Push-Sum on Random Graphs: Almost Sure Convergence and Convergence Rate

Pouya Rezaeinia [iD], Bahman Gharesifard [iD], Tamás Linder [iD], and Behrouz Touri [iD]

Abstract—In this paper, we study the problem of achieving average consensus over a random time-varying sequence of directed graphs by extending the class of so-called push-sum algorithms to such random scenarios. Provided that an ergodicity notion, which we term the directed infinite flow property, holds and the auxiliary states of agents are uniformly bounded away from zero infinitely often, we prove the almost sure convergence of the evolutions of this class of algorithms to the average of initial states. Moreover, for a random sequence of graphs generated using a so-called time-varying B-irreducible probability matrix, we establish convergence rates for the proposed push-sum algorithm.

Index Terms—Distributed optimization, optimization algorithms, random networks.

I. INTRODUCTION

Many distributed algorithms, executed with limited information over a network of agents, rely on estimating the average value of the initial state of the individual agents. These include consensus problems [2], distributed optimization protocols [3]–[13], distributed regret minimization algorithms in machine learning [14], and dynamics for fusion of information in sensor networks [15]. There is a large body of work devoted to the average consensus problem in directed time-varying communication settings, starting with the pioneering work [16], where the gossip-based *push-sum algorithm* is first introduced. The key differentiating factor of the push-sum algorithm from consensus dynamics is that it takes advantage of a paralleled scalar-valued agreement dynamics, initiated uniformly across the agents, that tracks the imbalances of the network and adjusts for them when estimating the consensus value. It is worth pointing out that the traditional average consensus protocols do not necessarily achieve the exact average when the communication network is directed and time varying [16].

In addition to the earlier work [16], several recent papers have studied the problem of average consensus, see for example [17], where other classes of algorithms based on weight adaptation are considered, ensuring convergence to the average on fixed directed graphs. The study of

convergence properties of push-sum (referred to as weighted gossip, as a generalization of pairwise gossip [18]) algorithms on time-varying deterministic sequences of directed graphs, to the best of our knowledge, was initiated in [19] and extended in [13], where push-sum protocols are intricately utilized to prove the convergence of a class of distributed optimization protocols on a sequence of time-varying directed graphs. The key assumption in [13] is the B-strongly-connectedness of the sequence, which means that in any window of size B the union of the underlying directed graphs over time is strongly connected. As we demonstrate, a by-product of our work in deterministic settings is the generalization of the sequences on which the convergence of the push-sum algorithms is valid given the infinite flow property; in this sense, this extension mimics the properties required for the convergence of consensus dynamics, along the lines of [20] and [21].

In large-scale networks links may fail at any time at random, and therefore, modeling these networks as random graphs is necessary to improve the robustness and accuracy of the model. This paper is concerned with the problem of average consensus for scenarios where communication between nodes is directed, time varying and possibly random. The convergence properties of consensus dynamics on random sequences of directed graphs are by this time well established, see for example [20], [22], [23]. Average consensus on random graphs has also been studied in [19], under the assumption that the corresponding random sequence of stochastic matrices is *stationary* and ergodic with positive diagonals and irreducible expectation. One of our main objectives in this paper is to extend these results to more general sequences of random stochastic matrices, in particular, beyond stationary. More importantly, we establish convergence rates for the push-sum algorithm on random sequences of directed graphs.

The remainder of this paper is organized as follows. Section II contains mathematical preliminaries. In Section III, we give a formal description of our consensus problem. In Section IV, we describe the push-sum algorithm. Section V studies the ergodicity of row-stochastic matrices, and Section VI contains our main convergence results. In Section VII, we derive convergence rates for the push-sum algorithm for a class of random column-stochastic matrices. Finally, we gather our conclusions and ideas for future directions in Section VIII.

II. MATHEMATICAL PRELIMINARIES

We start with introducing some notational conventions. Let $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real and integer numbers, respectively, and let $\mathbb{R}_{\geq 0}$ and $\mathbb{Z}_{\geq 0}$ denote the set of nonnegative real numbers and integers, respectively. For a set $\mathbb{A}$, we write $S \subset \mathbb{A}$ if S is a proper subset of $\mathbb{A}$, and we call the empty set and $\mathbb{A}$ trivial subsets of $\mathbb{A}$. The complement of S is denoted by $\bar{S}$. Let $|S|$ denote the cardinality of a finite set S. We view all vectors in $\mathbb{R}^n$ as column vectors, where $n \in \mathbb{Z}_{\geq 0}$. We denote by $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, the standard Euclidean norm, the 1-norm, and the infinity norm on $\mathbb{R}^n$, respectively. The ith unit vector in $\mathbb{R}^n$, whose ith component is 1 and all other components are 0, is denoted by e_i. The notation A' and v' will refer to the transpose of

the matrix A and the vector v, respectively. We will use the short-hand notation $\mathbf{1}_n = (1, \ldots, 1)'$ and $\mathbf{0}_n = (0, \ldots, 0)' \in \mathbb{R}^n$. A vector v is stochastic if its elements are nonnegative real numbers that sum to 1. We use $\mathbb{R}^{n \times n}_{\geq 0}$ to denote the set of $n \times n$ nonnegative real-valued matrices. A matrix $A \in \mathbb{R}^{n \times n}_{\geq 0}$ is row stochastic (column stochastic) if each of its rows (columns) sums to 1. For a given $A \in \mathbb{R}^{n \times n}_{\geq 0}$ and any nontrivial $S \subset [n] = \{1, , \ldots, n\}$, we let $A_{S\bar{S}} = \sum_{i \in S, j \in \bar{S}} A_{ij}$. Finally, a positive matrix is a real matrix with positive entries.

A. Graph Theory

A (weighted) *directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ consists of a node set $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and a weighted *adjacency matrix* $A \in \mathbb{R}^{n \times n}_{\geq 0}$, with $A_{ji} > 0$ if and only if $(v_i, v_j) \in \mathcal{E}$, in which case we say that v_i is connected to v_j. Similarly, given a matrix $A \in \mathbb{R}^{n \times n}_{\geq 0}$, one can associate to A a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $(v_i, v_j) \in \mathcal{E}$ if and only if $A_{ji} > 0$, and hence, A is the corresponding adjacency matrix for $\mathcal{G}$. The in-neighbors and the out-neighbors of v_i are the set of nodes $N_i^{\text{in}} = \{j \in [n] : A_{ij} > 0\}$ and $N_i^{\text{out}} = \{j \in [n] : A_{ji} > 0\}$, respectively. The out-degree of v_i is $d_i^{\text{out}} = |N_i^{\text{out}}|$. A path is a sequence of nodes connected by edges. A directed graph is *strongly connected* if there is a path between any pair of nodes. A directed graph is *complete* if every pair of distinct vertices is connected by an edge. If the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ is strongly connected, we say that A is irreducible.

B. Sequences of Random Stochastic Matrices

Let $\mathcal{S}_n^+$ be the set of $n \times n$ column-stochastic matrices that have positive diagonal entries, and let $\mathcal{F}_{\mathcal{S}_n^+}$ denote the Borel σ-algebra on $\mathcal{S}_n^+$. Given a probability space $(\Omega, \mathcal{B}, \mu)$, a measurable function $W : (\Omega, \mathcal{B}, \mu) \to (\mathcal{S}_n^+, \mathcal{F}_{\mathcal{S}_n^+})$ is called a random column-stochastic matrix, and a sequence $\{W(t)\}$ of such measurable functions on $(\Omega, \mathcal{B}, \mu)$ is called a random column-stochastic matrix sequence; throughout, we assume that $t \in \mathbb{Z}_{\geq 0}$. Note that for any $\omega \in \Omega$, one can associate a sequence of directed graphs $\{\mathcal{G}(t)(\omega)\}$ to $\{W(t)(\omega)\}$, where $(v_i, v_j) \in \mathcal{E}(t)(\omega)$ if and only if $W_{ji}(t)(\omega) > 0$. This in turn defines a sequence of random directed graphs on $\mathcal{V} = \{v_1, \ldots, v_n\}$, which we denote by $\{\mathcal{G}(t)\}$.

III. Problem Statement

Consider a network of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where node $v_i \in \mathcal{V}$ has an initial state (or opinion) $x_i(0) \in \mathbb{R}$; the assumption that this initial state is a scalar is without loss of generality, and our treatment can easily be extended to the vector case. The objective of each node is to achieve *average consensus*; that is to compute the average $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i(0)$ with the constraint that only limited exchange of information between nodes is permitted. The communication layer between nodes at each time $t \geq 0$ is specified by a sequence of random directed graphs $\{\mathcal{G}(t)\}$, where $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$. Specifically, at each time t, node v_i updates its value based on the values of its in-neighbors $v_j \in N_i^{\text{in}}(t)$. One standing assumption throughout this paper is that each node knows its out-degree at every time t; this assumption is indeed necessary, as shown in [24]. Our main objective is to show that the class of so-called push-sum algorithms can be used to achieve average consensus at every node, under the assumption that the communication network is random. This key point distinguishes our work from the existing results in the literature [16], [13], [17]. Another key objective that we pursue in this paper is to obtain rates of convergence for such algorithms. We start our treatment with reviewing the push-sum algorithm.

IV. Random Push-Sum

Consider a network of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where node $v_i \in \mathcal{V}$ has an initial state $x_i(0) \in \mathbb{R}$. The push-sum algorithm, proposed originally in [16], is defined as follows. Each node v_i maintains and updates, at each time $t \geq 0$, two state variables $x_i(t)$ and $y_i(t)$. The first state variable is initialized to $x_i(0)$ and the second one is initialized to $y_i(0) = 1$, for all $i \in [n]$. At time $t \geq 0$, node v_i sends $\frac{x_i(t)}{d_i^{\text{out}}(t)}$ and $\frac{y_i(t)}{d_i^{\text{out}}(t)}$ to its out-neighbors in the random directed graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$, which we assume to contain self-loops at each node for all $t \geq 0$. At time $(t + 1)$, node v_i updates its state variables according to

$$x_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{x_j(t)}{d_j^{\text{out}}(t)}$$

$$y_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{y_j(t)}{d_j^{\text{out}}(t)}. \tag{1}$$

It is useful to define another auxiliary variable $z_i(t+1) = \frac{x_i(t+1)}{y_i(t+1)}$; as we will show later, $z_i(t+1)$ is the estimate by node v_i of the average $\bar{x}$. One can rewrite this algorithm in a vector form; let the column-stochastic matrix $W(t)$ be a function of $\mathcal{E}(t)$ with entries

$$W_{ij}(t) = \begin{cases} \frac{1}{d_j^{\text{out}}(t)}, & \text{if } j \in N_i^{\text{in}}(t) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Using these weighted adjacency matrices, for every $t \geq 0$, we can rewrite the dynamics (1) as

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

$$y(t+1) = W(t)y(t) \tag{3}$$

where $x(t), y(t) \in \mathbb{R}^n$ with entries $x_i(t)$ and $y_i(t)$, respectively, with $i \in \{1, \ldots, n\}$.

V. Ergodicity

In this section, we establish some important auxiliary results regarding the convergence of products of matrices, which satisfy the so-called directed infinite flow property (c.f., Definition 3). We study the products of a class of matrices in a deterministic setting, which we then use to study the push-sum algorithm in the following section. We start by some definitions.

Definition 1 (Ergodicity [25], [20]): Let $\{A(t)\}$ be a sequence of row-stochastic matrices, and for $t \geq s \geq 0$, let $A(t : s)$ denote the product

$$A(t : s) = A(t)A(t-1) \cdots A(s) \tag{4}$$

where $A(s : s) = A(s)$. The sequence $\{A(t)\}$ is said to be *weakly ergodic*, if for all $i, j, l \in [n]$ and any $s \geq 0$, $\lim_{t \to \infty} (A_{il}(t : s) - A_{jl}(t : s)) = 0$. The sequence is said to be *strongly ergodic* if $\lim_{t \to \infty} A(t : s) = \mathbf{1}_n v'(s)$ for any $s \geq 0$, where $v(s) \in \mathbb{R}^n$ is a stochastic vector.

It can be shown that weak ergodicity and strong ergodicity are equivalent [25, Th. 1]. We will simply call such a sequence of row-stochastic matrices ergodic.

We first establish a sufficient condition for ergodicity of a sequence of row-stochastic matrices, Proposition 2, which we subsequently use in our convergence result for the push-sum algorithm. For this reason, we consider the following dynamical system:

$$x(t+1) = A(t)x(t), \qquad \text{for all } t \geq 0. \tag{5}$$

Let us start by two key definitions.

Definition 2 (Strong Aperiodicity [20]): We say that a sequence of matrices $\{A(t)\}$ is *strongly aperiodic* if there exists $\gamma > 0$ such that $A_{ii}(t) \geq \gamma$, for all $t \geq 0$ and $i \in [n]$.

Motivated by the *infinite flow property* [20, Definition 3.2.], we provide the following definition.

Definition 3 (Directed infinite flow property): We say that a sequence of matrices $\{A(t)\}$ has the *directed infinite flow property* (DIFP) if for any nontrivial $S \subset [n]$, $\sum_{t=0}^{\infty} A_{S\bar{S}}(t) = \infty$.

Consider now a sequence of matrices $\{A(t)\}$ that is strongly aperiodic and has the DIFP. Let $k_0 = 0$, and for any $q \geq 1$, define

$$k_q = \arg\min_{t' > k_{q-1}} \left(\min_{S \subset [n]} \sum_{t=k_{q-1}}^{t'-1} A_{S\bar{S}}(t) > 0 \right). \tag{6}$$

Note that k_q is the minimal time instance after k_{q-1}, such that there is nonzero information flow between any nontrivial subset of $\mathcal{V}$ and its complement; consequently, the directed graph associated with the product $A(k_q - 1 : k_{q-1})$ is strongly connected.

Proposition 1: If a sequence of matrices $\{A(t)\}$ has the DIFP, k_q is finite for all $q \geq 0$.

Proof: Suppose that k_q is not finite for some $q \geq 0$. Then, using (6), there exists a nontrivial subset $S \subset [n]$ such that $\sum_{t=k_{q-1}}^{\infty} A_{S\bar{S}}(t) = 0$. This implies that $\sum_{t=0}^{\infty} A_{S\bar{S}}(t) < \infty$, which contradicts the assumption that $\{A(t)\}$ has the DIFP. ■

To establish convergence results for the products of row-stochastic matrices satisfying Definition 3, we argue that in each time window where the underlying directed graph becomes strongly connected for n times, i.e., after $k_{qn} - k_{(q-1)n}$ time steps for some q, *significant mixing* will occur. To formalize this statement, let $\ell_0 = 0$ and

$$\ell_q = k_{qn} - k_{(q-1)n} \tag{7}$$

for $q \geq 1$. For $t > s \geq 0$, we also define

$$\mathbb{Q}_{t,s} = \{q : s \leq k_{(q-1)n}, k_{qn} \leq t\}.$$

We are now ready to state our first result.

Proposition 2: Consider the dynamics (5), where the sequence $\{A(t)\}$ is such that $A'(t)$ satisfies (2). Suppose, additionally, that $\{A(t)\}$ is strongly aperiodic and has the DIFP. Then, the following conditions exist:

i) there is a vector $\phi(s) \in \mathbb{R}^n$ such that, for all $i, j \in [n]$ and $t \geq s$

$$\left| [A(t:s)]_{ij} - \phi_j(s) \right| \leq \Lambda_{t,s}$$

where $\Lambda_{t,s} = \prod_{q \in \mathbb{Q}_{t,s}} \lambda_q$ and $\lambda_q = \left(1 - \frac{1}{n^{\ell_q}}\right) \in (0,1)$;
ii) if, for the sequence $\{\ell_q\}$ associated with $\{A(t)\}$, we have

$$\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}} = \infty \tag{8}$$

then the sequence $\{A(t)\}$ is ergodic.

Proof: We start by proving the first statement. By definition of k_q, we know that for all $q \geq 0$, $A(k_{q+1} - 1 : k_q)$ is irreducible. Since each $A(t)$ is strongly aperiodic, by Lemma A.1, the matrix

$$A(k_{n(q+1)} - 1 : k_{nq}) = A(k_{n(q+1)} - 1 : k_{n(q+1)-1})$$
$$\times \cdots \times A(k_{nq+2} - 1 : k_{nq+1}) \times A(k_{nq+1} - 1 : k_{nq})$$

which is the product of n irreducible matrices, is positive for all $q \geq 0$. Hence, by [6, Lemma 1], for all $i, j \in [n]$, we have

$$\left[A(k_{n(q+1)} - 1 : k_{nq}) \right]_{ij} \geq \frac{1}{n^{k_{n(q+1)} - k_{nq}}} = \frac{1}{n^{l_{q+1}}}.$$

Now, since $A(t:s) = A(t:s)I_n$ and for all $j \in [n]$, $\max_{i \in [n]}[I_n]_{ij} - \min_{i \in [n]}[I_n]_{ij} = 1$, using [26, Lemma 3], we obtain

$$\max_{i \in [n]}[A(t:s)]_{ij} - \min_{i \in [n]}[A(t:s)]_{ij} \leq \Lambda_{t,s}. \tag{9}$$

Note that if we let $\phi_j(s) = \min_{i \in [n]} A_{ij}(t:s)$ for all $j \in [n]$, we have

$$\left| [A(t:s)]_{ij} - \phi_j(s) \right| \leq \max_{i \in [n]}[A(t:s)]_{ij} - \min_{i \in [n]}[A(t:s)]_{ij}. \tag{10}$$

Using (9) and (10), we obtain the desired result.

We next prove part (ii); since $\lambda_q \in (0,1)$ for all $q \geq 1$, we have that $\ln(\lambda_q) \leq \frac{-1}{n^{\ell_q}}$, where we have used the fact that $\ln(\zeta) \leq \zeta - 1$ for all $\zeta > 0$. Summing over all $q \geq 1$ implies

$$\sum_{q=1}^{\infty} \ln(\lambda_q) \leq -\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}}. \tag{11}$$

On the other hand, we have

$$\lim_{t \to \infty} \Lambda_{t,0} = \lim_{t \to \infty} \prod_{q \in \mathbb{Q}_{t,0}} \lambda_q = \lim_{t \to \infty} \exp\left(\sum_{q \in \mathbb{Q}_{t,0}} \ln(\lambda_q) \right).$$

The definition of the sets $\mathbb{Q}_{t,s}$ implies that we can write the right-hand side as $\exp\left(\sum_{q=1}^{\infty} \ln(\lambda_q) \right)$, which gives

$$\lim_{t \to \infty} \Lambda_{t,0} = \exp\left(\sum_{q=1}^{\infty} \ln(\lambda_q) \right) = 0$$

where the last equality follows from (11) and the assumption $\sum_{q=0}^{\infty} \frac{1}{n^{\ell_q}} = \infty$. Using the fact that $\lim_{t \to \infty} \Lambda_{t,0} = 0$, we have that $\lim_{t \to \infty} \Lambda_{t,s} = 0$, for any $s > 0$. Hence, by Proposition 2, part (i), we conclude that $\{A(t)\}$ is ergodic. ■

Following similar steps as in Proposition 2, we obtain the following result for sequences of column-stochastic matrices of the form (2).

Proposition 3: Consider the dynamics (5) and assume that sequence of matrices $\{A(t)\}$ is strongly aperiodic and has the DIFP, where the $A(t)$ are weighted adjacency matrices in the form of (2). Then, the following conditions exist:

i) there is a vector $\phi(t) \in \mathbb{R}^n$ such that, for all $i, j \in [n]$ and $t \geq s$

$$\left| [A(t:s)]_{ij} - \phi_i(t) \right| \leq \Lambda_{t,s};$$

ii) if the sequence $\{\ell_q\}$ associated with $\{A(t)\}$ satisfies (8), then for all $j \in [n]$, $\lim_{t \to \infty} |[A(t:s)]_{ij} - \phi_i(t)| = 0$.

It is worth pointing out that in Proposition 2, since the $A(t)$ are row stochastic, $x(t)$ approaches a vector with identical entries. However, in Proposition 3 the $x(t)$ does not necessarily approach a fixed vector.

VI. CONVERGENCE OF PUSH-SUM

With all the pieces in place, we are now ready to study the behavior of the push-sum algorithm in a random setting (Th. 1 and Proposition 4).

Theorem 1: Consider the push-sum algorithm (3) and suppose that the sequence $\{W(t)\}$ has the DIFP, almost surely. Then, we have

$$|z_i(t+1) - \bar{x}| \leq \frac{2\|x(0)\|_1}{y_i(t+1)} \Lambda_{t,0} \tag{12}$$

almost surely, where $\Lambda_{t,0} = \prod_{q \in \mathbb{Q}_{t,0}} \lambda_q$ and $\lambda_q = \left(1 - \frac{1}{n^{\ell_q}}\right)$.

Proof: Define

$$D(t:s) \triangleq W(t:s) - \phi(t)\mathbf{1}'_n$$

where $\phi(t)$ is a (random) vector from part (i) of Proposition 3. In addition, under the push-sum algorithm we have that

$$x(t+1) = W(t:0)x(0)$$
$$y(t+1) = W(t:0)y(0)$$

for all $t \geq 0$. Following similar steps as in [13, proof of Lemma 1], we obtain

$$|z_i(t+1) - \bar{x}| \leq \frac{2\|x(0)\|_1}{y_i(t+1)} \left(\max_j |[D(t:0)]_{ij}|\right).$$

Using the upper bound in part (i) of Proposition 3, we obtain the desired result. ∎

Proposition 4: Consider the push-sum algorithm (3) and suppose that the sequence of random column-stochastic matrices $\{W(t)\}$ has the DIFP, almost surely. Moreover, suppose that the sequence $\{\ell_q\}$ associated with $\{W(t)\}$ satisfies (8), almost surely. If there exists $\delta > 0$, such that for any $t \geq 0$, there is $t' \geq t$ such that $y_i(t') \geq \delta$ for all $i \in [n]$, then

$$\lim_{t \to \infty} |z_i(t+1) - \bar{x}| = 0, \quad \text{almost surely.}$$

Remark 1: In the following section, we exhibit a class of random matrix sequences $\{W(t)\}$ that satisfy the conditions of Proposition 4 and, thus, admit average consensus almost surely.

Proof: Proof of this proposition is similar to the proof of [19, Th. 4.1], where the sequence $\{W(t)\}$ is assumed to be stationary; however, since we do not assume stationarity, we provide a proof. By Proposition 3 part (ii), for any $\varepsilon > 0$ there is a time t_ε such that for all $t \geq t_\varepsilon$ and $i \in [n]$

$$\sum_{j=1}^n \left| [W(t:0)]_{ij} - \frac{1}{n} \sum_{k=1}^n [W(t:0)]_{ik} \right| < \delta\varepsilon.$$

By assumption, there exists $t'_\varepsilon \geq t_\varepsilon$ such that $y(t'_\varepsilon) \geq \delta$, which implies that $f(t'_\varepsilon) < \varepsilon$, where $f(t)$ is defined as in Lemma A.2. Since by Lemma A.2, $f(t)$ is nonincreasing, $f(t) < \varepsilon$ for all $t \geq t'_\varepsilon$, meaning that $f(t)$ converges to zero as $t \to \infty$ and, hence, $\lim_{t \to \infty} |z_i(t+1) - \bar{x}| = 0$, almost surely. ∎

It is worth pointing out that for B-strongly-connected graphs [13], $k_q \leq Bq$, $l_q \leq nB$, and $\delta = \frac{1}{n^{nB}}$.

VII. B-IRREDUCIBLE SEQUENCES

In this section, we characterize a class of random column-stochastic matrices that admits average consensus and we provide a rate of convergence of the push-sum algorithm for this class. This class of random matrices, includes many interesting sequences including the stationary case considered in [19].

In the following discussion, we assume that the push-sum dynamics is generated by a column-stochastic matrix sequence $\{W(t)\}$ where

$$W_{ij}(t) = \frac{R_{ij}(t)}{\sum_{i=1}^n R_{ij}(t)} \qquad (13)$$

for all $i, j \in [n]$, where $R_{ij}(t)$ is 1 with probability $P_{ij}(t)$, and is 0 with probability $1 - P_{ij}(t)$ such that $\{R_{ij}(t) : i, j \in [n], t \geq 0\}$ are independent random variables. In other words, there is a random edge between node v_j and v_i at time t with probability $P_{ij}(t)$. Note that

$\{W(t)\}$ is a sequence of independent random column-stochastic matrices.

Furthermore, for the probability matrix sequence $\{P(t)\}_{t \geq 0}$, we assume that the following holds.

Assumption 1: $\{P(t)\}_{t \geq 0}$ is a sequence of $n \times n$ matrices with $P_{ij}(t) \in [0, 1]$. Additionally, we assume that $P_{ii}(t) = 1$, for all $v_i \in \mathcal{V}$. Also, for some constant $\epsilon > 0$, we assume that $P_{ij}(t) \geq \epsilon$ for all $i, j \in [n]$ and all $t \geq 0$ such that $P_{ij}(t) \neq 0$. Finally, we assume that the sequence $\{P(t)\}_{t \geq 0}$ is B-irreducible, i.e., for some integer $B > 0$

$$\sum_{t'=tB}^{(t+1)B-1} P(t)$$

is irreducible for all $t \geq 0$.

Assumption 1 is similar to B-strongly-connectedness of graphs [13], except that the edges between different nodes exist with a Bernoulli-like probability, chosen from the edges in an underlying B-strongly-connected (B-irreducible) sequence of graphs (probability matrices). Also, similar to B-strongly-connectedness in deterministic settings, this assumption is likely not necessary in general. The class of random matrices allowed by Assumption 1 includes many interesting sequences, for example, the stationary case considered in [19]. This assumption plays a key role in some of our upcoming proofs.

We next state the main result of this section.

Theorem 2: Consider the push-sum algorithm (3) and let $\{W(t)\}$ be a sequence of random column-stochastic matrices defined by (13), where $\{P(t)\}$ satisfies Assumption 1. Let $p = \epsilon^{2(n-1)}$. Then, for any $t \geq B + \frac{2nB}{p}$, where $n \geq 2$

$$\mathbb{E}\left[\ln\left(|z_i(t+1) - \bar{x}|\right)\right] \leq c_0 - c_1 t$$

where

$$c_0 = \ln\left(2\|x(0)\|_1\right) + \ln(n)\left(\frac{nB}{p} + B\right) + \ln(15)$$

$$c_1 = -\frac{p}{2nB} \ln\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right).$$

Note that the abovementioned result is similar to the so-called "diffusion speed" as defined in [16], which is an upper bound on the time after which $z_i(t)$ is at an $\epsilon > 0$ distance to $\bar{x}$ with probability arbitrary close to one, for all i.

Before giving the proof, let us illustrate the convergence of the push-sum algorithm on a sequence of random graphs by an example.

Example 1: Consider a network of nodes $\mathcal{V} = \{v_1, \dots, v_5\}$, each with a corresponding state initialized uniformly at random to a real number in [0,1]. For any $t > 0$, $W(t)$ is generated according to (13), with $P(t) = P$, where

$$P = 0.01(\mathbf{J}_5 - \mathbf{I}_5) + \mathbf{I}_5.$$

Here, $\mathbf{J}_5$ and $\mathbf{I}_5$ are a 5×5 matrix of all ones and the identity matrix, respectively. The nodes update their values using the push-sum protocol (3). Since P is irreducible, as we will prove in Corollary 1, $\{W(t)\}$ admits average consensus almost surely. The values of $\mathbb{E}\left[|\mathbf{z}_i(t+1) - \bar{x}|\right]$ are displayed in Fig. 1 for $1 \leq t \leq T = 500$ and all $v_i \in \mathcal{V}$. Note that in this example $B = 1$.

The proof of Theorem 2 relies on the following results.

Lemma 1: Let $\{W(t)\}$ be a sequence of random column-stochastic matrices defined by (13), where $\{P(t)\}$ satisfies Assumption 1. Let $\{k_q\}$ and $\{\ell_q\}$ be the sequences defined, respectively, in (6) and (7) along each sample path. Then, the following conditions exist:
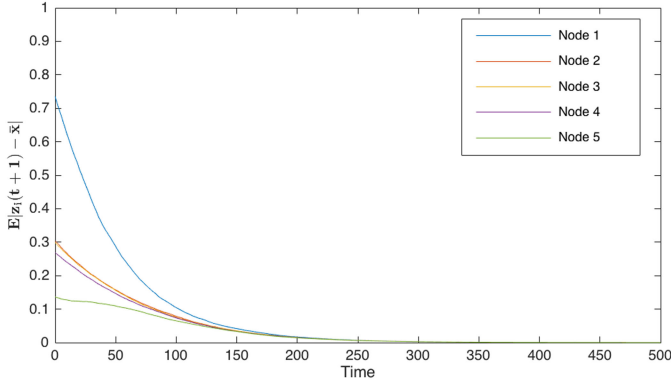i) the sequence $\{W(t)\}$ has the DIFP almost surely;

Fig. 1. Values of $\mathbb{E}\left[|z_i(t+1) - \bar{x}|\right]$ in Example 1 are displayed for all agents (nodes) $v_i \in \mathcal{V}$ and $0 \leq t \leq 500$. The states of the agents are randomly initialized and updated according to the push-sum protocol (3).

ii) for the sequence $\{\ell_q\}$, we have

$$\sum_{q=0}^{\infty} \frac{1}{n^{\ell_q}} = \infty, \quad \text{almost surely.}$$

Proof: We start by proving (i). For any $t \geq 0$, let us define the sequence of events

$$\mathcal{A}_t = \left\{ \sum_{t'=tB}^{(t+1)B-1} W(t') \text{ is irreducible} \right\}. \tag{14}$$

Note that for all $t \geq 0$, the events $\{\mathcal{A}_t\}_{t \geq 0}$ are independent and that $\mathcal{A}_t$ implies $\sum_{t'=tB}^{(t+1)B-1} W_{S\bar{S}}(t') > 0$, for any nontrivial $S \subset [n]$. Since $\min_{i,j \in [n]:P_{ij}(t)>0} P_{ij}(t) > \epsilon > 0$, for all $t \geq 0$, we have

$$\mathbb{P}(\mathcal{A}_t) \geq \epsilon^{2(n-1)}.$$

This follows from [27, Corollary 5.3.6] and the fact that $\{P(t)\}$ is B-irreducible and, hence, there is at least a subset of size $2(n-1)$ of the edges (v_j, v_i) that form a strongly connected graph and $P_{ij}(t') \geq \epsilon$ for some $t' \in [tB, (t+1)B-1]$.

Since the events $\mathcal{A}_t$ are independent, by the second Borel–Cantelli lemma [28, Th. 2.3.6], $\sum_{t'=tB}^{(t+1)B-1} W_{S\bar{S}}(t') > 0$ infinitely often, almost surely. Moreover, since every positive entry of $W(t)$ is bounded below by $\frac{1}{n}$, for any nontrivial $S \subset [n]$, $\sum_{t=0}^{\infty} W_{S\bar{S}}(t) = \infty$, almost surely, implying that $\{W(t)\}$ has the DIFP, almost surely. This also implies that k_q and ℓ_q are finite for all q, almost surely. This completes the proof of (i).

To prove (ii), let us define, for all $t \geq 0$

$$\mathcal{C}_t = \bigcap_{t'=tn}^{(t+1)n-1} \mathcal{A}_{t'} \tag{15}$$

where $\mathcal{A}_t$ is defined in (14). Since the $\mathcal{A}_t$ are independent, $\mathbb{P}(\mathcal{C}_t) = \prod_{t'=tn}^{(t+1)n-1} \mathbb{P}(\mathcal{A}_{t'}) \geq \epsilon^{2n(n-1)}$ for all $t \geq 0$. This implies that $\sum_{t=0}^{\infty} \mathbb{P}(\mathcal{C}_t) = \infty$. Again, since the $\mathcal{C}_t$ are independent, by the Borel–Cantelli lemma, $\mathcal{C}_t$ occurs infinitely often, almost surely. This implies that $\ell_q \leq nB$ infinitely often, almost surely. Hence, $\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}} = \infty$, almost surely. ∎

Lemma 2: In the push-sum algorithm (3), let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 1. Then, for all $t \geq 0$ there exists $t' \geq t$ such that for all $i \in [n]$, $y_i(t') \geq \frac{1}{n^{nB}}$.

Proof: Consider the event $\mathcal{C}_t$ defined in (15). At any time $\mathcal{C}_t$ occurs, by Lemma A.1, the product $W(tnB + nB - 1 : tnB)$ is positive; moreover, by [6, Lemma 1], $W_{ij}(tnB + nB - 1 : tnB) \geq \frac{1}{n^{nB}}$ for all $i, j \in [n]$. Since $W(t)$ is column-stochastic, we have $W_{ij}(tnB + nB - 1 : 0) \geq \frac{1}{n^{nB}}$. By Lemma 1, $\mathcal{C}_t$ occurs infinitely often, almost surely; therefore, for all $t \geq 0$ there exists $t' \geq t$ such that for all $i \in [n]$, $y_i(t') \geq \frac{1}{n^{nB}}$. ∎

To summarize, Lemmas 1 and 2 show that the existence of a time window (with length B) where the underlying aggregate graph is strongly connected with a large enough positive probability ($> \epsilon^{2(n-1)}$) allows us to attain the DIFP and (8). In addition, Assumption 1 guarantees the existence of δ in Proposition 4, as shown in Lemma 2. The preceding two lemmas and Proposition 4 imply the following.

Corollary 1: Let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 1. Then, $\{W(t)\}$ admits average consensus, almost surely.

The following two lemmas allow us to obtain the desired convergence rate stated in Theorem 2.

Lemma 3: Let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 1. Let $\{\ell_q\}$ be the sequence defined in (7) along each sample path. For all $t \geq B + \frac{2nB}{p}$, we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \exp\left(-\beta_t^2 \left(\frac{t}{B} - 2\right)\right) + 2\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}$$

where $\Lambda_{t,0} = \prod_{q \in \mathbb{Q}_{t,0}} (1 - \frac{1}{n^{\ell_q}})$, $\beta_t = \frac{p}{2} - \frac{2pB}{t}$, and $p = \epsilon^{2(n-1)}$.

Proof: Let $\chi_B(t)$ be the indicator of the event $\mathcal{A}_t$, i.e.,

$$\chi_B(t) = \begin{cases} 1, & \text{if } \sum_{t'=tB}^{(t+1)B-1} W(t') \text{ is irreducible} \\ 0, & \text{otherwise.} \end{cases}$$

By the preceding argument, we have $\mathbb{P}(\chi_B(t) = 1) \geq p > 0$. Note that the $\chi_B(t)$ are independent. We let $H_B(T) = \sum_{t=0}^{T} \chi_B(t)$ for all $T \geq 0$, and define

$$q_t \triangleq \max\{q : k_q \leq t\}.$$

By definition of $H_B(\cdot)$ and q_t, we have that

$$q_t \geq H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right). \tag{16}$$

Now, we have that

$$\mathbb{E}\left[\Lambda_{t,0}\right] = \mathbb{E}\left[\Lambda_{t,0} \Big| q_t \leq \frac{pt}{2B}\right] \mathbb{P}\left(q_t \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \Big| q_t > \frac{pt}{2B}\right] \mathbb{P}\left(q_t > \frac{pt}{2B}\right).$$

Since all terms on the right-hand side are less than or equal to 1, we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \mathbb{P}\left(q_t \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \Big| q_t > \frac{pt}{2B}\right].$$

Using (16), we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \mathbb{P}\left(H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \Big| q_t > \frac{pt}{2B}\right].$$

Let us consider the second term on the right-hand side. When $q_t > \frac{pt}{2B}$, we have $|\mathbb{Q}_{t,0}| \geq \lfloor \frac{pt}{2nB} \rfloor$. Using Lemma A.4 to maximize the second

term on the right-hand side over the choices of ℓ_q, we obtain

$$\mathbb{E}\left[\Lambda_{t,0}\bigg|q_t > \frac{pt}{2B}\right] \leq \left(1 - \frac{1}{n^{\left\lfloor\frac{pt}{2nB}\right\rfloor}}\right)^{\left\lfloor\frac{pt}{2nB}\right\rfloor}$$

$$\leq 2\left(1 - \frac{1}{n^{\left\lfloor\frac{t}{2nB}\right\rfloor}}\right)^{\frac{pt}{2nB}}. \qquad (17)$$

To further simplify the abovementioned inequality, we show that $\frac{t}{\left\lfloor\frac{pt}{2nB}\right\rfloor} \leq \frac{4nB}{p}$. To show this, we note that for all $t \geq \frac{2nB}{p} + B$, we have $\frac{pt}{2nB} > 1$ and, hence, $\left\lfloor\frac{pt}{2nB}\right\rfloor \geq 1$. Now, assume that $\xi = \left\lfloor\frac{pt}{2nB}\right\rfloor \geq 1$. We have $2nB\xi \leq pt \leq 2nB(\xi + 1)$. Therefore,

$$\frac{t}{\left\lfloor\frac{pt}{2nB}\right\rfloor} \leq \frac{2nB}{p}\left(\frac{\xi+1}{\xi}\right) \leq \frac{4nB}{p}$$

where the last inequality follows from the fact that $\xi \geq 1$.

Using this inequality in (17), we get

$$\mathbb{E}\left[\Lambda_{t,0}\bigg|q_t > \frac{pt}{2B}\right] \leq 2\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}. \qquad (18)$$

On the other hand, since $\mathbb{E}[\chi_B(t)] \geq p$ for all $t \geq B$, we have

$$\mathbb{P}\left(H\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right) \leq \frac{pt}{2B}\right)$$

$$= \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B\rfloor-1}\chi_B(t') - p\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right) \leq -\alpha_t\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right)\right)$$

$$\leq \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B\rfloor-1}(\chi_B(t') - \mathbb{E}[\chi_B(t')]) \leq -\alpha_t\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right)\right)$$

where

$$\alpha_t = \frac{p\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right) - \frac{pt}{2B}}{\left\lfloor\frac{t}{B}\right\rfloor - 1}. \qquad (19)$$

When $t \geq B + \frac{2nB}{p}$, $\alpha_t > 0$ and, hence, by Lemma A.3, we obtain

$$\mathbb{P}\left(H\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right) \leq \frac{pt}{2B}\right) \leq \exp\left(-\alpha_t^2\left(\left\lfloor\frac{t}{B}\right\rfloor - 1\right)\right)$$

$$\leq \exp\left(-\alpha_t^2\left(\frac{t}{B} - 2\right)\right). \qquad (20)$$

From (19), we have

$$\alpha_t > \frac{p\left(\frac{t}{B} - 2\right) - \frac{pt}{2B}}{\frac{t}{B}} = \frac{p}{2} - \frac{2pB}{t}.$$

If we let $\beta_t = \frac{p}{2} - \frac{2pB}{t}$, using (18) and (20), we conclude that

$$\mathbb{E}[\Lambda_{t,0}] \leq \exp\left(-\beta_t^2\left(\frac{t}{B} - 2\right)\right) + 2\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}$$

finishing the proof.   ∎

Lemma 4: In the push-sum algorithm (3) let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 1. We have, for all $i \in [n]$ and $t \geq 0$

$$\mathbb{E}\left[\ln\left(\frac{1}{y_i(t)}\right)\right] \leq \ln(n)\left(B\frac{n}{p} + B\right).$$

Proof: By [6, Lemma 1], for all $t < \frac{Bn}{p} + B$ and $i \in [n]$

$$[W(t:0)]_{ii} \geq \frac{1}{n^{B\frac{n}{p}+B}}$$

almost surely. This implies that

$$\mathbb{E}\left[\ln\left(\frac{1}{y_i(t)}\right)\right] \leq \ln(n)\left(B\frac{n}{p} + B\right)$$

for all $t < \frac{Bn}{p} + B$ and $i \in [n]$. If $t \geq \frac{Bn}{p} + B$, let $t = aB + b$, where $a, b \in \mathbb{Z}_{\geq 0}$ and $b < B$. Define

$$\tau_t = \begin{cases} \min\{T : \sum_{t=a-T}^{a-1}\chi_B(t) = n\}, & \text{if } \sum_{t=0}^{a-1}\chi_B(t) \geq n \\ a, & \text{otherwise.} \end{cases}$$

When $\tau_t = a$, $W_{ij}(t:0) \geq \frac{1}{n^{\tau_t B + B}}$, for all $i, j \in [n]$. When $\tau_t \neq a$, by Lemma A.1, $W(aB - 1 : (a - \tau_t)B)$ is a positive matrix and consequently by [6, Lemma 1], $W_{ij}(t : (a - \tau_t)B) \geq \frac{1}{n^{\tau_t B + B}}$ for all $i, j \in [n]$; in addition, since the $W(t)$ are column stochastic, we have $W_{ij}(t : 0) \geq \frac{1}{n^{\tau_t B + B}}$. Therefore, for all $t \geq 0$ we have

$$\ln\left(\frac{1}{W_{ij}(t:0)}\right) \leq \ln(n)(\tau_t B + B) \quad \text{for all } i, j \in [n].$$

Consider a sequence of independent Bernoulli trials Y_t, where in each trial the probability of success is p. The number of trials until n successes occur is a negative binomial random variable Z having parameters n and p. Since $\mathbb{P}(\tau_t \leq i) \geq \mathbb{P}(Z \leq i)$ for all $i \geq n$, we have $\mathbb{E}[\tau_t] \leq \mathbb{E}[Z]$. Since $\mathbb{E}[Z] = \frac{n}{p}$, we obtain $\mathbb{E}[\tau_t] \leq \frac{n}{p}$, and hence the result follows.   ∎

We are now in a position to prove Theorem 2.

Proof of Theorem 2 In (12), since both sides are positive, we have

$$\ln(|z_i(t+1) - \bar{x}|) \leq \ln\left(\frac{2\|x(0)\|_1}{y_i(t+1)}\Lambda_{t,0}\right)$$

$$= \ln(2\|x(0)\|_1) + \ln\left(\frac{1}{y_i(t+1)}\right)$$

$$+ \ln(\Lambda_{t,0}).$$

By taking expectations and using Lemma 4, we obtain

$$\mathbb{E}[\ln(|z_i(t+1) - \bar{x}|)] \leq \ln(2\|x(0)\|_1) + \ln(n)\left(\frac{nB}{p} + B\right)$$

$$+ \mathbb{E}[\ln(\Lambda_{t,0})]$$

$$\leq \ln(2\|x(0)\|_1) + \ln(n)\left(\frac{nB}{p} + B\right)$$

$$+ \ln(\mathbb{E}[\Lambda_{t,0}]) \qquad (21)$$

where the last inequality follows from Jensen's inequality. Now by Lemma 3, we have

$$\mathbb{E}[\Lambda_{t,0}] \leq \exp\left(-\beta_t^2\left(\frac{t}{B} - 2\right)\right) + 2\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}$$

where $\beta_t = \frac{p}{2} - \frac{2pB}{t}$. Let us consider the first term on the right-hand side; since $\beta_t \leq \frac{1}{2}$ we have

$$
\begin{aligned}
\exp\left(-\beta_t^2\left(\frac{t}{B}-2\right)\right) &\leq \exp\left(-\beta_t^2\frac{t}{B}+\frac{1}{2}\right) \\
&= \exp\left(-\frac{p^2 t}{4B}+2p^2+\frac{1}{2}-\frac{4p^2 B}{t}\right) \\
&\leq \exp\left(-\frac{p^2 t}{4B}+\frac{5}{2}\right) \\
&= 13\left(\exp\left(-\frac{pn}{2}\right)\right)^{\frac{pt}{2nB}}.
\end{aligned}
$$

Since $n \geq 2$, $\exp(-\frac{pn}{2}) \leq \exp(-p)$. On the other hand, $(1 - \frac{1}{n^{\frac{4nB}{p}}}) \geq (1 - \frac{1}{2^{\frac{8}{p}}})$ for all $n \geq 2$ and $B \geq 1$. It can be seen that for $p \in [0,1]$, $\exp(-p) \leq (1 - \frac{1}{2^{\frac{8}{p}}})$, and consequently $\exp(-\frac{pn}{2}) \leq (1 - \frac{1}{n^{\frac{4nB}{p}}})$. Hence,

$$
\mathbb{E}\left[\Lambda_{t,0}\right] \leq 15\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}. \tag{22}
$$

The result now follows using (21) and (22). ∎

VIII. CONCLUSION

We have studied the convergence properties of the push-sum algorithm for average consensus on sequences of random directed graphs. We have proved that this dynamics is convergent almost surely when some mild connectivity assumptions are met and the auxiliary states of agents are uniformly bounded away from zero infinitely often. We have shown that the latter assumption holds for sequences of random matrices constructed using a sequence of time-varying B-irreducible probability matrices. We have also obtained convergence rates for the proposed push-sum algorithm. Future work include extending Assumption 1 to more general random settings, finding the so-called diffusion speed, studying the implications in scenarios with link failure and in distributed optimization on random time-varying graphs.

APPENDIX

Lemma A.1: For $n \geq 2$, let $\{A(i)\}_{i=1}^{n-1}$ be a sequence of weighted adjacency matrices associated with the strongly connected directed graphs $\{\mathcal{G}(i)\}_{i=1}^{n-1}$ on the node set $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where $\mathcal{G}(i) = (\mathcal{V}, \mathcal{E}(i), A(i))$ and $A(i) \in S_n^+$ for all $i \in [n-1]$. Then, the matrix product $A(n-1:1)$ is positive.

Proof: Let $\mathcal{G}(k:1) = (\mathcal{V}, \mathcal{E}(k:1))$ indicate the directed graph associated with the product $A(k:1)$, where $k \in [n-1]$. Let $N_i^{\text{out}}(k:1)$ and $d_i^{\text{out}}(k:1)$ indicate the set of out-neighbors and out-degree of node $i \in [n]$ in directed graph $\mathcal{G}(k:1)$, respectively. Consider an arbitrary but fixed node $i \in [n]$. Since $A(1) \in \mathcal{S}_n^+$ and $\mathcal{G}(1)$ is strongly connected, we have

$$
d_i^{\text{out}}(1) \geq 2. \tag{23}
$$

Now, consider the directed graph $\mathcal{G}(k:1)$ and assume that $d_i^{\text{out}}(k:1) \leq n-1$ for some $k \in [n-1]$; we show that $d_i^{\text{out}}(k+1:1) > d_i^{\text{out}}(k:1)$. By [6, Lemma 1], we have $N_i^{\text{out}}(k:1) \subseteq N_i^{\text{out}}(k+1:1)$. Moreover, since $\mathcal{G}(k+1)$ is strongly connected and $d_i^{\text{out}}(k:1) \leq n-1$, there is $l \notin N_i^{\text{out}}(k:1)$ such that $l \in N_j^{\text{out}}(k+1)$ for some $j \in N_i^{\text{out}}(k:1)$; otherwise, there is no path between i and l in $\mathcal{G}(k+1)$, contradicting the strong connectivity of $\mathcal{G}(k+1)$. Hence, by [6, Lemma 1] $l \in N_i^{\text{out}}(k+1:1)$, implying that

$$
d_i^{\text{out}}(k+1:1) > d_i^{\text{out}}(k:1).
$$

This along with (23) imply that

$$
d_i^{\text{out}}(k:1) \geq k+1
$$

for all $k \in [n-1]$, which implies that $d_i^{\text{out}}(n-1:1) = n$. Since this statement holds for any $i \in [n]$, the matrix product $A(n-1:1)$ is positive. ∎

Lemma A.2 (Lemma 4.3 [19]): Consider the push-sum algorithm (3). Define

$$
f(t) = \max_{i \in [n]} \frac{\sum_{j=1}^n \left|[W(t:0)]_{ij} - \frac{1}{n}\sum_{k=1}^n [W(t:0)]_{ik}\right|}{y_i(t)}.
$$

Then, $f(t)$ is nonincreasing and

$$
\|z(t) - \bar{x}\mathbf{1}_n\|_\infty \leq \|x(0)\|_\infty f(t).
$$

Lemma A.3 (Hoeffding's inequality [29]): If $X_1, X_2, \ldots, X_n$ are independent random variables and $0 \leq X_i \leq 1$, for all $i \in [n]$, then for any $\alpha > 0$, we have

$$
\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \leq -\alpha n\right) \leq \exp\left(-2\alpha^2 n\right).
$$

The following is an application of Jensen's inequality.

Lemma A.4: For $n > 1$ and for all $l_1, l_2, \ldots, l_q \in \mathbb{Z}_{\geq 0}$, $q > 0$, we have

$$
\prod_{i=1}^q \left(1 - \frac{1}{n^{l_i}}\right) \leq \left(1 - \frac{1}{n^{\frac{t}{q}}}\right)^q
$$

where $t = l_1 + l_2 + \cdots + l_q$.

REFERENCES

[1] P. Rezaienia, B. Gharesifard, T. Linder, and B. Touri, "Convergence rate of push-sum algorithms on random graphs," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 4218–4223.

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

[3] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[4] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc. Symp. Inf. Process. Sensor Netw.*, Berkeley, CA, USA, Apr. 2004, pp. 20–27.

[5] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *J. Optim. Theory Appl.*, vol. 129, no. 3, pp. 469–488, 2006.

[6] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[7] P. Wan and M. D. Lemmon, "Event-triggered distributed optimization in sensor networks," in *Proc. Symp. Inf. Process. Sensor Netw.*, San Francisco, CA, USA, 2009, pp. 49–60.

[8] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[9] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM J. Control Optim.*, vol. 20, no. 3, pp. 1157–1170, 2009.

[10] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Oct. 2010, pp. 557–561.

[11] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proc. IEEE Conf. Decis. Control*, Orlando, FL, USA, 2011, pp. 3800–3805.

[12] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.

[13] A. Nedic and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[14] M. Akbari, B. Gharesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 417–428, Sep. 2017.

[15] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. IEEE Conf. Decis. Control*, Maui, HI, USA, 2012, pp. 5453–5458.

[16] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. IEEE Symp. Found. Comput. Sci.*, Washington, DC, USA, USA, Oct. 2003, pp. 482–491.

[17] A. D. Dominguez-Garcia and C. N. Hadjicostis, "Distributed matrix scaling and application to average consensus in directed graphs," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 667–681, Mar. 2013.

[18] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[19] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Int. Symp. Inf. Theory Proc.*, 2010, pp. 1753–1757.

[20] B. Touri, *Product of Random Stochastic Matrices and Distributed Averaging*. Berlin, Germany: Springer Science & Business Media, 2012.

[21] B. Touri and A. Nedić, "On backward product of stochastic matrices," *Automatica*, vol. 48, no. 8, pp. 1477–1488, 2012.

[22] B. Touri and C. Langbort, "On endogenous random consensus and averaging dynamics," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 241–248, Sep. 2014.

[23] B. Touri and A. Nedić, "Product of random stochastic matrices," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 437–448, Feb. 2014.

[24] J. M. Hendrickx and J. N. Tsitsiklis, "Fundamental limitations for anonymous distributed systems with broadcast communications," in *Proc. Allerton Conf. Commun., Control Comput.*, 2015, pp. 9–16.

[25] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *J. Appl. Probability*, vol. 14, no. 1, pp. 89–97, 1977.

[26] J. Hajnal and M. S. Bartlett, "Weak ergodicity in non-homogeneous Markov chains," *Math. Proc. Cambridge Philos. Soc.*, vol. 54, no. 2, pp. 233–246, 1958.

[27] J. Bang-Jensen and G. Z. Gutin, *Digraphs: Theory, Algorithms and Applications*. London, U.K.: Springer, 2009.

[28] R. Durrett, *Probability: Theory and Examples* (Series in Statistical and Probabilistic Mathematics), 4th ed. Cambridge, U.K: Cambridge Univ. Press, 2010.

[29] W. Hahn, *Theory and Application of Liapunov's Direct Method*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1963.