The asymptotical error of broadcast gossip averaging algorithms

Fabio Fagnani   Paolo Frasca[*]

December 16, 2010

Abstract

In problems of estimation and control which involve a network, efficient distributed computation of averages is a key issue. This paper presents theoretical and simulation results about the accumulation of errors during the computation of averages by means of iterative "broadcast gossip" algorithms. Using martingale theory, we prove that the expectation of the accumulated error can be bounded from above by a quantity which only depends on the mixing parameter of the algorithm and on few properties of the network: its size, its maximum degree and its spectral gap. Both analytical results and computer simulations show that in several network topologies of applicative interest the accumulated error goes to zero as the size of the network grows large.

1 Introduction

Distributed computation of averages is an important building block to solve problems of estimation and control over networks. As a reliable time-independent communication topology may be unlikely in the applications, a growing interest has been devoted to randomized "gossip" algorithms to compute averages. In such algorithms, at each time step, a random subset of the nodes communicates and performs computations. Unfortunately, some of these iterative algorithms do not deterministically ensure that the average is preserved through iterations, and due to the accumulation of errors, in general there is no guarantee that the typical algorithm realization will converge to a value which is close to the desired average.

In the present paper we focus on one of these randomized algorithm, the Broadcast Gossip Algorithm (BGA). In this algorithm, a node is randomly selected at each time step to broadcast its current value to its neighbors, which in turn update their values by a local averaging rule. Since these updates are not symmetric, it is clear that the average is not preserved, but instead is changed at each time step by some amount. In this paper we study how these errors accumulate, and how much the convergence value of the algorithm differs from the original average to be computed.

1.1 Contribution

In this paper, we study the bias, or asymptotical error, committed by a Broadcasting Gossip averaging algorithm, and we show that large neighborhoods and a large mixing parameter induce a large asymptotical error. As a theoretical contribution, we study the average of states as a martingale, and by this interpretation we prove that on symmetric graphs the expectation of the accumulated error can not be larger than a constant times $\dfrac{q}{1-q}\dfrac{d_{\max}^2}{N\lambda_1}$, where q is the "mixing" parameter of the algorithm, N is the network size, $d_{\max}$ is the maximum degree of the nodes, and λ_1 is the network spectral gap. For some families of graphs (e.g, expander graphs), this is enough to prove that the bias goes to zero as N goes to infinity. Further, by means of simulations we show that, on some example graph topologies, the mean

---

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

bias is an increasing function of the mixing parameter and is proportional, on large networks, to the ratio between degree and size of the network. In particular, whenever $d_{\max} = o(N)$, the simulated bias goes to zero as N goes to infinity.

1.2 Related works

The paper [9] provides a general theory for randomized linear averaging algorithms, and presents a few example algorithms, some of which do not preserve the average of the states. Among these algorithms, the Broadcast Gossip Algorithm, studied in the present paper, has been attracting a wide interest, for its natural application to wireless networks: main references include the paper [2] and the recent survey [6]. While it is simple to give conditions to ensure that the expectation of the convergence value is equal to the initial average, the problem of estimating the difference between expectation and realizations is harder, and has received partial answers in a few papers. In [1] the authors study a related communication model, in which the broadcasted values may not be received with a probability which depends on the transmitter and receiver nodes, and claim that "aggressive updating combined with large neighborhoods [...] result in more variance [of the convergence value] within the short time to convergence". This intuition extends to the Broadcast Gossip Algorithm which we are considering in this paper. Actually, in [2] the variance of the limit value has been estimated for general graphs, with an upper bound which is proportional to $\left(1 - \frac{\lambda_1}{\lambda_{N-1}} \frac{1}{1 - \frac{1}{2}\frac{q}{N}\lambda_{N-1}}\right)$, where λ_i is the i-th smallest eigenvalue of the graph Laplacian. This bound, however, is not useful to prove that the bias goes to zero as N grows, a fact which has been proved in [7] for sequences of Abelian Cayley graphs with bounded degree, using tools from algebra and Markov chain theory. Analogous problems can be studied for other randomized algorithms which do not preserve the average. For instance, in [8] two related algorithms are studied, in which node values are sent to one random neighbor only. If at each time step one random node sends its value, then the variance has an upper bound which is proportional to $\frac{q}{1-q}\frac{1}{N}$, while if at each time step every node sends its value, then the bound is proportional to $\frac{q^2}{\lambda_1 N}$.

2 Problem statement

Let a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ be given, together with $N = |\mathcal{V}|$ real numbers $\{x_v\}_{v\in\mathcal{V}} \subset [0, L]$. For every node $v \in \mathcal{V}$, we denote its out-neighborhood by $\mathcal{N}_v^+ = \{u \in \mathcal{V} : (u, v) \in \mathcal{E}\}$, and its in-neighborhood by $\mathcal{N}_v^- = \{u \in \mathcal{V} : (v, u) \in \mathcal{E}\}$. The following Broadcasting Gossip Algorithm (BGA) is run in order to estimate the average $x_{\mathrm{ave}} = N^{-1} \sum_{v\in\mathcal{V}} x_v$.

At each time step $t \in \mathbb{Z}_{\geq 0}$, one node v is sampled from a uniform distribution over $\mathcal{V}$. Then, node v broadcasts its state $x_v(t)$ to its neighbors $u \in \mathcal{N}_v^+$, which in turn update their states as

$$x_u(t+1) = (1-q)x_u(t) + qx_v(t). \tag{1}$$

The parameter $q \in (0, 1)$ is said to be the *mixing parameter* of the algorithm. If instead $u \notin \mathcal{N}_v^+$, there is no update: $x_u(t+1) = x_u(t)$.

It is known from [9, Corollary 3.2] that the BGA converges, in the sense that there exists a random variable x^* such that almost surely $\lim_{t\to+\infty} x(t) = x^*\mathbf{1}$. Let now $x_{\mathrm{ave}}(t) = N^{-1}\sum_{v\in\mathcal{V}} x_v(t)$. Although one can find conditions to ensure that $\mathbb{E}[x^*] = x_{\mathrm{ave}}(0)$, in general x^* is not equal to $x_{\mathrm{ave}}(0)$. Then, it is worth to ask how far the convergence value is from the initial average. To study this bias in the computation of the average, we define

$$\beta(t) = |x_{\mathrm{ave}}(t) - x_{\mathrm{ave}}(0)|^2.$$

The goal of this work is to study this quantity, and in particular its limit $\mathbb{E}[\beta(\infty)] := \lim_{t\to\infty} \mathbb{E}[\beta(t)]$, with a special attention to its dependence on the size of the network. In particular, we shall say that the algorithm is *asymptotically unbiased* if $\lim_{N\to+\infty} \mathbb{E}[\beta(\infty)] = 0$.

2

3 Analysis

3.1 A simplistic bound

Using (1), it is immediate to compute that, given v to be the broadcasting node at time t,

$$x_{\mathrm{ave}}(t+1) - x_{\mathrm{ave}}(t) = \frac{q}{N} \sum_{u \in \mathcal{N}_v^+} (x_v(t) - x_u(t)). \tag{2}$$

Then, we can obtain the following deterministic bound on the error introduced at each time step,

$$|x_{\mathrm{ave}}(t+1) - x_{\mathrm{ave}}(t)| \leq \frac{q}{N} d_v^+ L \leq \frac{q d_{\mathrm{max}}^+}{N} L, \tag{3}$$

where $d_v^+ = |\mathcal{N}_v^+|$ is the out-degree of node v, and $d_{\mathrm{max}}^+ = \max_{v \in \mathcal{V}} d_v^+$. This simple bound is worth some informal remarks. Indeed, Equation (3) suggests that choosing a low value of the mixing parameter q, and a graph with low degree d_{max}^+ and large size N, may ensure a small bias in the computation of the average. However, by choosing q, N or d_{max}^+, one affects the speed of convergence of the algorithm. Assume one is interested in an accurate computation, and chooses low values for q and d_{max}^+, compared to N. This choice would likely imply a slow convergence, and in turn a slow convergence may enforce to run the algorithms for a larger number of steps, in order to meet the same precision requirement. These extra steps, however, would introduce extra errors, thus possibly wasting the desired advantage in the accuracy. We argue from this discussion that there is a delicate *trade-off between speed and accuracy* for the BGA algorithm. The results presented in the next sections will shed light on this trade-off.

3.2 The average as a martingale

In this section, we shall derive a general bound on $\mathbb{E}[\beta(\infty)]$ in terms of the topology of the graph. The derivation is based on applying the theory of martingales to the stochastic processes $x(t)$ and $x_{\mathrm{ave}}(t)$. The reader can find the essentials of martingale theory in [11] or in [13].

Definition 3.1 *Given a sequence (filtration) of σ-algebras $\{\sigma_n\}_{n \in \mathbb{Z}_{\geq 0}}$, a sequence of random variables $\{M_n\}_{n \in \mathbb{Z}_{\geq 0}}$ is a σ_n-martingale if $\mathbb{E}[M_m | \sigma_n] = M_n$, for any $m \geq n$.*

Our first result states that $x_{\mathrm{ave}}(t)$ is a martingale with respect to the filtration induced by $x(t)$. Before the statement, we need some definitions. Let $d_v^+ = |\mathcal{N}_v^+|$ and $d_v^- = |\mathcal{N}_v^-|$ be the out-degree and in-degree of node v. The graph $\mathcal{G}$ is said to be *balanced* if $d_u^- = d_u^+$ for all $u \in \mathcal{V}$. Given a set of random variables X, we denote by $\sigma(X)$ the sigma-algebra generated by the random variables in X.

Proposition 3.1 *Let us consider the BGA algorithm and the filtration $\mathcal{F}_t = \sigma(\{x(s) : s \leq t\})$. If the graph $\mathcal{G}$ is balanced, then the sequence of random variables $\{x_{\mathrm{ave}}(t)\}_{t \in \mathbb{Z}_{\geq 0}}$ is a square-integrable $\mathcal{F}_t$-martingale.*

Proof: First, note that $x_{\text{ave}}(t)$ is $\mathcal{F}_t$-measurable. Moreover, Equation (2) implies that for all $t \geq 0$,

$$\mathbb{E}[x_{\text{ave}}(t+1) - x_{\text{ave}}(t)|\mathcal{F}_t] = \frac{1}{N} \sum_{v \in \mathcal{V}} \left(\frac{q}{N} \sum_{u \in \mathcal{N}_v^+} (x_v(t) - x_u(t)) \right)$$

$$= \frac{q}{N^2} \left[\sum_{v \in \mathcal{V}} d_v^+ x_v(t) - \sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{N}_v^+} x_u(t) \right]$$

$$= \frac{q}{N^2} \left[\sum_{v \in \mathcal{V}} d_v^+ x_v(t) - \sum_{u \in \mathcal{V}} d_u^- x_u(t) \right]$$

$$= \frac{q}{N^2} \left[\sum_{v \in \mathcal{V}} \left(d_v^+ - d_v^- \right) x_v(t) \right] = 0,$$

since we are assuming $d_u^- = d_u^+$ for every $u \in \mathcal{V}$. Then, the sequence of random variables $(x_{\text{ave}}(t))_{t \in \mathbb{Z}_{\geq 0}}$ is an $\mathcal{F}_t$-martingale. Moreover, the fact that $x_{\text{ave}}(t) \in [0, L]$ implies that the martingale is bounded in L^p for every $p \geq 1$, and in particular square-integrable. □

Let us define the distance from the agreement as

$$d(t) := \frac{1}{N} \sum_{v \in \mathcal{V}} (x_v(t) - x_{\text{ave}}(t))^2.$$

Using this definition, we can prove an inequality which is a refinement of (3). Let $d_{\max}^{\pm} = \max_v \{d_v^{\pm}\}$ and $d_{\max} = \max\{d_{\max}^-, d_{\max}^+\}$.

Lemma 3.2 *Let $\mathcal{G}$ be balanced. Then, the increments of the martingale $\{x_{\text{ave}}(t)\}_{t \in \mathbb{Z}_{\geq 0}}$ have bounded variance, in particular, for all $t \geq 0$,*

$$\mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2|\mathcal{F}_t] \leq 4 \frac{q^2 d_{\max}^2}{N^2} d(t). \tag{4}$$

Proof: By (2), we have

$$\mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2|\mathcal{F}_t] = \frac{1}{N} \sum_{v \in \mathcal{V}} \left(\frac{q}{N} \sum_{u \in \mathcal{N}_v^+} (x_v(t) - x_u(t)) \right)^2$$

$$\leq \frac{1}{N} \sum_{v \in \mathcal{V}} \frac{q^2}{N^2} d_v^+ \sum_{u \in \mathcal{N}_v^+} (x_v(t) - x_u(t))^2$$

$$\leq \frac{1}{N} \sum_{v \in \mathcal{V}} \frac{q^2}{N^2} 2 \left((d_v^+)^2 (x_v(t) - x_{\text{ave}}(t))^2 + d_v^+ \sum_{u \in \mathcal{N}_v^+} (x_{\text{ave}}(t) - x_u(t))^2 \right)$$

$$\leq 2 \left(\frac{q^2}{N^2} (d_{\max}^+)^2 d(t) + d_{\max}^+ d_{\max}^- d(t) \right)$$

$$\leq 4 \frac{q^2 d_{\max}^2}{N^2} d(t).$$

This completes the proof. □

We define the rate of convergence of the algorithm as

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

4

Then, there exists a positive constant C_R, depending on $x(0)$, such that $\mathbb{E}[d(t)] \leq C_R R^t$. This fact, combined with Lemma 3.2, implies that

$$\mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2] = \mathbb{E}\left[\mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2 | \mathcal{F}_t]\right]$$

$$\leq 4\, C_R \frac{q^2 d_{\max}^2}{N^2} R^t \tag{5}$$

We recall that the *spectral gap* of the graph $\mathcal{G}$ is the smallest (in modulus) non-zero eigenvalue of its Laplacian matrix, and we denote this quantity by λ_1. It is well-known that λ_1 relates to the mixing rate of Markov chains, and to the speed of convergence of gossip algorithms [3, 9]: the larger the spectral gap, the faster the convergence. In particular, let us assume that the graph $\mathcal{G}$ be *symmetric*, that is, such that $\mathcal{N}_v^+ = \mathcal{N}_v^-$ for all $v \in \mathcal{V}$. Then, we know from [7, Equation (18)] that

$$R \leq 1 - \frac{2q(1-q)}{N}\lambda_1. \tag{6}$$

Using these facts, we are going to prove the next result about the asymptotic behavior of the bias as $t \to +\infty$.

Proposition 3.3 *If $\mathcal{G}$ is symmetric, then there exists a constant $C > 0$ such that*

$$\mathbb{E}[\beta(\infty)] \leq C \frac{q}{1-q} \frac{d_{\max}^2}{N\lambda_1}.$$

Proof: Using the orthogonality of the increments of square-integrable martingales, we observe that

$$\lim_{t \to +\infty} \mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(0))^2] = \lim_{T \to +\infty} \mathbb{E}\left[\left(\sum_{t=0}^{T-1}(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))\right)^2\right]$$

$$= \lim_{T \to +\infty} \mathbb{E}\left[\sum_{t=0}^{T-1}(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2\right.$$

$$\left. +2\sum_{t=1}^{T-1}\sum_{s<t}(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))(x_{\text{ave}}(s+1) - x_{\text{ave}}(s))\right]$$

$$= \lim_{T \to +\infty} \left[\sum_{t=0}^{T-1}\mathbb{E}\left[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2\right]\right.$$

$$\left. +2\sum_{t=1}^{T-1}\sum_{s<t}\mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))(x_{\text{ave}}(s+1) - x_{\text{ave}}(s))]\right]$$

$$= \lim_{T \to +\infty} \sum_{t=0}^{T-1}\mathbb{E}\left[(x_{\text{ave}}(t+1) - x_{\text{ave}}(t))^2\right].$$

By applying Equation (5)

$$\lim_{t \to +\infty} \mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(0))^2] \leq 4\,C_R \frac{q^2 d_{\max}^2}{N^2} \lim_{T \to +\infty} \sum_{t=0}^{T-1} R^t = 4\,C_R \frac{q^2 d_{\max}^2}{N^2} \frac{1}{1-R}.$$

The inequality in (6) implies that $\lim_{t \to +\infty} \mathbb{E}[(x_{\text{ave}}(t+1) - x_{\text{ave}}(0))^2] \leq 2\,C_R \frac{q}{1-q} \frac{d_{\max}^2}{N\lambda_1}$. The thesis then follows, with $C = 2C_R$, by applying the dominated convergence theorem. $\qquad\square$

5

Note that the proof of Proposition 3.3 also implies that

$$\sup_t \mathbb{E}[(x_{\mathrm{ave}}(t+1) - x_{\mathrm{ave}}(0))^2] \leq C \frac{q}{1-q} \frac{d_{\max}^2}{N\lambda_1}.$$

On the other hand, for a convergent square-integrable martingale M_t, we know by Doob's maximal inequality that $\mathbb{E}[\sup_t M_t^2] \leq 4\mathbb{E}[\lim_t M_t^2]$. Then, we can immediately obtain the following finite-time counterpart of Proposition 3.3.

Theorem 3.4 *If $\mathcal{G}$ is symmetric, then there exists $C' > 0$ such that*

$$\mathbb{E}\left[\sup_{t\in\mathbb{N}} \beta(t)\right] \leq C' \frac{q}{1-q} \frac{d_{\max}^2}{N\lambda_1}.$$

Let us now consider a sequence of graphs $\mathcal{G}_N$ of increasing size N. In such a sequence, both $d_{\max}$ and λ_1 are functions of N. In this context, Proposition 3.3 implies the following corollary.

Corollary 3.5 *Let $I \subseteq \mathbb{N}$ and $(\mathcal{G}_N)_{N\in I}$ be a sequence of symmetric graphs such that $\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N)$ and $|\mathcal{V}_N| = N$. If*

$$\frac{d_{\max}^2}{\lambda_1} = o(N) \quad \text{as } N \to +\infty,$$

then the BGA algorithm is asymptotically unbiased.

Note that, since $x_{\mathrm{ave}}(t)$ converges a.s. and $\lim_{t\to+\infty} x_{\mathrm{ave}}(t) \in [0, L]$, it is trivial to find an upper bound on $\mathbb{E}[\beta(\infty)]$ which does not depend on N. The interest of the above corollary is in giving a sufficient condition for $\mathbb{E}[\beta(\infty)]$ to be $o(1)$ as $N \to \infty$.

Remark 3.6 Applying Markov's inequality, we see from Proposition 3.3 that for any $c > 0$,

$$\mathbb{P}[\beta(\infty) > c] \leq C \frac{q}{1-q} \frac{d_{\max}^2}{N\lambda_1} \frac{1}{c}$$

In the applications, one is often interested in computing an average because the average is the maximum likelihood estimator of the expectation of a random variable. In such context, the average enjoys the property that the mean square error, committed by approximating the expectation by the average of N samples, is equal to $1/N$. For this reason, one would like to ensure that the bias introduced by the Broadcast Gossip algorithm is not larger than $1/N$. If we take $c = 1/N$, we get

$$\mathbb{P}\left[\beta(\infty) > \frac{1}{N}\right] \leq C \frac{q}{1-q} \frac{d_{\max}^2}{\lambda_1}.$$

Then, provided the right-hand-side of this inequality does not diverge, we can choose the mixing parameter q so that with a positive given probability the bias is below $1/N$. In such case, if our purpose is distributed estimation of an expected value, averages which are approximated by a Broadcast Gossip Algorithm are as good as averages computed by a centralized method. $\square$

4   Examples

In this section, we show that the BGA is asymptotically unbiased on several example topologies which have been considered in the literature. Given a graph $\mathcal{G}_N$ of size N, we denote by $\lambda_1(N)$ and $d_{\max}(N)$ its spectral gap and maximum degree, respectively. We consider the following example sequences of graphs.

- **Expander graphs.**
  A sequence of graphs is said to be an *expander* sequence if there exist $d \in \mathbb{N}$ and $c > 0$ such that for every N, $\lambda_1(N) \geq c$ and $d_{\max}(N) \leq d$. In this case, $\frac{d_{\max}^2}{\lambda_1}$ is bounded, and this fact implies by Proposition 3.3 that

  $$\mathbb{E}[\beta(\infty)] = O\left(\frac{1}{N}\right) \text{ as } N \to +\infty,$$

  and then the BGA is asymptotically unbiased and Remark 3.6 applies. An example of an expander sequence is given by a sequence of *de Bruijn graphs* on n symbols of increasing dimensions k. A de Bruijn graph on n symbols of dimension k has n^k vertices and edges from any i to $ni, ni + 1, ni + 2, \ldots, ni + k - 1$ (all modulo n^k). Their expander properties have already been applied to efficient averaging algorithms in [5].

- **Hypercube graphs.**
  The n-dimensional hypercube graph is the graph obtained drawing the edges of a n-dimensional hypercube. It has $N = 2^n$ nodes which can be identified with the binary words of length n, and two nodes are neighbors if the corresponding binary words differ in only one component. For these graphs it is known, for instance from [10, Example 7], that $\lambda_1(N) = \Theta(1/\log N)$ and $d_{\max}(N) = O(\log N)$. Then,

  $$\mathbb{E}[\beta(\infty)] = O\left(\frac{\log^3 N}{N}\right) \text{ as } N \to +\infty$$

  and the BGA is asymptotically unbiased.

- **k-dimensional square lattices.**
  We consider square lattices obtained by tiling a k-dimensional torus, with $N = n^k$ nodes. For these graphs we know [4, Theorem 6], that $\lambda_1(N) = \Theta(1/N^{2/k})$ and $d_{\max}(N) = 2k$. Then,

  $$\mathbb{E}[\beta(\infty)] = O\left(\frac{1}{N^{1-2/k}}\right) \text{ as } N \to +\infty,$$

  and we argue that $k-$lattices are asymptotically unbiased if $k > 3$. However, we know that the BGA is asymptotically unbiased also if $k = 1, 2$: this has been proved in [7] using different techniques.

- **Random geometric graphs.**
  We can also consider random sequences of geometric graphs constructed as follows. We sample N points from a uniform distribution over the unit square, and we let nodes i and j be connected if the two corresponding points in the square are less than $r(N)$ far apart, with $r(N) = 1.1\sqrt{\frac{\log(N)}{N}}$. For these graphs, we know from [12] that with high probability $\lambda_1(N) = \Theta(1/N)$ and $d_{\max}(N) = O(\log N)$. Then, with high probability $\mathbb{E}[\beta(\infty)] = O\left(\log^2 N\right)$ as $N \to +\infty$, and we can not conclude asymptotical unbiasedness.

- **Complete graphs.**
  For these graphs, $\lambda_1(N) = N$ and $d_{\max}(N) = N - 1$.
  Then, we can not conclude from Proposition 3.3 that the BGA is asymptotically unbiased on complete graphs. Actually, in [9] it is shown that the BGA is *not* asymptotically unbiased on complete graphs, and in particular

  $$\mathbb{E}[\beta(\infty)] = \text{Var}(x(0))\frac{q}{2-q}\frac{N-1}{N}, \tag{7}$$

  where by $\text{Var}(x(0))$ we denote the (sample) variance of the initial condition.

5 Simulations

We have extensively simulated the evolution of BGA algorithm on sequences of graphs, and in particular on the example topologies presented in Section 4. In this section, we account for our results about the dependence of the bias $\beta(\infty)$ on the size N and on the parameter q.

Our simulation setup is as follows. Let q, N and the graph topology be chosen. For every run of the algorithm we generate a vector of initial conditions $x(0)$, sampling from a uniform distribution over $[0, 1]$.[1] Then, we run the algorithm until the disagreement $d(t)$ is below a small threshold ε, which we set at 10^{-4}. At this time $T^\varepsilon = \inf\{t \geq 0 : d(t) \leq \varepsilon\}$, the algorithm is stopped, and $\beta(T^\varepsilon)$ is evaluated. In order to simulate the expectation of $\mathbb{E}[\beta(\infty)]$, we average the outcome of 1000 realizations of $\beta(T^\varepsilon)$. Our results about the dependence on N are summarized in Figure 1, which plots the average bias against N in a log-log diagram. As expected, complete graphs are not asymptotically unbiased, while all other topologies, in which the degree is $o(N)$, are asymptotically unbiased. In particular, for de Bruijn graphs on 2 symbols, ring graphs and torus graphs, the bias is $\Theta(N^{-1})$, whereas for hypercubes and random geometric graphs the bias is $\Theta(\frac{\log(N)}{N})$. Overall, our set of simulations suggests that

$$\mathbb{E}[\beta(\infty)] = \Theta\left(\frac{d_{\max}}{N}\right) \quad \text{as } N \to \infty.$$

Results presented in Figure 2 confirm that this asymptotical law is independent of the choice of q, provided $q < 1$. Note indeed that if we run the BGA with $q = 1$ in the update (1), the convergence value is always one of the initial values, sampled according to a uniform distribution. This implies that, if $q = 1$, then $\mathbb{E}[\beta(\infty)] = \text{Var}(x_u(0)) = 1/12$. If instead $q \in (0, 1)$, simulations in Figure 3 show that the bias $\mathbb{E}[\beta(\infty)]$ is an increasing function of the mixing parameter.
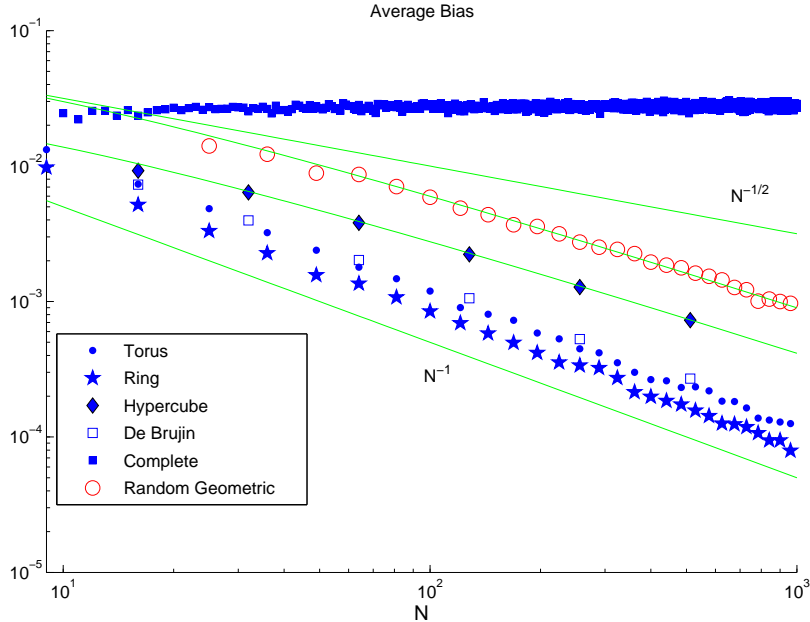

Figure 1: Average bias $\beta(\infty)$ as a function of the graph size N, for $q = 0.5$ and on different topologies (various marks). Solid lines represent $\Theta(N^{-1/2})$, $\Theta(\log N/N)$, $\Theta(\log N/N)$ and $\Theta(N^{-1})$, respectively.

---

[1]If the topology is random, namely a random geometric topology as described above, it is also sampled at this stage. Disconnected realizations are discarded: however, disconnected realizations are very few in our random geometric setting and their number decreases as N grows, so that they are less than 2% when $N > 50$.
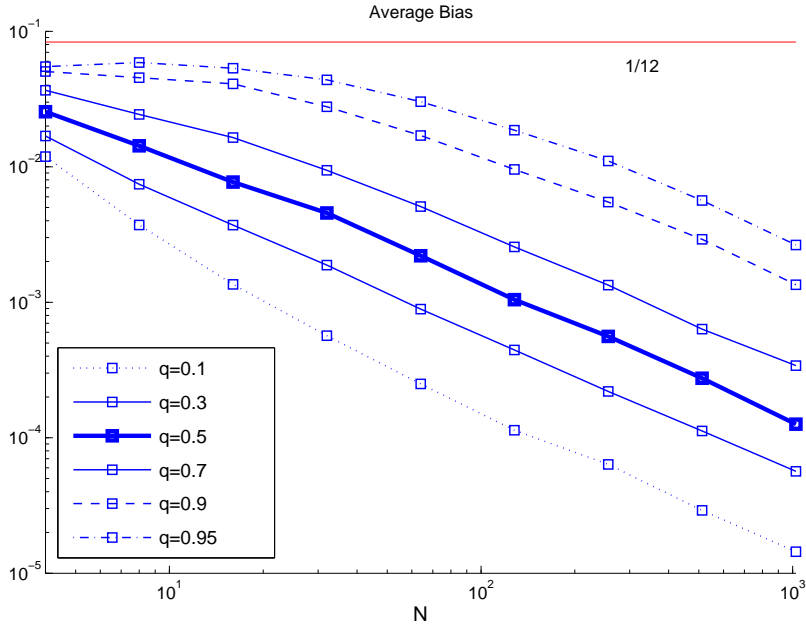
Figure 2: Average bias $\beta(\infty)$ as a function of the size N, on a sequence of de Bruijn graphs on 2 symbols, for different values of q. The solid horizontal line represents the theoretical value $\beta(\infty) = 1/12$ obtained for $q = 1$.

6   Conclusion

In this paper, we have proven that on any symmetric graph, $\mathbb{E}[\beta(\infty)] = O\left(\frac{d_{\max}^2}{N\lambda_1}\right)$, and in particular the BGA is asymptotically unbiased on expander graphs. On the other hand, simulations suggest that, on sequences of (almost) regular graphs with degree $d(N)$, the bias is such that $\mathbb{E}[\beta(\infty)] = \Theta\left(\frac{d(N)}{N}\right)$. Our future research will be devoted to find a bound on $\mathbb{E}[\beta(\infty)]$ which ensure asymptotic unbiasedness on a wider set of topologies, and more in general to study the trade-offs between speed and accuracy in gossip algorithms.

Acknowledgements

References

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

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

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.
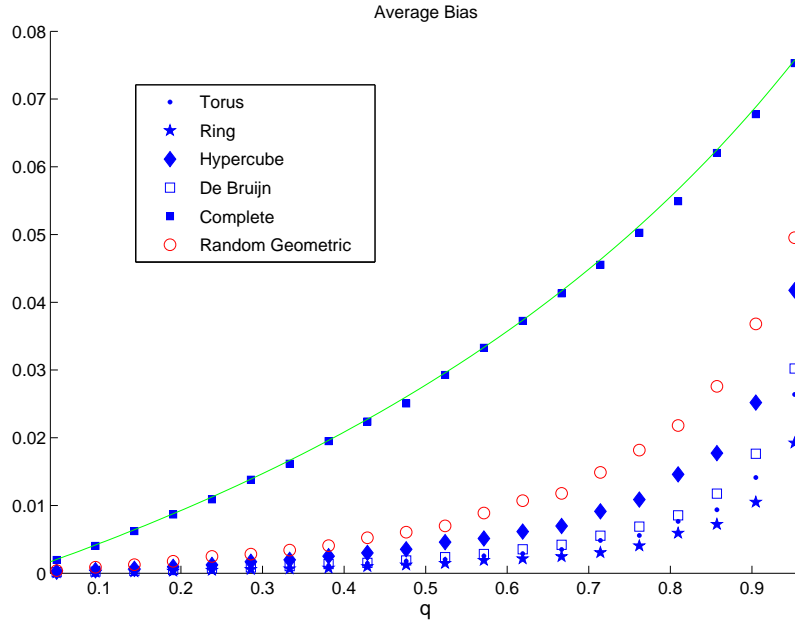
Figure 3: Average bias $\beta(\infty)$ as a function of the mixing parameter q, for $N = 64$ and on different topologies (various marks). The solid line is (7), confirming the theory in the complete case.

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

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

[6] 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*, 99(11):1847–1864, 2010.

[7] F. Fagnani and P. Frasca. Broadcast gossip averaging algorithms: interference and asymptotical error in large networks. *IEEE Journal of Selected Topics in Signal Processing*, May 2010. Submitted.

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

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

[10] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri. Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19(16):1787–1816, 2009.

[11] J. Jacod and P. Protter. *Probability Essentials*. Springer, 2003.

[12] M. Penrose. *Random Geometric Graphs*. Oxford Studies in Probability. Oxford University Press, 2003.

[13] A. N. Shiryaev. *Probability*. Graduate Texts in Mathematics. Springer, 2 edition, 1989.