Randomized consensus algorithms over large scale networks.

Fabio Fagnani and Sandro Zampieri [*] [†]

Abstract

Various randomized consensus algorithms have been proposed in the literature. In some case randomness is due to the choice of a randomized network communication protocol. In other cases, randomness is simply caused by the potential unpredictability of the environment in which the distributed consensus algorithm is implemented. Conditions ensuring the convergence of these algorithms have already been proposed in the literature. As far as the rate of convergence of such algorithms, two approaches can be proposed. One is based on a mean square analysis, while a second is based on the concept of Lyapunov exponent. In this paper, by some concentration results, we prove that the mean square convergence analysis is the right approach when the number of agents is large.

Differently from the existing literature, in this paper we do not stick to average preserving algorithms. Instead, we allow to reach consensus at a point which may differ from the average of the initial states. The advantage of such algorithms is that they do not require bidirectional communication among agents and thus they apply to more general contexts. Moreover, in many important contexts it is possible to prove that the displacement from the initial average tends to zero, when the number of agents goes to infinity.

1 Introduction

Suppose we have a (directed) graph $\mathcal{G}$ with set of nodes $V = \{1, \dots, N\}$ and a real quantity x_i for every node $i \in V$. The average consensus problem consists of computing the average $x_A = N^{-1} \sum_i x_i$ in an iterative and distributed way, exchanging information among nodes exclusively along the available edges in $\mathcal{G}$. This problem appears in a number of different contexts since the early 80's (decentralized computation [27], load balancing [8, 21, 9]) and, more recently, has attracted much attention for possible applications to sensor networks (data fusion problems [16, 14, 28, 17, 10], clock synchronization [18]) and to coordinated control of mobile autonomous agents [15, 19, 23, 20, 24, 26, 22]. Other places where consensus algorithms have been studied are [13, 3, 5, 4, 6, 11].

Different algorithms for average consensus have been proposed in the literature. They can be distinguished on the basis of the amount of communication and computation they require, of their scalability properties with respect to the number of nodes, of their adaptability to time-varying graphs, and, finally, on the basis of their deterministic or randomized operating protocol.

We now briefly review two of the possible applications of average consensus, to better understand the relevant features of this problem. In load balancing the nodes are processors or computers, and the edges are physical connections between the nodes. The corresponding communication graph presents in general some nice symmetry (e.g. a line, a ring, a torus, a hypercube, etc) and also a symmetry with respect to communication exchange (if i and j are connected by an edge, it means that i can send data to j and viceversa). In many situations the communication graph is fixed. The measurement x_i at each node is in this case the number of tasks which the processor i has to accomplish. The idea is that, in order to speed up the whole computation, processors should exchange tasks along the available edges in order to balance as much as possible the number of tasks among the various processors. The natural goal is to assign to each processor the same number of tasks to work on, namely a number of tasks close to the average x_A. There are two different approaches to this problem. In the first approach each processor evaluates the average x_A by means of an iterative consensus algorithm and, afterwards, there is a transfer of tasks among the processors. In the second approach, instead, the

[*] F. Fagnani is with Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy `fabio.fagnani@polito.it`

[†] S. Zampieri with the Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy `zampi@dei.unipd.it`

transfer of tasks occurs together with the evolution of the averaging algorithm. This procedure forces the algorithm to be inherently symmetric with respect to any pair of communicating processors.

In the context of sensor networks, nodes are sensors which are deployed (often randomly) in some geographical area. They typically communicate in a wireless fashion and, according to the most commonly adopted model, they are allowed to communicate with the sensors located within a certain distance. A good model for the communication graph is in this case the so called "random geometric graph". The quantities x_i that the nodes aim to average can be, in this case, some measurements taken by the sensors (e.g. temperature) and the averaging is done in order to increase precision, by filtering out the noise. In other cases, they need to average an internal state (e.g. cell charge) to obtain aggregate information on the whole net.

One of the key points in these applications is that, both computation and transmission are time and energy consuming tasks, and so their number has to be minimized. Also, it should be pointed out that in many practical applications a node can not simultaneously receive data from two different neighbor nodes (for instance collision can delete messages in wireless communication) and in some applications it cannot simultaneously transmit to more than one node (this happens for instance for processors nets). This fact makes the use of randomized algorithms quite appealing as it turns out that they allow to achieve better performance than deterministic ones with comparable complexity.

In the context of mobile autonomous agents, instead, the consensus problem often takes the form of the so called rendez-vous problem. Here x_i represents the position of node i and the goal is to drive the agents toward their centroid x_A. Mathematically, it appears as a similar problem. While the agents increase their precision in the evaluation of x_A, they also move towards it. However, the analogy is here a bit misleading. If we assume that each agent can only communicate within a given distance, we obtain a 'geometric' graph as for the sensor networks. However, in this case, the graph may change as agents move and this makes the analysis of the dynamics much more difficult. The research direction which inspired the present paper is related to the application context of static sensors and computer networks and not to the mobile agents scenario.

Deterministic (time-invariant and time-varying) consensus algorithms have been studied in many papers. Starting from the pioneering work [27], many variations can be found in the above cited literature. Most of the papers study the same algorithm. Every node runs a first order linear dynamical system to update its estimation and the systems are coupled through the available communication edges. Different schemes (higher order, with memory) however have shown up in the literature, see [9, 21, 6]. The problems typically considered in the literature concern necessary and sufficient conditions for convergence, speed of convergence and optimization issues. On the other hand, random linear schemes have been studied for instance in [16, 4, 10] under the name of gossip algorithms. In this case the evolution matrix of the algorithm changes randomly at every clock step. Convergence is now considered in a probabilistic sense and performance is studied in the mean square sense or in terms of a sort of contraction time. The algorithms studied in the literature assume symmetric communication graphs and lead in general to symmetric evolution matrices which preserve the global average over time. Symmetry is fundamental in certain applications as, for instance, the second approach to load balancing discussed above. However, in other situations, symmetry may not be so important and actually an undesirable constraint in situations where communication is asymmetric (this happens for instance in sensor networks). Also, the related property of achieving exactly the average can be a bit relaxed. Indeed, in some situations it may be sufficient to converge to some value sufficiently close to the average. In the literature also algorithms not converging to a consensus have been considered: see [6].

In this paper we will focus on random first order linear consensus algorithms as in [4]. However, differently from [4] we will not focus exclusively on average consensus. We will consider more general algorithms which do not necessarily converge to the average, but, under certain circumstances, to some good approximation of it. Our theory encompasses a number of examples which will be discussed in detail: asymmetric gossip and broadcasting random algorithms, classic deterministic consensus algorithms in the presence of random edge or node failures. For randomly time-varying algorithms, two different approaches have been proposed in the literature for the convergence analysis. One is based on the the concept of Lyapunov exponent, the second is based on a mean square analysis. The main theoretical contribution of this paper is to show, through a probabilistic concentration result, that, when the number of nodes is large with respect to time, the convergence rate of these algorithms is better described by the mean square analysis, while, when instead the number of nodes is small with respect to time, it is the Lyapunov exponent analysis which provides the right convergence description.

Therefore for large scale networks, namely for networks with a very large number of nodes, it is the mean square analysis to provide the most meaningful approach to the performance optimization.

We now briefly outline the content of the various sections. In Section 2 we formally describe the consensus problem in a random setting and we present an example which highlights that the Lyapunov exponent analysis may be not right tool for the proposed scopes. In Section 3 we propose a simple sufficient condition for a random algorithm to reach consensus. The mathematical core is a result by Cogburn [7]. We then describe various examples showing how this sufficient condition can be applied. Section 4 is devoted to the mean square analysis. For some specific examples, part of this has already appeared in the literature [4, 5, 11], but never in such a generality. We believe that this is one of the most important tool for the analysis and design of these algorithms. The proposed results are of potential application in the various examples presented in the paper. However, concrete application to such examples would require extra computational effort which is not done in this paper. Section 5 contains the most original theoretical results. Using Azuma's inequality for martingales we prove some concentration results on the algorithm performance. These results show that, for a number of important examples, if the time range for which we run our algorithms is sufficiently small with respect to the number of nodes, then the evolution tends to concentrate around their mean. This gives a strong motivation for pursuing the mean square analysis. An appendix devoted to the Lyapunov exponent analysis completes the paper.

Notation We introduce now some notation on linear algebra which will be used in the sequel. The symbols $\mathbb{R}^N$ and $\mathbb{R}^{M\times N}$ will denote the vector space of N dimensional column vectors and of $M \times N$ matrices with real entries. If M is a real valued matrix, M^* will denote the transpose of M. The symbol $\mathbb{1}$ will denote the column vector in $\mathbb{R}^N$ having all entries equal to 1 will e_i, $i = 1, \dots, N$ denotes the column vector in $\mathbb{R}^N$ having all entries equal to 0 except a 1 in position i. The symbol I denotes the identity matrix whose dimension is typically deducible from the context. Since in the sequel we will make frequent use of the $N \times N$ matrix $I - N^{-1}\mathbb{1}\mathbb{1}^*$, we will denote such a matrix with the symbol Ω. A matrix $P \in \mathbb{R}^{N\times N}$ is called a stochastic matrix if it has nonnegative entries and if $P(t)\mathbb{1} = \mathbb{1}$. A stochastic matrix P is called doubly stochastic if $\mathbb{1}^*P = \mathbb{1}^*$. The symbol $||\cdot||$ will denote the 2-norm on $\mathbb{R}^N$, and the induced 2-norm in $\mathbb{R}^{N\times N}$. The symbol $\mathrm{sr}(\cdot)$ means the spectral radius of a linear map, namely the maximum absolute value of its eigenvalues. Given a matrix $P \in \mathbb{R}^{N\times N}$, $\mathrm{diag}(P)$ will denote the diagonal part of P namely a diagonal matrix with the same diagonal entries of P. Moreover we define $\mathrm{out}(P) = P - \mathrm{diag}(P)$. A (directed) graph $\mathcal{G}$ is defined as a pair (V, E) where $V = \{1, \dots, N\}$ and $E \subseteq V \times V$. The set V is called the set of vertices and the set E is called the set of edges. If E if such that $(i, j) \in E$ implies $(i, j) \in E$, then the associated graph is called undirected.

2 Problem formulation

2.1 Linear consensus algorithms

The iterative consensus algorithms considered in this paper consist of N coupled linear dynamical systems

$$x_i(t+1) = \sum_{j=1}^{N} P_{ij}(t)x_j(t) \qquad i = 1, \dots, N$$

where $x_i(t) \in \mathbb{R}$ is the state of the i-th system at time t and $P_{ij}(t) \in \mathbb{R}$ coefficients which vary with the time t.

More compactly we can write

$$x(t+1) = P(t)x(t)\,, \tag{1}$$

where $x(t) \in \mathbb{R}^N$ and $P(t) \in \mathbb{R}^{N\times N}$. The sequence $P(t)$ is said to achieve the *consensus* if the following conditions are satisfied

(a) If $x(0) = \alpha\mathbb{1}$, where $\alpha \in \mathbb{R}$, then $x(t) = x(0)$ for every $t \in \mathbb{N}$.

(b) For any $x(0) \in \mathbb{R}^N$, there exists $\alpha \in \mathbb{R}$ such that

$$\lim_{t\to\infty} x(t) = \alpha\mathbb{1} \tag{2}$$

Moreover, if $\alpha = N^{-1}\mathbb{1}^* x(0)$, we say that *average consensus* is achieved.

In this paper we will assume to have statistical information on the matrices $P(t)$ and we will adopt a probabilistic approach to the problem instead of a worst case analysis considered in [27, 15, 20, 24, 3]. More precisely, in this paper we will assume that $P(t)$ is a sequence of i.i.d. matrix valued random variables and $x(t)$ is the stochastic process which is the solution of the equation (1). We say that the sequence $P(t)$ achieves the *probabilistic consensus* if condition (a) above holds while (b) is replaced by

(b') For any $x(0) \in \mathbb{R}^N$, there exists a scalar random variable α such that

$$\lim_{t\to\infty} x(t) = \alpha\mathbb{1} \quad \text{almost surely.} \tag{3}$$

If $\alpha = N^{-1}\mathbb{1}^* x(0)$ almost surely, we talk about *probabilistic average consensus*.

In this paper we will restrict to cases in which $P(t)$ are stochastic matrices. Notice that condition (a) is then clearly automatically satisfied. If, moreover, $P(t)$ is doubly stochastic, then the average is invariant, namely $N^{-1}\mathbb{1}^* x(t) = N^{-1}\mathbb{1}^* x(0)$ for every t and hence in this case, consensus implies average consensus.

Let

$$Q(t) = P(t-1)\cdots P(0)\,, \tag{4}$$

so that we can write $x(t) = Q(t)x(0)$. The random variable α in (3) is a linear function of the initial condition $x(0)$ so that we can write $\alpha = \rho^* x(0)$ for some random variable ρ taking values in $\mathbb{R}^N$ and such that $\mathbb{1}^*\rho = 1$. Therefore probabilistic consensus can be equivalently expressed by saying that there exists a random variable ρ taking values in $\mathbb{R}^N$ such that

$$\lim_{t\to\infty} Q(t) = \mathbb{1}\rho^* \tag{5}$$

almost surely. Notice that $\mathbb{1}\rho^*$ is a matrix whose rows are all equal to ρ^*. Notice that $x(\infty) = \rho^* x(0)$. We have probabilistic average consensus exactly when $\rho = N^{-1}\mathbb{1}$ almost surely.

2.2 Constraints on the algorithm: the communication graph

Given a matrix P of dimension $N \times N$, we can consider the directed graph $\mathcal{G}_P = (V, E)$ where $V = \{1, \dots, N\}$ and $E \subseteq V \times V$ is defined by

$$(j, i) \in E \;\Leftrightarrow\; P_{ij} \neq 0$$

$\mathcal{G}_P$ is called the directed graph associated with P. The graph $\mathcal{G}_{\mathrm{out}(P)}$ associated with $\mathrm{out}(P)$ is simply $\mathcal{G}_P$ without all possible self-loops.

Suppose we use a consensus algorithm $P(t)$. At time t this algorithms needs to use communications among all the edges of $\mathcal{G}_{\mathrm{out}(P(t))}$. The amount of non zero elements in $\mathrm{out}(P(t))$ is thus a measure of the number of communications that simultaneously have to take place in our network to implement such a scheme.

In many circumstances there is an a priori fixed communication skeleton, namely a fixed underlying directed graph $\mathcal{G} = (V, E)$, establishing which are the feasible communications among agents. We will say that the scheme $P(t)$ is adapted to $\mathcal{G}$ if $\mathcal{G}_{\mathrm{out}(P(t))}$ is a subgraph of $\mathcal{G}$ for every instant t.

For future use, we define some basic notation for graphs. Consider a directed graph $\mathcal{G} = (V, E)$ where $V = \{1, \dots, N\}$ and $E \subseteq V \times V$. For every $i \in V$ we put

$$N_i^+ = \{j \in V \setminus \{i\} \mid (i, j) \in E\}\,, \quad N_i^- = \{j \in V \setminus \{i\} \mid (j, i) \in E\}$$

Elements in N_i^+ (resp. in N_i^-) are called out-neighbors (resp. in-neighbors) of i. Moreover we put $\nu_i^+ = |N_i^+|$, $\nu_i^- = |N_i^-|$ which are called, respectively, the out-degree and the in-degree of the node i. Let e_i be the i-th element of the canonical basis of $\mathbb{R}^N$. The adjacency matrix of $\mathcal{G}$ is defined as

$$A_{\mathcal{G}} = \sum_{i\in V}\sum_{j\in N_i^+} e_i e_j^* = \sum_{i\in V}\sum_{j\in N_i^-} e_j e_i^* \tag{6}$$

namely $(A_{\mathcal{G}})_{ij} = 1$ if $(i,j) \in E$, otherwise it is zero. Notice that possible self loops have not been considered. This will make notation simpler further on. Notice moreover that

$$\begin{bmatrix} \nu_1^+ \\ \vdots \\ \nu_N^+ \end{bmatrix} = A_{\mathcal{G}}\mathbb{1} \qquad \text{and} \qquad \begin{bmatrix} \nu_1^- & \cdots & \nu_N^- \end{bmatrix} = \mathbb{1}^* A_{\mathcal{G}}$$

The out-degree and in-degree matrices are defined, respectively as

$$D_{\mathcal{G}+} = \sum_i \nu_i^+ e_i e_i^* \qquad D_{\mathcal{G}-} = \sum_i \nu_i^- e_i e_i^* \tag{7}$$

namely $D_{\mathcal{G}+}$ and $D_{\mathcal{G}-}$ are diagonal matrices with diagonal entries equal to ν_i^+ and ν_i^-, respectively. If the graph undirected, then $D_{\mathcal{G}+} = D_{\mathcal{G}-}$. In this case we will drop the superscript $\pm$ in the above notations.

A graph $\mathcal{G} = (V, E)$ is said to be strongly connected if for any two vertices $v_1, v_2 \in V$ there always exist a walk in $\mathcal{G}$ connecting v_1 to v_2.

We conclude this paragraph by introducing a first example of a random consensus algorithm. More examples will be presented in Section 3.

Example 2.1: We start from a undirected graph $\mathcal{G} = (V, E)$ and we assume that at every time instant t a node i is chosen randomly among the N possible nodes with probability $1/N$. This node then chooses randomly a node j among its ν_i neighbors with probability $1/\nu_i$, it establishes a bidirectional link with it. Finally i and j average their states $x_i(t)$ and $x_j(t)$. More precisely, let, for every $(i,j) \in E$,

$$R^{ij} := I - \frac{1}{2}(e_i - e_j)(e_i - e_j)^*$$

Then, $P(t)$ is concentrated on these matrices and

$$\mathbb{P}[P(t) = R^{ij}] = \mathbb{P}[P(t) = R^{ji}] = \frac{1}{N}\left[\frac{1}{\nu_i} + \frac{1}{\nu_j}\right]$$

We will see in Section 3 that, if $\mathcal{G}$ is strongly connected, then this algorithm always leads to probabilistic average consensus.

2.3 The measure of performance

In this section we assume that we have a fixed random algorithm $P(t)$ achieving probabilistic consensus so that (5) is satisfied with a suitable ρ.

In this paper, we will measure the performance of the algorithm $P(t)$ by considering two figures. The first figure we consider is a normalized version of the *distance from the consensus*

$$d(t) = \frac{1}{N}||x(t) - \mathbb{1}x_A(t)||^2 = \frac{1}{N}\sum_{i=1}^{N} |x_i(t) - x_A(t)|^2$$

where $x_A(t) = N^{-1}\mathbb{1}^* x(t)$ is the average of the components of $x(t)$. The second one is the *average displacement* from its initial value

$$\beta(t) = |x_A(t) - x_A(0)|^2$$

Of course in those situations where $P(t)$ is always doubly stochastic, we have that $\beta(t) = 0$ for every t. Notice moreover that

$$\frac{1}{N}||x(t) - \mathbb{1}x_A(0)||^2 = d(t) + \beta(t)$$

which shows that the evolution of $d(t)$ and $\beta(t)$ determines the evolution of $\frac{1}{N}||x(t) - \mathbb{1}x_A(0)||^2$. This coincides with the average distance from $x_i(t)$ to $x_A(0)$ and so it is the most important error parameter that typically one has to minimize.

We know that $d(t)$ converges to 0 almost surely. The main point is to evaluate the speed of this convergence. It turns out that convergence is always of exponential type, hence the interest is in evaluating the exponential rate of convergence for $t \to +\infty$. The reader could wonder why we did not consider, instead of $d(t)$, the apparently more natural $N^{-1}||x(t) - x(\infty)||$. The reason is that $d(t)$ is simpler to be analyzed while they are strictly linked to each other. This is shown in the next result which will useful later on.

Lemma 2.2. *The following inequalities hold, for any $x(0) \in \mathbb{R}^N$ and for any t*

$$d(t) \leq \frac{1}{N}||x(t) - x(\infty)||^2 \leq (1+\sqrt{N})^2 d(t) \tag{8}$$

Proof. Inequality on the left immediately follows from the identity

$$x(t) - \mathbb{1}x_A(t) = \Omega x(t) = \Omega(x(t) - \mathbb{1}\rho^* x(0))$$

and the fact that $||\Omega|| = 1$. To prove the right inequality, we start from

$$Q(t) = \Omega Q(t) + N^{-1}\mathbb{1}\mathbb{1}^* Q(t) \tag{9}$$

and

$$Q(s+t) = P(s+t-1)\cdots P(t)\Omega Q(t) + N^{-1}\mathbb{1}\mathbb{1}^* Q(t) \tag{10}$$

From (9) and (10) we finally obtain that

$$\begin{aligned} ||(Q(t) - Q(s+t))x(0)|| &= ||(I - P(s+t-1)\cdots P(t))\Omega Q(t)x(0)|| \\ &\leq (1+\sqrt{N})||x(t) - \mathbb{1}x_A(t)|| \end{aligned}$$

where the last inequality follows from the fact that the 2-norm of an $N \times N$ stochastic matrix is always $\leq \sqrt{N}$. By letting $s \to \infty$ we obtain the inequality on the right. ■

This result in particular implies that the exponential rate of convergence (formally defined below) of $||x(t) - \mathbb{1}x_A(t)||$ and of $||x(t) - x(\infty)||$ coincide.

Since,

$$d(t) = \frac{1}{N}||Q(t)\Omega x(0)||^2\,,$$

it is quite intuitive that the rate of convergence of $d(t)$ should be related to the second Lyapunov exponent of the sequence of random matrices $P(t)$ [25, 2] (the first Lyapunov exponent is clearly equal to one). In the appendix we will prove rigorously this fact, showing in particular that the rate of convergence of $d(t)$ is a constant, as formally described the the following theorem.

Theorem 2.3. *For almost every initial condition $x(0)$ (with respect to the standard N-dimensional Lebesgue measure), we have that*

$$\lim_{t\to+\infty} d(t)^{1/t} = \lambda^2\,, \quad \text{almost surely}\,, \tag{11}$$

where λ is a constant.

This result seems to suggest that the performance analysis of our models could simply rely on the classical theory of linear randomly switching systems [2] and to the computation of Lyapunov exponents. It is well known that computing Lyapunov exponents is a very difficult task in general, even though in particular situations it can be done. However, there is another much more important reason why the Lyapunov exponents analysis is not the right approach to the performance optimization problem. This is well illustrated in the following example.

Example 2.4: Consider the algorithm introduced in Example 2.1 in the special case when $\mathcal{G}$ is the complete graph and $N = 2^r$, for some $r \in \mathbb{N}$. In this case we have that

$$\lim_{t\to+\infty} d(t)^{1/t} = 0\,, \quad \text{almost surely}$$

To prove this fact first observe that, for every r, there exists a family of edges $(i_1, j_1), \cdots, (i_s, j_s) \in E$ such that

$$R^{i_1,j_1}\cdots R^{i_s,j_s} = N^{-1}\mathbb{1}\mathbb{1}^* \in \mathbb{R}^{N\times N} \tag{12}$$

Indeed, reasoning by induction on r, this fact is clearly true for $r = 1$. We divide the set of nodes N of $\mathcal{G}$ into two parts $N' = \{1, \ldots, N/2\}$ and $N'' = \{N/2+1, \ldots, N\}$. By induction we know that there exists $(i'_1, j'_1), \cdots, (i'_s, j'_s) \in N' \times N'$ and $(i''_1, j''_1), \cdots, (i''_s, j''_s) \in N'' \times N''$ such that

$$R^{i'_1,j'_1}\cdots R^{i'_s,j'_s} = \begin{pmatrix} (N/2)^{-1}\mathbb{1}\mathbb{1}^* & 0 \\ 0 & I \end{pmatrix} \qquad R^{i''_1,j''_1}\cdots R^{i''_s,j''_s} = \begin{pmatrix} I & 0 \\ 0 & (N/2)^{-1}\mathbb{1}\mathbb{1}^* \end{pmatrix}$$

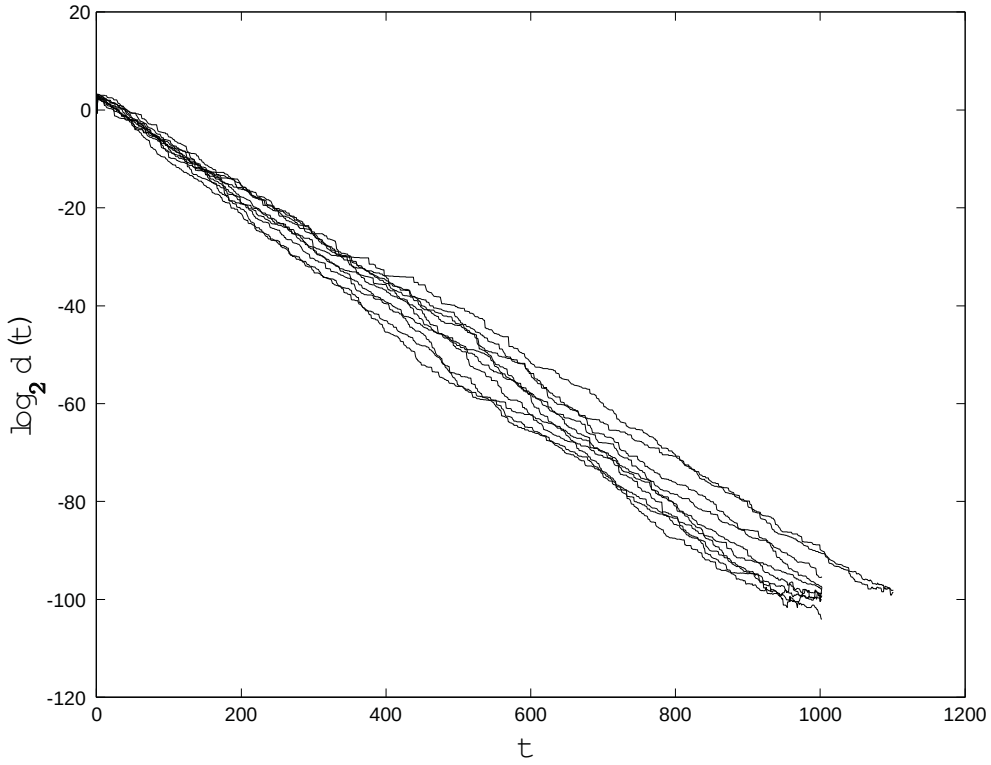

Figure 1: This figure shows the graphs of $\log_2 d(t)$ as a function of time t in 7 different simulations of the consensus algorithm presented in Example 2.4 in case $N = 16$.

Observe now that

$$R^{1,N/2+1}R^{2,N/2+2}\cdots R^{N/2,N} = \frac{1}{2}\begin{pmatrix} I & I \\ I & I \end{pmatrix}$$

and so we can argue that

$$R^{1,N/2+1}R^{2,N/2+2}\cdots R^{N/2,N}R^{i'_1,j'_1}\cdots R^{i'_s,j'_s}R^{i''_1,j''_1}\cdots R^{i''_s,j''_s} = N^{-1}\mathbb{1}\mathbb{1}^*$$

From the previous reasoning it follows that the consensus is achieved in finite time whenever the finite sequence of matrices $R^{i_1,j_1},\ldots,R^{i_s,j_s}$ appearing in (12) shows up in some point of the sequence $P(t)$. In all these cases $d(t)^{1/t} \to 0$. Since the probability of appearance of such event is small but strictly greater than 0, it follows from Theorem 11 that in this case $\lambda = 0$. Notice that, if we denote by T_r the number of matrices contained in the sequence in (12), it follows from the iterative construction above that $T_r = 2T_{r-1} + 2^{r-1}$ so $T_r = r2^{r-1}$.

A numerical simulation for the $d(t)$ in this case yields the behavior showed in Figure 1. The long time behavior is determined by the exponent $\lambda = 0$. Indeed, consensus is achieved in finite time with probability one. It is however evident from the graphs that, before achieving consensus, there is a transient behavior which is indeed exponential with rate $\simeq 0.93$. Clearly this will have an important role in the overall performance, since the time range in which this behavior exists is quite large. It can be seen that this time range increases with N. If we are interested in analyzing the behavior of the algorithm when N is big with respect to time, this behavior becomes the most significant. More quantitative conclusions will follow from the results in the rest of the paper.

3 Random algorithms achieving consensus

In this section we will first give some conditions which ensures the probabilistic consensus. Then we will present some examples of randomized consensus algorithms and we will prove that in all these cases the probabilistic consensus occurs.

3.1 Conditions for the probabilistic consensus

The conditions ensuring consensus in the deterministic case, namely when $P(t) = P$ for all t, are quite well-known [5] and are essentially related to the Perron Frobenius theorem applied to stochastic matrices. A simple sufficient condition is that $\mathcal{G}_P$ is strongly connected and its edge set contains all the self-loops (namely all the diagonal entries of P are positive). In this case P is a so-called aperiodic stochastic matrix and 1 is a simple eigenvalue and the remaining eigenvalues have norms strictly smaller than 1.

If we start from any strongly connected directed graph $\mathcal{G}$, a possible stochastic matrix P yielding $\mathcal{G}_P = \mathcal{G}$ is given by

$$P = qI + (1-q)D_{\mathcal{G}-}^{-1} A_{\mathcal{G}}^{*},$$

where $q \in (0,1)$ is an arbitrarily chosen parameter and where we used the notation introduced in (6) and (7). A slightly more complicated construction [5] actually allows to obtain a doubly stochastic matrix P such that average consensus is indeed achieved. In the simpler case when $\mathcal{G}$ is strongly connected and undirected, however, we can construct a symmetric stochastic matrix P by taking

$$0 < P_{ij} = P_{ji} < \min\{\nu_i^{-1}, \nu_j^{-1}\} \qquad \forall (i,j) \in E$$

$$P_{ij} = 0 \qquad \forall (i,j) \notin E$$

$$P_{ii} = 1 - \sum_j P_{ij}$$

It is well-known [23] that for all stochastic matrices P yielding consensus, the rate of convergence to the consensus is given by the essential spectral radius of P which is defined as follows

$$\mathrm{esr}(P) = \max\{|\lambda| : \lambda \text{ eigenvalue of } P \text{ different from } 1\} \tag{13}$$

More precisely we have that

$$\lim_{t\to\infty} d(t)^{1/t} = \mathrm{esr}(P)^2$$

Probabilistic consensus turns out to be an easily checkable property, namely as easily checkable as the deterministic consensus in the time-invariant case. The following result appears in [7].

Theorem 3.1. *The algorithm $P(t)$ achieves probabilistic consensus if and only if for every $i, j \in V$ we have that*

$$\mathbb{P}[\mathcal{E}_{ij}] = 1$$

where

$$\mathcal{E}_{ij} = \{\exists k\,,\ \exists t\ |Q_{ik}(t)Q_{jk}(t) > 0\}$$

To obtain a more handy condition, we need to impose a hypothesis which is however always satisfied in all the cases which are commonly considered in the literature. In the following result we will assume that all the diagonal elements of $P(t)$ are nonzero with probability 1. This simple condition ensuring probabilistic consensus is based on the expected value of $P(t)$ which will be denoted as $\overline{P} = \mathbb{E}[P(t)]$. We have the following result.

Corollary 3.2. *Assume that for any $i \in V$ we have that $P(t)_{ii} > 0$ almost surely. If $\mathcal{G}_{\overline{P}}$ is strongly connected, then $P(t)$ achieves probabilistic consensus.*

Proof. Let $i, j \in V$. Since $\mathcal{G}_{\overline{P}}$ is strongly connected, then there exists t such that $\overline{P}_{ij}^t > 0$. Consider the event

$$\mathcal{A}_{ij}(t) = \{Q_{ij}(t) > 0\}$$

Since $\overline{P}_{ij}^t > 0$, then $\mathbb{P}[\mathcal{A}_{ij}(t)] > 0$. Moreover, since

$$Q_{jj}(t) \geq P_{jj}(t-1) \cdots P_{jj}(1) P_{jj}(0)$$

then the fact that $P(t)_{jj} > 0$ almost surely ensures that $\mathbb{P}[\mathcal{A}_{jj}(t)] = 1$. Observe now that $\mathcal{A}_{ij}(t) \cap \mathcal{A}_{jj}(t) \subseteq \mathcal{E}_{ij}$, which implies that $\mathbb{P}[\mathcal{E}_{ij}] > 0$. On the other hand, it is immediate to check that $\mathcal{E}_{ij}$ is a tail event. Hence, for the Kolmogorov $0-1$ law, we must have $\mathbb{P}[\mathcal{E}_{ij}] = 1$. By Theorem 3.1, the thesis follows. ∎

Remark 3.3. Notice that there is a sort of weak converse to previous result. Indeed, assume $P(t)$ achieves probabilistic consensus. From the almost sure convergence $Q(t) \longrightarrow \mathbb{1}\rho^*$ and Lebesgue dominated convergence theorem it follows that $\mathbb{E}[Q(t)] \longrightarrow \mathbb{1}\mathbb{E}[\rho]^*$ and so $\overline{P}^t \longrightarrow \mathbb{1}\mathbb{E}[\rho]^*$. In other words, $\overline{P}$ achieves consensus. It can very well happen that $E\rho = \mathbb{1}$ even if ρ is not equal to $\mathbb{1}$ almost surely. In other terms, even if $\overline{P}$ achieves average consensus, not necessarily $P(t)$ will also achieve average probabilistic consensus. This will appear in the examples we will propose.

3.2 Examples of random consensus algorithms

We now present a number of examples of randomized consensus algorithms.

Example 3.4: The symmetric gossip model This algorithm generalizes Example 2.1. In a slightly more general way it can be introduced as follows. Fix a real number $q \in (0,1)$, an undirected graph $\mathcal{G} = (V, E)$ and a symmetric $N \times N$ matrix W with nonnegative entries adapted to $\mathcal{G}$ such that $\mathbb{1}^* W \mathbb{1} = 1$. At every time instant t the edge (j,i) is activated with probability W_{ij} and nodes i and j exchange their states and produce a new states according to the equations

$$\begin{aligned} x_i(t+1) &= (1-q)x_i(t) + qx_j(t) \\ x_j(t+1) &= qx_i(t) + (1-q)x_j(t) \end{aligned}$$

The other states remains unchanged. More formally for every $(i,j) \in E$, we let

$$R^{ij} = I - q(e_i - e_j)(e_i - e_j)^* ,$$

Then, $P(t)$ is concentrated on these matrices and

$$\mathbb{P}[P(t) = R^{ij}] = W_{ij}$$

We have that

$$\overline{P} = \sum_{(i,j)\in E} W_{ij} R^{ij}$$

Notice that, $\mathrm{out}(\overline{P}) = 2q\mathrm{out}(W)$. Since $\overline{P}$ is stochastic, this condition completely determines $\overline{P}$. Both W and q can be considered in principle as design parameters respect to which one can optimize the performance. Notice that, if $\mathcal{G}_W$ is strongly connected, then $\mathcal{G}_{\overline{P}}$ is automatically strongly connected. Moreover in this case all the diagonal elements of $P(t)$ are nonzero with probability 1. Applying Corollary 3.2 we can conclude that this algorithm yields the probabilistic consensus. Moreover, since the $P(t)$ are all symmetric we can conclude that in this case probabilistic average consensus is achieved. Notice finally that the case considered in Example 2.1 can be recovered by taking

$$W_{ij} = \frac{1}{2N}\left(\frac{1}{\nu_i} + \frac{1}{\nu_j}\right) , \; i \neq j , \quad q = \frac{1}{2}$$

Example 3.5: The asymmetric-gossip model In this case we start from a real number $q \in (0,1)$, a fixed directed graph $\mathcal{G} = (V, E)$ and an $N \times N$ matrix W with nonnegative entries adapted to $\mathcal{G}$ such that $\mathbb{1}^* W \mathbb{1} = 1$. At every time instant t the edge (j,i) is activated with probability W_{ij} and node j sends its state to i and i produces a new state according to the equation

$$x_i(t+1) = (1-q)x_i(t) + qx_j(t)$$

Formally, define, for every $(i,j) \in E$, $R^{ij} = I - qe_i(e_i - e_j)^*$ and let

$$\mathbb{P}[P(t) = R^{ij}] = W_{ij}$$

Notice that in this case $\mathrm{out}(\overline{P}) = q\mathrm{out}(W)$. Arguing as in previous example, we can apply Corollary 3.2 and conclude that this algorithm yields the probabilistic consensus. Since the matrices $P(t)$ are not doubly stochastic, average probabilistic consensus is not achieved in this case. However, $\overline{P}$ is doubly stochastic if W has been chosen such that $W^* \mathbb{1} = W\mathbb{1}$.

From the point of view of the practical implementation of this type of algorithms, an important issue is the possibility to parallelize a number of the atomic gossip averaging steps in order to fully use the power of the large scale network. This issue has been already investigated in [4] for the symmetric gossip case. The idea is to activate simultaneously a maximal number M of disconnected edges in the graph ('almost matching') and to perform the gossip symmetric algorithm simultaneously on all these edges. Since the corresponding matrices R^{ij} in this case commute, the final effect is equivalent to the effect of M consecutive application of the asymmetric gossip algorithm. What can change is only the statistics due to the way the 'almost matching' is selected.

This idea also works for the asymmetric gossip case. In this last case, however, there is another appealing possibility: at every time instant, each agent chooses a neighbor from which to get data and average with its own. Since in this case the matrices R^{ij} do not commute, this model is not equivalent to any possible serialized version of the asymmetric gossip algorithm, and we thus consider it as an autonomous example (see [5] for further details).

Example 3.6: The synchronous asymmetric gossip model We start from a real number $q \in (0,1)$, a directed graph $\mathcal{G} = (V, E)$ and an $N \times N$ stochastic matrix W adapted to $\mathcal{G}$. At every time instant t the N edges $(i, j_i) \in E$ for $i = 1, 2, \ldots, N$ are activated each with probability W_{i,j_i}. The node j_i sends its state to i and each i produces a new state according to the equation

$$x_i(t+1) \quad = \quad (1-q)x_i(t) + qx_{j_i}(t)$$

More formally fix for every $\mathbf{j} = (j_1, \ldots, j_N) \in V^N$ the matrix $R^{\mathbf{j}} = (1-q)I + q\sum_i e_i e_{j_i}^*$ and let

$$\mathbb{P}[P(t) = R^{\mathbf{j}}] = \prod_{i=1}^{N} W_{i,j_i}$$

In this case we obtain

$$\overline{P} = \sum_{\mathbf{j}} \prod_{i=1}^{N} W_{i,j_i} R^{\mathbf{j}} = (1-q)I + qW$$

As in previous examples, if $\mathcal{G}_W$ is strongly connected, we can apply Corollary 3.2 and conclude that this algorithm yields the probabilistic consensus. As in previous example the matrices $P(t)$ are in general not doubly stochastic and so the average probabilistic consensus is not achieved. Notice moreover that $\overline{P}$ is doubly stochastic if and only if W is doubly stochastic.

We now present a further example of a different nature.

Example 3.7: The broadcasting model We start from any directed graph $\mathcal{G} = (V, E)$ and we fix a vector $w = (w_1, \ldots, w_N)^*$ with $w_i \geq 0$ and $\sum w_i = 1$. We assume that at every time instant node i is chosen with probability w_i. This node i then broadcasts its state to all its out-neighbors which then average of their states with the received state. In this case $P(t)$ concentrated on the N matrices

$$R^i = I - q \sum_{j \in N_i^+} (e_j e_j^* - e_j e_i^*)$$

and we let $\mathbb{P}[P(t) = R^i] = w_i$. Hence

$$\overline{P} = I - q \sum_i w_i \sum_{j \in N_i^+} (e_j e_j^* - e_j e_i^*) = I - q[D - A_{\mathcal{G}}^* W]$$

where W is a diagonal matrix such that $W_{ii} = w_i$ for every i and D is a diagonal matrix such that $D_{ii} = \sum_{j \in N_i^-} w_j$. If $\mathcal{G}$ is strongly connected and each $w_i > 0$, automatically $\mathcal{G}_{\overline{P}}$ is strongly connected Moreover, all the diagonal elements of $P(t)$ are nonzero with probability 1. Applying Corollary 3.2 we can conclude that this algorithm yields the probabilistic consensus. The matrices $P(t)$ are never doubly stochastic. Notice that $\overline{P}$ is doubly stochastic if and only if

$$w^*(A_{\mathcal{G}} - D_{\mathcal{G}+}) = 0$$

In the special case when $w_i = N^{-1}$ for every i we obtain

$$\overline{P} = I - \frac{q}{N}[D_{\mathcal{G}-} - A_{\mathcal{G}}^*]$$

In this case $\overline{P}$ is doubly stochastic if and only if $D_{\mathcal{G}-} = D_{\mathcal{G}+}$.

3.3 Deterministic algorithms in a random media

As already mentioned, randomness can also be due environmental effects. In this section we describe some possible models for such situations.

Example 3.8: The packet drop model We start from a fixed stochastic matrix P such that $P_{ii} > 0$ for all i and such that the graph $\mathcal{G} = \mathcal{G}_P$ is strongly connected. Then we know that the algorithm

$$x(t+1) = Px(t) \tag{14}$$

yields consensus. In some situations there might be data loss in the communication between nodes. Here we model this by assuming that the data transmission over an edge (j,i) of $\mathcal{G}$ from the node j to the node i can occur with some probability p. More precisely, consider the family of independent binary random variables $L_{ij}(t)$, $t \in \mathbb{N}$, $i,j = 1,\dots,N$, $i \neq j$, such that

$$\mathbb{P}[L_{ij}(t) = 1] = p \quad \text{and} \quad \mathbb{P}[L_{ij}(t) = 0] = 1 - p \quad \text{if } i \neq j$$

We emphasize the fact that independence is assumed among all $L_{ij}(t)$ as i,j and t vary. Consider the random matrix $\tilde{A}(t)$ defined by $\tilde{A}_{ij}(t) = (A_{\mathcal{G}})_{ij} L_{ij}(t)$. Clearly, $\tilde{A}(t)$ is the adjacency matrix of a random graph $\tilde{\mathcal{G}}(t)$ obtained from $\mathcal{G}$ by deleting the edge (i,j) when $L_{ij}(t) = 0$. In general our initial matrix P is not going to be compatible with the graph $\tilde{\mathcal{G}}(t)$ and, as a consequence, the consensus algorithm has to be modified to consider this fact.

There are several ways to adapt P in order to take into account the missing data. We here describe one possibility and we refer to [12] for more considerations on this issue. We consider the following updating equation

$$x_i(t+1) = \left(P_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_{ij}(1 - L_{ij}(t)) \right) x_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_{ij} L_{ij}(t) x_j(t)$$

Roughly speaking, according to this method the weights P_{ij} are kept constant if $i \neq j$ while they are varied if $i = j$ in order to keep the stochasticity of $P(t)$. It is clear from the construction that, since P has nonzero diagonal elements, then $P(t)$ has nonzero diagonal elements with probability one. Observe moreover that, if $i \neq j$, then $\overline{P}_{ij} = pP_{ij}$ and so $\text{out}(\overline{P}) = p\text{out}(P)$. This fact ensures that since $\mathcal{G}_P$ is strongly connected, then $\mathcal{G}_{\overline{P}}$ is strongly connected as well. Applying Corollary 3.2 we can conclude that this algorithm yields the probabilistic consensus. Notice finally that, if P is doubly stochastic, then also $\overline{P}$ is doubly stochastic.

Example 3.9: The node failure model We start from the same model (14), but we now assume that at each time instant t there is some probability $1 - p$ that the node i fails and is unable to transmit. Formally, let $L_j(t)$, $t \in \mathbb{N}$, $j = 1,\dots,N$ be a family of independent binary random variables such that $\mathbb{P}[L_j(t) = 1] = p$. The event $\{L_j(t) = 1\}$ corresponds to have the node j active at time t. In this case, one possibility to adapt P to the graph constituted by the active nodes, similar to the choice in Example 3.8, is

$$x_i(t+1) = \left(P_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_{ij}(1 - L_j(t)) \right) x_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_{ij} L_j(t) x_j(t)$$

This model has the same properties of the model proposed before. In particular, since $P(t)$ has nonzero diagonal elements with probability one, and $\text{out}(\overline{P}) = p\text{out}(P)$, this algorithm yields the probabilistic consensus. Notice finally that, if P is doubly stochastic, then also $\overline{P}$ is doubly stochastic.

4 The mean square performance

In this section we will analyze the expectations of the two variables $d(t)$ and $\beta(t)$.

4.1 The distance from the consensus

We are interested in studying $\mathbb{E}[d(t)]$ and, in particular, its exponential rate of convergence

$$R = \sup_{x(0)} \limsup_{t \to +\infty} \mathbb{E}[d(t)]^{1/t} \tag{15}$$

A straightforward application of Jensen inequality yields $R \geq \lambda^2$ where λ^2 was defined in (11).

In the sequel, we characterize R as the spectral radius of a N^2-dimensional linear stochastic operator and we establish some useful bounds for it.

Notice that, using the fact that $\Omega^2 = \Omega$, we can argue that

$$\mathbb{E}[d(t)] = \frac{1}{N}\mathbb{E}[x^*(t)\Omega x(t)] = \frac{1}{N}x^*(0)\Delta(t)x(0)$$

where

$$\Delta(t) := \mathbb{E}[P(0)^*P(1)^* \cdots P(t-1)^*\Omega P(t-1)\cdots P(1)P(0)]$$

if $t \geq 1$ and where $\Delta(0) := \Omega$. A simple recursive argument shows that

$$\Delta(t+1) = \mathbb{E}[P(0)^*\Delta(t)P(0)]$$

This shows that $\Delta(t)$ is the evolution of a linear dynamical system which can be written in the form

$$\Delta(t+1) = \mathcal{L}(\Delta(t))$$

where $\mathcal{L} : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ is given by

$$\mathcal{L}(M) = \mathbb{E}[P(0)^*MP(0)]$$

If now we consider the reachable subspace $\mathcal{R}$ of the pair $(\mathcal{L}, \Omega)$, namely the smallest $\mathcal{L}$-invariant subspace of $\mathbb{R}^{N\times N}$ containing Ω, then we have that

$$R = \mathrm{sr}(\mathcal{L}_{|\mathcal{R}})$$

where $\mathrm{sr}(\cdot)$ means the spectral radius. It will be useful in the sequel to characterize the eigenspace of $\mathcal{L}$ relatively to the eigenvalue 1. We have the following result.

Proposition 4.1. *Assume that $P(t)$ achieves probabilistic consensus and let ρ be the random vector such that (5) holds. Then, the eigenspace of $\mathcal{L}$ with eigenvalue 1 is one-dimensional and $\mathbb{E}[\rho\rho^*]$ is the only eigenvector satisfying $\mathbb{1}^*\mathbb{E}[\rho\rho^*]\mathbb{1} = 1$.*

Proof. Notice that, since $\mathcal{L}^t(M) = \mathbb{E}[Q(t)^*MQ(t)]$ and since $Q(t) \to \mathbb{1}\rho^*$ almost surely and all elements are bounded in t, it follows, by a straightforward application of Lebesgue convergence theorem, that

$$\lim_{t\to\infty} \mathcal{L}^t(M) = (\mathbb{1}^*M\mathbb{1})\mathbb{E}[\rho\rho^*] \tag{16}$$

This proves the result. ■

Clearly the reachability subspace $\mathcal{R}$ will be contained in the subspace generated by the remaining eigenvectors; one could hope that the maximal eigenvalue on $\mathcal{R}$ coincides with the second dominant eigenvalue of $\mathcal{L}$. While we will not be able to prove such a result, we will be able to prove something similar. Notice that we can restrict the operator $\mathcal{L}$ to the subspace $\mathrm{Sym}(N)$ of symmetric $N \times N$ matrices. Consider now

$$\mathcal{S} = \{M \in \mathrm{Sym}(N) \mid M\mathbb{1} = 0\}$$

We have the following result.

Proposition 4.2.

$$R = \mathrm{sr}(\mathcal{L}_{|\mathcal{S}})$$

Proof. Clearly $\mathcal{R} \subseteq \mathcal{S}$ so that we only have to prove $\geq$.

First take a matric $M \in \mathcal{S}$ positive semidefinite. Then there exists W with $W\mathbb{1} = 0$ such that $M = W^*W$. Now observe that

$$\begin{aligned} x(0)^*\mathcal{L}^t(M)x(0) &= \mathbb{E}[||Wx(t)||^2] = \mathbb{E}[||W\Omega x(t)||^2] \\ &\leq ||W||^2\mathbb{E}[||\Omega x(t)||^2] = ||M||x(0)^*\mathcal{L}^t(\Omega)x(0) \end{aligned}$$

This yields

$$||\mathcal{L}^t(M)|| \leq ||M||\ ||\mathcal{L}^t(\Omega)||$$

Consider now any symmetric matrix $M \in \mathcal{S}$. Working with the orthonormal basis of eigenvectors of M, we can decompose it as $M = M_+ - M_-$, where $M_+, M_- \in \mathcal{S}$ are positive semidefinite such that $||M_+|| \leq ||M||$ and $||M_-|| \leq ||M||$. Using the previous result we obtain

$$||\mathcal{L}^t(M)|| \leq ||\mathcal{L}^t(M_+^2)|| + ||\mathcal{L}^t(M_-^2)|| \leq [||M_+|| + ||M_-||]||\mathcal{L}^t(\Omega)|| \leq 2||M||\ ||\mathcal{L}^t(\Omega)||$$

This implies that

$$||\mathcal{L}^t_{|\mathcal{S}}|| := \max_{M\in\mathcal{S}} \frac{||\mathcal{L}^t(M)||}{||M||} \leq 2||\mathcal{L}^t(\Omega)|| \leq 2||\mathcal{L}^t_{|\mathcal{R}}||$$

and hence

$$\mathrm{sr}(\mathcal{L}_{|\mathcal{S}}) = \lim_{t\to\infty} ||\mathcal{L}^t_{|\mathcal{S}}||^{1/t} \leq \lim_{t\to\infty} 2^{1/t}||\mathcal{L}^t_{|\mathcal{R}}||^{1/t} = \mathrm{sr}(\mathcal{L}_{|\mathcal{R}})$$

■

For computing R it is useful to introduce a matrix representation of the linear operator $\mathcal{L}$. Given a matrix $A \in \mathbb{R}^{N\times N}$ we define $\mathrm{vect}(A)$ to be the N^2 column vector having $A_{i,j}$ in position $(i-1)N+j$. Notice that $\mathrm{vect}(ABC) = (C^* \otimes A)\mathrm{vect}(B)$, where $\otimes$ is the Kronecker product of matrices. Using these facts and the properties of the Kronecker product we can argue that the linear operator $\mathcal{L}$ is represented by the matrix

$$\mathbf{L} := \mathbb{E}[P(0)^* \otimes P(0)^*] = \mathbb{E}[P(0) \otimes P(0)]^*$$

Using this notation we have that the rate of convergence R coincides with the absolute value of the dominant reachable eigenvalue of the pair $(\mathbf{L}, \mathrm{vect}(\Omega))$.

Notice that $\mathbf{L}$ has nonnegative elements and moreover

$$\mathbf{L}^*(\mathbb{1}\otimes\mathbb{1}) = \mathbb{E}[P(0)\otimes P(0)](\mathbb{1}\otimes\mathbb{1}) = \mathbb{E}[(P(0)\otimes P(0))(\mathbb{1}\otimes\mathbb{1})] = \mathbb{E}[(P(0)\mathbb{1})\otimes(P(0)\mathbb{1})] = \mathbb{1}\otimes\mathbb{1}$$

therefore $\mathbf{L}^*$ is stochastic. The following proposition allows to obtain more information on the spectral properties on $\mathbf{L}$.

Proposition 4.3. *Assume that $P(0)$ has nonzero diagonal with probability one. If $\mathcal{G}_{\overline{P}}$ is strongly connected then $\mathcal{G}_{\mathbf{L}}$ associated with $\mathbf{L}$ is strongly connected and aperiodic.*

Proof. First observe that

$$\mathbf{L}_{(i,j),(h,k)} := \mathbf{L}_{(i-1)N+j,(h-1)N+k} = \mathbb{E}[P_{hi}(0)P_{kj}(0)]$$

Since $P_{ii}(0) > 0$ with probability one, then $\mathbf{L}_{(i,j),(i,j)} = \mathbb{E}[P_{ii}(0)P_{jj}(0)] > 0$ and so $\mathcal{G}_{\mathbf{L}}$ contains all the self-loops. Notice moreover that, for the same reason, $\mathbf{L}_{(i,j),(i,k)} = \mathbb{E}[P_{ii}(0)P_{kj}(0)] > 0$ iff $\overline{P}_{kj} > 0$ and $\mathbf{L}_{(i,j),(h,j)} = \mathbb{E}[P_{hi}(0)P_{jj}(0)]$ iff $\overline{P}_{hi} > 0$. Therefore, if we take (i,j) and (h,k) arbitrarily, since $\mathcal{G}_{\overline{P}}$ is strongly connected, we can find a sequence $j = \nu_0, \nu_1, \ldots, \nu_s = k$ and another sequence $i = \mu_0, \mu_1, \ldots, \mu_r = h$ such that $\overline{P}_{\nu_{l-1},\nu_l} > 0$ and $\overline{P}_{\mu_{l-1},\mu_l} > 0$. This implies that the sequence

$$(i,j) = (i,\nu_0), (i,\nu_i), \ldots, (i,\nu_s) = (i,k) = (\mu_0,k), (\mu_1,k), \ldots, (\mu_s,k) = (h,k)$$

corresponds to an admissible path in $\mathcal{G}_{\mathbf{L}}$. ■

The previous proposition implies that, under mild hypotheses, $\mathbf{L}^*$ is an irreducible aperiodic stochastic matrix row and therefore the eigenvalue 1 has algebraic multiplicity 1.

The evaluation of the parameter R (defined in (15)) can be in general a quite heavy computational task because of the fact that $\mathcal{L}$ operates in a space of dimension N^2. It is thus useful the following result which furnish bounds to T in terms of the spectral structure of N-dimensional operators.

Proposition 4.4. *The following holds*

$$\mathrm{esr}(\overline{P})^2 \leq R \leq \mathrm{sr}(\mathcal{L}(\Omega))$$

where $\mathrm{esr}(\overline{P})$ *is the essential spectral radius of the stochastic matrix* $\overline{P}$ *as defined in (13).*

Proof. We start with the first inequality. Notice that

$$\mathbb{E}[x^*(t)\Omega x(t)] = \mathbb{E}[||\Omega x(t)||^2] = \mathbb{E}[||\Omega Q(t)x(0)||^2]$$

Now using the Jensen inequality we have that

$$\mathbb{E}[||\Omega Q(t)x(0)||^2] \geq ||\mathbb{E}[\Omega Q(t)x(0)]||^2 = ||\Omega \overline{P}^t x(0)||^2$$

which proves the result.

For the second inequality first observe that for all $x \in \mathbb{R}^N$ we have that

$$x^*\mathbb{E}[P^*(0)\Omega P(0)]x = \mathbb{E}[x^*\Omega P^*(0)\Omega P(0)\Omega x] \leq ||\mathbb{E}[P^*(0)\Omega P(0)]||x^*\Omega x$$

where we used the fact that, if M is a symmetric matrix, then $x^*Mx \leq ||M||x^*x$. This shows that $||\mathcal{L}(\Omega)||\Omega \geq \mathcal{L}(\Omega)$[1]. Observe moreover that $M_1 \geq M_2$ implies that $\mathcal{L}(M_1) \geq \mathcal{L}(M_2)$. From these facts it follows that

$$\mathcal{L}^t(\Omega) = \mathcal{L}^{t-1}(\mathcal{L}(\Omega)) \leq \mathcal{L}^{t-1}(||\mathcal{L}(\Omega)||\Omega) = ||\mathcal{L}(\Omega)||\mathcal{L}^{t-1}(\Omega)$$

Iterating this inequality we find that

$$\mathcal{L}^t(\Omega) \leq ||\mathcal{L}(\Omega)||^t\Omega$$

which implies the thesis. ■

Remark In the case of the symmetric gossip algorithm, the second inequality can be elaborated further. Indeed, in that case we have that all the values which $P(t)$ can assume are symmetric stochastic matrices. For this reason

$$\mathcal{L}(\Omega) = \mathbb{E}[P^*(t)\Omega P(t)] = \mathbb{E}[P^2(t)] - \frac{1}{N}\mathbb{1}\mathbb{1}^*$$

Moreover, if we choose $q = 1/2$, then we have that $P^2(t) = P(t)$ with and so we can argue that in this case $\mathrm{sr}(\mathcal{L}(\Omega)) = \mathrm{sr}(\overline{P} - N^{-1}\mathbb{1}\mathbb{1}^*) = \mathrm{esr}(\overline{P})$. This is exactly the bound proposed in [4].

We reconsider now some of the randomized consensus algorithms introduced above assuming that the supporting graph is the complete graph. In this case the mean square analysis is quite simple because the linear operator $\mathcal{L}$ keeps invariant the subspace generated by I and $N^{-1}\mathbb{1}\mathbb{1}^*$. Therefore, since $\Delta(0)$ belongs to this subspace, the dynamics of $\Delta(t)$ is determined by a 2×2 matrix.

Example 4.5: We reconsider the case presented in Example 2.4. As shown in [12, Proposition V.2] we have that $\Delta(t) = \alpha(t)I + \beta(t)N^{-1}\mathbb{1}\mathbb{1}^*$ and $\alpha(t)$ and $\beta(t)$ evolve according to the following linear dynamics

$$\begin{pmatrix} \alpha(t+1) \\ \\ \beta(t+1) \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{N-1} & 0 \\ \\ \frac{1}{N-1} & 1 \end{pmatrix} \begin{pmatrix} \alpha(t) \\ \\ \beta(t) \end{pmatrix}$$

Therefore $\Delta(t)$ converges to $N^{-1}\mathbb{1}\mathbb{1}^*$ and the speed of convergence is given by the second eigenvalue

$$R = 1 - \frac{1}{N-1}$$

In the specific case considered in the simulations in which $N = 16$ we have that $R = 14/15 \simeq 0.93$. This value coincides with the rate of convergence estimated from the graphs obtained in the simulations. This seems to suggest that R is the right parameter describing the convergence in the initial time range.

[1] In the proof of this proposition when we write $A \geq B$, where A, B are square matrices having the same dimension, we mean that $A - B$ is positive semidefinite, namely that $x^*(A - B)x \geq 0$ for all the column vectors x.

Observe moreover that

$$\overline{P} = \frac{N-1}{N} I + \frac{1}{N^2} \mathbb{1}\mathbb{1}^*$$

and so $\mathrm{esr}(\overline{P}) = 1 - N^{-1}$. Notice that $(1 - N^{-1})^2 \leq 1 - (N-1)^{-1}$ when $N > 2$ in accordance with Proposition 4.4. Finally we have that

$$\mathcal{L}(\Omega) = \left(1 - \frac{1}{N-1}\right) I + \left(\frac{1}{N-1} - 1\right) N^{-1} \mathbb{1}\mathbb{1}^*$$

and so $\mathrm{sr}(\mathcal{L}(\Omega)) = 1 - (N-1)^{-1} = R$, namely in this case the upper bound suggested by Proposition 4.4 is attained.

Example 4.6: We reconsider the case presented in Example 3.6 with the complete graph and $q = 1/2$. As shown in [5] we have that $\Delta(t) = \alpha(t) I + \beta(t) N^{-1} \mathbb{1}\mathbb{1}^*$ and $\alpha(t)$ and $\beta(t)$ evolve according to the following linear dynamics

$$\begin{pmatrix} \alpha(t+1) \\ \\ \beta(t+1) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4N} \\ \\ \frac{1}{2} & 1 - \frac{1}{4N} \end{pmatrix} \begin{pmatrix} \alpha(t) \\ \\ \beta(t) \end{pmatrix}$$

The eigenvalues are 1, with eigenvector $(1/4N \; 1/2)^*$, and $\frac{1}{2} - \frac{1}{4N}$ with eigenvector $(1 \; -1)^*$. The first eigenvector corresponds to the matrix $\Lambda := \frac{1}{4N} I + \frac{1}{2N} \mathbb{1}\mathbb{1}^*$, while the second eigenvalue corresponds to Ω. Therefore $\Delta(t)$ converges to zero and the speed of convergence is given by the second eigenvalue

$$R = \frac{1}{2} - \frac{1}{4N}$$

Example 4.7: We reconsider the case presented in Example 3.8. We assume that $P = N^{-1} \mathbb{1}\mathbb{1}^*$. The mean square analysis of this example has been done in [11]. It can be seen that also in this case $\Delta(t) = \alpha(t) I + \beta(t) N^{-1} \mathbb{1}\mathbb{1}^*$ and that $\alpha(t)$ and $\beta(t)$ evolve according to the following linear dynamics

$$\begin{pmatrix} \alpha(t+1) \\ \\ \beta(t+1) \end{pmatrix} = \begin{pmatrix} (1-p)\left(1 - p - \frac{2p}{N}\right) & \frac{2p(1-p)}{N^2} \\ \\ 1 - (1-p)\left(1 - p - \frac{2p}{N}\right) & 1 - \frac{2p(1-p)}{N^2} \end{pmatrix} \begin{pmatrix} \alpha(t) \\ \\ \beta(t) \end{pmatrix}$$

The eigenvalues are 1 and $(1-p)\left(1 - p + 2p\frac{N-1}{N^2}\right)$. The eigenvector relative to 1 is

$$\begin{pmatrix} \frac{2p(1-p)}{N^2} \\ \\ 1 - (1-p)\left(1 - p - \frac{2p}{N}\right) \end{pmatrix}$$

and corresponds to the matrix

$$\Lambda := \frac{2p(1-p)}{N^2} I + \left[1 - (1-p)\left(1 - p - \frac{2p}{N}\right)\right] \frac{1}{N} \mathbb{1}\mathbb{1}^*$$

while the eigenvector relative to the second eigenvalue is $(1 \; -1)^*$ which corresponds to the matrix Ω. Therefore, $\Delta(t)$ converges to zero and the speed of convergence is given by the second eigenvalue

$$R = (1-p)\left(1 - p + 2p\frac{N-1}{N^2}\right)$$

4.2 The average displacement

In the mean square analysis of the average displacement, first observe that, since $\beta(t)$ converges almost surely to $\beta(\infty) = |(\rho^* - N^{-1}\mathbb{1}^*)x(0)|^2$, then

$$\lim_{t \to \infty} \mathbb{E}[\beta(t)] = \mathbb{E}[\beta(\infty)]$$

Observe moreover that

$$\mathbb{E}[\beta(\infty)] = x(0)^* B x(0)$$

where

$$B = \mathbb{E}[\rho\rho^*] - 2N^{-1}\mathbb{E}[\rho]\mathbb{1}^* + N^{-2}\mathbb{1}\mathbb{1}^*$$

Notice that B is expressed in terms of $\mathbb{E}[\rho]$ and $\mathbb{E}[\rho\rho^*]$ which are the eigenvectors relative to the eigenvalue 1 of the operators $\overline{P}$ and $\mathcal{L}$, respectively. When $\overline{P}$ is doubly stochastic, we obtain in particular

$$B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^* \tag{17}$$

Notice moreover that, by (16), we can argue that

$$\mathbb{E}[\rho\rho^*] = \frac{1}{N}\lim_{t\to\infty}\mathcal{L}^t(I)$$

The importance of the positive semidefinite matrix B consists in the fact that through it we can also estimate $\mathbb{E}[\beta(t)]$ for all t. Indeed, we have that

$$\begin{aligned}\mathbb{E}[\beta(t)]^{1/2} &= \left\{\mathbb{E}[|x_A(t) - x_A(0)|^2]\right\}^{1/2} \leq \left\{\mathbb{E}[|x_A(t) - x_A(\infty)|^2\right\}^{1/2} + \left\{\mathbb{E}[|x_A(\infty) - x_A(0)|^2]\right\}^{1/2} \\ &\leq (1+\sqrt{N})\left\{\mathbb{E}[d(t)]\right\}^{1/2} + \left\{x(0)^* B x(0)\right\}^{1/2}\end{aligned}$$

where in the last inequality we used Lemma 2.2 and the Schwartz inequality.

Example 4.8: We reconsider the case studied in Example 4.6. As observed above we have that

$$B = \frac{1}{N}\left(\lim_{t\to\infty}\mathcal{L}^t(I) - N^{-1}\mathbb{1}\mathbb{1}^*\right)$$

By decomposing I along the two eigenvectors Λ and Ω of $\mathcal{L}$ one obtains

$$I = \frac{4N}{1+2N}\Lambda + \frac{2N}{1+2N}\Omega$$

and so

$$\lim_{t\to\infty}\mathcal{L}^t(I) = \frac{4N}{1+2N}\Lambda$$

From this we can argue that

$$B = \frac{1}{N(1+2N)}\Omega$$

and so

$$\mathbb{E}[\beta(\infty)] = \frac{1}{1+2N}\frac{1}{N}||\Omega x(0)||^2$$

If we consider initial conditions with bounded components $x_i(0)$, in such a way that $N^{-1}||x(0)||^2$ is bounded with respect to N, we have that $\mathbb{E}[\beta(\infty)]$ tends to zero as N tends to infinity. This means that, although the asymmetric gossip strategy does not yield average consensus, the error between the consensus point $x_i(\infty)$ and the initial average $x_A(0)$ becomes negligeable when N is sufficiently big.

Example 4.9: We reconsider the case studied in Example 4.7. By decomposing I along the two eigenvectors Λ and Ω of $\mathcal{L}$ one obtains that

$$I = \frac{N^2}{p[(2-p)N^2 + (2-2p)N + (2-2p)]}\Lambda + \frac{N[(2-p)N + (2-2p)]}{(2-p)N^2 + (2-2p)N + (2-2p)}\Omega$$

and so

$$\lim_{t\to\infty}\mathcal{L}^t(I) = \frac{N^2}{p[(2-p)N^2 + (2-2p)N + (2-2p)]}\Lambda$$

From this we can argue that

$$B = \frac{1}{N}\frac{2-2p}{(2-p)N^2 + (2-2p)N + (2-2p)}\Omega$$

and so

$$\mathbb{E}[\beta(\infty)] = \frac{2-2p}{(2-p)N^2 + (2-2p)N + (2-2p)}\frac{1}{N}||\Omega x(0)||^2$$

If we consider initial conditions with bounded components $x_i(0)$, in such a way that $N^{-1}||x(0)||^2$ is bounded with respect to N, also in this case we have that $\mathbb{E}[\beta(\infty)]$ tends to zero as N tends to infinity.

5 Concentration results

In this chapter we discuss some concentration results concerning the random variables $d(t)$ and $\beta(t)$ for large values of N. We will need to use in this section some basic results on martingales. We first recall some basic notation and concepts (see [1] for further details).

5.1 Martingales and Azuma's inequality

Consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathcal{F}_t$ $(t = 0, 1, \dots)$ of sub σ-fields of $\mathcal{F}$, namely,

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_t \subseteq \cdots$$

Let $X_t : \Omega \to \mathbb{R}$ be a sequence of random variables adapted to the filtration, namely such that X_t is $\mathcal{F}_t$-measurable. The sequence of random variables X_t is said to be a martingale (with respect to the given filtration) if it happens that

$$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = X_{t-1}\,, \quad \forall t \geq 1$$

If we start from any random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$, it is a straightforward fact that the stochastic process $X_t := \mathbb{E}[X|\mathcal{F}_t]$ is a martingale. The concept of martingale can also be defined in situations where the index t is restricted to an interval of positive integers. The basic result on martingales that we will need is the so called Azuma's inequality recalled below [1].

Theorem 5.1. Azuma's inequality *Let X_t be a martingale w.r. to some given filtration. Assume that $|X_t - X_{t-1}| \leq c$ for every t. Then, for every $\delta > 0$ we have that*

$$\mathbb{P}[|X_t - X_0| \geq \delta] \leq \exp\left(-\frac{\delta^2}{2c^2 t}\right)$$

In our setting the natural filtration to consider is the one induced by the sequence of random matrices $P(t)$. More precisely we let $\mathcal{F}_t$ to be the σ-field generated by the family of random variables $P(0), \dots, P(t-1)$. We assume moreover that $\mathcal{F}_0$ is the trivial σ-field.

The basic idea is to use Azuma's inequality applied to the martingales constructed, for any fixed t, by considering $X_s = \mathbb{E}[d(t)|\mathcal{F}_s]$, for $s = 0, \dots, t$. By proving a suitable bound on differences $|X_{s+1} - X_s|$ and noticing that

$$X_0 = \mathbb{E}[d(t)]\,, \quad X_t = d(t)\,,$$

we would in principle obtain a concentration result for our random variable $d(t)$ around its average. For some of the examples presented before (the gossip cases and the broadcasting model) this approach would work. However, in order to be able to obtain a result which can be applied to all the examples we have introduced so far, we will need to introduce a refinement of the above filtration.

Assume that the random stochastic matrices $P(t)$ are generated in the following way. Assume that there exist independent random variables $T_1(t), \dots, T_m(t)$ taking values in a finite set $\mathcal{T}$ and a function $\Gamma : \mathcal{T}^m \to \mathbb{R}^{N \times N}$ such that

$$P(t) = \Gamma(T_1(t), \dots, T_m(t)) \tag{18}$$

Roughly speaking, the random variables $T_1(t), \dots, T_m(t)$ generate all the randomness of $P(t)$. The function $\Gamma(T_1, \dots, T_m)$ is said to be a-sensitive if, when we vary just one of its variables T_k keeping fixed all the others, only at most a rows may vary.

Of course, this representation is generally not unique since one can always assemble some of the $T_k(t)$ into one random variable and decrease in such a way the index m. However, this will determine an increase in the sensitivity a. We will see that, because of the different roles that m and a will have in the concentration inequality we will propose, the sharpest results are obtained when $a^2 m$ is as small as possible. Typically this will be obtained when choosing representations with m as large as possible.

It is easy to see that the matrices $P(t)$ in our examples can always be generated as specified above. Specifically, in the following we list the representations in the various examples introduced in the previous section.

(1) *Symmetric gossip*: In this case $m = 1$ and $\mathcal{T} = E$. The variable $T_1(t)$ codifies the (undirected) edge which is activated and so $a = 2$.

(2) *Asymmetric gossip*: In this case $m = 1$ and $\mathcal{T} = E$. The variable $T_1(t)$ codifies the edge which is activated and so $a = 1$.

(3) *Synchronous gossip*: In this case $m = N$ and $\mathcal{T} = \{1, \dots, N\}$. The variable $T_k(t)$ codifies the neighbor node chosen by node k and so $a = 1$.

(4) *Broadcasting*: In this case $m = 1$ and $\mathcal{T} = \{1, \dots, N\}$. The variable $T_1(t)$ codifies the activated node and so $a = \max_k\{\nu_k^+\}$.

(5) *Packet drop*: In this case $m = N$ and $\mathcal{T} = \{0, 1\}^N$. The variable $T_k(t)$ codifies which of the incoming edges to agent k are active at time t and so $a = 1$.

(6) *Agent failure*: In this case $m = N$ and $\mathcal{T} = \{0, 1\}$. The variable $T_k(t)$ codifies the state of activity of agent k at time t and so $a = \max_k\{\nu_k^+\}$.

Once we have fixed a representation like (18), we can consider the following refinement of the filtration introduced above. For any $t \in \mathbb{N}$ and $k = 1, \dots, m$, we let $\mathcal{F}_{tm+k}$ be the σ-field generated by the random matrices $P(0), \dots, P(t-1)$ and the random variables $\{T_1(t), \dots, T_k(t)\}$. Assume moreover that $\mathcal{F}_0$ is the trivial σ-field.

5.2 Application of Azuma's inequality

We first want apply Azuma's inequality to the distance from the consensus $d(t)$. To this aim, any fixed t, consider the family of random variables $X_s = \mathbb{E}[d(t)|\mathcal{F}_s]$, for $s = 0, \dots, tm$. Notice that

$$X_0 = \mathbb{E}[d(t)]\,, \quad X_{tm} = d(t)\,,$$

The following result is obtained by applying Azuma's inequality to the martingale X_s, $s = 0, \dots, tm$. In the proof of the following results we will need to use an elementary inequality which we will highlight now.

Let x, y, z be independent random variables and let $\mathcal{F}_x$ and $\mathcal{F}_{x,y}$ be σ-fields generated by x and by x, y, respectively. Observe that $\mathbb{E}[f(x,y,z)|\mathcal{F}_x] = \mathbb{E}[f(x,y,z)|x]$ and $\mathbb{E}[f(x,y,z)|\mathcal{F}_{x,y}] = \mathbb{E}[f(x,y,z)|x,y]$, where $f(\cdot)$ is any function. Notice now that

$$\begin{aligned}\Big|\mathbb{E}(f(x,y,z)|x,y) - \mathbb{E}(f(x,y,z)|x)\Big| &= \Big|\mathbb{E}\Big[f(x,y,z) - \mathbb{E}(f(x,y,z)|x,z)\Big|x,y\Big]\Big| \\ &\leq \mathbb{E}\Big[|f(x,y,z) - \mathbb{E}(f(x,y,z)|x,z)|\Big|x,y\Big]\end{aligned}$$

Since we have that

$$\Big|f(x,y,z) - \mathbb{E}(f(x,y,z)|x,z)\Big| \leq \sup_{y',y''}\Big|f(x,y',z) - f(x,y'',z)\Big|$$

we can argue that

$$\Big|\mathbb{E}\big[f(x,y,z)|x,y] - \mathbb{E}[f(x,y,z)|x]\Big| \leq \mathbb{E}\Big[\sup_{y',y''}|f(x,y',z) - f(x,y'',z)|\Big|x\Big] \tag{19}$$

Proposition 5.2. *Assume that the random stochastic matrices $P(t)$ can be generated through an a-sensitive map Γ and discrete random variable $T_1(t), \dots, T_m(t)$ as above. Assume, moreover, that $\overline{P}$ is doubly stochastic. Then,*

$$\mathbb{P}[|d(t) - \mathbb{E}d(t)| \geq \delta] \leq \exp\left(-\frac{\delta^2 N^2}{128 a^2 m t ||x(0)||_\infty^4}\right)$$

Proof. Notice that, as a random variable $d(t)$ is function of the $P(s)$ for $s = 0, \dots, t-1$ and, ultimately, of the $T_j(\tau)$, $\tau = 0, \dots, t-1$, $j = 1, \dots, m$. Consider $l \in \{1, 2, \dots, tm\}$. Our aim is to determine a bound of $|X_l - X_{l-1}|$. Observe that, if $l = sm + k$, then the σ-field $\mathcal{F}_{l-1}$ will be generated by some of the random variables $T_j(\tau)$. We will collect these random variables in a random vector called x. The

σ-field $\mathcal{F}_l$ will be generated by x and by the random variable $T_k(s)$ which we will call y. Finally we collect the remaining random variables determining $d(t)$ in a random vector called z. In evaluating $|X_l - X_{l-1}|$ we can apply inequality (19). To do so we need to evaluate

$$\sup_{T_k(s)', T_k(s)''} \Big| ||\Omega P(t-1)\cdots P(s+1)\Gamma(T_1(s),\ldots,T_k(s)',\ldots,T_m(s))P(s-1)\cdots P(0)x(0)||^2 - \\ - ||\Omega P(t-1)\cdots P(s+1)\Gamma(T_1(s),\ldots,T_k(s)'',\ldots,T_m(s))P(s-1)\cdots P(0)x(0)||^2 \Big|$$

If we define the two vectors

$$\begin{aligned} w &:= \Gamma(T_1(s),\ldots,T_k(s)',\ldots,T_m(s))P(s-1)\cdots P(0)x(0) \\ w' &:= \Gamma(T_1(s),\ldots,T_k(s)'',\ldots,T_m(s))P(s-1)\cdots P(0)x(0) \end{aligned} \tag{20}$$

we have that

$$||w||_\infty \leq ||x(0)||_\infty, \quad ||w'||_\infty \leq ||x(0)||_\infty \tag{21}$$

and moreover there exist indices $j_1,\ldots,j_a$ such that

$$w' - w = \sum_{l=1}^{a} r_l e_{j_l}\,, \quad |r_l| \leq 2||x(0)||_\infty \tag{22}$$

From (21) and (22) we obtain

$$\begin{aligned} &\big| ||\Omega P(t-1)\cdots P(s+1)w||^2 - ||\Omega P(t-1)\cdots P(s+1)w'||^2 \big| \\ &\quad = |(w+w')^*(P(t-1)\cdots P(s+1))^*\Omega P(t-1)\cdots P(s+1)(w-w')| \\ &\quad \leq ||\Omega P(t-1)\cdots P(s+1)(w+w')||_\infty ||P(t-1)\cdots P(s+1)(w-w')||_1 \end{aligned} \tag{23}$$

Notice that

$$||\Omega P(t-1)\cdots P(s+1)(w+w')||_\infty \leq ||\Omega||_\infty ||P(t-1)\cdots P(s+1)||_\infty ||(w+w')||_\infty \leq 4||x(0)||_\infty$$

while

$$||P(t-1)\cdots P(s+1)(w-w')||_1 \leq 2||x(0)||_\infty \mathbb{1}^* P(t-1)\cdots P(s+1)\sum_{l=1}^{a} e_{j_l}$$

Now inserting inequality (23) inside (19), we finally obtain

$$\begin{aligned} |X_l - X_{l-1}| &\leq \frac{1}{N}\mathbb{E}\left[8||x(0)||_\infty^2 \mathbb{1}^* P(t-1)\cdots P(s+1)\sum_{l=1}^{a} e_{j_l}\Big|\mathcal{F}_{l-1}\right] \\ &= \frac{8||x(0)||_\infty^2}{N}\mathbb{1}^*\overline{P}^{t-s}\sum_{l=1}^{a} e_{j_l} = \frac{8a||x(0)||_\infty^2}{N} \end{aligned} \tag{24}$$

The result now follows from Azuma's inequality. ■

Similar concentration results can be obtained also for the average displacement $\beta(t)$. To prove this fact, we introduce now the family of random variables $X_s = \mathbb{E}[x_A(t)|\mathcal{F}_s]$, for $s = 0,\ldots,tm$, where the $\mathcal{F}_s$ are the σ-fields introduced above. Notice that

$$X_0 = \mathbb{E}[x_A(t)] = x_A(0)\,, \quad X_{tm} = x_A(t)\,,$$

The following result is obtained by applying Azuma's inequality to the martingale X_s, $s = 0,\ldots,tm$.

Proposition 5.3. *Assume that the random stochastic matrices $P(t)$ can be generated through an a-sensitive map Γ and discrete random variables $T_1(t),\ldots,T_m(t)$ as above. Assume, moreover, that $\overline{P}$ is doubly stochastic. Then,*

$$\mathbb{P}[\beta(t) \geq \delta] \leq \exp\left[-\frac{\delta N^2}{8||x(0)||_\infty^2 a^2 mt}\right]$$

Proof. The proof is very similar to the proof of the previous result. First notice that $\mathbb{P}[\beta(t) \geq \delta] = \mathbb{P}[|x_A(t) - x_A(0)| \geq \delta^{1/2}]$. Since $x_A(t) = N^{-1}\mathbb{1}^* P(t-1)\cdots P(0)x(0)$, in this case we need to evaluate

$$\sup_{T_k(s)', T_k(s)''} \Big| \mathbb{1}^* P(t-1)\cdots P(s+1)\Gamma(T_1(s), \ldots, T_k(s)', \ldots, T_m(s))P(s-1)\cdots P(0)x(0) - \\ -\mathbb{1}^* P(t-1)\cdots P(s+1)\Gamma(T_1(s), \ldots, T_k(s)'', \ldots, T_m(s))P(s-1)\cdots P(0)x(0)\Big|$$

If we define the two vectors w, w' as in (20), from (21) and (22) we obtain

$$|\mathbb{1}^* P(t-1)\cdots P(s+1)w - \mathbb{1}^* P(t-1)\cdots P(s+1)w'| \leq 2||x(0)||_\infty \mathbb{1}^* P(t-1)\cdots P(s+1)\sum_{l=1}^{a} e_{j_l} \quad (25)$$

Inserting inequality (25) inside (19), we finally obtain

$$|X_l - X_{l-1}| \quad \leq \quad \frac{1}{N}\mathbb{E}\left[2||x(0)||_\infty \mathbb{1}^* P(t-1)\cdots P(s+1)\sum_{l=1}^{a} e_{j_l}\Big|\mathcal{F}_{l-1}\right] = \frac{2a||x(0)||_\infty}{N} \quad (26)$$

The result now follows from Azuma's inequality. ■

5.3 The behavior for large N

In order to fully appreciate the previous results, it is convenient to consider sequences of consensus algorithms $P(t)$ of increasing dimension N. We assume from now on that all these algorithms satisfy the assumptions of Propositions 5.2 and 5.3, namely, that the random stochastic matrices $P(t)$ are generated through an a-sensitive map Γ and discrete random variable $T_1(t), \ldots, T_m(t)$ as in (18) and that $\overline{P}$ is doubly stochastic.

In this section we will use a notation which makes explicit the dependence on N. Namely, $x(t, N)$ will denote the state of the consensus algorithm, $d(t, N)$ the distance from the consensus, $\beta(t, N)$ the average displacement, $R(N)$ the exponential rate of convergence of $\mathbb{E}[d(t)]$, and, finally, $m(N)$, $a(N)$ come from the parametrization given in (18). We also need to say something on the way we choose the sequence of initial conditions $x(0, N)$. In this section we will assume that there exists L such that $||x(0, N)||_\infty \leq L$ for every N.

We now present a result showing how the typical behavior concentrate around the average behavior for large N.

Corollary 5.4. *Let T_N be a sequence such that*

$$T_N = O\left(\frac{N^{2-\epsilon}}{a(N)^2 m(N)}\right) \quad (27)$$

for some $\epsilon > 0$, (e.g. $T_N \left(\frac{N^{2-\epsilon}}{a(N)^2 m(N)}\right)^{-1}$ is bounded as $N \to \infty$). Then, almost surely,

$$\limsup_{N\to+\infty} \sup_{t \leq T_N} |d(t, N) - \mathbb{E}[d(t, N)]| = 0$$
$$\limsup_{N\to+\infty} \sup_{t \leq T_N} \beta(t, N) = 0\,.$$

Proof. It is a straightforward application of estimations in Propositions 5.2 and 5.3 and of the Borel-Cantelli lemma. ■

The above corollary says that, when the time is sufficiently small with respect to the dimension N, the typical behavior of the consensus distance $d(t, N)$ is quite well approximated by its mean $\mathbb{E}[d(t, N)]$ for large N. This provides a strong motivation for considering the rate of convergence of $\mathbb{E}[d(t, N)]$ for analysis and optimization purposes. A similar result holds true for the average displacement $\beta(t, N)$.

An important aspect to investigate now is related to the parameter T_N. We first notice that for all the models considered we can choose a sequence T_N which grows in N and tends to infinity as $N \to \infty$. The only exception in the examples we considered is given by the agent failure model when the maximal out degree is linear in N. Apart from this particular case, this means that the time

range in which the mean analysis is meaningful grows in N. Notice however that, typically, the rate of convergence $R(N)$ tends to one as $N \to \infty$. The effect of this is that it might happen that the time range T_N is not big enough to provide a sufficient convergence. In order to understand this fact, it is helpful to consider the so called the α-averaging time

$$T(\alpha, N) = \sup_{x(0,N)} \inf\{t \in \mathbb{N} \mid \mathbb{P}[||x(t,N) - x(\infty, N)||_2 \geq \alpha ||x(0,N)||^2] \leq \alpha\}$$

Roughly speaking the α-averaging time is the time needed to shrink the distance to average consensus of a factor α. This is the main performance index studied in the context of load balancing and also it is the one considered in [4].

In most of the applications it is necessary to run the algorithm till time $T(\alpha, N)$ for a given $\alpha < 1$. It is therefore of interest to understand if our concentration results can be applied over this time range. In other words, one can wonder whether $T(\alpha, N)$ satisfies condition (27). For the symmetric gossip model there are several situations in which we can estimate the α-averaging time and thus give an answer to the above question. Indeed, as shown in [4], we have that

$$T(\alpha, N) \asymp \left(\ln \frac{1}{R(N)}\right)^{-1} \ln \frac{1}{\alpha} \qquad (N \to \infty \quad \alpha \to 0) \tag{28}$$

(where $\asymp$ means that the ratio between the two functions tends to a finite non-zero limit). In Example 4.5 in which the graph is complete we have seen that $R(N) = 1 - (N-1)^{-1}$ and so in this case $T(\alpha, N) \asymp N \ln \frac{1}{\alpha}$. We can argue that we can take $T_N = T(\alpha, N)$. Some more general graphs can be treated by observing that, by Proposition 4.4 and the following Remark, equation (28) implies the following

$$T(\alpha, N) \asymp \left(\ln \frac{1}{\mathrm{esr}(\overline{P})}\right)^{-1} \ln \frac{1}{\alpha} \qquad (N \to \infty \quad \alpha \to 0) \tag{29}$$

The estimation of the essential spectral radius of a stochastic matrix as $\overline{P}$ is a classical problem which has been widely studied in the literature on Markov chains. There are cases in which this estimation is particularly simple. For instance, (see [4, 5]) for the hypercube graph we have that $1 - \mathrm{esr}(\overline{P}) \asymp N^{-1}$, while for the n-dimensional torus graph we have that $1 - \mathrm{esr}(\overline{P}) \asymp N^{-\frac{2+n}{n}}$. Using these estimations in (29) we thus obtain that $T(\alpha, N) \asymp N \ln \frac{1}{\alpha}$ for the hypercube graph and $T(\alpha, N) \asymp N^{\frac{2+n}{n}} \ln \frac{1}{\alpha}$ for the n-dimensional torus. The conclusion is that we can choose $T_N = T(\alpha, N)$ in Corollary 5.4 for the complete graph, for the hypercube graph and for the n-dimensional torus in case $n \geq 3$.

For the asymmetric gossip model, quite similar results can be obtained: they are however technically more complicated and will be presented elsewhere. Concerning the synchronous gossip model, at the moment we have a result for the special case when $\mathcal{G}$ is complete and $W = N^{-1}\mathbb{1}\mathbb{1}^*$ considered in Example 4.6. Since in this case we showed that $R(N)$ is bounded away from 0, it follows that that $T(\alpha, N)$ is asymptotically independent of N. Since $m(N) = N$ and $a(N) = 1$, also in this case we can choose $T_N = T(\alpha, N)$ in Corollary 5.4.

We are currently investigating the application of Corollary 5.4 to other graphs and to the other models discussed in this paper. We are also studying the application to random geometric graphs, using the the ideas suggested in [4].

Finally, we want to make some comments on the concentration result regarding $\beta(t, N)$ in Corollary 5.4. It clearly leaves open the analysis of $\beta(t, N)$ when t is large with respect to N. In this respect, the analysis of the asymptotic displacement in Section 4.2 is in some sense complementary with respect to the results in Corollary 5.4. In principle, it should be possible to combine the two analysis in order to prove a general convergence result on $\beta(t, N)$ valid for all time range. This will be discussed elsewhere.

6 Conclusions and future research

In this paper we presented, under a unified theory, a number of random consensus algorithms and of deterministic ones but embedded in a random environment. Some of these models had already appeared in the literature. Some are completely new. Scope of the work was to present a number of general results which can be applied to study the behavior of these different algorithms. We presented some general results concerning the mean square performance of these algorithms and proven some

concentration results which validate such analysis when the time range is sufficiently small with respect to the number of nodes. In many applications this is the most important range to be considered and we have shown with an example that the performance in such a range can differ considerably from performance studied through classical Lyapunov exponent theory. Many important questions remain open.

- The proposed mean square analysis needs to be concretely applied in many of the examples proposed. Detailed computation for the analysis and optimization of these algorithms will be carried on in an upcoming paper. We believe that quite rich and complete results can be obtained for a number of graphs possessing sufficient symmetries as for instance Cayley graphs. Another interesting family of graphs to work on would be the random geometrical graphs.
- Simulations seem to show quite stronger and wider range concentration results than those theoretically proven and we are currently investigating the possibility to improve them.
- The algorithms proposed in this paper are, by no means, the only one for consensus. In the literature there are also algorithms with memory, algorithms not based on stochastic matrices (see for instance [6]), algorithms with quantization structures. It is surely worth to investigate the possibility to extend our results to these broader class of examples.

7 Appendix

7.1 The proof of Theorem 2.3

We first recall some fundamental facts from ergodic theory (see [25, 2] for further details). Our basic tool is Kingman's subadditive ergodic theorem recalled below.

Theorem 7.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a measurable map $\sigma : \Omega \to \Omega$ such that $\sigma\mathbb{P} = \mathbb{P}$. Let $f_t : \Omega \to \mathbb{R} \cup \{-\infty\}$ be a sequence of measurable maps such that*

(a) $f_1^+ \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, where f_1^+ is defined by letting $f_1^+(x) := \max\{f_1(x), 0\}$ for all x.

(b) $f_{k+t}(x) \leq f_k(x) + f_t(\sigma^k x)$, for every $k, t \in \mathbb{N}$ and $x \in \Omega$.

Then, there exists $f : \Omega \to \mathbb{R} \cup \{-\infty\}$ which is measurable and σ-invariant such that

$$\lim_{t\to\infty} \frac{1}{t} f_t = f \quad \mathbb{P}\text{-almost surely}$$

To apply this result to our setting we suppose we have fixed a consensus algorithm $P(t)$ (of dimension N) achieving probabilistic consensus. The probability space on which all the random matrices $P(t)$ are defined, can be concretely constructed in the standard way using infinite sequence space as follows. We first consider the base probability space $(S, \mathcal{B}(S), \mu)$ where S is the compact space of all stochastic $N \times N$ matrices, $\mathcal{B}(S)$ is the corresponding Borel σ-algebra, and μ is the probability law of the random matrix $P(t)$ for any t. Consider now the infinite sequence space $(S^{\mathbb{N}}, \mathcal{B}(S)^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$ equipped with the product probability measure $\mu^{\otimes\mathbb{N}}$. On $S^{\mathbb{N}}$ we define the backward shift operator $\sigma : S^{\mathbb{N}} \to S^{\mathbb{N}}$: $(\sigma\omega)_t = \omega_{t+1}$ for every $\omega \in S^{\mathbb{N}}$ and $t \in \mathbb{N}$. It is immediate to check that $\sigma\mu^{\mathbb{N}} = \mu^{\mathbb{N}}$. We can think of the random matrices $P(t)$ as random variables defined on $(S^{\mathbb{N}}, \mathcal{B}(S)^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$. Using the notation $P(t, \omega)$ where $\omega \in S^{\mathbb{N}}$, we simply put $P(t, \omega) = \omega_t$. It holds $P(t, \omega) = P(0, \sigma^t\omega)$. Also the matrices $Q(t)$ are now random variables on the same space and we have that

$$Q(t, \omega) = P(0, \sigma^t\omega)P(0, \sigma^{t-1}\omega) \cdots P(0, \omega)$$

It is immediate to verify the following relation

$$Q(t + s, \omega) = Q(t, \sigma^s\omega)Q(s, \omega)\,, \quad \forall\omega \in S^{\mathbb{N}}\,,\ \forall s,\, t \in \mathbb{N} \tag{30}$$

Every sequence of random variable matrices satisfying (30) is said to be a cocycle. From the fact that $Q(t, \omega) \to \mathbb{1}\rho^*(\omega)$ for almost every ω together with the cocycle condition it immediately follows that

$$\rho^*(\sigma^s\omega)Q(s, \omega) = \rho^*(\omega) \tag{31}$$

for every $\omega \in S^{\mathbb{N}}$ and $s \in \mathbb{N}$.

Lemma 7.2. *$Q(t,\omega) - \mathbb{1}\rho^*(\omega)$ is a cocycle.*

Proof.

$$\begin{aligned}(Q(t,\sigma^s\omega) - \mathbb{1}\rho^*(\sigma^s\omega))(Q(s,\omega) - \mathbb{1}\rho^*(\omega)) &= Q(t,\sigma^s\omega)Q(s,\omega) - Q(t,\sigma^s\omega)\mathbb{1}\rho^*(\omega)\\ &- \mathbb{1}\rho^*(\sigma^s\omega)Q(s,\omega) + \mathbb{1}\rho^*(\sigma^s\omega)\mathbb{1}\rho^*(\omega)\\ &= Q(t+s,\omega) - \mathbb{1}\rho^*(\omega) - \mathbb{1}\rho^*(\omega) + \mathbb{1}\rho^*(\omega)\\ &= Q(t+s,\omega) - \mathbb{1}\rho^*(\omega)\end{aligned}$$

■

We have the following result encompassing Theorem 2.3.

Theorem 7.3. *There exists a non-negative constant λ such that*

$$\lim_{t\to+\infty} ||Q(t,\omega) - \mathbb{1}\rho^*(\omega)||^{1/t} = \lambda\,, \qquad \text{for almost every } \omega \tag{32}$$

Moreover, we have that

$$\lim_{t\to+\infty} ||x(t) - \mathbb{1}x_A(t)||^{1/t} = \lim_{t\to+\infty} ||x(t) - x(\infty)||^{1/t} = \lambda \tag{33}$$

for almost every pair $(\omega, x(0))$ with respect to the product measure $\mu^{\mathbb{N}}\otimes$Leb, where Leb *is the Lebesgue measure on $x(0)$.*

Proof. Consider

$$f_t(\omega) = \ln ||Q(t,\omega) - \mathbb{1}\rho^*(\omega)||$$

Because of Lemma 7.2, it follows that

$$\begin{aligned}f_{t+s}(\omega) &= \ln ||Q(t+s,\omega) - \mathbb{1}\rho^*(\omega)||\\ &= \ln ||(Q(t,\sigma^s\omega) - \mathbb{1}\rho^*(\sigma^s\omega))(Q(s,\omega) - \mathbb{1}\rho^*(\omega))||\\ &\le \ln ||(Q(t,\sigma^s\omega) - \mathbb{1}\rho^*(\sigma^s\omega))|| + \ln ||(Q(s,\omega) - \mathbb{1}\rho^*(\omega))||\\ &= f_t(\sigma^s\omega) + f_s(\omega)\end{aligned}$$

Applying Kingman's subadditive ergodic theorem together with the fact that measurable σ-invariant maps on $S^{\mathbb{N}}$ are constant almost surely, we obtain that there exists a constant θ (possibly equal to $-\infty$) such that $t^{-1}f_t \to \theta$ almost surely. This yields (32) with $\lambda = \exp(\theta) \ge 0$.

To complete the result, we now prove that

$$\lim_{t\to+\infty} \frac{\log\left(||(Q(t,\omega) - 1\rho(\omega)^*)x(0)||\right)}{t} = \theta \tag{34}$$

for almost every pair $(\omega, x(0))$. From this, taking exponentials and using the inequalities (8) the result will follow.

Consider the random subspace $V(\omega)$ which is the orthogonal to $\rho(\omega)$. Clearly, for every $x(0) \in V(\omega)$ we have that

$$(Q(t,\omega) - \mathbb{1}\rho^*(\omega))x(0) = Q(t,\omega)x(0)$$

In particular we have that $Q(t,\omega)x(0) \to 0$ for almost every ω and for every $x(0) \in V(\omega)$. Since $Q(t,\omega)\mathbb{1} = \mathbb{1}$, it follows that on $V(\omega)$ the rate of convergence to 0 is imposed by the second Lyapunov exponent [2] of $Q(t,\omega)$ (the first one being 1). More precisely, there exists a constant $\bar\theta$ and a proper subspace $\tilde V(\omega) \subseteq V(\omega)$ such that for almost every ω and for every $x(0) \in V(\omega) \setminus \tilde V(\omega)$, it holds

$$\begin{aligned}\lim_{t\to+\infty} \tfrac{\log(||(Q(t,\omega)-1\rho(\omega)^*)x(0)||)}{t} &= \lim_{t\to+\infty} \tfrac{\log(||(Q(t,\omega)x(0)||)}{t} = \bar\theta\,, \qquad \forall x(0) \in V(\omega)\setminus\tilde V(\omega)\\ \lim_{t\to+\infty} \tfrac{\log(||(Q(t,\omega)-1\rho(\omega)^*)x(0)||)}{t} &= \lim_{t\to+\infty} \tfrac{\log(||(Q(t,\omega)x(0)||)}{t} < \bar\theta\,, \qquad \forall x(0) \in \tilde V(\omega)\end{aligned}$$

Let $U(\omega)$ be the subspace generated by $\mathbb{1}$ and $\tilde V(\omega)$. If $x(0) \in \mathbb{R}^n \setminus U(\omega)$, we can write $x(0) = \alpha\mathbb{1} + v$ where $v \in V(\omega) \setminus \tilde V(\omega)$. Hence, for almost every ω, we also have that

$$\lim_{t\to+\infty} \frac{\log\left(||(Q(t,\omega) - 1\rho(\omega)^*)x(0)||\right)}{t} = \lim_{t\to+\infty} \frac{\log\left(||(Q(t,\omega) - 1\rho(\omega)^*)v||\right)}{t} = \bar\theta \tag{35}$$

On the other hand, if instead $x(0) \in U(\omega)$, for almost every ω we have that

$$\lim_{t\to+\infty} \frac{\log\left(||(Q(t,\omega) - 1\rho(\omega)^*)x(0)||\right)}{t} < \bar{\theta} \tag{36}$$

It is clear that $\bar{\theta} \leq \theta$. To prove that they are equal we proceed as follows. Let $x_{\omega,t} \in \mathbb{R}^N$ be such that for every ω and t,

$$||x_{\omega,t}|| = 1\,, \quad ||(Q(t,\omega) - \mathbb{1}\rho^*(\omega))x_{\omega,t}|| = ||Q(t,\omega) - \mathbb{1}\rho^*(\omega)||$$

For every ω, pick a sequence $t_n \to +\infty$ such that $x_{\omega,t_n} \to x_\omega$. We can estimate

$$\begin{aligned} ||(Q(t_n,\omega) - \mathbb{1}\rho^*(\omega))x_\omega|| &\geq ||Q(t_n,\omega) - \mathbb{1}\rho^*(\omega)|| - ||(Q(t_n,\omega) - \mathbb{1}\rho^*(\omega))(x_{\omega,t_n} - x_\omega)|| \\ &\geq ||Q(t_n,\omega) - \mathbb{1}\rho^*(\omega)||\,(1 - ||(x_{\omega,t_n} - x_\omega)||) \end{aligned}$$

Hence, for n sufficiently large we have that

$$||(Q(t_n,\omega) - \mathbb{1}\rho^*(\omega))x_\omega|| \geq \frac{1}{2}||Q(t_n,\omega) - \mathbb{1}\rho^*(\omega)|| \tag{37}$$

Combining (37) with (35) and (36), we immediately get that $\bar{\theta} \geq \theta$. This completes our result. ■

References

[1] N. Alon and J. Spencer. *The probabilistic method*. John Wiley, 2000.

[2] L. Arnold. *Random dynamical systems*. Springer-Verlag, 1998.

[3] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.. Tsitsiklis. Convergence in multiagent coordination, consensus and flocking. In *Proceedings of joint IEEE-CDC/ECC*, pages 2996–3000, 2005.

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

[5] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, 2007. To be published.

[6] B. Van Roy C.C. Moallemi. Consensus propagation. *IEEE Transactions on Information Theory*, 52(11):4753–4766, 2006.

[7] Robert Cogburn. On products of random stochastic matrices. In *Random matrices and their applications (Brunswick, Maine, 1984)*, volume 50 of *Contemp. Math.*, pages 199–213. Amer. Math. Soc., 1986.

[8] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *Journal of parallel and distributed computing*, 7:279–301, 1989.

[9] R. Diekmann, A. Frommer, and B. Monien. Efficient schemes for nearest neighbor load balancing. *Parallel computing*, 25:789–812, 1999.

[10] A. Dymakis, A.D. Sarwate, and M.J. Wainright. Geographic gossip: efficient aggregation on sensor networks. Technical report, EECS Berkeley, 2006.

[11] F. Fagnani and S. Zampieri. Average consensus with packet drop communication. 2006. submitted.

[12] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. In *Proceedings Information Theory and Application Workshop*, 2007.

[13] Y. Hatano and M. Mesbahi. Agreement of random networks. In *IEEE Conference on Decision and Control*, 2004.

[14] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *Proceedings ACM/IEEE Conf. Mobile Computing and Networking*, pages 56–67, 2000.

[15] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[16] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 1–10, 2003.

[17] S. Lall L. Xiao, S. Boyd. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. Int. Conf. Information Processing in Sensor Networks*, 2005.

[18] Q. Li and D. Rus. Global clock syncronization in sensor networks. *IEEE Transaction on Computers*, 55(2):214– 226, 2006.

[19] J. Lin, A. S. Morse, and B. D. O. Anderson. The multi-agent rendezvous problem: an extended summary. In A. S. Morse V. Kumar, N. E. Leonard, editor, *Proceedings of the 2003 Block Island Workshop on Cooperative Control*, volume 309 of *Lecture Notes in Control and Information Sciences*, pages 257–282. New York: Springer Verlag, 2004.

[20] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50:169–182, 2005.

[21] S. Muthukrishnan, B. Ghosh, and M. Schultz. First and second order diffusive methods for rapid, coarse, distributed load balancing. *Theory of computing systems*, 31:331–354, 1998.

[22] R. Olfati-Saber. Distributed kalman filter with embedded consensus filters. In *IEEE Conference on Decision and Control*, pages 63–70, 2005.

[23] 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.

[24] W. Ren and R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transaction on Automatic Control*, 50(5), 2005.

[25] D. Ruelle. Ergodic theory on differentiable dynamical systems. *IHES Publicationes Mathematiques*, 50:275–320, 1979.

[26] S. L. Smith, M. E. Broucke, and B. A. Francis. A hierarchical cyclic pursuit scheme for vehicle networks. *Automatica*, 41(6):1045–1053, 2005.

[27] J. Tsitsiklis. *Problems in decentralized decision making and computation.* PhD thesis, Department of EECs, MIT, 1984.

[28] J. Zhao, R. Govindan, and D. Estrin. Computing aggregates for monitoring wireless sensor networks. In *Proc. Int. Workshop on Sensor Net Protocols and Applications*, pages 139–148, 2003.