \tilde{a}_{k'_1,\ell'}\right]. \tag{15}$$

According to Eqs. (13)-(15), we need to calculate $\mathbb{E}[\tilde{a}_{k,\ell}]$ and $\mathbb{E}[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}]$.

Let us begin with the term $\mathbb{E}[\tilde{a}_{k,\ell}]$. For a given pair of $(k,\ell) \in [K]^2$, we have

$$\mathbb{E}[\tilde{a}_{k,\ell}] = a_{k,\ell} \mathbb{E}\left[s_k | c_\ell = 1\right] \mathbb{P}(c_\ell = 1). \tag{16}$$

We recall that s_k follows Bernoulli distribution with a success probability depending on the number of active neighbors $\tilde{K}_k$ and the function defined in Eq. (3). We only calculate $\mathbb{E}\left[s_k | c_\ell = 1\right]$ for the case where $a_{k,\ell} = 1$ as $\mathbb{E}[\tilde{a}_{k,\ell}]$ is non-zero only when $a_{k,\ell} = 1$. We thus obtain

$$\begin{aligned} \mathbb{E}[\tilde{a}_{k,\ell}] &= a_{k,\ell}(1-\gamma) \\ &\quad \cdot \sum_{m=1}^{K_k} \mathbb{E}\left[s_k | c_\ell = 1, \tilde{K}_k = m\right] \mathbb{P}(\tilde{K}_k = m | c_\ell = 1) \\ &= a_{k,\ell} \sum_{m=1}^{K_k} \binom{K_k - 1}{m-1} f(m)(1-\gamma)^m \gamma^{(K_k - m)}. \end{aligned} \tag{17}$$

The summation starts with 1 as node ℓ is assumed to be active, and nodes k and ℓ are connected. The random variable $(s_k | c_\ell = 1, \tilde{K}_k = m)$ thus follows Bernoulli distribution with success probability $f(m)$. We also use the fact that for $\ell \in [K]$, the activation of node ℓ follows Bernoulli distribution with success probability $1-\gamma$, making the random variable $(\tilde{K}_k | c_\ell = 1)$ binomial-distributed $\mathcal{B}(K_k, 1-\gamma)$, i.e. for $1 \leq m \leq K_k$,

$$\mathbb{P}(\tilde{K}_k = m | c_\ell = 1) = \binom{K_k - 1}{m-1} (1-\gamma)^{(m-1)} \gamma^{(K_k - m)}.$$

Let us continue with the term $\mathbb{E}[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}]$. For two pairs $(k,\ell) \in [K]^2$ and $(k',\ell') \in [K]^2$, we have

$$\mathbb{E}[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}] = a_{k,\ell} a_{k',\ell'} \mathbb{E}\left[s_k s_{k'} | c_\ell c_{\ell'} = 1\right] \mathbb{P}(c_\ell c_{\ell'} = 1).$$

Since s_k and $s_{k'}$ are independent when $k \neq k'$, and c_ℓ and $c_{\ell'}$ are independent when $\ell \neq \ell'$, we have

$$\mathbb{E}[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}] = \begin{cases} a_{k,\ell} a_{k',\ell'} \mathbb{E}[s_k | c_\ell c_{\ell'} = 1] \mathbb{E}[s_{k'} | c_\ell c_{\ell'} = 1] (1-\gamma)^2, \\ \qquad\qquad \text{if } k \neq k', \ell \neq \ell', \\ a_{k,\ell} a_{k,\ell'} \mathbb{E}[s_k | c_\ell c_{\ell'} = 1] (1-\gamma)^2, \text{if } k = k', \ell \neq \ell', \\ a_{k,\ell} a_{k',\ell} \mathbb{E}[s_k | c_\ell = 1] \mathbb{E}[s_{k'} | c_\ell = 1] (1-\gamma), \\ \qquad\qquad \text{if } k \neq k', \ell = \ell', \\ a_{k,\ell} \mathbb{E}[s_k | c_\ell = 1] (1-\gamma), \qquad \text{if } k = k', \ell = \ell'. \end{cases} \tag{18}$$

As $\mathbb{E}[s_k | c_\ell = 1]$ is calculated in (17), we only need to calculate $\mathbb{E}[s_k | c_\ell c_{\ell'} = 1]$ for the case $a_{k,\ell} = 1$ and $a_{k',\ell'} = 1$. We first consider the case that $k \neq k'$ and $\ell \neq \ell'$. If node k and ℓ' are not connected, i.e. $a_{k,\ell'} = 0$, we obtain

$$\mathbb{E}[s_k | c_\ell c_{\ell'} = 1] = \mathbb{E}[s_k | c_\ell = 1]. \tag{19}$$

If $a_{k,\ell'} = 1$, node k has at least two active neighbors, then

$$\mathbb{E}[s_k | c_\ell c_{\ell'} = 1] = \sum_{m=2}^{K_k} \binom{K_k - 2}{m-2} f(m)(1-\gamma)^{(m-2)} \gamma^{(K_k - m)}. \tag{20}$$

Therefore, when $k = k'$ and $\ell \neq \ell'$, we have $a_{k,\ell'} = a_{k',\ell'} = 1$, so Eq. (20) still holds. By combining both above-mentioned cases, we obtain the following result.

$$
\mathbb{E}[\tilde{a}_{k,\ell} \cdot \tilde{a}_{k',\ell'}] =
\begin{cases}
a_{k,\ell} a_{k',\ell'} F_{1+a_{k,\ell'}}(K_k) F_{1+a_{k',\ell}}(K_{k'})(1-\gamma)^2 \\
\qquad\qquad\qquad\qquad \text{if } k \neq k', \ell \neq \ell', \\
a_{k,\ell} F_1(K_k)(1-\gamma), \qquad \text{if } k = k', \ell = \ell', \\
a_{k,\ell} a_{k',\ell} F_1(K_k) F_1(K_{k'})(1-\gamma), \\
\qquad\qquad\qquad\qquad \text{if } k \neq k', \ell = \ell', \\
a_{k,\ell} a_{k,\ell'} F_2(K_k)(1-\gamma)^2, \quad \text{if } k = k', \ell \neq \ell'.
\end{cases}
\tag{21}
$$

where

$$
F_i(K) \triangleq \sum_{m=i}^{K} \binom{K-i}{m-i} f(m)(1-\gamma)^{(m-i)} \gamma^{(K-m)}. \tag{22}
$$

Finally, the term $\mathbb{E}[\mathbf{G}_t \otimes \mathbf{G}_t]$ can be obtained by putting Eq. (17) and Eqs. (21)-(22) into Eqs. (13)-(15). The term γ may be now obtained by calculating numerically the spectral radius of a matrix whose the closed-form expressions have been provided above.

V. Numerical Results

The considered underlying graphs are Random Geographical Graphs (RGG) [13], which are generated as follows: first, we select K points uniformly within the unit square $[0,1] \times [0,1]$ to represent the locations of the sensors. Then, we connect any two sensors with an undirected edge if they are within a specified radius r, which corresponds to the communication range.

Simulations are conducted with $K = 10$ nodes and $r = 0.4$. We generate 100 connected RGGs. On each graph, we run 2000 distributed estimation algorithm with the same $\mathbf{x}_0$. This operation is repeated 100 times by changing $\mathbf{x}_0$. We fix the target parameter $\theta = 2$ and the noise variance $\sigma^2 = 1$. For the function of communication success probability in Eq. (3), we choose $f(\bullet) = e^{-\alpha \max\{\bullet - 1, 0\}}$. It means that we consider an exponential decay with respect to the size of the neighbors in terms of performance. But perfect transmission is assumed when a node has at most one active neighbor since $f(1) = 1$. We evaluate the following metrics per underlying graph:

- **Slope**: obtained by doing a linear regression on $t \mapsto \log(\|\mathbf{x}_t - x_{avg} \cdot \mathbf{1}\|^2)$ by keeping these values below -2 when these values are obtained for one underlying graph but tested for the 100 initializations with 2000 realizations of the algorithm each.
- $\boldsymbol{\omega}$: obtained by numerically evaluating Eq. (11) through Eqs.(17) and (21) for one underlying graph.
- **Empirical** $\boldsymbol{\omega}$: obtained by numerically evaluating Eq. (11) by replacing the expectation with the empirical mean of $\mathbf{G}_t \otimes \mathbf{G}_t$.

In Fig 1, we plot the average over all the considered underlying graphs of the slope, ω and the empirical ω versus γ for $\alpha = 0.5$ (left) and $\alpha = 1$ (right). The curves of ω and the empirical ω well coincide validating our derivations

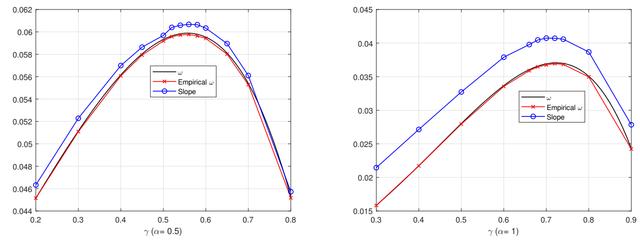

Fig. 1: Average slope, ω and empirical ω versus γ for $\alpha = 0.5$ (left) and $\alpha = 1$ (right).

of $\mathbb{E}[\mathbf{G} \otimes \mathbf{G}]$. We observe that ω has a big and similar impact of all the metrics. Therefore, we propose to select γ maximizing ω for a given underlying graph and a specific α. The maximization can be done through a simple 1-D search.

In Fig 2, we plot MSE'_t versus t for different γ with $\alpha = 0.5$ (top) and $\alpha = 1$ (bottom). The “Optimal γ” curve is obtained by maximizing ω for each graph. The ”Average Optimal γ” curve is obtained by maximizing the average ω over all the graphs. Both curves are of interest compared to arbitrary choices for γ. The “Average Optimal γ” albeit offering a loss compared to the best one is of great interest since it only depends on the statistics of the graph model and it is more practical in a distributed context.

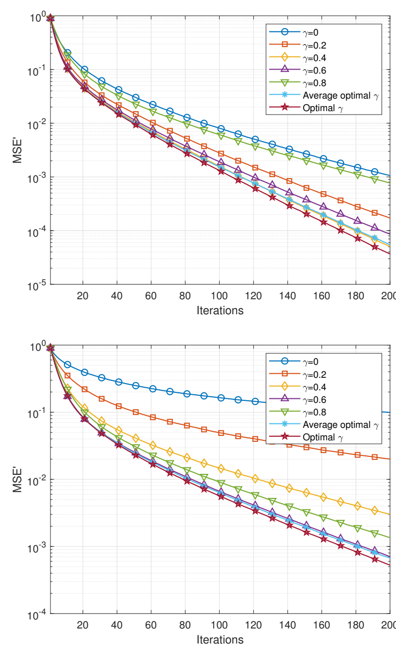

Fig. 2: MSE'_t vs t for different γ with $\alpha = 0.5$ (top) and $\alpha = 1$ (bottom).

VI. Conclusion

This paper focuses a consensus algorithm with stochastic activation scheme to avoid node collisions in synchronous setup. the activation rate is obtained mathematically and depends on the graph statistics. Extension to more complex schemes (rate per node, etc) are for future works.

REFERENCES

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Analysis and optimization of randomized gossip algorithms," in *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 5. IEEE, 2004, pp. 5310–5315.

[2] ——, "Randomized gossip algorithms," *IEEE transactions on information theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[3] 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, 2009.

[4] S. Wu and M. G. Rabbat, "Broadcast gossip algorithms for consensus on strongly connected digraphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3959–3971, 2013.

[5] D. Silvestre, J. P. Hespanha, and C. Silvestre, "Broadcast and gossip stochastic average consensus algorithms in directed topologies," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 2, pp. 474–486, Jun. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8361902/?arnumber=8361902

[6] S. Kar and J. M. 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, 2008.

[7] H. Wang, X. Liao, Z. Wang, T. Huang, and G. Chen, "Distributed parameter estimation in unreliable sensor networks via broadcast gossip algorithms," *Neural Networks*, vol. 73, pp. 1–9, 2016.

[8] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *2010 IEEE International Symposium on Information Theory*, Jun. 2010, pp. 1753–1757, iSSN: 2157-8117. [Online]. Available: https://ieeexplore.ieee.org/document/5513273/?arnumber=5513273

[9] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* Cambridge, MA, USA: IEEE Computer. Soc, 2003, pp. 482–491. [Online]. Available: http://ieeexplore.ieee.org/document/1238221/

[10] B. Gerencsér and J. M. Hendrickx, "Push-sum with transmission failures," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1019–1033, Mar. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8359298

[11] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, pp. 2802–2814, 2013.

[12] B. Gerencsér and L. Gerencsér, "Tight bounds on the convergence rate of generalized ratio consensus algorithms," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1669–1684, 2021.

[13] F. Iutzeler, "Distributed estimation and optimization in asynchronous networks," Ph.D. dissertation, Télécom ParisTech, 2013.