SIAM J. CONTROL OPTIM.
Vol. 48, No. 1, pp. 102–133

AVERAGE CONSENSUS WITH PACKET DROP COMMUNICATION*

FABIO FAGNANI† AND SANDRO ZAMPIERI‡

Abstract. Average consensus consists in the problem of determining the average of some quantities by means of a distributed algorithm. It is a simple instance of problems arising when designing estimation algorithms operating on data produced by sensor networks. Simple solutions based on linear estimation algorithms have already been proposed in the literature and their performance has been analyzed in detail. If the communication links which allow the data exchange between the sensors have some loss, then the estimation performance will degrade. In this contribution the performance degradation due to this data loss is evaluated.

Key words. average consensus, Cayley graphs, mean square analysis, packet drop, rate of convergence

AMS subject classifications. 05C80, 68W15, 68W20, 93A15

DOI. 10.1137/060676866

1. Introduction.

Average consensus problems have been widely studied in recent years [20, 12, 18, 2, 15, 9, 16], both in the context of coordination of mobile autonomous vehicles and in the context of distributed estimation. In fact, average consensus can be considered a simple paradigm for designing estimation algorithms implemented on sensor networks and working in a distributed way. More precisely, we assume in this setup that all sensors independently measure the same quantity with some error due to noise. A simple way to improve the estimate is to average all the measures. To do this, the sensors need to exchange their information. Energy limitations force transmission to take place directly along nearby sensors and also impose bounds on the amount of data an agent can process. A global description of the allowed exchange of information can be given by a directed graph in which the sensors are the nodes and in which an edge from agent i to agent j represents the possibility for i to send information to j. Algorithms which allow us to obtain this average are called average consensus algorithms. The performance of an average consensus algorithm may be measured by the speed of convergence toward the average. In [15] a simple algorithm is proposed which is based on a linear dynamical system. Moreover, in [15, 3] the relation between the performance of this algorithm and the degree of connectivity of the graph is also evaluated. In [20, 11, 13, 18] variations of this algorithm which handle time-varying communication graphs are considered. Since in these cases the analysis proposed is essentially a worst case analysis, the performance evaluation can be rather conservative. Different results can be obtained if the graphs vary in time randomly [2, 3]. In fact, randomly time-varying graphs typically yield improved performance.

In this paper we consider a more realistic model of the data exchange. In fact, in many practical applications, the data exchange between sensors takes place over a wireless communication network leading to the possibility that some packets get

*Received by the editors December 6, 2006; accepted for publication (in revised form) July 30, 2008; published electronically February 11, 2009.
http://www.siam.org/journals/sicon/48-1/67686.html

†Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy (fabio.fagnani@polito.it).
‡Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy (zampi@dei.unipd.it).

lost during the transmission. In this contribution this phenomenon is modeled by assuming that at every time instant the transmission of a number from one sensor to another can occur with a certain probability, and so there is a certain probability that the link will fail and the data will be lost. We can expect that this will produce a performance degradation. The main objective of this contribution is to provide some instruments that allow us to quantify this degradation as a function of the probability of the link failure. The problem is similar to the one considered in [9] where a more limited class of random graphs were considered and where only the convergence of the algorithm was considered. We have recently realized that other researchers [17] have independently studied similar problems; their results, however, are different from ours.

In section 2, after recalling classical average consensus algorithms, we propose two different adaptations of such algorithms which can cope with lossy links: the biased and the balanced compensation methods. The essential difference between the two methods is that in the biased version, local averaging weights at each node are kept fixed while, in the balanced case, weights are scaled depending on the available data at every instant. Both algorithms will be shown to converge (almost surely and in mean square sense) to a consensus value which in general may not coincide with the average of the initial states. For both cases, performance degradation will be analyzed through two figures showing the rate of convergence and the asymptotic displacement from the average consensus. Analysis will always be carried out in a mean square sense. Analysis of the degradation of the convergence rate is undertaken in section 3, where the problem is reduced to finding the largest eigenvalue of a suitable linear operator $\mathcal{L}$ acting on a space of N^2 dimensions (where N is the number of agents). This reduction, besides giving an important theoretical characterization, is amenable to efficient numerical analysis simulations. Sections 4 and 5 are devoted to the case when the network possesses symmetries, in particular when it can be modeled by an Abelian Cayley graph. In this case, the operator $\mathcal{L}$ can actually be substituted with an N-dimensional operator. This allows us to obtain deeper analytical results and, in particular, to obtain explicit solutions in special important cases (e.g., complete graph, cycle graph, and hypercube graph). A comparison of the two methods shows that, at least in some examples, the balanced method presents a better rate of convergence. Finally, in section 6, we analyze the asymptotic displacement from the average consensus due to packet drop and we prove that for the Abelian Cayley case, this displacement is infinitesimal in the number of agents for both methods. We will also show that with respect to the asymptotic displacement the biased method outperforms the other.

2. Problem formulation.

We assume that we have N agents. Each agent i measures a quantity $d_i \in \mathbb{R}$ and at each time instant t it can transmit a real number to some agents. The data exchange can be described by a directed graph $\mathcal{G}$ with vertices $\{1, \ldots, N\}$, in which there is an edge (j, i) if and only if the agent j can send data to agent i. The objective is to find a distributed algorithm which allows the agents to obtain a shared estimate of the average of the d_i's. An efficient algorithm solving this problem consists in the dynamic system

$$x_i(t+1) = \sum_{j=1}^{N} P_{ij} x_j(t), \qquad x_i(0) = d_i,$$

where P is a suitable matrix such that $P_{ij} = 0$ if (j, i) is not an edge in $\mathcal{G}$. We assume that $\mathcal{G}$ always includes all the self loop edges (i, i), meaning that each agent i has access to its own data. More compactly we can write

$$x^+ = Px, \qquad x(0) = d, \tag{1}$$

where $x, d \in \mathbb{R}^N$ and where x^+ is a shorthand notation for $x(t+1)$. According to this algorithm, the agent i needs to receive the value of $x_j(t)$ from the agent j to update the value of $x_i(t)$ only if $P_{ij} \neq 0$. In this case, we say that the agents reach the *consensus*, if for any initial condition $x(0) \in \mathbb{R}^N$, the closed loop system (1) yields

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

where $\mathbf{1} := (1, \ldots, 1)^*$ and where α is a scalar depending on $x(0)$ and P. Moreover, if α coincides with the average $N^{-1}\sum_{i=1}^N d_i = N^{-1}\mathbf{1}^* x(0)$, then we say that the agents reach the *average consensus*.

To make the concepts more precise it is useful to recall some notation and results on directed graphs (the reader can further refer to textbooks on graph theory such as [8] or [5]). Fix a directed graph $\mathcal{G}$ with a set of vertices V and a set of arcs $\mathcal{E} \subseteq V \times V$. The adjacency matrix E is a $\{0, 1\}$-valued square matrix indexed by the elements in V defined by letting $E_{ij} = 1$ if and only $(i, j) \in \mathcal{E}$. Define the out-degree of a vertex j as $\text{outdeg}(j) := \sum_i E_{ij}$ and the in-degree of a vertex i as $\text{indeg}(i) := \sum_j E_{ij}$. A graph is called in-regular (resp., out-regular) of degree k if each vertex has in-degree (resp., out-degree) equal to k. A path in $\mathcal{G}$ consists of a sequence of vertices $i_1 i_2 \ldots i_r$ such that $(i_\ell, i_{\ell+1}) \in \mathcal{E}$ for every $\ell = 1, \ldots, r-1$; i_1 (resp., i_r) is said to be the initial (resp., terminal) vertex of the path. A path is said to be closed if the initial and the terminal vertices coincide. A vertex i is said to be connected to a vertex j if there exists a path with initial vertex i and terminal vertex j. A directed graph is said to be connected if, given any pair of vertices i and j, either i is connected to j or j is connected to i. A directed graph is said to be strongly connected if, given any pair of vertices i and j, i is connected to j.

With an $N \times N$ matrix P we associate a directed graph $\mathcal{G}_P$ with a set of vertices $\{1, \ldots, N\}$ in which there is an arc from j to i whenever the element $P_{ij} \neq 0$. The graph $\mathcal{G}_P$ is said to be the *communication graph* associated with P. Conversely, given any directed graph $\mathcal{G}$ with the set of vertices $\{1, \ldots, N\}$, we say that a matrix P is *compatible* with $\mathcal{G}$ if $\mathcal{G}_P$ is a subgraph of $\mathcal{G}$. After introducing this notation we can make the consensus problem more precise. We say that the (average) consensus problem is solvable on a graph $\mathcal{G}$ if there exists a matrix P compatible with $\mathcal{G}$ and solving the (average) consensus problem.

As shown in [18, 15, 3], if $\mathcal{G}$ is strongly connected, it is always possible to choose P so as to obtain the consensus. Indeed, if P is a stochastic matrix (namely, $P_{ij} \geq 0$ for every i, j and $P\mathbf{1} = \mathbf{1}$), $\mathcal{G}_P$ is strongly connected, and $P_{ii} > 0$ for some i, then P solves the consensus problem. To obtain average consensus P needs to satisfy an extra condition: It must be doubly stochastic ($\mathbf{1}^* P = \mathbf{1}^*$). If $\mathcal{G}$ is strongly connected, a P also satisfying this last condition can be found even if the construction becomes, in general, more involved. There is an important case when the construction of such a P is quite simple—when all agents have the same out- and in-degree ν (without considering self loops). In this case, we can simply choose $P = kI + (1-k)\nu^{-1}E$ for any $k \in]0, 1[$. Undirected graphs are clearly an example which fits into this case and P in this case is actually symmetric.

In the following we will give an elementary example which casts the average consensus problem into the topic of distributed estimation.

Example 1 (estimation from distributed measures [14, 2]). Assume we have N sensors which measure a quantity $z \in \mathbb{R}$. However, due to noise, each sensor obtains different measures $y_i = z + v_i$, where v_i are independent random variables with zero mean and the same variance. It is well known that the average

$$\alpha = N^{-1} \sum_{i=1}^{N} y_i$$

provides the best possible linear estimate of z (in the sense of the minimum mean square error) from y_i. Running an average consensus problem with initial conditions $x_i(0) = y_i$ will lead to a distributed computation of α by every agent.

2.1. Packet drop consensus algorithms. We start from a fixed graph $\mathcal{G}$ and we assume that on each edge (j, i) of $\mathcal{G}$, communication from the node j to the node i can occur with some probability p. In order to describe this model more precisely, we introduce the family of independent binary random variables $L_{ij}(t)$, $t \in \mathbb{N}$, $i, j = 1, \ldots, N$, $i \neq j$, such that

$$\mathbb{P}[L_{ij}(t) = 1] = p, \qquad \mathbb{P}[L_{ij}(t) = 0] = 1 - p.$$

We emphasize the fact that independence is assumed among all $L_{ij}(t)$ as i, j and t vary. Let E be the adjacency matrix of $\mathcal{G}$, and let $H := E - I$. Consider the random matrix $\bar{E}(t) = I + \bar{H}(t)$, where $\bar{H}_{ij}(t) = H_{ij} L_{ij}(t)$. Clearly, $\bar{E}(t)$ is the adjacency matrix of a random graph $\bar{\mathcal{G}}(t)$ obtained from $\mathcal{G}$ by deleting the edge (i, j) when $L_{ij}(t) = 0$.

In this paper we will propose consensus strategies compatible with the random varying communication graphs $\bar{\mathcal{G}}(t)$; they will consist of a sequence of random stochastic matrices $P(t)$ such that $\mathcal{G}_{P(t)} \subseteq \bar{\mathcal{G}}(t)$ for all t.

Our construction always starts from the choice of a stochastic matrix P adapted to $\mathcal{G}$ yielding average consensus and we modify it in a way to compensate for the lack of some data. There is, in principle, more than one way to obtain this. We will propose two solutions. In the first, which will be called the *biased compensation method*, each agent, in updating the estimate of the average, adds the weights of the unavailable data to the weight it assigns to its own old estimate. In the second, which will be called the *balanced compensation method*, the compensation for the lack of data is done by modifying all the weights in a more balanced way. We want to emphasize the fact that we are assuming all agents to be time synchronized. As a consequence, at every time instant t, any agent i knows which data he has received; this means that agent i knows the value of $L_{ij}(t)$ for every neighbor j.

The biased compensation method. We consider the following updating law:

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

According to this strategy, the agent i, in computing the new estimate, compensates for the loss of data by accumulating the weights of the lost data with the weight assigned to its previous estimate. Intuitively, according to this method, the agent i

substitutes the unavailable $x_j(t)$ with $x_i(t)$ in the consensus algorithm. If we define the random matrices $D(t), Q(t)$ as

$$D_{ij}(t) := \begin{cases} P_{ii} + \sum_{j \neq i}(1 - L_{ij}(t))P_{ij} = 1 - \sum_{h \neq i} L_{ih}(t)P_{ih} & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

and

$$Q_{ij}(t) := \begin{cases} 0 & \text{if } i = j, \\ L_{ij}(t)P_{ij} & \text{if } i \neq j, \end{cases}$$

then we can describe this method through the stochastic system

$$x(t+1) = P(t)x(t), \tag{3}$$

where

$$P(t) := D(t) + Q(t).$$

The balanced compensation method. As opposed to the previous method, here we prefer to distribute the weights equally between the available data. The updating equation is thus

$$x_i(t+1) = \frac{1}{P_{ii} + \sum_{j \neq i} L_{ij}(t)P_{ij}} \left(P_{ii}x_i(t) + \sum_{j \neq i} L_{ij}(t)P_{ij}x_j(t) \right).$$

In this case it is convenient to define, for $i = 1, \ldots, N$, the binary random variable L_{ii} which is equal to 1 with probability 1. In this way, by defining

$$\nu_i(t) = \sum_{j=1}^{N} L_{ij}(t)P_{ij}$$

and introducing the diagonal matrix $D(t)$ having

$$D_{ii}(t) := \frac{1}{\nu_i(t)}$$

and the matrix $Q(t)$ such that

$$Q_{ij}(t) := L_{ij}(t)P_{ij},$$

we can more compactly write

$$x(t+1) = P(t)x(t)$$

with

$$P(t) := D(t)Q(t).$$

Our goal will be to evaluate the asymptotic behavior of system (3) in the two cases. The following result shows that, for both of the above methods, communication failures will never prevent us from reaching consensus. The proof is a simple consequence of Theorem 6 in [4] and is a particular instance of Corollary 3.2 in [6].

THEOREM 2.1. *Assume that P achieves the consensus and that $P_{ii} > 0$ for every i. Then, both the biased compensation method and the balanced compensation method yield consensus almost surely; namely,* (2) *almost surely holds for any initial condition $x(0)$ with an α which is general a random variable. Moreover, the convergence in* (2) *also holds in the mean square sense; namely,*

$$\lim_{t\to\infty} \mathbb{E}\left[||x(t) - \mathbf{1}\alpha||^2\right] = 0, \tag{4}$$

where $\mathbb{E}[\cdot]$ is the expected value and $||\cdot||$ is the 2-norm in $\mathbb{R}^N$.

Notice that the random variable α depends linearly on the initial condition $x(0)$. In other terms there exists an N-dimensional random vector v such that $\alpha = v^* x(0)$.

In spite of the previous theorem, we do expect a performance degradation due to communication failures. Degradation will show up in two ways; first, as a diminished convergence speed of the limit (4) and, second, as the deviation of the random variable α from the average. The aim of this paper is to quantify such a degradation. The next section will focus on the speed of convergence.

3. Mean square analysis. In this section we assume we have fixed a matrix P adapted to the graph $\mathcal{G}$ yielding consensus and such that $P_{ii} > 0$ for every node i. We then undertake a mean squared analysis of our stochastic models and we characterize their asymptotic rate of convergence. Precisely, our aim is to evaluate the exponential rate of convergence to 0 of $\mathbb{E}[||x(t) - \mathbf{1}\alpha||^2]$. We start with a preliminary result. Let

$$x_A(t) := \frac{1}{N}\sum_{i=1}^{N} x_i(t) = \frac{1}{N}\mathbf{1}^* x(t), \tag{5}$$

which coincides with the average of the current states.

PROPOSITION 3.1. *Assume that we have almost sure consensus, namely, that $x(t) \to \alpha\mathbf{1}$ almost surely. Then,*

$$\mathbb{E}[||x(t) - \mathbf{1}x_A(t)||^2] \le \mathbb{E}[||x(t) - \mathbf{1}\alpha||^2] \le (1+\sqrt{N})^2 \mathbb{E}[||x(t) - \mathbf{1}x_A(t)||^2]. \tag{6}$$

Proof. From

$$x(t) - \mathbf{1}x_A(t) = (I - N^{-1}\mathbf{1}\mathbf{1}^*)x(t) = (I - N^{-1}\mathbf{1}\mathbf{1}^*)(x(t) - \mathbf{1}\alpha),$$

we obtain

$$||x(t) - \mathbf{1}x_A(t)|| \le ||x(t) - \mathbf{1}\alpha||\,.$$

This proves the left inequality.

The following identity holds for every t and s:

$$x(t) - x(t+s) = (I - P(t+s-1)\cdots P(t))\,(I - N^{-1}\mathbf{1}\mathbf{1}^*)x(t)\,.$$

Using the fact that for any stochastic matrix P, $||P|| \le \sqrt{N}$, we obtain that

$$||x(t) - x(t+s)|| \le (1+\sqrt{N})||x(t) - \mathbf{1}x_A(t)||\,. \tag{7}$$

Letting $s \to \infty$ and taking the average of this yields the right inequality. $\square$

The proposition shows that $\mathbb{E}[||x(t)-\mathbf{1}\alpha||^2]$ and $\mathbb{E}[||x(t)-\mathbf{1}x_A(t)||^2]$ have the same exponential rate of convergence to zero or, in other words, that, for any initial condition $x(0)$, we have that

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

For this reason, in what follows we will study the right-hand expression, which turns out to be simpler to analyze. In order to have a single figure not dependent on the initial condition, we will concentrate on this worst case exponential rate of convergence:

$$R:=\sup_{x(0)}\limsup_{t\to+\infty}\mathbb{E}[||x(t)-\mathbf{1}x_A(t)||^2]^{1/t}\,.$$

Remark 1. Some of the considerations carried out above hold true even without a priori knowledge of almost sure consensus. This is true for (7) in the proof of Proposition 3.1 which, in any case, yields

$$(\mathbb{E}||x(t)-x(t+s)||^2)^{1/2}\le(1+\sqrt{N})(\mathbb{E}||x(t)-\mathbf{1}x_A(t)||^2)^{1/2}\,. \tag{8}$$

Hence, from the simple knowledge that $\mathbb{E}||x(t)-\mathbf{1}x_A(t)||^2$ converges to 0, we can deduce that $x(t)$ is a Cauchy sequence and so, for completeness arguments, $x(t)$ converges in mean square to some random vector $x(\infty)$. Notice, moreover, that for any vector $\zeta\in\mathbb{R}^N$ orthogonal to $\mathbf{1}$ we have that

$$|\zeta^*x(t)|=|\zeta^*(x(t)-\mathbf{1}x_A(t))|\le||\zeta||\,||x(t)-\mathbf{1}x_A(t)||\longrightarrow 0.$$

Since $\zeta^*x(t)\longrightarrow\zeta^*x(\infty)$, for the limit uniqueness we have that $\zeta^*x(\infty)=0$. This implies that $x(\infty)=\mathbf{1}\alpha$ for some random variable α. Hence, convergence of $\mathbb{E}||x(t)-\mathbf{1}x_A(t)||^2$ yields consensus in the mean square sense.

In order to study the behavior of $\mathbb{E}[||x(t)-\mathbf{1}x_A(t)||^2]$, the following characterization turns out to be very useful. Indeed, notice that

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

where

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

if $t\ge 1$ and where $\Delta(0):=I-N^{-1}\mathbf{1}\mathbf{1}^*$. Therefore we have that

$$R=\max_{ij}\limsup_{t\to+\infty}\Delta(t)_{ij}^{1/t}\,.$$

We now study the evolution of the matrices $\Delta(t)$.

Notice first that

$$\begin{aligned}
&\Delta(t+1)\\
&=\mathbb{E}[P(0)^*P(1)^*\cdots P(t-1)^*P(t)^*(I-N^{-1}\mathbf{1}\mathbf{1}^*)P(t)P(t-1)\cdots P(1)P(0)]\\
&=\mathbb{E}[\mathbb{E}[P(0)^*P(1)^*\cdots P(t-1)^*P(t)^*(I-N^{-1}\mathbf{1}\mathbf{1}^*)P(t)P(t-1)\cdots P(1)P(0)|P(0)]]\\
&=\mathbb{E}[P(0)^*\mathbb{E}[P(1)^*\cdots P(t-1)^*P(t)^*(I-N^{-1}\mathbf{1}\mathbf{1}^*)P(t)P(t-1)\cdots P(1)]P(0)]\\
&=\mathbb{E}[P(0)^*\Delta(t)P(0)],
\end{aligned}$$

where the last equality follows from the fact that, since the random matrices $P(t)$ are independent and identically distributed, the two sequences of random matrices $(P(0), \dots, P(t-1))$ and $(P(1), \dots, P(t))$ have the same probability distribution.

It is convenient to introduce the linear operator $\mathcal{L} : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ defined by

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

In this way $\Delta(t)$ is governed by the recursive relation

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

If we consider now 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

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

where $\mathcal{L}_{|\mathcal{R}}$ denotes the restriction of the operator $\mathcal{L}$ to the invariant subspace $\mathcal{R}$.

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 operator $\mathcal{L}$ has many interesting properties which have been studied in [6] in a more general context. It has been shown in particular that $\mathcal{L}$ can be interpreted as an aperiodic row-stochastic operator. As a consequence, 1 is an eigenvalue of algebraic multiplicity one. It is easy to find a corresponding eigenvector. Notice indeed that $x(0)^*\mathcal{L}^t(\Delta)x(0) = \mathbb{E}[x(t)^*\Delta x(t)]$. Since $x(t) \to \mathbf{1}v^*x(0)$ in mean square sense, it follows that

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

As a consequence,

$$\lim_{t\to+\infty} \mathcal{L}^t(\Delta) = (\mathbf{1}^*\Delta\mathbf{1})\mathbb{E}[vv^*]\,.$$

In particular, $\mathcal{L}(\mathbb{E}[vv^*]) = \mathbb{E}[vv^*]$. Clearly the reachability subspace $\mathcal{R}$ will be contained in the subspace generated by the eigenvectors different from $\mathbb{E}[vv^*]$.

In what follows we will write the operator $\mathcal{L}$ in a more explicit form. This will allow us to determine R numerically. To do this we now need to study the two cases separately.

3.1. The biased compensation method. For any matrix M we will denote diag (M) as the diagonal matrix with the same diagonal elements of M and out $(M) :=$ $M -$ diag (M) which is out-diagonal, namely, has zero diagonal elements.

PROPOSITION 3.2. *The sequence of matrices $\Delta(t)$ satisfies the recursive relation*

$$\begin{aligned} \Delta^+ &= [(1-p)I + pP]^*\Delta[(1-p)I+pP] \\ &\quad + p(1-p)\text{diag } \{\text{out } (P)\text{out } (P)^*\text{diag } (\Delta) + \text{out } (P)^*\text{diag } (\Delta)\text{out } (P)\} \\ &\quad - p(1-p)\big\{\text{diag } (\Delta)\text{out } (\tilde{P}) + \text{out } (\tilde{P})^*\text{diag } (\Delta)\big\}, \end{aligned} \tag{9}$$

where the matrix $\tilde{P}$ is defined by letting $\tilde{P}_{ij} := P_{ij}^2$.

Proof. Let $D := D(0)$ and $Q := Q(0)$. Notice, preliminarily, that $\mathbb{E}[Q] = p\,\text{out}\,(P)$ and that $\mathbb{E}[D] = (1-p)I + p\text{diag}\,(P)$. Notice, moreover, that

$$\mathbb{E}[D_{ii}D_{jj}] = \begin{cases} (1-p+pP_{ii})(1-p+pP_{jj}) & \text{if } i \neq j, \\ (1-p+pP_{ii})^2 + p(1-p)\sum_{k\neq i} P_{ik}^2 & \text{if } i = j, \end{cases} \tag{10}$$

and that

$$\mathbb{E}[D_{ii}Q_{ij}] = (1-p+pP_{ii})pP_{ij} - p(1-p)P_{ij}^2. \tag{11}$$

Notice now that

$$\Delta^+ = \mathbb{E}[D\Delta D] + \mathbb{E}[D\Delta Q] + \mathbb{E}[Q^*\Delta D] + \mathbb{E}[Q^*\Delta Q].$$

Using (10) we obtain that

$$\begin{aligned}&\mathbb{E}[D\Delta D]_{ij}\\ &= \mathbb{E}\left[D_{ii}D_{jj}\right]\Delta_{ij} = \begin{cases} (1-p+pP_{ii})(1-p+pP_{jj})\Delta_{ij} & \text{if } i \neq j,\\ (1-p+pP_{ii})^2\Delta_{ii} + p(1-p)\left(\sum_{k\neq i} P_{ik}^2\right)\Delta_{ii} & \text{if } i = j.\end{cases}\end{aligned}$$

More compactly, we can write

$$\begin{aligned}\mathbb{E}[D\Delta D] = &\ [(1-p)I + p\text{diag }(P)]\Delta[(1-p)I + p\text{diag }(P)]\\ &+ p(1-p)\text{diag}\{\text{out}(P)\text{out }(P)^*\}\text{diag }(\Delta).\end{aligned}$$

Notice now that

$$\mathbb{E}[D\Delta Q]_{ii} = \sum_{k\neq i}\mathbb{E}\left[D_{ii}\Delta_{ik}Q_{ki}\right] = \mathbb{E}\left[D_{ii}\right]\sum_{k\neq i}\Delta_{ik}\mathbb{E}\left[Q_{ki}\right] = p(1-p+pP_{ii})\sum_{k\neq i}\Delta_{ik}P_{ki}.$$

If instead $i \neq j$, then, using (11), we obtain

$$\begin{aligned}\mathbb{E}[D\Delta Q]_{ij} &= \sum_{k\neq j}\mathbb{E}\left[D_{ii}\Delta_{ik}Q_{kj}\right] = \mathbb{E}\left[D_{ii}\right]\sum_{\substack{k\neq i\\ k\neq j}}\Delta_{ik}\mathbb{E}\left[Q_{kj}\right] + \mathbb{E}\left[D_{ii}\Delta_{ii}Q_{ij}\right]\\ &= p(1-p+pP_{ii})\sum_{\substack{k\neq i\\ k\neq j}}\Delta_{ik}P_{kj} + p(1-p+pP_{ii})P_{ij}\Delta_{ii} - p(1-p)P_{ij}^2\Delta_{ii}.\end{aligned}$$

More compactly, we can write

$$\mathbb{E}[D\Delta Q] = p[(1-p)I + p\text{diag }(P)]\Delta\text{out }(P) - p(1-p)\text{diag }(\Delta)\text{out }(\tilde{P}).$$

Finally, observe that

$$\begin{aligned}\mathbb{E}[Q^*\Delta Q]_{ii} &= \sum_{\substack{h\neq i\\ k\neq i}}\mathbb{E}\left[Q_{hi}\Delta_{hk}Q_{ki}\right] = p^2\sum_{\substack{h\neq i\ \ k\neq i\\ h\neq k}}P_{hi}\Delta_{hk}P_{ki} + p\sum_{h\neq i}P_{hi}^2\Delta_{hh}\\ &= p^2\sum_{\substack{h\neq i\\ k\neq i}}P_{hi}\Delta_{hk}P_{ki} + p(1-p)\sum_{h\neq i}P_{hi}^2\Delta_{hh}.\end{aligned}$$

If instead $i \neq j$, then

$$\mathbb{E}[Q^*\Delta Q]_{ij} = \sum_{\substack{h\neq i\\ k\neq j}}\mathbb{E}\left[Q_{hi}\Delta_{hk}Q_{kj}\right] = p^2\sum_{\substack{h\neq i\\ k\neq j}}P_{hi}\Delta_{hk}P_{kj}.$$

More compactly, we can write

$$E[Q^*\Delta Q] = p^2\text{out }(P)^*\Delta\text{out }(P) + p(1-p)\text{diag }\{\text{out }(P)^*\text{diag }(\Delta)\text{out }(P)\}.$$

Putting all the pieces together we obtain relation (9). ∎

Following previous considerations, we are interested in evaluating the eigenvalues of the linear map which furnishes $\Delta(t+1)$ from $\Delta(t)$. These matrices are symmetric and so the linear dynamic system described in the previous proposition has a state space of dimension $\frac{N(N-1)}{2}$.

Remark 2. Numerical algorithms can clearly be employed to evaluate such eigenvalues. The following is a concrete way to achieve this. Given a matrix $A \in \mathbb{R}^{N\times N}$, we define vect$(A)$ to be the N^2 column vector having $A_{i,j}$ in position $(i-1)N+j$. Moreover, let

$$M := \begin{bmatrix} e_1e_1^* & 0 & \cdots & 0 \\ 0 & e_2e_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_Ne_N^* \end{bmatrix}, \tag{12}$$

where e_i is the vector with all zeros except for a 1 in the ith position. This matrix is such that $\text{vect}(\text{diag }(A)) = M\text{vect}(A)$. Finally, notice that $\text{vect}(ABC) = (C^* \otimes A)\text{vect}(B)$, where $\otimes$ is the Kronecker product of matrices. Using these facts and the properties of the Kronecker product we can argue that

$$\text{vect}(\Delta^+) = Z\text{vect}(\Delta),$$

where

$$\begin{aligned} Z = &\{(1-p)I + pP)^* \otimes ((1-p)I + pP)^*\} \\ &+ p(1-p)M\{I \otimes (\text{out }(P)\text{out }(P)^*) + \text{out }(P)^* \otimes \text{out }(P)^*\}M \\ &- p(1-p)\{\text{out }(\tilde{P})^* \otimes I + I \otimes \text{out }(\tilde{P})^*\}M. \end{aligned}$$

Then the rate of convergence R will coincide with the absolute value of the dominant reachable eigenvalue of the pair $(Z, \text{vect}(\Delta(0)))$.

Example 2. We apply the previous method for evaluating the rate of convergence for the following matrices*:*

$$P_1 = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 \\ 1/8 & 1/2 & 3/8 & 0 \\ 0 & 1/8 & 5/8 & 1/4 \\ 1/8 & 1/8 & 0 & 3/4 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix},$$

$$P_3 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}, \quad P_4 = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{bmatrix}.$$

The corresponding rate of convergence is illustrated in Figure 1. We will see in what follows that the same results can be found more easily for the matrices P_2, P_3, P_4.

3.2. The balanced compensation method. In the analysis of this case the following parameters will play a fundamental role:

$$\begin{aligned} \beta_{ih} &:= \mathbb{E}\left[\frac{P_{ih}L_{ih}}{\nu_i}\right] = \mathbb{E}[P_{ih}(0)], \\ \rho_{ihk} &:= \mathbb{E}\left[\frac{P_{ih}L_{ih}P_{ik}L_{ik}}{\nu_i^2}\right] = \mathbb{E}[P_{ih}(0)P_{ik}(0)]. \end{aligned}$$

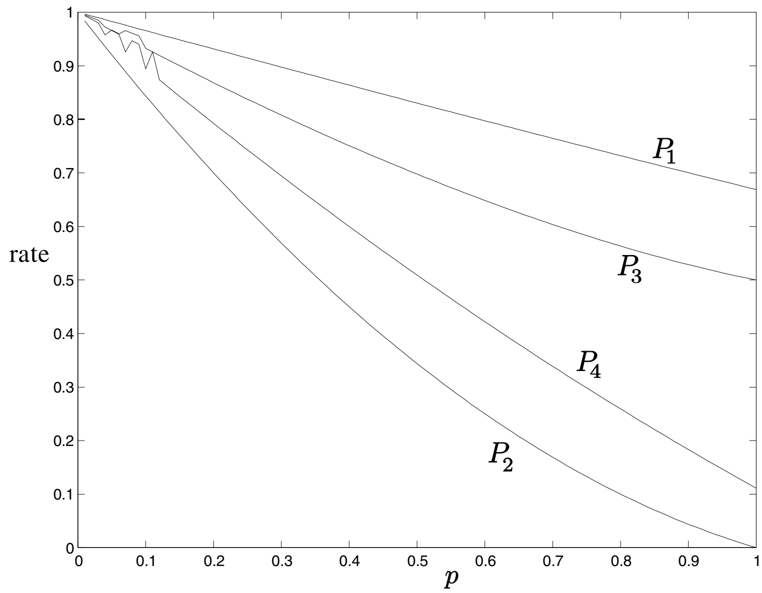

FIG. 1. *The graph of the rate of convergence for the matrices P_1, P_2, P_3, P_4 in Example 2.*

Notice that

$$\beta_{ih} = \sum_{\substack{v \in \{0,1\}^N \\ v_i = 1}} \frac{P_{ih} v_h}{\sum_s v_s P_{is}} p^{\mathbf{w}_H(v)-1}(1-p)^{N-\mathbf{w}_H(v)}$$

and

$$\rho_{ihk} = \sum_{\substack{v \in \{0,1\}^N \\ v_i = 1}} \frac{P_{ih} v_h P_{ik} v_k}{(\sum_s v_s P_{is})^2} p^{\mathbf{w}_H(v)-1}(1-p)^{N-\mathbf{w}_H(v)},$$

where $\mathbf{w}_H(v)$ is the Hamming weight. Therefore these parameters are polynomial functions of p of degree at most $N-1$. It is clear that $\rho_{ihk} = \rho_{ikh}$. The following lemma presents some other properties.

LEMMA 3.3. *The following relations hold true:*

$$\begin{aligned} &\sum_h \beta_{ih} = 1, \\ &\sum_{hk} \rho_{ihk} = 1, \qquad (13) \\ &\sum_k \rho_{ihk} = \beta_{ih}. \end{aligned}$$

Proof. We prove only the first one. The remaining relations can be proved in a similar way.

$$\begin{aligned} \sum_h \beta_{ih} &= \sum_h \mathbb{E}\left[\frac{P_{ih} L_{ih}}{\nu_i}\right] \\ &= \mathbb{E}\left[\frac{\sum_h P_{ih} L_{ih}}{\nu_i}\right] = 1. \qquad \square \end{aligned}$$

Now define the matrix

$$\bar{\beta} := \{\beta_{ij}\}.$$

Notice that $\bar{\beta} = \mathbb{E}[P(0)]$. By (13) this is a stochastic matrix and its graph coincides with the graph associated with the matrix P. Introduce, moreover, the linear operator $\bar{\rho}$ from the space of diagonal $N \times N$ matrices to the space of symmetric $N \times N$ matrices defined as follows:

$$\bar{\rho}(\text{diag } \{a_1, \dots, a_N\})_{ij} := \sum_k \rho_{kij} a_k.$$

Using these definitions, the relations proposed in the previous lemma can be translated to the following ones:

$$\bar{\beta}\mathbf{1} = \mathbf{1}, \qquad \mathbf{1}^* \bar{\rho}(e_i e_i^*)\mathbf{1} = 1, \qquad \bar{\rho}(e_i e_i^*)\mathbf{1} = \bar{\beta}^* e_i$$

for all $i = 1, \dots, N$, where e_i is the vector with all zeros except for a 1 in the ith position. The second condition is implied by the third and so can be eliminated. Moreover, the third condition is equivalent to the fact that for all diagonal matrices A, we have that

$$\bar{\rho}(A)\mathbf{1} = \bar{\beta}^* A \mathbf{1}.$$

The following result is less immediate to prove.

LEMMA 3.4. *If A is a nonnegative diagonal matrix, then $\bar{\rho}(A) - \bar{\beta}^* A \bar{\beta}$ is a positive semidefinite matrix.*

Proof. Notice that

$$\begin{aligned}
x^* \bar{\rho}(A) x &= \sum_i \sum_{hk} a_i \rho_{ihk} x_h x_k = \sum_i \sum_{hk} a_i \mathbb{E}\left[\frac{P_{ih} L_{ih} P_{ik} L_{ik}}{\nu_i^2}\right] x_h x_k \\
&= \sum_i a_i \mathbb{E}\left[\sum_{hk} \frac{P_{ih} L_{ih} P_{ik} L_{ik} x_h x_k}{\nu_i^2}\right] \\
&= \sum_i a_i \mathbb{E}\left[\left(\sum_h \frac{P_{ih} L_{ih} x_h}{\nu_i^2}\right)^2\right] \\
&\geq \sum_i a_i \mathbb{E}\left[\sum_h \frac{P_{ih} L_{ih} x_h}{\nu_i^2}\right]^2 = x^* \bar{\beta}^* A \bar{\beta} x. \qquad \square
\end{aligned}$$

We are now in a position to present the following result.

PROPOSITION 3.5. *The sequence of matrices $\Delta(t)$ satisfies the recursive relation*

$$\Delta^+ = \bar{\beta}^* \text{ out } (\Delta)\bar{\beta} + \bar{\rho}(\text{diag } (\Delta)). \tag{14}$$

Proof. Notice that

$$
\begin{aligned}
\mathbb{E}[P(0)^*\Delta P(0)]_{ij} = \mathbb{E}[Q^* D\Delta DQ]_{ij} &= \sum_h \sum_k \mathbb{E}\left[\frac{P_{hi}L_{hi}}{\nu_h}\frac{P_{kj}L_{kj}}{\nu_k}\right]\Delta_{hk} \\
&= \sum_{h\neq k} \mathbb{E}\left[\frac{P_{hi}L_{hi}}{\nu_h}\right]\left[\frac{P_{kj}L_{kj}}{\nu_k}\right]\Delta_{hk} + \sum_k \mathbb{E}\left[\frac{P_{ki}L_{ki}P_{kj}L_{kj}}{\nu_k^2}\right]\Delta_{kk} \\
&= \sum_{k\neq h} \beta_{hi}\Delta_{hk}\beta_{kj} + \sum_k \rho_{kij}\Delta_{kk} \\
&= \{\bar\beta^* \text{ out } (\Delta)\bar\beta + \bar\rho(\text{diag } (\Delta))\}_{ij}.
\end{aligned}
$$

This easily yields (14). ∎

Remark 3. Also in this case numerical algorithms can be employed to evaluate the rate of convergence. Introduce, moreover, the $N^2 \times N^2$ matrix T which is zero except in the following entries:

$$T_{(j-1)N+i,(s-1)N+s} = \rho_{sij}.$$

The matrix T is constructed in such a way that, for any diagonal matrix D, we have that

$$\text{vect}(\bar\rho(D)) = T\text{vect}(D).$$

Using the same arguments implemented in the previous remark we can argue that

$$\text{vect}(\Delta^+) = Z\text{vect}(\Delta),$$

where

$$Z = [\bar\beta^* \otimes \bar\beta^*](I - M) + TM,$$

and where the matrix M was defined in (12). Then the rate of convergence R will coincide with the absolute value of the dominant reachable eigenvalue of the pair $(Z, \text{vect}(\Delta(0)))$.

Example 3. We applied the previous method for evaluating the rate of convergence for the same matrices P_1, P_2, P_3, P_4 considered in Example 2. The corresponding rate of convergence is illustrated in Figure 2 and compared with the rates obtained by the biased compensation method. The balanced compensation method outperforms the biased compensation method for all the matrices except for P_3 in which the two methods coincide. We will see in what follows that the same results can be found more easily for the matrices P_2, P_3, P_4.

In what follows we will make further analytical developments assuming the graph $\mathcal{G}$ possesses some more symmetry; more precisely, we will work with Cayley graphs.

4. Cayley matrices over Abelian groups. For graphs possessing symmetries, the theoretical results obtained in the previous section can be refined quite a bit. In this paper we will deal with a special class of symmetric graphs: Abelian Cayley graphs [1].

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

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

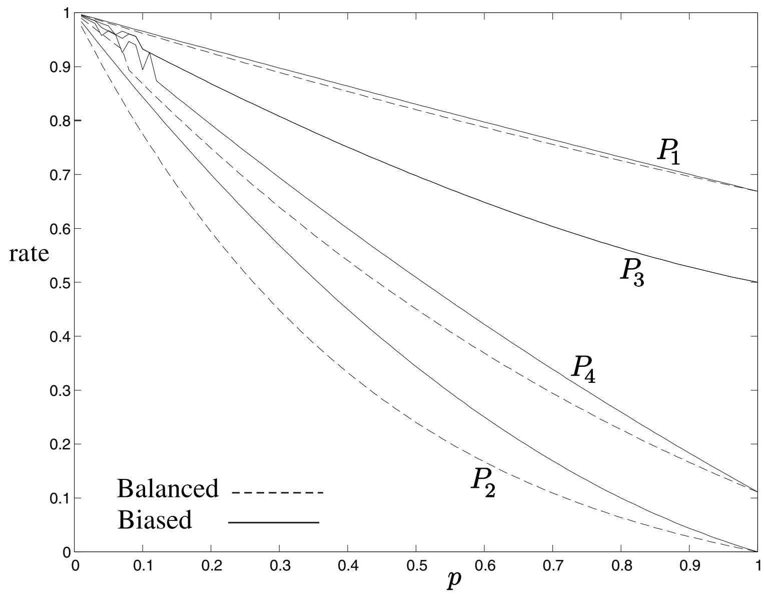

FIG. 2. *The graph of the rate of convergence for the matrices P_1, P_2, P_3, P_4. Biased compensation method rate of convergence is described by the continuous line; balanced compensation method rate of convergence is described by the dashed line.*

Notice that a Cayley graph is always in-regular and out-regular: Both the in-degree and the out-degree of each vertex are equal to $|S|$. Notice also that strong 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.

Symmetries can also be introduced 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 $P\mathbf{1} = \mathbf{1}$ if and only if $\mathbf{1}^*P = \mathbf{1}^*$. This implies that a Cayley stochastic matrix is automatically doubly stochastic.

4.1. Spectral properties and Fourier analysis of Cayley matrices over Abelian groups. In this subsection we will show that the spectral properties of Cayley matrices over Abelian groups are particularly simple to analyze. We briefly review the theory of Fourier transform over finite Abelian groups (see [19] for a comprehensive treatment of the topic). Let G be a finite Abelian group of order N as above, and let $\mathbb{C}^*$ be the multiplicative group of the nonzero complex numbers. A character on G is a group homomorphism $\chi : G \to \mathbb{C}^*$, namely, a function χ from G to $\mathbb{C}^*$ such that $\chi(g+h) = \chi(g)\chi(h)$ for all $g, h \in G$. Since we have that

$$\chi(g)^N = \chi(Ng) = \chi(0) = 1 \qquad \forall g \in G,$$

it follows that χ takes values on the Nth roots of unity. The character $\chi_0(g) = 1$ for every $g \in G$ is called the trivial character.

The set of all characters of the group G forms an Abelian group with respect to the pointwise multiplication. It is called the character group and denoted by $\hat{G}$. The trivial character χ_0 is the zero of $\hat{G}$. If we consider the vector space $\mathbb{C}^G$ of all functions from G to $\mathbb{C}$ with the canonical Hermitian form

$$\langle f_1, f_2 \rangle = \sum_{g \in G} f_1(g) f_2(g)^* ,$$

then it can be shown that the set $\{N^{-1/2}\chi \mid \chi \in \hat{G}\}$ is an orthonormal basis of $\mathbb{C}^G$.

The Fourier transform of a function $f : G \to \mathbb{C}$ is defined as

$$\hat{f} : \hat{G} \to \mathbb{C} , \quad \hat{f}(\chi) = \sum_{g \in G} \chi(-g) f(g) .$$

Now fix a Cayley matrix P on the Abelian group G generated by the function $\pi_P : G \to \mathbb{R}$. The spectral structure of P is very simple. Namely, it can be shown that the characters $\chi \in \hat{G}$ are eigenvectors of P and so P is diagonalizable. Moreover, the spectrum of P is given by the Fourier transform of the generator π_P of P:

$$\sigma(P) = \{\hat{\pi}_P(\chi) \mid \chi \in \hat{G}\} .$$

Notice that, if A, B are Cayley matrices with Fourier transforms $\hat{\pi}_A(\chi), \hat{\pi}_B(\chi)$, then

$$\hat{\pi}_{A+B}(\chi) = \hat{\pi}_A(\chi) + \hat{\pi}_B(\chi), \qquad \hat{\pi}_{AB}(\chi) = \hat{\pi}_A(\chi)\hat{\pi}_B(\chi).$$

Moreover, observe that, if A is a Cayley matrix, then diag (A) and out (A) are also Cayley and we have

$$\text{diag } (A) = N^{-1} \text{trace } (A) \; I = N^{-1} \sum_{\bar{\chi}} \hat{\pi}_A(\bar{\chi}) \; I .$$

This implies that, for every $\chi \in \hat{G}$,

$$\hat{\pi}_{\text{diag } (A)}(\chi) = N^{-1} \sum_{\bar{\chi} \in \hat{G}} \hat{\pi}_A(\bar{\chi}),$$
$$\hat{\pi}_{\text{out } (A)}(\chi) = \hat{\pi}_A(\chi) - N^{-1} \sum_{\bar{\chi}} \hat{\pi}_A(\bar{\chi}).$$

5. Mean square analysis for Cayley matrices. In this section we will show that when P is a Cayley matrix, the analysis proposed above simplifies considerably. Let G be a finite Abelian of order N, and let P be a Cayley matrix with respect to G. It easily follows from Propositions 3.2 and 3.5 that $\Delta(t)$ are Cayley matrices. This in particular implies that the matrices $\Delta(t)$ admit a common orthonormal basis of eigenvectors. In other words, there exists an $N \times N$ unitary matrix U such that $U^*\Delta(t)U = \tilde{\Delta}(t)$ is diagonal for every t. We can then write

$$\tilde{\Delta}(t+1) = \mathbb{E}[U^*P(0)^*U\tilde{\Delta}(t)U^*P(0)U]. \tag{15}$$

This shows that there exists a linear operator $\tilde{\mathcal{L}}$ such that $\tilde{\Delta}(t+1) = \tilde{\mathcal{L}}(\tilde{\Delta}(t))$ for every t. It is clear that

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

where $\tilde{\mathcal{R}}$ is the reachable subspace of the pair $(\tilde{\mathcal{L}}, \tilde{\Delta}(0))$, where

$$\tilde{\Delta}(0) = U(I - N^{-1}\mathbf{1}\mathbf{1}^*)U^* = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Notice that (15) is an evolution equation on the eigenvalues of the matrices $\Delta(t)$ which we know to be given by the Fourier transforms $\hat{\pi}_{\Delta(t)}(\chi)$ of the generating function $\pi_{\Delta(t)}(g)$. We will use these representations to express, in a more explicit form, the operator $\tilde{\mathcal{L}}$. As before we will handle the two cases separately.

5.1. The biased compensation method. In the event that P is a Cayley matrix, for the biased compensation method the evolution of the eigenvalues of $\Delta(t)$ is described by the following proposition. First notice that, since P is a Cayley matrix, $\tilde{P}$ and P^*P are also Cayley matrices.

PROPOSITION 5.1. *For all $\chi \in \hat{G}$ we have that*

$$\hat{\pi}_{\Delta(t+1)}(\chi) = A(\chi)\hat{\pi}_{\Delta(t)}(\chi) + B(\chi)N^{-1}\sum_{\bar{\chi}\in\hat{G}}\hat{\pi}_{\Delta(t)}(\bar{\chi}), \tag{16}$$

where

$$A(\chi) = |1 - p + p\hat{\pi}_P(\chi)|^2 \tag{17}$$

and

$$B(\chi) = 2p(1-p)\left\{\hat{\pi}_{\tilde{P}}(\chi_0) - \Re\left[\hat{\pi}_{\tilde{P}}(\chi)\right]\right\}. \tag{18}$$

Proof. We start with formula (9). First notice that, since the matrix diag (Δ) is Cayley and diagonal, it is a scalar multiple of the identity, namely, diag $(\Delta) = xI$. This implies that

$$\begin{aligned}\Delta^+ &= [(1-p)I + pP]^*\Delta[(1-p)I + pP] \\ &\quad + p(1-p)\Big\{\text{diag }[\text{out }(P)\text{out }(P)^* + \text{out }(P)^*\text{out }(P)] - \text{out }(\tilde{P}) \\ &\quad - \text{out }(\tilde{P})^*\Big\}\,\text{diag }(\Delta) \\ &= [(1-p)I + pP]^*\Delta[(1-p)I + pP] + p(1-p)\Big\{\text{diag }[PP^* + P^*P] - \tilde{P} - \tilde{P}^*\Big\}\,xI.\end{aligned}$$

Notice now that

$$\text{diag }[PP^* + P^*P] = 2\hat{\pi}_{\tilde{P}}(\chi_0)\ I$$

and that

$$x = N^{-1} \sum_{\bar{\chi} \in \hat{G}} \hat{\pi}_{\Delta}(\bar{\chi}).$$

These facts yield (16). $\square$

Remark 4. Notice that $\tilde{P}$ is an irreducible nonnegative Cayley matrix and so $\hat{\pi}_{\tilde{P}}(\chi_0)$ is its spectral radius. This implies that $B(\chi) \geq 0$ and that $B(\chi) = 0$ if and only if $\chi = \chi_0$.

The linear dynamic system described in (16) can finally be rewritten in a more compact way as follows. Enumerate in some way the characters of G, $\hat{G} = \{\chi_0, \chi_1, \dots, \chi_{N-1}\}$, and define the column vector in $\mathbb{R}^N$ as

$$\Pi(t) := \begin{bmatrix} \hat{\pi}_{\Delta(t)}(\chi_0) \\ \vdots \\ \hat{\pi}_{\Delta(t)}(\chi_{N-1}) \end{bmatrix}. \tag{19}$$

Define, moreover, the column vector B in $\mathbb{R}^N$ such that for all $i = 0, 1, \dots, N-1$, we have that $B_i := B(\chi_i)$ and the diagonal matrix A such that for all $i = 0, 1, \dots, N-1$, we let $A_{ii} := A(\chi_i)$. Then we can write the linear dynamic system (16) as follows:

$$\Pi(t+1) = \left(A + N^{-1} B \mathbf{1}^*\right) \Pi(t).$$

Notice that both A and B depend on the probability p and so in some cases we will write $A(p)$ and $B(p)$ to make this dependence evident. Notice that $A_{ii}(p) < 1$ if $p > 0$, while $A_{ii}(0) = 1$ for all i. Moreover, we have that $B_0(p) = 0$ for all i and $0 < B_i(p) < 1$ if $i \neq 0$ and $0 < p < 1$, while $B_i(0) = B_i(1) = 0$.

We have the following result.

PROPOSITION 5.2. *We have the following properties:*

(a) *The matrix $A + N^{-1} B \mathbf{1}^*$ has nonnegative entries.*

(b) *It has the structure*

$$A + N^{-1} B \mathbf{1}^* = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ X_{21} & & X_{22} & \end{pmatrix}, \tag{20}$$

where $X_{21} \in \mathbb{R}^{1 \times (N-1)}$ and $X_{22} \in \mathbb{R}^{(N-1) \times (N-1)}$ have nonnegative entries.

(c) $R = \max\{|\lambda| \ : \ \lambda \text{ eigenvalue of } \ X_{22}\} \geq \max\{A_{ii} : 1 = 1, \dots, N-1\}$.

(d) *The eigenvector of $A + N^{-1} B \mathbf{1}^*$ relative to the eigenvalue* 1 *has a nonzero first component.*

Proof. (a) follows from the previous remark.

(b) can be proven by inspection.

(c) Notice that $\Pi(0) = (0, 1, \dots, 1)^*$. Because of (a), this shows that the reachability subspace $\tilde{\mathcal{R}}$ of the pair $(X, \Pi(0))$ has the structure $\tilde{\mathcal{R}} = \{0\} \times \tilde{\mathcal{R}}_2$, where $\tilde{\mathcal{R}}_2$ is the reachability subspace of $(X_{22}, \mathbf{1})$. Since X_{22} has nonnegative entries, its spectral radius is achieved by a nonnegative eigenvalue λ with a corresponding nonnegative eigenvector $\bar{w}$ [7, p. 66]. Clearly, we can write $\mathbf{1} = a\bar{w} + w'$ for some $a > 0$ and another nonnegative vector w'. Hence,

$$X_{22}^t \mathbf{1} = a \lambda^t \bar{w} + X_{22}^t w' \geq a \lambda^t \bar{w}\,.$$

From this it immediately follows that $R \geq \lambda$. This clearly proves the equality in (c). Finally, the fact that $R \geq \max\{A_{ii} : 1 = 1, \dots, N-1\}$ follows from the fact that B has nonnegative entries [10, Corollary 8.1.19].

(d) Finally, consider any eigenvector $w \in \mathbb{R}^N$ of $A+N^{-1}B\mathbf{1}^*$ relative to the eigenvalue 1. If the first component w_0 of w were zero, then the vector $(w_1, \dots, w_{N-1})^*$ would be an eigenvector of X_{22} relative to the eigenvalue 1. This could not be possible, however, since we know that 1 is an eigenvalue of $A + N^{-1}B\mathbf{1}^*$ with algebraic multiplicity equal to 1. □

Proposition 5.2 reduces the computation of R to the computation of the second dominant eigenvalue of the $N \times N$ matrix (20). An explicit expression for the characteristic polynomial of (20) can be obtained through the following lemma.

LEMMA 5.3.

$$\det\left(A + N^{-1}B\mathbf{1}^*\right) = \prod_{j=0}^{N-1} A_{jj} + N^{-1}\sum_{i=0}^{N-1} B_i \prod_{\substack{j=0\\ j\neq i}}^{N-1} A_{jj}$$

Proof. Notice that

$$\begin{aligned}&\det\left(A + N^{-1}B\mathbf{1}^*\right)\\ &= \det(A)\det\left(I + N^{-1}B\mathbf{1}^*A^{-1}\right) = \det(A)\left(1 + N^{-1}\mathbf{1}^*A^{-1}B\right). \qquad \Box\end{aligned}$$

From this lemma we can argue that

$$\begin{aligned}F(z,p) &:= \det\left(zI - A(p) - N^{-1}B(p)\mathbf{1}^*\right)\\ &= \prod_{j=0}^{N-1}(z - A_{jj}(p)) - N^{-1}\sum_{i=0}^{N-1} B_i(p)\prod_{\substack{j=0\\ j\neq i}}^{N-1}(z - A_{jj}(p)).\end{aligned}$$

The polynomial $F(z,p)$ has degree N in z and degree $2n$ in p. The stability analysis of this polynomial can be in general quite complicated. We will investigate this problem through some examples.

We start with a couple of examples in which the eigenvalues can be determined exactly and also some natural optimization design can be carried out.

Example 4. Consider the matrix $P = (1-k)I + \frac{k}{N}\mathbf{1}\mathbf{1}^*$. It is clear that the matrix P is in this case a Cayley matrix over the group $\mathbb{Z}_N$ and with $S = \mathbb{Z}_N$. After some computation we can find that

$$A_{ii}(k,p) = \begin{cases} 1 & \text{if } i = 0,\\ (1-kp)^2 & \text{if } i \neq 0\end{cases} \tag{21}$$

and

$$B_i(k,p) = \begin{cases} 0 & \text{if } i = 0,\\ \frac{2p(1-p)k^2}{N} & \text{if } i \neq 0,\end{cases} \tag{22}$$

and so the eigenvalues are

$$\begin{aligned}\bar{z}_0(k,p) &= 1,\\ \bar{z}_1(k,p) &= (1-kp)^2 + 2p(1-p)k^2\frac{N-1}{N^2},\\ \bar{z}_i(k,p) &= (1-kp)^2 \qquad i = 2,\dots,N-1.\end{aligned}$$

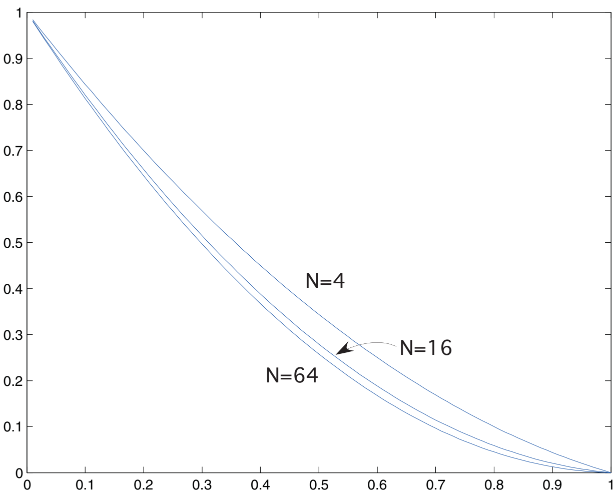

FIG. 3. *The graph of the rate of convergence in Example* 4 *for* $N = 4, 16, 64$.

The rate of convergence to the consensus is determined by the eigenvalue $\bar{z}_1(k, p)$. In this case the optimal k yielding the fastest convergence can be computed analytically. Indeed, it can be seen that

$$k = \frac{N^2}{N^2 p + 2(1-p)(N-1)}.$$

For large N we have that $k \simeq 1/p$. In Figure 3 we show the graph of the rate of convergence as a function of the probability p for $N = 4, 16, 64$ when $k = 1$. It can be shown that the graph relative to the case $N = 4$ coincides up to numerical errors with the one obtained in Example 2 for the matrix P_2.

Example 5. Consider the case in which the group is $\mathbb{Z}_N$ and $S = \{0, 1\}$. Consider a matrix P with generator $\pi_P(0) = 1 - k, \pi_P(1) = k$, and $\pi_P(g) = 0$ for all $g \neq 0, 1$. In this case we have that

$$\begin{aligned} \hat{\pi}_P(\chi_i) &= 1 - k + k e^{j\frac{2\pi}{N}i}, \\ \hat{\pi}_{\tilde{P}}(\chi_i) &= (1-k)^2 + k^2 e^{j\frac{2\pi}{N}i}. \end{aligned}$$

From this we can argue that

$$A_{ii}(p) = 1 - 2pk(1 - pk)\left(1 - \cos\left(\frac{2\pi}{N}i\right)\right), \tag{23}$$

$$B_i(p) = 2p(1-p)k^2\left(1 - \cos\left(\frac{2\pi}{N}i\right)\right). \tag{24}$$

With fixed probability p one can find the optimal k yielding the fastest convergence. We did this numerically for $N = 5, 10, 20$. The graph showing the optimal k as a function of p is illustrated in Figure 4. In Figure 5 we show the graph of the rate of

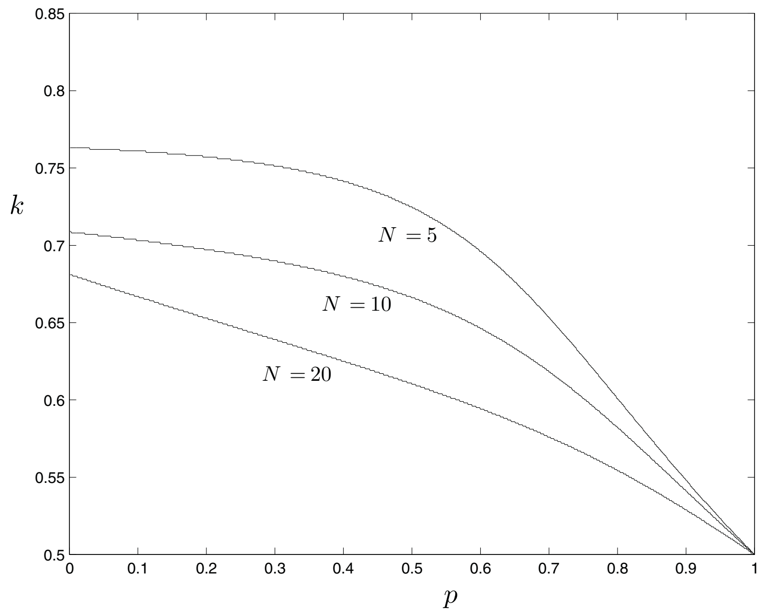

FIG. 4. *The graph of the optimal k as a function of the probability p in Example 5 for $N = 5, 10, 20$.*

convergence as a function of the probability p for $N = 2, 4, 8$ when $k = 1/2$. It can be shown that the graph relative to the case $N = 4$ coincides up to numerical errors with the one obtained in Examples 2 and 3 for the matrix P_3.

The next example relates instead to the hypercube graph. We present only the analytical computation of the eigenvalues.

Example 6. Consider the case in which the group is $\mathbb{Z}_2^n$ and

$$S = \{0, e_1, \ldots, e_n\},$$

where e_i is the vector with all zeros except for a 1 in the ith position. Let E be the adjacency matrix of the graph defined in this way and consider the matrix $P := \frac{1}{n+1}E$. This means that given $u, v \in \mathbb{Z}_2^n$ we have that

$$P_{u,v} = \begin{cases} \frac{1}{n+1} & \text{if } u + v \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that, in this case, we have that $\tilde{P} = \frac{1}{n+1}P$. It can be shown that for all $v \in \mathbb{Z}_2^n$ we have that

$$\hat{\pi}_P(v) = 1 - \frac{2}{n+1}\mathbf{w}_H(v),$$

where $\mathbf{w}_H(v)$ is the Hamming weight of v, namely, the number of 1's. From this we can argue that

$$A_v = \left(1 - \frac{2p}{n+1}\mathbf{w}_H(v)\right)^2,$$

$$B_v = \frac{4p(1-p)}{(n+1)^2}\mathbf{w}_H(v).$$

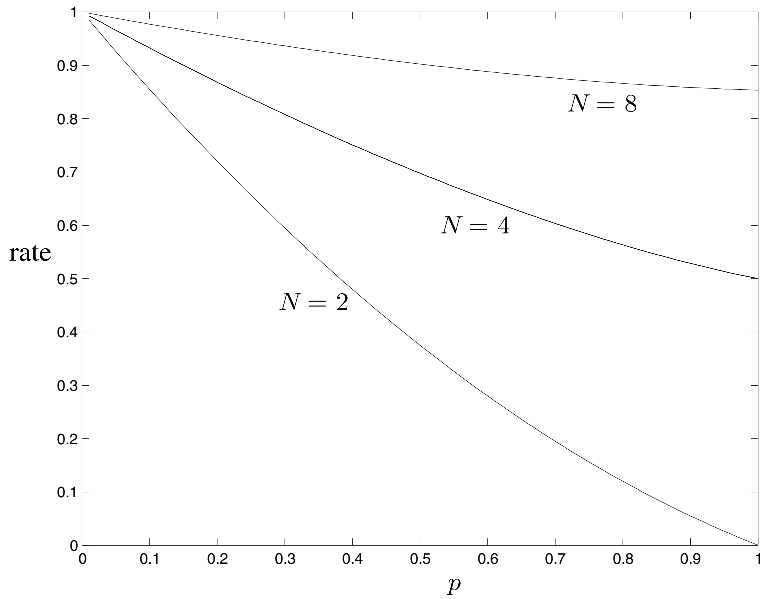

FIG. 5. *The graph of the rate of convergence in Example* 5 *for* $N = 2, 4, 8$.

From this it follows that

$$F(z,p) = \prod_{h=0}^{n} (z - A_h)^{\binom{n}{h}-1} \left\{ \prod_{h=0}^{n} (z - A_h) - N^{-1} \sum_{k=0}^{n} \binom{n}{k} B_k \prod_{\substack{h=0 \\ h \neq k}}^{n} (z - A_h) \right\},$$

where, for $h = 0, 1, \ldots, n$, we let

$$A_h = \left(1 - \frac{2p}{n+1} h\right)^2,$$
$$B_h = \frac{4p(1-p)}{(n+1)^2} h.$$

This implies that $N - n$ eigenvalues coincide with $A_h(p)$, $h = 0, 1, \ldots, n$, while the remaining n are the roots of

$$\prod_{h=0}^{n} (z - A_h) - N^{-1} \sum_{k=0}^{n} \binom{n}{k} B_k \prod_{\substack{h=0 \\ h \neq k}}^{n} (z - A_h).$$

Figure 7 shows the graph of the rate of convergence as a function of the probability p for $n = 2, 4$. It can be shown that the graph relative to the case $n = 2$ coincides up to numerical errors with the one obtained in Example 2 for the matrix P_4.

Now we present an example where instead only numerical results can be obtained.

5.2. The balanced compensation method. First notice that, if P is a Cayley matrix, we have the following result.

LEMMA 5.4. *If P is a Cayley matrix, then $\bar{\beta}, \bar{\rho}(I)$ are also Cayley matrices.*

Proof. Notice that

$$\begin{aligned}\beta_{i+l,h+l} &= \sum_{\substack{v\in\{0,1\}^N\\ v_{i+l}=1}} \frac{P_{i+l,h+l}v_{h+l}}{\sum_s v_s P_{i+l,s}} p^{\mathbf{w}_H(v)-1}(1-p)^{N-\mathbf{w}_H(v)}\\ &= \sum_{\substack{v\in\{0,1\}^N\\ v_{i+l}=1}} \frac{P_{i,h}v_{h+l}}{\sum_s v_s P_{i,s-l}} p^{\mathbf{w}_H(v)-1}(1-p)^{N-\mathbf{w}_H(v)}.\end{aligned}$$

We now define for any $v \in \{0,1\}^N$ a $u \in \{0,1\}^N$ such that $v_s = u_{s-l}$. Then

$$\begin{aligned}\beta_{i+l,h+l} &= \sum_{\substack{u\in\{0,1\}^N\\ u_i=1}} \frac{P_{i,h}u_h}{\sum_s u_{s-l} P_{i,s-l}} p^{\mathbf{w}_H(u)-1}(1-p)^{N-\mathbf{w}_H(u)}\\ &= \sum_{\substack{u\in\{0,1\}^N\\ u_i=1}} \frac{P_{i,h}u_h}{\sum_s u_s P_{i,s}} p^{\mathbf{w}_H(u)-1}(1-p)^{N-\mathbf{w}_H(u)} = \beta_{i,h}.\end{aligned}$$

In a similar way we can prove that $\rho_{k,i,j} = \rho_{k+l,i+l,j+l}$. From this it follows that $\bar\rho(I)$ is a Cayley matrix. ∎

In this case we have the following proposition.

PROPOSITION 5.5. *For all $\chi \in \hat G$ we have that*

$$\hat\pi_{\Delta(t+1)}(\chi) = A(\chi)\hat\pi_{\Delta(t)}(\chi) + B(\chi)N^{-1}\sum_{\bar\chi\in\hat G}\hat\pi_{\Delta(t)}(\bar\chi), \tag{25}$$

where

$$A(\chi) = \left|\hat\pi_{\bar\beta}(\chi)\right|^2 \tag{26}$$

and

$$B(\chi) = \hat\pi_{\bar\rho(I)}(\chi) - \left|\hat\pi_{\bar\beta}(\chi)\right|^2 . \tag{27}$$

Proof. We start from the equation

$$\Delta^+ = \bar\beta^* \text{out } (\Delta)\bar\beta + \bar\rho(\text{diag } (\Delta)).$$

Notice now that diag (Δ) is a multiple of the identity, namely diag $(\Delta) = xI$. This implies that

$$\Delta^+ = \bar\beta^*\Delta\bar\beta + \{\bar\rho(I) - \bar\beta^*\bar\beta\}x.$$

Considering the fact that

$$x := N^{-1}\sum_{\bar\chi}\hat\pi_\Delta(\bar\chi),$$

we obtain the thesis. ∎

As in the previous case, using the notation (19), we can rewrite (25) as

$$\Pi(t+1) = (A + N^{-1}B\mathbf{1}^*)\Pi(t),$$

where B is the column vector in $\mathbb{R}^N$ such that for all $i = 0, 1, \ldots, N-1$, we have that $B_i := B(\chi_i)$ and A is the diagonal matrix such that $A_{ii}(\chi) := A(\chi_i)$. As in the previous case we will use the notation $A(p)$ and $B(p)$ whenever we want to underline the dependence on p.

Notice that, as observed above, when $p > 0$, the matrix $\bar{\beta}$ is an irreducible stochastic matrix. This implies that $A_{ii}(p) < 1$ if $p > 0$. On the other hand, since when $p = 0$ we have $\bar{\beta} = \bar{\rho}(I) = I$, then $A_{ii}(0) = 1$ and $B_i(0) = 0$ for all i. Finally, when $p = 1$ we have that $\bar{\beta} = P$ and $\bar{\rho}(I) = P^*P$, and so $B_i(1) = 0$.

From Lemma 3.3 we can argue that $\hat{\pi}_{\bar{\rho}(I)}(\chi_0) = \hat{\pi}_{\bar{\beta}}(\chi_0) = 1$ and so $B_0(p) = 0$ for all p. Using Lemma 3.4 it can be shown that Proposition 5.2 still holds and as above we can argue that the eigenvalues of $A(p) + N^{-1}B(p)\mathbf{1}^*$ coincide with the roots of the polynomial

$$F(z,p) = \prod_{j=0}^{N-1}(z - A_{jj}(p)) - N^{-1}\sum_{i=0}^{N-1} B_i(p) \prod_{\substack{j=0 \\ j\neq i}}^{N-1}(z - A_{jj}(p)).$$

In some cases some further simplifications can be introduced. Consider a Cayley graph $\mathcal{G}$. Since each node of $\mathcal{G}$ has exactly the same number n (excluding self loops) of incoming edges and outgoing edges, we can introduce a Cayley matrix $\bar{P}$ compatible with $\mathcal{G}$ by letting

$$\bar{P}_{ij} = \begin{cases} 1/n & \text{if } (j,i) \text{ is an edge of the graph,} \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, let

$$P := (1-k)I + k\bar{P}. \tag{28}$$

In this way we obtained a family of Cayley matrices P compatible with the graph $\mathcal{G}$. In this case the parameters β_{ih}, ρ_{ihk} become simpler to evaluate. Indeed, let E be the adjacency matrix of the graph and $H := E - I$. Moreover, let b_k be a binomial random variable, namely, a random variable taking value on the nonnegative integers with law

$$\mathbb{P}[b_k = i] = \binom{k}{i} p^i (1-p)^{k-i}, \quad i = 0, 1, \ldots, k.$$

After some simple but lengthy calculations it can be shown that

$$\begin{array}{ll} \beta_{ii} = \alpha & \forall\, i, \\ \beta_{ih} = \beta H_{ih} & \forall\, i,h \text{ such that } i \neq h, \\ \rho_{iii} = \gamma, & \\ \rho_{iih} = \rho_{ihi} = \delta H_{ih} & \forall\, i,h \text{ such that } i \neq h, \\ \rho_{ihh} = \xi H_{ih} & \forall\, i,h \text{ such that } i \neq h, \\ \rho_{ihk} = \rho H_{ih} H_{ik} & \forall\, i,h,k \text{ that are different from each other,} \end{array}$$

where

(29)

$$\alpha := \mathbb{E}\left[\frac{n-nk+k}{n-nk+k+kb_{n-1}}\right], \qquad \beta := p\mathbb{E}\left[\frac{k}{n-nk+2k+kb_{n-2}}\right],$$

$$\gamma := \mathbb{E}\left[\frac{(n-nk+k)^2}{(n-nk+k+kb_{n-1})^2}\right], \qquad \delta := p\mathbb{E}\left[\frac{(n-nk+k)k}{(n-nk+2k+kb_{n-2})^2}\right],$$

$$\xi := p\mathbb{E}\left[\frac{k^2}{(n-nk+2k+kb_{n-2})^2}\right], \qquad \rho := p^2\mathbb{E}\left[\frac{k^2}{(n-nk+3k+kb_{n-3})^2}\right].$$

These parameters depend on k, n, and p. The relations (13) become

$$\begin{aligned}&\alpha+\beta(n-1)=1,\\&\gamma+\delta(n-1)=\alpha,\\&(n-2)\rho=\beta-\xi-\delta.\end{aligned}$$

It is clear that $\bar{\beta}=\alpha I+\beta H$. Moreover, after some computations, it can be shown that, if D is any diagonal matrix, then in this case

$$\bar{\rho}(D)=\rho H^*DH+(\xi-\rho)\mathrm{diag}\,(H^*DH)+\gamma D+\delta(H^*D+DH). \tag{30}$$

Under these assumptions we can write

$$A(\chi)=|\alpha+\beta\hat{\pi}_H(\chi)|^2$$

and

$$B(\chi)=\rho\,|\hat{\pi}_H(\chi)|^2+(\xi-\rho)(n-1)+\gamma+2\delta\,\Re\,[\hat{\pi}_H(\chi)]-|\alpha+\beta\hat{\pi}_H(\chi)|^2\,.$$

We now want to compare the two compensation methods proposed here through the examples presented previously. Notice that in Example 5 the two compensation methods coincide.

Example 7. Now consider the matrix $P=(1-k)I+N^{-1}\mathbf{1}\mathbf{1}^*$ considered in Example 4. After some computation we can find that, for $i=1,\dots,N-1$,

$$A_{ii}(k,p)=(1-N\beta)^2, \quad B_i(k,p)=(1-N\beta)N\beta+N(\xi-\delta) \tag{31}$$

and so the eigenvalues are

$$\begin{aligned}\bar{z}_0(k,p)&=1,\\\bar{z}_1(k,p)&=(1-N\beta)(1-\beta)+(N-1)(\xi-\delta),\\\bar{z}_i(k,p)&=(1-\beta N)^2, \qquad i=2,\dots,N-1.\end{aligned}$$

The rate of convergence to the consensus is determined by $\bar{z}_1(k,p)$. Figure 6 shows the graph of the dominant eigenvalue $\bar{z}_1(1,p)$ as a function of the probability p for $N=4,16,64$ when $k=1$. (We are not making any optimization in this case.) It can be shown that the graph relative to the case $N=4$ coincides up to numerical errors with the one obtained in Example 2 for the matrix P_2.

Example 8. Consider the same matrix P introduced in Example 6. It can be shown that for all $v\in\mathbb{Z}_2^n$ we have that

$$\begin{aligned}A_v&=(1-2\beta\mathbf{w}_H(v))^2,\\b_v&=4\mathbf{w}_H(v)(\delta-\rho+\mathbf{w}_H(v)(\rho-\beta^2)),\end{aligned}$$

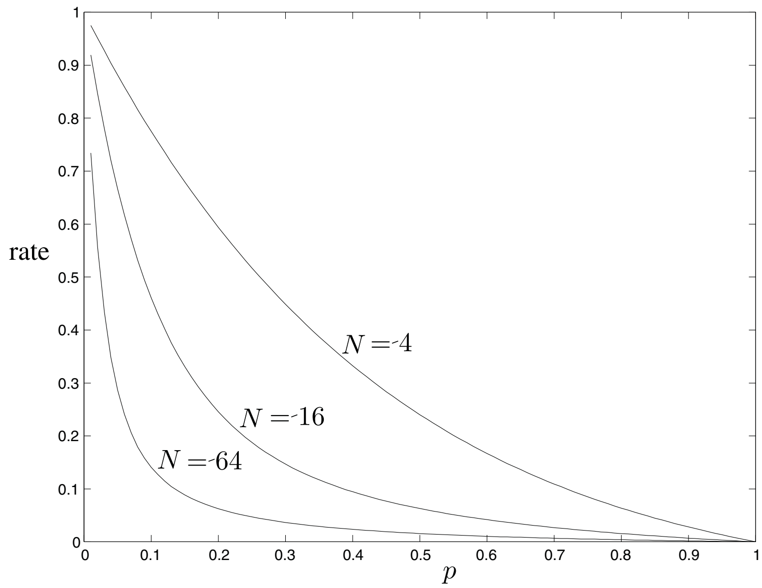

FIG. 6. *The graph of the rate of convergence in Example 7 for $N = 4, 16, 64$.*

where $\mathbf{w}_H(v)$ is the Hamming weight of v, namely, the number of 1's. From this it follows that

$$F(z,p) = \prod_{h=0}^{n} (z - A_h)^{\binom{n}{h}-1} \left\{ \prod_{h=0}^{n} (z - A_h) - N^{-1} \sum_{k=0}^{n} \binom{n}{k} b_k \prod_{\substack{h=0 \\ h\neq k}}^{n} (z - A_h) \right\},$$

where, for $h = 0, 1, \ldots, n$, we let

$$\begin{aligned} A_h &= (1 - 2\beta h)^2, \\ b_h &= 4h(\delta - \rho + h(\rho - \beta^2)). \end{aligned}$$

This implies that $N - n$ eigenvalues coincide with $A_h(p)$, $h = 0, 1, \ldots, n$ while the remaining n are the roots of

$$\prod_{h=0}^{n} (z - A_h) - N^{-1} \sum_{k=0}^{n} \binom{n}{k} b_k \prod_{\substack{h=0 \\ h\neq k}}^{n} (z - A_h)$$

and can be estimated when $p \simeq 1$ by the method proposed above. Figure 7 shows the graph of the rate of convergence as a function of the probability p for $N = 2, 4$ in the case of the biased compensation and the balanced compensation methods. It can be shown that the graph relative to the case $n = 2$ coincides up to numerical errors with the one obtained in Example 2 for the matrix P_4.

6. Average consensus. Even if the original algorithm was chosen to solve the average consensus problem, in general the perturbed solutions due to packet drops will no longer satisfy this property. In this section we will show how to estimate the distance of the consensus point from the average of the initial conditions. From

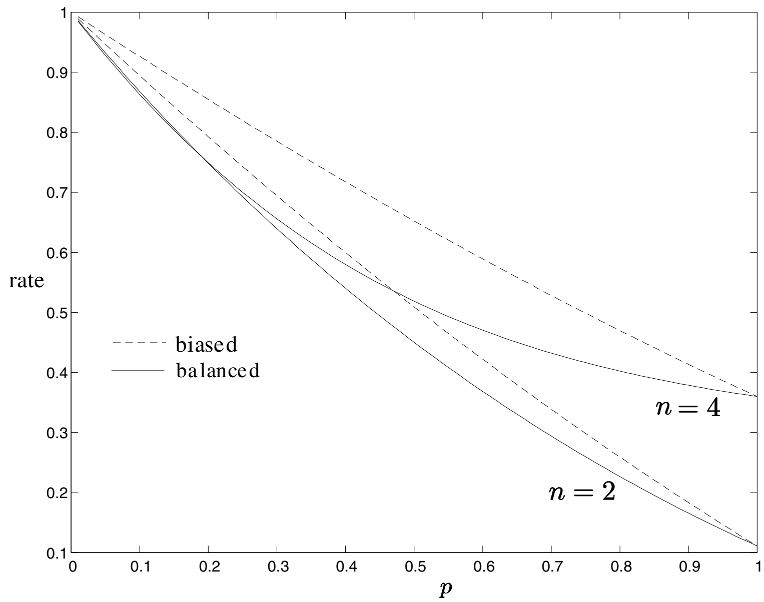

FIG. 7. *The graph of the rate of convergence in Examples* 6 *and* 8 *for* $n = 2, 4$ *in the case of the biased compensation method and the balanced compensation method.*

now on we will assume that the matrix P is doubly stochastic so that $P\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^*P = \mathbf{1}^*$. Consider $x_A(t)$ as defined in (5) and let

$$\begin{aligned} D &:= \sup_{||x(0)||\le 1} \mathbb{E}[|x_A(\infty) - x_A(0)|^2] = \sup_{||x(0)||\le 1} \mathbb{E}[|(v^* - N^{-1}\mathbf{1}^*)x(0)|^2] \\ &= \sup_{||x(0)||\le 1} x(0)^*\mathbb{E}[(v - N^{-1}\mathbf{1})(v - N^{-1}\mathbf{1})^*]x(0) \\ &= \max\{|\lambda| \ : \ \lambda \text{ eigenvalue of } \mathbb{E}[(v - N^{-1}\mathbf{1})(v - N^{-1}\mathbf{1})^*]\}\,. \end{aligned}$$

Notice that D is expressed in terms of the random vector v which in general may not be explicitly available. A further step, however, allows us to write

$$\mathbb{E}[(v - N^{-1}\mathbf{1})(v - N^{-1}\mathbf{1})^*] = \mathbb{E}[vv^*] - N^{-1}\mathbb{E}[v]\mathbf{1}^* - \mathbf{1}N^{-1}\mathbb{E}[v]^* + N^{-2}\mathbf{1}\mathbf{1}^*. \tag{32}$$

We now recall that $\mathbb{E}[vv^*]$ is the dominant eigenvector of the positive operator $\mathcal{L}$ and can thus be computed using standard techniques. As far as $\mathbb{E}[v]$ is concerned, notice that, since $x(t) \to v^*x(0)\mathbf{1}$ almost surely, it follows that

$$\mathbb{E}[x(t)] = \mathbb{E}[P(0)]^t x(0) \to \mathbb{E}[v]^*x(0)\mathbf{1}.$$

Since $\mathbb{E}[P(0)]$ is an aperiodic stochastic matrix, it follows that $\mathbb{E}[v]$ coincides with the the dominant left eigenvector of $\mathbb{E}[P(0)]$ and thus it is computable using standard techniques.

We now start to analyze the Cayley setting for which more precise results can be obtained. First, it can be checked that, for both the biased and the balanced compensation methods, the matrix $\overline{P}$ is Cayley. As a consequence, in this case we have $\mathbb{E}[v] = N^{-1}\mathbf{1}$ and hence

$$\mathbb{E}[(v - N^{-1}\mathbf{1})(v - N^{-1}\mathbf{1})^*] = \mathbb{E}[vv^*] - N^{-2}\mathbf{1}\mathbf{1}^*.$$

What remains to be computed is the normalized dominant eigenvector of $\mathcal{L}$ or, equivalently, the normalized dominant eigenvector of the matrix $A + N^{-1}B\mathbf{1}^*$ introduced above, where the matrix A is diagonal such that $A_{ii} = A(\chi_i)$ and B is a column vector such that $B_i = B(\chi_i)$. The quantities $A(\chi)$ and $B(\chi)$ have been defined in (17) and (18) for the biased case, and in (26) and (27) for the balanced case. Notice that, in both cases, we have that

$$\begin{aligned} A(\chi_0) = 1\,,\ 0 \le A(\chi_i) < 1 \ \ \forall i = 1, \dots, N-1,\\ B(\chi_0) = 0\,,\ B(\chi_i) \ge 0 \ \ \forall i = 1, \dots, N-1\,. \end{aligned}$$

We have the following result.

LEMMA 6.1. *The vector $w \in \mathbb{R}^N$ with components*

$$w_0 = 1\,,\ \ w_h = \frac{N^{-1}}{1 - N^{-1}\sum\limits_{i=1}^{N-1}\frac{B_i}{1-A_{ii}}}\frac{B_h}{1-A_{hh}} \ \ \forall h = 1, \dots, N-1$$

is an eigenvector of $(A + N^{-1}B\mathbf{1}^*)$ relative to the eigenvalue 1*.*

Proof. Notice first that, since $A + N^{-1}B\mathbf{1}^*$ is a nonnegative matrix, there exists an eigenvector $w \in \mathbb{R}^N$ of $A + N^{-1}B\mathbf{1}^*$ relative to the eigenvalue 1 with nonnegative entries. By Proposition 5.2 we can argue that the first component w_0 of w must be positive. Notice now that the relation $(A+N^{-1}B\mathbf{1}^*)w = w$ is equivalent to the $N-1$ linear conditions

$$(1 - A_{hh})w_h = N^{-1}B_h\mathbf{1}^*w\,,\ h = 1, \dots, N-1\,. \tag{33}$$

Since, as noticed above, $A_{hh} < 1$, we have that

$$w = \left(w_0, \lambda\frac{B_1}{1-A_{11}}, \dots, \lambda\frac{B_{N-1}}{1-A_{N-1\,N-1}}\right)^*, \tag{34}$$

where $\lambda = N^{-1}\mathbf{1}^*w$. This implies that

$$w_0 + \lambda\sum_{i=1}^{N-1}\frac{B_i}{1-A_{ii}} = N\lambda,$$

which is equivalent to

$$\left(1 - N^{-1}\sum_{i=1}^{N-1}\frac{B_i}{1-A_{ii}}\right)\lambda = N^{-1}w_0.$$

Finally, notice that, since $w_0 > 0$,

$$1 > \frac{N^{-1}\sum_{i=1}^{N-1}w_i}{\lambda} = N^{-1}\sum_{i=1}^{N-1}\frac{B_i}{1-A_{ii}},$$

which implies that $1 - N^{-1}\sum_{i=1}^{N-1}\frac{B_i}{1-A_{ii}} < 1$ and so, by taking $w_0 = 1$, we obtain the thesis. ◻

Notice that, if we go back to the matrix form, the corresponding eigenmatrix of $\mathcal{L}$ is given by

$$W = N^{-1}\sum_{i=0}^{N-1}w_i\chi_i\chi_i^*\,.$$

To find the right normalization constant, notice that $\mathbf{1}^*W\mathbf{1} = N$. This implies that

$$\mathbb{E}[vv^*] = N^{-2}\sum_{i=0}^{N-1} w_i\chi_i\chi_i^*\,. \tag{35}$$

Notice that, since $\mathbb{E}[vv^*]$ is positive semidefinite, surely all $w_i \geq 0$. We can now state the following proposition.

Proposition 6.2. *Assume P to be a Cayley matrix. Then, for both the biased and the balanced compensation methods, we have that*

$$D = \frac{N^{-2}}{1 - N^{-1}\sum\limits_{i=1}^{N-1}\frac{B_i}{1-A_{ii}}}\max_{h=1,\dots,N-1}\left\{\frac{B_h}{1-A_{hh}}\right\}.$$

Proof. Notice that, from (32), we obtain

$$\begin{aligned}&\mathbb{E}[(v - N^{-1}\mathbf{1})(v - N^{-1}\mathbf{1})^*]\\ &= N^{-2}\mathbf{1}\mathbf{1}^* + N^{-2}\sum_{i=1}^{N-1} w_i\chi_i\chi_i^* - 2N^{-2}\mathbf{1}\mathbf{1}^* + N^{-2}\mathbf{1}\mathbf{1}^* = N^{-2}\sum_{i=1}^{N-1} w_i\chi_i\chi_i^*\,.\end{aligned}$$

Notice now that

$$D = \max\{N^{-1}w_i \mid i = 1,\dots,N-1\}\,.$$

This proves the result. $\square$

Let us make explicit computations in the examples considered above.

Example 9. Consider the matrix $P = (1-k)I + \frac{k}{N}\mathbf{1}\mathbf{1}^*$ introduced in Examples 4 and 7. In the biased compensation method, using computation (21) and (22), we obtain that for any $h \neq 0$,

$$\frac{B_h}{1-A_{hh}} = \frac{2p(1-p)k^2}{N}\frac{1}{1-(1-p+p(1-k))^2} = N^{-1}\frac{2(1-p)k}{2-pk}\,.$$

Hence,

$$D = \frac{N^{-2}}{1 - N^{-2}(N-1)\frac{2(1-p)k}{2-pk}}N^{-1}\frac{2(1-p)k}{2-pk} = N^{-3}\frac{2(1-p)k}{2-pk-2N^{-2}(N-1)(1-p)k}\,.$$

As expected for $p \to 1$ we have that $D \to 0$. More interestingly, note also that for $N \to +\infty$, we have that $D \to 0$ as N^{-3}.

Consider now the balanced case. We limit the analysis to the optimal case in which we let $k = 1$. Using computation (31), we obtain that for any $h \neq 0$,

$$\frac{B_h}{1-A_{hh}} = \frac{(1-N\beta)N\beta}{1-(1-N\beta)^2} = \frac{1-N\beta}{2-N\beta}\,.$$

Hence,

$$D = \frac{N^{-2}}{1 - N^{-1}(N-1)\frac{1-N\beta}{2-N\beta}}\frac{1-N\beta}{2-N\beta} = N^{-2}\frac{1-N\beta}{1+N^{-1}(1-N\beta)}\,.$$

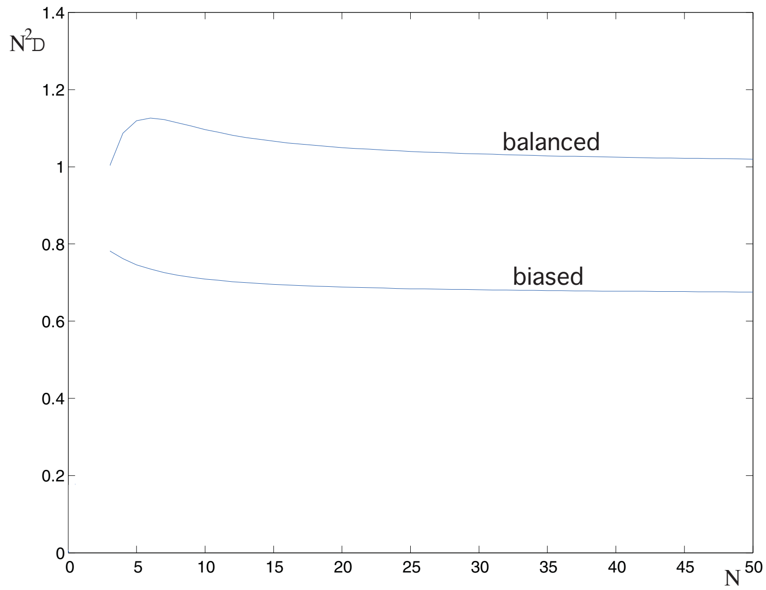

FIG. 8. *The graphs of N^3D as a function of N in the biased and balanced cases, both assuming that $k = 1$ and $p = 1/2$ as in Example* 9.

Even in this case, for $p \to 1$ we have that $D \to 0$ since it is easy to see that $N\beta \to 1$. Also the convergence to 0 for $N \to \infty$ is maintained. In Figure 8 we plot the graphs showing N^3D as a function of N in the biased and balanced cases, both assuming that $k = 1$ and $p = 1/2$. In both cases we notice that D converges to zero as fast as N^{-3}, and that the biased compensation method outperforms the balanced compensation method.

Example 10. Now consider Example 5, where the group is $\mathbb{Z}_N$ and $S = \{0, 1\}$. As we already noticed, for this example the biased and the balanced methods coincide. From the computations of the matrix A and of the vector B we obtain

$$\frac{B_h}{1 - A_{hh}} = \frac{2p(1-p)k^2\left(1 - \cos\left(\frac{2\pi}{N}i\right)\right)}{2pk(1-pk)\left(1 - \cos\left(\frac{2\pi}{N}i\right)\right)} = \frac{(1-p)k}{1-pk}\,.$$

Hence,

$$D = \frac{N^{-1}}{1 - \frac{(1-p)k}{1-pk}} \frac{(1-p)k}{1-pk} = N^{-2}\frac{k(1-p)}{1 - k + N^{-1}(1-p)k}\,.$$

Also in this case, for $p \to 1$, or for $N \to +\infty$, we have that $D \to 0$. Notice that the speed of convergence to 0 with respect to N is lower than in the complete case.

In order to have a clearer insight into the behavior of D we make some further estimations by analyzing the two cases separately.

Let us start with the biased case. Notice that we have

$$\frac{B_h}{1 - A_{hh}} = \frac{B(\chi_h)}{1 - A(\chi_h)} = \frac{2p(1-p)[\hat{\pi}_{\tilde{P}}(\chi_0) - \Re[\hat{\pi}_{\tilde{P}}(\chi_h)]]}{1 - |1 - p + p\hat{\pi}_P(\chi_h)|^2}\,.$$

Assume that $\pi_P(j) = k_j$, and notice that we have

$$\hat{\pi}_P(\chi_h) = \sum_j k_j \chi_h(-j)\,,\ \hat{\pi}_{\tilde{P}}(\chi_h) = \sum_j k_j^2 \chi_h(-j)\,.$$

Hence,

$$\hat{\pi}_{\tilde{P}}(\chi_0) - \Re[\hat{\pi}_{\tilde{P}}(\chi_h)] = \sum_j k_j(1 - \Re[\chi_h(-j)])\,,$$

while

$$\begin{aligned} 1 - |1 - p + p\hat{\pi}_P(\chi_h)|^2 &= -p^2 + 2p - p^2|\hat{\pi}_P(\chi_h)|^2 - 2p(1-p)\sum_j k_j \Re\chi_h(j) \\ &= 2p(1-p)\sum_j k_j(1 - \Re[\chi_h(j)]) + p^2(1 - |\hat{\pi}_P(\chi_h)|^2)\,. \end{aligned}$$

We have thus obtained that

$$\frac{B_h}{1 - A_{hh}} = \frac{2(1-p)\sum_j k_j^2(1 - \Re[\chi_h(-j)])}{2(1-p)\sum_j k_j(1 - \Re[\chi_h(-j)]) + p(1 - |\hat{\pi}_P(\chi_h)|^2)}\,. \tag{36}$$

This explicit expression allows us to estimate D. We have the following result.

PROPOSITION 6.3. *Consider a Cayley matrix P and let $\pi_P(j)$ be its generator. Let $M = \max\{\pi_P(j) \,|\, j = 1, \dots, N-1\}$. Then*

$$D \leq N^{-2}\frac{M}{1-M}\,.$$

Proof. From (36) we can argue that

$$\frac{B_h}{1 - A_{hh}} \leq \frac{\hat{\pi}_{\tilde{P}}(\chi_0) - \Re[\hat{\pi}_{\tilde{P}}(\chi_h)]}{1 - \Re[\hat{\pi}_P(\chi_h)]}.$$

Assume that $\pi_P(j) = k_j$, and notice that we have

$$\hat{\pi}_P(\chi_h) = \sum_j k_j \chi_h(-j)\,,\ \hat{\pi}_{\tilde{P}}(\chi_h) = \sum_j k_j^2 \chi_h(-j)\,.$$

Hence,

$$\frac{B_h}{1 - A_{hh}} \leq \frac{\sum_j k_j^2(1 - \Re[\chi_h(j)])}{\sum_j k_j(1 - \Re[\chi_h(j)])} \leq M\,.$$

The thesis now simply follows from Proposition 6.2. ∎

The key point of Proposition 6.3 is that if we have a sequence of consensus strategies indexed by N, for which M is bounded away from 1, then D will converge to 0 at least as fast as N^{-2}. Notice that this is in agreement with the two examples considered above.

We now proceed to analyze the balanced case. We can prove the following result.

PROPOSITION 6.4. *Denote*

$$M = \pi_{\bar{P}(I)}(0);$$

then

$$D \leq N^{-2}\frac{1}{1-M}\,. \tag{37}$$

Proof. In the balanced case A_{hh} and B_h are defined in (26) and (27). We obtain

$$\frac{B_h}{1 - A_{hh}} = \frac{\hat{\pi}_{\bar{\rho}(I)}(\chi_h) - \left|\hat{\pi}_{\bar{\beta}}(\chi_h)\right|^2}{1 - \left|\hat{\pi}_{\bar{\beta}}(\chi_h)\right|^2}. \tag{38}$$

It follows from (38), using the inequality $0 \leq \hat{\pi}_{\bar{\rho}(I)}(\chi_h) \leq 1$, that

$$\frac{B_h}{1 - A_{hh}} \leq \hat{\pi}_{\bar{\rho}(I)}(\chi_h) \leq 1\,. \tag{39}$$

Notice now that

$$N^{-1} \sum_{j=1}^{N-1} \frac{B_j}{1 - A_{jj}} \leq N^{-1} \sum_{j=1}^{N-1} \hat{\pi}_{\bar{\rho}(I)}(\chi_j) \leq N^{-1} \sum_{j=0}^{N-1} \hat{\pi}_{\bar{\rho}(I)}(\chi_j) = \pi_{\bar{\rho}(I)}(0)\,. \tag{40}$$

The thesis now follows from Proposition 6.2 and estimations (39) and (40). ∎

Notice that M is strictly smaller than 1. It clearly depends on the matrix P but also, as opposed to the biased case, on the probability p. Let us analyze a simple case in more detail. Assume that the matrix P is defined as in (28). In this case, using (30), we have that

$$\begin{aligned} M &= \pi_{\bar{\rho}(I)}(0) = N^{-1}\text{trace } \bar{\rho}(I) \\ &= N^{-1}\text{trace } (\rho H^* H + (\xi - \rho)\text{diag } (H^* H) + \gamma I + \delta(H + H^*)) \\ &= (n-1)\xi + \gamma. \end{aligned}$$

Therefore M depends on n and p and k, but it does not depend on N. It thus follows that δ converges to 0 as fast as N^{-2} in the biased case.

7. Conclusions. In this paper we proposed some tools which allow us to evaluate the performance degradation due to failing transmission links in the average consensus algorithm. Though the tools proposed here seem to be very effective for the evaluation of the effect of packet drop in the data transmission between the agents in a consensus seeking problem, many problems are still to be investigated, such as the following:

1. The analysis of convergence has been carried out in a mean square sense. Concentration results can be obtained in certain cases (see [6]) and it would be important to study them in the context of packet drop models.
2. Many problems are still open in the general (non-Cayley) case, such as the evaluation of the mean distance of the limit from the average as a function of the number N of agents.
3. The analysis is still quite intricate and it is difficult to use in design. We expect that some interesting simplifications could occur when N tends to infinity. It is important to determine whether this is really the case and to exploit these simplifications in the design process.
4. The average consensus algorithm we considered is somehow memoryless. We expect that algorithms with memory in principle could yield better performance (consider for instance an algorithm which, when data is lost, can substitute it with its past version). It is important to understand whether adding memory will improve the performance or not.

REFERENCES

[1] L. Babai, *Spectra of Cayley graphs*, J. Combin. Theory Ser. B, 27 (1979), pp. 180–189.
[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.
[3] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in the average consensus problem*, Automatica, 44 (2008), pp. 671–684.
[4] R. Cogburn, *On products of random stochastic matrices*, in Random Matrices and Their Applications (Brunswick, Maine, 1984), Contemp. Math. 50, AMS, Providence, RI, 1986, pp. 199–213.
[5] R. Diestel, *Graph Theory*, Grad. Texts in Math. 173, Springer-Verlag, Heidelberg, 2005.
[6] F. Fagnani and S. Zampieri, *Randomized consensus algorithms over large scale networks*, IEEE J. Sel. Areas Commun., 26 (2008), pp. 634–649.
[7] F. R. Gantmacher, *The Theory of Matrices*, Chelsea, New York, 1959.
[8] C. Godsil and G. Royle, *Algebraic Graph Theory*, Grad. Texts in Math. 207, Springer-Verlag, New York, 2001.
[9] Y. Hatano and M. Mesbahi, *Agreement over random networks*, IEEE Trans. Automat. Control, 50 (2005), pp. 1867–1872.
[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.
[11] A. Jadbabaie, J. Lin, and A. S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1001.
[12] Z. Lin, B. Francis, and M. Maggiore, *State agreement for continuous-time coupled nonlinear systems*, SIAM J. Control Optim., 46 (2007), pp. 288–307.
[13] L. Moreau, *Stability of multiagent systems with time-dependent communication links*, IEEE Trans. Automat. Control, 50 (2005), pp. 169–182.
[14] R. Olfati-Saber, *Distributed Kalman filter with embedded consensus filters*, in Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference, 2005, pp. 8179–8184.
[15] R. Olfati-Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Control, 49 (2004), pp. 1520–1533.
[16] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proceedings of the IEEE, 95 (2007), pp. 215–233.
[17] S. Patterson, B. Bamieh, and A. El Abbadi, *Distributed average consensus with stochastic communication failures*, in Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 4215–4220.
[18] W. Ren and R. W. Beard, *Consensus seeking in multiagent systems under dynamically changing interaction topologies*, IEEE Trans. Automat. Control, 50 (2005), pp. 655–661.
[19] A. Terras, *Fourier analysis on finite groups and applications*, London Math. Soc. Stud. Texts 43, Cambridge University Press, Cambridge, UK, 1999.
[20] J. Tsitsiklis, *Problems in Decentralized Decision Making and Computation*, Ph.D. thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1984.