On Approximations and Ergodicity Classes in Random Chains

Behrouz Touri and Angelia Nedić

Abstract

We study the limiting behavior of a random dynamic system driven by a stochastic chain. Our main interest is in the chains that are not necessarily ergodic but rather decomposable into ergodic classes. To investigate the conditions under which the ergodic classes of a model can be identified, we introduce and study an ℓ_1-approximation and infinite flow graph of the model. We show that the ℓ_1-approximations of random chains preserve certain limiting behavior. Using the ℓ_1-approximations, we show how the connectivity of the infinite flow graph is related to the structure of the ergodic groups of the model. Our main result of this paper provides conditions under which the ergodicity groups of the model can be identified by considering the connected components in the infinite flow graph. We provide two applications of our main result to random networks, namely broadcast over time-varying networks and networks with random link failure.

Index Terms

Ergodicity, ergodicity classes, infinite flow, product of random matrices.

I. Introduction

The dynamic systems driven by stochastic matrices have found their use in many problems arising in decentralized communication [5], [8], [29], [3], decentralized control [16], [27], [31], [20], [21], distributed optimization [38], [39], [30], [25], [17], and information diffusion in social networks [14], [1]. In many of these applications, the ergodicity plays a central role in ensuring that the local "agent" information diffuses eventually over the entire network of agents. The conditions under which the ergodicity happens have been subject of some recent

Department of Industrial and Enterprise Systems Engineering, University of Illinois, Urbana, IL 61801, Email: {touri1,angelia}@illinois.edu. This research is supported by the National Science Foundation grant CMMI 07-42538.

studies [34], [35]. However, the limiting behavior of the dynamics driven by time-varying chains has not been studied much for the case when the ergodicity is absent. In this context, a notable work is [19] where the asymptotic stability of the Hegselmann-Krause model [15] for opinion dynamics has been studied. In particular, it was shown that while the consensus may not be reached globally among all the agents, the consensus is always reached locally within different agent's subgroups under some mild assumptions on the agent communications. Studying the stability of such models is important for understanding group formation in both deterministic and random time-varying networks, such as multiple-leaders/multiple-followers networked systems.

The main objective of this paper is to investigate the limiting behavior of the linear dynamics driven by random independent chains of stochastic matrices in the absence of ergodicity. Our goal is to study the conditions under which the ergodic groups are formed and to characterize these groups. To do so, we introduce an ℓ_1-approximation and the infinite flow graph of a random model, and we study the properties of these objects. Using the established properties, we extend the main result of the previous work in [36] to a broader class of independent random models. We then proceed to show that for certain random models, although the ergodicity might not happen, the dynamics of the model still converges and partial ergodicity happens almost surely. In other words, under certain conditions, *ergodic groups* are formed and we characterize these groups through the connected components of the infinite flow graph. We then apply the results to a broadcast-gossip algorithm over a time-varying network and to a time-varying network with random link failures.

The work in this paper is related to the literature on ergodicity of random models. A discussion on the ergodicity of deterministic (forward and backward) chains can be found in [33]. The earliest occurrence of the study of random models dates back to the work of Rosenblatt [32], where the algebraic and topological structure of the set of stochastic matrices is employed to investigate the limiting behavior of the product of independent identically distributed (i.i.d.) random matrices. Later, in [22], [23], [10], such a product is studied extensively under a more general assumption of stationarity, and a necessary and sufficient condition for the ergodicity is developed. In [7], [34] the class of i.i.d. random models with almost sure positive diagonal entries were studied. In particular, in [34] it has been showed that such a random model is ergodic if and only if its expected chain is ergodic. Later, this result has been extended to the stationary ergodic models in [35].

Unlike the work on the i.i.d. models or stationary processes in [32], [22], [23], [12], [34], [35], the work in this paper is on independent random models that are not necessarily identically distributed. This work is a continuation of our work in [36], where for a class of independent random models, we showed that the ergodicity is equivalent to the connectivity of the *infinite flow graph* of the random model or its expected model. Furthermore, unlike the studies that provide conditions for ergodicity of deterministic or random chains, such as [9], [38], [16], [6], [12], [34], [35], [36], the work presented in this paper considers the limiting behavior of deterministic and random models that are not necessarily ergodic.

The main contribution of this work is in the following aspects: (1) The establishment of *conditions on random models under which the ergodicity classes are fully characterized*. This result not only implies the stability of certain random dynamics, but also provides the structure of the equilibrium points. The structure is revealed through the connectivity topology of the infinite flow graph of the model. Although the model is not ergodic, the ergodicity happens *locally for groups of indices*, which are characterized as the vertices in the same connected component of the infinite flow graph (Theorem 5). (2) The introduction and study of some perturbations (ℓ_1-approximations) of models that preserve ergodicity classes (as seen in Lemma 1). (3) The introduction of random models (with finite total variation) for which the ergodicity can be fully characterized by the infinite flow property (Theorem 3). This class encircles many of the known ergodic deterministic and random models, as discussed in Section IV.

The structure of this paper is as follows: in Section II, we present the problem of our interest and define some notions. In Section III, we define and investigate ℓ_1-approximations of random models which play an important role in the development. In Section IV, we study the ergodicity a class of random models and we extend the infinite flow theorem of [36] to this class. In Section V, we study the stability of certain random models and characterize their ergodicity classes. These classes are identified using infinite flow graph concept and analysis that combines the results of Sections III and IV. In Section VI, we apply our main result to two different random models, and we conclude in Section VII.

Notation and Basic Terminology. We view all vectors as columns. For a vector x, we write x_i to denote its ith entry, and we write $x \geq 0$ ($x > 0$) when all its entries are nonnegative (positive). We use x^T for the transpose of a vector x. For a vector $x \in \mathbb{R}^m$, we use $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$ for $p \geq 1$ and $\|x\|$ when $p = 2$. For a matrix A, we write $\|A\|_p$ to denote the matrix norm induced by $\|\|_p$ vector norm. We use e_i to denote the vector with the ith entry equal to 1 and all other entries equal to 0. We write e to denote the vector with all entries equal to 1. We write $\{x(k)\}$ to denote a sequence $x(0), x(1), \ldots$ of some elements, and $\{x(k)\}_{k \geq t}$ to denote the truncated sequence $x(t), x(t+1), \ldots$ for $t > 0$. For a given set C and a subset S of C, we write $S \subset C$ when S is a proper subset of C. A set $S \subset C$ with $S \neq \emptyset$ is a *nontrivial* subset of C. We use $[m]$ for the integer set $\{1, \ldots, m\}$. We let $\bar{S}$ denote the complement of a given set $S \subseteq [m]$ with respect to $[m]$.

We denote the identity matrix by I. For a finite collection $A_1, \ldots, A_\tau$ of square matrices, we write $A = \mathrm{diag}(A_1, \ldots, A_\tau)$ to denote the block diagonal matrix with rth diagonal block being A_r for $1 \leq r \leq \tau$. For a matrix W, we write W_{ij} to denote its (i, j)th entry, W^i to denote its ith column vector, and W^T to denote its transpose. For an $m \times m$ matrix W, we let $\sum_{i<j} W_{ij} = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} W_{ij}$. For such a matrix and a nontrivial subset $S \subset [m]$, we define $W_S = \sum_{i \in S, j \in \bar{S}} (W_{ij} + W_{ji})$. A vector $v \in \mathbb{R}^m$ is stochastic if $v \geq 0$ and $\sum_{i=1}^m v_i = 1$. A matrix W is *stochastic* when all its rows are stochastic, and it is *doubly stochastic* when both W and W^T are stochastic. We let $\mathbb{S}^m$ denote the set of $m \times m$ stochastic matrices. We refer to a sequence $\{W(k)\}$ of matrices as *model* or *chain* interchangeably.

We write $\mathsf{E}[X]$ to denote the expected value of a random variable X. For an event $\mathscr{A}$, we use $\mathsf{Pr}(\mathscr{A})$ to denote its probability. For a given probability space, we say that a *statement* R holds almost surely if the set of realizations for which *the statement* R holds is an event and $\mathsf{Pr}(\{\omega \mid \text{statement } R \text{ holds}\}) = 1$. We often abbreviate "almost surely" by *a.s.*

II. Problem Formulation and Motivation

In this section, we describe the problems of interest and introduce some background concepts. We also provide an example that motivates the further development.

A. *Problem of Interest and Basic Concepts*

We consider a linear dynamic system given by

$$x(k+1) = W(k)x(k) \qquad \text{for } k \geq t_0, \tag{1}$$

where $\{W(k)\}$ is a random stochastic chain, t_0 is an initial time and $x(t_0) \in \mathbb{R}^m$ is an initial state of the system. It is well known that, for an ergodic random chain $\{W(k)\}$, the dynamics in (1) is convergent almost surely for any initial time t_0 and any initial state $x(t_0)$ (see [9]). Furthermore, the limiting value of each coordinate $x_i(k)$ is the same, which is often referred to as consensus, agreement, or synchronization. In this case, the dynamics in (1) is asymptotically stable and its equilibrium point lies on the line spanned by the vector e.

A natural question arises: what happens if $\{W(k)\}$ is not ergodic? In particular, what can we say about the stability and the equilibrium points (if any) of dynamics (1)? Can we determine the ergodicity classes based on the properties of the matrices $W(k)$? Our motivation in this paper is to answer these questions.

To formalize the problem of interest, we revisit several notions related to random chains. Let $(\Omega, \mathcal{F}, \mathsf{Pr}(\cdot))$ be a probability space and let $\mathbb{Z}^+$ be the set of non-negative integers. Let $W : \Omega \times \mathbb{Z}^+ \to \mathbb{S}^m$ be a random matrix process, i.e., $W_{ij}(k)$ is a Borel-measurable function for all $i, j \in [m]$ and $k \geq 0$ (which makes it possible to discuss the events of ergodicity and consensus event among others). We refer to such a process as a *random chain* or a *random model*, and denote it by its coordinate representation $\{W(k)\}$. If matrices $W(k)$ are independent, the model is independent. In addition, if $W(k)$s are identically distributed, we say that $\{W(k)\}$ is an *independent identically distributed (i.i.d.)* random model (or chain).

We now introduce the concept of ergodicity. We first define it for a deterministic chain $\{A(k)\}$, which can be viewed as a special independent random chain by setting $\Omega = \{\omega\}$, $\mathcal{F} = \{\{\omega\}, \emptyset\}$, $\mathsf{Pr}(\{\omega\}) = 1$ and $W(k)(\omega) = A(k)$. Then, the dynamic system in Eq. (1) is deterministic and we have the following definition.

Definition 1: A chain $\{A(k)\}$ is *ergodic* if $\lim_{k\to\infty}(x_i(k) - x_j(k)) = 0$ for all $i, j \in [m]$, all starting times $t_0 \geq 0$ and all starting points $x(t_0) \in \mathbb{R}^m$. The chain *admits consensus* if the above assertion is true for $t_0 = 0$.

As pointed out in [9], [16], a chain $\{A(k)\}$ is ergodic if and only if $\lim_{k\to\infty} A(k)\cdots A(t_0) = ev^T(t_0)$ for some stochastic vector $v(t_0) \in \mathbb{R}^m$ and for all $t_0 \geq 0$. Similarly, $\{A(k)\}$ admits consensus if and only if $\lim_{k\to\infty} A(k)\cdots A(0) = ev^T$ for some stochastic vector $v \in \mathbb{R}^m$.

Note that, for a random model $\{W(k)\}$, the subsets of Ω on which the ergodicity and consensus happen, respectively, are events. More precisely, let $\mathscr{E}$ be the subset of Ω over which ergodicity happens. Then, by the definition of ergodicity, we have $\omega \in \mathscr{E}$ if and only if

$\lim_{k\to\infty}(x_i(k,\omega)-x_j(k,\omega))=0$ for all $t_0\ge 0$ and all $x(t_0,\omega)\in\mathbb{R}^m$, where $x(k+1,\omega)=W(k,\omega)x(k,\omega)$ for $k\ge t_0$. Since the dynamics (1) is linear and m-dimensional, the set $\mathscr{E}$ is equivalently described by considering only the starting points $x(t_0,\omega)=e_\ell$ for all $\ell\in[m]$, i.e.,

$$\mathscr{E}=\bigcap_{t_0=0}^{\infty}\left(\bigcap_{\ell=1}^{m}\{\omega\in\Omega\mid\lim_{k\to\infty}(x_i(k,\omega)-x_j(k,\omega))=0\ \text{ for all } i,j\in[m],\ x(t_0,\omega)=e_\ell\}\right),$$

Letting $\mathscr{C}$ denote the subset of Ω over which consensus happens, we can similarly obtain

$$\mathscr{C}=\bigcap_{\ell=1}^{m}\{\omega\in\Omega\mid\lim_{k\to\infty}(x_i(k,\omega)-x_j(k,\omega))=0\ \text{ for all } i,j\in[m],\, x(0)=e_\ell\}.$$

Since $W_{ij}(k)$ are Borel-measurable, the variable $x_i(k)-x_j(k)$ is random for all $i,j\in[m]$ implying that $\mathscr{E}$ and $\mathscr{C}$ are events. We say that a random chain is *ergodic* (*admits consensus*) if the event $\mathscr{E}$ ($\mathscr{C}$) happens almost surely.

The ergodicity of certain random models is closely related to the *infinite flow property*, as shown in [36], [37]. We will use this property, so we recall its definition below.

Definition 2: (*Infinite Flow Property*) A deterministic chain $\{A(k)\}$ has infinite flow property if $\sum_{k=0}^{\infty}A_S(k)=\infty$ for all nonempty $S\subset[m]$. A random model $\{W(k)\}$ has infinite flow property if it has infinite flow property almost surely.

As in the case of consensus and ergodicity events, since $W_{ij}(k)$s are Borel-measurable, the subset of Ω over which the infinite flow happens is an event. We denote this event by $\mathscr{F}$.

Example 1: As an example, consider the 2×2 static chain $\{A(k)\}$ given by $A(k)=A$ for all $k\ge 0$, where

$$A=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}.$$

It can be seen that

$$\lim_{k\to\infty}A(k)\cdots A(t_0)=\lim_{k\to\infty}A^{k-t_0}(0)=A.$$

Thus, $\{A(k)\}$ is ergodic and also admits consensus. Furthermore, there are only two non-trivial subsets S, $S=\{1\}$ and $S=\{2\}$. For either of them, we have $A_S(k)=A_{12}+A_{21}=1$. Therefore, $\sum_{k=0}^{\infty}A_S(k)=\infty$ for all non-trivial subsets $S\subset[m]$ implying that $\{A(k)\}$ has infinite flow property. $\square$

In our development, we also use some additional properties of random models such as weak feedback property and a common steady state in expectation, as introduced in [36]. For convenience, we provide them in the following definition.

Definition 3: Let $\{W(k)\}$ be a random model. We say that the model has:

(a) *Weak feedback property* if there exists $\gamma > 0$ such that

$$\mathsf{E}\big[W^i(k)^T W^j(k)\big] \geq \gamma(\mathsf{E}[W_{ij}(k)] + \mathsf{E}[W_{ji}(k)]) \qquad \text{for all } k \geq 0 \text{ and } i \neq j,\ i,j \in [m].$$

(b) A *common steady state π in expectation* if $\pi^T \mathsf{E}[W(k)] = \pi^T$ for some stochastic vector π and all $k \geq 0$.

Basically, a chain has weak feedback property if the cross-correlations of any two columns $W^i(k)$ and $W^j(k)$ can be bounded below by a time-independent (positive) factor of the term $\mathsf{E}[W_{ij}(k)] + \mathsf{E}[W_{ji}(k)]$ at all times k. This condition is analogous to the aperiodicity condition for homogeneous Markov chains, and it plays a similar role in our development. As an example, any random model with $W_{ii}(k) \geq \gamma > 0$ almost surely for all $k \geq 0$ and $i \in [m]$ has weak feedback property. Also, the i.i.d. models with almost sure positive diagonal entries have this property, as seen in [36]. As an example of a model with a common steady state π in expectation, consider $\{W(k)\}$ where each $W(k)$ is doubly stochastic almost surely, for which we have $\pi = \frac{1}{m}\, e$. More generally, any i.i.d. random model $\{W(k)\}$ has this property, where the vector π can be taken as the (normalized) left-eigenvector of the matrix $\mathsf{E}[W(k)]$ associated with the eigenvalue 1.

With a given random model, we associate an undirected graph which we refer to as *infinite flow graph*. We define this graph as a simple undirected graph with links that have sufficient information flow (simple graph is a graph without self-loops and multiple edges).

Definition 4: (*Infinite Flow Graph*) The infinite flow graph of a random model $\{W(k)\}$ is the graph $G^\infty = ([m], \mathcal{E}^\infty)$, where $\{i, j\} \in \mathcal{E}^\infty$ if and only if $\sum_{k=0}^{\infty}(W_{ij}(k) + W_{ji}(k)) = \infty$ almost surely.

As an example, consider the infinite flow graph of the chain discussed in Example 1. Its infinite flow graph is the graph with vertices 1 and 2, and with an edge connecting them.

The infinite flow graph has been (silently) used in [38] mainly to establish the ergodicity of certain deterministic chains. Here, however, we make use of this graph to establish ergodicity classes for some independent random chains. As an example of what we want to accomplish, consider the static chain of Example 1 and the following static chain

$$P(k) = P \qquad \text{with } P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{for all } k \geq 0.$$

This chain also has the infinite flow property as it has $S = \{1\}$ and $S = \{2\}$ as the non-trivial subsets, with each satisfying $P_S(k) = P_{12} + P_{21} = 2$. Hence, this chain and the chain of Example 1 have the same infinite flow graph, consisting of the graph with vertices 1 and 2, and an edge $\{1, 2\}$. However, while the chain of Example 1 is ergodic, the chain $\{P(t)\}$ is not (as it permutes the entries of the starting vector $x(t_0)$ for all $t_0 \geq 0$). Thus, the dynamics driven by the chain $\{A(k)\}$ of Example 1 is stable, while the dynamics driven by chain $\{P(t)\}$ is not, and yet these chains have the same infinite flow graph.

The first question that arises is *when is the infinite flow graph informative about the stability of the dynamics*? A partial answer to this question is provided by the infinite flow result established in [36], which is given below.

Theorem 1: (*Infinite Flow Theorem*) Let a random model $\{W(k)\}$ be independent, and have a common steady state $\pi > 0$ in expectation and weak feedback property. Then, the following statements are equivalent:

(a) The model is ergodic.
(b) The model has infinite flow property.
(c) The expected model has infinite flow property.
(d) The expected model is ergodic.

Using Theorem 1, one can find that the chain $\{A(k)\}$ of Example 1 is ergodic. However, Theorem 1 provides no information for the chain $\{P(t)\}$, as it fails to satisfy the conditions of the theorem. To further illustrate the limitations of Theorem 1 consider the following example.

Example 2: Let $\{A(k)\}$ be a 3×3 static chain given by $A(k) = A$ for all $k \geq 0$, where

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then, it can be seen that for all $t_0 \geq 0$,

$$\lim_{k\to\infty} A(k) \cdots A(t_0) = \lim_{k\to\infty} A^{k-t_0}(0) = A.$$

Thus, the dynamics driven by this chain is always stable, but we cannot conclude this from Theorem 1, as the chain does not have the infinite flow property since $A_{\{3\}}(k) = A_{12} + A_{13} + A_{21} + A_{31} = 0$. However, the chain satisfies all the other conditions of the theorem, as it has the common steady state vector $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ and weak feedback property with $\gamma = \frac{1}{2}$ (can be seen from $A_{ii} \geq \frac{1}{2}$ for all i). $\square$

Our objective in this paper is to extend the result of Theorem 1 to dynamics that fail to satisfy the infinite flow property. In particular, we would like to *provide conditions for stability of the dynamics* in the absence of the infinite flow property. We do this by investigating the connected components in the infinite flow graph. In the sequel, *a connected component of a graph will always be maximal* with respect to the set inclusion, i.e., a connected subgraph of the given graph is not properly contained in any other connected subgraph.

B. Ergodic Indices

As opposed to a global view of ergodicity, we adopt a local view where ergodicity is defined for each index and extended to index pairs. These refinements of the ergodicity concept are defined as follows.

Definition 5: Let $\{x(k)\}$ in (1) be the dynamics driven by a deterministic chain $\{A(k)\}$. We say that:

(a) The index $i \in [m]$ is an *ergodic index* for the chain if $\lim_{k\to\infty} x_i(k)$ exists for all starting times $t_0 \geq 0$ and all initial points $x(t_0) \in \mathbb{R}^m$. The chain is *asymptotically stable* when each $i \in [m]$ is ergodic.

(b) Two indices $i, j \in [m]$ are *mutually weakly ergodic indices* for the chain if $\lim_{k\to\infty}(x_i(k) - x_j(k)) = 0$ for all initial times $t_0 \geq 0$ and all initial points $x(t_0) \in \mathbb{R}^m$. We write $i \leftrightarrow_A j$ when i and j are mutually weakly ergodic for the chain $\{A(k)\}$.

(c) Two indices $i, j \in [m]$ are *mutually ergodic indices* if i and j are ergodic indices and $i \leftrightarrow_A j$ for all initial times $t_0 \geq 0$ and all initial points $x(t_0) \in \mathbb{R}^m$. We write $i \Leftrightarrow_A j$ when i and j are mutually ergodic for the chain $\{A(k)\}$.

To illustrate these concepts, consider the chain $\{A(k)\}$ of Example 2 for which we have

$$\lim_{k\to\infty} A(k) \cdots A(t_0) = A \qquad \text{for all } t_0 \geq 0,$$

with

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In this case, all indices are ergodic as the dynamics is stable, i.e., for all $t_0 \geq 0$ and $x(t_0) \in \mathbb{R}^3$, we have

$$\lim_{k\to\infty} x(k) = \lim_{k\to\infty} A(k) \cdots A(t_0)x(t_0) = Ax(t_0).$$

Furthermore, we have for all $x(t_0) \in \mathbb{R}^3$,

$$\lim_{k\to\infty} x_1(k) = \lim_{k\to\infty} x_2(k) = \frac{1}{2}\left(x_1(t_0) + x_2(t_0)\right),$$

implying that the indices i and j are mutually weakly ergodic and also mutually ergodic, i.e., $1 \leftrightarrow_A 2$ and $1 \Leftrightarrow_A 2$. However, since $x_3(k) = x_3(t_0)$ for all $k \geq t_0$, starting the dynamics at $x(t_0) = (0, 0, 1)^T$ we obtain

$$\lim_{k\to\infty} x(k) = (0, 0, 1)^T,$$

thus implying that 1 and 3 are not mutually weakly ergodic pairs, and the same is true for 2 and 3, i.e., $1 \not\leftrightarrow_A 3$ and $2 \not\leftrightarrow_A 3$.

The relation $\leftrightarrow_A$ is an equivalence relation on $[m]$ and we can consider its equivalence classes. We refer to the equivalence classes of this relation as *ergodicity classes* and we refer to the resulting partitioning of $[m]$ as the *ergodicity pattern* of the model. In the light of Definition 5, by recalling the definition of the ergodicity (Definition 1), we see that a chain is ergodic if and only if its ergodicity class is a singleton or, equivalently, its ergodicity pattern is $\{[m]\}$. For the chain of Example 2, from the preceding discussion we can see that the ergodicity classes are $\{\{1, 2\}, \{3\}\}$.

Definition 5 extends naturally to a random model. Specifically, if any of the properties in Definition 5 holds almost surely for a random model $\{W(k)\}$, we say that the model $\{W(k)\}$ has the corresponding property. In the further development, when unambiguous, we will omit the explicit dependency of the relation $\leftrightarrow$ and $\Leftrightarrow$ on the underlying chain.

For a random model $\{W(k)\}$, the set of realizations for which $i \leftrightarrow j$ (or $i \Leftrightarrow j$) is a measurable set; hence, an event. When the model $\{W(k)\}$ is independent, these events are tail events and, therefore, each of these events happens with probability either zero or one. Hence, for an independent random model $\{W(k)\}$, we write $i \leftrightarrow j$ when $\mathsf{Pr}\left(i \leftrightarrow j\right) = 1$ and

$i \not\leftrightarrow j$ when $\Pr(i \leftrightarrow j) = 0$. Analogously, we define $i \Leftrightarrow j$ and $i \not\Leftrightarrow j$. Thus, the ergodicity pattern of any independent random model is well-defined.

III. Approximation of chains

Here, we develop a chain approximation concept that plays a key role in our study of non-ergodic chains. In particular, it enables investigating the conditions under which the ergodicity classes of two different chains are the same. To be precise, we say that two chains $\{W(k)\}$ and $\{U(k)\}$ have the *same ergodicity classes* if there exists a bijection $\theta : [m] \to [m]$ between the indices of $\{W(k)\}$ and $\{U(k)\}$ such that:

(a) $i \leftrightarrow_W j$ if and only if $\theta(i) \leftrightarrow_U \theta(j)$, and

(b) i is an ergodic index for $\{W(k)\}$ if and only if $\theta(i)$ is an ergodic index for $\{U(k)\}$.

When θ is a bijection, then the indices of one of the chains can be permuted according to the bijection θ, so that the bijection θ can always be taken as identity. We assume that this is the case for the rest of the paper.

We want to determine a perturbation of a chain that does not affect the ergodicity classes of the chain. It turns out that a perturbation that is small in ℓ_1-norm has such a property. To formally set up the framework, we next define the concept of ℓ_1-approximation.

Definition 6: A deterministic chain $\{B(k)\}$ is an ℓ_1-approximation of a chain $\{A(k)\}$ if $\sum_{k=0}^{\infty} |A_{ij}(k) - B_{ij}(k)| < \infty$ for all $i, j \in [m]$.

As an example, consider two chains $\{A(k)\}$ and $\{B(k)\}$ that may be different only for finitely many instances, i.e., there exists some time $t \geq 0$ such that $W(k) = U(k)$ for all $k \geq t$. Since every entry in each of the matrices $A(k)$ and $B(k)$ is in the interval $[0, 1]$, it follows that $\sum_{k=0}^{\infty} |A_{ij}(k) - B_{ij}(k)| \leq t$. Hence, the two models are ℓ_1-approximation of each other. As a more concrete example, we have the following.

Example 3: Consider the chain $\{A(k)\}$ of Example 2, and let the chain $\{B(k)\}$ be defined by

$$B(k) = \begin{bmatrix} \frac{1}{2} - \frac{1}{k^2+1} & \frac{1}{2} & \frac{1}{k^2+1} \\ \frac{1}{2} & \frac{1}{2} - \frac{1}{k^2+1} & \frac{1}{k^2+1} \\ \frac{1}{k^{3/2}+1} & 0 & 1 - \frac{1}{k^{3/2}+1} \end{bmatrix} \qquad \text{for all } k \geq 0.$$

For all i and j, the value $|B_{ij}(k) - A_{ij}(k)|$ is bounded by $\frac{1}{k^{3/2}+1}$ for all k. Thus, $\sum_{k=0}^{\infty} |A_{ij}(k) - B_{ij}(k)| < \infty$ for all i, j implying that $\{B(k)\}$ is an ℓ_1-approximation of $\{A(k)\}$. $\square$

Definition 6 extends to random chains by requiring that ℓ_1-approximation is almost sure. Specifically, a random chain $\{U(k)\}$ is an ℓ_1-approximation of a random chain $\{W(k)\}$ if $\sum_{k=0}^{\infty} |W_{ij}(k) - U_{ij}(k)| < \infty$ almost surely for all $i, j \in [m]$.

We have some remarks for Definition 6. First, we note that ℓ_1-approximation is an equivalence relation for deterministic chains, since the set of all absolutely summable sequences in $\mathbb{R}$ is a vector space over $\mathbb{R}$. This is also true for independent random chains $\{W(k)\}$ and $\{U(k)\}$ that are adapted to the same sigma-field. In this case, we have $\sum_{k=0}^{\infty} |W_{ij}(k) - U_{ij}(k)| < \infty$ for all $i, j \in [m]$ with either probability zero or one, due to Kolmogorov's 0-1 law ([11], page 61). Thus, $\{W(k)\}$ and $\{U(k)\}$ are ℓ_1-approximations of each other with either probability zero or one. Second, we note that there are alternative formulations of ℓ_1-approximation. Since the matrices have a finite dimension, if $\sum_{k=0}^{\infty} |A_{ij}(k) - B_{ij}(k)| < \infty$ for all $i, j \in [m]$, then $\sum_{k=0}^{\infty} \|A(k) - B(k)\|_p < \infty$ for any $p \geq 1$. Thus, an equivalent definition of ℓ_1-approximation is obtained by requiring that $\sum_{k=0}^{\infty} \|A(k) - B(k)\|_p < \infty$ for some $p \geq 1$.

Now, we present the central result of this section which will serve as one of the main tools in our study of ergodicity classes for non-ergodic chains.

Lemma 1: (*Approximation Lemma*) Let $\{B(k)\}$ and $\{A(k)\}$ be deterministic chains that are an ℓ_1-approximation of each other. Then, the chains have the same ergodicity classes.

Proof: Suppose that $i \leftrightarrow_B j$. Let $t_0 = 0$ and let $x(0) \in [0, 1]^m$. Also, let $\{x(k)\}$ be the dynamics as defined in Eq. (1) by matrices $\{A(k)\}$. For any $k \geq 0$, we have

$$x(k+1) = A(k)x(k) = (A(k) - B(k))x(k) + B(k)x(k).$$

Since $|x_i(k)| \leq 1$ for any $k \geq 0$ and any $i \in [m]$, it follows that for all $k \geq 0$,

$$\|x(k+1) - B(k)x(k)\|_\infty \leq \|A(k) - B(k)\|_\infty. \tag{2}$$

We want to show that $i \leftrightarrow_A j$, or equivalently that $\lim_{k\to\infty}(x_i(k) - x_j(k)) = 0$. To do so, we let $\epsilon > 0$ be arbitrary but fixed. Since $\{B(k)\}$ is an ℓ_1-approximation of $\{A(k)\}$, there exists time $N_\epsilon \geq 0$ such that $\sum_{k=N_\epsilon}^{\infty} \|A(k) - B(k)\|_\infty \leq \epsilon$. Let $\{z(k)\}_{k \geq N_\epsilon}$ be the dynamics given by Eq. (1), which is driven by $\{B(k)\}$ and started at time N_ϵ with the initial vector $z(N_\epsilon) = x(N_\epsilon)$. We next show that

$$\|x(k+1) - z(k+1)\|_\infty \leq \sum_{t=N_\epsilon}^{k} \|A(t) - B(t)\|_\infty \qquad \text{for all } k \geq N_\epsilon. \tag{3}$$

We use the induction on k, so we consider $k = N_\epsilon$. Then, by Eq. (2), we have $\|x(N_\epsilon+1) - B(N_\epsilon)x(N_\epsilon)\|_\infty \leq \|A(N_\epsilon) - B(N_\epsilon)\|_\infty$. Since $z(N_\epsilon) = x(N_\epsilon)$, it follows that $\|x(N_\epsilon+1) - z(N_\epsilon+1)\|_\infty \leq \|A(N_\epsilon) - B(N_\epsilon)\|_\infty$. We now assume that $\|x(k) - z(k)\|_\infty \leq \sum_{t=N_\epsilon}^{k-1} \|A(t) - B(t)\|_\infty$ for some $k > N_\epsilon$. Using Eq. (2) and the triangle inequality, we have

$$\begin{aligned}\|x(k+1) - z(k+1)\|_\infty &= \|A(k)x(k) - B(k)z(k)\|_\infty \\ &= \|(A(k) - B(k))x(k) + B(k)(x(k) - z(k))\|_\infty \\ &\leq \|(A(k) - B(k))\|_\infty \|x(k)\|_\infty + \|B(k)\|_\infty \|(x(k) - z(k))\|_\infty.\end{aligned}$$

By the induction hypothesis and relation $\|B(k)\|_\infty = 1$, which holds since $B(k)$ is a stochastic matrix, it follows that $\|x(k+1) - z(k+1)\|_\infty \leq \sum_{t=N_\epsilon}^{k} \|A(t) - B(t)\|_\infty$, thus showing relation (3).

Recalling that the time $N_\epsilon \geq 0$ is such that $\sum_{k=N_\epsilon}^{\infty} \|A(k) - B(k)\|_\infty \leq \epsilon$ and using relation (3), we obtain for all $k \geq N_\epsilon$,

$$\|x(k+1) - z(k+1)\|_\infty \leq \sum_{t=N_\epsilon}^{k} \|A(t) - B(t)\|_\infty \leq \sum_{t=N_\epsilon}^{\infty} \|A(t) - B(t)\|_\infty \leq \epsilon. \tag{4}$$

Therefore, $|x_i(k) - z_i(k)| \leq \epsilon$ and $|z_j(k) - x_j(k)| \leq \epsilon$ for any $k \geq N_\epsilon$, and by the triangle inequality we have $|(x_i(k) - x_j(k)) + (z_i(k) - z_j(k))| \leq 2\epsilon$ for any $k \geq N_\epsilon$. Since $i \leftrightarrow_B j$, it follows that $\lim_{k\to\infty}(z_i(k) - z_j(k)) = 0$ and $\limsup_{k\to\infty} |x_i(k) - x_j(k)| \leq 2\epsilon$. The preceding relation holds for any $\epsilon > 0$, implying that $\lim_{k\to\infty}(x_i(k) - x_j(k)) = 0$. Furthermore, the same analysis would go through when t_0 is arbitrary and the initial point $x(0) \in \mathbb{R}^m$ is arbitrary with $\|x(0)\|_\infty \neq 1$. Thus, we have $i \leftrightarrow_A j$.

Using the same argument and inequality (4), one can see that, if i is ergodic index for $\{B(k)\}$, then it is also an ergodic index for $\{A(k)\}$. Since ℓ_1-approximation relation is symmetric, the result follows. ■

As an example, by applying Lemma 1, we can conclude that the chains $\{B(k)\}$ and $\{A(k)\}$ of Example 3, have the same ergodicity classes, namely $\{1,2\}$ and $\{3\}$. However, the lemma does not say that the limiting values for the chains have to be the same for any of the ergodic classes. Within an ergodic class, the limiting values for the chains may differ.

Approximation lemma is a tight result with respect to the choice of ℓ_1-norm in the sense that the lemma need not hold if we consider ℓ_p-approximation with $p > 1$. To see this, consider the following 2×2 chain:

$$A(k) = \begin{bmatrix} 1 - \frac{1}{k+1} & \frac{1}{k+1} \\ \frac{1}{k+1} & 1 - \frac{1}{k+1} \end{bmatrix} \qquad \text{for all } k \geq 0.$$

The chain is doubly stochastic and $A_{ii}(k) \geq \frac{1}{2}$ for $i = 1, 2$ and all $k \geq 0$. Thus, it has weak feedback property with a feedback coefficient $\gamma = \frac{1}{2}$. Also, since $\sum_{k=0}^{\infty} \frac{1}{k+1} = \infty$, by the infinite flow theorem (Theorem 1), the chain is ergodic. For any $p > 1$, the identity chain $\{I\}$ is an ℓ_p-approximation of $\{A(k)\}$, i.e., $\sum_{k=0}^{\infty} |A_{ij}(k) - I_{ij}|^p = \sum_{k=0}^{\infty} \frac{1}{(k+1)^p} < \infty$ for $i, j = 1, 2$. However, the chain $\{I\}$ is not ergodic and, therefore, ℓ_1-norm in the approximation lemma cannot be replaced by any ℓ_p-norm for $p > 1$.

We now present a more involving consequence of Lemma 1, which relates mutual weak ergodicity of the indices to the infinite flow graph of the model. The result provides another critical tool for our characterization of the ergodicity classes of non-ergodic chains.

Lemma 2: Let $\{A(k)\}$ be a deterministic chain and let G^{∞} be its infinite flow graph. Then, $i \leftrightarrow_A j$ implies that i and j belong to the same connected component of G^{∞}.

Proof: To arrive at a contradiction, suppose that i and j belong to two different connected components $S, T \subset [m]$ of G^{∞}. Therefore, $T \subset \bar{S}$ implying that $\bar{S}$ is not empty. Also, since S is a connected component of G^{∞}, it follows that $\sum_{k=0}^{\infty} A_S(k) < \infty$. Without loss of generality, we assume that $S = \{1, \ldots, i^*\}$ for some $i^* < m$, and consider the chain $\{B(k)\}$ defined by

$$B_{ij}(k) = \begin{cases} A_{ij}(k) & \text{if } i \neq j \text{ and } i, j \in S \text{ or } i, j \in \bar{S}, \\ 0 & \text{if } i \neq j \text{ and } i \in S, j \in \bar{S} \text{ or } i \in \bar{S}, j \in S, \\ A_{ii}(k) + \sum_{\ell \in \bar{S}} A_{i\ell}(k) & \text{if } i = j \in S, \\ A_{ii}(k) + \sum_{\ell \in S} A_{i\ell}(k) & \text{if } i = j \in \bar{S}. \end{cases} \tag{5}$$

The above approximation simply sets the cross terms between S and $\bar{S}$ to zero, and adds the corresponding values to the diagonal entries to maintain the stochasticity of the matrix $B(k)$. Therefore, for the stochastic chain $\{B(k)\}$ we have

$$B(k) = \begin{bmatrix} B_1(k) & 0 \\ 0 & B_2(k) \end{bmatrix},$$

where $B_1(k)$ and $B_2(k)$ are respectively $i^* \times i^*$ and $(m - i^*) \times (m - i^*)$ matrices for all $k \geq 0$. By the assumption $\sum_{k=0}^{\infty} A_S(k) < \infty$, the chain $\{B(k)\}$ is an ℓ_1-approximation of $\{A(k)\}$.

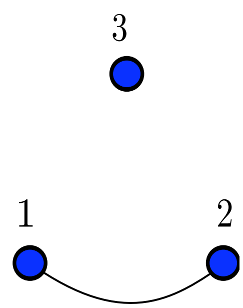

Fig. 1. The infinite flow graph for the chain $\{A(k)\}$ in Example 2.

Now, let u_{i^*} be the vector which has the first i^* coordinates equal to one and the rest equal to zero, i.e., $u_{i^*} = \sum_{\ell=1}^{i^*} e_\ell$. Then, $B(k)u_{i^*} = u_{i^*}$ for any $k \geq 0$ implying that $i \not\leftrightarrow_B j$. By approximation lemma (Lemma 1) it follows $i \not\leftrightarrow_A j$, which is a contradiction. ■

As an illustration of the result of Lemma 2, let us revisit Example 2. As discussed earlier, we only have $1 \leftrightarrow 2$, which by Lemma 2 implies that 1 and 2 must be in the same connected component of the infinite flow graph of $\{A(k)\}$. By looking at the infinite flow graph of this chain, as depicted in Figure 1, we find that 1 and 2 indeed belong to the same connected component, as predicted by the lemma.

Note that Lemma 2 applies to independent random models as well. As a special consequence of Lemma 2, we obtain that *the infinite flow property is necessary for the ergodicity* (Theorem 1 in [36]). To see this, we note that by Lemma 2, the ergodic classes of an independent random model $\{W(k)\}$ are subsets of the connected components of its infinite flow graph. When the model is ergodic, its infinite flow graph is connected, which implies that the model has infinite flow property.

The converse result of Lemma 2 is not true in general. For example, let $A(k) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ for all $k \geq 0$. In this case, the infinite flow graph is connected while the model is not ergodic. In the resulting dynamics, agents 1 and 2 keep swapping their initial values $x_1(0)$ and $x_2(0)$.

IV. Ergodicity in a class of independent random models

As discussed, by Lemma 2 the infinite flow property is necessary for ergodicity of independent random models. Also, by Theorem 1 we have that the infinite flow property is *also sufficient* for ergodicity for independent random models characterized by a common steady state $\pi > 0$ in expectation and weak feedback property. In this section, we introduce a larger class of independent random models for which the infinite flow property is also sufficient for ergodicity. This larger class is defined through a finite total variation property, which is defined as follows.

Definition 7: We say that an independent random model $\{W(k)\}$ has a finite total (expected) variation if the dynamic system (1) is such that for any $t_0 \geq 0$ and any $x(t_0) \in \mathbb{R}^m$,

$$\sum_{k=t_0}^{\infty} \sum_{i<j} H_{ij}(k) \mathsf{E}\big[(x_i(k) - x_j(k))^2\big] < \infty, \tag{6}$$

where $H(k) = \mathsf{E}\big[W^T(k)W(k)\big]$.

The terminology "total variation" is slightly abused in this definition. Specifically, relation (6) would correspond to a weighted total expected variation if all $H_{ij}(k)$ are positive. However, as we allow some of these entries to be zero, relation (6) is in fact more restrictive that the requirement that the "actual" total expected variation of the entries of $x(k)$ is finite.

When a model has a common steady state $\pi > 0$ in expectation, then, almost surely we have ([36], Theorem 5):

$$\sum_{k=0}^{\infty} \sum_{i<j} \bar{H}_{ij}(k)(x_i(k) - x_j(k))^2 < \infty,$$

where $\bar{H}(k) = \mathsf{E}\big[W^T(k)\mathrm{diag}(\pi)W(k)\big]$, where $\mathrm{diag}(\pi)$ denotes the diagonal matrix with entries π_i on its diagonal. Thus, $\min_{i\in[m]} \pi_i \, \mathsf{E}\big[W^T(k)W(k)\big] \leq \bar{H}(k)$, implying that any independent random model with a common steady state $\pi > 0$ satisfies relation (6). Thus, the class of chains with a finite total variation includes the class of chains to which the infinite flow theorem is applicable.

We further observe that the class of chains with a finite total variation includes chains that do not necessarily have a common steady state $\pi > 0$ in expectation. To see this, consider the class of deterministic chains $\{A(k)\}$ that satisfy a bounded-connectivity condition and have a uniform lower-bound on their positive entries, such as those discussed in [38], [16], [24], [25], [26]. In these models, the sequence $d(x(k)) = \max_{i\in[m]} x_i(k) - \min_{j\in[m]} x_j(k)$ is (sub)geometric and, thus, it is absolutely summable. Furthermore, $H_{ij}(k) = [A^T(k)A(k)]_{ij} \leq m$, which together with relation $|x_i(k) - x_j(k)| \leq d(x(k))$ for all $i, j \in [m]$, implies that $\sum_{k=0}^{\infty} \sum_{i<j} [A^T(k)A(k)]_{ij}(x_i(k) - x_j(k))^2 < \infty$, i.e., relation Eq. (6) holds. In conclusion, *the class of chains with a finite total variation is strictly larger than the the class of chains to which the infinite flow theorem is applicable*.

The next result establishes some properties of an ℓ_1-approximation of a chain that satisfies the assumption of the infinite flow theorem. The result will yield an immediate extension of the infinite flow theorem to some of the chains with finite total variation.

Lemma 3: Let $\{W(k)\}$ be an independent random model with a common steady state $\pi > 0$ in expectation. Let an independent random model $\{U(k)\}$ be an ℓ_1-approximation of $\{W(k)\}$. Then, $\{U(k)\}$ has a finite total variation. Furthermore, if $\{W(k)\}$ has weak-feedback property, then

$$\sum_{k=0}^{\infty}\sum_{i<j}\mathsf{E}\left[(U_{ij}(k)+U_{ji}(k))(x_i(k)-x_j(k))^2\right] < \infty.$$

Proof: Define function $V(x) = \sum_{i=1}^{m}\pi_i(x_i - \pi^T x)^2$. Let $D = \mathrm{diag}(\pi)$, $H(k) = \mathsf{E}\left[U^T(k)U(k)\right]$ and $L(k) = \mathsf{E}\left[W^T(k)DW(k)\right]$ for $k \geq 0$. Also, let $x(0) \in [0,1]^m$ and $\{x(k)\}$ be the dynamics driven by the chain $\{U(k)\}$. Then, we have for all $k \geq 0$,

$$\begin{aligned}\mathsf{E}[V(x(k+1))|x(k)] &= \mathsf{E}\left[x^T(k+1)(D-\pi\pi^T)x(k+1)|x(k)\right]\\ &= \mathsf{E}\left[(W(k)x(k)+y(k))^T(D-\pi\pi^T)(W(k)x(k)+y(k))|x(k)\right] \qquad (7)\\ &\leq \mathsf{E}[V(W(k)x(k))|x(k)] + 2\mathsf{E}\left[x(k+1)^T(D-\pi\pi^T)y(k)|x(k)\right],\end{aligned}$$

where $y(k) = (U(k)-W(k))x(k)$ and the last inequality follows by positive semi-definiteness of the matrix $D-\pi\pi^T$. By Theorem 4 in [36], we also have $\mathsf{E}[V(W(k)x(k))|x(k)] \leq V(x(k)) - \sum_{i<j}L_{ij}(k)(x_i(k)-x_j(k))^2$. Therefore, it follows that for all $k \geq 0$,

$$\begin{aligned}\mathsf{E}[V(x(k+1))|x(k)] \leq V(x(k)) &- \sum_{i<j}L_{ij}(k)(x_i(k)-x_j(k))^2\\ &+ 2\mathsf{E}\left[x(k+1)^T(D-\pi\pi^T)y(k)|x(k)\right]. \qquad (8)\end{aligned}$$

Since each $U(k)$ is stochastic, we have $x(k) \in [0,1]^m$ implying that $\|y(k)\|_\infty \leq \|W(k) - U(k)\|_\infty$. The sum of the absolute values of the entries in the ith row of $D-\pi\pi^T$ is equal to $\pi_i(2-\pi_i) \leq 1$ for all i. Thus, we have $\|(D-\pi\pi^T)y(k)\|_\infty \leq \|y(k)\|_\infty \leq \|W(k)-U(k)\|_\infty$. By $x(k) \in [0,1]^m$ for all k, we have $\|x(k)\|_1 \leq m$ implying that $x(k+1)^T(D-\pi\pi^T)y(k) \leq m\|W(k)-U(k)\|_\infty$. Since $U(k)$ and $W(k)$ are independent of $x(k)$, we obtain

$$\mathsf{E}\left[x(k+1)^T(D-\pi\pi^T)y(k)|x(k)\right] \leq m\mathsf{E}[\|W(k)-U(k)\|_\infty],$$

which when combined with Eq. (8) yields

$$\mathsf{E}[V(x(k+1))|x(k)] \leq V(x(k)) - \sum_{i<j} L_{ij}(k)(x_i(k)-x_j(k))^2 + 2m\mathsf{E}[\|W(k)-U(k)\|_\infty]\,. \tag{9}$$

Note that $\sum_{k=0}^{\infty} \mathsf{E}[\|W(k)-U(k)\|_\infty] < \infty$ since $\{U(k)\}$ is an ℓ_1-approximation of $\{W(k)\}$. Thus, by the Robbins-Siegmund theorem ([28], page 164),

$$\sum_{k=0}^{\infty}\sum_{i<j} L_{ij}(k)(x_i(k)-x_j(k))^2 < \infty \qquad \text{almost surely.}$$

The last step is to show that the difference between the sums $\sum_{k=0}^{\infty}\sum_{i<j} L_{ij}(k)(x_i(k)-x_j(k))^2$ and $\sum_{k=0}^{\infty}\sum_{i<j} H_{ij}(k)(x_i(k)-x_j(k))^2$ is finite. Since $|W_{\ell i}(k)-U_{\ell i}(k)| \leq \|W(k)-U(k)\|_\infty$ for any $i,\ell \in [m]$, using the definitions of $H(k)$ and $L(k)$, we obtain

$$\begin{aligned}
\pi_{\min} H_{ij}(k) &\leq \sum_{\ell=1}^{m} \mathsf{E}[\pi_\ell U_{\ell i}(k) U_{\ell j}(k)] \\
&= \sum_{\ell=1}^{m} \mathsf{E}[\pi_\ell (W_{\ell i}(k) + [W(k)-U(k)]_{\ell i})(W_{\ell j}(k) + [W(k)-U(k)]_{\ell j})] \\
&\leq L_{ij}(k) + \mathsf{E}\left[\|W(k)-U(k)\|_\infty \sum_{\ell=1}^{m} \pi_\ell (W_{\ell i}(k) + W_{\ell j}(k) + 1)\right],
\end{aligned} \tag{10}$$

where $\pi_{\min} = \min_{i\in[m]} \pi_i$. The last inequality is obtained by using the triangle inequality and the following

$$[W(k)-U(k)]_{\ell i}[W(k)-U(k)]_{\ell j} \leq \|W(k)-U(k)\|_\infty^2 \leq \|W(k)-U(k)\|_\infty,$$

which follows by $U_{\ell i}(k)$, $W_{\ell j}(k) \in [0,1]$. Further, from relation (10), $W_{\ell j}(k) \in [0,1]$ and the fact that π is a stochastic vector, we have

$$\pi_{\min} H_{ij}(k) \leq L_{ij}(k) + 3\mathsf{E}[\|W(k)-U(k)\|_\infty]\,.$$

Therefore,

$$\begin{aligned}
&\pi_{\min} \sum_{k=0}^{\infty}\sum_{i<j} H_{ij}(k)(x_i(k)-x_j(k))^2 \\
&\qquad \leq \sum_{k=0}^{\infty}\sum_{i<j} L_{ij}(k)(x_i(k)-x_j(k))^2 + 3\sum_{k=0}^{\infty}\sum_{i<j} \mathsf{E}[\|W(k)-U(k)\|_\infty]\,(x_i(k)-x_j(k))^2 \\
&\qquad \leq \sum_{k=0}^{\infty}\sum_{i<j} L_{ij}(k)(x_i(k)-x_j(k))^2 + 3m^2 \sum_{k=0}^{\infty} \mathsf{E}[\|W(k)-U(k)\|_\infty]\,,
\end{aligned}$$

where the last inequality holds by $(x_i(k) - x_j(k))^2 \leq 1$. Since $\sum_{k=0}^{\infty} \mathsf{E}[\|W(k) - U(k)\|_{\infty}] < \infty$ and $\pi_{\min} > 0$, and we have shown that $\sum_{k=0}^{\infty} \sum_{i<j} L_{ij}(k)(x_i(k) - x_j(k))^2 < \infty$ almost surely, it follows that $\sum_{k=0}^{\infty} \sum_{i<j} H_{ij}(k)(x_i(k) - x_j(k))^2 < \infty$ almost surely. Then, by the monotone convergence theorem ([11], page 225), we conclude that relation (6) holds, thus showing that $\{U(k)\}$ has finite total variation.

If $\{W(k)\}$ has weak feedback property, then by Eq. (9) and monotone convergence theorem ([11], page 225), we obtain

$$\gamma \sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}\left[(\bar{W}_{ij}(k) + \bar{W}_{ji}(k))(x_i(k) - x_j(k))^2\right] \leq \mathsf{E}\left[\sum_{k=0}^{\infty} \sum_{i<j} L_{ij}(k)(x_i(k) - x_j(k))^2\right]$$
$$\leq V(x(0)) + 2m \sum_{k=0}^{\infty} \mathsf{E}[\|W(k) - U(k)\|_{\infty}] < \infty.$$

Since $\{W(k)\}$ is an independent chain, it follows

$$\mathsf{E}\left[(\bar{W}_{ij}(k) + \bar{W}_{ji}(k))(x_i(k) - x_j(k))^2\right] = \mathsf{E}\left[(W_{ij}(k) + W_{ji}(k))(x_i(k) - x_j(k))^2\right].$$

Thus,

$$\sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}\left[(W_{ij}(k) + W_{ji}(k))(x_i(k) - x_j(k))^2\right] < \infty. \tag{11}$$

Then, for the chain $\{U(k)\}$ we have

$$\sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}\left[(U_{ij}(k) + U_{ji}(k))(x_i(k) - x_j(k))^2\right]$$
$$\leq \sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}\left[|U_{ij}(k) + U_{ji}(k) - (W_{ij}(k) + W_{ji}(k))|(x_i(k) - x_j(k))^2\right]$$
$$+ \sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}\left[(W_{ij}(k) + W_{ji}(k))(x_i(k) - x_j(k))^2\right]. \tag{12}$$

The last term on the right hand side of equation (12) is finite by relation (11). The first term on the right hand side of (12) can be bounded from above, by using $(x_i(k) - x_j(k))^2 \leq 1$ and the following:

$$\sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}[|U_{ij}(k) + U_{ji}(k) - (W_{ij}(k) + W_{ji}(k))|] \leq 2 \sum_{k=0}^{\infty} \mathsf{E}[\|U(k) - W(k)\|_1] < \infty.$$

where the finiteness follows from $\{U(k)\}$ being an ℓ_1-approximation of $\{W(k)\}$. Thus, we obtain $\sum_{k=0}^{\infty} \sum_{i<j} \mathsf{E}[(U_{ij}(k) + U_{ji}(k))(x_i(k) - x_j(k))^2] < \infty$. ■

We now provide a characterization of chains for which the infinite flow property is both necessary and sufficient. We use this later to extend the domain of infinite flow theorem.

Theorem 2: Let $\{W(k)\}$ be a random chain such that

$$\mathsf{E}\left[\sum_{k=t_0}^{\infty}\sum_{i<j}\left(W_{ij}(k)+W_{ji}(k)\right)\left(x_i(k)-x_j(k)\right)^2\right]<\infty,$$

for all starting times $t_0 \geq 0$ and all starting points $x(t_0) \in \mathbb{R}^m$. Then, the infinite flow property is both necessary and sufficient for ergodicity of $\{W(k)\}$. In particular, the infinite flow property is both necessary and sufficient for ergodicity of any independent random chain $\{W(k)\}$ with a finite total variation and weak feedback property.

Proof: The necessity of the infinite flow property follows by Lemma 2. For the converse, assume that the model has the infinite flow property. Let $t_0 \geq 0$ and let $x(t_0) \in \mathbb{R}^m$ be arbitrary and suppose that we have

$$\mathsf{E}\left[\sum_{k=t_0}^{\infty}\sum_{i<j}(W_{ij}(k)+W_{ji}(k))(x_i(k)-x_j(k))^2\right]<\infty,$$

which implies

$$\sum_{k=t_0}^{\infty}\sum_{i<j}(W_{ij}(k)+W_{ji}(k))(x_i(k)-x_j(k))^2<\infty \qquad \text{almost surely.}$$

Since $\{W(k)\}$ has the infinite flow property, by Lemma 3 in [37], we have $\lim_{k\to\infty}(x_{\max}(k)-x_{\min}(k))=0$ almost surely for all $t_0 \geq 0$ and $x(t_0) \in \mathbb{R}^m$, implying that the chain is ergodic.

Now, let $\{W(k)\}$ be an independent random chain with finite total variation and with weak feedback property. Then, we have

$$\sum_{k=0}^{\infty}\sum_{i<j}H_{ij}(k)\mathsf{E}\left[(x_i(k)-x_j(k))^2\right]<\infty,$$

where $H(k)=\mathsf{E}\left[W^T(k)W(k)\right]$. Due to the weak feedback property, we have $\mathsf{E}\left[W^i(k)^T W^j(k)\right] \geq \gamma\mathsf{E}[W_{ij}(k)+W_{ji}(k)]$ for all $k \geq 0$, all $i,j \in [m]$ and some $\gamma > 0$. Using the independence of the model and relation $H_{ij}(k)=\mathsf{E}\left[W^i(k)^T W^j(k)\right]$, we obtain

$$\begin{aligned}&\sum_{k=0}^{\infty}\sum_{i<j}\mathsf{E}\left[(W_{ij}(k)+W_{ji}(k))(x_i(k)-x_j(k))^2\right]\\&=\mathsf{E}\left[\sum_{k=0}^{\infty}\sum_{i<j}(W_{ij}(k)+W_{ji}(k))(x_i(k)-x_j(k))^2\right]<\infty,\end{aligned}$$

where the equality holds by $(W_{ij}(k) + W_{ji}(k))(k)(x_i(k) - x_j(k))^2 \geq 0$ and the monotone convergence theorem ([13], page 50). Therefore, by the preceding part, we conclude that the infinite flow property is both necessary and sufficient for ergodicity. ∎

Using Theorem 2 we extend the domain of the infinite flow theorem (Theorem 1) to a class of independent random models with weak feedback property and finite total variation. This class is larger than the class of models to which Theorem 1 applies, as the existence of a common steady state vector $\pi > 0$ in expectation is not required. The extension is given in the following theorem.

Theorem 3: Theorem 1 applies to any random model that has weak feedback property and finite total variation.

Proof: By Theorem 2, (d) implies (a). The implications (a) ⇒ (b) ⇒ (c) ⇒ (d) are true for any independent random model as proven in [36] (Theorem 7). ∎

With this theorem we conclude our discussion on ergodic models, and we shift our focus on the models that are not necessarily ergodic.

V. Ergodicity Classes and Infinite Flow Graph

In this section, we study models with a common steady state $\pi > 0$ in expectation and weak feedback property that are not ergodic, which is equivalent to *not having the infinite flow property*. We investigate the limiting behavior of such models and characterize their ergodicity classes. We do this by considering the infinite flow graph of a model and defining another approximation of the model, namely diagonal approximation.

Basically, for a given chain, the idea of the diagonal approximation is to construct a chain of block-diagonal matrices such that the ergodicity classes of the original chain and its diagonal approximation are the same. This can be formally done by considering the infinite flow graph of the chain and by removing the links between the connected components, at each instance of time. At the same time, in order to preserve the stochasticity of the block-diagonal matrices, the weights of the removed links are added to the self-feedback weights of corresponding agents.

We now formalize the idea discussed above. Consider an independent random model $\{W(k)\}$ and let G^∞ be its infinite flow graph. Assume that G^∞ has $\tau \geq 1$ connected components, and let $S_1, \ldots, S_\tau \subset [m]$ be the sets of vertices of the connected components in G^∞. Let $S_1 = \{1, \ldots, a_1\}$, $S_2 = \{a_1 + 1, \ldots, a_2\}, \ldots,$ $S_\tau = \{a_{\tau-1} + 1, \ldots, a_\tau = m\}$ for $1 \le a_1 \le \ldots \le a_\tau = m$, and let $m_r = |S_r| = a_r - a_{r-1}$ be the number of vertices in the rth component, where $a_0 = 0$. Using the connected components of G^∞, we define the *diagonal approximation* $\{\tilde{W}(k)\}$ of $\{W(k)\}$, as follows.

Definition 8: (*Diagonal Approximation*) Let $\{W(k)\}$ be a random model. For $1 \le r \le \tau$, let the random model $\{W^{(r)}(k)\}$ in $\mathbb{R}^{m_r}$ be given as follows: for $i, j \in [m_r]$,

$$W_{ij}^{(r)}(k) = \begin{cases} W_{(i+a_{r-1})(i+a_{r-1})}(k) + \sum_{\ell \in \bar{S}_r} W_{(i+a_{r-1})\ell}(k) & \text{if } j = i, \\ W_{(i+a_{r-1})(j+a_{r-1})}(k) & \text{if } j \ne i. \end{cases} \tag{13}$$

The diagonal approximation of the model $\{W(k)\}$ is the model $\{\tilde{W}(k)\}$ defined by

$$\tilde{W}(k) = \text{diag}(W^{(1)}(k), \ldots, W^{(\tau)}(k)) = \begin{pmatrix} W^{(1)}(k) & 0 & \cdots & 0 \\ 0 & W^{(2)}(k) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & W^{(\tau)}(k) \end{pmatrix}.$$

We note that the diagonal approximation of a random model is unique.

As an example of a diagonal approximation, consider the chain $\{B(k)\}$ of Example 3. The diagonal approximation of $\{B(k)\}$ is the chain $\{\tilde{B}(k)\}$, where

$$\tilde{B}(k) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{for all } k \ge 0.$$

Note that the diagonal approximation $\{\tilde{B}(k)\}$ coincides with the chain $\{A(k)\}$ of Example 2, which is an ℓ_1-approximation of the chain $\{B(k)\}$. This is no accident, but rather a property of diagonal approximation, as seen in our next result. The result shows that the coupling between the connected components is weak enough to guarantee that the diagonal approximation is in fact an ℓ_1-approximation. At the same time, the coupling within each of the diagonal sub-models is rather strong, as each sub-model possesses the infinite flow property.

Lemma 4: Let $\{W(k)\}$ be an independent random model and $\{\tilde{W}(k)\}$ be its diagonal approximation. Then, $\{\tilde{W}(k)\}$ is an ℓ_1-approximation of $\{W(k)\}$. Furthermore, for every $r = 1, \ldots, \tau$, the random model $\{W^{(r)}(k)\}$ of Eq. (13) has the infinite flow property.

Proof: First we show that $\tilde{W}(k)$ is a stochastic matrix for any $k \ge 0$. Due to the diagonal structure of the matrix $\tilde{W}(k)$ (Eq. (13)), it suffices to show that $W^{(r)}(k)$ is stochastic for each

r, $1 \leq r \leq \tau$. Let r be arbitrary but fixed. By the definition of $W^{(r)}(k)$, we have $W^{(r)}(k) \geq 0$. Also, for all $i \in [m_r]$, we have

$$\begin{aligned}\sum_{j=1}^{m_r} W_{ij}^{(r)}(k) &= W_{ii}^{(r)}(k) + \sum_{j \neq i, j \in [m_r]} W_{ij}^{(r)}(k) \\ &= W_{(i+a_{r-1})(i+a_{r-1})}(k) + \sum_{\ell \in \bar{S}_v} W_{(i+a_{r-1})\ell}(k) + \sum_{\ell \neq i+a_{r-1}, \ell \in S_r} W_{(i+a_{r-1})\ell}(k) \\ &= \sum_{\ell=1}^{m} W_{(i+a_r)\ell}(k) = 1.\end{aligned}$$

Now, let $i \in S_r$. Then, for any $j \neq i$, we have two cases:
(i) If $j \in S_r$, then $W_{ij}(k) = \tilde{W}_{ij}(k)$ by the definition of $\tilde{W}(k)$. Hence, $|W_{ij}(k) - \tilde{W}_{ij}(k)| = 0$.
(ii) If $j \notin S_r$, then $\tilde{W}_{ij}(k) = 0$ implying $|W_{ij}(k) - \tilde{W}_{ij}(k)| = W_{ij}(k)$.
For $j = i$, we have $\tilde{W}_{ii}(k) = W_{ii}(k) + \sum_{j \notin S_r} W_{ij}(k)$. Hence, $|W_{ij}(k) - \tilde{W}_{ij}(k)| = \sum_{j \notin S_r} W_{ij}(k)$, implying that $\sum_{j=1}^{m} |W_{ij}(k) - \tilde{W}_{ij}(k)| = 2\sum_{j \notin S_r} W_{ij}(k)$. By summing these relations over all $i \in S_r$, we obtain

$$\sum_{i \in S_r} \sum_{j=1}^{m} |W_{ij}(k) - \tilde{W}_{ij}(k)| = 2 \sum_{i \in S_r} \sum_{j \notin S_r} W_{ij}(k) \leq 2W_{S_r}(k).$$

Now, by summing the preceding inequalities over $r = 1, \ldots, \tau$, we further obtain

$$\sum_{r=1}^{\tau} \sum_{i \in S_r} \sum_{j=1}^{m} |W_{ij}(k) - \tilde{W}_{ij}(k)| \leq 2 \sum_{r=1}^{\tau} W_{S_r}(k).$$

Note that $\sum_{r=1}^{\tau} \sum_{i \in S_r} \sum_{j=1}^{m} |W_{ij}(k) - \tilde{W}_{ij}(k)| = \sum_{i,j \in [m]} |W_{ij}(k) - \tilde{W}_{ij}(k)|$. Since $S_1, \ldots, S_\tau$ are the sets of vertices of the connected components of G^∞, it follows that $\sum_{k=0}^{\infty} \sum_{r=1}^{\tau} W_{S_r}(k) < \infty$ almost surely. Therefore, we conclude that

$$\sum_{k=0}^{\infty} \sum_{i,j \in [m]}^{m} |W_{ij}(k) - \tilde{W}_{ij}(k)| < \infty \quad a.s.,$$

which proves that $\{\tilde{W}(k)\}$ is an ℓ_1-approximation of $\{W(k)\}$.

To prove that each sub-model $\{W^{(r)}(k)\}$ has the infinite flow property, let $V \subset S_r$ be nonempty but arbitrary. Since S_r is the set of vertices of the rth connected component of G^∞, there is an edge $\{i, j\} \in \mathcal{E}^\infty$ such that $i \in V$ and $j \in \bar{V}$. By the definition of $W^{(r)}(k)$, for $i_r = i - a_{r-1}$ and $j_r = j - a_{r-1}$, we have $W_{i_r j_r}^{(r)}(k) + W_{j_r i_r}^{(r)}(k) = W_{ij}(k) + W_{ji}(k)$. Since $\{i, j\} \in \mathcal{E}^\infty$, it follows

$$\sum_{k=0}^{\infty} (W_{i_r j_r}^{(r)}(k) + W_{j_r i_r}^{(r)}(k)) = \sum_{k=0}^{\infty} (W_{ij}(k) + W_{ji}(k)) = \infty,$$

thus showing that the infinite flow graph of $\{W^{(r)}(k)\}$ is connected. Hence, $\{W^{(r)}(k)\}$ has infinite flow property. ■

Lemma 4, together with approximation lemma (Lemma 1), provides us with basic tools for studying the relations between the ergodicity classes and the infinite flow graph. Recall that in Lemma 2, we showed that if i and j are mutually weakly ergodic, then i and j belong to the same connected component of the infinite flow graph. Since mutual ergodicity is more restrictive than mutual weak ergodicity, Lemma 2 is valid when i and j are mutually ergodic. We next characterize the models for which the converse statement holds.

Theorem 4: Let $\{W(k)\}$ be an independent random model with a common steady state $\pi > 0$ in expectation and weak feedback property. Then, $i \Leftrightarrow j$ if and only if i and j are in the same connected component of the infinite flow graph G^{∞} of the model.

Proof: Since mutual ergodicity implies mutual weak ergodicity, the "if" part follows from Lemma 2. To show the "only if" part, suppose that G^{∞} has τ connected component and let $\{\tilde{W}(k)\}$ be the diagonal approximation of $\{W(k)\}$. By Lemma 4, $\{\tilde{W}(k)\}$ is also an ℓ_1- approximation of $\{W(k)\}$. Since $\{W(k)\}$ has weak feedback property and $\{\tilde{W}(k)\}$ is its ℓ_1- approximation, by Lemma 3 we see that the dynamics $\{x(k)\}$ driven by $\{\tilde{W}(k)\}$ satisfies the following relation:

$$\sum_{k=0}^{\infty}\sum_{i<j}\mathsf{E}\left[(\tilde{W}_{ij}(k)+\tilde{W}_{ji}(k))(x_i(k)-x_j(k))^2\right]<\infty,$$

which implies

$$\sum_{k=0}^{\infty}\sum_{i<j}(\tilde{W}_{ij}(k)+\tilde{W}_{ji}(k))(x_i(k)-x_j(k))^2<\infty \qquad a.s.$$

Since $\{\tilde{W}(k)\}$ consists of τ decoupled dynamics, it follows that

$$\sum_{k=0}^{\infty}\sum_{\substack{i<j\\ i,j\in[m_r]}}(W_{ij}^{(r)}(k)+W_{ji}^{(r)}(k))(x_i^{(r)}(k)-x_j^{(r)}(k))^2<\infty \qquad a.s.$$

for all $r \in [\tau]$, where $x^{(r)}$ consists of the entries of $x(k)$ that are governed by $\{W^{(r)}(k)\}$. Furthermore, by Lemma 4 each $\{W^{(r)}(k)\}$ has the infinite flow property. Thus, by Theorem 2 it follows that the chain $\{W^{(r)}(k)\}$ is ergodic almost surely for all $r \in [\tau]$. This implies that $i \Leftrightarrow_{\tilde{W}} j$ for all i, j belonging to the same connected component of G^{∞}. Therefore, by ℓ_1- approximation lemma (Lemma 1) we conclude that $i \Leftrightarrow_{W} j$ for any i, j belonging to the same connected component of G^{∞}. ■

Theorem 4 implies that the model satisfying the conditions of the theorem is asymptotically stable. Furthermore, it shows that the ergodicity classes of such a model can be fully characterized by considering the connected components in the infinite flow graph of the model.

Our next result further strengthens Theorem 4 by showing that this theorem also applies to an ℓ_1-approximation of a model that satisfies the conditions of the theorem.

Theorem 5: (*Extended Infinite Flow Theorem*) Let an independent random model $\{W(k)\}$ be an ℓ_1-approximation of an independent random model with a common steady state $\pi > 0$ in expectation and weak feedback property. Let G^{∞} be the infinite flow graph of $\{W(k)\}$ and $\bar{G}^{\infty}$ be the infinite flow graph of the expected model $\{\bar{W}(k)\}$, where $\bar{W}(k) = \mathsf{E}[W(k)]$. Then, $\{W(k)\}$ is asymptotically stable almost surely and the following statements are equivalent:

(a) $i \Leftrightarrow_W j$.

(b) $i \Leftrightarrow_{\bar{W}} j$.

(c) i and j belong to the same connected component of $\bar{G}^{\infty}$.

(d) i and j belong to the same connected component of G^{∞}.

Proof: Since $\{W(k)\}$ is independent, (a) implies (b) by the dominated convergence theorem ([11] page 15). By Lemma 2, (b) implies (c). Since $0 \leq W_{ij}(k) \leq 1$ and the model is independent, by Kolmogorov's three series theorem ([7], page 63), $\sum_{k=0}^{\infty} W_{ij}(k) < \infty$ holds a.s. only if $\sum_{k=0}^{\infty} \mathsf{E}[W_{ij}(k)] < \infty$, so (c) implies (d). Finally, by Theorem 4, (d) and (a) are equivalent. ■

By Theorem 5, we have that any dynamics driven by a random model satisfying the assumptions of the theorem converges almost surely. This, however, need not be true if either $\pi > 0$ or the week feedback assumption is violated, as seen in the following examples.

Example 4: Let matrices $W(k)$ be given by

$$W(k) = \begin{bmatrix} 1 & 0 & 0 \\ u_1(k) & u_2(k) & u_3(k) \\ 0 & 0 & 1 \end{bmatrix},$$

where $u(k) = (u_1(k), u_2(k), u_3(k))^T$ are i.i.d. random vectors distributed uniformly in the probability simplex of $\mathbb{R}^3$. Then, starting from the point $x(0) = (0, \frac{1}{2}, 1)^T$, the dynamics will not converge. This model has infinite flow property and satisfies all assumptions of Theorem 5 except for the assumption $\pi > 0$. □

Example 5: Consider the random permutation model. Specifically, let $W(k)$ be the i.i.d. model with each $W(k)$ randomly and uniformly chosen from the set of permutation matrices in $\mathbb{R}^m$. Starting from any initial point, this model just permutes the coordinates of the initial point. Therefore, the dynamic is not converging for any $x(0)$ that lies outside the subspace spanned by the vector e. The model has the infinite flow property and has the common steady state $\pi = \frac{1}{m}e$ in expectation. However, the model does not have weak feedback property, since $\mathsf{E}\left[W^i(k)^T W^j(k)\right] = 0$ for $i \neq j$ while $\mathsf{E}[W_{ij}(k)] + \mathsf{E}[W_{ji}(k)] > 0$. □

VI. Applications

Here, we consider some applications of Theorem 1 and its extended variant to ergodicity classes in Theorem 5. First, we discuss the broadcast-gossip model for a time-varying network and, then, we consider a link failure process on random networks.

A. Broadcast Gossip Algorithm on Time-Varying Networks

Broadcast gossip algorithm has been presented and analyzed in [2], [3] for consensus over a static network. Here, we propose broadcast gossip algorithm for time-varying networks and provide a necessary and sufficient condition for ergodicity. Suppose that we have a network with m nodes and a sequence of simple undirected graphs $\{G(k)\}$, where $G(k) = ([m], \mathscr{E}(k))$ and the edge set $\mathscr{E}(k)$ represents the topology of the network at time k. The sequence $\{G(k)\}$ is assumed to be deterministic. Suppose that at time k, agent $i \in [m]$ wakes up with probability $\frac{1}{m}$ (independently of the past) and broadcasts its value to its neighboring agents $N_i(k) = \{j \in [m] \mid \{i, j\} \in \mathscr{E}(k)\}$. At this time, each agent $j \in N_i(k)$ updates its estimate as follows:

$$x_j(k+1) = \gamma(k) x_i(k) + (1 - \gamma(k)) x_j(k),$$

where $\gamma(k) \in (0, \gamma]$ is a mixing parameter of the system at time k and $\gamma \in (0, 1)$. The other agents keep their values unchanged, i.e., $x_j(k+1) = x_j(k)$ for $j \notin N_i(k)$. Therefore, in this case the vector $x(k)$ of agents' estimates $x_i(k)$ evolves in time according to (1) where

$$W(k) = I - \gamma(k) \sum_{j \in N_i(k)} e_j (e_j - e_i)^T \qquad \text{with probability } \tfrac{1}{m}. \tag{14}$$

Let G_b^∞ be the infinite flow graph of the broadcast gossip model, and suppose that this graph has τ connected components, namely $S_1, \ldots, S_\tau$. Using Theorem 5, we have the following result.

Lemma 5: The time-varying broadcast gossip model of (14) is asymptotically stable almost surely. Furthermore, any two agents are in the same ergodicity class if and only if they belong to the same connected component of G_b^{∞}. In particular, the model is ergodic if and only if G_r^{∞} is connected.

Proof: In view of Theorem 5, it suffices to show that the broadcast gossip model has a common steady state $\pi > 0$ in expectation and weak feedback property. Since each agent is chosen uniformly at any time instance and the graph $G(k)$ is undirected, the (random) entries $W_{ij}(k)$ and $W_{ji}(k)$ have the same distribution. Therefore, the expected matrix $\mathsf{E}[W(k)]$ is a doubly stochastic matrix for all $k \geq 0$. Since $\gamma(k) \leq \gamma < 1$, it follows that $W_{ii}(k) \geq 1 - \alpha(k) \geq \gamma$ for all $i \in [m]$ and all $k \geq 0$. When a model satisfies $W_{ii}(k) \geq \gamma > 0$ for all i and k, then the model has weak feedback property with $\frac{\gamma}{m}$, as implied by Lemma 7 in [36]. ■

The above result shows that no matter how the underlying network evolves. When the broadcast gossip algorithm is applied to a time-varying network the stability of the algorithm is guaranteed. In fact, we can provide a characterization of the connected components S_r for the infinite flow graph G_b^{∞}. By Theorem 5, it suffices to determine the infinite flow graph $\bar{G}_b^{\infty}$ of the expected model. A link $\{i, j\}$ is in the edge-set of the graph $\bar{G}_b^{\infty}$ if and only if $\sum_{k=0}^{\infty} (\mathsf{E}[W_{ij}(k)] + \mathsf{E}[W_{ji}(k)]) = \infty$. By (14), we have $\mathsf{E}[W_{ij}(k)] = \frac{1}{m}\gamma(k)$ if $j \in N(k)$ and otherwise $\mathsf{E}[W_{ij}(k)] = 0$. Thus, $\{i, j\} \in \bar{G}_b^{\infty}$ if and only if $\sum_{k:\{i,j\}\in\mathscr{E}(k)} \gamma(k) = \infty$.

Two instances of the time-varying broadcast gossip algorithm that might be of practical interest are: (1) The case when $G(k) = G$ for all $k \geq 0$. Then, the random model is ergodic if and only if G is connected and $\sum_{k=0}^{\infty} \gamma(k) = \infty$. (2) The case when the sequence $\{\gamma(k)\}$ is also bounded below i.e., $\gamma(k) \in [\gamma_b, \gamma]$ with $0 < \gamma_b \leq \gamma < 1$. Then, the model is ergodic if and only if, in the sequence $\{G(k)\}$, there are infinitely many edges between S and $\bar{S}$ for all nonempty $S \subset [m]$.

B. Link Failure Models

The application in this section is motivated by the work in [18] where the ergodicity of a random link failure model has been considered. However, the link failure model in [18] corresponds to just a random model in our setting. Here, we assume that we have an underlying random model and that there is another random process that models link failure in the random model. We use $\{W(k)\}$ to denote the underlying random model, as in Eq. (1). We let $\{F(k)\}$ denote a link failure process, which is independent of the underlying model $\{W(k)\}$. Basically, the failure process *reduces the information flow between agents* in the underlying random model $\{W(k)\}$. For the failure process, we have either $F_{ij}(k) = 0$ or $F_{ij}(k) = 1$ for all $i, j \in [m]$ and $k \geq 0$, so that $\{F(k)\}$ is a binary matrix sequence. We define the *link-failure model* as the random model $\{U(k)\}$ given by

$$U(k) = W(k) \cdot (ee^T - F(k)) + \text{diag}([W(k) \cdot F(k)]e), \tag{15}$$

where "$\cdot$" denotes the element-wise product of two matrices. To illustrate this model, suppose that we have a random model $\{W(k)\}$ and suppose that each entry $W_{ij}(k)$ is set to zero (fails), when $F_{ij}(k) = 1$. In this way, $F(k)$ induces a failure pattern on $W(k)$. The term $W(k) \cdot (ee^T - F(k))$ in Eq. (15) reflects this effect. Thus, $W(k) \cdot (ee^T - F(k))$ does not have some of the entries of $W(k)$. This lack is compensated by the feedback term which is equal to the sum of the failed links, the term $\text{diag}([W(k) \cdot F(k)]e)$. This is the same as adding $\sum_{j \neq i} [W(k) \cdot F(k)]_{ij}$ to the self-feedback weight $W_{ii}(k)$ of agent i at time k in order to ensure the stochasticity of $U(k)$.

Now, let us define *feedback property*. A random model $\{W(k)\}$ has feedback property if there is $\gamma > 0$ such that $\mathsf{E}[W_{ii}(k)W_{ij}(k) + W_{jj}(k)W_{ji}(k)] \geq \gamma \mathsf{E}[W_{ij}(k) + W_{ji}(k)]$ for any $k \geq 0$ and $i, j \in [m]$ with $i \neq j$. In general, this property is stronger than weak feedback property, as proved in [36].

Our discussion will be focused on a special class of link failure processes, which are introduced in the following definition.

Definition 9: A *uniform link-failure process* is a process $\{F(k)\}$ such that:

(a) The random variables $\{F_{ij}(k) \mid i, j \in [m],\ i \neq j\}$ are binary i.i.d. for every fixed $k \geq 0$.

(b) The process $\{F(k)\}$ is an independent process in time.

Note that the i.i.d. condition in Definition 9 is assumed for a fixed time. Therefore, the uniform link-failure model can have a time-dependent distribution but for any given time the distribution of the link-failure should be identical across the different edges.

For the uniform-link failure process, we have the following result.

Lemma 6: Let $\{W(k)\}$ be an independent model with a common steady state $\pi > 0$ in expectation and feedback property. Let $\{F(k)\}$ be a uniform-link failure process that is independent of $\{W(k)\}$. Then, the failure model $\{U(k)\}$ is ergodic if and only if $\sum_{k=0}^{\infty}(1-p_k)W_S(k)=\infty$ for all nonempty $S\ \subset [m]$, where $p_k=\Pr(F_{ij}(k)=1)$.

Proof: By the definition of $\{U(k)\}$ in (15), the failure model $\{U(k)\}$ is also independent since both $\{W(k)\}$ and $\{F(k)\}$ are independent. Then, for $i\neq j$ and for all $k\geq 0$, we have

$$\mathsf{E}[U_{ij}(k)]=\mathsf{E}[W_{ij}(k)(1-F_{ij}(k))]=(1-p_k)\mathsf{E}[W_{ij}(k)]\,, \tag{16}$$

where the last equality holds since $W_{ij}(k)$ and $F_{ij}(k)$ are independent, and $\mathsf{E}[F_{ij}(k)]=p_k$. By summing the relations in (16) over $j\neq i$ for a fixed i, we obtain $\sum_{j\neq i}\mathsf{E}[U_{ij}(k)]=(1-p_k)\sum_{j\neq i}\mathsf{E}[W_{ij}(k)]$, which by stochasticity of $W(k)$ implies $\sum_{j\neq i}\mathsf{E}[U_{ij}(k)]=(1-p_k)(1-\mathsf{E}[W_{ii}(k)])$. Since $U(k)$ is stochastic, it follows that

$$\mathsf{E}\left[U_{ii(k)}\right]=1-\sum_{j\neq i}\mathsf{E}[U_{ij}(k)]=p_k+(1-p_k)\mathsf{E}[W_{ii}(k)]\,.$$

From the preceding relation and Eq. (16), in matrix notation, the following relation holds:

$$\mathsf{E}[U(k)]=p_kI+(1-p_k)\mathsf{E}[W(k)]\qquad\text{for all } k. \tag{17}$$

Since π is a common steady state of $\{\mathsf{E}[W(k)]\}$, from (17) we obtain $\pi^T\mathsf{E}[U(k)]=\pi^T$, thus showing that $\pi>0$ is also a common steady state for $\{U(k)\}$ in expectation.

We next show that $U(k)$ has feedback property. By the definition of $U(k)$, $U_{ii}(k)\geq W_{ii}(k)$ for all $i\in[m]$ and $k\geq 0$. Hence, $\mathsf{E}[U_{ii}(k)U_{ij}(k)]\geq\mathsf{E}[W_{ii}(k)U_{ij}(k)]$. Since $\{F(k)\}$ and $\{W(k)\}$ are independent, we have

$$\begin{aligned}\mathsf{E}[W_{ii}(k)U_{ij}(k)]&=\mathsf{E}[\mathsf{E}[W_{ii}(k)U_{ij}(k)\mid F_{ij}(k)=0]]=\mathsf{E}[\mathsf{E}[W_{ii}(k)W_{ij}(k)\mid F_{ij}(k)=0]]\\&=(1-p_k)\mathsf{E}[W_{ii}(k)W_{ij}(k)]\,.\end{aligned}$$

A similar relation holds for $\mathsf{E}[U_{jj}(k)U_{ji}(k)]$. By the feedback property of $\{W(k)\}$, we have

$$\mathsf{E}[U_{ii}(k)U_{ij}(k)+U_{jj}(k)U_{ji}(k)]\geq(1-p_k)\gamma\mathsf{E}[W_{ij}(k)+W_{ji}(k)]=\gamma\mathsf{E}[U_{ij}(k)+U_{ji}(k)]\,,$$

where the last equality follows from (17) and the fact that $\gamma>0$ is a feedback constant for $\{W(k)\}$. Thus, $\{U(k)\}$ has feedback property with the same constant γ as the model $\{W(k)\}$. Hence, $\{U(k)\}$ satisfies the assumptions of Theorem 1, so the model $\{U(k)\}$ is ergodic if and only if $\sum_{k=0}^{\infty}\mathsf{E}[U_S(k)]=\infty$ for all nontrivial $S\subset[m]$. By Eq. (17) we have $\mathsf{E}[U_S(k)]=(1-p_k)\mathsf{E}[W_S(k)]$, implying that $\{U(k)\}$ is ergodic if and only if $\sum_{k=0}^{\infty}(1-p_k)\mathsf{E}[W_S(k)]=\infty$ for all nontrivial $S\subset[m]$. ■

Lemma 6 shows that the severity of a uniform link failure process cannot cause instability in the system. When the failure probabilities p_k are bounded away from 1 uniformly, i.e., $p_k \leq \bar{p}$ for all k and some $\bar{p} < 1$, it can be seen that $\sum_{k=0}^{\infty}(1-p_k)\mathsf{E}[W_S(k)] = \infty$ if and only if $\sum_{k=0}^{\infty}\mathsf{E}[W_S(k)] = \infty$. In this case, by Lemma 6 the following result is valid: *the failure model* $\{U(k)\}$ *is ergodic if and only if the original model* $\{W(k)\}$ *is ergodic*.

VII. Conclusion

In this paper, we have studied the limiting behavior of time-varying dynamics driven by random stochastic matrices. We have introduced the concept of ℓ_1-approximation of a chain and have shown that such approximations preserve the limiting behavior of the original chains, under some conditions. We have also introduced the class stochastic chains with finite total (expected) variation to which the infinite flow theorem is applicable. This non-trivially extends the class of models originally covered by the infinite flow theorem in [36]. Finally, we have identified a certain class of independent random models that are asymptotically stable almost surely. Moreover, we have characterized the equilibrium points of these models by considering their infinite flow graphs. Finally, we have applied our main result to a broadcast gossip algorithm over a time-varying network and to a link-failure model.

References

[1] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, *Bayesian learning in social networks*, Working Paper 14040, National Bureau of Economic Research, May 2008.

[2] T.C. Aysal, M.E. Yildiz, A.D. Sarwate, and A. Scaglione, *Broadcast gossip algorithms: Design and analysis for consensus*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008, pp. 4843–4848.

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

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Gossip algorithms: Design, analysis, and applications*, Proceedings IEEE Infocom, 2005, pp. 1653–1664.

[5] ———, *Randomized gossip algorithms*, IEEE Transactions on Information Theory **52** (2006), no. 6, 2508–2530.

[6] M. Cao, D.A. Spielman, and A.S. Morse, *A lower bound on convergence of a distributed network consensus algorithm*, Proceedings of 44th IEEE Conference on Decision and Control, 2005, pp. 2356 – 2361.

[7] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and S. Zampieri, *Average consensus on networks with transmission noise or quantization*, Proceedings of 26th IEEE American Control Conference, 2007, pp. 4189–4194.

[8] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in the average consensus problem*, Automatica **44** (2008), no. 3, 671–684.

[9] S. Chatterjee and E. Seneta, *Towards consensus: Some convergence theorems on repeated averaging*, Journal of Applied Probability **14** (1977), no. 1, 89–97.

[10] R. Cogburn, *On products of random stochastic matrices*, In Random matrices and their applications (1986), 199–213.

[11] R. Durrett, *Probability: Theory and examples*, 3rd ed., Curt Hinrichs, 2005.

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

[13] G. B. Folland, *Real analysis : modern techniques and their applications*, 2nd ed., Wiley, 1999.

[14] B. Golub and M.O. Jackson, *Naive learning in social networks: Convergence, influence and the wisdom of crowds*, The American Economic Journal: Microecnomics **2** (2010), no. 1, 112–149.

[15] R. Hegselmann and U. Krause, *Opinion dynamics and bounded confidence: models, analysis and simulation*, Journal of Artificial Societies and Social Simulation **5** (2002), no. 3.

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

[17] B. Johansson, *On distributed optimization in networked systems*, Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden, 2008.

[18] S. Kar and J.M.F. Moura, *Distributed consensus algorithms in sensor networks: Link failures and channel noise*, IEEE Transactions on Signal Processing **57** (2009), no. 1, 355–369.

[19] J. Lorenz, *A stabilization theorem for continuous opinion dynamics*, Physica A: Statistical Mechanics and its Applications **355** (2005), 217–223.

[20] W. Lu and T. Chen, *Synchronization analysis of linearly coupled networks of discrete time systems*, Physica D: Nonlinear Phenomena **198** (2004), no. 1–2, 148 – 168.

[21] ______, *Synchronization of coupled connected neural networks with delays*, IEEE Transactions on Circuits and Systems I **51** (2004), no. 12, 2491 – 2503.

[22] K. Nawrotzki, *Discrete open systems on markov chains in a random environment. I*, Elektronische Informationsverarbeitung und Kybernetik **17** (1981), 569–599.

[23] ______, *Discrete open systems on markov chains in a random environment. II*, Elektronische Informationsverarbeitung und Kybernetik **18** (1982), 83–98.

[24] A. Nedić and A. Ozdaglar, *On the rate of convergence of distributed subgradient methods for multi-agent optimization*, Proceedings of 46th IEEE Conference on Decision and Control, 2007, pp. 4711–4716.

[25] ______, *Distributed subgradient methods for multi-agent optimization*, IEEE Transactions on Automatic Control **54** (2009), no. 1, 48–61.

[26] ______, *Asynchronous consensus with delays*, Journal of Global Optimization **47** (2010), no. 3, 437–456.

[27] R. Olfati-Saber and R.M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control **49** (2004), no. 9, 1520–1533.

[28] A.S. Poznyak, *Advanced mathematical tools for automatic control engineers: Stochastic techniques*, Elsevier, 2009.

[29] P. Frasca R. Carli, F. Fagnani and S. Zampieri, *Gossip consensus algorithms via quantized communication*, Automatica **46** (2010), no. 1, 70–80.

[30] S.S. Ram, A. Nedić, and V.V. Veeravalli, *Distributed stochastic subgradient algorithm for convex optimization*, To appear in JOTA, Available at http://arxiv.org/abs/0811.2595, 2008.

[31] W. Ren and R.W. Beard, *Consensus seeking in multi-agent systems under dynamically changing interaction topologies*, IEEE Transactions on Automatic Control **50** (2005), no. 5, 655–661.

[32] M. Rosenblatt, *Products of independent identically distributed stochastic matrices*, Journal of Mathematical Analysis and Applications **11** (1965), no. 1, 1–10.

[33] E. Seneta, *Non-negative Matrices and Markov Chains*, 3 ed., Springer Series in Statistics, 2006.

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

[35] ______, *Consensus over ergodic stationary graph processes*, IEEE Transactions on Automatic Control **55** (2010), no. 1, 225–230.

[36] B. Touri and A. Nedić, *On ergodicity, infinite flow and consensus in random models*, `http://arxiv.org/PS_cache/arxiv/pdf/1001/1001.1890v3.pdf`, 2010.

[37] ______, *When infinite flow is sufficient for ergodicity*, Proceedings of 49th IEEE Conference on Decision and Control, 2010, pp. 7479–7486.

[38] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1984.

[39] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.