On Consensus Over Random Networks

Alireza Tahbaz-Salehi and Ali Jadbabaie

Abstract— **We consider the decentralized consensus problem over random information networks. In such networks, the underlying graph of the network at a given time instance is random but independent of all other times. For such a framework, we present a simple necessary and sufficient criteria for asymptotic consensus using simple ergodicity and probabilistic arguments. Finally, we investigate a special case for which the decentralized consensus algorithm converges to the average of the initial values.**

I. Introduction

Over the past few years, algorithms for solving different variants of the *decentralized consensus problem* have received a significant amount of attention. In general, the main goal of such algorithms is to achieve a global objective over a network of agents using purely local interactions. Due to absence of a centralized computational entity, possible lack of information on global topology and limited energy resources, one desirable property of such algorithms is robustness to occurrence of discontinuous changes in the topology of the multiagent network. These topology changes may occur naturally due to communication constraints and/or movement of agents.

Earlier instances of the consensus problem first appeared in [1] and [2] as an estimation-modification process among a group of experts, followed by similar results of [3] and [4] in the context of distributed processing and computation. Recently, many other variants of the problem from diverse fields of study have been widely investigated as well. Problems such as motion coordination of autonomous agents [5], [6], computation of averages and least squares among sensors in a distributed manner [7]–[10], rendezvous problems [11] and distributed locational and geometric optimization [12] are among the wide range of consensus problems addressed by the researchers over the past few years. In all these cases, essentially a distributed iterative scheme known as the *consensus algorithm* is used, according to which the set of agents exchange information with a limited number of other agents and perform a local computation to achieve asymptotic agreement over agents' state values.

More recently, there has been some interest in the random variant of the consensus problem, in which either the topology of the network or the parameters of the consensus algorithm change randomly over time. These randomness assumptions offer a natural framework for situations of practical interest. Although the general form of the consensus problem over random networks has not been addressed yet, some special cases of the problem have been subject of some interest over the past few years. For example, [8] investigates the randomized version of *gossip algorithms* for a network with fixed topology, where each agent exchanges information with only one randomly chosen agent at any time. As another example, in [13] the authors address the problem of reaching consensus when the underlying graph of the network follows the Erdős-Rényi random graph model, where two agents can communicate with each other with a constant probability p, independently of the others.

This paper presents a more general framework - compared to the already existing results - for almost sure convergence of the agreement protocols for a randomly changing network. Our framework contains both the random gossip algorithm with synchronous time model and the Erdős-Rényi random graph model as special cases. Moreover, the results in this paper are more general comparing to the ones derived in [14] which only provides a sufficient condition for reaching consensus in probability, while our theorems provide a necessary and sufficient condition for almost sure convergence. The main assumption behind our model is that the underlying random graphs of the network are mutually independent over time and have the same distribution. We also assume that the parameters of the distributed consensus algorithm at any given time step are independent of the parameters at all other times and are drawn from the same probability distribution.

In order to solve the consensus problem, we address the equivalent problem of ergodicity of left infinite products of stochastic matrices. This equivalent problem is investigated in the literature [15], [16] for the deterministic case, and our main contribution is presenting an easily checkable necessary and sufficient condition for ergodicity of a random sequence of stochastic matrices.

This note is organized as follows: In section II-A we present the general model of information share-and-update process among the nodes of the network. The randomness setup of the network and the weight allocation is presented in section II-B. Section III deals with the ergodicity problem of random stochastic matrices, followed by our main results indicating a criteria for reaching a consensus in section IV. Finally, section V contains our conclusions.

II. Problem Setup

In this section we setup a model for the decentralized consensus problem. Our setup can be used to model different instances of the problem arising in various fields of study

This research is supported in parts by the following grants: ARO/MURI DAAD19-02-1-0383, ONR/YIP-542371 and NSF-ECS-0347285

Alireza Tahbaz-Salehi and Ali Jadbabaie are with GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6228, USA. {atahbaz, jadbabai}@seas.upenn.edu

such as opinion dynamics [17], alignment of autonomous agents [5] and sensor networks [7].

Consider a group of n agents, labeled 1 through n, each of which with an internal state that can be updated over time. We denote the internal state of agent i at time $t > 0$ by $x_i(t)$ and by $x_i(0)$ we mean its initial state value. Also assume that these agents are interconnected pairwise via an information network, i.e. there exists a (directed) communication link between any two agents i and j. We further assume that the information links of the network can disappear and reappear only in discrete time steps $k \in \mathbb{N}$, the set of natural numbers. At any time step, each agent transmits its internal state value over the information network. Agent j is capable of receiving agent i's state information at a given time only if the directed link from i to j is active at that instance of time. In such a case, we say agent j has *access* to agent i.

Our goal is to show whether the state values of the agents can reach a common steady state value by sharing their state information through the information links available to them at each time step. In order to answer this question we use the time dependent directed graph $\mathbb{G}(k)$, with $k \in \mathbb{N}$, to model the topology of the network of agents. Each node (also called a vertex) of $\mathbb{G}(k)$ represents an agent, while pair (i, j) is an edge of the graph if and only if j has access to i at time k. Also we use the notion $\mathcal{N}_j(k)$ to indicate the set of nodes i for which (i, j) is an edge of $\mathbb{G}(k)$. If both (i, j) and (j, i) are edges of $\mathbb{G}(k)$, we say i and j *communicate* at time k. Clearly, the communication relation at a given time among the nodes of the graph is an equivalence relation and defines classes on the vertex set. If all the nodes outside a specific communication class, say class α_r, have no access to any node in that class, α_r is called a *final* class.

A. Update Scheme

We assume the following standard and simple linear dynamics used by the agents (nodes) to update their states: Each agent sets its state at time step k to be a self-chosen weighted average of the states of the agents it has access to and itself at the previous time step $k-1$. In other words, for any $1 \le i \le n$, the state update scheme can be written as

$$x_i(k) = W_{ii}(k)x_i(k-1) + \sum_{j\in\mathcal{N}_i(k)} W_{ij}(k)x_j(k-1) \quad (1)$$

where the values $W_{ij}(k)$ can be considered as nonnegative weights at time k satisfying

$$W_{ii}(k) = 1 - \sum_{j\in\mathcal{N}_i(k)} W_{ij}(k),$$

with the convention that $W_{ii}(k)$ is strictly positive for all nodes i and all times $k \in \mathbb{N}$. Note that the update process (1) possesses the characteristics of a distributed algorithm: it is both iterative and decentralized in the sense that each node only uses the local information available to it.

By defining $W_{ij}(k) = 0$ for $j \notin \mathcal{N}_i(k) \cup \{i\}$ and setting $x(k) = \left[x_1(k) \cdots x_n(k)\right]^T$, one can rewrite (1) in the vector form as

$$x(k) = W_k x(k-1), \quad (2)$$

where the *weight matrix* W_k is a square matrix of order n which has $W_{ij}(k)$ as its (i, j) entry. Such nonnegative matrix W_k with unit row sums is called a *stochastic* matrix in the linear algebra literature. If in addition to row sums, the column sums are also equal to one, the matrix W_k is said to be *doubly stochastic*.

Having defined the weight matrices, we have

$$x(k) = W_k \cdots W_2 W_1 x(0). \quad (3)$$

Now the question is whether (2) can result in an eventual agreement in state value among all nodes. In other words, we are interested in sufficient conditions for

$$\lim_{k\to\infty} |x_i(k) - x_j(k)| = 0$$

to hold for all $1 \le i, j \le n$ and all initial conditions $x(0)$.

B. Random Networks

The randomness in the evolution of the agents' model presented earlier can have two distinct causes. Simply, one can assume that a probability distribution governs the activeness of the pairwise information links between agents (the edges of the graph modeling the network). In other words, the directed edges can disappear and reappear randomly over time either independent from or dependent on each other. But one can also turn the evolution of the system into a random process by simply allowing each node to allocate weights to the states of accessible nodes randomly, during the state update process. In the remaining part of this section, we setup a general framework for randomness in the information network and the update scheme, which captures both of these sources of randomness.

Let $(\Omega_0, \mathcal{B}, \mu)$ be a probability space, where $\Omega_0 = \mathrm{S}_n = \{$set of stochastic matrices of order n with strictly positive diagonals$\}$, $\mathcal{B}$ is the Borel σ-algebra of Ω_0, and μ is a probability measure on Ω_0. Define the product probability space as $(\Omega, \mathcal{F}, \mathbb{P}) = \prod_{k=1}^{\infty}(\Omega_0, \mathcal{B}, \mu)$. Kolmogorov's extension theorem [18] guarantees that $\mathbb{P}$ is the unique probability measure on $(\Omega, \mathcal{F})$ that makes its coordinates stochastically independent while preserving the marginal distributions. By definition, the elements of the product space have the following forms:

$$\begin{aligned} &\Omega = \{(\omega_1, \omega_2, \cdots) : \omega_k \in \Omega_0\} \\ &\mathcal{F} = \mathbb{B} \times \mathbb{B} \times \cdots \\ &\mathbb{P} = \mu \times \mu \times \cdots \end{aligned}$$

Now let the mapping $W_k : \Omega \to \mathrm{S}_n$ be the kth coordinate function, which for all $\omega \in \Omega$ is defined as

$$W_k(\omega) = \omega_k.$$

This guarantees that $W_k(\omega), k \in \{1, 2, \cdots\}$ are independent random stochastic matrices with common distribution μ. Now if the random matrix $W_k(\omega)$ is used as the weight matrix in equation (2), we have a random iterative decentralized update scheme for our network of agents. Note that this setup models both sources of randomness discussed earlier. In fact, while the positivity of the entries of $W_k(\omega)$ characterizes the pairwise accessibilities, the actual value of the (i, j) entry designates the random weight which node i has assigned to the state of node j.

Under the above definitions, we have a time-homogeneous model of the information network and the update process with mutually independent weight matrices. Moreover, the state vectors $x(k)$ form a sequence of random vectors and as a result, the question of reaching eventual state agreement requires a probabilistic discussion, which is done in the following sections. For the rest of the paper, we will denote the weight matrices by W_k for simplicity even in the cases that we are referring to random weight matrices.

III. ERGODICITY

As it is evident from (3), in order to address the eventual state consensus over the network, one needs to investigate the behavior of infinite products of stochastic matrices. This motivates us to borrow the concept of *weak ergodicity* of a sequence of stochastic matrices from the Markov chain theory.

Definition 1: A sequence of (either random or deterministic) $n \times n$ stochastic matrices $\{W_k\}_{k=1}^{\infty}$ is weakly ergodic, if for all $i, j, s = 1, \cdots, n$

$$\left(U_{i,s}^{(k)} - U_{j,s}^{(k)}\right) \to 0$$

as $k \to \infty$, where $U^{(k)} = W_k \cdots W_2 W_1$ is the left product of the matrices in the sequence.

As the definition suggests, a sequence of stochastic matrices is weakly ergodic if the rows of the product matrix converge to each other, as the number of terms in the product grows. A closely related concept is *strong ergodicity* of a matrix sequence.

Definition 2: A sequence of $n \times n$ stochastic matrices $\{W_k\}_{k=1}^{\infty}$ is strongly ergodic, if for all $i, s = 1, \cdots, n$

$$U_{i,s}^{(k)} \to d_s$$

as $k \to \infty$, where $U^{(k)}$ is the left product and d_s is a constant not depending on i.

To be more precise, the above statements are definitions of weak and strong ergodicity *in the backward direction* respectively. But since this is the only type of ergodicity that we concern, we will simply refer to those properties as ergodicity.

One can easily see that weak and strong ergodicity both describe a tendency to consensus. If either type of ergodicity (weak or strong) holds for the matrix sequence $\{W_k\}_{k=1}^{\infty} = W_1, W_2, \cdots$, the pairwise differences between rows of the product matrix $U^{(k)}$ converge to zero. Therefore, the states of the nodes converge to each other as time goes by for all initial values. Also, they would not reach a consensus in state for all initial state values $x(0)$, unless the sequence of weight matrices is (weakly or strongly) ergodic.

At the first glance, it may seem that there exists a difference between weak and strong ergodicity. In the case of weak ergodicity, the states of every two nodes converge to each other, but the state of each single node does not necessarily converge to some limit. On the other hand, in the presence of strong ergodicity, not only the difference between state values converges to zero, but also all nodes' states enjoy a common limit. In other words, in the case of strong ergodicity, $\lim_{k\to\infty} U^{(k)}$ exists and equals to a fixed rank one matrix. Although one may consider the difference to be an important one, that is not the case here because of the following theorem:

Theorem 1: Given a matrix sequence $\{W_k\}_{k=0}^{\infty}$ and their left products $U^{(k)} = W_k \cdots W_1$, weak and strong ergodicity are equivalent.

Proof: See [2]. ∎

Thus, we shall need only speak of weak ergodicity of the weight matrix sequence in the rest of the paper, and be sure that the limit of the product actually exists. More formally, when $\{W_k\}_{k=0}^{\infty}$ is weakly ergodic, we have

$$\lim_{k\to\infty} U^{(k)} = \mathbf{1} d^T, \tag{4}$$

where d is a vector of size n and $\mathbf{1}$ is a vector with all entries equal to one.

Before stating our first result we need to define the *tail σ-field* of the sequence of random weight matrices.

Definition 3: If $\mathcal{F}'_k = \sigma(W_k, W_{k+1}, \cdots)$, then the σ-field defined as

$$\mathcal{T} = \bigcap_{k=1}^{\infty} \mathcal{F}'_k$$

is the tail σ-field of the sequence $\{W_k\}_{k=1}^{\infty}$.

Intuitively, the event A is in $\mathcal{T}$ if and only if changing a finite number of matrices W_k does not affect the occurrence of A. One important result concerning the tail σ-field of independent random matrices (which is the case here) is the well-known Kolmogorov's 0-1 law. The statement and proof of this theorem can be found in many probability related books and articles, e.g. see [18]. Here, we restate it:

Theorem 2 (Kolmogorov's 0-1 law): If $W_1, W_2, \cdots$ are independent and $A \in \mathcal{T}$ is a tail event, then $\mathbb{P}(A) = 0$ or 1.

In other words, any event in the tail σ-field of a sequence of independent random variables occurs either almost surely or almost never. Now by having the definition of the tail σ-field $\mathcal{T}$, and the statement of Kolmogorov's 0-1 law, we can state our first result regarding the ergodicity of a sequence of independent random stochastic matrices.

Lemma 1: The weak ergodicity of the sequence $W_1, W_2, \cdots$ is a *trivial* event.

Proof: Let k be a positive integer. Define the event

$$A_k = \{\text{The sequence } W_k, W_{k+1}, \cdots \text{ is weakly ergodic.}\}$$

which is clearly an event in $\mathcal{F}'_k = \sigma(W_k, W_{k+1}, \cdots)$. These events form a *decreasing sequence* of events, satisfying

$$A_1 \supseteq A_2 \supseteq \cdots. \tag{5}$$

As a result, by definition of the tail σ-field, we have

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{T},$$

and by Kolmogorov's 0-1 law, $\cap_k A_k$ is a trivial event. Therefore, once we show that $\mathbb{P}(A_1) = \mathbb{P}(\cap_k A_k)$, we have shown that the event $\{\{W_k\}_{k=1}^{\infty}$ is weakly ergodic$\}$ has either probability 0 or 1:

$$\begin{aligned}\mathbb{P}(\bigcap_{k=1}^{\infty} A_k) &= \lim_{j\to\infty} \mathbb{P}(\bigcap_{k=1}^{j} A_k) \\ &= \lim_{j\to\infty} \mathbb{P}(A_j) = \mathbb{P}(A_1)\end{aligned}$$

The first equality is because of continuity of the probability measure while (5) holds and the last equality holds because the distribution of W_k does not depend on k. This completes the proof. ■

Lemma 1 indicates that eventual global state agreement over the network under the decentralized update scheme (2) with random independent identically distributed (i.i.d.) weight matrices obeys a 0-1 law. In other words, the agents' states eventually reach a consensus either almost surely or almost never, which is a discontinuous behavior. The immediate question following this discussion is where this discontinuity in behavior takes place, or in other words, how one can distinguish the two cases of almost sure and almost never ergodicity apart. This is the topic of the next section.

IV. Main Results: Necessary and Sufficient Conditions For Ergodicity

Many papers have addressed the convergence problem of consensus algorithms for the cases that the evolution of the underlying graph of the network $\mathbb{G}(k)$ and the weight assignments by the nodes are deterministic processes. In this section, we present a simple necessary and sufficient condition for convergence of the update scheme (2) given that the weight matrices are i.i.d. stochastic matrices with positive diagonals as constructed in section II. But first, it is constructive to review some existing results regarding convergence of such algorithms for the non-random case.

The following theorem, taken from [16], indicates a sufficient condition for weak ergodicity of a deterministic sequence of stochastic matrices.

Theorem 3: For a stochastic matrix W, define

$$\tau(W) = 1 - \delta(W) = 1 - \max_j(\min_i W_{ij}). \tag{6}$$

Then the sequence $\{W_k\}_{k=1}^{\infty}$ is weakly ergodic if there exists a strictly increasing sequence $\{k_r\}, r = 1, 2, \cdots$ of the positive integers such that

$$\sum_{r=1}^{\infty} \big(1 - \tau(W_{k_{r+1}} \cdots W_{k_r+1})\big) = \infty. \tag{7}$$

The above theorem is a special case of a more general result in [16]. In fact, the function $0 \le \tau(\cdot) \le 1$ as defined in (6) is just one specific case of many possible functions, generally known as *coefficients of ergodicity*, which by satisfying (7) will lead to weak ergodicity [2], [16]. Condition (7) and similar ones derived by using other coefficients of ergodicity have a graph theoretic interpretation as well. As first addressed in [3] and later in [5], such conditions are in fact equivalent to *joint connectivity* of the network over time intervals $[k_r+1, k_{r+1}]$ for $r \in \mathbb{N}$. For $a, b \in \mathbb{N}, a \le b$ we say the network is jointly connected over time interval $[a, b]$ if the graph whose edge set is the union of the edge sets of the graphs in $\{\mathbb{G}(a), \mathbb{G}(a+1), \cdots, \mathbb{G}(b)\}$ (the *union graph*) is connected. Such infinitely often joint connectivity over time intervals $[k_r + 1, k_{r+1}]$, $r \in \mathbb{N}$, guarantees the information flow from each node to any other node over time, and hence the convergence.

We now generalize these results for the case that the weight matrices in (2) are random i.i.d. stochastic matrices with positive diagonals. The next two theorems provide us with a simple necessary and sufficient condition in order to distinguish the two cases of almost sure divergence and almost sure convergence. The theorems indicate that just by looking at the element of the spectrum of the average weight matrix with the second largest modulus, the two cases are distinguishable. More precisely, if $\mathbb{E}W_k$ is the average weight matrix at time k with n eigenvalues satisfying

$$0 \le |\lambda_n(\mathbb{E}W_k)| \le \cdots \le |\lambda_2(\mathbb{E}W_k)| \le |\lambda_1(\mathbb{E}W_k)| = 1,$$

one only needs to know the value of $|\lambda_2(\mathbb{E}W_k)|$ to check if the sequence is ergodic. Note that since the weight matrices at different times have the same distribution, $\mathbb{E}W_k$ does not depend on k.

Before stating our main results regarding weak ergodicity of such sequences, we need to state the following lemma, the proof of which can be found in [19].

Lemma 2: Suppose that W is a stochastic matrix for which its underlying graph has s communication classes named $\alpha_1, \cdots, \alpha_s$. Class α_r is final, if and only if the spectral radius of $\alpha_r[W]$ equals to one, where $\alpha_r[W]$ is the submatrix of W corresponding to the nodes in the class α_r.

Having the above lemma, we continue to our main results. The following theorem provides a sufficient condition for almost never weak ergodicity.

Theorem 4: The random sequence $\{W_k\}_{k=0}^{\infty}$ of stochastic matrices with positive diagonals is (weakly) ergodic almost never, if $|\lambda_2(\mathbb{E}W_k)| = 1$.

Proof: Since all $\omega_k \in \Omega_0$ have positive diagonals, $\mathbb{E}W_k$ has positive diagonals as well. Hence, if $\mathbb{E}W_k$ is irreducible, then it is aperiodic and as a result of theorem 6.6.1 in [20], $|\lambda_2(\mathbb{E}W_k)| < 1$ which is in contradiction with our assumption. Therefore, $|\lambda_2(\mathbb{E}W_k)| = 1$ implies that $\mathbb{E}W_k$ is reducible. As a result, without loss of generality, one can label the agents such that $\mathbb{E}W_k$ gets the following block triangular form

$$\mathbb{E}W_k = \begin{bmatrix} Q_{11} & 0 & \cdots & 0 \\ Q_{21} & Q_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Q_{s1} & Q_{s2} & \cdots & Q_{ss} \end{bmatrix}, \tag{8}$$

where each Q_{ii} is an irreducible matrix and represents the nodes in the equivalence class α_i. Since $|\lambda_2(\mathbb{E}W_k)| = 1$, at least two of the classes have unit spectral radius (note that because of irreducibility and aperiodicity of Q_{ii}'s, the multiplicity of the unit-modulus eigenvalue of each one of them cannot be more than one). Therefore, since $\mathbb{E}W_k$ is a stochastic matrix, lemma 2 implies,

$$\exists i \neq j \text{ s.t. } \alpha_i \text{ and } \alpha_j \text{ are both final classes,}$$

or equivalently, $Q_{ir} = 0$ for all $r \neq i$ and $Q_{jl} = 0$ for all $l \neq j$. Since Ω_0 is a subset of nonnegative matrices, W_k has the same type (zero block pattern) as $\mathbb{E}W_k$ for all time k almost surely. Therefore, $U^{(k)} = W_k \cdots W_2 W_1$ has two orthogonal rows almost surely for any k which means that the random sequence $\{W_k\}$ is weakly ergodic almost never. ■

Theorem 4 provides a necessary condition for reaching consensus using the consensus algorithm with random weight matrices. Our next theorem suggests that this condition is also sufficient for reaching consensus.

Theorem 5: The random sequence $\{W_k\}_{k=0}^{\infty}$ of stochastic matrices with positive diagonals is (weakly) ergodic almost surely, if $|\lambda_2(\mathbb{E}W_k)| < 1$.

Proof: Since $|\lambda_2(\mathbb{E}W_k)|$ is subunit, lemma 2 implies that the underlying graph of matrix $\mathbb{E}W_k$ has exactly one final class. Therefore, the matrix is either irreducible or is reducible but with only one final class. We investigate the two cases separately.

Case 1: Suppose $\mathbb{E}W_k$ is irreducible. Since its largest eigenvalue has multiplicity one, the matrix is primitive [19]. Hence,

$$\exists m \text{ s.t. } [\mathbb{E}W_k]^m > 0,$$

where by > 0 for a matrix, we mean the entry-wise nonnegativity of that matrix. Independence over time implies

$$\mathbb{E}(W_m \cdots W_1) = [\mathbb{E}W_k]^m > 0.$$

As a result, for all $1 \leq i, j \leq n$ the (i, j) entry of $U^{(m)} = W_m \cdots W_2 W_1$ is positive with some nonzero probability, say $p_{ij} > 0$. Therefore, since the weight matrices are i.i.d. with positive diagonals, all the entries of the matrix $W_{n^2 m} \cdots W_2 W_1$ is positive with at least probability $\prod_{i,j} p_{ij} > 0$, i.e. the event $\{W_{n^2 m} \cdots W_2 W_1 > 0\}$ has nonzero probability. [1] Hence, by the second Borel-Cantelli lemma (page 49 of [18]), we have

$$\mathbb{P}(W_{(r+1)n^2 m} \cdots W_{rn^2 m+1} > 0 \text{ for infinitly many } r) = 1.$$

Therefore, picking $k_r = rn^2 m$ results in

$$\delta(W_{k_{r+1}} \cdots W_{k_r+1}) > 0 \quad \textit{i.o. a.s.}$$

where $\delta(\cdot)$ is defined in (6). This guarantees that (7) holds with probability one, and theorem 3 provides us with almost sure ergodicity.

Case 2: Suppose $\mathbb{E}W_k$ is reducible. Therefore, its underlying graph has only one final class. This means that without loss of generality, $\mathbb{E}W_k$ can be written as in (8) with all Q_{ii} irreducible and α_1 (the class corresponding to submatrix Q_{11}) the only final class of nodes. Since α_1 is the only final class, there exists a positive integer m such that all nodes of the graph have access to a node in α_1 (e.g. say node 1) with at most m intermediate nodes in between. i.e. any node in the network is at most an m-hop neighbor of node 1. This combined with the fact that $\mathbb{E}W_k$ has strictly positive diagonals guarantees that the first column of $[\mathbb{E}W_k]^m$ is strictly positive. That with independence, as in case 1, implies the positivity of the first column of $\mathbb{E}(W_m \cdots W_1)$. Therefore, for all $1 \leq j \leq n$ the $(j, 1)$-entry of the matrix $W_m \cdots W_1$ is nonzero with positive probability. Hence, again in parallel to the discussion in case 1, we have,

$$\mathbb{P}\left(\delta(W_{nm} \cdots W_2 W_1) > 0\right) > 0.$$

Now if we set $k_r = rnm$, once again the second Borel-Cantelli lemma guarantees that

$$\mathbb{P}\left(\delta(W_{k_{r+1}} \cdots W_{k_r+1}) > 0 \text{ for infinitly many } r\right) = 1.$$

As a result, the sum $\sum_{r=1}^{\infty} \delta(W_{k_{r+1}} \cdots W_{k_r+1})$ diverges with probability one. Now, theorem 3 implies that the random sequence $\{W_k\}$ is weakly ergodic almost surely. ■

The above theorem provides us with a sufficient condition for almost sure ergodicity and hence consensus over the network. In fact this theorem shows why the iterative update (1) ends up in consensus when the network evolves randomly for some special cases discussed in the literature. For example, the authors of [8] address random gossip algorithms under which each node picks only one neighboring node in random and the two set their states to be the average of their states at the previous time step. For such a model, since the average weight matrix is irreducible, theorem 5 guarantees convergence to the a common limit.

Another widely used model for network's random evolution is Erdős-Rényi random graph model. In such a model as used in [13], an information link is active with a constant probability p and inactive with probability $1 - p$ independently from other edges and other time instances and the weights are allocated evenly among accessible nodes. Therefore when we use Erdős-Rényi as the random graph evolution model, every entry of $\mathbb{E}W_k$ is strictly positive and therefore, the conditions of theorem 5 hold.

Also note that since convergence to a limit almost surely is a more general case of convergence in probability, our results contain the results obtained in [14] as a special case.

A. Interpretation

Theorems 4 and 5 provide a simple criteria to distinguish two very different behaviors of the iterative distributed update (2) with i.i.d. weight matrices. In fact as these theorems suggest, the information in the average assigned weights, rather than the whole information in the probability distribution, is sufficient to predict the long-run behavior of the multiagent system. But this should not come as a surprise to the reader.

In the case that $|\lambda_2(\mathbb{E}W_k)| = 1$ holds, the underlying graph of the network cannot become jointly connected over any time interval with positive probability. In fact there exists at least one communication class in which no node has

[1] Selecting $n^2 m$ as the number of matrices in the product in order to make the product matrix entry-wise positive is a conservative pick. The product matrix becomes positive with nonzero probability with less terms, but such a pick suffices our purpose.

access to any node outside of that class and vice versa. Therefore, one can expect that in such a case the system reaches consensus with probability zero for an arbitrary initial condition. This is exactly what theorem 4 says.

On the other hand, when $\mathbb{E}W_k$ has exactly one eigenvalue with unit modulus, the underlying directed graph of the network becomes jointly connected (not considering the direction of the edges) over time with positive probability, and therefore joint connectivity over time occurs infinitely often with probability one. This guarantees the convergence of the consensus algorithm. This is exactly why theorem 5 holds. The two distinct cases that we dealt with separately are of great importance as well. In case 1, the underlying graph of the network becomes *strongly connected* over non-overlapping time intervals $[k_r + 1, k_{r+1}]$ infinitely often, i.e. the union graphs over such intervals contain a directed spanning tree. Therefore, the final common state of the agents depends on the initial states of all n agents. That is why in the context of moving autonomous agents this situation is called a *leaderless coordination* [5]. On the contrary, in case 2, the union of the underlying directed graphs of the network over the non-overlapping time intervals contains an undirected spanning tree (and not a directed one), a situation called *weak connectivity* over time. This leads to what is called *leader following* by authors of [5] for the case of autonomous agents. In this scenario, there is a non-empty subset of agents whose headings are not affected by the headings of the agents outside that set. But, they have influence on all agents in the network. Given this one-sided information flow, this group of agents can be considered as a collective leader for the whole network (hence the naming).

B. A Special Case

As stated in theorem 1 once we have weak ergodicity, strong ergodicity automatically follows. Therefore, under the conditions of theorem 5, all the nodes reach an agreement in the limit, i.e. the states of all agents converge to a common fixed limit for all initial conditions. Clearly, this limiting state is a random variable which depends on the vector of initial states $x(0)$ and the random sequence of weight matrices $\{W_k\}_{k=0}^{\infty}$. In other words, for known initial conditions, the distribution of d in (4) depends on the distribution of the weight matrices.

For the special case that the weight matrices are doubly stochastic almost surely, the mass of the limit value is concentrated at a single point. In fact, in this case the common final value is equal to the average of the initial values. The following theorem formalizes this statement.

Theorem 6: Suppose $S'_n \subset S_n$ is the set of doubly stochastic matrices of order n. Given the initial condition $x(0)$, if $|\lambda_2(\mathbb{E}W_k)| < 1$ and $\mu(S_n - S'_n) = 0$ hold, then $\lim_{k\to\infty} x(k) = \frac{1}{n}\left(\mathbf{1}^T x(0)\right)\mathbf{1}$ almost surely.

Proof: Our proof follows the same lines used in [8] to prove a similar result. As the first step, note that $|\lambda_2(\mathbb{E}W_k)| < 1$ implies

$$c = |\lambda_2\left(\mathbb{E}(W_k^T W_k)\right)| < 1. \tag{9}$$

Now, define $x_{ave} = \frac{1}{n}\mathbf{1}^T x(0)$ as the average of the initial state values and $y(k) = x(k) - x_{ave}\mathbf{1}$ as the difference between the state vector and the average value. Since all the weight matrices are doubly stochastic with $\mathbf{1}$ both as left and right eigenvectors, $\mathbf{1}^T y(k) = 0$ for all k. Moreover,

$$\begin{aligned} y(k) &= x(k) - x_{ave}\mathbf{1} \\ &= W_k\left(x(k-1) - x_{ave}\mathbf{1}\right) \\ &= W_k y(k-1), \end{aligned}$$

which means that $y(k)$ has the same dynamics of $x(k)$ but with different initial conditions. Also as a result of Chebychev's inequality, for any $\epsilon > 0$,

$$\mathbb{P}(\|x(k) - x_{ave}\mathbf{1}\| > \epsilon) \leq \frac{\mathbb{E}\|y(k)\|^2}{\epsilon^2}$$

where $\|\cdot\|$ represents the l_2-norm of the vectors. Based on this inequality, in order to bound $\mathbb{P}(\|x(k) - x_{ave}\mathbf{1}\| > \epsilon)$ from above, we need to find an upper bound for $\mathbb{E}\|y(k)\|^2$. By defining the σ-field $\mathcal{F}_{k-1} = \sigma(W_1, W_2, \cdots, W_{k-1})$, we have

$$\begin{aligned} \mathbb{E}\|y(k)\|^2 &= \mathbb{E}\left[y(k-1)^T W_k^T W_k y(k-1)\right] \\ &= \mathbb{E}\mathbb{E}\left[y(k-1)^T W_k^T W_k y(k-1)|\mathcal{F}_{k-1}\right]. \end{aligned}$$

But since $y(k-1) \in \mathcal{F}_{k-1}$ and W_k is independent from $\mathcal{F}_{k-1}$, one can see that

$$\mathbb{E}\|y(k)\|^2 = \mathbb{E}\left[y(k-1)^T \mathbb{E}(W_k^T W_k) y(k-1)\right]$$

holds. Therefore, by $y(k-1)\perp\mathbf{1}$ and $\mathbb{E}(W_k^T W_k)\mathbf{1} = \mathbf{1}$ we have

$$\mathbb{E}\|y(k)\|^2 \leq c\mathbb{E}\|y(k-1)\|^2.$$

Hence, applying induction on k results in

$$\mathbb{E}\|y(k)\|^2 \leq c^k \|y(0)\|^2$$

Now (9) implies that for all $\epsilon > 0$,

$$\sum_{k=0}^{\infty} \mathbb{P}(\|x(k) - x_{ave}\mathbf{1}\| > \epsilon) \leq \frac{\|y(0)\|^2}{\epsilon^2(1-c)} < \infty$$

which by using the first Borel-Cantelli lemma (page 46 of [18]) leads to

$$\mathbb{P}(\|x(k) - x_{ave}\mathbf{1}\| > \epsilon \text{ i.o.}) = 0.$$

Therefore,

$$\lim_{k\to\infty} x(k) = x_{ave}\mathbf{1} \quad a.s.$$

and the proof is complete. ■

V. Conclusion

In this paper, we showed how the problem of reaching consensus over a network can be reduced to the problem of ergodicity of a sequence of matrices. In particular, for the case of i.i.d. matrices, we showed that ergodicity is a trivial event. Our main result provides a necessary and sufficient condition for reaching consensus over the random network. The current assumption of independence of graphs over time is not valid once we consider a network of moving agents with geometric graph model for agents in which the existence of an edge is distance-dependent. We plan to extend these results for the more general case of dependence over time to include the case of agents with motion.

REFERENCES

[1] M. H. DeGroot, "Reaching a consensus," *Journal of American Statistical Association*, vol. 69, no. 345, pp. 118–121, Mar. 1974.

[2] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *Journal of Applied Probability*, vol. 14, no. 1, pp. 89–97, Mar. 1977.

[3] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[4] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stocahstic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. AC-31, no. 9, pp. 803–812, Sept. 1986.

[5] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[6] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, Dec. 2005, pp. 2996–3000.

[7] S. Boyd, A. Gosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proceedings of IEEE INFOCOM 2005*, vol. 3, Miami, Mar. 2005, pp. 1653–1664.

[8] ——, "Randomized gossip algorithms," *Special issue of IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[9] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proceedings of the 4th International Conference on Information Processing in Sensor Networks*, Apr. 2005, pp. 63–70.

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

[11] J. Cortes, S. Martinez, and F. Bullo, "Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions," *IEEE Transactions on Automatic Control*, vol. 51, no. 8, 2006, to appear.

[12] ——, "Analysis and design tools for distributed motion coordination," in *Proceedings of the American Control Conference*, Portland, OR, June 2005, pp. 1680–1685.

[13] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[14] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Transactions on Automatic Control*, vol. 51, no. 7, pp. 1207–1210, July 2006.

[15] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," in *Proceedings of the American Mathematical Society*, vol. 14, no. 5, 1963, pp. 733–737.

[16] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed. New York: Springer, 1981.

[17] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis and simulation," *Journal of Societies and Social Simulation*, vol. 5, no. 3, 2002.

[18] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont, CA: Duxbury Press, 2005.

[19] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic Press, 1979.

[20] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford, UK: Oxford University Press, 2001.