Randomized consensus algorithms over large scale networks.

Fabio Fagnani
Dipartimento di Matematica
Politecnico di Torino
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
Email: fabio.fagnani@polito.it

Sandro Zampieri
Department of Information Engineering
Università di Padova
Via Gradenigo 6/a, 35131 Padova, Italy
Email: zampi@dei.unipd.it

***Abstract*— Suppose we have a directed graph G with set of nodes $V = \{1, \dots, N\}$ and a measure x_i for every node $i \in V$. The average consensus problem consists in computing the average $x_A = N^{-1} \sum_i x_i$ in an iterative way, exchanging information among nodes exclusively along the available edges in G. This problem appears in a number of different contexts since the 80's (decentralized computation, load balancing, clock syncronization) and, recently, has attracted much attention for possible applications to sensor networks (data fusion problems) and to coordinated control for mobile autonomous agents. Several algorithms for average consensus can be found in the literature: they differentiate on the basis of the amount of communication and computation they use, on their scalability with respect to the number of nodes, on their adaptability to time-varying graphs, and, finally, they can be deterministic or random. In this presentation we will focus on random algorithms: we will review some algorithms present in the literature and we will propose some new ones. We will present some performance results which will allow to make some comparison. Finally, we will establish some probabilistic concentration results which will give a stronger significance to previous results.**

I. Introduction

Suppose we have a (directed) graph $\mathcal{G}$ with set of nodes $V = \{1, \dots, N\}$ and a measure x_i for every node $i \in V$. The average consensus problem consists in computing the average $x_A = N^{-1} \sum_i x_i$ in an iterative 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 [1], load balancing [2], [3], [4]) and, recently, has attracted much attention for possible applications to sensor networks (data fusion problems [5], [6], [7], [8], [9], clock syncronization [10]) and to coordinated control for mobile autonomous agents [11], [12], [13], [14], [15], [16], [17]. Other places where consensus algorithms have been studied in general are [18], [19], [20], [21], [22], [23]

Several algorithms for average consensus can be found in the literature: they differentiate on the basis of the amount of communication and computation they use, on their scalability with respect to the number of nodes, on their adaptability to time-varying graphs, and, finally, they can be deterministic or random.

We now briefly review two of the possible applications of average consensus, to better understand which are the important issues of this problem. In load balancing the nodes can be thought as identical processors, computers, and the edges as physical connections among themselves. 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 measure x_i at each node is in this case the amount of tasks (all considered to be equal) 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 equilibrate as much as possible the tasks among the various processors. The natural goal is that each processor will have at the end the same quantity of tasks to work on namely a quantity of task close to the average x_A. There are two different approaches to this problem. In the first approach each processor evaluate the average x_A with an iterative consensus algorithm and afterwords, there is a physical movements of tasks among processors. In the second approach instead the movement of tasks is coupled to the evolution of the iterative algorithm: this physical transfer forces the iterative algorithm to be inherently symmetric with respect to any pair of communicating processors.

In the context of sensor networks, nodes are sensors deployed (often randomly) in some geographical area. They typically transmit in a wireless fashion and the common adapted model is that they can communicate to the other sensors within a distance R. The communication graph obtained in this case is typically a random graph: a good model is the geometric graph. The quantities x_i they want to average can be in this case some measurement all the sensors have done (e.g. a temperature) and the averaging is done in order to increase precision, by filtering out the noise. In other cases they may want 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 which have to be kept as low as possible. 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 environment) and in some applications it cannot simultaneously transmit to more than a node (this happens for instance for processors nets). Thus, even in case where the communication graph is quite dense, algorithms should take into considerations these fundamental limitations. This fact makes the use of random 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 represent the position of node i and the goal is to make physically the agents meet in 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. Indeed, a natural model for the communication graph is the geometric one as for sensor networks: each agent can only talk to the other agents within a given distance. However, since the agents change their position, their communication graph also changes: as a consequence, the resulting dynamical system is in general much more complicated. Except the studies in [24] where the authors consider algorithms which deliberately prevent the breaking of previous established communication edges, we believe there does not exist rigorous mathematical analysis of these models except for quite simplified scenarios where the graph variation is decoupled from the dynamics.

For the purpose of our paper the applications context we have in mind are those connected to sensor and computer networks and not to mobile agents scenario.

Deterministic (time-invariant and time-varying) consensus algorithms have been studied in many papers: starting from the pioneering work [1], many variations can be found in above cited literature. Most of 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 [4], [3], [22]. The type of problem typically faces in the literature are: necessary and sufficient conditions for convergence, speed of convergence, optimization issues. On the other hand random linear schemes have been studied for instance in [5], [21], [9] under the name of gossip algorithms. In this case the evolution matrix of the algorithm is changed randomly at every clock step: convergence is now considered in a probabilistic sense and performance is studied in mean square sense or in terms of a sort of contraction time. The algorithms studied in the literature assume symmetric communication graph and lead in general to symmetric evolution matrices which preserve the global average over time. Symmetry is fundamental in certain applications as, fr instance, the second approach to load balancing discussed above. However in other situations, symmetry may not be so important and actually an undesirable feature in situations where communications are asymmetric (this happens for instance in sensor networks). Also the related property of achieving exactly the average can be a bit relaxed: in some situations it may be sufficient to converge to some value sufficiently close to the average.

In this paper we will focus to random first order linear consensus algorithms as in [21]. However, differently from [21] we will not focus exclusively on average consensus: we will consider more general consensus algorithms which do not converge to the average, but, under certain circumstances, to some good approximation of it. In this paper we will focus on three examples. The first one is the symmetric gossip model studied in [21]: at each clock step, a communication edge is randomly chosen and the two nodes average their estimations of the average. This algorithm clearly needs symmetric links and maintain the global average. The second model is what we called in-gossip: at each time step, each node receives data from one randomly chosen neighbor and perform the average. In this case symmetric communication is not needed. However it does not preserve the global average. Finally, the third proposed model is a sort of broadcasting one: at each time step a randomly chosen node activates and sends its data to all its neighbors. As the in-gossip model, it is asymmetric and does not preserve the global average.

For the various algorithms we propose a mean square analysis. We first analyze the special case when the communication graph is the complete graph: in this case we obtain complete results in term of speed of convergence and average displacement which allow to make sensible comparison among the three schemes.

We then pass to consider more general graphs. Here we focus on the family of Abelian Cayley graphs which already contains some interesting graph architectures (ring, torus, hypercube). We show that for such graphs analytical computations of mean square quantities can in principle be obtained. Results in this case are however only at a primitive steps. We propose some numerical evaluations of the speed of convergence in the case of a ring communication graph.

We then present a probabilistic concentration analysis which shows how with probability one the behavior of our system converges to the mean square analysis when the number of nodes $N \to +\infty$ while time t remains fixed or is anyhow of type $t = o(N)$. This means that the mean square analysis is meaningful in the scenario $N >> t$. Therefore our analysis is particularly meaningful for large scale networks.

II. Problem formulation

A. Linear consensus algorithms

The iterative consensus algorithms we will study in this paper consist of N coupled linear dynamical systems (as many as the nodes)

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

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 $x(0) \in \mathcal{I}$ then $x(t) = x(0)$ for every $t \in \mathbb{N}$.

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

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

where $\mathbb{1} := (1, \dots, 1)^T$ and $\mathcal{I}$ is the subspace generated by $\mathbb{1}$. 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.

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, almost surely,

$$\lim_{t\to\infty} x(t) = \alpha \mathbb{1}\,. \tag{3}$$

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: namely we assume that $P(t)_{ij} \geq 0$ for every i and j and $P(t)\mathbb{1} = \mathbb{1}$. Notice that condition (a) is then clearly automatically satisfied. If, moreover, $\mathbb{1}^* P(t) = \mathbb{1}^*$, we say that $P(t)$ is doubly stochastic. In this case we have that the average is invariant: $\mathbb{1}^* x(t) = \mathbb{1}^* x(0)$ for every t. Hence, if (b) or (b') holds, then automatically $\alpha = N^{-1}\mathbb{1}^* x(0)$.

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 with all rows equal to ρ^*. ρ is called the random asymptotic weight associated with $P(t)$. We have probabilistic average consensus exactly when $\rho = N^{-1}\mathbb{1}$.

B. 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. If we use a consensus algorithm $P(t)$, we are assuming that at instant t all communications along the edges of $\mathcal{G}_{P(t)}$ are feasible. The amount of non zero elements in $P(t)$ is thus a 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 $\mathcal{G}_{P(t)}$ is a subgraph of $\mathcal{G}$ for every instant t. In the sequel we will assume that every self loop is always in $\mathcal{G}$ (we assume that every agent has always access to its own data).

For future utility we need to set some basic notation on 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

$$\begin{aligned} N_i^+ &= \{j \in V \setminus \{i\} \mid (i, j) \in E\} \\ N_i^- &= \{j \in V \setminus \{i\} \mid (j, i) \in E\}\,. \end{aligned}$$

Elements in N_i^+ (resp. in N_i^-) are called out-neighbors (resp. in-neighbors) of i. Moreover we put $\nu_i^+ = |N_i^+|$, $\nu_i^- = |N_i^-|$. 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

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

Notice that self loops have not been considered: this will make notation simpler further on. The in-degree and out-degree matrices are defined, respectively as

$$D_{\mathcal{G}-} = \sum_i \nu_i^- e_i e_i^*\,. \quad D_{\mathcal{G}+} = \sum_i \nu_i^+ e_i e_i^*\,.$$

If it happens that, whenever $(i, j) \in E$, then also, $(j, i) \in E$, we will call the graph symmetric (or undirected). In this case we will drop the superscript $\pm$ in the above notations.

All the Examples considered in this paper will deal with symmetric graphs. However most of the theoretical results we will present do actually apply also to directed graphs.

III. Random algorithms achieving consensus

A. Conditions for the probabilistic consensus

We start recalling some well known facts on consensus algorithms in case when $P(t) = P$ is constant (see [20] for more details).

Proposition III.1. *Let P be an $N \times N$ stochastic matrix. The following conditions are equivalent:*

(1) *P achieves consensus.*

(2) *For every $i, j = 1, \dots, N$, there exist k, t such that*

$$P_{ik}^t P_{jk}^t > 0\,.$$

(3) *$\mathcal{G}_P$ is a connected directed graph such that the tree of the strongly connected components has just one source which is aperiodic.*

If we start from a graph $\mathcal{G}$ satisfying property (3) above, it is very simple to construct a P achieving consensus: it is sufficient to make sure that $\mathcal{G}_P = \mathcal{G}$. One possibility is for instance to consider

$$P = k_0 I + (1-k_0) D_{\mathcal{G}^-}^{-1} A_{\mathcal{G}}\,,$$

where $k_0 \in (0,1)$ is an arbitrarily chosen parameter. If, moreover, $\mathcal{G}$ is strongly connected, we can also make sure that we obtain a doubly stochastic matrix so that average consensus is indeed achieved: this is slightly more complicated (see [20] for details). In the simple case when $\mathcal{G}$ is strongly connected and symmetric, however we can construct a symmetric solution P as follows. Put

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

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

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

Such a P achieves average consensus.

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 [25]:

Theorem III.2. *The algorithm $P(t)$ achieves probabilistic consensus if and only if for every $i, j \in V$ we have that*

$$\mathbb{P}(\Omega_{ij}) = 1$$

where

$$\Omega_{ij} = \{\exists k\,,\ \exists t\,,\ |Q_{ik}(t) Q_{jk}(t) > 0\}\,.$$

To obtain a more handy condition, we consider the average $\bar{P} = \mathbb{E}(P(t))$ and the average dynamics $m(t+1) = \bar{P} m(t)$. We have the following result:

Corollary III.3. *The following conditions are equivalent:*

(1) *$P(t)$ achieves probabilistic consensus.*
(2) *$\bar{P}$ achieves consensus*

Proof: (1)⇒(2) immediately follows from Lebesgue dominated convergence theorem.

(2)⇒(1): If $\bar{P}$ achieves consensus it follows from Proposition III.1 that for every i, j, there exist k, t such that

$$\bar{P}^t_{ik} \bar{P}^t_{jk} > 0\,.$$

As a consequence, $\mathbb{P}(\Omega_{ij}) > 0$. On the other hand it is immediate to check that Ω_{ij} is a tail event. Hence, for the Kolmogorov $0-1$ law, we must have $\mathbb{P}(\Omega_{ij}) = 1$. By previous Theorem, (1) then follows. ■

Notice that, if $Q(t) \to \mathbb{1}\rho^*$, we have that $\bar{P}^t = \mathbb{E}(Q(t)) \to \mathbb{1}\mathbb{E}\rho^*$. It can very well happen that $E\rho = \mathbb{1}$ even if ρ is not equal to $\mathbb{1}$ a.s. In other terms even if $\bar{P}$ achieves average consensus, not necessarily, $P(t)$ will also achieve average probabilistic consensus. This will appear in the examples we will propose.

B. Examples

We now present a number of examples on which most of the paper will be focused on.

Example III.4: The symmetric gossip model This is the example studied in [21]. We start from a symmetric graph $\mathcal{G} = (V, E)$ and we assume that at every time instant a node among the N possible is chosen randomly. This node then chooses also randomly one of its neighbors, it establishes a bidirectional link with it and they average their quantity. More precisely, let, for every $(i,j) \in E$,

$$R^{ij} = I - k_0 (e_i - e_j)(e_i - e_j)^*\,,$$

(where $k_0 \in (0,1)$). 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 have that

$$\bar{P} = \sum_{(i,j)\in E} \frac{1}{2N}\left[\frac{1}{\nu_i} + \frac{1}{\nu_j}\right]\left[I - k_0 (e^i - e^j)(e^i - e^j)^*\right]$$

Notice that, if $(h,k) \in E$ and $h \neq k$,

$$\bar{P}_{hk} = \frac{1}{N}\left[\frac{1}{\nu_h} + \frac{1}{\nu_k}\right] k_0\,.$$

A remark regarding the parameter k_0: it has to be considered as a sort of design parameter respect to which optimize performance. This will also appear in the next two examples.

Example III.5: The in-gossip model This example was studied in [20]. We start from any directed graph $\mathcal{G} = (V, E)$ and we assume that at every instant, each system receives the state of another system chosen randomly and independently at each time instant among its possible in-neighbors.

This is modelled in the following way. Fix, for every $i \in V$ an in-neighbor λ_i. To the configuration $\lambda = (\lambda_1, \ldots, \lambda_N)$ we then associate

$$E^\lambda = \sum_i e_i e^*_{\lambda_i}\,.$$

$P(t)$ is this time uniformly concentrated on the matrices

$$(1-k_0) I + k_0 E^\lambda\,.$$

In this case we obtain

$$\bar{P} = \sum_\lambda \frac{1}{\prod_i \nu_i^-}[(1-k_0)I + k_0 E^\lambda] = (1-k_0)I + k_0 D_{\mathcal{G}^-}^{-1} A_{\mathcal{G}}\,.$$

Example III.6: The broadcasting model We start from any directed graph $\mathcal{G} = (V, E)$ and we assume that that at every time instant a node among the N possible is chosen randomly. This node then broadcast its value to all its out-neighbors. $P(t)$ is this time uniformly concentrated on the N matrices

$$I + k_0 \sum_{j \in N_i^+} (e_j e_i^* - e_j e_j^*)\,.$$

In this case we have

$$\bar{P} = I + \frac{k_0}{N}\sum_i \sum_{j\in N_i^+}(e_j e_i^* - e_j e_j^*) = I + \frac{k_0}{N}[A_{\mathcal{G}} - D_{\mathcal{G}-}]\,.$$

Corollary III.7. *If the graph $\mathcal{G}$ is strongly connected, then the algorithms in Examples III.4, III.5, and III.6 all achieve probabilistic consensus.*

Notice that Example III.4 yields doubly stochastic matrices: average in this case is always preserved. Examples III.5 and III.6 instead do not have this feature even if the communication graph is symmetric.

IV. The mean square performance

We will measure the performance of a particular algorithm $P(t)$ achieving probabilistic consensus by considering two figures. The first figure we consider is a normalized version of the distance from the consensus

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

Consider now the centroid $x_A(t) = N^{-1}\sum_i x_i(t)$. The second figure we will consider is the centroid displacement from its initial value

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

These two figures will be now analyzed by considering their expectations: $\mathbb{E}[d(t)]$, $\mathbb{E}[\beta(t)]$.

A. Evolution of $\mathbb{E}[d(t)]$

We are interested in studying $\mathbb{E}[||x(t) - \mathbb{1}x_A(t)||^2]$ and, in particular, their exponential rate of convergence:

$$R = \limsup_{t\to+\infty}(\mathbb{E}[||x(t) - \mathbb{1}x_A(t)||^2])^{1/t}\,.$$

Notice that

$$\begin{aligned}\mathbb{E}[||x(t) - \mathbb{1}x_A(t)||^2] &= \mathbb{E}[x^*(t)(I - N^{-1}\mathbb{1}\mathbb{1}^*)x(t)]\\ &= x^*(0)\Delta(t)x(0)\end{aligned}$$

where

$$\begin{aligned}\Delta(t) := \;& \mathbb{E}[P(0)^*P(1)^*\cdots P(t-1)^*\cdot\\ &\cdot(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(1)P(0)]\end{aligned}$$

if $t \geq 1$ and where $\Delta(0) := (I - N^{-1}\mathbb{1}\mathbb{1}^*)$. 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

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

If now we consider the reachable subspace $\mathcal{R}$ of the pair $(\mathcal{L}, \Delta(0))$, namely the smallest $\mathcal{L}$-invariant subspace of $\mathbb{R}^{N\times N}$ containing $\Delta(0)$, we clearly have that

$$R = \max\{|\lambda| \;:\; \lambda\,\text{eigenvalue of}\;\; \mathcal{L}_{|\mathcal{R}}\}\,.$$

Consider the canonical basis of $\mathbb{R}^{N\times N}$: $\{e_{hk}\}$ (which is the matrix with all elements equal to 0 except a 1 in position (h,k)) we have that $\mathcal{L}$ is represented by a non-negative matrix. Indeed,

$$(\mathcal{L}(e_{hk}))_{ij} = \mathbb{E}[P_{hi}(0)P_{kj}(0)] \geq 0\,.$$

Moreover,

$$\sum_{(i,j)}(\mathcal{L}(e_{hk}))_{ij} = 1\,,$$

namely the transpose matrix is also stochastic. 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 IV.1. *Assume that $P(t)$ achieves consensus with random asymptotic weight ρ. Then, the eigenspace of $\mathcal{L}$ relative to 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 $x(0)^*\mathcal{L}^t(\Delta)x(0) = \mathbb{E}[x(t)^*\Delta x(t)]$. Since $x(t) \to \mathbb{1}\rho^*x(0)$ in mean square sense, it follows that

$$\mathbb{E}[x(t)^*\Delta x(t)] \to x(0)^*\mathbb{1}^*\Delta\mathbb{1}\mathbb{E}[\rho\rho^*]x(0)\,.$$

Hence,

$$\lim \mathcal{L}^t(\Delta) = (\mathbb{1}^*\Delta\mathbb{1})\mathbb{E}[\rho\rho^*]\,.$$

This proves the result. ∎

B. The average displacement

For what concerns the average displacement, we are interesting in evaluating the following figure

$$\begin{aligned}\delta &= \mathbb{E}[|x_A(\infty) - x_A(0)|^2]\\ &= \mathbb{E}[|(\rho^* - N^{-1}\mathbb{1}^*)x(0)|^2] = x(0)^*Bx(0)\end{aligned} \tag{6}$$

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 eigenvectors of $\bar{P}$ and $\mathcal{L}$, respectively.

In the case when $\bar{P}$ achieves the average consensus, in particular we obtain

$$B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^*\,. \tag{7}$$

V. Analysis of the Examples for the complete graph

In this section we analyze the simplest case when actually no communication constraint is pre-imposed at the communication level. We will show that in this case a complete analysis can be carried on for previous examples: both for the mean square evolution and for the average displacement. A fundamental fact which is common to all our examples is that, for the complete graph, the operator $\mathcal{L}$ keeps invariant the subspace generated by I and $N^{-1}\mathbb{1}\mathbb{1}^*$. Everything thus reduce to a 2×2 matrix. We will use the following trivial fact:

Lemma V.1. *Let*

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}$$

be a matrix with non-negative elements such that $a+c=1$ and $b+d=1$. Then, its eigenvalues are 1 and $d-c$. Moreover, $(b,c)^*$ is an eigenvector of 1 .

In the sequel of this chapter we assume that $\mathcal{G}$ is the complete graph, so that $A_{\mathcal{G}} = \mathbb{1}\mathbb{1}^* - I$ and $\nu_j^+ = \nu_j^- = N-1$ for every $j \in V$.

A. The symmetric gossip model

We have the following result:

Proposition V.2. *Assume that $\Delta = \alpha I + \beta N^{-1}\mathbb{1}\mathbb{1}^*$, then,*

$$\Delta^+ = \mathbb{E}[P(t)^*\Delta P(t)] = \alpha^+ I + \beta^+ N^{-1}\mathbb{1}\mathbb{1}^*$$

where

$$\begin{pmatrix}\alpha^+\\ \beta^+\end{pmatrix} = \begin{pmatrix}1-\frac{4k_0(1-k_0)}{(N-1)} & 0\\ \frac{4k_0(1-k_0)}{N-1} & 1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

Proof: Notice that

$$\begin{aligned}
&\mathbb{E}[P(t)^*P(t)] = \tfrac{1}{N(N-1)}\textstyle\sum_{i\neq j}[I-k_0(e_i-e_j)(e_i-e_j)^*]^2\\
&= I - 2k_0\tfrac{1}{N(N-1)}\textstyle\sum_{i\neq j}(e_i-e_j)(e_i-e_j)^*\\
&+2k_0^2\tfrac{1}{N(N-1)}\textstyle\sum_{i\neq j}(e_i-e_j)(e_i-e_j)^*\\
&= I-2k_0(1-k_0)\tfrac{1}{N(N-1)}\textstyle\sum_{i\neq j}(e_ie_i^*+e_je_j^*-e_je_i^*-e_ie_j^*)\\
&= I-2k_0(1-k_0)\tfrac{1}{N(N-1)}(2(N-1)I-2\mathbb{1}\mathbb{1}^*+2I)\\
&= (1-2k_0(1-k_0)\tfrac{2}{(N-1)})I + 2k_0(1-k_0)\tfrac{2}{N(N-1)}\mathbb{1}\mathbb{1}^*.
\end{aligned}$$

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

This implies the claim. ■

In this case we have that $\Delta(t)$ converges to $N^{-1}\mathbb{1}\mathbb{1}^*$ and the speed of convergence is given by the second eigenvalue

$$R = 1-4k_0(1-k_0)\frac{1}{(N-1)}$$

which is minimum for $k_0 = 1/2$ for which we obtain

$$R = 1-\frac{1}{(N-1)}.$$

We recall that in this case, all matrices are doubly stochastic. This yields in particular $\delta = 0$.

B. The in-gossip model

We have the following result:

Proposition V.3. *Assume that $\Delta = \alpha I + \beta N^{-1}\mathbb{1}\mathbb{1}^*$, then,*

$$\Delta^+ = \mathbb{E}[P(t)^*\Delta P(t)] = \alpha^+ I + \beta^+ N^{-1}\mathbb{1}\mathbb{1}^*$$

where

$$\begin{pmatrix}\alpha^+\\ \beta^+\end{pmatrix} = \begin{pmatrix}(1-k_0)^2+k_0^2-\frac{2k_0(1-k_0)}{(N-1)} & \frac{k_0^2(N-2)}{(N-1)^2}\\ \frac{2k_0(1-k_0)N}{(N-1)} & 1-\frac{k_0^2(N-2)}{(N-1)^2}\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}. \tag{8}$$

Proof: Notice that

$$\begin{aligned}
&\mathbb{E}[P(t)^*P(t)]\\
&= \tfrac{1}{(N-1)^N}\sum_\lambda[(1-k_0)I+k_0E^{\lambda*}][(1-k_0)I+k_0E^\lambda]\\
&= (1-k_0)^2I + k_0(1-k_0)\tfrac{1}{(N-1)^N}\sum_\lambda[E^{\lambda*}+E^\lambda]+\\
&k_0^2\tfrac{1}{(N-1)^N}\sum_\lambda E^{\lambda*}E^\lambda
\end{aligned}$$

Now,

$$\begin{aligned}
&\tfrac{1}{(N-1)^N}\sum_\lambda E^{\lambda*}E^\lambda = \tfrac{1}{(N-1)^N}\sum_\lambda\sum_i e_{\lambda_i}e_{\lambda_i}^*\\
&= \tfrac{1}{(N-1)}\sum_i(I-e_ie_i^*) = \tfrac{N}{N-1}I - \tfrac{1}{N-1}I = I.
\end{aligned}$$

Substituting we obtain

$$\begin{aligned}
&\mathbb{E}[P(t)^*P(t)] = (1-k_0^2)I + 2k_0(1-k_0)\tfrac{1}{N-1}[\mathbb{1}\mathbb{1}^*-I]+k_0^2I\\
&= [(1-k_0^2)+k_0^2-2k_0(1-k_0)\tfrac{1}{N-1}]I+2k_0(1-k_0)\tfrac{1}{N-1}\mathbb{1}\mathbb{1}^*
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}[P(t)^*\mathbb{1}\mathbb{1}^*P(t)] = (1-k_0)^2\mathbb{1}\mathbb{1}^*\\
&+k_0(1-k_0)\tfrac{1}{(N-1)^N}\sum_\lambda[E^{\lambda*}\mathbb{1}\mathbb{1}^*+\mathbb{1}\mathbb{1}^*E^\lambda]\\
&+k_0^2\tfrac{1}{(N-1)^N}\sum_\lambda E^{\lambda*}\mathbb{1}\mathbb{1}^*E^\lambda.
\end{aligned}$$

We have

$$\begin{aligned}
&\tfrac{1}{(N-1)^N}\sum_\lambda E^{\lambda*}\mathbb{1}\mathbb{1}^*E^\lambda = \tfrac{1}{(N-1)^N}\sum_\lambda\sum_{i,j}e_{\lambda_i}e_{\lambda_j}^*\\
&= \tfrac{1}{(N-1)^2}\textstyle\sum_{i\neq j}(\sum_{\lambda_i}e_{\lambda_i})(\sum_{\lambda_j}e_{\lambda_j})^* + \tfrac{1}{(N-1)}\sum_i\sum_{\lambda_i}e_{\lambda_i}e_{\lambda_i}^*\\
&= \tfrac{1}{(N-1)^2}\textstyle\sum_{i\neq j}(\mathbb{1}-e_i)(\mathbb{1}^*-e_j^*) + \tfrac{1}{(N-1)}\sum_i(I-e_ie_i^*)\\
&= \mathbb{1}\mathbb{1}^* - \tfrac{2}{N-1}\mathbb{1}\mathbb{1}^* + \tfrac{1}{(N-1)^2}(\mathbb{1}\mathbb{1}^*-I)+I\\
&= (\tfrac{N}{N-1}-\tfrac{1}{(N-1)^2})I + (1-\tfrac{2}{N-1}+\tfrac{1}{(N-1)^2})\mathbb{1}\mathbb{1}^*
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}[P(t)^*\mathbb{1}\mathbb{1}^*P(t)] = (1-k_0)^2\mathbb{1}\mathbb{1}^*\\
&+k_0(1-k_0)\tfrac{1}{(N-1)}[(\mathbb{1}\mathbb{1}^*-I)\mathbb{1}\mathbb{1}^*+\mathbb{1}\mathbb{1}^*(\mathbb{1}\mathbb{1}^*-I)]\\
&+k_0^2(1-\tfrac{1}{(N-1)^2})I + k_0^2(\tfrac{N}{N-1}-\tfrac{2}{N-1}+\tfrac{1}{(N-1)^2})\mathbb{1}\mathbb{1}^*\\
&= \tfrac{k_0^2N(N-2)}{(N-1)^2}I + [1-\tfrac{k_0^2(N-2)}{(N-1)^2}]\mathbb{1}\mathbb{1}^*
\end{aligned}$$

This proves the result. ■

It follows from Lemma V.1 that in this case

$$R = 1-\frac{k_0^2(N-2)}{(N-1)^2}-\frac{2k_0(1-k_0)N}{N-1}.$$

Instead B can be computed as follows:

$$\begin{aligned}
&B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^* = \frac{k_0^2(N-2)}{N(N-1)^2(\frac{k_0^2(N-2)}{(N-1)^2}+\frac{2k_0(1-k_0)N}{N-1})}\\
&\cdot[I-N^{-1}\mathbb{1}\mathbb{1}^*] \simeq \tfrac{k_0}{2(1-k_0)}\tfrac{1}{N^2}[I-N^{-1}\mathbb{1}\mathbb{1}^*].
\end{aligned}$$

Hence,

$$\delta \le \frac{k_0}{(1-k_0)}\frac{1}{N^2}||x(0)||^2 \le \frac{k_0}{(1-k_0)}\frac{1}{N}||x(0)||_\infty^2.$$

This in particular shows that if we assume that all initial measures are bounded by a common fixed value M, the average displacement goes to 0 as $1/N$ when $N \to +\infty$.

C. The broadcasting model

We have the following result:

Proposition V.4. *Assume that $\Delta = \alpha I + \beta N^{-1}\mathbb{1}\mathbb{1}^*$, then,*

$$\Delta^+ = \mathbb{E}[P(t)^*\Delta P(t)] = \alpha^+ I + \beta^+ N^{-1}\mathbb{1}\mathbb{1}^*$$

where

$$\begin{pmatrix}\alpha^+\\ \beta^+\end{pmatrix} = \begin{pmatrix}1-2k_0(1-k_0) & k_0^2\\ 2k_0(1-k_0) & 1-k_0^2\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

Proof: Notice that

$$\begin{aligned}&\mathbb{E}[P(t)^*P(t)] = \tfrac{1}{N}\sum_i (I + k_0 \sum_{j\neq i}(e_je_i^* - e_je_j^*)^*)\cdot\\ &\cdot(I + k_0 \sum_{j'\neq i}(e_{j'}e_i^* - e_{j'}e_{j'}^*))\\ &= I + \tfrac{2k_0}{N}(\mathbb{1}\mathbb{1}^* - I - (N-1)I)\\ &+\tfrac{k_0^2}{N}\sum_i\sum_{j\neq i}\sum_{j'\neq i}(e_ie_j^* - e_je_j^*)(e_{j'}e_i^* - e_{j'}e_{j'}^*)\\ &= I + \tfrac{2k_0}{N}(\mathbb{1}\mathbb{1}^* - NI)\\ &+\tfrac{k_0^2}{N}\sum_i\sum_{j\neq i}(e_ie_i^* - e_ie_j^* - e_je_i^* + e_je_j^*)\\ &= I + \tfrac{2k_0}{N}(\mathbb{1}\mathbb{1}^* - NI) + \tfrac{k_0^2}{N}(2(N-1)I - 2(\mathbb{1}\mathbb{1}^* - I))\\ &= [1 - 2k_0(1-k_0)]I + \tfrac{2k_0(1-k_0)}{N}\mathbb{1}\mathbb{1}^*\end{aligned}$$

Similarly,

$$\begin{aligned}&\mathbb{E}[P(t)^*\mathbb{1}\mathbb{1}^*P(t)] = \mathbb{1}\mathbb{1}^* + \tfrac{2k_0}{N}(\mathbb{1}\mathbb{1}^* - I - (N-1)I)\mathbb{1}\mathbb{1}^*\\ &+\tfrac{k_0^2}{N}\sum_i\sum_{j\neq i}\sum_{j'\neq i}(e_ie_j^* - e_je_j^*)\mathbb{1}\mathbb{1}^*(e_{j'}e_i^* - e_{j'}e_{j'}^*)\\ &= \mathbb{1}\mathbb{1}^* + \tfrac{k_0^2}{N}\sum_i\sum_{j\neq i}\sum_{j'\neq i}(e_ie_i^* - e_ie_{j'}^* - e_je_i^* + e_je_{j'}^*)\\ &= \mathbb{1}\mathbb{1}^* + \tfrac{k_0^2}{N}[(N-1)^2I - 2(N-1)(\mathbb{1}\mathbb{1}^* - I)\\ &+(N-2)(\mathbb{1}\mathbb{1}^* - I) + (N-1)I\end{aligned}$$

From this the thesis easily follows. ∎

In this case, using Lemma V.1 we obtain that $R = (1-k_0)^2$ independent of N. We now evaluate B:

$$B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^* = \frac{k_0}{2-k_0}\frac{1}{N}[I - N^{-1}\mathbb{1}\mathbb{1}^*]\,.$$

In this case it is easy to see that there exist bounded initial conditions $x(0)$ for which the average displacement is not infinitesimal for $N \to +\infty$. However we can estimate as follows:

$$\delta \le \frac{2k_0}{2-k_0}||x(0)||_\infty^2\,.$$

The case $k_0 = 0$ corresponds to the identity evolution (indeed in this case $R = 1$ and $\delta = 0$). On the other hand, the case $k_0 = 1$ yields convergence to consensus in one step (in this case indeed $R = 0$ but it yields the largest possible δ. Varying k_0 between these two extremes we are trading off speed of convergence against precision in the evaluation of the average.

VI. More general communication graphs: the Cayley case

In this chapter we will move to more challenging examples for the communication graphs. We will consider Abelian Cayley graphs: this class of graphs include already many interesting examples (the circuit, the torus, the hypercube) but retains some fundamental structure which allows to obtain some theoretical, though not yet conclusive results.

A. Abelian Cayley graphs

Let G (with an addition $+$) be any finite Abelian group of order $|G| = N$, and let S be a subset of G containing zero. The Cayley graph $\mathcal{G}(G,S)$ is the directed graph with vertex set G and arc set

$$\mathcal{E} = \{(g,h) : h - g \in S\}\,.$$

Notice that for a Cayley graph both the in-degree and the out-degree of each vertex are equal to $|S|$. Notice also that strongly connectivity can be checked algebraically. Indeed, it can be seen that a Cayley graph $\mathcal{G}(G,S)$ is strongly connected if and only if the set S generates the group G, which means that any element in G can be expressed as a finite sum of (not necessarily distinct) elements in S. If S is such that $-S = S$ then the graph obtained is symmetric. See [26] for more results on these graphs.

Symmetries can be introduced also on matrices. Let G be any finite Abelian group of order $|G| = N$. A matrix $P \in \mathbb{R}^{G\times G}$ is said to be a Cayley matrix over the group G if

$$P_{i,j} = P_{i+h,j+h} \qquad \forall\ i,j,h \in G\,.$$

It is clear that for a Cayley matrix P there exists a $\pi : G \to \mathbb{R}$ such that $P_{i,j} = \pi(i-j)$. The function π is called the generator of the Cayley matrix P. Notice that, if π and π' are generators of the Cayley matrices P and P' respectively, then $\pi + \pi'$ is the generator of $P + P'$ and $\pi * \pi'$ is the generator of PP', where $(\pi*\pi')(i) := \sum_{j\in G}\pi(j)\pi'(i-j)$ for all $i \in G$. This in particular shows that P and P' commute. It is easy to see that for any Cayley matrix P we have that $P\mathbb{1} = \mathbb{1}$ if and only if $\mathbb{1}^*P = \mathbb{1}^*$. This implies that a Cayley stochastic matrix is automatically doubly stochastic.

From now on we assume we have fixed a Cayley graph $\mathcal{G} = (G,E)$ which we assume to be symmetric. Whenever we talk of a Cayley matrix, we assume to be such with respect to the group G.

We have this basic fact whose proof is by inspection.

Proposition VI.1. *For the three examples above we have the following result:*

(1) *$\bar{P}$ is a Cayley matrix.*
(2) $\mathbb{E}[P(t)_{i+h,j+h}P(t)_{n+h,m+h}] = \mathbb{E}[P(t)_{i,j}P(t)_{n,m}]$.

Corollary VI.2. *For the three examples above we have that the sequence of matrices $\Delta(t)$ are all Cayley.*

Proof: By induction on t: it is an immediate consequence of (2) of previous proposition. ∎

B. The symmetric gossip model

We have the following result:

Proposition VI.3. *Assume that $\mathcal{G} = (G,E)$ is a symmetric Cayley graph with degree ν. Assume that Δ is a symmetric Cayley matrix and put $\Delta^+ = \mathbb{E}[P(t)\Delta P(t)^*]$. Then,*

$$\begin{aligned}\pi_{\Delta^+}(\lambda) &= \tfrac{1}{N}[(N-4k_0) + 4k_0^2\nu^{-1}\pi_{A_\mathcal{G}}(\lambda)]\pi_\Delta(\lambda)\\ &+ \tfrac{4k_0}{N}\nu^{-1}\pi_{A_\mathcal{G}\Delta}(\lambda) - \tfrac{4k_0^2}{N}\pi_{\Delta(0)}\nu^{-1}\pi_{A_\mathcal{G}}(\lambda)\\ &+ \tfrac{4k_0^2}{N}[\pi_{\Delta(0)} - <\nu^{-1}\pi_{A_\mathcal{G}},\pi_\Delta>\delta_0(\lambda)]\end{aligned}$$

Let us analyze in detail an example.

Example VI.4: Consider the symmetric circuit graph $\mathcal{G}$ on N elements:

$$\pi_{A_\mathcal{G}}(\lambda) = \begin{cases}1 & \text{if } \lambda = \pm 1\\ 0 & \text{otherwise}\end{cases} \qquad \nu = 2\,. \tag{9}$$

In this case we have that

$$\begin{aligned}&\pi_{A_\mathcal{G}\Delta}(\lambda) = \pi_\Delta(\lambda-1) + \pi_\Delta(\lambda+1)\\ &<\nu^{-1}\pi_{A_\mathcal{G}},\pi_\Delta> = \tfrac{1}{2}[\pi_\Delta(1) + \pi_\Delta(-1)]\,.\end{aligned} \tag{10}$$

From Proposition VI.3 we obtain:

$$\begin{array}{rcl}\pi_{\Delta^+}(0) & = & [1-\frac{4k_0(1-k_0)}{N}]\pi_\Delta(0)\\ & + & \frac{2k_0(1-k_0)}{N}[\pi_\Delta(1)+\pi_\Delta(-1)]\\ \pi_{\Delta^+}(1) & = & [1-\frac{2k_0(2-k_0)}{N}]\pi_\Delta(+1)\\ & + & \frac{2k_0(1-k_0)}{N}\pi_\Delta(0)+\frac{2k_0}{N}\pi_\Delta(2)\\ \pi_{\Delta^+}(-1) & = & [1-\frac{2k_0(2-k_0)}{N}]\pi_\Delta(-1)\\ & + & \frac{2k_0(1-k_0)}{N}\pi_\Delta(0)+\frac{2k_0}{N}\pi_\Delta(-2)\\ \pi_{\Delta^+}(\lambda) & = & [1-\frac{4k_0}{N}]\pi_\Delta(\lambda)\\ & + & \frac{2k_0}{N}[\pi_\Delta(\lambda-1)+\pi_\Delta(\lambda+1)]\\ (\lambda\neq 0,\pm 1) & & \end{array}$$

C. The in-gossip model

We have the following result:

Proposition VI.5. *Assume that $\mathcal{G}=(G,E)$ is a symmetric Cayley graph with degree ν. Assume that Δ is a symmetric Cayley matrix and put $\Delta^+=\mathbb{E}[P(t)^*\Delta P(t)]$. Then,*

$$\pi_{\Delta^+}(\lambda)=(1-k_0)^2\pi_\Delta(\lambda)+2k_0(1-k_0)\nu^{-1}\pi_{A_\mathcal{G}\Delta}(\lambda)$$
$$+k_0^2\nu^{-2}\pi_{A_\mathcal{G}^2\Delta}(\lambda)+k_0^2\pi_\Delta(0)(\delta_0(\lambda)-\nu^{-2}\pi_{A_\mathcal{G}^2}(\lambda))\,.$$

Let us analyze in detail an example.

Example VI.6: Consider the symmetric circuit graph $\mathcal{G}$ on N elements as before. Using (9) and (10) as well

$$\pi_{A_\mathcal{G}^2}(\lambda)=\begin{cases}1 & \text{if } \lambda=\pm 2\\ 2 & \text{if } \lambda=0\\ 0 & \text{otherwise}\end{cases}\,. \tag{11}$$

we obtain

$$\pi_{\Delta^+}(\lambda)=[(1-k_0)^2+\tfrac{k_0^2}{2}]\pi_\Delta(\lambda)+k_0(1-k_0)\cdot$$
$$\cdot[\pi_\Delta(\lambda-1)+\pi_\Delta(\lambda+1)]+\tfrac{k_0^2}{4}[\pi_\Delta(\lambda-2)+\pi_\Delta(\lambda+2)]$$
$$+k_0^2\pi_\Delta(0)(\delta_0(\lambda)-\nu^{-2}\pi_{A_\mathcal{G}^2}(\lambda))\,.$$

From this we obtain,

$$\begin{array}{rcl}\pi_{\Delta^+}(0) & = & [(1-k_0)^2+k_0^2]\pi_\Delta(0)+k_0(1-k_0)\cdot\\ & \cdot & [\pi_\Delta(-1)+\pi_\Delta(1)]+\frac{k_0^2}{4}[\pi_\Delta(-2)+\pi_\Delta(2)]\\ \pi_{\Delta^+}(+2) & = & [(1-k_0)^2+\frac{k_0^2}{2}]\pi_\Delta(2)\\ & + & k_0(1-k_0)[\pi_\Delta(1)+\pi_\Delta(3)]+\frac{k_0^2}{4}\pi_\Delta(4)\\ \pi_{\Delta^+}(-2) & = & [(1-k_0)^2+\frac{k_0^2}{2}]\pi_\Delta(-2)\\ & + & k_0(1-k_0)[\pi_\Delta(-1)+\pi_\Delta(-3)]+\frac{k_0^2}{4}\pi_\Delta(-4)\\ \pi_{\Delta^+}(\lambda) & = & [(1-k_0)^2+\frac{k_0^2}{2}]\pi_\Delta(\lambda)\\ & + & k_0(1-k_0)[\pi_\Delta(\lambda-1)+\pi_\Delta(\lambda+1)]\\ & + & \frac{k_0^2}{4}[\pi_\Delta(\lambda-2)+\pi_\Delta(\lambda+2)]\quad(\lambda\neq 0,\pm 2)\end{array}$$

D. The broadcasting model

We have the following result:

Proposition VI.7. *Assume that $\mathcal{G}=(G,E)$ is a symmetric Cayley graph with degree ν. Assume that Δ is a symmetric Cayley matrix and put $\Delta^+=\mathbb{E}[P(t)^*\Delta P(t)]$. Then,*

$$\pi_{\Delta^+}(\lambda)=(1-\tfrac{2k_0\nu}{N}+\tfrac{k_0^2}{N}\pi_{A_\mathcal{G}^2}(\lambda))\pi_\Delta(\lambda)$$
$$+2(\tfrac{k_0}{N}+\tfrac{k_0^2}{N}\pi_{A_\mathcal{G}}(\lambda))\pi_{A_\mathcal{G}\Delta}(\lambda)+\tfrac{k_0^2}{N}<\pi_{A_\mathcal{G}},\pi_{A_\mathcal{G}\Delta}>\delta_0(\lambda)\,.$$

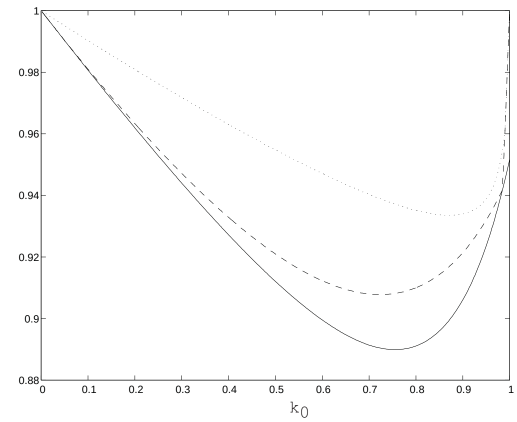

Fig. 1. The graph illustrating of the rate of convergence in the Examples VI.4 (dashed line), VI.6 (dotted line) and VI.8 (solid line) for $N=20$. In order to make a fair comparison in the Examples VI.4 and VI.8 the rates are powered to N.

Let us analyze in detail the symmetric circulant example.

Example VI.8: Consider the symmetric circuit graph $\mathcal{G}$ on N elements as before. From relations (9), (10), (11) and from

$$<\pi_{A_\mathcal{G}},\pi_{A_\mathcal{G}\Delta}>=2\pi_\Delta(0)+\pi_\Delta(+2)+\pi_\Delta(-2)$$

we obtain this time

$$\begin{array}{rcl}\pi_{\Delta^+}(\lambda) & = & (1-\frac{2k_0\nu}{N}+\frac{k_0^2}{N}\pi_{A_\mathcal{G}^2}(\lambda))\pi_\Delta(\lambda)\\ & + & 2(\frac{k_0}{N}-\frac{k_0^2}{N}\pi_{A_\mathcal{G}}(\lambda))[\pi_\Delta(\lambda-1)+\pi_\Delta(\lambda+1)]\\ & + & \frac{k_0^2}{N}[2\pi_\Delta(0)+\pi_\Delta(+2)+\pi_\Delta(-2)]\delta_0(\lambda)\,.\end{array}$$

which leads to

$$\pi_{\Delta^+}(0)=(1-\tfrac{4k_0}{N}+\tfrac{4k_0^2}{N})\pi_\Delta(0)+\tfrac{2k_0}{N}[\pi_\Delta(-1)+\pi_\Delta(+1)]$$
$$+\tfrac{k_0^2}{N}[\pi_\Delta(-2)+\pi_\Delta(+2)]$$
$$\pi_{\Delta^+}(+1)=(1-\tfrac{4k_0}{N})\pi_\Delta(+1)+2(\tfrac{k_0}{N}-\tfrac{k_0^2}{N})[\pi_\Delta(0)+\pi_\Delta(2)]$$
$$\pi_{\Delta^+}(-1)=(1-\tfrac{4k_0}{N})\pi_\Delta(-1)+2(\tfrac{k_0}{N}-\tfrac{k_0^2}{N})[\pi_\Delta(0)+\pi_\Delta(-2)]$$
$$\pi_{\Delta^+}(+2)=(1-\tfrac{4k_0}{N}+\tfrac{k_0^2}{N})\pi_\Delta(2)+2\tfrac{k_0}{N}[\pi_\Delta(1)+\pi_\Delta(3)]$$
$$\pi_{\Delta^+}(-2)=(1-\tfrac{4k_0}{N}+\tfrac{k_0^2}{N})\pi_\Delta(-2)+2\tfrac{k_0}{N}[\pi_\Delta(-1)+\pi_\Delta(-3)]$$
$$\pi_{\Delta^+}(\lambda)=(1-\tfrac{4k_0}{N})\pi_\Delta(\lambda)+2\tfrac{k_0}{N}[\pi_\Delta(\lambda-1)+\pi_\Delta(\lambda+1)]$$
$$(\lambda\neq 0,\pm 1,\pm 2)$$

We computed the rate of convergence in the Examples VI.4, VI.6 and VI.8 for $N=20$ as function of the parameter k_0. Figure 1 shows the the rate of convergence in these three Examples. Notice that, in Examples VI.4 and VI.8 essentially one node communicate at each time while in Example VI.6 all nodes communicate. Therefore the synchronous in-gossip will be always faster than the other two. Therefore, in order to make a fair comparison, in the figure the rates relative to Examples VI.4 and VI.8 are powered to N.

VII. Concentration results

In this chapter we discuss some initial concentration results obtained through the theoretical tool of Azuma's inequality for martingale which we recall below.

Theorem VII.1. Azuma's inequality *Let $X_0, X_1, \dots X_m$ be a martingale such that $|X_{i+1} - X_i| \leq c$ for every i such that $0 \leq i < m$. Then, for every $\theta > 0$ we have that*

$$\mathbb{P}[|X_m - X_0| > \theta] \leq e^{-\frac{\theta^2}{2c^2 m}}\,.$$

Now we assume that our random stochastic matrices can be generated in the following way. We assume the existence of independent random variables $T_1(t), \dots, T_{m_N}(t)$ taking values in a finite set $\mathcal{T}$. These r.v. contain all the randomness to construct $P(t)$ in the sense that there exists a function $\Gamma : \mathcal{T}^{m_N} \to \mathbb{R}^{N\times N}$ yielding

$$P(t) = \Gamma(T_1(t), \dots, T_{m_N}(t))\,.$$

It is easy to see that the matrices $P(t)$ in our examples can indeed be generated in this way. Specifically, in the symmetric gossip case we have that $m_N = 1$ and $\mathcal{T} = |E|$: $T_1(t)$ simply codifies the edge which is activated. In the in-gossip model instead we have that $m_N = N$ and $\mathcal{T} = \{1, \dots, N\}$: $T_j(t)$ simply codifies the neighbor which is chosen by agent j. Finally, in the broadcasting model we have $m_N = 1$ and $\mathcal{T} = \{1, \dots, N\}$: $T_1(t)$ simply codifies the node which is activated

We can now state and proof the following general result.

Proposition VII.2. *Assume that the random stochastic matrices $P(t)$ can be generated through the discrete r.v. $T_1(t), \dots, T_{m_N}(t)$ as above and that they possess the following further properties:*

1. *There exist $a \in \mathbb{N}$, independent of N, such that when in $\Gamma(t_1, \dots, t_{m_N})$ we vary just one of its variables t_j keeping fixed all the others, only at most a rows may vary.*
2. *$\bar{P}$ is doubly stochastic.*

Then,

$$\mathbb{P}[|d(t,N) - \mathbb{E}d(t,N)| \geq \delta] \leq e^{-\frac{\delta^2 N^2}{36||x(0)||_\infty^4 a^2 (t+1) m_N}}$$

Proof: Assume t is fixed and consider the following doubly indexed martingale:

$$X_{s,k} = \mathbb{E}[d(t,N)\,|\,P(0), \dots, P(s-1), T_1(s), \dots, T_k(s)]$$

(where $s = 0, \dots, t$ and $k = 1, \dots N$).

It follows from conditions (1) and (2) that

$$\begin{aligned}
&|X_{s,k} - X_{s,k-1}| = \tfrac{1}{N}|\mathbb{E}_{P(s+1),\dots P(t)}\{\mathbb{E}_{T_{k+1}(s),\dots T_{m_N}(s)}\\
&\{\mathbb{E}_{T_k(s)}\{||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)\\
&\Gamma(T_1(s), \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)||^2\\
&-||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)\\
&\Gamma(T_1(s), \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)||^2\}\}\}|\\
&\leq \tfrac{1}{N}\mathbb{E}_{P(s+1),\dots P(t-1)}\{\mathbb{E}_{T_{k+1}(s),\dots T_{m_N}(s)}\{\max_{\sigma,\sigma'}\\
&|\,||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)\\
&\Gamma(T_1(s), \dots, \sigma, \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)||^2-\\
&-||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)\\
&\Gamma(T_1(s), \dots, \sigma', \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)||^2\}|\}\}
\end{aligned} \tag{12}$$

If we consider the two vectors

$$\begin{aligned}
y &= \Gamma(T_1(s), \dots, \sigma, \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)\\
y' &= \Gamma(T_1(s), \dots, \sigma', \dots T_{m_N}(s))P(s-1)\cdots P(0)x(0)\,,
\end{aligned}$$

we have that

$$||y||_\infty \leq ||x(0)||_\infty\,, \quad ||y'||_\infty \leq ||x(0)||_\infty \tag{13}$$

and there exist indices $j_1, \dots, j_a$ such that

$$y' - y = \sum_{l=1}^{a} r_l e_{j_l}\,, |r_l| \leq 2||x(0)||_\infty \tag{14}$$

From estimation (12), using (13) and (14), we obtain

$$\begin{aligned}
&|X_{s,k} - X_{s,k-1}| \leq \tfrac{1}{N}\mathbb{E}_{P(s+1),\dots P(t-1)}\{\sup_{y:||y||_\infty \leq ||x(0)||_\infty}\\
&\{\max_{j_1,\dots j_a}\{\sup_{r_l : r_l \leq 2||x(0)||_\infty}\\
&|\,||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)y||^2-\\
&||(I - N^{-1}\mathbb{1}\mathbb{1}^*)P(t-1)\cdots P(s+1)(y + \textstyle\sum_{l=1}^{a} r_l e_{j_l})||^2\,|\}\}\}\\
&\leq \tfrac{1}{N}\mathbb{E}\{\sup_{y:||y||_\infty \leq ||x(0)||_\infty}\{\max_{j_1,\dots j_a}\{\sup_{r_l : r_l \leq 2||x(0)||_\infty}\\
&|\, < (I - N^{-1}\mathbb{1}\mathbb{1}^*)Q(t-s)(\sum_{l=1}^{a} r_l e_{j_l} - 2y),\\
&, (I - N^{-1}\mathbb{1}\mathbb{1}^*)Q(t-s)(\sum_{l=1}^{a} r_l e_{j_l}) > \,|\}\}\}\\
&\leq \tfrac{6||x(0)||_\infty^2}{N}\mathbb{E}\{< Q(t-s)\mathbb{1}, Q(t-s)(\sum_{l=1}^{a} e_{j_l}) >\}\\
&\leq \tfrac{6||x(0)||_\infty^2}{N} < \mathbb{1}, \bar{P}^{t-s}(\sum_{l=1}^{a} e_{j_l}) >= \tfrac{6a||x(0)||_\infty^2}{N}
\end{aligned} \tag{15}$$

Result now follows from Azuma's inequality. ■

Let us comment on the application to our examples. In the symmetric gossip case, we have that $a = 2$ and $m_N = 1$ so that we have concentration around the mean average for fixed t and $N \to +\infty$ or, more generally, as long as $N^2/t \to +\infty$. In the in-gossip case instead we have that $a = 1$ and $m_N = N$, so that we obtain a similar concentration result but under the condition $N/t \to +\infty$. In the broadcast model instead $a = \nu \leq N$ (in the Cayley case) and $m_N = 1$ so that we obtain a similar result.

Similar concentration results can be obtained also for the average displacement $\beta(t)$.

VIII. Conclusions

In this paper we have analyzed three different randomized consensus algorithms. One of them needs symmetric communication, while the other two can also be utilized in asymmetric communication scenario. On the other hand, while the first one preserves the global average, the other two do not. We have analyzed the mean square convergence of the three schemes for Abelian Cayley communication graphs and we have proved a concentration result. We have concretely analyzed two cases: the complete graph and the circulant graph. For the complete case we could carry on a detailed theoretical analysis of the speed of convergence and also of the average displacement. For the circulant case, we only have partial results and some numerical simulations. It was not our goal to prove that one scheme was 'better' than the others, rather we wanted to put into evidence that there are many possible schemes available for average consensus or at least for consensus quite close to average: they differ for complexity in their implementation, speed of convergence, average displacement. What is preferable will mostly depend on the specific application. The broadcasting example for instance seems to show a better speed of convergence, while it has probably the worse average displacement. Our analysis is however only at the beginning and much more needs to be done in this area. For instance we conjecture that the in-gossip models always have (at least for Cayley graphs) the average displacement which is infinitesimal with respect to the number of nodes. Also analysis of other graphs would be of interest. In particular, we would like to extend our analysis to the geometric graph: this would be of interest in the area of sensor networks.

References

[1] Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Department of EECs, MIT, 1984.

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

[3] B. G. S. Muthukrishnan and M. Schultz, "First and second order diffusive methods for rapid, coarse, distributed load balancing," *Theory of computing systems*, vol. 31, pp. 331–354, 1998.

[4] A. F. R. Diekmann and B. Monien, "Efficient schemes for nearest neighbor load balancing," *Parallel computing*, vol. 25, pp. 789–812, 1999.

[5] A. D. D. Kempe and J. Gehrke, "Gossip-based computation of aggregate information," in *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, 2003.

[6] D. E. C. Intanagonwiwat, R. Govindan, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in *Proceedings ACM/IEEE Conf. Mobile Computing and Networking*, 2000.

[7] D. E. J. Zhao, R. Govindan, "Computing aggregates for monitoring wireless sensor networks," in *Proc. Int. Workshop on Sensor Net Protocols and Applications*, 2003, pp. 139–148.

[8] S. L. 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.

[9] A. S. A. Dymakis and M. Wainright, "Geographic gossip: efficient aggregation on sensor networks," EECS Berkeley, Tech. Rep., 2006.

[10] D. R. Q. Li, "Global clock syncronization in sensor networks," Tech. Rep., 2006.

[11] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

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

[13] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

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

[15] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transaction on Automatic Control*, vol. 50, no. 5, 2005.

[16] S. L. Smith, M. E. Broucke, and B. A. Francis, "A hierarchical cyclic pursuit scheme for vehicle networks," *Automatica*, vol. 41, no. 6, pp. 1045–1053, 2005.

[17] R. Olfati-Saber, "Distributed kalman filter with embedded consensus filters," in *IEEE Conference on Decision and Control*, 2005, pp. 63–70.

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

[19] A. O. J. T. V.D. Blondel, J.M. Hendrickx, "Convergence in multiagent coordination, consensus and flocking," in *Proceedings of joint IEEE-CDC/ECC*, 2005.

[20] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, "Communication constraints in the average consensus problem," 2005, submitted.

[21] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[22] B. V. R. C.C. Moallemi, "Consensus propagation," *IEEE Transactions on nformation Theory*, vol. 52, no. 11, pp. 4753–4766, 2006.

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

[24] F. B. J. Cortes, S. Martinez, "Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions," *IEEE Transactions on Automatic Control*, vol. 51, pp. 1289–1298, 2006.

[25] R. Cogburn, "On products of random stochastic matrices," in *Random matrices and their applications (Brunswick, Maine, 1984)*, ser. Contemp. Math. Amer. Math. Soc., 1986, vol. 50, pp. 199–213.

[26]