2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Large network consensus is robust to packet losses and interferences

Paolo Frasca Julien M. Hendrickx

Abstract— **In many randomized consensus algorithms, the constraint of average preservation may not be enforced at every time step, resulting in an error between the average of the initial conditions and the current average. We have recently shown that under mild conditions on the distribution of the update matrices, the mean square error has an upper bound inversely proportional to the size of the network. In this work, we consider the case of consensus with packet losses and interferences. Using an extension of our results taking correlations into account, we show that the MSE induced by losses and interferences can be estimated by such a bound: hence we argue that larger networks are naturally more robust, in terms of accuracy, to packet losses and interferences. Our results hold for general networks, without restrictive assumptions on its topology.**

I. Introduction

Robustness against link failures, either temporary or permanent, has been recognized as a key feature in evaluating networked control systems. This problem has been formalized in several ways in the context of coordination and consensus algorithms, leading to both probabilistic and worst-case approaches to switching networks, since the seminal work in [14]. In this paper, we consider an average consensus algorithm running on a fixed communication network whose links may for various reasons occasionally fail to transmit the information among the nodes. It will be clear that link failures and consequent message losses prevent the consensus algorithm from converging –as required– to the exact average of the initial conditions. We analyze the expected deviation from the average, providing explicit bounds which depend on certain features of the network and on the specific algorithm at hand. These estimates in particular imply that the error decays to zero when the number of nodes grows.

Contribution and relation with previous works

We consider two models of lossy communication in consensus systems. In Section III we consider a standard linear consensus algorithm on a network whose links are affected by packet losses. The occurrence of a loss is assumed to be independent among links: this modeling approach is taken from [7], although our analysis includes more general update rules. In Section IV we consider an algorithm involving information broadcast from random nodes, along the lines of [1], [3]. Such a communication model has been originally proposed in [5] to study the effect of interferences on consensus problems. Closely related issues are studied in [4], [13], [16].

The two above models show important differences. Compared with the former, the latter framework leads to more correlation between the updates, resulting in a more complicated analysis. In both cases, we show that the mean square deviation can be explicitly bounded by a quantity which is inversely proportional to the size of the network. Hence, we argue that the error induced by losses vanishes as the network grows larger. Our results are based on the application and the extension of a general technique, presented in Section II and recently developed by the authors in the analysis of related classes of randomized consensus algorithms [9]. An extension is indeed needed to treat the correlations occurring in the model between the losses in the interference model studied in Section IV.

Related results, leading to the same favorable conclusion on the robustness of some of the systems we consider, were proved in [7] (for independent losses) and in [5] (for interference losses), but only under restrictive assumptions on the topology of the network. These papers assume indeed certain symmetries in the network structure, namely that the network is the Cayley graph of an Abelian group: this assumption leads to results which are insightful but may not be directly applied to many networks of interest. We refer the reader to [10] for a discussion on the meaning of this assumption. In contrast, our results hold irrespective of the topology, and are thus of immediate and general application.

Note that in this work, noise in communication is modeled by the loss of data packets. Other modeling approaches are available in the literature, and have been applied to consensus: without giving a complete review, we mention data rate limitations [8], [12], additive analog noise [17], and bit erasures [2]. Finally, we leave out of the scope of this paper any issue related to conditions for convergence to consensus and estimates of the rate of convergence: indeed, there is a solid literature on these problems, including [6], [7], [11], [15], to which we refer the interested reader.

Notation and preliminaries

We will use the notion of (weighted directed) graph, which we define as a pair $G = (I, A)$, where I is a finite set

Paolo Frasca is with the Dipartimento di Scienze Matematiche, Politecnico di Torino, corso Duca degli Abruzzi 24, 10126 Torino, Italy; paolo.frasca@polito.it. His work was partly supported by MIUR under grant PRIN-20087W5P2K.

Julien M. Hendrickx is with the ICTEAM Institute at the Université catholique de Louvain, Avenue Georges Lemaitre 4, B-1348 Louvain-la-Neuve, Belgium; julien.hendrickx@uclouvain.be. His work is supported by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian Science Policy Office. The research is also supported by the Concerted Research Action (ARC) Large Graphs and Networks of the French Community of Belgium. The scientific responsibility rests with its authors.

The authors wish to thank Prof. Sandro Zampieri for pointing out the interest of the packet loss application.

whose elements are called *nodes* and $A \in \mathbb{R}^{I\times I}$ is a matrix with nonnegative entries. Resorting to more standard graph-theoretic jargon, we may equivalently think of an implicit edge set $E = \{(i,j) \in I \times I \,:\, A_{ij} > 0\}$ and say that i is connected to j when $A_{ij} > 0$. We assume that the graphs have no self-loops, and that $A_{ii} = 0$ for every $i \in I$. Given a nonnegative matrix A, we can define an associated Laplacian matrix $L(A) \in \mathbb{R}^{I\times I}$ as the matrix such that $[L(A)]_{ij} = -A_{ij}$ if $i \neq j$ and $[L(A)]_{ii} = \sum_{j:j\neq i} A_{ij}$. We also note that the map $A \mapsto L(A)$ is a linear operator. Observe that $L(A)+L(A)^*$ is positive semidefinite and that $L(A)\mathbf{1} = 0$, provided we denote by $\mathbf{1}$ the vector of suitable size whose components are all 1. If A is stochastic matrix, that is, $\sum_j A_{ij} = 1$, then $L(A) = I - A$. Assume now $\mathbf{1}^* L(A) = \mathbf{1}^*$. Then, there holds $x^*(L(A)+L(A)^*)x = \sum_{i,j} a_{ij}(x_j - x_i)^2$. If additionally we denote $\alpha = \max_i \sum_j A_{ij}$, one can show that $L(A)^*L(A) \leq \alpha(L(A)+L(A)^*)$ –see [9, Eq. (5)]– and that

$$L(A\cdot A) + L(A\cdot A)^* \leq \alpha\left(L(A)+L(A)^*\right),$$

where the symbol $\cdot$ is used to denote entry-wise product of matrices, and the inequality is to be understood as the semi-definite positiveness of the difference.

II. Randomized averaging algorithms

Given a set of nodes I of finite cardinality N, we consider a distributed state $x(t) \in \mathbb{R}^I$ evolving according to a stochastic discrete-time system of the form

$$x_i(t+1) = \sum_{j\in I} a_{ij}(t)x_j(t) \quad \text{for all } i \in I, \quad t \in \mathbb{Z}_{\geq 0}, \quad (1)$$

where the $a_{ij}(t) \geq 0$ satisfy $\sum_{\ell\in I} a_{i\ell}(t) = 1$ for all $t \geq 0$, and the matrices $A(t)$ defined by $[A(t)]_{ij} = a_{ij}(t)$ are assumed to be independent and identically distributed random variables. System (1) is run with the goal for the state of each node to provide a good estimate of the initial average $\frac{1}{N}\sum_{i\in I} x_i(0)$. Note that $x(0)$ is unknown but given, and that all our results need to be valid for any $x(0) \in \mathbb{R}^I$. System (1) can also be conveniently rewritten as

$$x_i(t+1) = x_i(t) + \sum_{j\in I} a_{ij}(t)(x_j(t) - x_i(t)) \quad \forall\, i \in I, t \geq 0,$$

or in matrix form as

$$x(t+1) = x(t) - L(t)x(t) \qquad t \in \mathbb{Z}_{\geq 0}, \qquad (2)$$

where we remind the reader that $L_{ij}(t) = -a_{ij}(t)$ if $i \neq j$ and $L_{ii}(t) = \sum_{j\neq i} a_{ij}(t)$. Namely, $L(t) = I - A(t)$ is the Laplacian matrix of a weighted graph $(I, A(t))$.

Let $\bar{x}(t) = N^{-1}\sum_{i\in I} x_i(t)$. Our goal in this section is to study how large is the deviation between $\bar{x}(0)$ and $\bar{x}(t)$ for any $t \geq 0$, and in the limit for $t \to +\infty$. First of all, by conditioning at each time step upon the current state $x(t)$, we remark that

$$\begin{aligned}\mathbb{E}[\bar{x}(t+1)] =& \mathbb{E}[\mathbb{E}[\bar{x}(t+1)|x(t)]] \\ =& N^{-1}\mathbb{E}[\mathbf{1}^* x(t) - \mathbf{1}^*\mathbb{E}[L(t)]x(t)].\end{aligned}$$

Then, the average is preserved in expectation if and only if $\mathbf{1}^*\mathbb{E}[L(t)] = 0$. In view of this result, we restrict our attention to systems such that $\mathbf{1}^*\mathbb{E}[L(t)] = 0$, that is, systems whose expected update matrix $\mathbb{E}[A(t)]$ is doubly stochastic and thus average preserving. We are then left with the problem of studying the second-order moment of $\bar{x}(t)$, that is, the variance

$$\mathbb{E}[(\bar{x}(t) - \bar{x}(0))^2]. \qquad (3)$$

The following general result, proved in [9], provides conditions under which the increase of the deviation is bounded proportionally to the decrease of the disagreement.

Theorem 1 (Accuracy condition): Let x be an evolution of system (2), and denote

$$V(t) = \frac{1}{N}\sum_{i\in I}\left(x_i(t) - \bar{x}(t)\right)^2.$$

If $\mathbf{1}^*\mathbb{E}[L(t)] = 0$ and there exists $\gamma > 0$ such that

$$\mathbb{E}[L(t)^*\mathbf{1}\mathbf{1}^*L(t)] \leq \gamma\,\mathbb{E}[L(t) + L(t)^* - L(t)^*L(t)], \quad (4)$$

then for every $t \geq 0$, there holds

$$\mathbb{E}[(\bar{x}(t) - \bar{x}(0))^2] \leq \frac{\gamma}{\gamma + N}V(0).$$

If moreover the system converges to consensus, then

$$\mathbb{E}\left[(x_\infty - \bar{x}(0))^2\right] \leq \frac{\gamma}{\gamma + N}V(0).$$

In what follows, when the context prevents all ambiguities, we omit the explicit specification of the index set I in summations and the dependence on t of the coefficients $a_{ij}(t)$ and corresponding matrices.

The claims collected in the following Lemma will be useful to verify the condition of Theorem 1; their proof can be found in [9, Lemma 3].

Lemma 2 (Laplacian stochastic inequalities): Consider system (2) and let $a^{\mathrm{r}}_{\max}$ be a positive constant such that almost surely $\sum_{j:j\neq i} a_{ij} \leq a^{\mathrm{r}}_{\max}$ for all $i \in I$.

(i) If $\mathbf{1}^*\mathbb{E}(L) = 0$ and there exists $\beta > 0$ such that

$$\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L) \leq \beta\,\mathbb{E}(L + L^*),$$

then the condition of Theorem 1 holds for

$$\gamma = \frac{\beta}{1 - a^{\mathrm{r}}_{\max}}.$$

(ii) Assume that is a_{ij} and a_{kl} are uncorrelated when $i \neq k$. If $\mathbf{1}^*\mathbb{E}(L) = 0$, then the condition of Theorem 1 holds for

$$\gamma = \frac{a^{\mathrm{r}}_{\max}}{1 - a^{\mathrm{r}}_{\max}}.$$

The results above hold independently of the network topology. They are however mainly motivated by situations where the system converges to consensus, and convergence to consensus does depend on the network topology. In particular, convergence is guaranteed almost surely if and only if the graph corresponding to $\mathbb{E}(L)$ is connected.

III. SYNCHRONOUS CONSENSUS WITH PACKET LOSSES

In this section we analyze the effect of random temporary link failures, and consequent message losses, in classical linear consensus algorithms. In this purpose, we need to introduce the nominal consensus algorithm, model the phenomenon of packet losses, and propose an adaptation of the nominal consensus algorithm to lossy communication.

Let us start with a doubly stochastic consensus matrix Q adapted to the given graph $G = (I, E)$. This matrix defines a nominal discrete-time consensus algorithm

$$x(t+1) = Qx(t) \qquad t \geq 0,$$

which can be equivalently rewritten as

$$x(t+1) = x(t) - L(Q)x(t) \qquad t \geq 0,$$

where $L(Q)$ denotes the Laplacian associated to the stochastic matrix Q.

We then assume to model the unreliability of actual communication links by a stochastic process, according to which each link has an independent probability $(1-p)$ of failure at every time. More formally, we associate to each time $t \in \mathbb{Z}_{\geq 0}$ a matrix $S(t) \in \{0,1\}^{I \times I}$ such that the entries $S_{ij}(t)$ are independent across t, i, j, with $\mathbb{P}(S_{ij}(t) = 1) = p$ for all i, j, t. This loss process should be interpreted in the following way: if $S_{ij}(t) = 1$, then the link between i and j is active at time t; otherwise, the link is failing at time t, and the transmitted data are lost. Note that this definition is meaningful only if $(i,j) \in E$: otherwise, the values of S_{ij} have no effect.

A. A simple compensation rule

There is a natural and simple way to cope with the data loss described above. Let i be an agent: whenever she does not receive an expected incoming message, she uses her own value in place of the lost message. This compensation rule, which results in increasing the weight which is given to self information, is referred to as *biased compensation rule* in [7]. More general compensation rules will be studied later on. Under this compensation model, the resulting consensus algorithm is defined as

$$x(t+1) = x(t) - L(Q \cdot S(t))x(t) \qquad t \geq 0 \qquad (5)$$

where $\cdot$ denotes entrywise product. This evolution can clearly be framed in the context of Section II by writing $L(t) = L(Q \cdot S(t))$. The following Lemma, proved in the Appendix, shows how to compute the relevant moments of $L(t)$ for the purpose of applying Theorem 1.

Lemma 3: If $L(t) = L(Q \cdot S(t))$ as defined in (5), then

$$\begin{aligned}
\mathbb{E}[L(t)] &= pL(Q)\\
\mathbb{E}[L(t)^* L(t)] &= p^2 L(Q)^* L(Q)\\
&\quad + p(1-p)(L(Q \cdot Q) + L(Q \cdot Q)^*)\\
\mathbb{E}[L(t)^* \mathbf{1}\mathbf{1}^* L(t)] &= p(1-p)(L(Q \cdot Q) + L(Q \cdot Q)^*).
\end{aligned}$$

We now directly obtain the following result.

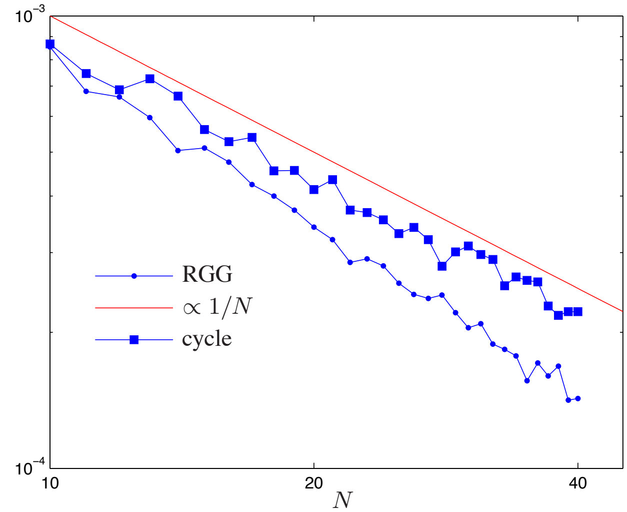

Fig. 1. Simulation of the mean square error (3) at convergence as a function of N for cycles and random geometric graphs. Rings are defined as (I, E) such that $I = \{0, \dots, N-1\}$ and $(i,j) \in E$ if $|i - j| = 1 \mod N$; random geometric graphs are defined by generating for each node a uniform random variable in the square $[0,1]^2$, and connecting two nodes if their Euclidean distance is below a constant times $\sqrt{\log(N)/N}$. For each graph, the nominal Q is defined with Metropolis weights $q_{ij} = (1 + \max\{d_i, d_j\})^{-1}$, where d_i and d_j are the degrees of i and j. Expectation is sampled by averaging over 1000 runs, assuming $p = 0.8$.

Proposition 4 (Biased compensation): The algorithm (5) satisfies the conditions of Theorem 1 with

$$\gamma = \frac{q}{1-q}(1-p),$$

where we denote $q = \max_i \sum_{j \neq i} Q_{ij}$.

Proof: Using Lemma 3, we see that γ must satisfy

$$\begin{aligned}
&p(1-p)(L(Q \cdot Q) + L(Q \cdot Q)^*) \leq\\
&\qquad \gamma\big(p(L(Q) + L(Q)^*) - p^2 L(Q)^* L(Q)\\
&\qquad\quad - p(1-p)\left(L(Q \cdot Q) + L(Q \cdot Q)^*\right)\big)
\end{aligned}$$

If, according to the properties stated in the Preliminaries, we notice that $L(Q)^* L(Q) \leq q(L(Q) + L(Q)^*)$ and $L(Q \cdot Q) + L(Q \cdot Q)^* \leq q\left(L(Q) + L(Q)^*\right)$, we see that a sufficient condition for γ is to satisfy

$$\big((1-p)q - \gamma(1 - qp - (1-p)q)\big)\left(L(Q) + L(Q)^*\right) \leq 0,$$

which holds when $\gamma \geq \frac{q(1-p)}{1-q}$ since $L(Q) + L(Q)^* \geq 0$. ■

Proposition 4 implies that the deviation is decreasing with the size N, and the decrease is at least as fast as N^{-1}. This fact is consistent with previous results on the topic [7], and is confirmed by simulations run on sequences of graphs with increasing N, as shown in Fig. 1. We also remark from simulations in Fig. 2 that the bound in Proposition 4 captures well the dependence of the deviation on the loss probability p. Note that the value q appearing in Proposition 4 is the maximal sum of weights given simultaneously by a node to all her neighbors. The denominator $1-q$ corresponds thus to the smallest weights that a node gives to herself, i.e. the smallest self-confidence.

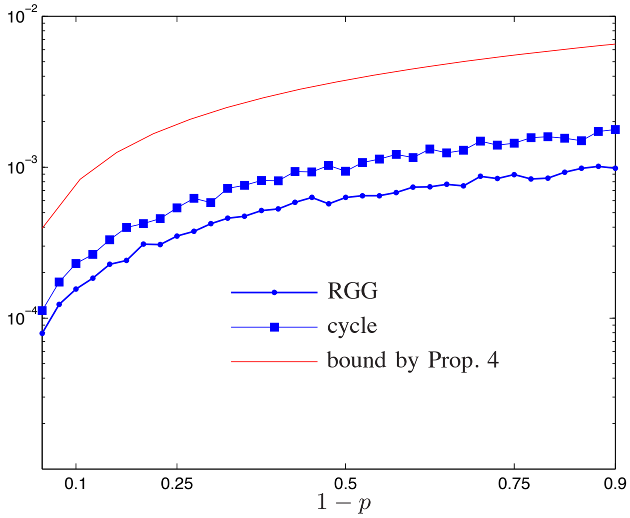

Fig. 2. Simulation of the mean square error (3) at convergence as a function of p on random geometric and cycle graphs with $N = 20$ and Metropolis weights ($q_{ij} = (1+\max\{d_i, d_j\})^{-1}$). The mean is sampled by averaging over 1000 realizations. These results suggest that the bound in Proposition 4 qualitatively captures how the mean square error depends on p.

B. Other compensation rules

We must note that the update rule proposed in (5) is by far not the only possible one. Indeed, the updating agent may compensate the data loss by using any combination of the available data (incoming messages and own state). An example is provided by the *balanced compensation rule* introduced in [7] as

$$L(t) = I - \text{diag}((S(t) \cdot Q)\mathbf{1})^{-1}(S(t) \cdot Q),$$

which distributes the weights of the nodes for which the information is missing on all other nodes, proportionally to their initial weights.

In this subsection, we consider the most general class of compensation rules coping with the loss process $S(t)$, which we describe as follows. Every agent i may choose any vector $(L_{ij})_{j\in I}$, provided it satisfies the following natural assumptions: $0 \le L_{ij}(t) \le 1$ if $i \ne j$; (ii) $L_{ij}(t) = 0$ if $S_{ij}(t)Q_{ij} = 0$ and $i \ne j$; and (iii) $\sum_j L_{ij}(t) = 0$. In order to analyze this class of rules, we may resort to the assumption of independence which we made on the underlying data loss process $S(t)$.

Proposition 5 (Any compensation): For any compensation rule against independent random losses satisfying the natural assumptions above and such that $\mathbf{1}^*\mathbb{E}[L(t)] = 0$ and $L_{ii}(t) \ge -1 + \varepsilon$ for every $i \in I$ and $t \ge 0$, the statement of Theorem 1 holds true with

$$\gamma = \frac{1-\varepsilon}{\varepsilon}.$$

Proof: In view of the independence of the losses on different links, the update rules which are taken by two different agents are statistically independent, that is, $L_{ij}(t)$ is independent from $L_{kl}(t)$ whenever $i \ne k$. Then, the result follows from Lemma 2(ii) noticing that in this case $a^{\mathrm{r}}_{\max} = 1 - \varepsilon$. ■

This result is interesting as it shows us that, for any network and under a very mild assumption on the compensation rule, the error due to packet losses is at most inversely proportional to the size of the network, and thus goes to zero for large networks. Observe also that the bound is again inversely proportional to the minimal self-confidence ε of the nodes.

Remark 1 (Limitations of Prop. 5): The result in Proposition 5 is very general thanks to the fact that it does not use any information about the process $S(t)$, except the independence between components. As a consequence, it may be looser than bounds available on specific compensation models. Indeed, in the case of (5), $\epsilon = 1-q$ and Proposition 5 implies $\gamma \ge \frac{q}{1-q}$. This bound is looser than the one in Proposition 4, and does not capture the role of p in the performance.

IV. Collision Broadcast Gossip Algorithm

We now consider the more complex case of packet losses caused by interferences between multiple broadcasted messages. The idea is that nodes occasionally broadcast their state, and that interferences result in the loss of all incoming information for a node when two or more of her neighbors are simultaneously broadcasting, or when the node is broadcasting herself.

More particularly, consider a symmetric bi-directional network G. At each time step, each node j may broadcast her state $x_j(t)$ to all her neighbors with a probability p, independently of the other nodes. If a node i is not broadcasting her state at time t, and if exactly one of her neighbors, j, broadcasts her state, node i updates her state to $x_i(t+1) = x_i(t) + q_i(x_j(t) - x_i(t))$. In all other cases, node i receives no information, and $x_i(t+1) = x_i(t)$.

This algorithm has been introduced and studied for regular graphs and a uniform $q_i = q$ in [5]. For general graphs, q_i must depend on p and on the degree d_i in order to guarantee that the protocol preserves the average on expectation, i.e., that $\mathbf{1}^*\mathbb{E}(L) = 0$. A sufficient condition for the latter condition to hold is the symmetry of expectations $\mathbb{E}(a_{ij}) = \mathbb{E}(a_{ji})$. Observe that $a_{ij}(t) = q_i$ if j broadcasts at time t while neither i nor any of her $d_i - 1$ other neighbors broadcasts at the same time, which happens with a probability $p(1-p)^{d_i}$, and $a_{ij}(t) = 0$ otherwise. Therefore, we have $\mathbb{E}(a_{ij}) = q_i p(1-p)^{d_i}$. Taking $q_i = q_0(1-p)^{-d_i}$ for every i, for an arbitrary q_0 for which $q_i < 1$ for all i, we obtain $\mathbb{E}(a_{ij}) = q_0 p$, for any i, j connected by an edge in G, which clearly satisfies the symmetry condition. This rule requires the nodes to know p and their degrees, but the degree can easily be estimated, and a uniform p would be known to every node.

Analyzing such a system is challenging because the simultaneous updates can be numerous, and correlated. Indeed, if i receives a message from j, every other neighbor of j will also be sent that message. As a result, if a node ℓ has also sent a message to a neighbor k of j, node k will not receive any message. This translates into a correlation between a_{ij} and $a_{k\ell}$, even though the corresponding edges do not share any common node. Theorem 1 has so far been particularized to classes of systems with few simultaneous updates, and to systems where update coefficients could be separated in mutually uncorrelated groups of small sizes (see Lemma 2(ii) on uncorrelated updates, and [9] for more examples), which is not the case here. In what follows, we propose a new particularization of Theorem 1 to systems with weak correlations, and then apply it to the Collision Broadcast Gossip Algorithm.

Proposition 6 (Weak correlations): Consider system (2) and suppose that $\mathbf{1}^*\mathbb{E}(L) = 0$. Define for every i, j, k, l the covariance $\sigma_{ij,kl} = \mathbb{E}(a_{ij}a_{kl}) - \mathbb{E}(a_{ij})\mathbb{E}(a_{kl})$. If there exists a $\rho_{\max} \geq 0$ such that $\sum_{k,l} |\sigma_{ij,kl}| \leq \rho_{\max}\mathbb{E}[a_{ij}]$ for all i, j, and a positive constant $a^{\mathrm{r}}_{\max}$ such that $\sum_{j:j\neq i} a_{ij} \leq a^{\mathrm{r}}_{\max}$ for all $i \in I$, then Theorem 1 holds true for

$$\gamma = \frac{\rho_{\max}}{1 - a^{\mathrm{r}}_{\max}}. \tag{6}$$

Our proof of Proposition 6 requires the next Lemma, which follows from Cauchy-Schwarz inequality.

Lemma 7: Let S be a finite set, and k, x, y functions $S \to \mathbb{R}$ taking at $\alpha \in S$ the respective values $k_\alpha, x_\alpha, y_\alpha$. Then

$$\left|\sum_{\alpha\in S} k_\alpha x_\alpha y_\alpha\right| \leq \sqrt{\sum_{\alpha\in S} |k_\alpha|\, x_\alpha^2}\sqrt{\sum_{\alpha\in S} |k_\alpha|\, y_\alpha^2}.$$

Proof of Proposition 6: Observe that $x^*L^*\mathbf{1}\mathbf{1}^*Lx = (\mathbf{1}Lx)^2 = \left(\sum_{i,j} a_{ij}(x_j - x_i)\right)^2$. Moreover, since $\mathbf{1}^*\mathbb{E}(L) = 0$, there holds $\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L] = \mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L] - \mathbb{E}[L^*]\mathbf{1}\mathbf{1}^*[L]$, and therefore $x^*\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L]x =$

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i,j} a_{ij}(x_j - x_i)\right)^2 - \left(\sum_{i,j} \mathbb{E}(a_{ij})(x_j - x_i)\right)^2 \\
&= \sum_{i,j,k,l} \left(\mathbb{E}(a_{ij}a_{kl}) - \mathbb{E}(a_{ij})\mathbb{E}(a_{kl})\right)(x_j - x_i)(x_l - x_k) \\
&= \sum_{i,j,k,l} \sigma_{ij,kl}\,(x_j - x_i)(x_l - x_k).
\end{aligned}$$

Applying Lemma 7 to the latter summation (on the index $\alpha = (i, j, k, l)$) leads to $x^*\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L]x \leq$

$$\sqrt{\sum_{i,j,k,l} |\sigma_{ij,kl}|\,(x_j - x_i)^2}\sqrt{\sum_{i,j,k,l} |\sigma_{ij,kl}|\,(x_l - x_k)^2}.$$

By relabeling the second factor, and taking into account the symmetry $\sigma_{ij,kl} = \sigma_{kl,ij}$, we obtain

$$x^*\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L]x \leq \sum_{i,j,k,l} |\sigma_{ij,kl}|\,(x_j - x_i)^2.$$

The existence of $\rho_{\max}$ implies then that $x^*\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L]x$ is bounded by

$$\sum_{i,j} \rho_{\max}\mathbb{E}(a_{ij})\,(x_j - x_i)^2 = \rho_{\max}x^*\mathbb{E}\left(L + L^*\right)x.$$

The result then follows from Lemma 2 (i). ■

Remark 2: Since $a_{ij} \in [0, 1]$, one can prove that $\sigma_{ij,kl} \leq \frac{1}{4}\mathbb{E}a_{ij}$. Therefore, when no coefficient a_{ij} is correlated with more than $C - 1$ other coefficients, Proposition 6 can always be applied with $\rho_{\max} = C/4$.

We now apply this new result to the Collision Broadcast Gossip Algorithm described at the beginning of this section. Let $d_{\max}$ be the largest degree in G. For any i, j, $a_{ij} \leq q_i = q_0(1 - p)^{-d_i} \leq q_0(1 - p)^{-d_{\max}}$. Moreover, for a given i, at most one coefficient $a_{ij}(t)$ $(j \neq i)$ can be positive at the time. Therefore, there holds $\sum_j a_{ij}(t) \leq q_0(1 - p)^{-d_i}$, and the latter quantity is a valid $a^{\mathrm{r}}_{\max}$ for the denominator in (6).

We now compute a value of $\rho_{\max}$. Observe that a_{ij} is determined by the broadcast "decisions" of j, of i, and of all other neighbors of i. It depends thus on the (independent) broadcast events at all the nodes at distance less than or equal to 1 from i, and no other nodes. As a consequence, if i and k are distant by 3 or more on the network G, the coefficients a_{ij} and a_{kl} depend on no common event and are uncorrelated. The number of coefficients correlated to a_{ij} is thus the number of (directed) edges leaving nodes at a distance less than 3 from i. There are at most $d_{\max}^2(d_{\max} - 1)$ such edges. Indeed, i has at most $d_{\max}$ neighbors, and at most $d_{\max}(d_{\max} - 1)$ neighbors of neighbors, and each node is incident to at most $d_{\max}$ edges. Remark 2 implies then that $\frac{1}{4}d_{\max}^3$ is a valid $\rho_{\max}$. Using Proposition 6 and the value of $a^{\mathrm{r}}_{\max}$ that we have computed above, we obtain the following result, which establishes that the Collision Broadcast Gossip Algorithm is asymptotically unbiased on networks with bounded or sufficiently slowly growing degrees.

Proposition 8 (Collision Broadcast): If the Collision Broadcast Gossip Algorithm is implemented with a broadcast probability p and coefficients $q_i = q_0(1 - p)^{-d_i}$ for each node i on a network with degrees bounded by $d_{\max}$, then Theorem 1 holds true for

$$\gamma = \frac{1}{4}\frac{d_{\max}^3}{1 - q_0(1 - p)^{-d_{\max}}}.$$

This analysis of the error is confirmed by simulations in Fig. 3. Note that for the sake of simplicity of exposition, we have used the conservative approach of Remark 2 that allows us to just compute a bound on the number of coefficients to which a_{ij} may be correlated. More accurate bounds could be obtained by a more precise estimation of $\rho_{\max}$.

V. Discussion and conclusion

In this paper, we have provided estimates of the mean square error due to packet losses in consensus algorithms, which are consistent with the trends observed experimentally. These results rely on general bounds that we have recently proved, and on an extension of these bounds that takes correlations into account. Under the assumption that the algorithms converge to consensus, our results imply that the effect of packet drops on the mean square error of the consensus value decreases no slower than the inverse of the network size. In the limit for $N \to \infty$, we thus claim that packet losses have negligible effects, but we want to stress that our bounds hold for any network size. Since they also hold for any topology, we conclude that they are relevant for network analysis in practical cases. It would of course

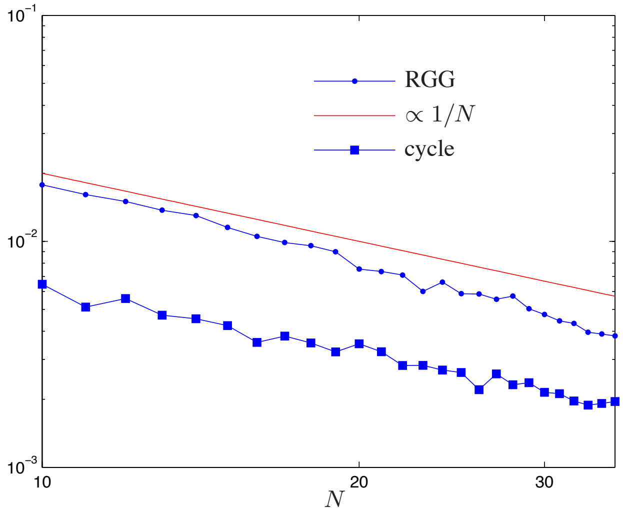

Fig. 3. Simulation of the mean square error (3) as a function of N for cycle and random geometric graphs (cf. caption of Fig. 1 for more details). Expectation is sampled by averaging over 1000 runs. On RGGs we set $p = 0.2$ and $q_i \leq 0.9$ for all $i \in I$; on cycles we set $p = 0.3$ and $q_i \leq 0.5$.

be interesting to design "auto-correcting" algorithms, which are able to effectively compensate for the loss of packets. However, that would come at the price of more complicate dynamics. Instead, in this paper we have provided conditions under which there is limited need for such a design and implementation effort.

References

[1] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *IEEE Transactions on Signal Processing*, 57(7):2748–2761, 2009.

[2] R. Carli, G. Como, P. Frasca, and F. Garin. Distributed averaging on digital erasure networks. *Automatica*, 47(1):115–121, 2011.

[3] 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*, 98(11):1847–1864, 2010.

[4] O. Dousse, F. Baccelli, and P. Thiran. Impact of interferences on connectivity in Ad Hoc Networks. *IEEE-ACM Transactions on Networking*, 13(2):425–436, 2005.

[5] F. Fagnani and P. Frasca. Broadcast gossip averaging: interference and unbiasedness in large Abelian Cayley networks. *IEEE Journal of Selected Topics in Signal Processing*, 5(4):866–875, 2011.

[6] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, 26(4):634–649, 2008.

[7] F. Fagnani and S. Zampieri. Average consensus with packet drop communication. *SIAM Journal on Control and Optimization*, 48(1):102–133, 2009.

[8] P. Frasca. Continuous-time quantized consensus: Convergence of Krasovskii solutions. *Systems & Control Letters*, 61(2):273–278, 2012.

[9] P. Frasca and J. M. Hendrickx. On the mean square error of randomized averaging algorithms. *Automatica*, 2013, to appear.

[10] F. Garin and S. Zampieri. Mean square performance of consensus-based distributed estimation over regular geometric graphs. *SIAM Journal on Control and Optimization*, 50(1):306–333, 2012.

[11] J. M. Hendrickx and J. N. Tsitsiklis. Convergence of type-symmetric and cut-balanced consensus seeking systems. *IEEE Transactions on Automatic Control*, 58(1):214–218, 2013.

[12] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: Quantized data and random link failures. *IEEE Transactions on Signal Processing*, 58(3):1383–1400, 2010.

[13] B. Nazer, A. G. Dimakis, and M. Gastpar. Local interference can accelerate gossip algorithms. In *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, September 2008.

[14] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

[15] A. Tahbaz-Salehi and A. Jadbabaie. Consensus over ergodic stationary graph processes. *IEEE Transactions on Automatic Control*, 55(1):225–230, 2010.

[16] S. Vanka, M. Haenggi, and V. Gupta. Convergence speed of the consensus algorithm with interference and sparse long-range connectivity. *IEEE Journal of Selected Topics in Signal Processing*, 5(4):855–865, 2011.

[17] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007.

Appendix

Proof of Lemma 3: First of all, we note that $\mathbb{E}[S(t)] = p\mathbf{1}\mathbf{1}^*$, implying by linearity that $\mathbb{E}[L(Q \cdot S(t))] = pL(Q)$. The assumption of statistical independence between the entries of S implies that assuming $i \neq j$ and $k \neq l$ it holds

$$\mathbb{E}[a_{ij}a_{kl}] = \begin{cases} \mathbb{E}[a_{ij}^2] = p\,Q_{ij}^2 & \text{if } i = k \text{ and } j = l \\ \mathbb{E}[a_{ij}]\mathbb{E}[a_{kl}] = p^2 Q_{ij}Q_{kl} & \text{otherwise} \end{cases}$$

These formulas can be used to compute $\mathbb{E}[x^* L^* L x]$

$$\begin{aligned}
&= \mathbb{E}\Big[\sum_i \Big(\sum_j a_{ij}(x_i - x_j)\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_i \Big(\sum_j a_{ij}(x_i - x_j) \sum_k a_{ik}(x_i - x_k)\Big)\Big] \\
&= \sum_i \sum_{j,k} \mathbb{E}[a_{ij}a_{ik}](x_i - x_j)(x_i - x_k) \\
&= \sum_i \sum_{j,k} p^2 Q_{ij} Q_{ik}(x_i - x_j)(x_i - x_k) \\
&\qquad + \sum_i \sum_j \left(-p^2 Q_{ij}^2 + pQ_{ij}^2\right)(x_i - x_j)^2 \\
&= p^2 x^* L(Q)^* L(Q) x \\
&\qquad + p(1-p) x^* \left(L(Q \cdot Q) + L(Q \cdot Q)^*\right) x.
\end{aligned}$$

Similarly, $\mathbb{E}[x^* L^* \mathbf{1}\mathbf{1}^* L x]$

$$\begin{aligned}
&= \mathbb{E}\Big[\Big(\sum_{i,j} a_{ij}(x_i - x_j)\Big)\Big(\sum_{k,l} a_{kl}(x_k - x_l)\Big)\Big] \\
&= \sum_{i,j,k,l} \mathbb{E}\left[a_{ij}a_{kl}\right](x_i - x_j)(x_k - x_l) \\
&= \sum_{i,j,k,l} p^2 Q_{ij} Q_{kl}(x_i - x_j)(x_k - x_l) \\
&\qquad + \sum_{i,j} \left(-p^2 Q_{ij}^2 + p\,Q_{ij}^2\right)(x_i - x_j)^2 \\
&= p^2 x^* L(Q)^* \mathbf{1}\mathbf{1}^* L(Q) x \\
&\qquad + p(1-p) x^* \left(L(Q \cdot Q) + L(Q \cdot Q)^*\right) x \\
&= p(1-p) x^* \left(L(Q \cdot Q) + L(Q \cdot Q)^*\right) x,
\end{aligned}$$

where we have used $\mathbf{1}^* L(Q) = 0$ for the last equality. ∎