On the mean square error of randomized averaging algorithms

Paolo Frasca ∗ Julien M. Hendrickx †

November 22, 2011

Abstract

This paper regards randomized discrete-time consensus systems that preserve the average on expectation. As a main result, we provide an upper bound on the mean square deviation of the consensus value from the initial average. Then, we particularize our result to systems where the interactions which take place simultaneously are few, or weakly correlated; these assumptions cover several algorithms proposed in the literature. For such systems we show that, when the system size grows, the deviation tends to zero, not slower than the inverse of the size. Our results are based on a new approach, unrelated to the convergence properties of the system: this independence questions the relevance in this context of the spectral properties of the matrices related to the graph of possible interactions, which have appeared in some previous results.

1 Introduction

In modern control and signal processing applications, effective and easy-to-implement distributed algorithms for computing averages are an important tool. As a significant and motivating example, we recall that, in estimation problems, the law of large numbers allows using the average as the estimator of an expectation. In that context, the average is an unbiased estimator, and its mean square error decreases as the inverse of the number of samples. In a distributed setting, the sample values are available at the nodes of a communication network, and the average needs to be approximated by running an iterative consensus system having the sample data as the initial condition. Clearly one has to ensure that along the iterations of the consensus system, little (or no) deviation from the correct average is introduced. This requirement can be achieved by a local symmetry assumption or by some global constraints on the update, which may be difficult to enforce when updates are performed asynchronously, possibly following a random scheme. Then, a weaker requirement for stochastic updates is preserving the average on expectation: such systems are known to converge to consensus under mild conditions, but their consensus value is in general different from the average.

In this paper, we consider linear randomized asynchronous averaging algorithms, and we analyze the mean square deviation of the consensus value from the initial average. We want to ensure that this error is small, so that averages can be effectively computed. In particular, we aim at providing conditions under which the mean square error tends to zero when the number of samples, i.e. the number of nodes, grows. We will refer to this property as to the asymptotical accuracy of the algorithm.

Literature review

The opportunity of using randomized systems to compute averages has already attracted a significant interest, as testified by recent surveys and special issues [11, 4]. Convergence theories for randomized linear averaging algorithms have been developed by Fagnani and Zampieri [9] and with more generality by Tahbaz-Salehi and Jadbabaie [12] and by Matei and Baras [10]. As we will formally define later,

∗ Paolo Frasca is with the Dipartimento di Matematica at the Politecnico di Torino, corso Duca degli Abruzzi 24, 10126 Torino, Italy; paolo.frasca@polito.it. His work is partly supported by MIUR under grant PRIN-20087W5P2K.
† Julien M. Hendrickx is with the ICTEAM Institute at the Université catholique de Louvain, Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium; julien.hendrickx@uclouvain.be. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.
random linear averaging algorithms consist in multiplying the node-indexed state by a random update matrix. In principle, the variance of the consensus value can be exactly computed by the formula in [12, Eq. (7)], which involves the dominant eigenvectors of the first two moments of the update matrix. Unfortunately, little is known about these eigenvectors, and in particular explicit formulas are not available, so that these results are difficult to apply. A few papers, on the other hand, have focused on specific examples of randomized algorithms, obtaining results which are interesting, although partial, from our perspective [8, 2, 6, 7]. Typically, these results are obtained as a by-product of a convergence analysis and involve the eigenvalues of the update matrices, which are fairly well known for many families of communication graphs. We come back to these results in Section 3 when discussing the examples.

### Contribution

In this paper, we consider discrete-time consensus systems with random updates that preserve the average in expectation, and we provide new bounds on the mean square deviation of the current average from the initial average. We show that under certain conditions the expected increase of the deviation is bounded proportionally to the expected decrease of the disagreement. This approach leads to bounds on the total deviation which are proportional to the initial disagreement and, unlike previous results, are actually independent of the convergence properties: indeed they hold at every time regardless of convergence. Compared to those already available in the literature, our bounds typically result in less conservative estimates of the deviation error and, remarkably, they are independent of the global properties of the communication network, like connectivity or graph spectrum and eigensystem. Instead, only local network properties, like the degree, play a role in the examples. By contrast, we recall that results in the literature about convergence, and speed of convergence, depend on global network properties. Our estimates show that deviation tends to zero when the number of nodes grows under weak assumptions on the update law, in particular for

i) systems where few updates take place simultaneously; and

ii) systems where the updates have small statistical dependence across the network.

Thanks to their generality and to their dependence on only local network properties, our results offer effective and easy-to-implement guidelines to the designer who needs to choose a network and an algorithm to solve an estimation problem.

### Notation and preliminaries

In this work, we use the notion of *(weighted directed) graph*, which we define as a pair $G = (I, A)$, where $I$ is a finite set 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 $A_{ij} > 0$. For simplicity, we will sometimes assume that a graph may have no loops, that is $A_{ii} = 0$ for every $i \in I$. The column-degree $d_{col}^i$ of $i$ is the number of off-diagonal positive elements in the $i$-th column of $A$, i.e. the cardinality of $\{j \neq i : a_{ij} > 0\}$. The graph is said to be strongly connected if for every node $i$ and $j$, there exists a sequence $i = i_0, i_1, \ldots, i_p = j$ of nodes such that $A_{i_k,i_{k+1}} > 0$ for $k = 0, \ldots, p - 1$. Given a graph, that is, 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 $[L(A)]_{ii} = \sum_{j: j \neq i} A_{ij}$. Observe that $L(A)$ is positive semidefinite and that $L(A)1 = 0$, provided we denote by $1$ the vector of suitable size whose components are all $1$. Besides, to any matrix $L$ satisfying $L1 = 0$ with nonpositive off-diagonal elements, one can associate a corresponding weighted graph.

### 2 Problem statement and main result

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

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

(1)
where for every $i, j \in I$, we assume $a_{ij}(t)$ to be a sequence of independent and identically distributed random variables such that $a_{ij}(t) \geq 0$ and $\sum_{t \in I} a_{ii}(t) = 1$ for all $t \geq 0$. System (1) is run with the goal for the state of each node to provide a good estimate of the initial average $\frac{1}{N} \sum_{i} x_i(0)$. 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

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

or in matrix form as

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

(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, $L(t)$ is the Laplacian matrix of a weighted graph $(I, A(t))$ where the entries of $A(t)$ are defined as $[A(t)]_{ij} = a_{ij}(t)$. The convergence of (2) has been addressed in the literature; the next proposition provides a handy sufficient condition.

**Proposition 1** (Convergence [9]). Consider system (2). If the graph induced by $\mathbb{E}[L(t)]$ is strongly connected and there exists $i \in I$ such that almost surely $L_{ii}(t) < 1$, then there exists a scalar random variable $x_\infty$ such that $x(t)$ converges almost surely to $x_\infty 1$.

Rather than in convergence, we are interested in the quality of the convergence value, in terms of its distance from the initial average: this issue is investigated in the rest of this paper. For our convenience, we denote the average of the $x_i(t)$ by

$$\bar{x}(t) := \frac{1}{N} \sum_{i} x_i(t)$$

and the second moment by $\bar{x}^2(t) = \frac{1}{N} \sum_{i} x_i^2(t)$. The next result provides a necessary and sufficient condition for the average to be preserved in expectation, that is, for the process $\bar{x}(t)$ to be a martingale with respect to the filtration induced by $x(t)$.

**Proposition 2** (Average preservation). Consider system (2) and let $\{F_t\}_{t \in \mathbb{Z}_{\geq 0}}$ denote the filtration of $\sigma$-algebras generated by the process $x(t)$. Then, $\mathbb{E}[\bar{x}(t+1)|F_t] = \bar{x}(t)$ if and only if $1^*\mathbb{E}[L(t)] = 0$.

**Proof.** Since with our assumptions $L(t)$ is independent from $F_t$, the result immediately follows from $\bar{x}(t+1) = \bar{x}(t) - 1^*L(t)x(t)$. \qed

In view of this result, we restrict our attention to systems that preserve the average on expectation, that is we will assume $1^*\mathbb{E}[L(t)] = 0$, implying that $\mathbb{E}[\bar{x}(t)] = \bar{x}(0)$ for all $t \geq 0$. Consequently, we are left with the problem of studying the variance of $\bar{x}(t)$, that is $\mathbb{E}[(\bar{x}(t) - \bar{x}(0))^2]$. We will derive all our bounds from the following general result, which establishes that, under some conditions, the increase of the deviation is bounded proportionally to the decrease of the disagreement.

**Theorem 3** (Accuracy condition). Let $x$ be an evolution of system (2), and denote

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

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

$$\mathbb{E}[L(s)^*11^*L(s)] \leq \gamma \mathbb{E}[L(s) + L(s)^* - L(s)^*L(s)],$$

(3)

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

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

If moreover the system converges to consensus, then

$$\mathbb{E}[(x_\infty - \bar{x}(0))^2] \leq \frac{\gamma}{N} V(0).$$
Proof. We compute the increase of the deviation up to time $t$

$$
\mathbb{E}\left[ (\bar{x}(t) - \bar{x}(0))^2 \right] = \mathbb{E}\left[ \left( \sum_{s=0}^{t-1} (\bar{x}(s+1) - \bar{x}(s)) \right)^2 \right]
$$

we condition upon the filtration generated by $x(t)$

$$
= \sum_{s=0}^{t-1} \mathbb{E}\left[ (\bar{x}(s+1) - \bar{x}(s))^2 | F_s \right]
+ 2 \sum_{s=0}^{t-1} \sum_{u=0}^{s-1} \mathbb{E}\left[ (\bar{x}(s+1) - \bar{x}(s))(\bar{x}(u+1) - \bar{x}(u)) | F_s \right]
$$

and since $\mathbb{E}\left[ (\bar{x}(s+1) - \bar{x}(s))(\bar{x}(u+1) - \bar{x}(u)) | F_s \right] = 0$ for $u < s$ by Proposition 2, we obtain

$$
= \sum_{s=0}^{t-1} \mathbb{E}\left[ (\bar{x}(s+1) - \bar{x}(s))^2 | F_s \right]
$$

and finally since $\mathbb{E}[\bar{x}(s+1)\bar{x}(s)|F_s] = \bar{x}(s)^2$

$$
= \sum_{s=0}^{t-1} \mathbb{E}\left[ \bar{x}(s+1)^2 - \bar{x}(s)^2 | F_s \right].
$$

Then, we wish to study the increase in the squared average at each iteration. To this goal, a straightforward manipulation gives the following useful relations, which are valid for every $s \geq 0$

$$
N \left( \bar{x}^2(s+1) - \bar{x}^2(s) \right) = -x(s)^* \left[ (L(s) + L(s))^* - L(s)^*L(s) \right] x(s)
$$

$$
N^2 \left( \bar{x}^2(s+1) - \bar{x}^2(s) \right) = (1^*L(s)x(s))^2 - 21^*x(s)1^*L(s)x(s)
$$

By taking the conditional expectation, we also obtain that

$$
N \mathbb{E}\left[ \bar{x}^2(s+1) - \bar{x}^2(s) | F_s \right] = -x(s)^* \mathbb{E}[L(s) + L(s)^* - L(s)^*L(s)] x(s)
$$

$$
N^2 \mathbb{E}\left[ \bar{x}^2(s+1) - \bar{x}^2(s) | F_s \right] = x(s)^* \mathbb{E}[L(s)^*11^*L(s)] x(s)
$$

These two formulas and (3) allow us to relate the change in the square of the average with the change in the average of the squares, implying that

$$
\mathbb{E}\left[ \bar{x}^2(s+1) - \bar{x}^2(s) | F_s \right] \leq \frac{\gamma}{N} \mathbb{E}\left[ \bar{x}^2(s) - \bar{x}^2(s+1) | F_s \right],
$$

and consequently the increase in the deviation is upper bounded as follows:

$$
\mathbb{E}\left( \bar{x}(t) - \bar{x}(0) \right)^2 \leq \sum_{s=0}^{t-1} \frac{\gamma}{N} \mathbb{E}\left[ \bar{x}^2(s) - \bar{x}^2(s+1) | F_s \right]
$$

$$
= \frac{\gamma}{N} \sum_{s=0}^{t-1} \mathbb{E}\left[ \bar{x}^2(s) - \bar{x}^2(s+1) \right]
$$

$$
= \frac{\gamma}{N} \left( \bar{x}^2(0) - \mathbb{E}[\bar{x}^2(t)] \right)
$$

$$
\leq \frac{\gamma}{N} \bar{x}^2(0).
$$

Observe now that the algorithm is invariant under translation (addition of a constant to each component of the state). By applying the last inequality to $\dot{x}(t) = x(t) - \bar{x}(0)$, we obtain

$$
\mathbb{E}\left( \dot{x}(t) - \bar{x}(0) \right)^2 = \mathbb{E}\left( \langle \dot{x}(t) - \bar{x}(0) \rangle \right)^2 \leq \frac{\gamma}{N} \bar{x}^2(0) = \frac{\gamma}{N} \left( \frac{1}{N} \sum x_i(0) - \bar{x}(0) \right)^2 = \frac{\gamma}{N} V(0).
$$

$\square$
3 Applications and examples

In this section we see classes of systems of type (2) for which we can apply Theorem 3, that is, we can find $\gamma$ satisfying (3). Before presenting these example systems, we prove in the next subsection a general lemma which simplifies the search for $\gamma$.

3.1 Bounds on deterministic and stochastic Laplacians

The idea of the proofs of our results will always be to bound $E(L^*11^*L)$ and $E(L^*L)$ in terms of $E(L + L^*)$. In this purpose, the following result will be useful.

**Lemma 4** (Laplacian bounds). Let $L$ be the Laplacian of a weighted directed graph, and $a_{r \text{max}} > 0$ be such that

$$\sum_{j:j \neq i} a_{ij} \leq a_{r \text{max}}^r$$

for all $i$.

(i) If $1^*L = 0$, then

$$L^*L \leq a_{r \text{max}}^r (L + L^*). \quad (4)$$

Let now $L$ be a random variable such that the upper bound $a_{r \text{max}}^r$ is valid almost surely.

(ii) If $1^*E(L) = 0$, then

$$E(L^*L) \leq a_{r \text{max}}^r E(L + L^*). \quad (5)$$

(iii) If $1^*E(L) = 0$ and moreover there exists $\beta > 0$ such that

$$E(L^*11^*L) \leq \beta E(L + L^*),$$

then $E[L^*11^*L] \leq \gamma E[L + L^* - L^*L]$ holds for any $\gamma \geq \frac{\beta}{1 - a_{r \text{max}}^r}$.

Before the proof, we need the following preliminary lemma.

**Lemma 5.** Suppose that the coefficients $c_1, \ldots, c_m$ are nonnegative. Then, there holds

$$\left( \sum_{i=1}^m c_i z_i \right)^2 \leq \left( \sum_{i=1}^m c_i \right) \sum_{i=1}^m c_i z_i^2$$

**Proof.** Let $u, v \in \mathbb{R}^m$ be defined by $u_i = \sqrt{c_i}$ and $v_i = \sqrt{c_i} z_i$. It follows from Cauchy-Schwartz inequality that

$$\left( \sum_{i=1}^m c_i z_i \right)^2 = (u^*v)^2 \leq (||u||_2 ||v||_2)^2 = \left( \sum_{i=1}^m u_i^2 \right) \left( \sum_{i=1}^m v_i^2 \right) = \left( \sum_{i=1}^m c_i \right) \sum_{i=1}^m c_i z_i^2.$$

\[ \square \]

**Proof of Lemma 4.** Along this proof let $x \in \mathbb{R}^I$ be arbitrary but fixed. To prove claim (i), we note that $(Lx)_i = \sum_j a_{ij} (x_i - x_j)$. Therefore,

$$x^*L^*Lx = \sum_i \left( \sum_{j:j \neq i} a_{ij} (x_i - x_j) \right)^2.$$

For every $i$, since $a_{r \text{max}}^r \geq \sum_{j:j \neq i} a_{ij}$, it follows then from Lemma 5 that

$$\left( \sum_{j:j \neq i} a_{ij} (x_j - x_i) \right)^2 \leq a_{r \text{max}}^r \sum_{j:j \neq i} (x_j - x_i)^2,$$
and by summing on \( i \) that

\[
x^* L^* L x \leq a^r_{\text{max}} \sum_i \sum_{j: j \neq i} a_{ij} (x_i - x_j)^2.
\]

Also note that

\[
\sum_i \sum_{j: j \neq i} a_{ij} (x_j - x_i)^2 = \sum_i x_i^2 \sum_{j: j \neq i} a_{ij} - 2 \sum_i \sum_{j: j \neq i} a_{ij} x_i x_j + \sum_j x_j^2 \sum_{i: i \neq j} a_{ij}.
\]

Since \( 1^* L = 0 \), we have \( \sum_{i: i \neq j} a_{ij} = \sum_{j: j \neq i} a_{ij} \). Therefore, a relabeling of the third term leads to

\[
\sum_i \sum_{j: j \neq i} a_{ij} (x_j - x_i)^2 = 2 \sum_i \left( \sum_{j: j \neq i} a_{ij} \right) x_i^2 - 2 \sum_i \sum_{j: j \neq i} a_{ij} x_i x_j = 2 x^* L x = x^* (L + L^*) x,
\]

where we have used \( x^* L^* x = (x^* L x)^* = x^* L x \). Using (6), statement (i) follows.

We now turn to prove statement (ii). It follows from (6) that

\[
x^* \mathbb{E}(L^* L) x = \mathbb{E}(x^* L^* L x) \leq \mathbb{E} \left[ a^r_{\text{max}} \sum_i \sum_{j: j \neq i} a_{ij} (x_i - x_j)^2 \right] = a^r_{\text{max}} \sum_i \sum_{j: j \neq i} \mathbb{E}(a_{ij})(x_i - x_j)^2.
\]

Since \( \mathbb{E}(L) \) is a (deterministic) Laplacian and \( 1^* \mathbb{E}(L) = 0 \), we can apply the same argument leading to (4) in order to argue that

\[
x^* \mathbb{E}(L^* L) x \leq a^r_{\text{max}} x^* \mathbb{E}(L + L^*) x,
\]

which implies (5). Finally, we prove the last claim (iii). It follows from (5) that \( -a^r_{\text{max}} \mathbb{E}(L + L^*) \leq -\mathbb{E}(L^* L) \). Therefore, the existence of \( \beta \) implies that for any \( \gamma \geq \frac{\beta}{1 - a^r_{\text{max}}} \), there holds

\[
\mathbb{E}(L^* 11^* L) \leq \beta \mathbb{E}(L + L^*) \leq \gamma \mathbb{E}(L + L^*) - \gamma a^r_{\text{max}} \mathbb{E}(L + L^*) \leq \gamma \mathbb{E}(L + L^* - L^* L).
\]

The quantity \( 1 - a^r_{\text{max}} \) appearing in Lemma 4(iii) actually corresponds to a lower bound on the “self-confidence” \( a_{ii}(t) = 1 - \sum_{j \in J} a_{ij}(t) \) of the nodes. For a constant \( \beta \), the bound on the mean square error is thus inversely proportional to the minimal self-confidence. This is consistent with the intuition that, when \( a_{ii}(t) \) is very small, the information held by some nodes may be almost entirely “forgotten” in one iteration, possibly resulting in large variations of the average.

### 3.2 Limited simultaneous updates

In this section, we show that a suitable \( \gamma \) to satisfy the condition in Theorem 3 can be found when the number, or at least the contribution, of the simultaneous updates is small. The next result has the following interpretation: the mean square deviation can be bounded proportionally to the ratio between “strength” of the interactions in the system and “self-confidence” of each node. Note that from now on, when studying the evolution of system (2), we will for brevity forget to write the dependence on time of the random variables \( a_{ij} \) and \( L \), whenever this causes no confusion.

**Theorem 6 (Limited updates).** Consider system (2) and let \( a^\text{all}_{\text{max}} \) and \( a^r_{\text{max}} \) be two positive constants such that almost surely \( \sum_i \sum_{j: j \neq i} a_{ij} \leq a^\text{all}_{\text{max}} \) and \( \sum_{j: j \neq i} a_{ij} \leq a^r_{\text{max}} \) for all \( i \in I \). If \( 1^* \mathbb{E}(L) = 0 \), then the condition of Theorem 3 holds for all

\[
\gamma \geq \frac{a^\text{all}_{\text{max}}}{1 - a^r_{\text{max}}}.
\]

**Proof.** It follows from Lemma 5 that

\[
x^* L^* 11^* L x = \left( \sum_i \sum_{j: j \neq i} a_{ij} (x_j - x_i) \right)^2 \leq a^\text{all}_{\text{max}} \sum_i \sum_{j: j \neq i} a_{ij} (x_j - x_i)^2.
\]
Therefore,
\[
\mathbb{E}(x^* L^* 11^* L x) \leq a_{\text{max}}^{\text{all}} \sum_i \sum_{j \neq i} \mathbb{E}(a_{ij}) (x_j - x_i)^2 = a_{\text{max}}^{\text{all}} x^T (L + L^*) x,
\]
where we have used Lemma 4(i), so that \( \mathbb{E}(L^* 11^* L) \leq a_{\text{max}}^{\text{all}} (L + L^*) \). The result follows then from Lemma 4(iii).

\[\square\]

Theorem 6 can be applied to several particular cases involving small number of edges or small interactions: we discuss here two of them, drawn from the literature. The following is the simplest example of a randomized averaging algorithm.

**Example 1** (Asynchronous Asymmetric Gossip Algorithm (AAGA)). Let a graph \( G = (I, W) \) and \( q \in (0, 1) \) be given, such that \( 1^* W 1 = 1 \). For every \( t \geq 0 \), one edge \((i,j)\) is sampled from a distribution such that the probability of selecting \((i,j)\) is \( W_{ij} \). Then, \( x_i(t+1) = (1-q) x_i(t) + q x_j(t) \) and \( x_k(t+1) = x_k(t) \) for \( k \neq i \).

We note that Proposition 1 applies to Example 1, as well as to all following examples in this paper.

**Corollary 7** (AAGA is asymptotically accurate). Consider the AAGA system (2) of Example 1 and assume that and \( W1 = W^* 1 \). Then Theorem 3 holds for any \( \gamma \geq \frac{q}{1-q} \).

**Proof.** Note that \( \mathbb{E}[L(t)] = q L(W) \); then \( 1^* \mathbb{E}[L(t)] = 0 \) by the assumption on \( W \). To apply Theorem 6 we observe that \( a_{\text{max}}^{\text{all}} = a_{\text{max}}^{\text{all}} = q \) since only one node is sending her state to another. \[\square\]

This result implies that the expected deviation is not larger than \( \frac{q}{1-q} \frac{1}{N} V(0) \): extensive numerical simulations indicate that this bound captures the correct dependence on the size. We summarize this evidence in Figure 1 by taking two exemplary sequences of graphs: ring graphs and de Bruijn graphs. We recall their definitions and significant properties in the next example.

**Example 2** (Ring and de Bruijn graphs). A graph \( G = \{0, \ldots, N-1\}, A \) is a ring when \( A \) is such that for every \( i \), it holds \( A_{ij} > 0 \) if and only if either \( j = i+1 \) or \( j = i-1 \) (modulo \( N \)). Instead, \( G \) is a de Bruijn graph on \( n \) symbols of dimension \( k \) when \( N = n^k \) nodes and every node \( i \) is connected to \( ni, ni+1, ni+2, \ldots, ni+n-1 \) (all modulo \( n^k \)). Their global connectivity properties are very different, especially in large networks: rings are “poorly” connected, whereas de Bruijn graphs are “very well” connected. This difference can be highlighted by looking at the spectra of their Laplacians. Assuming for simplicity that \( A \in \{0,1\}^{I \times I} \), we compute the smallest eigenvalue of the Laplacian, denoted by \( \lambda_1 \). For rings, \( \lambda_1 = 2 - 2 \cos \left( \frac{2 \pi}{N} \right) \leq \frac{4\pi^2}{N^2} \), which goes to zero as \( N \) grows. For de Bruijn graphs, \( \lambda_1 \) is equal to \( n \). Then, we can consider a sequence of de Bruijn graphs with \( n = 2 \) and increasing size \( 2^k \), for \( k \in \mathbb{N} \). On such a sequence, \( \lambda_1 \) does not go to zero as \( N \) grows, and the degree is not larger than 2, which is the degree of the ring (and is actually exactly 2 for all nodes if self-loops are taken into account). The notable spectral properties of these two sequences are also discussed in [3], in connection with the convergence speed of averaging algorithms.

The next example, which applies very naturally to wireless networks, has attracted a significant attention [9, 2].

**Example 3** (Broadcast Gossip Algorithm (BGA)). Let \( q \in (0,1) \) a graph \( G = (I, W) \) be given, such that \( W \in \{0,1\}^{I \times I} \). For every \( t \geq 0 \), one node \( j \) is sampled from a uniform distribution over \( I \). Then, \( x_i(t+1) = (1-q) x_i(t) + q x_j(t) \) if \( W_{ij} > 0 \) and \( x_i(t+1) = x_i(t) \) otherwise. In other words, one randomly selected node broadcasts her value to all her neighbors, which update their values accordingly.

A few partial results about the mean square deviation of this algorithm are available in the literature: we provide a summary in the following remark.

---

1The AAGA system is also studied in [8, Section 4], where it is proved that the mean square deviation of the limit value is not larger than \( \frac{q}{1-q} \frac{1}{N^2} V(0) \). We note that although the authors assume that the initial condition is a random variable, their results on the deviation actually remain valid for arbitrary \( x(0) \). Our bound from Corollary 7 is slightly larger, but asymptotically equivalent.
Figure 1: Logarithmic plots of the simulated mean square deviation $E[(x_{\infty} - \bar{x}(0))^2]$ against the graph size $N$, for AAGA and BGA systems with $q = 0.5$ running over ring graphs and de Bruijn graphs on 2 symbols. See text for precise graph definitions. Convergence to consensus is approximated by the condition $N^{-1}\|x(t) - \bar{x}(t)\|_2^2 \leq 10^{-4}$ and expectation is simulated by averaging over 500 runs. Each initial condition is sampled from independent uniform distributions over $[0, 1]$.

**Remark 1** (Earlier results on BGA). Previous results about the deviation of BGA are dependent on the topology of the network. In [2, Proposition 3] it is proved that the mean square deviation is upper bounded by $V(0)\left(1 - \frac{\lambda_1}{\lambda_{N-1}} - \frac{1}{2N^2}\right)$, where $\lambda_i$ is the $i$-th smallest non-zero eigenvalue of the Laplacian of the graph $G$. This bound, however, does not imply that the deviation goes to zero as $N$ grows. In [6, Proposition 3.3] the authors obtain the upper bound $2V(0)\frac{q}{1-q}\frac{d_{\text{max}}^2}{N\lambda_1}$, where $d_{\text{max}}$ is the maximum degree of the graph. We can see that on a ring graph $\frac{d_{\text{max}}^2}{N\lambda_1} \geq \frac{N}{2}$: in such a case, one cannot argue that the deviation goes to zero. Although for rings and other sequences of graphs asymptotic accuracy is shown in [7], using Markov chain theory results from [5], a general proof of accuracy is not available in the literature.

Simulations show that the global topology of the graph plays a limited role: as a demonstration, we plot in Figure 1 the results for ring and de Bruijn graphs. Based on similar simulations, it was conjectured in [6] that the mean square error of the BGA is proportional to the ratio between the degree and the number of nodes. This fact can actually be proved by applying Theorem 6.

**Corollary 8** (BGA is asymptotically accurate). Consider the BGA system (2) of Example 3 and assume that $W_1 = W^*1$. Then Theorem 3 holds for any $\gamma \geq \frac{q}{1-q}d_{\text{col}}^\text{max}$, where $d_{\text{col}}^\text{max}$ is the maximum column degree of the graph.

**Proof.** Note that $E[L(t)] = \frac{1}{d^\text{all}}L(W)$; then $1^*E[L(t)] = 0$ by the assumption on $W$. To apply Theorem 6 we observe that $a_{\text{max}}^r = q$ and $a_{\text{max}}^\text{all} = qd_{\text{col}}^\text{max}$, since one node may send her value to at most $d_{\text{col}}^\text{max}$ neighbors.

### 3.3 Uncorrelated updates

In this section we show that a small $\gamma$ can still be found even if there are many simultaneous updates, provided that the correlation between the updates is sufficiently small. Indeed, the argument presented in the next remark shows that for an algorithm such that all entries $a_{ij}$ are uncorrelated, the conclusion of Theorem 3 holds.

**Remark 2** (No correlation implies accuracy). Observe that, using the convention that $a_{ii} = 0$, we have

$$x^*L^*11^*Lx = \left(\sum_{i,j \in I, i \neq j} a_{ij}(x_j - x_i)\right)^2 = \sum_{i,j,k,l \in I} a_{ij}a_{kl}(x_j - x_i)(x_l - x_k)$$

(7)
Suppose now that all $a_{ij}$ are uncorrelated, then

$$
\mathbb{E}(x^* L^{11^*} L x) = \sum_{i,j,k,l \in I} \mathbb{E}(a_{ij}) \mathbb{E}(a_{kl}) (x_j - x_i) (x_l - x_i) + \sum_{i,j \in I} (\mathbb{E}(a_{ij}^2) - \mathbb{E}(a_{ij})^2) (x_j - x_i)^2.
$$

It follows from (7) and from $1^* \mathbb{E}(L) = 0$ that the first term is 0. Then, if $a_{ij} \leq a_{\max}^{\text{ind}} \leq 1$ for every $i,j$, we obtain

$$
\mathbb{E}(x^* L^{11^*} L x) \leq a_{\max}^{\text{ind}} \sum_{i,j \in I} \mathbb{E}(a_{ij}) (x_j - x_i)^2 = a_{\max}^{\text{ind}} x^* \mathbb{E}(L + L^*) x,
$$

where the last equality holds thanks to $1^* \mathbb{E}(L) = 0$. Lemma 4(iii) implies then that Theorem 3 holds with $\gamma = \frac{a_{\max}^{\text{ind}}}{1 - a_{\max}^{\text{ind}}}$. $\square$

Clearly, a strong assumption such as the absence of correlation between the $a_{ij}$ is rarely met in practical algorithms. However, in this spirit we can look for results assuming a “small” degree of correlation. The first result assumes that the lines of $L(t)$ are uncorrelated, corresponding to the absence of correlation between the update behavior of the different nodes. It implies in particular that any scheme (preserving the average $\bar{x}$ on expectation) where nodes update their values independently and have a minimal self-confidence is asymptotically accurate.

**Theorem 9** (Uncorrelated updates). Consider system (2) and let $a_{\max}^r$ be a positive constant such that almost surely $\sum_{i,j \neq i} a_{ij} \leq a_{\max}^r$ for all $i \in I$. Assume that is $a_{ij}$ and $a_{kl}$ are uncorrelated when $i \neq k$. If $1^* \mathbb{E}(L) = 0$, then the condition of Theorem 3 holds for any

$$
\gamma \geq \frac{a_{\max}^r}{1 - a_{\max}^{\text{ind}}}.
$$

*Proof.* To take advantage of the decorrelation assumption, we first analyze the expression $\mathbb{E}[L^* (11^* - I_N) L]$. Let $l_i \bullet$ be the $i$-th row of $L$, i.e. $L = ([l_1 \bullet, l_2 \bullet, \ldots, l_n \bullet]^*)$. Our assumption implies that $l_i \bullet$ and $l_j \bullet$ are uncorrelated when $i \neq j$. Therefore, there holds

$$
\mathbb{E}(L^* (11^* - I_N) L) = \mathbb{E} \left( \sum_{i,j \neq i} l_i^* l_j \bullet \right) = \sum_{i,j \neq i} \mathbb{E}[l_i]^* \mathbb{E}[l_j] = \mathbb{E}[L^*] (11^* - I_N) \mathbb{E}[L].
$$

Since $1^* \mathbb{E}(L) = 0$, this implies that

$$
\mathbb{E}(L^* (11^* - I) L) = - (\mathbb{E} L)^* (\mathbb{E} L) \leq 0.
$$

Therefore, using first this last inequality and then Lemma 4(ii), we obtain

$$
\mathbb{E}(L L^*) = \mathbb{E}(L^* (11^* - I) L) + \mathbb{E}(L^* L) \leq \mathbb{E}(L^* L) \leq a_{\max}^r \mathbb{E}(L + L^*).
$$

The result follows then from Lemma 4(iii). $\square$

A natural example of uncorrelated updates is as follows.

**Example 4** (Synchronous Asymmetric Gossip Algorithm (SAGA)). Let $q \in (0,1)$ and a graph $G = (I, W)$ be given, such that $W 1 = 1$. For every $t \geq 0$, and every $i \in I$ one edge $(i,j_i)$ is sampled from a distribution such that the probability of selecting $(i,j_i)$ is $W_{i,j_i}$. Then, for every $i \in I$,

$$
x_i(t + 1) = (1 - q) x_i(t) + q x_{j_i}(t).
$$

In other words, every node chooses one neighbor, reads her value, and updates her own value accordingly.

Previous results on SAGA are only able to guarantee asymptotical accuracy on certain sequences of graphs.
The inequality (9) implies thus 

\[ E \] 

so that 

\[ \text{bound is asymptotically equivalent, as } N \to \infty, \text{ to } \frac{1}{1-q} \frac{1}{2N} V(0), \] 

where \( \text{esr}(W) \) is the second-largest absolute value of the eigenvalues of \( W \). This result fails to prove asymptotical accuracy for some sequences of graphs. For instance, on a ring graph with positive \( W_{ij} \)s equal to \( 1/2 \), we have 

\[ \frac{1}{2N} \frac{1}{1-\text{esr}(W)} = \frac{1}{2N} \frac{1}{\cos(\frac{\pi}{N})} V(0) \geq \frac{q}{1-q} \frac{N}{4\pi^2} V(0). \]

Simulations in Figure 2 suggest instead that asymptotical accuracy is a general property of SAGA: this fact is proved in the next result.

**Remark 3** (Earlier results on SAGA). This algorithm is also studied in [8, Section 5], where the authors derive an upper bound on the deviation of the limit value. When \( W \) is symmetric, this bound is asymptotically equivalent, as \( N \to \infty \), to 

\[ \frac{1}{2N} \frac{1}{1-\text{esr}(W)} V(0), \] 

where \( \text{esr}(W) \) is the second-largest absolute value of the eigenvalues of \( W \). This result fails to prove asymptotical accuracy for some sequences of graphs. For instance, on a ring graph with positive \( W_{ij} \)s equal to \( 1/2 \), we have 

\[ \frac{1}{2N} \frac{1}{1-\text{esr}(W)} = \frac{1}{2N} \frac{1}{\cos(\frac{\pi}{N})} V(0) \geq \frac{q}{1-q} \frac{N}{4\pi^2} V(0). \]

A second theorem, which is a sense dual to Theorem 9, assumes that the columns of \( L(t) \) are uncorrelated, corresponding to the fact that the transmission of information from one node to her neighbors is not correlated with the transmissions from other nodes.

**Corollary 10** (SAGA is asymptotically accurate). Consider the SAGA system (2) of Example 4 and assume \( 1^*W = 1^* \). Then Theorem 3 holds for any \( \gamma \geq \frac{q}{1-q} \).

**Proof.** Note that \( E[L(t)] = qL(W) \); then \( 1^*E[L(t)] = 0 \) by the assumption on \( W \). To apply Theorem 9 we observe that \( a_{\text{max}} = q \) since every node receives exactly one value.

Theorem 11 (Uncorrelated transmissions). Consider system (2) and let \( a_{\text{max}}^c \) be a positive constant such that almost surely \( a_{\text{max}}^c \geq \sum_{i,j \neq} a_{ij} \) for all \( j \in I \). Assume that \( a_{ij} \) and \( a_{kl} \) are uncorrelated if \( l \neq j \). If \( 1^*E(L) = 0 \), then the condition of Theorem 3 holds for any 

\[ \gamma \geq \frac{a_{\text{max}}^c}{1 - a_{\text{max}}}. \]

**Proof.** Let \( A \) be such that \( L = \text{diag}(A1) - A \). Let then \( K = \text{diag}(A^*1) - A^* \). One can verify that \( K \) is the Laplacian of the weighted graph obtained by reversing all edges in the graph encoded by \( L \). In particular, its off-diagonal elements are non-positive, and \( K1 = \text{diag}(A^*1)1 - A^*1 = A^*1 - A^*1 = 0 \).

Observe that 

\[ 1^*K = 1^*\text{diag}(A1^*) - 1^*A^* = 1^*A - 1^*A^* = 1^* - 1^*\text{diag}(A1) = -1^*L, \]

so that \( 1^*E(K) = -1^*E(L) = 0 \). Moreover, the rows of \( K \) are uncorrelated by construction, and \( a_{\text{max}}^c \geq \sum_{i,j \neq} a_{ij} = \sum_{j,j \neq} a_{ji} \) for all \( i \) and all realizations. By the argument of Theorem 9, we have thus 

\[ E(K^*11^*K) \leq a_{\text{max}}^c E(K + K^*). \]

Now, (8) implies that \( K^*11^*K = L^*11^*L \). And, since \( 1^*E(L) = 0 \), there holds 

\[ 1^*E(A) = 1^*\text{diag}(E(A)1) = 1^*E(A^*), \]

so that \( E(A)1 = E(A^*)1 \). As a result, 

\[ E(K + K^*) = E(\text{diag}(A^*1)) - E(A^*) + E(\text{diag}(A1)) - E(A) \]

\[ = E(\text{diag}(A1)) - E(A^*) + E(\text{diag}(A1)) - E(A) \]

\[ = E(L + L^*). \]

The inequality (9) implies thus \( E(L^*11^*L) \leq a_{\text{max}}^c E(L + L^*), \) and the result follows from Lemma 4(iii).

**Example 5** (Reverse Synchronized Asymmetric Gossip Algorithm (RSAGA)). Let graph \( G = (I, W) \) be given, such that \( 1^*W = 1^* \). At each time step, every node \( j \) sends her value to one neighbor \( i_j \), chosen with probability \( W_{ij,j} \). Every node \( i \) then updates her value to 

\[ x_i(t+1) = x_i(t) + q \sum_{j,i_j = i} (x_j(t) - x_i(t)), \]

where \( q \in (0, 1/a_{\text{max}}^c) \) and \( a_{\text{max}}^c \) is the maximum column degree of the graph. In other words, every node sends her value to one of her neighbors, and then updates her value using all the values that she has received.
we observe that a bijection (A biased algorithm) Example 6 showing that systems with numerous and correlated updates may not be asymptotically accurate. There are few simultaneous updates or when the updates are uncorrelated. We now present an example.

We have seen in the previous subsections that asymptotically accurate systems are obtained when values of the first \( n \) and of the latter \( n \) nodes respectively. One can verify that the evolution of \( x_A(t) \) and \( x_B(t) \) and their (common) limit value \( x_\infty \) is actually independent of \( n \). Since the initial variance is \( V(0) = \frac{1}{2N} \sum (x_i(0) - \bar{x}(0))^2 = \frac{1}{2} (x_B(0) - x_A(0))^2 \), the ratio between \( E(x_\infty - \bar{x}(0))^2 \) and \( V(0) \) is also independent of \( n \). In fact, it can be computed that \( E(x_\infty - \bar{x}(0))^2 = \frac{1}{3} V(0) \). The mean square error does thus not decrease when \( n \) grows.

Observe that the columns of the update matrix are uncorrelated in the RSAGA system, as every node chooses independently to whom she is going to send her value. Hence, Theorem 11 implies the following result.

**Corollary 12 (RSAGA is asymptotically accurate).** Consider the RSAGA system of Example 5 and assume that \( W1 = 1 \). Theorem 3 holds for any \( \gamma \geq \frac{q}{1-q d_{\text{max}}^\text{col}} \).

**Proof.** Note that \( E[L(t)] = qL(W/2); \) then \( 1^\ast E[L(t)] = 0 \) by the assumption on \( W \). To apply Theorem 11 we observe that \( a_{\text{max}}^\text{col} = q \) since every node sends her value to exactly one other node, and \( a_{\text{max}}' = a_{\text{max}}^\text{col}q \) since a node can receive up to \( a_{\text{max}}^\text{col} \) values simultaneously.

Observe that Corollaries 10 and 12 ensure that SAGA and RSAGA are asymptotically accurate, while an analysis based on the number of simultaneous updates (that is, on Theorem 6) would suggest values of \( \gamma \) which are proportional to \( N \) and thus do not guarantee asymptotic accuracy.

### 3.4 Simultaneous correlated updates

We have seen in the previous subsections that asymptotically accurate systems are obtained when there are few simultaneous updates or when the updates are uncorrelated. We now present an example showing that systems with numerous and correlated updates may not be asymptotically accurate.

**Example 6 (A biased algorithm).** Consider a system on \( N = 2n \) nodes. At every time step \( t \), one selects a bijection \( f_t : \{1, \ldots, n\} \to \{n+1, \ldots, 2n\} \), i.e. a function \( f_t \) such that, for every \( j \in \{n+1, \ldots, 2n\} \), there exists a unique \( i = f_t^{-1}(j) \in \{1, \ldots, n\} \) for which \( f_t(i) = j \). At every time step, with a probability \( 1/2 \), every node \( i \in \{1, \ldots, n\} \) updates its value by \( x_i(t+1) = \frac{1}{2} (x_i(t) + x_{f_t(i)}(t)) \) while the nodes \( j \in \{n+1, \ldots, 2n\} \) keep their values unchanged; otherwise, every node \( j \in \{n+1, \ldots, 2n\} \) updates its value by \( x_j(t+1) = \frac{1}{2} (x_j(t) + x_{f_t^{-1}(j)}(t)) \), while the nodes \( i \in \{1, \ldots, n\} \) keep their values unchanged. The system clearly preserves the average on expectation. Suppose now that \( x_1(t) = x_2(t) = \cdots = x_{n}(t) \) and \( x_{n+1}(t) = x_{n+2}(t) = \cdots = x_{2n}(t) \) hold for \( t = 0 \). Then, a recurrence argument shows that these equalities hold for all \( t \). Let \( x_A(t) \) and \( x_B(t) \) be the common values of the first \( n \) nodes and of the latter \( n \) nodes respectively. One can verify that the evolution of \( x_A(t) \) and \( x_B(t) \) and their (common) limit value \( x_\infty \) is actually independent of \( n \). Since the initial variance is \( V(0) = \frac{1}{2N} \sum (x_i(0) - \bar{x}(0))^2 = \frac{1}{4} (x_B(0) - x_A(0))^2 \), the ratio between \( E(x_\infty - \bar{x}(0))^2 \) and \( V(0) \) is also independent of \( n \). In fact, it can be computed that \( E(x_\infty - \bar{x}(0))^2 = \frac{1}{3} V(0) \). The mean square error does thus not decrease when \( n \) grows.
However, one should not conclude that every system with unbound ed and not strictly uncorrelated updates must not be asymptotically accurate. In particular, small mean square errors can still occur for systems where the updates follow some more complex probability law, presenting some partial correlations. An example is the following algorithm, which generalizes the BGA and has been proposed in [1].

**Example 7** (Probabilistic Broadcast Gossip Algorithm (PBGA)). Let \( q \in (0, 1) \) and \( G = (I, W) \). At each time step, one node \( j \), sampled from a uniform distribution over \( I \), broadcasts her current value. Every node \( i \) receives the value with a probability \( W_{ij} \in [0, 1] \). When a node \( i \) does receive the value from \( j \), she updates her value to \( x_i(t+1) = x_i(t) + q(x_j(t) - x_i(t)) \).

**Proposition 13** (PBGA is asymptotically accurate). Assume that \( W = W^* \). Then, Theorem 3 holds with
\[
\gamma \geq (W_{\text{max}} + 1) \frac{q}{1 - q},
\]
where \( W_{\text{max}} = \max_{i \in I} \sum_{j \in I} W_{ij} \).

**Proof.** From [1, Lemma 2] we can quickly derive the following formulas for every \( t \geq 0 \),
\[
\begin{align*}
E[L(t)] &= \frac{q}{N} L(W) \\
E[L(t)^* L(t)] &= 2 \frac{q^2}{N} L(W) \\
E[L(t)^* 11^* L(t)] &= \frac{q^2}{N} L(W)^2 + 2 \frac{q^2}{N} L(W) - 2 \frac{q^2}{N} L(W \cdot W),
\end{align*}
\]
where \( W \cdot W \) denotes entrywise product. The assumption on \( W \) implies that \( 1^* \mathbb{E}[L(t)] = 0 \), and in order to apply Theorem 3 we have to find \( \gamma \) which satisfies the inequality
\[
\frac{q^2}{N} L(W)^2 + 2 \frac{q^2}{N} L(W) - 2 \frac{q^2}{N} L(W \cdot W) \leq \gamma \left( 2 \frac{q}{N} L(W) - 2 \frac{q^2}{N} L(W) \right),
\]
that is
\[
L(W)^2 - 2L(W \cdot W) \leq 2 \left( \frac{1 - q}{q} - 1 \right) L(W).
\]
Since any Laplacian – and in particular \( L(W \cdot W) \) – is positive semidefinite, a sufficient condition for the previous inequality to hold is
\[
L(W)^2 \leq 2 \left( \frac{1 - q}{q} - 1 \right) L(W).
\]
As Gershgorin’s disk lemma implies that the spectral radius of \( L(W) \) is not larger than \( 2W_{\text{max}} \), the statement of the result follows. \( \square \)

4 Conclusion and perspectives

We have developed a new way of evaluating the mean square error of decentralized consensus protocols that preserve the average on expectation. Unlike previous approaches, which relied on the convergence speed of these systems, our results are based on the fact that the increase of the error can be bounded proportionally to the decrease of the disagreement. As such, they are independent of the speed at which the system converges, and therefore of the spectral properties of the network, which determines this speed. Notably, many of our bounds only involve local quantities such as the degree of the nodes or the weight that they give to their neighbors’ values, as opposed to global ones such as the eigenvalues of the network Laplacian. As local quantities are much easier to control in distributed systems, our results are of immediate application in design.

Our method can be applied to several known protocols: although we have sometimes been very conservative when deriving our bounds, our method provides bounds that are more accurate than those available in the literature in almost all cases treated, and closely match the experimental results.
from algorithm simulations, capturing the qualitative dependence on the network size. Indeed, our results ensure that, under mild conditions, distributed averaging can be performed via asymmetric and asynchronous algorithms, with a loss in the quality of the estimate which vanishes when increasing the number of samples (and nodes). This fact strongly supports the application of such algorithms to large networks.

In the interest of concision and simplicity, we have limited the number of possible particular cases of our results: there exists thus many possibilities of extending our results to more complex protocols. Overall, two classes of systems were proved to be asymptotically accurate: those with sufficiently few or small simultaneous updates, and those with sufficiently uncorrelated simultaneous updates. These two apparently complementary situations actually present strong similarities; remember indeed that the updates taking place at different times are assumed to be uncorrelated. This suggests that the real parameter determining the mean square error is the level of correlation between the updates taking place across the history of the system. Further work could be devoted to formalizing and quantifying this intuition on the importance of correlations between the updates. Finally, we note that the distribution of final values for processes that do not preserve the average on expectation has, to the best of our knowledge, not been studied so far.

References

[1] T. C. Aysal, A. D. Sarwate, and A. G. Dimakis. Reaching consensus in wireless networks with probabilistic broadcast. In Allerton Conf. on Communications, Control and Computing, pages 732–739, Monticello, IL, September 2009.

[2] 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.

[3] J.-C. Delvenne, R. Carli, and S. Zampieri. Optimal strategies in the average consensus problem. Systems & Control Letters, 58(10-11):759–765, 2009.

[4] 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.

[5] F. Fagnani and J.-C. Delvenne. Democracy in Markov chains and its preservation under local perturbations. In IEEE Conf. on Decision and Control, pages 6620–6625, Atlanta, GA, December 2010.

[6] F. Fagnani and P. Frasca. The asymptotical error of broadcast gossip averaging algorithms. In IFAC World Congress, pages 10027–10031, Milan, Italy, August 2011.

[7] 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.

[8] F. Fagnani and S. Zampieri. Asymmetric randomized gossip algorithms for consensus. In IFAC World Congress, pages 9052–9056, 2008.

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

[10] I. Matei and J. S. Baras. Convergence results for the agreement problem on Markovian random topologies. In IFAC World Congress, Milan, Italy, August 2011.

[11] A. Scaglione, M. Coates, M. Gastpar, J. N. Tsitsiklis, and M. Vetterli. Special issue on gossiping algorithms design and applications. IEEE Journal of Selected Topics in Signal Processing, 5(4):645–648, 2011.

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