Ergodic Randomized Algorithms and Dynamics over Networks

Chiara Ravazzi, *Member, IEEE*, Paolo Frasca, *Member, IEEE*, Roberto Tempo, *Fellow, IEEE*, and Hideaki Ishii, *Senior Member, IEEE*

***Abstract*—Algorithms and dynamics over networks often involve randomization and randomization can induce oscillating dynamics that fail to converge in a deterministic sense. Under assumptions of independence across time and linearity of the updates, we show that the oscillations are ergodic if the expected dynamics is stable. We apply this result to three problems of network systems, namely the estimation from relative measurements, the PageRank computation, and the dynamics of opinions in social networks. In these applications, the randomized dynamics is the asynchronous counterpart of a deterministic (stable) synchronous one. By ergodicity, the deterministic limit can be recovered via a time-averaging operation, which can be performed locally by each node of the network.**

I. Introduction

Randomization has proved to be a useful ingredient for effective algorithms in control and optimization, as reviewed in [1]. In network dynamics, randomization is specially natural, either by the uncertain nature of the network at hand, or by a design aimed at improving performance and robustness.

In this work, we focus on a class of randomized affine dynamics that do not have equilibria but are stable on average. This stability property ensures that the dynamics, although it features persistent random oscillations, has an ergodic behavior. Our main contribution is precisely to prove two ergodicity results that can be readily applied to several network-based dynamics, where randomization apparently prevents convergence. As a consequence, the desired convergence property – holding in expectation– can be recovered by each node through a process of time-averaging. Remarkably, time-averages can be computed locally by each node and, in some cases, even without access to a common clock.

In this paper, we consider dynamics where nodes interact in randomly chosen pairs, following a *gossip* approach which has been popularized in the field of control by [2] and has been followed in several papers. Indeed, many network algorithms can be randomized in such a way that the randomized dynamics converges (almost surely) to the same limit of the synchronous dynamics. Notable examples include consensus algorithms, studied in many papers as [3], [4], [5], [6], and other algorithms for estimation and classification [7] and for optimal deployment of robotic networks [8]. Nevertheless, examples of randomized algorithms that do not converge also have recently appeared in the literature. Such algorithms require some sort of additional "smoothing" operation in order to converge: in our approach, this goal is achieved by time-averaging.

The first instance we consider involves the problem of distributed estimation from relative measurements, which has applications from self-localization in robotic networks to synchronization in networks of clocks and to phase estimation in power grids. This problem was formulated in the context of clock synchronization [9] and then studied in much detail in [10], [11], [12], [13], [14], [15], where both fundamental performance limitations and distributed algorithms have been presented. More recently, randomized algorithms have been proposed by several researchers [16], [17]. Regarding this problem, our contribution includes a randomized asynchronous algorithm, in which nodes update in pairs in a gossip fashion. A related but different randomized algorithm for least-squares estimation has been recently proposed in [18].

The second example is PageRank computation, which has attracted much attention in recent years for the importance of its applications, e.g., in the search engine of Google [19], [20], [21], and for its similarities with the consensus problem illustrated in [22]. Randomized algorithms for PageRank computation have been studied in a series of papers, including [23], [24], [25], [26]. Other recent references on PageRank are listed in [27], [28], [29]. Our contribution provides a general convergence result for randomized algorithms, which we apply to a novel pair-wise gossip algorithm in Section IV.

The third example comes from social sciences and specifically from the mechanisms of opinion evolution. Indeed, opinion dynamics models, where agents have some degree of obstinacy and interactions are randomized, give rise to ergodic oscillations. This observation has been made in [30] and here we extend it to the opinion model [31]. We propose in Section V a gossip mechanism of opinion update, which allows us to interpret the classical opinion dynamics –which makes simplistic assumptions on the communication process among individuals– as the "average" evolution of our randomized model. This observation answers an open question on modeling the communication process that was raised in the original paper [31].

In the area of systems and control, time-averaging has been

This work was supported in part by the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n.257462 HYCON2 Network of Excellence and in part by JST, CREST.

Chiara Ravazzi is with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Italy. E-mail: chiara.ravazzi@polito.it

Paolo Frasca is with the Department of Applied Mathematics, University of Twente, The Netherlands. E-mail: p.frasca@utwente.nl

Roberto Tempo is with CNR-IEIIT, Politecnico di Torino, Italy. E-mail: roberto.tempo@polito.it

Hideaki Ishii is with the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Japan. E-mail: ishii@dis.titech.ac.jp

widely employed in optimization problems (see, e.g., [32]). For distributed optimization algorithms such as those in [33], [18], error bounds are established for the time-average, which can be computed from locally available information only. In the randomized distributed PageRank algorithm [23] discussed above, time-average of local states has also been used in a slightly different manner; this work has motivated us to study from a more general viewpoint other multi-agent type problems to which the technique can be applied.

Preliminary versions of some of our results have been reported in the proceedings of conferences as [34], [35], regarding relative localization, and as [36], regarding opinion dynamics. The current presentation incorporates and builds upon the previous ones, includes the PageRank computation, and most importantly embeds them into a comprehensive framework that is suitable for the study of other applications.

A. Paper outline

In Section II we study asynchronous dynamics over networks obtained by randomization of deterministic synchronous dynamics. The main results, Theorem 1 and Theorem 2, provide sufficient conditions for ergodicity and are subsequently used for three applications coming from different areas: relative localization (Section III), PageRank computation (Section IV), and opinion dynamics (Section V). Section VI contains the technical derivation of our main results. Additional remarks and research outlooks are given in a concise Section VII.

B. Notation and preliminaries

Throughout this paper, we use the following notation. Real and nonnegative integer numbers are denoted by $\mathbb{R}$ and $\mathbb{Z}_+$, respectively. The symbol $|\cdot|$ denotes either the cardinality of a set or the absolute value of a real number. The symbol e_i is the vector with the i-th entry equal to 1 and all the other elements equal to 0, and we write $\mathbf{1}$ for the vector with all entries equal to 1. A vector x is stochastic if its entries are nonnegative and $\sum_i x_i = 1$. A matrix A is row-stochastic (column-stochastic) when its entries are nonnegative and $M\mathbf{1} = \mathbf{1}$ ($M^\top \mathbf{1} = \mathbf{1}$). A matrix is doubly stochastic when it is both row and column-stochastic. A matrix P is said to be Schur stable if the absolute value of all its eigenvalues is smaller than 1. A directed graph is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. We say that $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph if $(u, v) \in \mathcal{E}$ implies that (v, u) is also an edge in $\mathcal{E}$. To avoid trivialities, we implicitly assume that graphs have at least three nodes, i.e., $|\mathcal{V}| > 2$. A directed graph $\mathcal{G}$ is called strongly connected if there is a path from each vertex in the graph to every other vertex. To any matrix $P \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ with non-negative entries, we can associate a directed graph $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}_P)$ by putting $(i, j) \in \mathcal{E}_P$ if and only if $P_{ij} > 0$. The matrix P is said to be adapted to graph $\mathcal{G}$ if $\mathcal{G}_P \subseteq \mathcal{G}$.

II. Ergodic randomized dynamics over networks

Consider the affine dynamics representing a time-invariant discrete-time dynamical system over a network, described by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n nodes, with state $x(k) \in \mathbb{R}^{\mathcal{V}}$, $k \in \mathbb{Z}_+$

$$x(k+1) = Px(k) + u, \qquad (1)$$

where the matrix $P \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is adapted to the graph $\mathcal{G}$ and $u \in \mathbb{R}^{\mathcal{V}}$ is a constant input. We have the following simple fact.

Proposition 1. *If P is Schur stable, then the dynamics in* (1) *converges to*

$$x^\star = (I - P)^{-1} u$$

for any initial conditions $x(0) = x_0$.

In this paper we are interested in suitable randomized versions of the dynamics in (1). More precisely, let $\{\theta(k)\}_{k\in\mathbb{Z}_+}$ be a sequence of independent identically distributed (i.i.d.) random variables taking values in a finite set Θ. Given a realization $\theta(k)$, $k \in \mathbb{Z}_+$, we associate to it a matrix $P(k) = P(\theta(k)) \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ and an input vector $u(k) = u(\theta(k)) \in \mathbb{R}^{\mathcal{V}}$, obtaining a time-varying discrete-time dynamical system of the form

$$x(k+1) = P(k)x(k) + u(k), \qquad (2)$$

with initial condition $x(0) \in \mathbb{R}^{\mathcal{V}}$. We observe that the state $\{x(k)\}_{k\in\mathbb{Z}_+}$ is a Markov process because, given the current position of the chain, the conditional distribution of the future values does not depend on the past values.

It may happen that the dynamics (2) oscillates persistently and fails to converge in a deterministic sense: this behavior is apparent in the examples we show in the next section. In view of this fact, we give simple conditions which guarantee probabilistic convergence (formally defined subsequently) to the vector $x^\star$ given in Proposition 1. We say that the process $\{x(k)\}_{k\in\mathbb{Z}_+}$ is *ergodic* if there exists a random variable $x_\infty \in \mathbb{R}^{\mathcal{V}}$ such that almost surely

$$\lim_{k\to\infty} \frac{1}{k} \sum_{\ell=0}^{k-1} x(\ell) = \mathbb{E}[x_\infty]. \qquad (3)$$

The time-average in (3) is called Cesáro average or Polyak average in some contexts [32]. The closely related definition of *mean-square ergodicity* instead requires $\lim_{k\to\infty} \mathbb{E}\left[\left\|\frac{1}{k}\sum_{\ell=0}^{k-1} x(\ell) - \mathbb{E}[x_\infty]\right\|_2^2\right] = 0$. In what follows, we mostly focus on almost-sure ergodicity: indeed, it is often possible to deduce mean-square convergence from almost sure convergence. We now establish our first result for randomized dynamics over networks.

Theorem 1 (Ergodicity of affine dynamics)**.** *Consider the random process $\{x(k)\}_{k\in\mathbb{Z}_+}$ defined in (2), where $\{P(k)\}_{k\in\mathbb{Z}_+}$ and $\{u(k)\}_{k\in\mathbb{Z}_+}$ are i.i.d. and have finite first moments. If there exists $\alpha \in (0, 1]$ such that*

$$\mathbb{E}[P(k)] = (1-\alpha)I + \alpha P, \qquad \mathbb{E}[u(k)] = \alpha u, \qquad (4)$$

where P and u are given in Proposition 1, then

1) *$x(k)$ converges in distribution to a random variable x_∞, and the distribution of x_∞ is the unique invariant distribution for* (2)*;*
2) *the process is* ergodic*;*

3) *the limit random variable satisfies* $\mathbb{E}[x_\infty] = x^\star$.

We postpone the proof of Theorem 1 to Section VI-A, but we make the following insightful observation. Under the assumptions of the theorem, P is Schur stable in Proposition 1 and so is $\mathbb{E}[P(k)]$. Consequently,

$$\begin{aligned}\mathbb{E}[x(k+1)] &= \mathbb{E}[P(k)]\mathbb{E}[x(k)] + \mathbb{E}[u(k)]\\ &= ((1-\alpha)I + \alpha P)\mathbb{E}[x(k)] + \alpha u,\end{aligned}$$

and $\lim_{k\to\infty} \mathbb{E}[x(k)] = x^\star$. The expected dynamics of the process (2) can indeed be interpreted as a "lazy" (slowed down) version of the synchronous dynamics (1) associated to the matrix P.

In our applications, we aim at approximating $x^\star$ but we do not necessarily have the opportunity to sample $x(k)$ at all times. The following refinement of Theorem 1 –proved in Section VI-B– allows for sampling over a (random) subsequence of times and is thus specially useful to our purposes.

Theorem 2 (Ergodicity of affine dynamics on random subsequences)**.** *Consider the random process* $\{x(k)\}_{k\in\mathbb{Z}_+}$ *defined in (2), where* $\{P(k)\}_{k\in\mathbb{Z}_+}$ *and* $\{u(k)\}_{k\in\mathbb{Z}_+}$ *are i.i.d. and have finite first moments. Let* $\{\omega(k)\}_{k\in\mathbb{Z}_+} \in \{0,1\}^{\mathbb{Z}_+}$ *be an i.i.d. random sequence such that, for all k,* $\omega(k)$ *is independent of* $P(\ell)$ *for all* $\ell < k$ *and* $\omega(k) \neq 0$ *with positive probability. If there exists* $\alpha \in (0,1]$ *such that*

$$\mathbb{E}[P(k)] = (1-\alpha)I + \alpha P, \qquad \mathbb{E}[u(k)] = \alpha u, \tag{5}$$

where P *and* u *are given in Proposition 1, then almost surely*

$$\lim_{k\to\infty} \frac{1}{\sum_{i=0}^{k-1}\omega(i)} \sum_{\ell=0}^{k-1} \omega(\ell)x(\ell) = x^\star.$$

In the following sections, we apply Theorem 1 and Theorem 2 to specific randomized dynamics in sensor localization, PageRank computation, and opinion dynamics. Even though these applications are quite diverse, we show that classical algorithms for their solutions can be represented by the affine dynamics (1) and their randomized versions by (2).

III. Sensor localization in wireless networks

In sensor localization in wireless networks, we seek to estimate the relative position of sensors using noisy relative measurements. We formulate the problem using an oriented graph[1] $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node i in $\mathcal{V}$ has to estimate its own variable s_i, knowing only noisy measurements $b_{(i,j)}$ of some difference with neighboring edges

$$b_{(i,j)} = s_i - s_j + \eta_{(i,j)} \qquad \text{if } (i,j) \in \mathcal{E} \text{ or } (j,i) \in \mathcal{E}$$

where $\eta_{(i,j)}$ is additive noise. The graph topology is encoded in the incidence matrix $A \in \{0, \pm 1\}^{\mathcal{E}\times\mathcal{V}}$ defined by

$$A_{ei} = \begin{cases} +1 & \text{if } e = (i,j)\\ -1 & \text{if } e = (j,i)\\ 0 & \text{otherwise}\end{cases}$$

[1] An oriented graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a directed graph such that $(i,j) \in \mathcal{E}$ only if $i < j$. $\mathcal{G}$ is said to be weakly connected if the graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ where $\mathcal{E}' = \{(h,k) \in \mathcal{V}\times\mathcal{V} : \text{ either } (h,k) \in \mathcal{E} \text{ or } (k,h) \in \mathcal{E}\}$ has a path which connects every pair of nodes.

for every $e \in \mathcal{E}$. We can collect all the measurements and variables in vectors $b \in \mathbb{R}^{\mathcal{E}}$ and $s \in \mathbb{R}^{\mathcal{V}}$, so that

$$b = As + \eta,$$

where $\eta \in \mathbb{R}^{\mathcal{E}}$. A least-squares approach can be used to determine the optimal estimate $x^\star_{\text{loc}} = \min_{z\in\mathbb{R}^{|\mathcal{V}|}} \|Az - b\|_2^2$ of the state s based on the measurements b. Given a weakly connected oriented graph $\mathcal{G}$ with incidence matrix A, the least square estimation of s is given by $x^\star_{\text{loc}} = L^\dagger A^\top b$, where $L^\dagger$ denotes the Moore-Penrose pseudo-inverse of the Laplacian $L := A^\top A$. Notice that the Laplacian L is not full rank, hence we need $L^\dagger$. The solution $x^\star_{\text{loc}}$ can be easily computed by an iterative gradient algorithm [34], which takes the following form. Given a parameter $\tau > 0$ and the initial condition $x(0) = 0$, we let

$$x(k+1) = (I - \tau L)x(k) + \tau A^\top b \tag{6}$$

where the matrix $I - \tau L$ is doubly stochastic. After some manipulations, equation (6) can be written in the form in (1) taking

$$P = (I - \tau L)\Omega \quad \text{and} \quad u = \tau A^\top b, \tag{7}$$

with $\Omega = I - |\mathcal{V}|^{-1}\mathbf{1}\mathbf{1}^\top$. The gradient descent algorithm in (6) with $x(0) = 0$ converges to the optimal least-squares solution $x^\star_{\text{loc}}$ if $\tau < 2/|\mathcal{V}|$ and $\mathcal{G}$ is weakly connected. The proof is a straightforward application of Proposition 1 and can be found in [34].

We now consider the randomized algorithm of [34] to solve the sensor localization problem. For each node $i \in \mathcal{V}$, the algorithm involves a triple of states $(x_i, \kappa_i, \tilde{x}_i)$, which depend on a discrete time index $k \in \mathbb{Z}_+$. These three variables play the following roles: $x_i(k)$ is the "raw" estimate of s_i obtained by i at time k through communications with its neighbors, $\kappa_i(k)$ counts the number of updates performed by i up to time k, and $\tilde{x}_i(k)$ is the "smoothed" estimate obtained through time-averaging. The algorithm is defined by choosing a scalar parameter $\gamma \in (0,1)$ and a sequence of random variables $\{\theta(k)\}_{k\in\mathbb{Z}_+}$ taking values in $\mathcal{E}$. The state variables are initialized as $(x_i(0), \kappa_i(0), \tilde{x}_i(0)) = (0,0,0)$ for all i, and at each time $k > 0$, provided that $\theta(k) = (i,j)$, the states are updated according to the following recursions, namely the raw estimates as

$$\begin{aligned} x_i(k+1) &= (1-\gamma)x_i(k) + \gamma x_j(k) + \gamma b_{(i,j)}\\ x_j(k+1) &= (1-\gamma)x_j(k) + \gamma x_i(k) - \gamma b_{(i,j)}\\ x_\ell(k+1) &= x_\ell(k) \qquad \text{if } \ell \notin \{i,j\};\end{aligned} \tag{8a}$$

the local times as

$$\begin{aligned} \kappa_i(k+1) &= \kappa_i(k) + 1\\ \kappa_j(k+1) &= \kappa_j(k) + 1\\ \kappa_\ell(k+1) &= \kappa_\ell(k) \qquad \text{if } \ell \notin \{i,j\};\end{aligned} \tag{8b}$$

and the time-averages as

$$\begin{aligned} \widetilde{x}_i(k+1) &= \frac{1}{\kappa_i(k+1)}\big(\kappa_i(k)\widetilde{x}_i(k) + x_i(k+1)\big)\\ \widetilde{x}_j(k+1) &= \frac{1}{\kappa_j(k+1)}\big(\kappa_j(k)\widetilde{x}_j(k) + x_j(k+1)\big)\\ \widetilde{x}_\ell(k+1) &= \widetilde{x}_\ell(k) \qquad \text{if } \ell \notin \{i,j\}.\end{aligned} \tag{8c}$$

Next, we assume the sequence $\{\theta(k)\}_{k\in\mathbb{Z}_+}$ to be i.i.d. and its probability distribution to be *uniform*, i.e.,

$$\mathbb{P}[\theta(k)=(i,j)]=\frac{1}{|\mathcal{E}|}, \qquad \forall k\in\mathbb{Z}_+. \tag{9}$$

Remark 1 (Local and global clocks)**.** *It should be noted that the time index k counts the number of updates which have occurred in the network, whereas for each $i\in\mathcal{V}$ the variable $\kappa_i(k)$ is the number of updates involving i up to the current time. Hence, κ_i is a local variable which is inherently known to agent i, even if a common clock k is unavailable. Therefore, this algorithm is totally asynchronous and fully distributed, in the sense that the updates, including the time-averaging process, do not require the nodes to be aware of a common clock. This feature is especially attractive if the algorithm is applied to clock synchronization problems. These problems have recently attracted much interest in systems and control: see for instance [37], [38], [39].*

The dynamics in (8a) oscillates persistently and fails to converge in a deterministic sense, as shown in [35]. However, the oscillations asymptotically concentrate around the solution of the least-squares problem, as it is formally stated in the following result, which shows that the sample dynamics is well-represented by the average one. This indicates that $\widetilde{x}_i(k)$ is "the right variable" to approximate the optimal estimate $x^\star_{\rm loc}$ because the process $x(k)$ is ergodic. In the proof of the proposition, we show that the dynamics in equation (8a) can be written in terms of the more general process (2).

Proposition 2 (Ergodicity of sensor localization)**.** *The dynamics in* (8) *with uniform selection* (9) *is such that* $\lim_{k\to\infty}\widetilde{x}(k)=x^\star_{\rm loc}$ *almost surely.*

Proof: We rewrite the dynamics of (8a) as

$$x(k+1)=Q(k)x(k)+u(k) \tag{10}$$

and, provided $\theta(k)=(i,j)$, we define

$$Q(k)=I-\gamma(e_i-e_j)(e_i-e_j)^\top$$

and $u(k)=b_{\theta(k)}(e_i-e_j)$, where the vector e_i is defined in the preliminaries. We note that for all k the matrix $Q(k)$ is doubly stochastic and the sum of the elements in $u(k)$ is zero: in particular, given $x(0)=0$, then $\mathbf{1}^\top x(k)=0$ for each $k\in\mathbb{Z}_+$. These observations further imply that the dynamics of $x(k)$ is equivalently described by the iteration

$$x(k+1)=Q(k)\Omega x(k)+u(k). \tag{11}$$

Letting $P(k)=Q(k)\Omega$, the dynamics of the algorithm is cast in the form of (2). Next, using the uniform distribution (9), we compute

$$\mathbb{E}[P(k)]=\left(I-\gamma\frac{L}{|\mathcal{E}|}\right)\Omega, \quad \mathbb{E}\left[u(k)\right]=\gamma\frac{A^\top b}{|\mathcal{E}|},$$

and observe that $\mathbb{E}[P(k)]$ satisfies the ergodicity condition in Theorem 1 with P and u defined in (7), $\alpha=1$ and $\tau=\gamma/|\mathcal{E}|$. If we define, for all $i\in\mathcal{V}$ and all $k\in\mathbb{Z}_+$,

$$\omega_i(k)=\begin{cases}1 & \text{if } \theta(k)=(i,j) \text{ or } \theta(k)=(j,i)\\ 0 & \text{otherwise}\end{cases}$$

then $\kappa_i(k+1)=\kappa_i(k)+\omega_i(k)=\sum_{\ell=0}^k\omega_i(\ell)$ and

$$\widetilde{x}_i(k+1)=\frac{1}{\sum_{\ell=0}^k\omega_i(\ell)}\sum_{\ell=0}^k\omega_i(\ell)x_i(\ell).$$

Since $\{\omega(k)\}_k$ is an i.i.d. random sequence, by Theorem 2 we conclude our argument. ■

It is also true that $\widetilde{x}(k)$ converges to $x^\star_{\rm loc}$ in the mean-square sense. A proof can be obtained with similar arguments as in [34] and is not detailed here.

IV. PageRank computation in Google

In this section, we study a network consisting of web pages [19]. This network can be represented by a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where the set of vertices correspond to the web pages and edges represent the links between the pages, i.e., the edge $(i,j)\in\mathcal{E}$, if page i has an outgoing link to page j, or in other words, page j has an incoming link from page i.

The goal of the PageRank algorithm is to provide a measure of relevance of each web page: the PageRank value of a page is a real number in $[0,1]$, which is defined next. Let us denote $\mathcal{N}_i=\{h\in\mathcal{V}\,:\,(i,h)\in\mathcal{E}\}$ and $n_i=|\mathcal{N}_i|$, for each node $i\in\mathcal{V}$, and $A\in\mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ the matrix such that

$$A_{ij}=\begin{cases}1/n_j & \text{if } j\in\mathcal{N}_i\\ 0 & \text{otherwise.}\end{cases}$$

Let $m\in(0,1)$ and recall $n=|\mathcal{V}|$, and define

$$M=(1-m)A+\frac{m}{n}\mathbf{1}\mathbf{1}^\top. \tag{12}$$

The PageRank of the graph $\mathcal{G}$ is the vector $x^\star_{\rm pgr}$ such that $Mx^\star_{\rm pgr}=x^\star_{\rm pgr}$ and $\sum_i x^\star_{{\rm pgr}\,i}=1$.

Given the initial condition such that $\mathbf{1}^\top x(0)=1$ (i.e., it is a stochastic vector), the PageRank vector can be computed through the recursion

$$x(k+1)=Mx(k)=(1-m)Ax(k)+\frac{m}{n}\mathbf{1}. \tag{13}$$

In this case, we observe that the PageRank vector can be represented in terms of the affine dynamics (1) simply taking

$$P=(1-m)A \quad \text{and} \quad u=\frac{m}{n}\mathbf{1}. \tag{14}$$

Using Proposition 1, it can be proved that for any initial condition $x(0)\in\mathbb{R}^{\mathcal{V}}$ such that $\mathbf{1}^\top x(0)=1$, the sequence in (13) converges to

$$x^\star_{\rm pgr}=(I-(1-m)A)^{-1}\frac{m}{n}\mathbf{1}. \tag{15}$$

Further detail on this convergence result can be found in [40].

We now describe a new example of an "edge-based" randomized gossip algorithm. Its motivation comes from the interest in reducing the coordination effort required by the network at each iteration: if only one edge is activated at each time, this effort is minimal. Each node $i\in\mathcal{V}$ holds a couple of states $(x_i,\overline{x}_i)$. For every time step k an edge $\theta(k)$ is sampled from a uniform distribution over $\mathcal{E}$ (note that sampling is independent at each time k). Then, the states are updated as follows:

$$x_i(k+1) = (1-r)\left(1-\frac{1}{n_i}\right)x_i(k)+\frac{r}{n} \tag{16a}$$

$$x_j(k+1) = (1-r)\left(x_j(k)+\frac{1}{n_i}x_i(k)\right)+\frac{r}{n} \tag{16b}$$

$$x_h(k+1) = (1-r)\,x_h(k)+\frac{r}{n} \qquad \text{if } h\neq i,j \tag{16c}$$

and

$$\overline{x}_\ell(k+1)=\frac{kx_\ell(k)+x_\ell(k+1)}{k+1} \quad \forall\,\ell\in\mathcal{V} \tag{17}$$

where $r\in(0,1)$ is a design parameter to be determined. The update in (16) can also be formally rewritten in vector-wise form as (2) with

$$P(k)=(1-r)A(k), \quad u(k)=\frac{r}{n}\mathbf{1}.$$

Here $A(k)$ and $P(k)$ are random matrices which are determined by the choice of $\theta(k)=(i,j)$

$$A(k)=I+\frac{1}{n_i}(e_je_i^\top-e_ie_i^\top).$$

Then, $A(k)$ is uniformly distributed over the set of matrices $\{I+\frac{1}{n_i}(e_je_i^\top-e_ie_i^\top)\,:\,(i,j)\in\mathcal{E}\}$.

Remark 2 (Local and global clocks)**.** *We note that, contrary to* (8)*, the dynamics described by* (16)-(17) *does require the nodes to access the global time variable k. The reason for this synchrony requirement comes from the need to preserve the stochasticity of the vector $x(k)$. We believe this is a reasonable assumption, because these algorithms need to be implemented on webpages or domain servers which are typically endowed with clocks.*

In the next result, we state convergence of this algorithm.

Proposition 3 (Ergodic PageRank convergence)**.** *Let us consider the dynamics* (16)-(17) *with*

$$r=\frac{m}{m-|\mathcal{E}|m+|\mathcal{E}|}$$

where $x(0)$ is a stochastic vector. Then, the sequence $\{\overline{x}(k)\}_{\mathbb{Z}_+}$ is such that $\lim_{k\to\infty}\overline{x}(k)=x^\star_{\text{pgr}}$ almost surely.

Proof: For each $k\in\mathbb{Z}_+$, we have

$$\begin{aligned}\mathbb{E}[A(k)] &= \frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}}\left(I+\frac{1}{n_i}(e_je_i^\top-e_ie_i^\top)\right)\\ &= I+\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}}\frac{1}{n_i}e_je_i^\top-\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}}\frac{1}{n_i}e_ie_i^\top\\ &= I+\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}}\frac{1}{n_i}e_je_i^\top-\frac{1}{|\mathcal{E}|}\sum_{i\in V}\sum_{j\in\mathcal{N}_i}\frac{1}{n_i}e_ie_i^\top\\ &= I+\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}}A_{ji}e_je_i^\top-\frac{1}{|\mathcal{E}|}\sum_{i\in V}\frac{n_i}{n_i}e_ie_i^\top\\ &= \left(1-\frac{1}{|\mathcal{E}|}\right)I+\frac{1}{|\mathcal{E}|}A.\end{aligned}$$

It should be noted that, setting $\alpha=(m-m|\mathcal{E}|+|\mathcal{E}|)^{-1}$ and P and u as in (14),

$$\mathbb{E}[P(k)]=(1-r)\mathbb{E}[A(k)]=(1-\alpha)\,I+\alpha P,$$

and $\mathbb{E}[u(k)]=\alpha\frac{m}{n}\mathbf{1}=\alpha u$. From Theorem 1 we conclude the almost sure convergence. ■

The almost sure convergence can also be proved by techniques from stochastic approximation. Such techniques have already been effectively applied to other algorithms for PageRank computation [26].

Since $x(k)$ are stochastic vectors, they are uniformly bounded and by the Dominated Convergence Theorem we conclude the convergence in the mean-square sense. Mean-square ergodicity of randomized PageRank was already proved in [40] under assumptions which are equivalent to those in Theorem 1.

V. Opinion dynamics in social networks

In this application, we study a classical model introduced in [31] to describe the effect of social influence and prejudices in the evolution of opinions in a population in the presence of the so-called stubborn agents. We briefly review and cast this model into the general framework of affine dynamics (1).

We consider a finite population $\mathcal{V}$ of interacting agents, whose *social network* of potential interactions is encoded by an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$, endowed with a self-loop (i,i) at every node. At time $k\in\mathbb{Z}_+$, each agent $i\in\mathcal{V}$ holds a *belief or opinion* about an underlying state of the world. We denote the vector of beliefs as $x(k)\in\mathbb{R}^{\mathcal{V}}$. An edge $(i,j)\in\mathcal{E}$ means that agent j may directly influence the opinion of agent i. Let $W\in\mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ be a nonnegative matrix which defines the strength of the interactions ($W_{ij}=0$ if $(i,j)\notin\mathcal{E}$) and Λ be a diagonal matrix describing how sensitive each agent is to the opinions of the others based on interpersonal influences. We assume that W is row-stochastic, i.e., $W\mathbf{1}=\mathbf{1}$ and we set $\Lambda=I-\mathrm{diag}(W)$, where $\mathrm{diag}(W)$ collects the self-weights given by the agents. The dynamics of opinions $x(k)$ proposed in [31] is

$$x(k+1)=\Lambda Wx(k)+(I-\Lambda)v \tag{18}$$

where $x(0)=v$ and $v\in\mathbb{R}^{\mathcal{V}}$. The vector v, which corresponds to the individuals' preconceived opinions, also appears as an input at every time step. This model falls under the class of affine dynamics (1) by simply taking

$$P=\Lambda W \quad \text{and} \quad u=(I-\Lambda)v. \tag{19}$$

The limit behavior of (18) is described in [36]. In particular, if we assume that in the graph associated to W for any node $\ell\in\mathcal{V}$ there exists a path from ℓ to a node i such that $W_{ii}>0$. Then, the opinions converge to

$$x^\star_{\text{opd}}=(I-\Lambda W)^{-1}(I-\Lambda)v. \tag{20}$$

We remark that the assumption on the existence of the path implies that each agent is influenced by at least one stubborn agent and is automatically satisfied if the graph is strongly connected. In practice, it is reasonable to think that most agents in a social network will have some (positive) level of obstinacy $W_{ii} > 0$.

We introduce now a more realistic model of the communication process among the agents. Each agent $i \in \mathcal{V}$ possesses an initial belief $x_i(0) = v_i \in \mathbb{R}$, as in the model (18). At each time $k \in \mathbb{Z}_+$ a link is randomly sampled from a uniform distribution over $\mathcal{E}$. If the edge (i, j) is selected at time k, agent i meets agent j and updates its belief to a convex combination of its previous belief, the belief of j, and its initial belief. Namely,

$$\begin{aligned} x_i(k+1) &= h_i\big((1-\Gamma_{ij})x_i(k) + \Gamma_{ij}x_j(k)\big) + (1-h_i)v_i \\ x_\ell(k+1) &= x_\ell(k) \qquad \forall \ell \in \mathcal{V} \setminus \{i\}, \end{aligned} \tag{21}$$

where the weighting coefficients h_i and Γ_{ij} are defined as

$$h_i = \begin{cases} 1-(1-\lambda_i)/d_i & \text{if } d_i \neq 1 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

$$\Gamma_{ij} = \begin{cases} \frac{d_i(1-h_i)+h_i-(1-\lambda_i W_{ii})}{h_i} & \text{if } i = j,\ d_i \neq 1 \\ \frac{\lambda_i W_{ij}}{h_i} & \text{if } i \neq j,\ d_i \neq 1 \\ 1 & \text{if } i = j,\ d_i = 1 \\ 0 & \text{if } i \neq j,\ d_i = 1 \end{cases} \tag{23}$$

where the matrices W and Λ are those in (18), λ is the main diagonal of Λ, and $d_i = |\{h : (i,h) \in \mathcal{E}\}|$ is the degree of node i. Recall that $d_i \geq 1$ by the presence of self-loops. It is immediate to observe that (a) $h_i \in [0,1]$ for all $i \in \mathcal{V}$; (b) Γ is adapted to the graph $\mathcal{G}$; (c) Γ is row-stochastic; and (d) at all times the opinions of the agents are convex combinations of their initial prejudices.

We now study the convergence properties of the gossip opinion dynamics and we show that the opinions converge to the same value $x^\star_{\text{opd}}$ given in (20). In the proof of the result, we show that the dynamics (21) can be written in terms of the more general process (2).

Proposition 4 (Ergodic opinion dynamics)**.** *Assume that in the graph associated to W for any node $\ell \in \mathcal{V}$ there exists a path from ℓ to a node i such that $W_{ii} > 0$. Then, the dynamics* (21) *is ergodic, and the time-averaged opinions defined in* (3) *converge to $x^\star_{\text{opd}}$.*

Proof: Provided the edge $\theta(k) = (i, j)$ is chosen at time k, the dynamics (21) can be rewritten in vector form as

$$\begin{aligned} x(k+1) =& (I - e_i e_i^\top (I-H))\left(I + \Gamma_{ij}(e_i e_j^\top - e_i e_i^\top)\right) x(k) \\ &+ e_i e_i^\top (I-H)v. \end{aligned}$$

If we define the matrices

$$\begin{aligned} P^{ij} =& (I - e_i e_i^\top (I-H))\left(I + \Gamma_{ij}(e_i e_j^\top - e_i e_i^\top)\right) \\ u^{ij} =& e_i e_i^\top (I-H)v, \end{aligned}$$

then the dynamics is $x(k+1) = P^{ij}x(k) + u^{ij}$. Note that the expressions in (22) and (23) imply

$$\begin{aligned} D(I-H) &= I - \Lambda \\ D(I-H) + H(I-\Gamma) &= I - \Lambda W \end{aligned}$$

where $H = \text{diag}\{h_1, h_2, \ldots, h_n\}$ and $D = \text{diag}\{d_1, d_2, \ldots, d_n\}$. Consequently, one can compute the generic entries of the expected matrix $\mathbb{E}[P(k)] = \frac{1}{|\mathcal{E}|}\sum_{(\ell,m)\in\mathcal{E}} P^{\ell m}$ as $\mathbb{E}[P(k)]_{ij} = \frac{1}{|\mathcal{E}|}h_i\Gamma_{ij} = \frac{1}{|\mathcal{E}|}\lambda_i W_{ij}$ if $i \neq j$, and

$$\begin{aligned} \mathbb{E}[P(k)]_{ii} &= 1 - \frac{1}{|\mathcal{E}|}\big(d_i(1-h_i) + h_i(1-\Gamma_{ii})\big) \\ &= \Big(1 - \frac{1}{|\mathcal{E}|}\Big) + \frac{1}{|\mathcal{E}|}\lambda_i W_{ii}. \end{aligned}$$

From these formulas, we conclude that $\mathbb{E}[P(k)] = (1 - \frac{1}{|\mathcal{E}|})I + \frac{1}{|\mathcal{E}|}\Lambda W$ and $\mathbb{E}[u(k)] = \frac{1}{|\mathcal{E}|}(I-\Lambda)v$. Then, using (19) the claim follows by Theorem 1. ■

Since the opinions are uniformly bounded, by the Dominated Convergence Theorem we also conclude the convergence in the mean-square sense. Mean-square ergodicity is also proved in [36] under assumptions which are equivalent to those in Theorem 1. The ergodicity of the opinion dynamics is illustrated by the simulations in [36]. Furthermore, we observe that also Theorem 2 applies to this dynamics, so that time averages can be performed asynchronously by the nodes via a mechanism like (8).

VI. Proofs of the ergodicity results

A. Proof of Theorem 1

The proof is based on techniques for iterated random functions, which we recall from [41]. These techniques require, in order to study the random process (2), to consider the associated *backward* process $\overleftarrow{x}(k)$, which we define below.

For any time instant k, consider the random matrices $P(k)$ and $u(k)$ and define the matrix product

$$\overrightarrow{P}(\ell, m) := P(m)P(m-1)\cdots P(\ell+1)P(\ell) \tag{24}$$

with $\ell \in \{0, \ldots, m\}$. Then, the iterated affine system in (11) can be rewritten as

$$x(k+1) = \overrightarrow{P}(0,k)x(0) + \sum_{0 \leq \ell \leq k} \overrightarrow{P}(\ell+1,k)u(\ell).$$

The corresponding *backward process* is defined by

$$\overleftarrow{x}(k+1) = \overleftarrow{P}(0,k)x(0) + \sum_{0 \leq \ell \leq k} \overleftarrow{P}(0,\ell-1)u(\ell),$$

where

$$\overleftarrow{P}(\ell, m) := P(\ell)P(\ell+1)\cdots P(m-1)P(m) \tag{25}$$

with $\ell \in \{0, \ldots, m\}$. Crucially, the backward process $\overleftarrow{x}(k)$ has at every time $k \in \mathbb{Z}_+$ the same probability distribution as $x(k)$. The main tool to study the backward process is the following result. Let $\|\cdot\|$ denote any norm.

Lemma 1 (Theorem 2.1 in [41])**.** *Let us consider the Markov process $\{x(k)\}_{k\in\mathbb{Z}_+}$ defined by*

$$x(k+1) = P(k)x(k) + u(k) \quad k \in \mathbb{Z}_+$$

where $P(k) \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ and $u(k) \in \mathbb{R}^{\mathcal{V}}$ are i.i.d. random variables. Let us assume that

$$\mathbb{E}[\log\|P(k)\|] < \infty \qquad \mathbb{E}[\log\|u(k)\|] < \infty. \tag{26}$$

The corresponding backward random process $\overleftarrow{x}(k)$ converges almost surely to a finite limit x_∞ if and only if

$$\inf_{k>0} \frac{1}{k}\mathbb{E}\left[\log \|P(1)\cdots P(k)\|\right] < 0. \tag{27}$$

If (27) *holds, then the distribution of x_∞ is the unique invariant distribution for the Markov process $x(k)$.*

This result provides conditions for the backward process to converge to a limit. Although the forward process has a different behavior compared to the backward process, the forward and backward processes have the same distribution. This fact allows us to determine, by studying the backward process $\overleftarrow{x}(k)$, whether the sequence of random variables $\{x(k)\}_{k\in\mathbb{Z}_+}$ converges in distribution to the invariant distribution of the Markov process in (2). This analysis is done in the following result.

Lemma 2. *Consider the random process $x(k)$ defined in* (2)*, where $P(k)$ and $u(k)$ are i.i.d. and have finite first moments $\mathbb{E}[P(k)]$ and $\mathbb{E}[u(k)]$. If there exists $\alpha \in (0,1]$ such that $\mathbb{E}[P(k)] = (1-\alpha)I + \alpha P$ where P is Schur stable, then, $\overleftarrow{x}(k)$ converges almost surely to a finite limit x_∞, and the distribution of x_∞ is the unique invariant distribution for $x(k)$.*

Proof: In order to apply Lemma 1, let us compute

$$\begin{aligned}
&\inf_{k>0} \frac{1}{k}\mathbb{E}\left[\log \|\overleftarrow{P}(0,k-1)\|_1\right]\\
&\quad\le \inf_{k>0}\frac{1}{k}\log\mathbb{E}\left[\|\overleftarrow{P}(0,k-1)\|_1\right]\\
&\quad= \inf_{k>0}\frac{1}{k}\log\mathbb{E}\left[\max_{j\in\mathcal{V}}\sum_{i\in\mathcal{V}}(\overleftarrow{P}(0,k-1))_{ij}\right]\\
&\quad\le \inf_{k>0}\frac{1}{k}\log\mathbb{E}\left[\sum_{j\in\mathcal{V}}\sum_{i\in\mathcal{V}}(\overleftarrow{P}(0,k-1))_{ij}\right]\\
&\quad\le \inf_{k>0}\frac{1}{k}\log\sum_{j\in\mathcal{V}}\sum_{i\in\mathcal{V}}\mathbb{E}\left[\overleftarrow{P}(0,k-1)_{ij}\right]\\
&\quad\le \inf_{k>0}\frac{1}{k}\log\left(n\left\|\mathbb{E}\left[\overleftarrow{P}(0,k-1)\right]\right\|_\infty\right)\\
&\quad= \inf_{k>0}\frac{1}{k}\log\left(n\left\|\prod_{h=0}^{k-1}\mathbb{E}\left[P(h)\right]\right\|_\infty\right).
\end{aligned}$$

Let q be the number of distinct eigenvalues of $\mathbb{E}[P(k)]$, denoted as $\{\lambda_\ell\}_{\ell=1}^q$, and consider the Jordan canonical decomposition $\mathbb{E}\left[P(k)\right] = UJU^{-1}$. Then $\left\|\prod_{h=0}^{k-1}\mathbb{E}\left[P(h)\right]\right\|_\infty \le \|U\|_\infty\|J^k\|_\infty\|U^{-1}\|_\infty$. Denoting by s_ℓ the size of the largest Jordan block corresponding to λ_ℓ, we observe that

$$\|J^k\|_\infty = \max_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}}(J^k)_{ij} = \max_{\ell=1,\dots,q}\sum_{m=0}^{s_\ell-1}\lambda_\ell^{k-m}\binom{k}{m}$$

and deduce

$$\begin{aligned}
\|J^k\|_\infty &\le \max_{\ell=1,\dots,q}|\lambda_\ell|^k\sum_{m=0}^{s_\ell-1}|\lambda_\ell|^{-m}\binom{k}{m}\\
&\le \max_{\ell=1,\dots,q}|\lambda_\ell|^k k^n\sum_{m=0}^{s_\ell-1}|\lambda_\ell|^{-m} \le \chi\rho^k k^n,
\end{aligned}$$

where χ is a constant independent of k and ρ is the spectral radius of $\mathbb{E}[P(k)] = (1-\alpha)I + \alpha P$, which is known to be smaller than 1 because P is Schur stable. We conclude that there exists a constant $C = \|U\|_\infty\|U^{-1}\|_\infty\chi$, independent of k, such that $\mathbb{E}\left[\log\|\overleftarrow{P}(0,k-1)\|_1\right] \le \log\left(nC\rho^k k^n\right)$ and, consequently,

$$\begin{aligned}
\inf_{k>0}\frac{1}{k}\mathbb{E}\left[\log\|\overleftarrow{P}(0,k-1)\|_1\right] &\le \lim_{k\to\infty}\frac{\log(Cnk^n\rho^k)}{k}\\
&= \log\rho < 0.
\end{aligned} \tag{28}$$

The claim then follows from Lemma 1. ■

As a consequence, we deduce that also the (forward) random process $x(k)$ converges in distribution to a limit x_∞, and the distribution of x_∞ is the unique invariant distribution for $x(k)$. We are now ready to verify the ergodicity of $x(k)$ under the assumptions of Theorem 1. Let $z(0)$ be a random vector independent from $x(0)$ with the same distribution as x_∞. Let $\{z(k)\}_{k\in\mathbb{Z}_+}$ be the sequence such that

$$z(k) = \overrightarrow{P}(0,k-1)z(0) + \sum_{0\le\ell\le k-1}\overrightarrow{P}(\ell+1,k-1)u(\ell)$$

where $\overrightarrow{P}(\ell+1,k-1)$ is defined as in (24). Since the process $z(k)$ is stationary and the invariant measure is unique we can apply the Birkhoff Ergodic Theorem (see for instance [42, Chapter 6] or [43, Chapter 5] for a tutorial introduction) and conclude that with probability one $\lim_{k\to\infty}\frac{1}{k}\sum_{\ell=0}^{k-1}z(\ell) = \mathbb{E}[x_\infty]$. On the other hand, we can compute

$$\begin{aligned}
&\mathbb{P}\left(\|x(k)-z(k)\|_1 \ge \varepsilon^k\right)\\
&\qquad\le \frac{\mathbb{E}\left[\|\overrightarrow{P}(0,k-1)(z(0)-x(0))\|_1\right]}{\varepsilon^k}\\
&\qquad\le \frac{\mathbb{E}[\|\overrightarrow{P}(0,k-1)\|_1]\mathbb{E}\left[\|z(0)-x(0)\|_1\right]}{\varepsilon^k}\\
&\qquad\le \frac{Cnk^n\rho^k}{\varepsilon^k}\mathbb{E}\left[\|z(0)-x(0)\|_1\right],
\end{aligned} \tag{29}$$

where we have used (28). If we choose $\varepsilon \in (\rho,1)$, then the Borel-Cantelli Lemma [44, Theorem 1.4.2] implies that with probability one $\|x(k)-z(k)\|_1 < \varepsilon^k$ for all but finitely many values of k. Therefore, almost surely $\frac{1}{k}\sum_{\ell=0}^{k-1}\|x(\ell)-z(\ell)\|_1$ converges to zero as k goes to infinity, and $\lim_{k\to\infty}\frac{1}{k}\sum_{\ell=0}^{k-1}x(\ell) = \mathbb{E}[x_\infty]$. To complete the proof, we only need to observe that $\mathbb{E}[x_\infty] = \lim_{k\to\infty}\mathbb{E}[x(k)]$, which is equal to $x^\star$ as argued after the statement of Theorem 1.

B. Proof of Theorem 2

The argument is similar to [18, Theorem 4.1]. Let us define for all i and k in $\mathbb{Z}_+$

$$\xi_{ki} = \begin{cases}\omega(i)/\sum_{\ell=0}^{k-1}\omega(\ell) & \text{if } i\le k\\ 0 & \text{if } i>k.\end{cases}$$

Since $\lim_{k\to\infty}\sum_{\ell=0}^{k-1}\omega(\ell) = \infty$ almost surely, $\{\xi_{ki}\}_{k,i\in\mathbb{Z}_+}$ forms a Toeplitz array with probability one. Since by (29)

$\lim_{k\to\infty}\|x(k)-z(k)\|_1=0$, we can apply Silverman-Toeplitz Theorem [45] to conclude that almost surely

$$\lim_{k\to\infty}\sum_{i=0}^{\infty}\xi_{ki}\|x(i)-z(i)\|_1 = \lim_{k\to\infty}\frac{1}{\sum_{\ell=0}^{k-1}\omega(\ell)}\sum_{i=0}^{k-1}\omega(i)\|x(i)-z(i)\|_1=0.$$

This equality implies that almost surely

$$\begin{aligned}\lim_{k\to\infty}&\frac{1}{\sum_{\ell=0}^{k-1}\omega(\ell)}\sum_{i=0}^{k-1}\omega(i)x(i)\\ &=\lim_{k\to\infty}\frac{1}{\sum_{\ell=0}^{k-1}\omega(\ell)}\sum_{i=0}^{k-1}\omega(i)(x(i)-z(i))\\ &\quad+\lim_{k\to\infty}\frac{k}{\sum_{\ell=0}^{k-1}\omega(\ell)}\frac{1}{k}\sum_{i=0}^{k-1}\omega(i)z(i)\\ &=\frac{1}{\mathbb{E}[\omega(0)]}\lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\omega(i)z(i),\end{aligned}$$

where the last equality comes from the law of large numbers. Again by Birkhoff Ergodic Theorem, $\{(\omega(k)^\top,z(k)^\top)^\top\}_{k\in\mathbb{Z}_+}$ is a stationary and ergodic process and we finally conclude

$$\lim_{k\to\infty}\frac{1}{\sum_{\ell=0}^{k-1}\omega(\ell)}\sum_{i=0}^{k-1}\omega(i)x(i)=\frac{1}{\mathbb{E}[\omega(0)]}\mathbb{E}[\omega(0)z(0)]=x^\star,$$

thanks to the independence between $\omega(0)$ and $z(0)$.

VII. Concluding remarks

In this work, we have proposed time-averaging as a tool for smoothing oscillations in randomized network systems. Other authors have proposed different solutions, which essentially damp the system inputs in the long run: this goal is achieved through "under-relaxations", that is, by using gains that decrease along time. The analysis of the resulting dynamics is often based on tools from stochastic approximation [46] or semi-martingale theory [47, Ch. 2]. Similarly to our asynchronous sampling, also the choice of decreasing gains can be performed asynchronously and without coordination, albeit at the price of a more complex analysis [46, Ch. 7] [48].

Our method of time-averaging, together with its analysis based on ergodicity, has three advantages: (i) it is simple to apply as it requires minimal assumptions, (ii) it allows for a unified treatment of different algorithms, and (iii) it gives a qualitative insight into the stochastic processes of interest. However, the use of time-averaging is not itself free from drawbacks. Indeed, convergence of time-averages is not exponential, as for the original synchronous dynamics, but polynomial: more precisely, for large times k the distance from the limit value is proportional to $1/k$. This fact can be observed by inspecting the proof of Theorem 1 or by performing a mean-square convergence analysis, as in [23] and [34]. This drawback, which is shared by the over-relaxation approaches, stimulates research towards exponentially-fast algorithms. Likely, effective algorithms can be constructed by endowing the nodes with some memory capabilities: an example is provided in [17] for the localization problem. More generally, their design can be based on the so-called asynchronous iteration method from numerical analysis [49, Section 6.2]: for instance, the application of this method to PageRank computation is discussed in [23, Section VII].

For our examples, we have chosen three problems from the literature and three specific algorithms for their solution. This selection does not cover the spectrum of possible applications. For instance, in the proposed gossip updates, nodes are sampled according to uniform distributions, but the approach may in principle be extended to other distributions if required by the specific application. Also regarding the choice of the problems, we expect that our results can be applied to a much wider range of problems in network systems.

Acknowledgements

The authors are grateful to Noah E. Friedkin for posing the problem studied in Section V and to Keyou You for finding an error in a preliminary version of this paper [35] and for providing a copy of [18]. The authors would also like to thank Francesco Bullo, Giacomo Como, and Fabio Fagnani for interesting conversations on the topics of this paper.

References

[1] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems, with Applications*. Springer, 2013.

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

[3] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, 2008.

[4] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 791–795, 2008.

[5] I. Matei, J. S. Baras, and C. Somarakis, "Convergence results for the linear consensus problem under Markovian random graphs," *SIAM Journal on Control and Optimization*, vol. 51, no. 2, pp. 1574–1591, 2013.

[6] B. Touri and A. Nedic, "Product of random stochastic matrices," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 437–448, 2014.

[7] A. Chiuso, F. Fagnani, L. Schenato, and S. Zampieri, "Gossip algorithms for simultaneous distributed estimation and classification in sensor networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 691–706, 2011.

[8] F. Bullo, R. Carli, and P. Frasca, "Gossip coverage control for robotic networks: Dynamical systems on the space of partitions," *SIAM Journal on Control and Optimization*, vol. 50, no. 1, pp. 419–447, 2012.

[9] A. Giridhar and P. R. Kumar, "Distributed clock synchronization over wireless networks: Algorithms and analysis," in *Proceedings of the IEEE Conference on Decision and Control*, San Diego, CA, USA, Dec. 2006, pp. 4915–4920.

[10] P. Barooah and J. P. Hespanha, "Estimation from relative measurements: Algorithms and scaling laws," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 57–74, 2007.

[11] ——, "Estimation from relative measurements: Electrical analogy and large graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2181–2193, 2008.

[12] ——, "Error scaling laws for linear optimal estimation from relative measurements," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5661–5673, 2009.

[13] S. Bolognani, S. Del Favero, L. Schenato, and D. Varagnolo, "Consensus-based distributed sensor calibration and least-square parameter identification in WSNs," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 2, pp. 176–193, 2010.

[14] W. S. Rossi, P. Frasca, and F. Fagnani, "Transient and limit performance of distributed relative localization," in *Proceedings of the IEEE Conference on Decision and Control*, Maui, HI, USA, Dec. 2012, pp. 2744–2748.

[15] ——, "Limited benefit of cooperation in distributed relative localization," in *Proceedings of the IEEE Conference on Decision and Control*, Florence, Italy, Dec. 2013, pp. 5427–5431.

[16] N. M. Freris and A. Zouzias, "Fast distributed smoothing of relative measurements," in *Proceedings of the IEEE Conference on Decision and Control*, Maui, HI, USA, Dec. 2012, pp. 1411–1416.

[17] A. Carron, M. Todescato, R. Carli, and L. Schenato, "An asynchronous consensus-based algorithm for estimation from noisy relative measurements," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 283–295, 2014.

[18] K. You, S. Song, and L. Qiu, "Randomized incremental least squares for distributed estimation over sensor networks," in *IFAC World Congress*, Cape Town, South Africa, Aug. 2014, pp. 7424–7429.

[19] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, 1998.

[20] K. Bryan and T. Leise, "The $25,000,000,000 eigenvector: The linear algebra behind Google," *SIAM Review*, vol. 48, no. 3, pp. 569–581, 2006.

[21] A. Langville and C. Meyer, *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2006.

[22] H. Ishii and R. Tempo, "The pagerank problem, multiagent consensus, and web aggregation: A systems and control viewpoint," *IEEE Control Systems Magazine*, vol. 34, no. 3, pp. 34–53, 2014.

[23] ——, "Distributed randomized algorithms for the PageRank computation," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 1987–2002, 2010.

[24] A. V. Nazin and B. T. Polyak, "Randomized algorithm to determine the eigenvector of a stochastic matrix with application to the PageRank problem," *Automation and Remote Control*, vol. 72, no. 2, pp. 342–352, 2011.

[25] H. Ishii, R. Tempo, and E. W. Bai, "PageRank computation via a distributed randomized approach with lossy communication," *Systems & Control Letters*, vol. 61, no. 12, pp. 1221–1228, 2012.

[26] W.-X. Zhao, H.-F. Chen, and H. Fang, "Convergence of distributed randomized PageRank algorithms," *IEEE Transactions on Automatic Control*, vol. 58, no. 12, pp. 3255–3259, 2013.

[27] H. Ishii, R. Tempo, and E. Bai, "A web aggregation approach for distributed randomized PageRank algorithms," *IEEE Transactions on Automatic Control*, vol. 57, pp. 2703–2717, 2012.

[28] O. Fercoq, M. Akian, M. Bouhtou, and S. Gaubert, "Ergodic control and polyhedral approaches to PageRank optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 134–148, 2013.

[29] B. C. Csáji, R. M. Jungers, and V. D. Blondel, "PageRank optimization by edge selection," *Discrete Applied Mathematics*, vol. 169, pp. 73–87, 2014.

[30] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[31] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinion change," in *Advances in Group Processes*, E. J. Lawler and M. W. Macy, Eds. JAI Press, 1999, vol. 16, pp. 1–29.

[32] B. Polyak and A. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 838–855, 1992.

[33] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[34] C. Ravazzi, P. Frasca, H. Ishii, and R. Tempo, "A distributed randomized algorithm for relative localization in sensor networks," in *Proceedings of the European Control Conference*, Zürich, Switzerland, Jul. 2013, pp. 1776–1781.

[35] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Almost sure convergence of a randomized algorithm for relative localization in sensor networks," in *Proceedings of the IEEE Conference on Decision and Control*. Florence, Italy, Dec. 2013, pp. 4778–4783.

[36] P. Frasca, C. Ravazzi, R. Tempo, and H. Ishii, "Gossips and prejudices: Ergodic randomized dynamics in social networks," in *Proceedings of the IFAC Workshop on Estimation and Control of Networked Systems*, Koblenz, Germany, Sep. 2013, pp. 212–219.

[37] N. M. Freris, S. R. Graham, and P. R. Kumar, "Fundamental limits on synchronizing clocks over networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 1352– 1364, 2011.

[38] S. Bolognani, R. Carli, E. Lovisari, and S. Zampieri, "A randomized linear algorithm for clock synchronization in multi-agent systems," in *Proceedings of the IEEE Conference on Decision and Control*, Dec. 2012, pp. 20–25.

[39] Y. Kadowaki and H. Ishii, "Event-based distributed clock synchronization for wireless sensor networks," in *Proceedings of the IEEE Conference on Decision and Control*, Dec. 2013, pp. 6747–6752.

[40] H. Ishii, R. Tempo, E. W. Bai, and F. Dabbene, "Distributed randomized PageRank computation based on web aggregation," in *Proceedings of the IEEE Conference on Decision and Control*, Shanghai, China, Dec. 2009, pp. 3026–3031.

[41] P. Diaconis and D. Freedman, "Iterated random functions," *SIAM Review*, vol. 41, pp. 45–76, 1999.

[42] D. W. Stroock, *Probability Theory: An Analytic View*. Cambridge University Press, 1993.

[43] M. Hairer, "Ergodic properties of Markov processes," 2006. [Online]. Available: http://www.hairer.org/notes/Markov.pdf

[44] V. Borkar, *Probability Theory: An Advanced Course*. Springer, 1995.

[45] O. Toeplitz, "Über allgemeine lineare Mittelbildungen," *Prace Matematyczno-Fizyczne*, vol. 22, pp. 113–119, 1911.

[46] V. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[47] B. Polyak, *Introduction to Optimization*. New York, NY: Optimization Software, Inc., 1987.

[48] A. Nedic, "Asynchronous broadcast-based convex optimization over a network," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1337–1351, 2011.

[49] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

Chiara Ravazzi (M'13) received the B.Sc. and M.Sc., and Ph.D. degrees in applied mathematics from Politecnico di Torino, Italy, in 2005 and 2007, and 2011 respectively. She is currently a Postdoctoral Associate at the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Italy. During 2010, she was a visiting student at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge. Specific themes of current interest include the mathematics of control and information theory, coding theory, and signal processing.

Paolo Frasca (M'13) received the Ph.D. degree in Mathematics for Engineering Sciences from Politecnico di Torino, Italy, in 2009. Between 2008 and 2013, he has held research and visiting positions at the University of California, Santa Barbara, USA, at the IAC-CNR, Rome, and at the Politecnico di Torino. Since June 2013, he is an Assistant Professor with the Department of Applied Mathematics, University of Twente, the Netherlands.

His research interests are in the theory of network systems and cyber-physical systems, with applications to robotic, sensor, infrastructural, and social networks. On these topics, Dr. Frasca has (co)authored more than forty journal and conference papers and has given several invited talks at international institutes and events, both in Europe and in the USA. He is a recipient of the 2013 SIAG/CST Best SICON Paper Prize.

Dr. Frasca is currently serving as a Subject Editor for the International Journal of Robust and Nonlinear Control and in the Conference Editorial Boards of the IEEE Control Systems Society and of the European Control Association (EUCA).

Roberto Tempo (F'00) is currently a Director of Research of Systems and Computer Engineering at CNR-IEIIT, Politecnico di Torino, Italy. He has held visiting positions at Chinese Academy of Sciences in Beijing, Kyoto University, The University of Tokyo, University of Illinois at Urbana-Champaign, German Aerospace Research Organization in Oberpfaffenhofen and Columbia University in New York. His research activities are focused on the analysis and design of complex systems with uncertainty, and various applications within information technology. On these topics, he has published more than 180 research papers in international journals, books and conferences. He is also a co-author of the book Randomized Algorithms for Analysis and Control of Uncertain Systems, Springer-Verlag, London, published in two editions in 2005 and 2013.

He is a Fellow of the IEEE and a Fellow of the IFAC. He is a recipient of the IFAC Outstanding Paper Prize Award for a paper published in Automatica and of the Distinguished Member Award from the IEEE Control Systems Society. He is a Corresponding Member of the Academy of Sciences, Institute of Bologna, Italy, Class Physical Sciences, Section Technical Sciences.

In 2010 Dr. Tempo has served the IEEE Control Systems Society as President. Since January 2015 he is serving as Editor-in-Chief of Automatica. He has been Editor for Technical Notes and Correspondence of the IEEE Transactions on Automatic Control in 2005-2009 and a Senior Editor of the same journal in 2011-2014. He is a member of the Advisory Board of Systems & Control: Foundations & Applications, Birkhauser. He was General Co-Chair for the IEEE Conference on Decision and Control, Florence, Italy, 2013 and Program Chair of the first joint IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, 2005.

Hideaki Ishii (M'02-SM'12) received the M.Eng. degree in applied systems science from Kyoto University, Kyoto, Japan, in 1998, and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 2002.

He was a Postdoctoral Research Associate with the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, USA, from 2001 to 2004, and a Research Associate with the Department of Information Physics and Computing, The University of Tokyo, Tokyo, Japan, from 2004 to 2007. Currently, he is an Associate Professor in the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama, Japan. He is a co-author of the book Limited Data Rate in Control Systems with Networks, Lecture Notes in Control and Information Sciences (Springer, 2002). He is an Associate Editor for Automatica. His research interests are in networked control systems, multiagent systems, hybrid systems, cyber security of power systems, and probabilistic algorithms.

Dr. Ishii has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is the Chair of the IFAC Technical Committee on Networked Systems.