Broadcast gossip averaging: interference and unbiasedness in large Abelian Cayley networks

Fabio Fagnani Paolo Frasca

***Abstract*—In this paper we study two related iterative randomized algorithms for distributed computation of averages. The first algorithm is the Broadcast Gossip Algorithm, in which at each iteration one randomly selected node broadcasts its own state to its neighbors. The second algorithm is a novel variation of the former, in which at each iteration every node is allowed to broadcast: hence this algorithm, which we call Collision Broadcast Gossip Algorithm (CBGA), is affected by interference among messages. The performance of both algorithms is evaluated in terms of rate of convergence and asymptotical error: focusing on large Abelian Cayley networks, we highlight the role of topology and of design parameters. We show that on fully-connected graphs the rate of convergence is bounded away from one, whereas the asymptotical error is bounded away from zero. On the contrary, on sparse graphs the rate of convergence goes to one and the asymptotical error goes to zero, as the size of the network grows larger. Our results also show that the performance of the CBGA is close to the performance of the BGA: this indicates the robustness of broadcast gossip algorithms to interferences.**

I. Introduction

When it comes to perform control and monitoring tasks through networked systems, a crucial role has to be played by algorithms for distributed estimation, that is algorithms to collectively compute aggregate information from locally available data. Among these problems, a prototypical one is the distributed computation of averages, also known as the average consensus problem. In the average consensus problem each node of a network is given a real number, and the goal for the nodes is to iteratively converge to a good estimate of the average of these initial values, by repeatedly communicating and updating their states. Recently, an increasing interest has been devoted within the control and signal processing communities to *randomized* algorithms able to solve the average consensus problem, because randomized algorithms may offer better performance and robustness with respect to their deterministic counterparts. As well, randomized algorithms may require less or no synchronization among the nodes, a property which is often difficult to ensure in the applications. These facts are especially relevant when communication is obtained through a wireless network. For these reasons, in this paper we study the performance of two notable algorithms based on random broadcast communication.

Related works

In latest years, the interest for wireless networks has lead the researchers to consider averaging algorithms based on *broadcast* communication over networks. The paper [3] is devoted to study the so-called Broadcast Gossip Algorithm (BGA): at each time step one node, randomly selected from a uniform distribution over the nodes, broadcasts its current value to its neighbors. Each of these neighbors, in turn, updates its value to a convex combination of its previous value and the received one. In [3], the authors prove that the BGA converges almost surely to a consensus value, which is, in expectation, the average of the initial node values. They also show that the mixing parameter of the algorithm can be suitably used to trade-off between convergence rate and accuracy of the computation in a mean squared error sense. More recently, the paper [2] considers a related communication model, leading to a Probabilistic Broadcast Gossip Algorithm (PBGA): the broadcasted values are received or not with a probability which depends on the transmitter and receiver nodes, or equivalently on the graph edge. In [2] it is shown that also the PBGA converges almost surely to a consensus value whose expectation is the average of initial node values. These results suggest the robustness of BGA to independent random communication failures. A few other randomized "gossip" algorithms have been proposed and studied in the literature, including [5], [4], [11], [18], [9], [14]; see [10] for a recent survey, and [1] for general theoretical results.

The BGA is distributed and requires minimal synchronization: indeed, it is observed in [3] that the BGA communication model is equivalent, up to a suitable scaling of time, to assuming that each node broadcasts at time instants selected by a private Poisson process. Nevertheless, this equivalence is no longer true if broadcasting takes a finite duration of time. If this happens, a node can be with non-zero probability the target of more than one simultaneous communication, and destructive interference may occur. This issue is mostly relevant in wireless networks, which have to share their communication medium. Hence, the practical applicability of this algorithm in a distributed system resides either on the validity of the assumption that communications are instantaneous, or on the possibility to incorporate some nontrivial collision detection scheme. The role of interference, message collisions and packet losses in consensus problems, and the effectiveness of countermeasures, has already been investigated in [8] for consensus problems in a finite state space, and for real-valued consensus in a few papers, including [19], [22], [15]. However, to our knowledge there is no contribution yet about the role of interference in broadcast gossip algorithms.

Fabio Fagnani and Paolo Frasca are with the Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. {fabio.fagnani,paolo.frasca}@polito.it

Contribution

In this paper, we study averaging algorithms based on broadcasting communication. In order to investigate the effect of interference in the BGA, we propose and study a novel averaging algorithm, which we call Collision Broadcast Gossip Algorithm (CBGA). In this algorithm, we allow more than one node to broadcast at the same time, possibly causing the destructive interference of attempted communications, and message loss. Instead, when a node properly receives a message, it updates its state as in the BGA. It is of note that this communication model can not be understood as a special case of the PBGA in [2], because in the CBGA the events of ineffective communication are not independent among edges. Our results show that, just like the BGA, also the CBGA algorithm converges almost surely to a random variable, whose expectation is the average of initial values.

Besides convergence, we are then interested in the speed and the accuracy of the algorithms. In particular, we call *bias* the variance of the limit value, and we want to know whether the algorithms are *asymptotically unbiased*, that is whether the accuracy can be arbitrarily improved by taking a *larger* network. This analysis question is motivated because, although in most applications the size of the network is fixed during operation, network design typically requires to choose the size and topology of the network. In classical (centralized) estimation, average is a natural estimator, and we know that increasing the number of samples is a way to improve the accuracy of the estimate. In a network, when each sample is available to only one node and the average is computed by a distributed procedure, we possibly introduce an error which depends on the network and in particular on its size. Then, if such error does not decrease to zero when the size grows, we are wasting the advantage of using more samples – see [17] for a related discussion.

Our results are based on the technical assumption that the network topology has certain structural symmetries, namely that the network is the Cayley graph of an Abelian group. Under this assumption we prove that on sparse graphs with bounded degree, both the BGA and the CBGA are asymptotically unbiased. Instead, both algorithms are biased on complete graphs. As a byproduct of our analysis, we prove a decomposition formula for the mean square evolution of the BGA and CBGA, which shows that the performance of the two algorithms is close on large networks: this suggests that the BGA is robust to non-independent communication failures due to interferences.

Paper structure

After presenting in Section II the averaging problem and the algorithms under consideration, we develop our analysis tools in Section III, including the mean square analysis theory and the Abelian Cayley graph model. Later, Sections IV and V are devoted to analyze the Broadcast Gossip Algorithm and the novel Collision Broadcast Gossip Algorithm, respectively. Some concluding remarks are presented in Section VI.

Notations

Given a set $\mathcal{V}$ of finite cardinality $|\mathcal{V}| = N$, we define a graph on this set as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ (we exclude the presence of self-loops, namely edges of type (u,u)). Given $u, v \in \mathcal{V}$, if $(v,u) \in \mathcal{E}$, we shall say that v is an in-neighbor of u, and conversely u is an out-neighbor of v. We will denote by $\mathcal{N}_u^+$ and $\mathcal{N}_u^-$, the set of, respectively, the out-neighbors and the in-neighbors of u. Also, $d_u^+ = |\mathcal{N}_u^+|$ and $d_u^- = |\mathcal{N}_u^-|$ are said to be the out-degree and the in-degree of node u, respectively. A graph whose nodes all have in-degree k is said to be k-regular. A graph is said to be (strongly) connected if for any pair of nodes (u,v), one can find a path, that is an ordered list of edges, from u to v. A graph is said to be symmetric if $(u,v) \in \mathcal{E}$ implies $(v,u) \in \mathcal{E}$. In a symmetric graph, being the neighborhood relation symmetrical, there is no distinction between in- and out-neighbors and we will drop, consequently, the index $+$ or $-$. We let $\mathbf{1}$ be the $N-$vector whose entries are all 1, I be the $N \times N$ identity matrix, and $\Omega := I - N^{-1}\mathbf{1}\mathbf{1}^*$. Given a N-vector a, we denote by $\mathrm{diag}(a)$ the diagonal matrix whose diagonal is equal to a. The adjacency matrix of the graph $\mathcal{G}$, denoted by $A_{\mathcal{G}}$, is the matrix in $\{0,1\}^{\mathcal{V}\times\mathcal{V}}$ such that $A_{\mathcal{G}uv} = 1$ if and only if $(v,u) \in \mathcal{E}$. We also define the out-degree matrix as $D_{\mathcal{G}}^+ := \mathrm{diag}\left(A_{\mathcal{G}}^* \mathbf{1}\right)$, the in-degree matrix as $D_{\mathcal{G}}^- := \mathrm{diag}\left(A_{\mathcal{G}} \mathbf{1}\right)$, and the Laplacian matrix as $L_{\mathcal{G}} = D_{\mathcal{G}}^- - A_{\mathcal{G}}$. Given a matrix $M \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$, we define the graph $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}_M)$ by putting $(v,w) \in \mathcal{E}_M$ iff $v \neq w$ and $M_{wv} \neq 0$. A matrix M is said to be *adapted* to the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $\mathcal{G}_M \subseteq \mathcal{G}$, that is if $\mathcal{E}_M \subseteq \mathcal{E}$. When it comes to compare two sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, we shall write that $a_n = o(b_n)$ if $\limsup_n \frac{|a_n|}{|b_n|} = 0$, that $a_n = O(b_n)$ if $\limsup_n \frac{|a_n|}{|b_n|} < +\infty$, and that $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. Given a linear operator $\mathcal{L}$ from a vector space to itself, for instance represented by a square matrix, we denote by $\mathrm{sr}(\mathcal{L})$ its spectral radius, that is the modulus of its largest in magnitude eigenvalue. Whenever $\mathrm{sr}(\mathcal{L}) = 1$, we shall define as $\mathrm{esr}(\mathcal{L})$ the modulus of the second largest in magnitude eigenvalue.

II. Broadcast gossip averaging algorithms

In this section we present the averaging problem and the algorithms we are going to study. Let us be given a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and one real value y_v for each node $v \in \mathcal{V}$. Then, the averaging problem consists in approximating the average $\frac{1}{N}\sum_{v\in\mathcal{V}} y_v$, with the constraint that at every time step each node v can communicate to its out-neighbors only. Typically, this is solved by linear iterative algorithms such that $x(0) = y$, and for all $t \in \mathbb{Z}_{\geq 0}$, $x(t+1) = P(t)x(t)$, where the matrix $P(t) \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is adapted to $\mathcal{G}$.

We start by recalling the Broadcast Gossip Algorithm as presented in [14], [3]. In this algorithm, at every time step one node, randomly selected from a uniform distribution over the nodes, broadcasts its current value to its neighbors. Its neighbors, in turn, update their values to a convex combination of their previous values and the received ones. More formally, we can write the algorithm as follows. Note that the only design parameter is the weight given to the received value in the convex update, which we call *mixing parameter*.

Broadcast Gossip Algorithm – Parameters: $q \in (0,1)$

For all $t \in \mathbb{Z}_{\geq 0}$,

Sample a node v from a uniform distribution over $\mathcal{V}$
for $u \in \mathcal{V}$ **do**
  if $u \in \mathcal{N}_v^+$ **then**
    $x_u(t+1) = (1-q)x_u(t) + qx_v(t)$
  else
    $x_u(t+1) = x_u(t)$
  end if
end for

This algorithm can be written in the form of iterated matrix multiplication. Let v be the broadcasting node which has been sampled at time t. Then, $x(t+1) = P(t)x(t)$, where

$$P(t) = I + q \sum_{u \in \mathcal{N}_v^+} (e_u e_v^* - e_u e_u^*), \tag{1}$$

and e_i is the i-th element of the canonical basis of $\mathbb{R}^{\mathcal{V}}$. Clearly, at each time t, the matrix $P(t)$ is the realization of a uniformly distributed random variable, depending on the stochastic choice of the broadcasting node. The Broadcast Gossip algorithm has been thoroughly studied in [3], under the assumption that the communication graph is symmetric. In this paper we shall focus on one specific analysis question, regarding the accuracy of the algorithm. We know that the expectation of the convergence value is the initial average: which is its variance? In particular, how does this variance depend on the size and topology of the graph?

As we noted in the introduction, the practical interest of the BGA algorithm depends on the assumption that the transmissions are instantaneous, and reliable. In an effort towards more realistic communication models, we propose a modification of the Broadcast Gossip Algorithm, which has the feature of dealing with the issue of finite-length transmissions, and consequent message losses due to collisions. At every time step, each node is allowed to wake up, independently with probability p, and broadcast its current state to all its out-neighbors. It is clear that some nodes can be the target of more than one message: in this case, we assume that a destructive collision occurs, and no message is actually received by these nodes. As well, interference prevents the broadcasting nodes from hearing any others[1]. If a node $u \in \mathcal{V}$ is able to receive a message from node v, it updates its state to a convex combination with the received value, similarly to the standard BGA. More formally, the algorithm is as follows.

Collision Broadcast Gossip Algorithm – Parameters: $q \in (0,1)$, $p \in (0,1)$

For all $t \in \mathbb{Z}_{\geq 0}$,

let Act be the random set defined by: for every $v \in \mathcal{V}$, $\mathbb{P}[v \in \mathrm{Act}] = p$
let $\mathrm{Rec} := \{u \in \mathcal{V} \,:\, |\mathcal{N}_u^- \cap \mathrm{Act}| = 1, u \notin \mathrm{Act}\}$
for all $u \in \mathrm{Rec}$, let $\sigma(u)$ be the only $v \in \mathcal{V}$ such that $v \in \mathrm{Act} \cap \mathcal{N}_u^-$
for $u \in \mathcal{V}$ **do**
  if $u \in \mathrm{Rec}$ **then**
    $x_u(t+1) = (1-q)x_u(t) + qx_{\sigma(u)}(t)$
  else
    $x_u(t+1) = x_u(t)$
  end if
end for

[1]This *half-duplex* constraint is assumed throughout the paper: however, dropping it would imply minimal changes in the analysis.

Also the latter algorithm can be written as matrix multiplication, defining

$$P(t) = I + q \sum_{(v,u) \in \mathrm{Act} \times \mathrm{Rec}} (e_u e_v^* - e_u e_u^*). \tag{2}$$

Both algorithms can actually be rewritten in the following graph-theoretic way. Let $\mathcal{G}(t)$ be the subgraph of $\mathcal{G}$ depicting the communications taking place at a certain instant t: the pair (u,v) is an edge in $\mathcal{G}(t)$ iff v successfully receives a message from u at time t. Denote by $A(t)$, $D(t)$, $L(t)$ the adjacency, degree and Laplacian matrices, respectively, of $\mathcal{G}(t)$. Clearly, for both algorithms:

$$P(t) = I - qL(t). \tag{3}$$

Several questions are natural for the collision-prone CBGA algorithm, in comparison with the BGA. Does the algorithm converge? How fast? Does it preserve the average of states? If not, how far it goes? Is performance poorer because of interferences? We are going to answer the analysis questions we have posed, via a mean square analysis of the algorithm. Our interest will be mostly devoted to the properties of algorithms for large networks. To this goal, we shall often assume to have a sequence of graphs $\mathcal{G}_N$ of increasing order $N \in \mathbb{N}$, and we shall consider, for each $N \in \mathbb{N}$, the corresponding matrix $P(t)$, which depends on $\mathcal{G}$ and then on N. Thus we will focus on studying the asymptotical properties of the algorithms as N goes to infinity.

III. Mathematical models and techniques

In this section we lay down some mathematical tools that can be used to analyze gossip and other randomized algorithms. Namely, in Section III-A we review the Mean Square Analysis (MSA) in [14], which is going to be applied to the BGA and CBGA algorithms in Sections IV and V, respectively. Later, in Section III-B we introduce Abelian Cayley graphs and their properties.

A. Mean square analysis

Motivated by the interpretation of the broadcast algorithms as iterated multiplications by random matrices, given in Equations (1) and (2), in this subsection we shall recall from [14] some definitions and results for the analysis of a generic algorithm, in which the vector of states $x(t) \in \mathbb{R}^{\mathcal{V}}$ evolves in time following an iterate $x(t+1) = P(t)x(t)$, where $\{P(t)\}_{t \in \mathbb{Z}_{\geq 0}}$ is a sequence of i.i.d. stochastic-matrix-valued random variables. Consequently, $x(t)$ is a stochastic process. In this context, the sequence $P(t)$ is said to achieve *probabilistic consensus* if for any $x(0) \in \mathbb{R}^{\mathcal{V}}$, it exists a scalar random variable α such that almost surely $\lim_{t \to \infty} x(t) = \alpha \mathbf{1}$.

Assuming that the algorithm achieves probabilistic consensus, we describe its speed of convergence as follows. We define the current average $x_{\mathsf{ave}}(t) := \frac{1}{N}\sum_{v\in\mathcal{V}} x_v(t)$, the disagreement $d(t) := N^{-1}\|x(t) - x_{\mathsf{ave}}(t)\mathbf{1}\|_2^2$, and the rate of convergence as

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

If we consider the (linear) operator $\mathcal{L} : \mathbb{R}^{\mathcal{V}\times\mathcal{V}} \to \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ such that

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

we can notice that $\mathbb{E}[d(t)] = \mathbb{E}[x^*(t)\Omega x(t)] = x^*(0)\Delta(t)x(0)$, with $\Delta(t) = \mathcal{L}^t(\Omega)$ defined by applying t times the operator $\mathcal{L}$. Then, R is the spectral radius of $\mathcal{L}$, and it has been proved in [14, Proposition 4.4] that

$$\mathrm{esr}(\overline{P})^2 \le R \le \mathrm{sr}(\mathcal{L}(\Omega)), \tag{4}$$

where $\bar{P} = \mathbb{E}[P(t)]$. It is clear that, if all $P(t)$ matrices are doubly stochastic, $x(t)$ converges to the initial average of states $x_{\mathsf{ave}}(0)$. If they are not, as in the cases we are studying in this paper, it is worth asking how far is the convergence value from the initial average. To study this bias in the estimation of the average, we let $\beta(t) = |x_{\mathsf{ave}}(t) - x_{\mathsf{ave}}(0)|^2$, and we define a matrix B such that

$$\lim_{t\to\infty} \mathbb{E}[\beta(t)] = x(0)^* B x(0). \tag{5}$$

Let $Q(t) = P(t-1)\cdots P(0)$. Then, there exists a random variable ρ, taking values in $\mathbb{R}^{\mathcal{V}}$, such that $\lim_{t\to\infty} Q(t) = \mathbf{1}\rho^*$ and $\alpha = \rho^* x(0)$. This implies that

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

where $\mathbb{E}[\rho]$ and $\mathbb{E}[\rho\rho^*]$ are the eigenvectors relative to 1 of $\bar{P}$ and $\mathcal{L}$, respectively. In particular, if $\bar{P}$ is doubly stochastic, then

$$B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbf{1}\mathbf{1}^* = \frac{1}{\mathbf{1}^*\Delta\mathbf{1}} \lim_{t\to\infty} \mathcal{L}^t(\Delta) - N^{-2}\mathbf{1}\mathbf{1}^*, \tag{6}$$

for any N-dimensional matrix Δ. Instead of computing B, it may be easier and significant to obtain results about some functional of B, for instance the spectral norm $\|B\|_2$ or the trace $\mathrm{tr}(B)$. The latter figure is of interest because, if we assume that the initial values $x_i(0)$ are i.i.d. random variables with zero mean and variance σ^2, then

$$\mathbb{E}[x(0)^* B x(0)] = \sigma^2\, \mathrm{tr}\,(B). \tag{7}$$

Motivated by our interest in the properties of the algorithms on large networks, and by Equation (7), we state the following definition.

Definition III.1 *Given a sequence of graphs $\mathcal{G}_N$, a randomized algorithm $P(t)$ is said to be* asymptotically unbiased *if* $\lim\limits_{N\to\infty} \mathrm{tr}(B) = 0.$

One would like to use the MSA results in [3] to obtain unbiasedness results for the BGA on some sequences of graphs. If we plug the formulas from [3, Lemma 2 and Lemma 4] into the bound presented in [3, Proposition 3] we obtain that

$$\mathrm{tr}B \le \left(1 - \frac{\lambda_1}{\lambda_{N-1}} \frac{1}{1 - \frac{1}{2}\frac{q}{N}\lambda_{N-1}}\right), \tag{8}$$

where $\{\lambda_i\}_{i=1}^{N-1}$ are the eigenvalues of L, with $0 = \lambda_0 < \lambda_1 \le \ldots \le (1-q)\lambda_{N-1} \le 2N$. By the latter inequality, the right-hand-side in (8) can be lower bounded by $1 - \frac{\lambda_1}{\lambda_{N-1}}$. Since on several graphs of applicative interest, including rings, bidimensional grids and random geometric graphs, the ratio $\frac{\lambda_1}{\lambda_{N-1}}$ does not go to one as the size grows, the bound in (8) can not be effectively used to prove asymptotical unbiasedness. In view of these remarks, and in order to prove unbiasedness results, in this paper we apply the mean square analysis to class of graphs with structural symmetries, the Abelian Cayley graphs, which we define in the next section.

B. *Abelian Cayley graphs*

A special family of graphs is that of Abelian Cayley graphs, which are graphs representing a group, as follows. Let G be an Abelian group, considered with the additive notation, and let S be a subset of G. Then, the *Abelian Cayley graph* generated by S in G is the graph $\mathcal{G}(G, S)$ having G as node set and $\mathcal{E} = \{(g,h) \in G\times G : \ h - g \in S\}$ as edge set. In words, two nodes -i.e. two group elements- are neighbors if their difference is in S. As well, a notion of Abelian Cayley *matrix* can be defined. Given a group G and a generating vector π of length $|G|$, we shall define the Cayley matrix generated by π as $\mathrm{cayl}(\pi)_{hg} = \pi_{h-g}$. Correspondingly, for a given Cayley matrix M, we shall denote by π^M the generating vector of the Cayley matrix M. Clearly, the adjacency matrices of G-Cayley graphs are G-Cayley matrices.

Cayley graphs have a long history in abstract mathematics, and have been recently used in control theoretical applications, for instance in [21], [17], to describe communication networks. Assuming Abelian Cayley topologies is motivated both by their algebraic structure, which allows a formal mathematical treatment –we refer the reader to [23], [6] for more details–, and also by their potential applications. Indeed, Abelian Cayley graphs are a simplified and idealized version of communications scenarios of practical interest. In particular, they capture the effects on performance of the strong constraint that, for many networks of interest, communication is local, not only in the sense of a little number of neighbors, but also with a bound on the geometric distance among connected nodes. This constraint is abstracted into the edge set definition given above. These constraints are especially relevant for wireless networks: indeed, Abelian Cayley graphs have been related, for instance in [5], [20], [7], to other models for wireless networks, as random geometric graphs or disk-graphs [16]. Abelian Cayley graphs encompass several important examples.

Example III.1 Let $\mathbb{Z}_n$ denote the cyclic group of integers modulo n.

1) The *complete* graph on N nodes, that is the graph where each node is directly connected with every other node, is $\mathcal{G}(\mathbb{Z}_N, \mathbb{Z}_N \setminus \{0\})$;

2) The *circulant* graphs (resp. matrices) are Abelian Cayley graphs (resp. matrices) on the group $\mathbb{Z}_N$; we denote the circulant matrix generated by π as $\mathrm{circ}(\pi)$. For instance, he *ring* graph is the circulant graph $\mathcal{G}(\mathbb{Z}_N, \{-1,1\})$; its adjacency matrix is $A = \mathrm{circ}([0,1,0,\dots,0,1])$ and its Laplacian is $L = \mathrm{circ}([2,-1,0,\dots,0,-1])$. For a ring, the eigenvalues of L are $\{2(1-\cos\left(\frac{2\pi}{N}l\right))\}_{l\in\mathbb{Z}_N}$, and in particular $\lambda_1 = \frac{4\pi^2}{N^2} + o\left(\frac{1}{N^3}\right)$ as $N\to+\infty$.
3) The square *grids* on a d-dimensional torus are $\mathcal{G}(\mathbb{Z}_n^d, \{e_i, -e_i\}_{i\in\{1,\dots,d\}})$, where e_i are elements of the canonical basis of $\mathbb{R}^d$. In particular, the n-*dimensional hypercube* graph is $\mathcal{G}(\mathbb{Z}_2^n, \{e_i\}_{i\in\{1,\dots,n\}})$.

Notice how all examples above are naturally forming a sequence of graphs indexed by the number of nodes N. Special cases for which we will be able to prove asymptotical unbiasedness in the following, are when the generating set S is finite and "kept fixed" as in the ring graph. Precisely we consider the following general example.

Example III.2 Start from an infinite lattice $\mathcal{V} = \mathbb{Z}^d$ and fix a finite $S \subseteq \mathbb{Z}^d \setminus \{0\}$ generating $\mathbb{Z}^d$ as a group. For every integer n, let $V_n = [-n,n]^d$ considered as the Abelian group $\mathbb{Z}_{2n+1}^d$ and let $\mathcal{G}^{(n)}$ be the Cayley Abelian graph generated by $S_n = S \cap [-n,n]^d$. Notice that all graphs $\mathcal{G}^{(n)}$ have the same generating set S for n sufficiently large, in particular they have the same degree. Rings and grids fit in this framework.

A G-Cayley structure for the communication graph $\mathcal{G}$ has important consequences on the mean square analysis of randomized consensus algorithms. In particular, it is easy to see that if C is a G-Cayley matrix, then also $\mathcal{L}(C)$ is G-Cayley. Then, for every $t\in\mathbb{Z}_{\geq 0}$, the matrices $\Delta(t) = \mathcal{L}^t(\Omega)$ are G-Cayley. Thus, the sequence $\Delta(t)$ can be equivalently seen as the sequence of the corresponding generating vectors $\pi(t) = e_0^*\Delta(t)$, where e_0 is the indicator vector corresponding to the group element 0. We shall refer to the vector $\pi(t)$ as the *MSA vector*. Since $\mathcal{L}$ is linear, the MSA vector evolution can be written as a matrix multiplication $\pi(t+1) = M\pi(t)$. Clearly, $R = \mathrm{esr}(M)$. Moreover, M is $*$-stochastic, and if we let π' to be the invariant vector of M, that is

$$\begin{cases}\pi' = M\pi' \\ \mathbf{1}^*\pi' = 1,\end{cases}$$

then, Equation (6), implies that

$$B = \frac{1}{N}\,\mathrm{cayl}\left(\pi' - \frac{1}{N}\mathbf{1}\right), \quad \mathrm{tr}\, B = \pi'_0 - \frac{1}{N} \tag{9}$$

where π'_0 is the component of π' corresponding to $0\in G$.

The following simple result will be useful later on.

Lemma III.3 *Let $\mathcal{G}$ be G-Cayley, and suppose that $P(t)_{uu} \geq \delta > 0$ almost surely. Then,*

$$\mathcal{G}_A \subseteq \mathcal{G}_{M^*} \subseteq \mathcal{G}_{A+A^*+A^*A}$$

Proof: A straightforward computation shows that

$$M_{uv} = \sum_k \mathbb{E}[P(t)_{k+v,u}P(t)_{k0}]$$

Hence, $M_{uv} > 0$ implies that there exists k such that $P(t)_{k+v,u} > 0$ and $P(t)_{k0}$. If $k=0$, this yields $A_{vu} > 0$. If $k+v=u$, then $A_{u-v,0} > 0$ or also $A_{uv} > 0$. Finally, if both cases above do not happen, then, $A_{k+v,u} > 0$ and $A_{k+v,v} > 0$ which yields $(A^*A)_{uv} > 0$. The second inclusion is thus proven. To prove the first one, notice that, by the assumption made, $M_{uv} \geq \delta\,\mathbb{E}[P(t)_{vu}]$. This completes the proof. ∎

Notice that in the BGA and CBGA examples we can always apply Lemma III.3 with $\delta = 1-q$.

IV. Broadcast without collisions

In this section we present a comprehensive analysis of the Broadcasting Gossip Algorithm, in terms of both rate of convergence and bias. The following result characterizes the convergence properties of the algorithm, extending [3, Lemmas 2 and 4] to directed networks. We omit a detailed proof, which can be obtained by direct computation.

Proposition IV.1 (Convergence of BGA algorithm) *Consider the BGA algorithm. Let $\mathcal{G}$ be any connected graph, and let A and L respectively denote its adjacency and Laplacian matrices. Then*

$$\bar{P} = I - qN^{-1}L \tag{10}$$

$$\begin{aligned}\mathcal{L}(\Omega) = {} & \Omega - q(1-q)N^{-1}(L+L^*) + qN^{-2}(L^*\mathbf{1}\mathbf{1}^* + \mathbf{1}\mathbf{1}^*L) \\ & - q^2N^{-2}(D^+ - A)(D^+ - A^*).\end{aligned} \tag{11}$$

In particular, the BGA algorithm achieves probabilistic consensus.

The next result provides explicit bounds on the convergence rate, assuming the communication graph to be symmetric.

Corollary IV.2 ([3], Lemma 4) *Under the assumptions of Proposition IV.1, if $\mathcal{G}$ is symmetric, then*

$$\begin{aligned}\bar{P} &= I - qN^{-1}L \\ \mathcal{L}(\Omega) &= \Omega - 2q(1-q)N^{-1}L - q^2N^{-2}L^2.\end{aligned}$$

In particular, the convergence rate can be estimated as

$$1 - \frac{2q}{N}\lambda_1 \leq R \leq 1 - \frac{2q(1-q)}{N}\lambda_1,$$

where λ_1 is the smallest positive eigenvalue of L.

For the rest of this section, we focus on *Abelian Cayley graphs*. The next result shows that the linear operator $\mathcal{L}$, which encodes the MSA evolution of the algorithm, can be decomposed into a "simple" operator, which is essentially the Laplacian of the communication graph, plus a perturbation, which is a sparse matrix if the graph is sparse.

Lemma IV.3 *Consider the BGA algorithm and let the communication graph $\mathcal{G}$ be an Abelian Cayley graph generated by $S\subset G$, with degree $d = |S|$. Then, the MSA vector π evolves as*

$$\pi(t+1) = (C+T)\pi(t), \tag{12}$$

where
$$C = I - \frac{q}{N}(L + L^*)$$
and T is a matrix such that $T = \frac{q^2}{N}\tilde{T}$ where $\tilde{T}$ does not depend neither on q nor explicitly on N but only on S, and is such that the number of non-zero rows is at most d^2 and non-zero columns is at most $d^2 - d + 1$.

Proof: Using the fact that all matrices are Abelian Cayley and the notation in (3), we obtain

$$\begin{aligned}\Delta(t+1) =& \mathbb{E}[P(t)^*\Delta(t)P(t)]\\ =&\Delta(t) - q\mathbb{E}[L(t)^*\Delta(t)] - q\mathbb{E}[\Delta(t)L(t)]\\ &+ q^2\mathbb{E}[L(t)^*\Delta(t)L(t)]\\ =& \left(I - \frac{q}{N}(L+L^*)\right)\Delta(t)\\ &+ \frac{q^2}{N}\sum_{g\in G}\sum_{h:h-g\in S}\sum_{k:k-g\in S}(e_g e_h^* - e_h e_h^*)\Delta(t)(e_k e_g^* - e_k e_k^*)\end{aligned}$$

Since $\pi(t) = \Delta(t)e_0$ (being $\Delta(t)$ symmetric for every t), we easily obtain from above that

$$\begin{aligned}\pi(t+1) =&(I - q\frac{1}{N}(L+L^*))\pi(t)\\ &+ \frac{1}{N}q^2\sum_{g\in G}\sum_{h:h-g\in S}\sum_{k:k-g\in S}\pi_{h-k}(t)\\ &\times(e_g e_g^* - e_g e_k^* - e_h e_g^* + e_h e_k^*)e_0\\ =&(I - q\frac{1}{N}(L+L^*)\pi(t)\\ &+ \frac{1}{N}q^2[\sum_{h:h\in S}\sum_{k:k\in S}\pi_{h-k}(t)(e_0 - e_h)\\ &+ \sum_{g\in -S}\sum_{h:h-g\in S}\pi_h(t)(e_h - e_g)]\,.\end{aligned}$$

From this we immediately see that the non-zero elements of $\tilde{T}$ have row indices in $(S-S)\cup S$ and column indices in $S-S$, where $S - S = \{g \in G : \exists s_1, s_2 \in S$ such that $g = s_1 - s_2\}$. Hence the result follows.

■

This lemma has a few interesting consequences. Note that the entries of the matrix T in (12) are proportional to N^{-1} and, moreover, the number of the non-zero entries is upper bounded by $(1 + d^2)^2$, where d is the degree of $\mathcal{G}$. This implies that, if the degree is small, i.e., the graph is sparse, then also the matrix T is *sparse*. In particular, if we consider a sequence of graphs of increasing size with fixed degree, we expect that, as N diverges, T would become negligible, and the MSA would depend on the matrix C only[2]. This would imply the unbiasedness of the algorithm, because it is immediate to remark that the invariant vector of C is $N^{-1}\mathbf{1}$. This property of asymptotical unbiasedness can be actually stated as the following result.

[2]Note that, in general, C is not a stochastic matrix since it may be negative on the diagonal, however for large enough N (and d fixed) C is a stochastic matrix.

Theorem IV.4 (Unbiasedness of BGA) *Fix a finite $S \subseteq \mathbb{Z}^d\setminus\{0\}$ generating $\mathbb{Z}^d$ as a group. For every integer n, let $V_n = [-n,n]^d$ considered as the Abelian group $\mathbb{Z}^d_{2n+1}$ and let $\mathcal{G}^{(n)}$ be the Cayley Abelian graph generated by $S_n = S \cap [-n,n]^d$. Then on the sequence of $\mathcal{G}^{(n)}$ the BGA is asymptotically unbiased.*

Proof: The idea is to apply the perturbation result Theorem A.1 to the sequence of matrices C^* and $(C+T)^*$. Notice that $\mathcal{G}_{C^*} = \mathcal{G}^{(n)}$ while $\mathcal{G}_{(C+T)^*} \supseteq \mathcal{G}^{(n)}$ by Lemma III.3. Hence, $\mathcal{G}_{C^*}$ and $\mathcal{G}_{(C+T)^*}$ are both strongly connected. This also implies that the limit graph on $\mathbb{Z}^d_n$ of the two sequences $\mathcal{G}_{C^*}$ and $\mathcal{G}_{(C+T)^*}$ both contain $\mathcal{G}^{(\infty)}$ which is simply the Abelian Cayley graph on $\mathbb{Z}^d_n$ generated by S which is strongly connected by the assumption made. Finally, notice that C^* is Abelian Cayley, hence obviously weakly democratic, while Lemma IV.3 guarantees that $(C+T)^*$ is a finite perturbation of C^* in the sense of Appendix A. Hence also $(C+T)^*$ is weakly democratic. This yields, by (9),
$$\mathrm{tr}\,B = |N^{-1} - \pi'_0| \leq N^{-1} + \pi'_0 \to 0.$$

■

Theorem IV.4 tells us that the BGA is asymptotically unbiased on sparse Abelian Cayley graphs. More precise results can be obtained by computing the matrix T in some examples: for instance, we consider ring graphs and complete graphs, which show opposite behaviors.

Example IV.5 (Ring graph) On ring graphs, Corollary IV.2 implies that, for N large enough,
$$1 - q\frac{8\pi^2}{N^3} \leq R \leq 1 - q(1-q)\frac{8\pi^2}{N^3},$$
and namely $R = 1 - \Theta(\frac{1}{N^3})$. Specializing the proof of Lemma IV.3, the evolution of $\pi(t)$ can be written as
$$\begin{aligned}\pi_j(t+1) =&\Big(1 - \frac{4q}{N} + \frac{q^2}{N}\pi_j^{A^2}\Big)\pi_j(t)\\ &+ 2\Big(\frac{q}{N} - \frac{q^2}{N}\pi_j^A\Big)(\pi_{j-1}(t) + \pi_{j+1}(t))\\ &+ \frac{q^2}{N}\left(2\pi_0(t) + \pi_2(t) + \pi_{-2}(t)\right)\delta_{0j},\end{aligned}$$
that is $\pi(t+1) = (C+T)\pi(t)$, with
$$C = \mathrm{circ}\,(1 - \frac{4q}{N}, 2\frac{q}{N}, 0, \ldots, 0, 2\frac{q}{N})$$
and
$$T = \frac{q^2}{N}\begin{pmatrix} 4 & 0 & 1 & 0 & \ldots & 0 & 1 & 0\\ -2 & 0 & -2 & 0 & \ldots & & \ldots & 0\\ 0 & 0 & 1 & 0 & \ldots & & \ldots & 0\\ 0 & \ldots & & & & & \ldots & 0\\ 0 & \ldots & & & & & \ldots & 0\\ 0 & \ldots & & & & & \ldots & 0\\ 0 & \ldots & & & \ldots & 0 & 1 & 0\\ -2 & 0 & \ldots & & \ldots & 0 & -2 & 0\end{pmatrix}.$$

Thanks to these explicit formulas, we can numerically compute the rate and bias. Namely, about the rate we obtain that $\mathrm{esr}(C) < \mathrm{esr}(C+T)$, and $\mathrm{esr}(C+T) = 1 - \Theta(N^{-3})$. This means that the perturbation T does not significantly affect the rate for large N. Moreover, since $\mathrm{esr}(C) = 1 - \Theta(N^{-3})$ and $\mathrm{esr}(C+T) - \mathrm{esr}(C) = 1 - \Theta(N^{-4})$ we argue that actually

$$R = 1 - q\frac{8\pi^2}{N^3} + O\left(\frac{1}{N^4}\right) \qquad \text{as } N \to \infty. \tag{13}$$

On the other hand, about the bias we obtain that $\mathrm{tr}\,(B) = \Theta(N^{-1})$. This is confirmed by the simulations reported in Fig. 1. □

Example IV.6 (Complete graph) Let $\mathcal{G}$ be a complete graph, and consider the BGA algorithm. In this case, the degree is proportional to N, and Theorem IV.4 does not apply. From [13] we know that

$$R = 1 - q(2-q)$$
$$B = \frac{q}{2-q}\frac{1}{N}(I - \frac{\mathbf{1}\mathbf{1}^*}{N}).$$

Then, $\mathrm{tr}\,B = \frac{q}{2-q}\left(1 - \frac{1}{N}\right)$ and hence the *BGA is not asymptotically unbiased on the complete graph.* □

In order to take into account the locality constraint on connectivity in real-world networks, several models of random geometric graphs have been proposed, as accounted in [16]. Such models are not in general Abelian Cayley: hence in the next example we study one such model by means of simulations.

Example IV.7 (Random geometric graph) In this example we consider sequences of random geometric graphs, based on the following construction. For all $N \in \mathbb{N}$, we sample N points $\{z_i\}_{i\in\mathbb{Z}_N}$ from a uniform distribution over a unit square $[0,1]^2$, and we draw an edge (i,j) between nodes $i,j \in \mathbb{Z}_N$ when $\|z_i - z_j\| \leq 0.8\sqrt{\frac{\log N}{N}}$. On these realizations we run the BGA algorithm until convergence is reached, up to a small tolerance threshold, and in this way we compute an approximation of $\lim_{t\to\infty}\beta(t)$. The results are plotted in Fig. 1, in comparison with the analogous quantity on the complete and ring graph. It appears from simulations that, as N diverges, β is $\Theta(1)$ on the complete graph, whereas it is $\Theta(N^{-1})$ on the ring graph, and $O(N^{-1/2})$ on the random geometric graph. These evidences are in accordance with the theoretical results, and suggest their extension to other families of geometric graphs. □

V. Broadcast with collisions

In this section we present a comprehensive analysis, in terms of both rate of convergence and bias, of the Collision Broadcast Gossip Algorithm. We recall that in this algorithm every node is allowed to broadcast at every time step with some probability p. After proving a general convergence result, we focus on Abelian Cayley graphs in Subsection V-A, and study ring and complete graphs as examples in Subsection V-B. Our main finding is that the behavior of the CBGA, in terms of both speed and bias, is close to the behavior of the BGA: in this sense we may claim the robustness of broadcast gossip algorithms to local interferences.

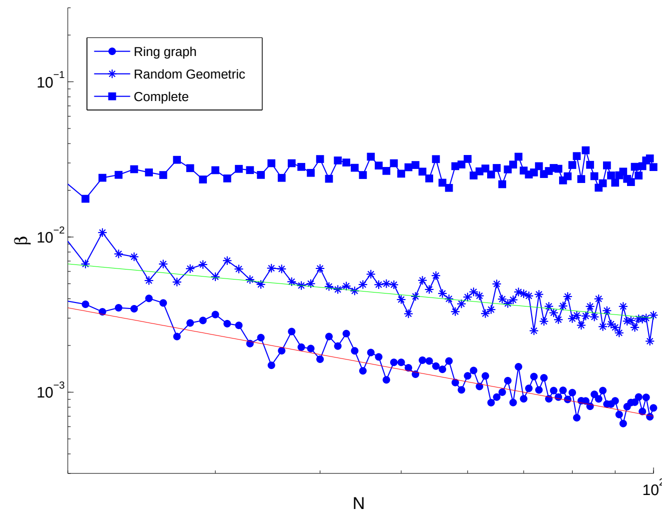

Fig. 1. The plot shows the asymptotical bias β as a function of N, computed by simulations on sequences of complete, random geometric, and ring graphs. Plotted values are the average over 1000 runs. See text for more information.

Proposition V.1 (Convergence of CBGA algorithm) *Consider the CBGA algorithm. Let $\mathcal{G}$ be any connected graph, and L its Laplacian matrix. Then,*

$$\bar{P} = I - qp(1-p)^{D^-}L, \tag{14}$$

where $((1-p)^{D^-})_{ij} = (1-p)^{D^-_{ij}}$ for every i and j. In particular, the CBGA algorithm achieves probabilistic consensus.

Proof: The probability of having at time t a successful transmission from v to u is $\mathbb{P}[A_{uv}(t) = 1] = p(1-p)^{d^-_u}A_{uv}$. Hence,

$$\mathbb{E}(A(t)) = p(1-p)^{D^-}A\,. \tag{15}$$

Now, $\bar{P} = \mathbb{E}[P(t)] = \mathbb{E}[I + q(A(t) - D(t))] = I + qp(1-p)^{D^-}[A - D^-]$. Note that the induced graph $\mathcal{G}_{\bar{P}}$ is strongly connected: by [14, Corollary 3.2], we can conclude the convergence of the CBGA. ■

Formula (14) is simpler if the graph is d-regular, because in that case

$$\bar{P} = I - qp(1-p)^d L.$$

Then, provided the graph is symmetric, $\mathrm{esr}(\bar{P}) = 1 - qp(1-p)^d\lambda_1$, and this together with (4) leads to

$$R \geq 1 - 2qp(1-p)^d\lambda_1$$

as a lower bound for the rate of convergence. As a function of p, this lower bound is minimal if p is equal to $p^* = \frac{1}{d+1}$.

Hence, natural questions are: is this bound tight? is p^* the best choice to improve the convergence rate? Section V-B will answer positively these questions for complete and ring graphs.

A. Abelian Cayley graphs

From now on we consider the CBGA on Abelian Cayley graphs. The next result, analogous to Lemma IV.3, characterizes the mean square analysis of the algorithm, giving a decomposition into a "simple" operator plus a perturbation.

Lemma V.2 *Consider the CBGA algorithm and let the communication graph $\mathcal{G}$ be Abelian Cayley with degree d. Then, the MSA vector π evolves as*

$$\pi(t+1) = (C+T)\pi(t), \tag{16}$$

where

$$C = I - qp(1-p)^d(L+L^*) + q^2p^2(1-p)^{2d}LL^*$$

and $T = q^2\tilde{T}$, with $\tilde{T}$ does not depend neither on q nor explicitly on N, and is such that the number of non-zero rows is at most $d^2(d+1)^2$ and non-zero columns is at most d^2.

Proof: Using commutativity of Abelian Cayley matrices and (15), we obtain that

$$\Delta(t+1) = (I - qp(1-p)^d(L+L^*))\Delta(t) + q^2\mathbb{E}[L(t)^*\Delta(t)L(t)]$$

Passing to the generating vector,

$$\pi(t+1) = (I - qp(1-p)^d(L+L^*))\pi(t) + q^2 f(\pi(t)) \tag{17}$$

where

$$\begin{aligned} f(\pi(t)) :=& \mathbb{E}[L(t)^*\Delta(t)L(t)]e_0 \\ =& \underbrace{\mathbb{E}[L(t)]^*\Delta(t)\mathbb{E}[L(t)]e_0}_{=:f_1(\pi(t))} \\ &+ \underbrace{\mathbb{E}[L(t)^*\Delta(t)L(t)]e_0 - \mathbb{E}[L(t)]^*\Delta(t)\mathbb{E}[L(t)]e_0}_{=:f_2(\pi(t))} \end{aligned}$$

Now,

$$f_1(\pi(t)) = p^2(1-p)^{2d}L^*L\pi(t) \tag{18}$$

and

$$\begin{aligned} [f_2(\pi(t))]_l =& \sum_{hk} \big[\mathbb{E}[L(t)_{hl}L(t)_{k0}] \\ &- \mathbb{E}[L(t)_{hl}]\mathbb{E}[L(t)_{k0}]\pi(t)_{h-k}\big] \\ =& \sum_{tk} \big[\mathbb{E}[L(t)_{k+t,l}L(t)_{k0}] \\ &- \mathbb{E}[L(t)_{k+t,l}]\mathbb{E}[L(t)_{k0}]\pi(t)_t\big] \end{aligned} \tag{19}$$

Notice now that, by the way the model has been defined, we have that $A_{ij}(t)$ and $A_{hk}(t)$ are independent whenever $N^-(i)\cap N^-(h) = \emptyset$ or equivalently $i-h \notin S-S$. In this case, also $L_{ij}(t)$ and $L_{hk}(t)$ are independent. Therefore, the double summation in (19) can be restricted to $k \in S\cup\{0\}$ and $t \in S-S$. Consequently, the values of l for which $[f_2(\pi(t))]_l \neq 0$ can be restricted to $(S\cup\{0\})+S-S-(S\cup\{0\})$. Plugging (18) into (17) and using the information on the structure of $f_2(\pi(t))$ obtained above, the result follows. ∎

Note that, as for the BGA, the matrix T is responsible for the bias of the algorithm, and the mixing parameter q plays the same role in both algorithms. The matrix T is roughly proportional to q^2, and (for small q) the matrix $I-C$ is proportional to q. This implies that by changing q we can trade-off q between speed and accuracy of both algorithms. For the BGA, this trade-off has been studied in [3], and we argue from Lemma V.2 that such analysis can be promptly extended to the CBGA. The simplicity of this extension relates to the robustness of BGA to collisions. Moreover, Lemma V.2 states the sparsity property of T, which is the key to infer an unbiasedness result analogous to Theorem IV.4.

Theorem V.3 (Unbiasedness of CBGA) *Fix a finite $S \subseteq \mathbb{Z}^d \setminus \{0\}$ generating $\mathbb{Z}^d$ as a group. For every integer n, let $V_n = [-n,n]^d$ considered as the Abelian group $\mathbb{Z}^d_{2n+1}$ and let $\mathcal{G}^{(n)}$ be the Cayley Abelian graph generated by $S_n = S\cap[-n,n]^d$. On the sequence of $\mathcal{G}^{(n)}$ the CBGA is asymptotically unbiased.*

Proof: The idea is to apply the perturbation result Theorem A.1 to the sequence of matrices C^* and $(C+T)^*$. Notice that $\mathcal{G}_{C^*} = \mathcal{G}_{A+A^*+A^*A} \supseteq \mathcal{G}^{(n)}$ while $\mathcal{G}_{(C+T)^*} \supseteq \mathcal{G}^{(n)}$ by Lemma III.3. Hence, $\mathcal{G}_{C^*}$ and $\mathcal{G}_{(C+T)^*}$ are both strongly connected. As in the proof of Theorem IV.4, connectivity implies that the limit graph on Z^d_n of the two sequences $\mathcal{G}_{C^*}$ and $\mathcal{G}_{(C+T)^*}$ are also strongly connected. Finally, notice that C^* is Cayley Abelian, hence obviously weakly democratic, while Lemma IV.3 guarantees that $(C+T)^*$ is a finite perturbation of C^* in the sense of Appendix A. Hence also $(C+T)^*$ is weakly democratic. This yields $\operatorname{tr} B = |N^{-1} - \pi'_0| \leq N^{-1} + \pi'_0 \to 0$. ∎

With this result, we have shown that also the CBGA is asymptotically unbiased on sparse Abelian Cayley graphs.

B. Ring and complete graphs

To begin with, we specialize the results in Section V-A to the case of ring graphs.The following result can be proven by computing T; the detailed derivation is omitted.

Proposition V.4 (Ring graph - Rate) *Given a ring graph and the CBGA algorithm, we have*

$$\begin{aligned} \bar{P} =& I - qp(1-p)^2L \\ \mathcal{L}(\Omega) =& \Omega - 2q(1-q)p(1-p)^2L - q^2p(1-p)^2N^{-1}L^2 \\ &+ q^2p(1-p)^2N^{-1}p\operatorname{circ}(\tau), \end{aligned}$$

where

$$\begin{aligned} \tau =& [2(p-2), 6-4p+p^2, -3(2-2p+p^2), 2-4p+3p^2, \\ & 0,\dots,0, 2-4p+3p^2, -3(2-2p+p^2), 6-4p+p^2]. \end{aligned}$$

In particular, for N large enough,

$$1 - qp(1-p)^2\frac{8\pi}{N^2} \leq R \leq 1 - q(1-q)p(1-p)^2\frac{8\pi}{N^2}.$$

This result in particular shows that the bound on the convergence rate based on $\bar{P}$ is asymptotically *tight* for the ring graph. Note that the speed of convergence for the CBGA is one order faster than the BGA: this is not surprising, since in the former case the average number of activated nodes per round is Np, instead of 1.

Remark V.5 (Large N) Based on the formulas in Proposition V.4, the performance for large N can be numerically investigated, showing that $\mathrm{esr}(C) < \mathrm{esr}(C+T)$, and $\mathrm{esr}(C+T) = 1 - \Theta(N^{-2})$. This means that the perturbation T does not significantly affect the rate for large N. Moreover, since $\mathrm{esr}(C) = 1-\Theta(N^{-2})$ and $\mathrm{esr}(C+T)-\mathrm{esr}(C) = 1-\Theta(N^{-3})$, we argue that actually

$$R = 1 - qp(1-p)^2 \frac{8\pi^2}{N^2} + O\left(\frac{1}{N^3}\right).$$

This formula is very close to Eq. (13) about the BGA. On the other hand, one has $\mathrm{tr}(B) = \Theta(N^{-1})$, that is the asymptotical error has the same dependence on N as for the BGA.

Remark V.6 (Optimization - Ring) Remarkably, for large N both the upper and the lower bound on the rate in Proposition V.4 show the same dependence on p. Thus, they can be simultaneously optimized by taking $p^* = 1/3$. Instead, the dependence on p of the asymptotical error is negligible. This implies, from the design point of view, that p can be chosen to be $p^* = 1/3$, optimizing the convergence rate, whereas by choosing q we trade off asymptotical error and convergence rate, as done for the BGA in [3].

As for the BGA, a more precise analysis can be pursued on complete graphs.

Proposition V.7 (Complete graph) *Let $x(t)$ evolve following the CBGA algorithm. Then,*

$$\begin{aligned} R &= 1 - q(2-q)Np(1-p)^{N-1} \\ B &= \frac{q}{2-q}\frac{1}{N}(I - \frac{\mathbf{1}\mathbf{1}^*}{N}). \end{aligned}$$

Namely, $\mathrm{tr}(B) = \frac{q}{2-q}\left(1 - \frac{1}{N}\right)$ *and then the* CBGA is not asymptotically unbiased on the complete graph.

Proof: It is immediate that in the complete graph either one node communicates to every others, or no node communicates. Then $\mathbb{P}[P(t) = I] = 1 - Np(1-p)^{N-1}$ and $\mathbb{P}[P(t) = P_b] = Np(1-p)^{N-1}$, where $P_b = I + q\sum_{u\neq v}(e_u e_v^* - e_u e_u^*)$, and v is the realization of a random variable uniformly distributed over the nodes. We note that

$$\mathbb{E}[P_b] = I + q\frac{\mathbf{1}\mathbf{1}^* - NI}{N},$$

which implies that

$$\begin{aligned} &\mathbb{E}[P(t)^*P(t)] \\ &\quad = Np(1-p)^{N-1}\mathbb{E}[P_b^*P_b] + (1 - Np(1-p)^{N-1})I \\ &\quad = [1 - 2q(1-q)Np(1-p)^{N-1}]I \\ &\qquad + 2q(1-q)Np(1-p)^{N-1}\frac{\mathbf{1}\mathbf{1}^*}{N}. \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[P(t)^*\mathbf{1}\mathbf{1}^*P(t)] = &\, q^2N^2p(1-p)^{N-1}I \\ &+ (1 - q^2Np(1-p)^{N-1})\mathbf{1}\mathbf{1}^*. \end{aligned}$$

Hence the application of $\mathcal{L}$ keeps invariant the subspaces generated by I and $\mathbf{1}\mathbf{1}^*$, and the linear operator $\mathcal{L}$ can be represented by the matrix

$$\begin{pmatrix} 1 - 2q(1-q)Np(1-p)^{N-1} & q^2Np(1-p)^{N-1} \\ 2q(1-q)Np(1-p)^{N-1} & 1 - q^2Np(1-p)^{N-1} \end{pmatrix}.$$

The eigenvalues of this matrix are 1 and $R = 1 - q(2-q)Np(1-p)^{N-1}$, and the eigenspace relative to eigenvalue 1 is spanned by the vector $v^{(1)} = qNp(1-p)^{N-1}\left(q, 2(1-q)\right)^*$. Since $E[\rho\rho^*]$ belongs to this eigenspace, and $\mathbf{1}^*E[\rho\rho^*]\mathbf{1} = 1$, we conclude that

$$B = E[\rho\rho^*] - N^{-2}\mathbf{1}\mathbf{1}^* = \frac{q}{2-q}\frac{1}{N}\big(I - \frac{\mathbf{1}\mathbf{1}^*}{N}\big).$$

■

Some remarks are in order about the parameters p, q in the CBGA algorithm on complete graphs.

Remark V.8 (Optimization - Complete) The convergence rate R as a function of p is optimal for $p^* = 1/N$. Note that $R(p^*) = 1 - q(2-q)(1 - \frac{1}{N})^{N-1} \to 1 - q(2-q)\frac{1}{e}$ when N goes to infinity, while if we fix $p = \bar{p} \in (0,1)$, then $R(\bar{p}) \to 1$. On the other hand, B is independent of p. From the design point of view, it is clear that p has to be chosen equal to N^{-1}, optimizing the speed. Instead, choosing q we trade off speed and asymptotic displacement: if we recall the formulas for the BGA in Example IV.6, it is clear that the optimization problem is the same for both algorithms.

VI. Conclusion

This paper has been devoted to study gossip algorithm for the estimation of averages, based on iterated broadcasting of current estimates. We presented a novel broadcast gossip algorithm, dealing with communication interference, whose effect is studied in the paper. Our results, obtained under symmetry assumptions about the network topology, allow us to conjecture an interesting picture of the performance of broadcast gossip algorithms on real world networks, in terms of accuracy and of robustness to interference. In broad terms, we claim that the BGA is robust to interferences. As expected, interferences have a negative effect on the rate of convergence, which can be mitigated by a suitable choice of the broadcasting probability p. Instead, interferences have on the asymptotical error a small effect, which is negligible on large networks. The size of the network is also important for accuracy: on large highly connected graphs, both algorithms provide biased estimations, whereas on sparse graphs the estimation bias goes to zero as the network grows larger. On the other hand, the rate of convergence degrades on large sparse graphs, which is a general feature of consensus algorithms based on diffusion.

The results of this paper have been obtained under two simultaneous technical assumptions: the graphs are Abelian Cayley, thus in particular vertex-transitive, and in the CBGA each node broadcasts with the same probability. Future work should consider non-vertex-transitive networks of nodes with non-uniform broadcasting probabilities. A better understanding of the role of the network topology in the trade-off between speed and achievable precision may come from such extension.

Acknowledgements: The authors wish to thank Sandro Zampieri for many fruitful discussions on the issues studied in this paper.

APPENDIX

In this appendix we recall a perturbation result from [12] about the limit of the invariant vectors of sequences of stochastic matrices, which is used in our paper to estimate the trace of the matrix B.

We assume we have fixed an infinite universe set $\mathcal{V}$, an increasing sequence V_n of finite cardinality subsets of $\mathcal{V}$ such that $\cup_n V_n = \mathcal{V}$ and a sequence of irreducible stochastic matrices $P^{(n)}$ on the state spaces V_n with the following stabilizing property: for every $i \in \mathcal{V}$, there exist $n(i) \in \mathbb{N}$ such that $i \in V_{n(i)}$ and

$$P_{ij}^{(n)} = P_{ij}^{(n(i))}\,, \quad \forall n \geq n(i)\,,\, \forall j \in V_{n(i)}.$$

This property allows us to define, in a natural way, a limit stochastic matrix on $\mathcal{V}$. For every $i, j \in \mathcal{V}$, we define

$$P_{ij}^{(\infty)} = \begin{cases} P_{ij}^{(n(i))} & \text{if } j \in V_{n(i)} \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

The sequence of stochastic matrices $P^{(n)}$ is said to be *weakly democratic* if the corresponding invariant vectors $\pi^{(n)}$ are such that, for all $i \in \mathcal{V}$, $\pi_i^{(n)} \to 0$ for $n \to +\infty$. Fix now a finite subset $W \subseteq \cap_n V_n$ and another sequence of irreducible stochastic matrices $\tilde{P}^{(n)}$ on V_n such that

$$\tilde{P}_{ij}^{(n)} = P_{ij}^{(n)} \quad \forall i \in V_n \setminus W\,,\, \forall j \in V_n$$

$$\tilde{P}_{ij}^{(n)} = \tilde{P}_{ij}^{(1)} \quad \forall\, i \in W\,,\, \forall j \in V_1$$

In other terms, $\tilde{P}^{(n)}$ can be seen as a perturbed version of $P^{(n)}$ with the perturbation confined to the fixed subset W and stable (it does not change as n increases). Also for this perturbed sequence we can define, following (20), the asymptotic chain $\tilde{P}^{(\infty)}$. The following result has been proven in [12].

Theorem A.1 *Suppose that $P^{(\infty)}$ and $\tilde{P}^{(\infty)}$ are both irreducible. Then, if $P^{(n)}$ is weakly democratic, also $\tilde{P}^{(n)}$ is weakly democratic.*

REFERENCES

[1] T. C. Aysal and K. E. Barner. Convergence of consensus models with stochastic disturbances. *IEEE Transactions on Information Theory*, 56(8):4101–4113, 2010.

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

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

[4] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli. Gossip along the way: Order-optimal consensus through randomized path averaging. In *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, September 2007.

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

[6] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, 44(3):671–684, 2008.

[7] R. Carli, F. Garin, and S. Zampieri. Quadratic indices for the analysis of consensus algorithms. In *Information Theory and Applications Workshop*, San Diego, CA, February 2009.

[8] G. Chockler, M. Demirbas, S. Gilbert, N. Lynch, C. Newport, and T. Nolte. Consensus and collision detectors in radio networks. *Distributed Computing*, 21(1):55–84, 2008.

[9] P. Denantes, F. Benezit, P. Thiran, and M. Vetterli. Which distributed averaging algorithm should I choose for my sensor network? In *IEEE Conference on Computer Communications (INFOCOM)*, pages 986–994, April 2008.

[10] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, 2010.

[11] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright. Geographic gossip: Efficient averaging for sensor networks. *IEEE Transactions on Signal Processing*, 56(3):1205–1216, 2008.

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

[13] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. In *Information Theory and Applications Workshop*, San Diego, CA, January 2007.

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

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

[16] M. Franceschetti and R. Meester. *Random networks for communication*. Cambridge University Press, 2007.

[17] F. Garin and S. Zampieri. Performance of consensus algorithms in large-scale distributed estimation. In *European Control Conference*, Budapest, Hungary, August 2009.

[18] D. Mosk-Aoyama and D. Shah. Fast distributed algorithms for computing separable functions. *IEEE Transactions on Information Theory*, 54(7):2997–3007, 2008.

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

[20] S. Rai. The spectrum of a random geometric graph is concentrated. *Journal of Theoretical Probability*, 20(2):119–132, 2007.

[21] B. Recht and R. D'Andrea. Distributed control of systems over discrete groups. *IEEE Transactions on Automatic Control*, 49(9):1446–1452, 2004.

[22] V. Saligrama, M. Alanyali, and O. Savas. Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Transactions on Signal Processing*, 54(11):4118–4132, 2006.

[23] A. Terras. *Fourier analysis on finite groups and applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, MA, 1999.