Locally Interacting Markov Chains on Heterogeneous and Random Graphs

Advisor: Kavita Ramanan
Second Reader: Stuart Geman
Graduate Mentor: Yin-Ting Liao

Miriam Gordin

April 2020

Brown University
Division of Applied Mathematics

Undergraduate Senior Honors Thesis

Contents

Abstract

Locally interacting Markov chains (LIMC) describe the joint evolution of a large collection of particles, in which the evolution of each particle depends only on its neighbors with respect to a specified underlying interaction graph. We obtain a novel characterization of the neighborhood dynamics (of a typical particle) for such models on deterministic heterogeneous graphs that are comprised of sparsely-connected dense structures. Furthermore, we investigate LIMC models for opinion dynamics on deterministic and random graphs, such as Erdős-Rényi random graphs and Galton-Watson trees, through simulations. We study and compare the efficacy of classical mean field approximations and more recently established local equation approximations, for the dynamics of the neighborhood of a typical vertex on such graphs. Our work presents further questions about understanding the marginal dynamics of LIMC on heterogeneous graph structures and the analytic and computational implications of these characterizations.

Acknowledgements. Thank you to Professor Kavita Ramanan for her guidance, support, and mentorship of my research and learning throughout my years at Brown and the completion of this thesis. Thank you to Yin-Ting Liao for his guidance and interesting discussions. Thank you to Juniper Cocomello for giving feedback on the exposition of an earlier draft.

1 Introduction

1.1 Background and Motivation

Stochastic models are useful for representing many physical, biological, social, and technological systems. In particular, interacting particle systems (IPS) are a class of high-dimensional stochastic processes, which model a large number of random particles whose local interaction is governed by an underlying graph structure. First studied in the 1960s by F. Spitzer in the United States and R. L. Dobrushin in the Soviet Union, the theory of IPS gave rise to many new questions. While the entire collection of particles in an IPS evolves as a high-dimensional Markov chain, the marginal dynamics in a small neighborhood of a typical vertex are described by a non-Markovian process. Thus, the analysis of such marginal dynamics of IPS requires different techniques than the existing theory of IPS.

The study of IPS was initiated in the context of statistical mechanics, where, for example, the stochastic Ising model represents the evolution of magnetic spins in a medium over time. In subsequent decades, IPS have been well-studied in other areas of application as well. The voter model represents the evolution of opinions in a population over time, where each particle represents a person's opinion of a particular party or viewpoint over time. The contact process models disease propagation, in which particles represent individuals or computers that may be healthy or infected at a particular time. These three models give only a sampling of the variety of applications for IPS, and in more recent years, further applications, including from neuroscience and technological systems have been explored. In all cases, IPS are effective representations for large systems that evolve randomly and have some local structure, such as physical proximity in the case of atoms in a piece of iron, or social connections in the case of the voter model. [1]

1.2 Main Contributions

In this thesis, we will consider sequences of discrete-time interacting particle systems, also called locally interacting Markov chains (LIMC) or stochastic cellular automata, whose interactions are governed by both homogeneous and heterogeneous deterministic graphs as well as random graphs. A question of interest for such a system is the marginal dynamics of a single particle, or finite number of particles, over a finite time as the total number of particles, n, goes to infinity. The main aim is to characterize the limit of the system (with respect to a suitable topology) as $n \to \infty$ and to understand the marginal dynamics at a neighborhood of a typical particle of this limit system.

From an analytic perspective, there are many properties of random systems characterized by the neighborhood dynamics that are compelling to study. The non-Markovian nature of local dynamics for sparse systems require new tools for their analysis, compared to existing characterizations for dense systems. Computationally, the neighborhood dynamics of the limit system provide an approximation for the behavior of the neighborhood of a corresponding particle in a large finite system. Finding such approximations that do not scale with n is useful for simulations, since in principle the distribution of a single particle could, in general, depend on every node in the system.

Sequences of IPS are well-characterized for densely connected interaction graph sequences and recent work by Lacker, Ramanan, and Wu provides new results for the limits and neighborhood dynamics for sequences of sparsely interacting systems. [2],[3] This report addresses the novel setting of interaction graphs that are neither dense nor sparse, but may have heterogeneous connectivity properties. In Section 4, we introduce an interaction graph structure that combines dense and sparse motifs and apply and extend existing techniques for analysis of discrete IPS to this new setting to characterize the limit and neighborhood dynamics of the heterogeneous system.

This resport is structured as follows. In Section 2, we state the definition of a LIMC and precisely define the interaction graph sequences considered. In Section 3, we give an overview of existing characterizations of random dynamics for dense and sparse graph sequences. In Section 4, we introduce the heterogeneous ring cluster model and characterize its limit and marginal dynamics. In Section 5, we provide the results of simulations of LIMC modeling opinion dynamics. Finally, in Section 6, we formulate further questions of interest and discuss the results in this report.

2 Locally Interacting Markov Chains

2.1 Notation

Given a graph $G = (V, E)$, for any vertex $v \in V$, let ∂v denote its neighborhood and set $\overline{\partial v} := \{v\} \cup \partial v$. The degree of a vertex in a graph is equal to $|\overline{\partial v}|$, the cardinality of its neighborhood, and a graph is said to be locally finite if the degree of each vertex is finite. Unless otherwise noted, we always considers graphs $G = (V, E)$ to be undirected, simple (no self-loops or multiedges), and locally finite graph with finite or countably infinite vertex set V. Often, we will consider a sequence of graphs $\{G^n\}_{n\in\mathbb{N}}$, with $G^n = (V_n, E_n)$ and $|V_n| \to \infty$ as $n \to \infty$.

Definition 2.1 (space of symmetric terminating sequences)**.** For a metric space $(\mathcal{S}, ||\cdot||)$ and any $k \in \mathbb{N}_0$, let $S^k(\mathcal{S})$ denote the space of unordered $\mathcal{S}$-valued sequences of length k, with the convention that $S^0(\mathcal{S}) = \{0\}$. Then

$$S^{\sqcup}(\mathcal{S}) := \bigsqcup_{k=0}^{\infty} S^k(\mathcal{S})$$

denotes the space of symmetric terminating sequences. [2]

For a discrete set $\mathcal{S}$, let $\mathcal{P}(\mathcal{S})$ denote the set of all probability distributions on $\mathcal{S}$. For probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ and μ on $\mathcal{S}$, we say that μ_n converges to μ weakly and write $\mu_n \Rightarrow \mu$ if for every bounded function $f : \mathcal{S} \to \mathcal{S}$,

$$\lim_{n\to\infty} \mathbb{E}_{\mu_n}[f] = \mathbb{E}_{\mu}[f].$$

Moreover, for $\mathcal{S}$-valued random variables $\{X_n\}_{n\in\mathbb{N}}$ and X, we say X_n converges to X in law and write $X_n \Rightarrow X$ if for the distributions $\{\mu_n\}_{n\in\mathbb{N}}$ and μ, of $\{X_n\}_{n\in\mathbb{N}}$ and X, respectively, $\mu_n \Rightarrow \mu$.

2.2 Main Definitions

The fundamental mathematical object that we will consider is the discrete interacting particle system, also known as a locally interacting Markov chain (LIMC) or stochastic cellular automaton, which is a high-dimensional Markov chain that evolves locally with respect to a certain underlying interaction graph.

Definition 2.2 (LIMC)**.** Given a graph $G = (V, E)$, discrete state space $\mathcal{S}$, $\{\xi_v(t)\}_{v\in V, t\in\mathbb{N}}$ independent, identically-distributed (i.i.d.) noise taking values in $[0, 1]$, a family of maps $F^0_{v,\lambda} : [0, 1] \to \mathcal{S}$, $\lambda \in \mathbb{R}, v \in V$, that govern the initial conditions, and measurable transition function $F : \mathcal{S} \times S^{\sqcup}(\mathcal{S}) \times [0, 1] \to \mathcal{S}$, the evolution dynamics of the corresponding LIMC (G, X), where $X = \{X_v(t)\}_{v\in V, t\in\mathbb{N}}$, is given by

$$\begin{aligned} X_v(0) &= F^0_{v,\lambda}(\xi_v(0)), \\ X_v(t+1) &= F(X_v(t), X_{\overline{\partial v}}(t), \xi_v(t+1)), \text{ for } t \geq 0 \end{aligned}$$

for every $v \in V$, where for any set $A \subset V$, we use $X_A(t)$ to denote the collection of random variables $\{X_v(t)\}_{v\in A}$.

2.3 Graph Structures

For examples and for simulations, we will consider a variety of deterministic and random graph structures. Recall that we take $\{G^n\}_{n\in\mathbb{N}}$ to denote a sequence of graphs with $|V_n| \to \infty$ as $n \to \infty$.

Definition 2.3. We say that a (deterministic) graph sequence $\{G^n\}$ is *dense* if $\inf_{v\in V_n} |\overline{\partial v}|$ is unbounded as $n \to \infty$.

Definition 2.4. We say that a (deterministic) graph sequence $\{G^n\}$ is *sparse* if $\sup_{v\in V_n} |\overline{\partial v}|$ is uniformly bounded for all n.

For the random graph sequences we consider, we use the following the characterizations of sparse and dense sequences.

Definition 2.5. We say that a random graph sequence $\{G^n\}$ is *dense* if

$$\inf_{n\in\mathbb{N}} \mathbb{E}_{G^n,\rho}[|\overline{\partial\rho}|] = \infty,$$

where the expectation is taken over all possible graph structures and vertices ρ chosen uniformly at random. In other words, the graph sequence is dense if the average typical degree of a node within the graph is unbounded with n.

Definition 2.6. We say that a random graph sequence $\{G^n\}$ is *sparse* if

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{G^n,\rho}[|\overline{\partial\rho}|] < \infty,$$

where the expectation is taken over all possible graph structures and vertices ρ chosen uniformly at random. In other words, the graph sequence is sparse if the average typical degree of a node within the graph is uniformly bounded with n.

The simplest deterministic examples of dense and sparse sequences are the complete and cycle (or ring) graphs on n vertices, respectively.

Definition 2.7 (complete graph)**.** The sequence of complete graphs on n vertices, $\{G^n\}_{n\in\mathbb{N}}$, also frequently denoted as $\{K_n\}$, is the sequence of graphs on vertex sets $V_n := \{1, \dots, n\}$ with edge sets $E_n := V_n \times V_n$ (all possible pairs of vertices).

Definition 2.8 (cycle graph)**.** The sequence of cycle graphs on n vertices, $\{G^n\}_{n\in\mathbb{N}}$, is the sequence of graphs on vertex set $V_n := \{1, \dots, n\}$ with edge set:

$$E_n = \{(u, v) : u, v \in V_n, |u - v| \mod n \leq 1\}.$$

Given a graph $G = (V, E)$, a cycle is a subset of the edges $\{(v, v_1), (v_1, v_2), \dots, (v_m, v)\} \subset E$ for some $m \in \mathbb{N}$ that forms a path where some vertex v is both the first and the last vertex in the path and $v, v_1, \dots, v_m$ are distinct. A tree is defined to be an acyclic graph, i.e. a graph which has no cycles. An important class of sparse graph sequences is that of finite d-regular trees.

Definition 2.9. For any $d, N \in \mathbb{N}$, we call a finite d-regular tree of N levels a tree where the root has d children and each offspring subsequently has d children for $N-1$ further generations.

We will also consider random graphs of two types: Erdős-Rényi random graphs and Galton-Watson trees.

Definition 2.10 (Erdős-Rényi random graph)**.** Given a $(0,1)$-valued sequence $\{p_n\}$, we define a sequence of Erdős-Rényi random graphs $\{\mathcal{G}(n, p_n)\}_{n\in\mathbb{N}}$ as follows. Let $\{\eta_{ij}\}_{i\geq j=1}^{n}$ be i.i.d. Bernoulli(p_n). Then for any $i, j \in V_n$, $i \in \overline{\partial j}$ if $\eta_{ij} = 1$.

Definition 2.11 (truncated Galton-Watson tree)**.** Given a probability distribution $\rho : \mathbb{N}_0 \to [0,1]$, $N \in \mathbb{N}$, $GW(\rho, N)$ denotes a random tree with a special vertex $\varnothing$, called the root, and each vertex v at a distance i with $i < N$ from the root has offspring (or neighbors at a distance $i+1$ from the root) drawn independently from ρ and the vertices at distance N have no offspring. Such a tree will be called an N-truncated Galton-Watson (GW) tree.

Definition 2.12 (truncated unimodular Galton-Watson tree)**.** Given a probability distribution $\rho : \mathbb{N}_0 \to [0,1]$, $N \in \mathbb{N}$, let $UGW(\rho, N)$ denote a random tree with diameter N with root offspring distribution ρ and subsequent offspring having distribution $\hat{\rho}$, with

$$\hat{\rho}(k) = \frac{(k+1)\rho(k+1)}{\sum_{n\in\mathbb{N}} n\rho(n)}$$

for $k \in \mathbb{N}_0$. Such a tree will be called an N-truncated unimodular Galton-Watson (UGW) tree.

Remark. It is easy to check that $\rho = \hat{\rho}$ for a UGW if and only if ρ is a Poisson distribution.

When studying the properties of random dynamics on random graph sequences, there are two sources of randomness and it is important to distinguish between quenched and annealed models. These terms originate in the statistical physics literature. In the quenched case, one studies the randomness of the dynamics, conditioned on the structure of the graph, while in the annealed setting, one averages over the randomness of both the graph structure and the dynamics. In this report, we will restrict ourselves to considering the quenched case.

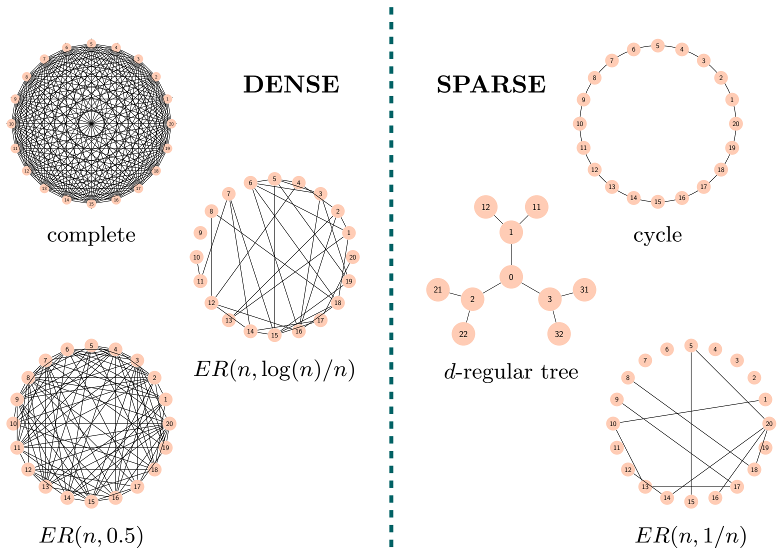


Figure 1: Diagram showing representative examples or realizations from dense and sparse graph sequences, respectively.

3 Survey of Prior Work

The main objectives of this thesis are to characterize the limits of sequences of LIMC as the size of the system grows to infinity and to provide an autonomous characterization of the marginal dynamics of a single or finite collection of particles of the limiting system. In this section, we give an overview of existing results for deterministic dense and sparse graphs sequences.

3.1 Mean Field Limits of Dense LIMC Sequences

The limits of dense LIMC are well described by mean field theory. The study of mean field interactions was introduced by McKean and Vlasov in 1966 to describe physical models of plasma, in which each particle interacts weakly with a large number of other particles [4]. Such systems can be represented by stochastic processes continuous in time and space, and are thus governed by a system of stochastic differential equations (known as the McKean-Vlasov). In the literature since then, such a mean field analysis has been carried out for a variety of other models, including LIMC, where the dynamics are discrete in time and space [5].

The key idea of the mean field limit is the concept of asymptotic independence: intuitively, the infinitesimal dynamics of each particle in a large system depend symmetrically on its neighbors, so the influence of each neighbor on the particle is inversely proportional to the degree of that vertex. Therefore, for dynamics on dense sequences of graphs, the dependence vanishes in the limits, the particles become asymptotically independent, the global empirical measure converges to a deterministic limit, and the trajectory of a single particle is described by a so-called nonlinear Markov process.

More precisely, we will state the result in [5] for the mean field limit of a discrete interacting particle systems on complete interaction graphs. In the model considered in [5], the transition function is assumed to depend symmetrically on the neighboring particles and transitions are allowed to depend on a finite memory.

Let $\mathcal{S} = \{1, \ldots, S\}$ for some $S \in \mathbb{N}$ and let $X^n = \{X_j^n(t)\}_{j\in\{1,\ldots,n\},t\in\mathbb{N}_0}$ represent the states of n interacting particles with each $X_j^n(t)$ an $\mathcal{S}$-valued random variable with dynamics given as follows. Define the empirical measure $\mu^n(t) \in \mathcal{P}(\mathcal{S})$ as follows

$$\mu_i^n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j^n(t)=i\}}, \quad i \in \mathcal{S}$$

For each $t \in \mathbb{N}_0$, let $R^n(t)$ be an $\mathbb{R}^d$-valued random variable representing the memory of the process. Let $g : \mathbb{R}^d \times \mathcal{P}(\mathcal{S}) \times \mathbb{R}^d$ be the memory update function:

$$R^n(t+1) = g(R^n(t), \mu^n(t+1)).$$

At $t = 0$, let $X^n(0)$ and $R^n(0)$ be distributed according to some probability distributions on $\mathcal{S}^n$ and $\mathbb{R}^{d\times n}$, respectively. Then let $F : \mathcal{S} \times \mathbb{R}^d \times [0, 1] \to \mathcal{S}$ be the transition function such that for all $j \in \{1, \ldots, n\}, t \in \mathbb{N}_0$,

$$X_j^n(t+1) = F(X_j^n(t), R^n(t), \xi_j(t+1))$$

where $\{\xi_v(t)\}_{v\in V,t\in\mathbb{N}_0}$ are i.i.d. random noises taking values in $[0, 1]$ with common distribution ν and F is measurable. Notice that if $g(R^n(t), \mu^n(t+1)) = g(\mu^n(t+1))$, that is, the memory update does not depend on the past memory of the particle, then this particle system is a special case of the definition of an LIMC (Definition 2.2) on a complete graphs sequence, with a dependence on the neighboring particles only through the empirical distribution.

For $a, b \in \mathcal{S}$, define the transition probability $K_{a,b} : \mathbb{R}^d \to [0, 1]$ from a to b by

$$K_{a,b}(r) = \mathbb{P}\big(X_j^n(t+1) = b \mid R^n(t) = r, X_j^n(t) = a\big), \qquad r \in \mathbb{R}^d,$$

where $r \mapsto K_{a,b}$ is continuous.

Theorem 3.1 (Theorem 4.1 of [5])**.** *If there exist $\mu(0) \in \mathcal{P}(\mathcal{S}), R(0) \in \mathbb{R}^d$ such that almost surely $\mu^n(0) \Rightarrow \mu(0)$ and $R^n \to \rho(0)$ as $n \to \infty$, then for every $t \in \mathbb{N}_0$ almost surely*

$$\mu^n(t) \Rightarrow \mu(t) \text{ and } R^n(t) \to R(t) \text{ as } n \to \infty.$$

where μ, R are defined by the evolution

$$\mu(t+1) = \mu(t) \cdot K(R(t)), \qquad R(t+1) = g(R(t), \mu(t+1)).$$

As a consequence, we can characterize the autonomous limit of a single particle as follows. Let $\{X^n\}_{n\in\mathbb{N}}$ be defined as above. Suppose there exist $\mu(0) \in \mathcal{P}(\mathcal{S}), R(0) \in \mathbb{R}^d$ such that almost surely $\mu^n(0) \Rightarrow \mu(0)$ and $R^n \to \rho(0)$ as $n \to \infty$. For v chosen uniformly at random from $\{1, \dots, n\}$, if we define recursively the $\mathcal{S}$-valued stochastic process $\{X(t)\}_{t\in\mathbb{N}_0}$ by $X(0) = X^n_v(0)$ and for all $t \geq 0$,

$$X(t+1) = F(X(t), R(t), \xi(t+1))$$

where $\{\xi(t)\}_{t\in\mathbb{N}_0}$ are i.i.d. random noises taking values in $[0,1]$ with common distribution ν and

$$R(t) = \begin{cases} g(R(0), \mu(0)) & \text{if } t = 0, \\ g(R(t-1), \mathrm{Law}(X(t))) & \text{if } t > 0. \end{cases}$$

Then by the continuity of $r \mapsto K_{a,b}$, we conclude that then $X^n_v(t) \Rightarrow X(t)$ for each $t \in \mathbb{N}_0$.

3.2 Limits of Sparse LIMC Sequences

In comparison to the dense case, an understanding of the limits of sparse LIMC sequences has developed only very recently. Specifically the work of Ramanan, Lacker, and Wu, [3], identifies the limits of large sparse systems as well as autonomously characterize the law of the evolution of a typical particle and its neighbors in the limit, in terms of a certain non-Markovian process [3], [6].

The limits of sparse systems are established in the sense of local convergence. First, we introduce some necessary notation. For a graph $G = (V, E)$ and $A \subset V$, let G_A denote the subgraph of G induced by the set A. For a "root vertex" $\rho \in V$, let $B_r(\rho) \subset V$ be the set of all vertices accessible by paths of length at most r from the vertex ρ within the graph G. Finally, we say that two graphs $G_i = (V_i, E_i)$ for $i = 1, 2$ with roots ρ_1, ρ_2, respectively, are isomorphic if there exists a bijection $\varphi : V_1 \to V_2$ such that $\varphi(\rho_1) = \rho_2$ and $(\varphi(u), \varphi(v)) \in E_2$ if and only if $(u, v) \in E_1$.

Definition 3.2 (local graph convergence)**.** We say that a sequence of connected rooted graphs $\{G^n, \rho^n\}$ converges to a graph (G, ρ) locally if for every radius $r \in \mathbb{N}$ there exists an $N = N(r)$ such that for all $m \geq N$, $G^n_{B_r(\rho^n)}$ is isomorphic to $G_{B_r(\rho)}$.

Remark. Since by our convention the limiting graph G is also locally finite, local graph convergence makes sense only for sparse graph sequences.

Example 3.3. A sequence of cycle graphs converges locally to the infinite one-dimensional lattice. (See Figure 2.) For fixed $d \in \mathbb{N}$, a sequence of d-regular trees of increasing levels converges locally to an infinite d-regular tree.

We would like to characterize limits of sparse LIMC which converge locally both in graph structure and in random dynamics, motivating the following definitions. Given a finite state space $\mathcal{S}$, let a marked rooted graph (G, ρ, x) be a graph $G = (V, E)$ with root ρ and a state $x_v \in \mathcal{S}^V$. We think of (G, ρ, x) as representing the inital conditions of a LIMC (G, X). Then we say two marked graphs (G_i, x_i) for $i = 1, 2$ with roots ρ_1, ρ_2, respectively, are isomorphic if there is an isomorphism φ from G_1 to G_2 and $(x_1)_v = (x_2)_{\varphi(v)}$ for each $v \in V_1$.

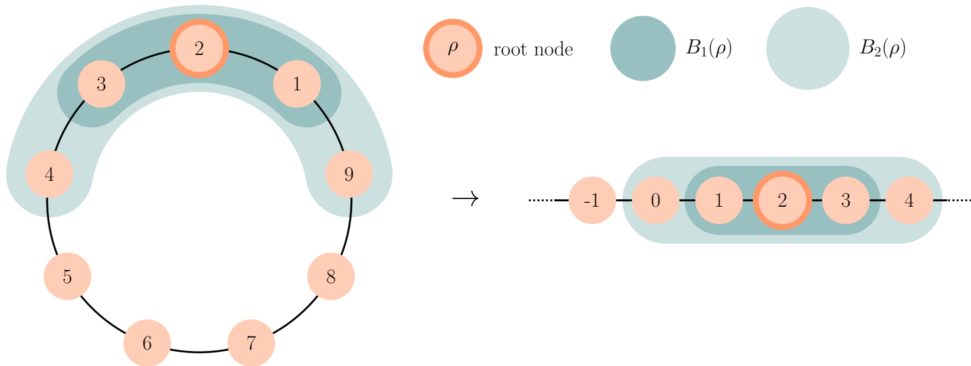


Figure 2: For any root node in a sequence of cycle graphs, as long as one chooses a large enough cycle, a ball of finite radius around the root is isomorphic to the corresponding ball on a one-dimensional infinite lattice.

Definition 3.4 (local convergence of marked graphs)**.** A sequence of marked rooted connected graphs (G^n, ρ^n, x^n) converges locally to (G, ρ, x) for every radius $r \in \mathbb{N}$, and $\varepsilon > 0$, there exists an $N = N(\rho, r, \varepsilon)$ such that for all $m \geq N$, there exists an isomorphism $\varphi : G^m_{B_r(\rho)} \to G_{B_r(\rho)}$ and

$$\sup_{v \in G^m_{B_r(\rho)}} |v - \varphi(v)| < \varepsilon.$$

In fact, the space $\mathcal{G}_*[\mathcal{S}]$ of such marked graphs has an associated metric which is compatible with this sense of convergence. Thus, for a sequence of graphs with random marks, one can say that $(G^n, X^n) \Rightarrow (G, X)$ in the sense of local convergence.

We mention two of the key results of Lacker, Ramanan, and Wu in [6] and [2]. Let (G^n, X^n) be a sequence of LIMC with sparse interaction graph sequence $\{G^n\}$. Let G be the local limit of $\{G_n\}$, and X random dynamics on G given by the same transition dynamics as X^n.

Theorem 3.5 (corresponding to Theorem 3.2 of [6] and Theorem 2.13 of [2])**.** *If F is continuous, $(G^n, X^n(0)) \Rightarrow (G, X(0))$, and G^n is a $UGW(\rho, n)$ tree, then $(G^n, X^n) \Rightarrow (G, X)$. Moreover, the trajectory of the root particle $X^n_{\rho^n}$ converges in law in $\mathcal{S}^\infty$ to the root particle X_ρ of the limiting system.*

A key intermediate result that is used in the proof of Theorem 3.5 for deterministic graphs is a certaing conditional independence property or Markov random field property, which we preface with a definition.

Definition 3.6. For any stochastic process $Y = \{Y(t)\}_{t \in \mathbb{N}}$, we denote the history, or trajectory, of Y through time t by $Y[t]$ where $Y[t] = \{Y(s)\}_{s=0}^t$.

Furthermore, for a graph $G = (V, E)$ and $A \subset V$, the boundary of A is given by

$$\partial A := \{u \in V \setminus A : (u, v) \in E \text{ for some } v \in A\}$$

and the double boundary of A by

$$\partial^2 A := \partial A \cup \partial(A \cup \partial A).$$

Theorem 3.7 (Theorem 3.2 of [2]). *Let $\{X_v(0)\}_{v \in V}$ be i.i.d. with common distribution ν. Then for each $t \in \mathbb{N}$ and for any $A \subset V$, $X_A[t] \perp X_{V \setminus (A \cup \partial^2 A)}[t] \mid X_{\partial^2 A}[t]$. In other words, the trajectories of X form a second-order Markov random field.*

The proof of Theorem 3.5 for random graphs uses a more involved conditional independence result, which also involves the randomness of the structure of the tree. We refer the interested reader to [2].

4 Heterogeneous Graph Structures

In section 3, we mentioned existing results for scaling limits and local characterizations of dynamics on dense structures and sparse structures. A novel setting for investigating these limiting properties is on heterogeneous structures which are neither dense nor sparse. Such structures arise in many systems of interest, from social networks to neuroscience [7], [8].

4.1 Ring Cluster System

To begin an investigation of heterogeneous structures, we introduce a graph structure which combines two motifs, the cycle and the complete graph. We define the ring cluster graph sequence, a sequence of growing rings decorated by densely connected clusters, which simultaneously grow in size, and study random dynamics on this graph sequence.

4.1.1 Definitions

Let $\alpha \in \mathbb{N}$ denote the fixed period of the occurrence of clusters on the cycle and let $k : \mathbb{N} \to \mathbb{N}$ be a function representing the size of a cluster such that $k(N) \to \infty$ as $N \to \infty$, where $N \in \mathbb{N}$ denotes the number of clusters. Notice that α does not depend on N.

Definition 4.1. We define the ring cluster graph structure $G^N = G(\alpha, N, k)$, as the graph on the vertex set, V_N, with vertices

$$V_N = R_N \cup \left(\bigcup_{i=0}^{N-1} C_N^i \right).$$

where $R_N := \{0, 1, 2, \ldots, \alpha N - 1\} \subset V$ represents the set of vertices on the ring and for $i = 0, \ldots N - 1$,

$$C_N^i = \{i\alpha\} \cup \{(i\alpha, n) : n = 1, \ldots, k(N) - 1\}$$

denotes the set of vertices in the ith cluster.

The (undirected) edge set E_N is given by the following: we say that $(u, v) \in E_N$, or $u \sim v$, for some $u, v \in V_N$ if one of the following conditions is met:

(i) $u, v \in R_N$ and $|u - v| \equiv 1 \mod \alpha N$.

(ii) $u, v \in C_N^i$ for some i.

See Figure 3 for a diagram of the graph structure.

Definition 4.2 (Ring cluster dynamics)**.** For a ring cluster graph $G^N = G(\alpha, N, k)$ as defined above, let $X^N = \{X_v^N(t)\}_{v \in V}$ be the discrete-time stochastic process representing the state of the dynamics on G^N, which we refer to as the *ring cluster system.* Each $X_v^N(t)$ takes values on a finite state space $\mathcal{S} := \{1, \ldots, S\}$ for some $S \in \mathbb{N}$.

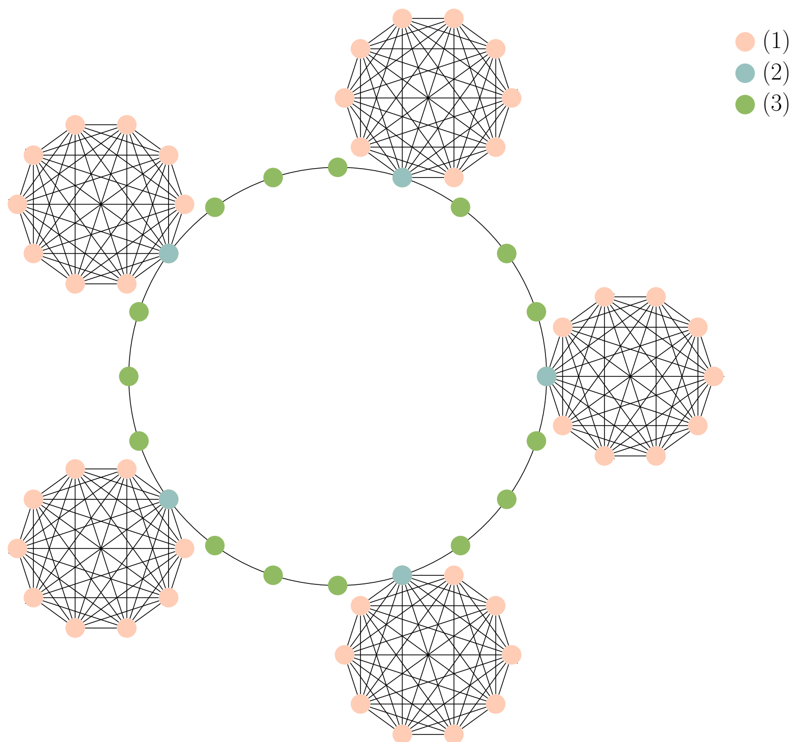


Figure 3: Diagram showing the ring cluster graph structure for $N = 5$, $\alpha = 4$, and $k(5) = 10$, with color-coding to mark (1) interior cluster, (2) cluster-ring interface, and (3) ring nodes.

Let $\lambda_v \in \mathcal{P}(\mathcal{S})$ denote the distribution of $X_v^N(0)$ for each $v \in V$. We now specify the state transitions, which have a different form for vertices in each of the following subsets: (1) interior cluster, (2) cluster-ring interface, and (3) ring. Given measurable transition functions $F_1 : \mathcal{S} \times \mathcal{P}(\mathcal{S}) \times [0,1] \to \mathcal{S}$, $F_2 : \mathcal{S} \times \mathcal{S}^3 \times \mathcal{P}(\mathcal{S}) \times [0,1] \to \mathcal{S}$, and $F_3 : \mathcal{S} \times \mathcal{S}^3 \times [0,1] \to \mathcal{S}$, we define the process X^N by the following recursive relations, for $i \in \{0, \ldots, N-1\}$,

$$\begin{aligned} X_{i\alpha,n}^N(t+1) &= F_1(X_{i\alpha,n}^N(t), \mu^{(N,i)}(t), \xi_{i\alpha,n}(t+1)) \quad \text{for } n \in \{1, \ldots, k(N)-1\}, \\ X_i^N(t+1) &= F_2(X_i^N(t), X_{\overline{\partial^R i}}^N(t), \mu^{(N,i)}(t), \xi_i(t+1)) \quad \text{if} \quad i \equiv 0 \mod \alpha, \\ X_i^N(t+1) &= F_3(X_i^N(t), X_{\overline{\partial^R i}}^N(t), \xi_i(t+1)) \quad \text{if} \quad i \not\equiv 0 \mod \alpha, \end{aligned}$$

where $\overline{\partial^R i} := \overline{\partial i} \cap R_N$ is the neighborhood of i within R_N, and

$$\mu^{(N,i)}(t) = \frac{1}{k(N)} \sum_{j=0}^{k(N)-1} \delta_{X_{i\alpha,j}^N(t)},$$

with the indexing convention that $X_{i\alpha,0}^N = X_{i\alpha}^N$. In general, we take $\{\xi_v(t)\}_{v \in V, t \in \mathbb{N}}$ to be i.i.d. with common distribution ξ, and assume without loss of generality that $\xi \sim \mathrm{Unif}([0,1])$.

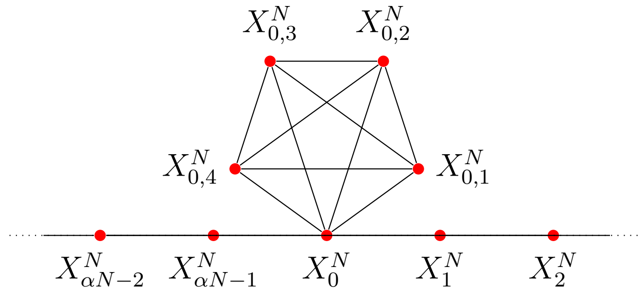


Figure 4: Schematic representation of a part of the ring cluster system for $k(N) = 5$ showing the indexing convention.

See Figure 3 for a diagram of the ring cluster graph structure and Figure 4 for a schematic showing the indexing convention.

4.1.2 Convergence Results

The convergence proof follows two main steps: first we use a mean-field approach to show the convergence of the empirical measure of each cluster to a deterministic limit, and to give an autonomous characterization of the evolution of a single particle in the interior of a cluster. Second, we identify the limiting (heterogeneous) dynamics of the particles lying on the ring and give an autonomous characterization of the evolution of the law of a finite subset particles in the limit, representative of the typical particles on the ring within the heterogeneous system, in terms of a certain non-Markovian process. Together, the arguments in these two steps completely characterize the limit dynamics of all types of particles in the ring cluster system and the evolution of the marginal dynamics for typical particles of each type.

Step 1. First under suitable conditions on the initial conditions, we show convergence of the empirical measure of each cluster. Specifically, for each $i = 0, 1, \dots, N-1$, we will show that almost surely for every $t \in \mathbb{N}_0$, there exists a (deterministic) $\mu(t) \in \mathcal{P}(\mathcal{S})$ such that

$$\mu^{(N,i)}(t) \Rightarrow \mu(t).$$

Argument for Step 1. For any $j \in \mathcal{S}$ and $\mu \in \mathcal{P}(\mathcal{S})$, denote as $P^1_{j,\cdot}(\mu)$ the transition probability distribution on $\mathcal{S}$ given by

$$P^1_{j,\cdot}(\mu) := \mathrm{Law}(F_1(j, \mu, \xi)), \tag{4.1}$$

or in other words, for any $i \in \{0, 1, \dots, N-1\}$, $n \in \{1, \dots, k(N)-1\}$ and $k \in \mathcal{S}$,

$$P^1_{j,k}(\mu) = \mathbb{P}(F_1(j, \mu, \xi) = k)\, \mathbb{P}(X^N_{i,n}(t+1) = k \mid X^N_{i,n}(t) = j, \mu^{(N,i)}(t) = \mu).$$

Similarly, for $j, k, l \in \mathcal{S}$, define

$$P^2_{j,k,l,\cdot}(\mu) := \mathrm{Law}(F_2(j, \{k, j, l\}, \mu, \xi)).$$

Assumption 4.3. For each $j,k,l \in \mathcal{S}$, $\mu \mapsto P^1_{j,\cdot}(\mu)$ and $\mu \mapsto P^2_{j,k,l,\cdot}(\mu)$ are continuous in μ, where $\mu \in \mathcal{P}(\mathcal{S})$.

If almost surely $\mu^{(N,i)}(0) \Rightarrow \mu(0)$ as $N \to \infty$ for some deterministic limit $\mu(0) \in \mathcal{P}(\mathcal{S})$, define for all $t > 0$ the candidate limit empirical measure process $\mu(t) = \{\mu_k(t)\}_{k\in\mathcal{S}} \in \mathcal{P}(\mathcal{S})$ as follows:

$$\mu_k(t+1) = \sum_{j\in\mathcal{S}} \mu_j(t) P^1_{j,k}(\mu(t)). \tag{4.2}$$

To establish the desired convergence, we will use a coupling argument. We first construct the process $\tilde{X}^N = \{\tilde{X}^N_v(t)\}_{v\in V, t\in\mathbb{N}}$ that is driven by the same noise processes and with the same interaction graphs as X^N, with similar dynamics, but for which it will be easier to show convergence of the empirical measure. For $t = 0$ and all $v \in V$, set $\tilde{X}^N_v(0) := X^N_v(0)$, and for all $t > 0$,

$$\begin{aligned}
\tilde{X}^N_{i\alpha,n}(t+1) &= F_1(\tilde{X}^N_{i\alpha,n}(t), \mu(t), \xi_{i\alpha,n}(t+1)) \text{ for } i \in \{0,\dots,N-1\}, n \in \{1,\dots,N-1\},\\
\tilde{X}^N_i(t+1) &= F_2(\tilde{X}^N_i(t), \tilde{X}^N_{\partial R_i}(t), \mu(t), \xi_i(t+1)) \text{ for } i \in \{\alpha j : j = 0,1,2,\dots,N-1\},\\
\tilde{X}^N_i(t+1) &= F_3(\tilde{X}^N_i(t), \tilde{X}^N_{\partial R_i}(t), \xi_i(t+1)) \text{ for } i \in R_N \setminus \{\alpha j : j = 0,1,2,\dots,N-1\},
\end{aligned}$$

where recall that $R_N = \{0,1,2,\dots,\alpha N-1\}$ the $\{\xi_i(t)\}_{i\in V_N, t\in\mathbb{N}_0}$ are the same noises as in the definition of the original process X. Let $\tilde{\mu}^{(N,i)}$ be the empirical measure of the process $\tilde{X}^N$, that is,

$$\tilde{\mu}^{(N,i)}(t) := \frac{1}{k(N)} \sum_{j=0}^{k(N)-1} \delta_{\tilde{X}^N_{i\alpha,j}(t)}.$$

Then by Lemma 8.1 in [5], we have that for any i, if almost surely $\mu^{(N,i)}(0) \Rightarrow \mu(0)$ as $N \to \infty$ for some deterministic limit $\mu(0)$, then for any $t \geq 0$, almost surely,

$$\tilde{\mu}^{(N,i)}(t) \Rightarrow \mu(t) \text{ as } N \to \infty.$$

Lemma 4.4. *(modification of Lemma 8.3 in [5]) Suppose $\mu^{(N,i)}(0) \Rightarrow \mu(0)$ as $N \to \infty$ for some deterministic limit $\mu(0)$ and that Assumption 4.3 holds. Then almost surely*

$$\lim_{N\to\infty} \frac{1}{k(N)} \sum_{j=0}^{k(N)-1} 1_{\{X^N_{i,j}(t) \neq \tilde{X}^N_{i,j}(t)\}} = 0, \tag{4.3}$$

and almost surely

$$\mu^{(N,i)}(t) \Rightarrow \mu(t) \text{ as } N \to \infty.$$

Proof. We follow closely the proof of Lemma 8.3 in [5]. We prove the result by induction on t. For $t = 0$, (4.3) follows trivially from the construction of $\tilde{X}$ and the fact that

$\mu^{(N,i)}(0) \Rightarrow \mu(0)$ as $N \to \infty$ by assumption. Now suppose we have that the result (4.3) holds for some $t \in \mathbb{N}_0$. Then, for any $s \in \mathcal{S}$,

$$\begin{aligned}
\frac{1}{k(N)} & \sum_{n=0}^{k(N)-1} 1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1)\}} \\
&= \frac{1}{k(N)} \sum_{n=0}^{k(N)-1} \left(1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1), X_{i,n}^N(t) \neq \tilde{X}_{i,n}^N(t)\}} + 1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1), X_{i,n}^N(t) = \tilde{X}_{i,n}^N(t)\}} \right) \\
&\leq \frac{1}{k(N)} \sum_{n=0}^{k(N)-1} 1_{\{X_{i,n}^N(t) \neq \tilde{X}_{i,n}^N(t)\}} + \sum_{s \in \mathcal{S}} A_s^N,
\end{aligned}$$

where,

$$A_s^N := \frac{1}{k(N)} \sum_{n=0}^{k(N)-1} 1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1), X_{i,n}^N(t) = \tilde{X}_{i,n}^N(t) = s\}}.$$

By the inductive hypothesis, almost surely

$$\lim_{N \to \infty} \frac{1}{k(N)} \sum_{n=0}^{k(N)-1} 1_{\{X_{i,n}^N(t) \neq \tilde{X}_{i,n}^N(t)\}} = 0,$$

so it remains to show only that $\sum_{s \in \mathcal{S}} A_s^N \to 0$ almost surely as $N \to \infty$. We have that:

$$A_s^N = \frac{1}{k(N)} 1_{\{X_{i,0}^N(t+1) \neq \tilde{X}_{i,0}^N(t+1), X_{i,0}^N(t) = \tilde{X}_{i,0}^N(t) = s\}} + \frac{1}{k(N)} \sum_{n=1}^{k(N)-1} 1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1), X_{i,n}^N(t) = \tilde{X}_{i,n}^N(t) = s\}},$$

Since the first term is bounded by $\frac{1}{k(N)}$, as $N \to \infty$ the first term goes to 0, so we need only to focus on the summation in the second term. Note since the $\{X_{i,n}^N(t+1)\}_{n=1,\ldots,N-1}$ are identically distributed, as are the $\{\tilde{X}_{i,n}^N(t+1)\}_{n=1,\ldots,N-1})$, the same method of proof as for the homogeneous system on the complete graph analyzed in [5] applies.

In particular, for each $s \in \mathcal{S}$ and $n = 1, \ldots, k(N) - 1$, we would like to rewrite the quantity $1_{\{X_{i,n}^N(t+1) \neq \tilde{X}_{i,n}^N(t+1), X_{i,n}^N(t) = \tilde{X}_{i,n}^N(t) = s\}}$ as a function of the $\{\xi_{i,n}(t+1)\}_{n \in \{1,\ldots,k(N)-1\}}$. Recalling the definition of the transition probabilities $P_{l',s}^1$ in (4.1), define:

$$\begin{aligned}
L_{s,l}^N &= \min\left(\sum_{l'=1}^{l} P_{l',s}^1(\mu^{(N,s)}(t)), \sum_{l'=1}^{l} P_{l',s}^1(\mu(t)) \right), \\
V_{s,l}^N &= \max\left(\sum_{l'=1}^{l} P_{l',s}^1(\mu^{(N,i)}(t)), \sum_{l'=1}^{l} P_{l',s}^1(\mu(t)) \right), \\
I_s^N &= \bigcup_{s=1}^{S} \left(L_{s,l}^N, V_{s,l}^N \right].
\end{aligned}$$

Thus, I_s^N is the subset of $[0,1]$ on which the cumulative distribution functions, conditional on $X_{i,n}^N(t) = \tilde{X}_{i,n}^N(t) = s$, of $X_{i,n}^N(t+1)$ and $\tilde{X}_{i,n}^N(t+1)$ differ, so we have that:

$$A_s^N = \frac{1}{k(N)} 1_{\{X_{i,0}^N(t+1) \neq \tilde{X}_{i,0}^N(t+1), X_{i,0}^N(t) = \tilde{X}_{i,0}^N(t) = s\}} + \frac{1}{k(N)} \sum_{n=1}^{k(N)-1} 1_{\{\xi_{i,n}(t+1) \in I_s^N\}}.$$

Let $G^{(N,i)}(\cdot)$ denote the empirical distribution function of the $\{\xi_{i,n}(t+1)\}_{n \in \{1,\dots,N-1\}}$, that is:

$$G^{(N,i)}(x) = \frac{1}{N-1} \sum_{n=1}^{k(N)-1} 1_{\{\xi_{i,n}(t+1) \leq x\}}$$

Then:

$$A_s^N \leq \frac{1}{k(N)} 1_{\{X_{i,0}^N(t+1) \neq \tilde{X}_{i,0}^N(t+1), X_{i,0}^N(t) = \tilde{X}_{i,0}^N(t) = s\}} + \sum_{n=1}^{k(N)-1} (G^{(N,i)}(V_{s,l}^N) - G^{(N,i)}(L_{s,l}^N))$$

By the triangle inequality,

$$\begin{aligned} A_s^N &\leq \frac{1}{k(N)} 1_{\{X_{i,0}^N(t+1) \neq \tilde{X}_{i,0}^N(t+1), X_{i,0}^N(t) = \tilde{X}_{i,0}^N(t) = s\}} \\ &\quad + \sum_{n=1}^{k(N)-1} \left(|G^{(N,i)}(V_{s,l}^N) - V_{s,l}^N| + |G^{(N,i)}(L_{s,l}^N) - L_{s,l}^N| + |V_{s,l}^N - L_{s,l}^N| \right) \\ &\leq \frac{1}{k(N)} 1_{\{X_{i,0}^N(t+1) \neq \tilde{X}_{i,0}^N(t+1), X_{i,0}^N(t) = \tilde{X}_{i,0}^N(t) = s\}} + 2S \sup_{x \in [0,1]} |G^{(N,i)}(x) - x| + \sum_{n=1}^{k(N)-1} |V_{s,l}^N - L_{s,l}^N| \end{aligned} \tag{4.4}$$

By the Glivenko-Cantelli lemma, almost surely

$$\lim_{N \to \infty} \sup_{x \in [0,1]} |G^{(N,i)}(x) - x| = 0$$

Moreover, by our induction hypothesis, we have that $\mu^{(N,i)}(t) \Rightarrow \mu(t)$ as $N \to \infty$, so since the transition probability $P_{n,k}^1(\mu)$ as defined in (4.1) is continuous in μ by our Assumption 4.3, we have that the third term in (4.4) converges to 0 almost surely. Therefore, we have that $A_s^N \to 0$ almost surely and thus (4.3) holds for $t+1$.

Moreover,

$$||\mu^{(N,i)}(t+1) - \tilde{\mu}^{(N,i)}(t+1)|| \leq \frac{2}{k(N)} \sum_{j=0}^{k(N)-1} 1_{\{X_{i,j}^N(t+1) \neq \tilde{X}_{i,j}^N(t+1)\}},$$

where the norm on the left is taken to be the L^1 norm. The left-hand side converges to 0 almost surely by our previous argument, so almost surely

$$\lim_{N \to \infty} \mu^{(N,i)}(t+1) = \mu(t+1).$$

$\square$

Thus, we have shown that in the limit, the evolution of the law of any vertex in a cluster is autonomously defined, with no influence from vertices in the ring.

STEP 2. We now study the limiting dynamics of vertices on the ring. This is similar in spirit to the limits of dynamics on sparse graphs considered in Lacker, Ramanan, and Wu [2], [6] with the difference that the dynamics are heterogeneous. Indeed, the ring consists of two types of vertices, those which sit at the interface of the ring and cluster, and those which do not. Thus, the construction of the limit system is more complicated, as is the proof of a key correlation decay property that is used in the proof of convergence (see Lemma 4.7).

ARGUMENT FOR STEP 2.

Construction 4.5 (Limit System)**.** Given $\alpha, F_1, F_2, F_3, \{\lambda_v\}_{v\in V}, \{\xi_v(t)\}_{v\in V, t\in\mathbb{N}}$ as in Definition 4.2, we define the (infinite) limit system $X = \{X_v(t)\}_{t\in\mathbb{N}, v\in V}$ as follows. Let $V = R\cup C$, where $R = \{r_i\}_{i\in\mathbb{Z}}$ and $C = \{c_i\}_{i\in\mathbb{Z}}$. For all $v \in V$, $X_v(0) \sim \lambda_v$. For all $t > 0$, we define $X_v(t+1)$ recursively as follows:

$$\begin{aligned}
X_{c_i}(t+1) &= F_1(X_{c_i}(t), \mu(t), \xi_{c_i,n}(t+1)) \quad \text{ where } \mu(t) = \mathrm{Law}(X(t)),\\
X_{r_i}(t+1) &= F_2(X_{r_i}(t), X_{\overline{\partial r_i}}(t), \mu(t), \xi_{r_i}(t+1)) \quad \text{ if } \quad i \equiv 0 \mod \alpha,\\
X_{r_i}(t+1) &= F_3(X_{r_i}(t), X_{\overline{\partial r_i}}(t), \xi_{r_i}(t+1)) \quad \text{ if } \quad i \not\equiv 0 \mod \alpha,
\end{aligned}$$

where μ is well-defined by Lemma 4.4.

We first state a technical lemma whose proof is straightforward intuitively, but somewhat notationally involved. The idea is to show that the trajectory of a fixed particle on the ring up to a finite time t depends only on initial conditions within some finite radius, as well as the cluster empirical distributions and noise within a "triangular area" of the past.

Recall that R_N denotes the set of vertices that lie in the ring structure G^N. For a radius $r \in \mathbb{N}$, and root $\rho \in R_N$, let $B_r(\rho) \subset R_N$ denote the set of all vertices in R_N connected by paths of length at most r to ρ, in other words,

$$B_r(\rho) = \{v \in V_N : |v - \rho| \leq r\} \cap R_N$$

where $|v - \rho|$ is the shortest path between v and ρ in the edge set E_n. Similarly, let $B_r(\rho) \subset R$ denote the set

$$B_r(\rho) = \{v \in R : |v - \rho| \leq r\}$$

for R as given by Construction 4.5.

Definition 4.6. Given $t \in \mathbb{N}$ and root $\rho \in \mathbb{Z}$, denote as $T(t, \rho)$ the "triangular" subset of $\mathbb{N} \times \mathbb{Z}$ given by the rule:

$$T(t, \rho) = \{(s, r) \in \mathbb{N} \times \mathbb{Z} : s \leq t, r \in [\rho - t + s, \rho + t - s]\}$$

Lemma 4.7. *Fix $t \in \mathbb{N}$ and root $\rho \in R$. Then there exists a measurable function $\varphi_{(t,\rho)} : \mathcal{S}^{2t+1} \times \mathcal{P}(\mathcal{S})^{T(t,\rho)} \times [0,1]^{T(t,\rho)} \to \mathcal{S}^t$ such that*

$$X^N_\rho[t] = \varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,. \tag{4.5}$$

Moroever, given Assumption 4.3, the mapping

$$\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)} \mapsto \mathrm{Law}\big(\varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\big) \tag{4.6}$$

is continuous.

Proof. We prove both statements by induction on t. For $t = 0$, the lemma is trivially true. Suppose we have that (4.5) holds and the mapping (4.6) is continuous for some $t \geq 0$.

Case 1. If $\rho \equiv 0 \mod \alpha$, we have that

$$X^N_\rho(t+1) = F_2(X^N_\rho(t), X^N_{\partial\rho}(t), \mu^{(N,\rho)}(t), \xi_\rho(t+1)).$$

Then by our assumption,

$$X^N_\rho[t] = \varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_r(s)\}_{(s,v)\in T(t,\rho)}\big)$$

and

$$\begin{aligned} X^N_{\partial\rho}[t] = \big\{&\varphi_{(t,\rho-1)}\big(X^N_{B_t(\rho-1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho-1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho-1)}\big)\,, \\ &\varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\ &\varphi_{(t,\rho+1)}\big(X^N_{B_t(\rho+1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho+1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho+1)}\big)\big\}\,. \end{aligned}$$

Noting that $T(t+1,\rho) = T(t,\rho-1) \cup T(t,\rho) \cup T(t,\rho+1) \cup \{(t+1,\rho)\}$ and $B_{t+1}(\rho) = B_t(\rho) \cup B_t(\rho-1) \cup B_t(\rho+1)$, we define:

$$\begin{aligned} &\varphi_{(t+1,\rho)}\big(X^N_{B_{t+1}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t+1,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t+1,\rho)}\big) \\ &\quad:= F_2\Bigg(\varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\ &\qquad \big\{\varphi_{(t,\rho-1)}\big(X^N_{B_t(\rho-1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho-1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho-1)}\big)\,, \\ &\qquad \varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\ &\qquad \varphi_{(t,\rho+1)}\big(X^N_{B_t(\rho+1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho+1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho+1)}\big)\big\}\,, \mu^{(N,\rho)}(t), \xi_\rho(t+1)\Bigg). \end{aligned}$$

By Assumption 4.3, $\mu \mapsto \mathrm{Law}(F_2(j,\{k,j,l\},\mu,\xi))$ is continuous for any $j,k,l \in \mathcal{S}$ and by our inductive assumption, $\mathrm{Law}(\varphi_{(t,\rho)}), \mathrm{Law}(\varphi_{(t,\rho)}), \mathrm{Law}(\varphi_{(t,\rho+1)})$ are continuous in the arguments $\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho-1)}, \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho+1)}$, respectively, so by composition, $\mathrm{Law}(\varphi_{(t+1,\rho)})$ is continuous in the argument $\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t+1,\rho)}$.

Case 2. If $\rho \not\equiv 0 \mod \alpha$, we have that

$$X^N_\rho(t+1) = F_3(X^N_\rho(t), X^N_{\partial\rho}(t), \xi_\rho(t+1)).$$

By a similar reasoning to Case 1, we define:

$$\begin{aligned}
&\varphi_{(t+1,\rho)}\big(X^N_{B_{t+1}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t+1,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t+1,\rho)}\big) \\
&:=F_3\Bigg(\varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\
&\quad \big\{\varphi_{(t,\rho-1)}\big(X^N_{B_t(\rho-1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho-1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho-1)}\big)\,, \\
&\quad \varphi_{(t,\rho)}\big(X^N_{B_t(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\
&\quad \varphi_{(t,\rho+1)}\big(X^N_{B_t(\rho+1)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho+1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho+1)}\big)\big\}\,, \xi_\rho(t+1)\Bigg).
\end{aligned}$$

By our inductive assumption, $\mathrm{Law}(\varphi_{(t,\rho)}), \mathrm{Law}(\varphi_{(t,\rho)}), \mathrm{Law}(\varphi_{(t,\rho+1)})$ are continuous in the arguments $\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho-1)}, \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,\rho+1)}$, respectively, so by composition, $\mathrm{Law}(\varphi_{(t+1,\rho)})$ is continuous in the argument $\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t+1,\rho)}$. Thus, $\varphi_{(t+1,\rho)}$ is defined for any ρ. Thus, we have shown that (4.5) holds and the mapping (4.6) is continuous when t is replaced with $t+1$. Thus, by the principle of induction, (4.5) holds and the mapping (4.6) is continuous for all t, which completes the proof. □

Corollary 4.7.1. *Fix $t \in \mathbb{N}$ and root $\rho \in R$. Then for $\varphi_{(t,\rho)}$ as defined in Lemma 4.7 and X as given by Construction 4.5,*

$$X_\rho[t] = \varphi_{(t,\rho)}\big(X_{B_t(\rho)}(0), \{\mu(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,.$$

Moroever, given Assumption 4.3, the mapping

$$\{\mu(s)\}_{(s,i\alpha)\in T(t,\rho)} \mapsto \mathrm{Law}\big(\varphi_{(t,\rho)}\big(X_{B_t(\rho)}(0), \{\mu(s)\}_{(s,i\alpha)\in T(t,\rho)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\big)$$

is continuous.

Proof. The proof is nearly identical to that of Lemma 4.7. □

In the following corollary, we extend the result of Lemma 4.7 to show that the trajectories up through time t of particles within a ball of radius r on the ring are a function of initial conditions within a finite radius and noise and cluster empirical distributions within a finite trapezoidal subset of the past. Let us extend the notation introduced in Definition 4.6 as follows

$$T(t, B_r(\rho)) := \bigcup_{v\in B_r(\rho)} T(t,v),$$

for $t \in \mathbb{N}_0$, root $\rho \in \mathbb{Z}$, and $r \in \mathbb{N}_0$.

Corollary 4.7.2. *In addition, fix a radius $r \in \mathbb{N}_0$. Then there exists a measurable function $\Phi_r : \mathcal{S}^{2r+2t+1} \times \mathcal{P}(\mathcal{S})^{T(t,B_r(\rho))} \times [0,1]^{T(t,B_r(\rho))} \to \mathcal{S}^{B_r(\rho)}$ such that*

$$X^N_{B_r(\rho)}[t] = \Phi_r\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)$$

and

$$\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))} \mapsto Law\big(\Phi_r\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\big) \tag{4.7}$$

is continuous.

Proof. We prove the corollary by induction on r. For $r = 0$, both statements reduce to the result in Lemma 4.7. Now assume that there exists such a Φ_r for some $r \geq 0$. Then

$$\begin{aligned} X^N_{B_{r+1}(\rho)}[t] &= \{X^N_{\rho-r-1}[t], X^N_{B_r(\rho)}[t], X^N_{\rho+r+1}[t]\} \\ &= \{\varphi_{(t,\rho-r-1)}\big(X^N_{B_t(\rho-r-1)}(0), \{\mu^{(N,i)}(s)\}_{(s,\pi i)\in T(t,\rho-r-1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\,, \\ &\quad \Phi_r\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\,, \\ &\quad \varphi_{(t,\rho+r+1)}\big(X^N_{B_t(\rho+r+1)}(0), \{\mu^{(N,i)}(s)\}_{(s,\pi i)\in T(t,\rho+r+1)}, \{\xi_v(s)\}_{(s,v)\in T(t,\rho)}\big)\} \\ &=: \Phi_{r+1}\big(X^N_{B_{r+t+1}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_{r+1}(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big) \end{aligned}$$

By Lemma 4.7 and the inductive assumption, for Φ_{r+1} defined as such, the mapping in (4.7.3) is continuous. □

Corollary 4.7.3. *In addition, fix a radius $r \in \mathbb{N}_0$. Then for $\Phi_r : \mathcal{S}^{2r+2t+1} \times \mathcal{P}(\mathcal{S})^{T(t,B_r(\rho))} \times [0,1]^{T(t,B_r(\rho))} \to \mathcal{S}^{B_r(\rho)}$ as defined in 4.7.2 and X as given by Construction 4.5,*

$$X_{B_r(\rho)}[t] = \Phi_r\big(X_{B_{r+t}(\rho)}(0), \{\mu(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)$$

and

$$\{\mu(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))} \mapsto Law\big(\Phi_r\big(X_{B_{r+t}(\rho)}(0), \{\mu(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\big)$$

is continuous.

Proof. The proof is nearly identical to that of Corollary 4.7.2. □

Theorem 4.8. *As $N \to \infty$, $(R^N, X^N_R) \Rightarrow (R, X_R)$ in the sense of local convergence (see Definition 3.4).*

We follow a similar argument to the proof of Theorem 3.5 from [2].

Proof. Fix any root $\rho \in R^N$. It is enough to show that for any time $t \in \mathbb{N}_0$, $(R^N, X^N_R[t]) \Rightarrow (R, X_R[t])$. The local convergence of the ring graph to the infinite 1-dimensional lattice is easy to show; the proof is thus omitted. It remains to show that for any $r \in \mathbb{N}_0$, $X^N_{B_r(\rho)}[t] \Rightarrow X_{B_r(\rho)}[t]$. For N large enough, i.e. for $N > 2t + 2r + 1$, $X_{B_t(v)}(0) \overset{(d)}{=} X^N_{B_t(v)}(0)$ for any $v \in B_r(\rho)$. Notice that

$$\bigcup_{v\in B_r(\rho)} B_t(v) = B_{r+t}(\rho).$$

By Corollary 4.7.2, there exists a measurable function $\Phi : \mathcal{S}^{2r+2t+1} \times \mathcal{P}(\mathcal{S})^{T(t,B_r(\rho))} \times [0,1]^{T(t,B_r(\rho))} \to \mathcal{S}^{B_r(\rho)}$ such that

$$X^N_{B_r(\rho)}[t] = \Phi\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)$$

and

$$\{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))} \mapsto \mathrm{Law}\big(\Phi\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\big)$$

is continuous. Then we have that, by 4.4,

$$\begin{aligned} &\lim_{N\to\infty} \mathrm{Law}\left(\Phi\big(X^N_{B_{r+t}(\rho)}(0), \{\mu^{(N,i)}(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\right) \\ &\qquad = \mathrm{Law}\left(\Phi\big(X_{B_{r+t}(\rho)}(0), \{\mu(s)\}_{(s,i\alpha)\in T(t,B_r(\rho))}, \{\xi_v(s)\}_{(s,v)\in T(t,B_r(\rho))}\big)\right). \end{aligned}$$

By Corollary 4.7.3 , this limit is precisely $\mathrm{Law}(X_{B_r(\rho)}[t])$. Thus, $X^N_{B_r(\rho)}[t] \Rightarrow X_{B_r(\rho)}[t]$. □

4.1.3 Local Equations for the Ring Cluster System

In this section, we focus on obtaining an autonomous characterization of the dynamics for a finite subset of nodes in the limit system defined in Construction 4.5. As shown in the proof of Proposition 4.10, this relies on a key conditional independence property which can be deduced from results in Lacker, Ramanan, and Wu [2] (see also [3]), but for which we provide a self-contained proof below. To do so, we construct a non-Markovian stochastic process that is equal in law to the corresponding marginal dynamics of the limit system.

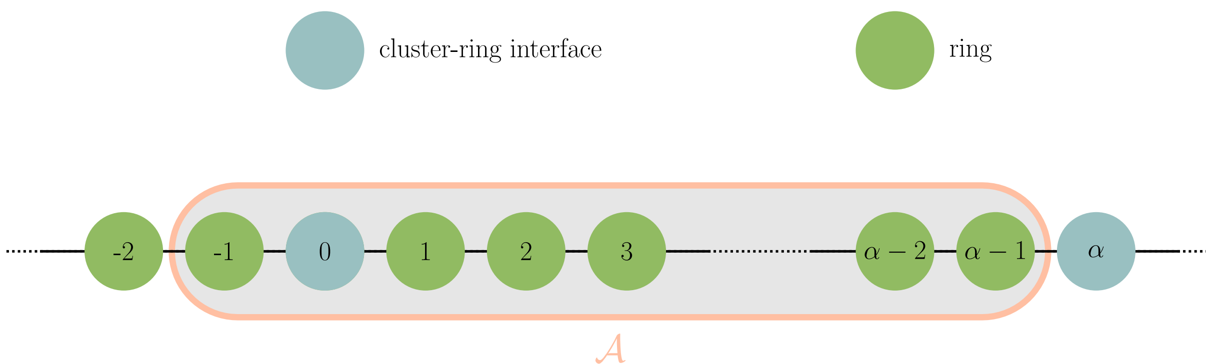


Figure 5: Diagram of the subsection of the ring structure of the limit system containing the vertices $\mathcal{A}$.

For $\alpha > 1$, let $\mathcal{A} = \{-1, 0, 1, \ldots, \alpha-1\}$, the set of node indices for which we will compute the law of the marginal dynamics. (See Figure 5.) In Construction 4.9, we provide a recursive definition of the stochastic process $\overline{X} = \{\overline{X}_v(t)\}_{v\in\mathcal{A},t\in\mathbb{N}_0}$. In Proposition 4.10, we show that $\overline{X}$ is equal in law to $X_{\mathcal{A}}$.

Construction 4.9. Fix $\alpha, F_1, F_2, F_3, \{\lambda_v\}_{v\in V}, \{\xi_v(t)\}_{v\in V,t\in\mathbb{N}}$ as in Definition 4.2. For simplicity, assume that λ_v are identical across v, meaning that we have i.i.d. initial conditions at $t = 0$. For each $t \in \mathbb{N}_0$, let $\mathcal{J}^t_{\mathcal{A}} \in \mathcal{P}(\mathcal{S}^{|\mathcal{A}|\times(t+1)})$ be the law of $\overline{X}_{\mathcal{A}}[t]$, i.e.

$$\mathcal{J}^t_{\mathcal{A}}(x_{\mathcal{A}}[t]) := \mathbb{P}(X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t])$$

for all $x_{\mathcal{A}}[t] \in \mathcal{S}^{|\mathcal{A}|\times(t+1)}$. We define $\mathcal{J}^t_{\mathcal{A}}$ recursively over $t \in \mathbb{N}_0$ as follows.

(i) Set $\mathcal{J}^0_{\mathcal{A}} = \otimes_{v\in\mathcal{A}}\lambda_v$.

(ii) For $t > 0$, assume $\mathcal{J}^t_{\mathcal{A}}$ is known. For $x_2(t) \in \mathcal{S}, x_{1,0}[t] \in \mathcal{S}^{2\times(t+1)}, x_0(t) \in \mathcal{S}, x_{2,1}[t] \in \mathcal{S}^{2\times(t+1)}$, let

$$\begin{aligned} c^t_{-2}(x_2(t), x_{1,0}[t]) &:= \mathbb{P}\big(\overline{X}_2(t) = x_2(t) \mid \overline{X}_{1,0}[t] = x_{1,0}[t]\big)\,, \\ c^t_{\alpha}(x_0(t), x_{2,1}[t]) &:= \mathbb{P}\big(\overline{X}_0(t) = x_0(t) \mid \overline{X}_{2,1}[t] = x_{2,1}[t]\big)\,, \end{aligned}$$

where c^t_{-2} is given in terms of $\mathcal{J}^t_{\mathcal{A}}$ as follows

$$c^t_{-2}(x_2(t), x_{1,0}[t]) = \frac{\sum_{\{y\in\mathcal{S}^{|\mathcal{A}|\times t}: y_2(t)=x_2(t), y_{1,0}[t]=x_{1,0}[t]\}} \mathcal{J}^t_{\mathcal{A}}(y_{\mathcal{A}}[t])}{\sum_{\{y\in\mathcal{S}^{|\mathcal{A}|\times t}: y_{1,0}[t]=x_{1,0}[t]\}} \mathcal{J}^t_{\mathcal{A}}(y_{\mathcal{A}}[t])},$$

and c^t_{α} can be represented similarly. (The reason for the notation of c^t_{-2}, c^t_{α} will become clear in Proposition 4.10.) Now, for $x_{\mathcal{A}}[t] \in \mathcal{S}^{|\mathcal{A}|\times(t+1)}$ define

$$\begin{aligned} P(x_{-1}(t+1) \mid x_{\mathcal{A}}[t]) = \sum_{x\in\mathcal{S}} &\mathbb{P}(F_3(x_{-1}(t), \{x, x_{-1}(t), x_0(t)\}, \xi_{-1}(t+1)) = x_{-1}(t+1)) \\ &\cdot c^t_{-2}(x, x_{-1,0}[t]) \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} P(x_{\alpha-1}(t+1) \mid x_{\mathcal{A}}[t]) = \sum_{x\in\mathcal{S}} &\mathbb{P}(F_3(x_{\alpha-1}(t), \{x, x_{\alpha-1}(t), x_{\alpha-2}(t)\}, \xi_{\alpha-1}(t+1)) = x_{\alpha-1}(t+1)) \\ &\cdot c^t_{\alpha}(x, x_{\alpha-1,\alpha-2}[t]), \end{aligned} \tag{4.9}$$

where the probabilities on the right depend only on the law of $\xi_v(t+1)$, which is given.

Furthermore, define

$$P(x_0(t+1) \mid x_{\mathcal{A}}[t]) := \mathbb{P}(F_2(x_0(t), x_{\partial 0}(t), \mu(t), \xi_0(t+1)) = x_0(t+1)), \tag{4.10}$$
$$P(x_v(t+1) \mid x_{\mathcal{A}}[t]) := \mathbb{P}(F_3(x_v(t), x_{\partial v}(t), \xi_v(t+1)) = x_v(t+1)) \tag{4.11}$$

for $v \in \{1, \dots, \alpha-2\}$, $x_{\mathcal{A}}[t] \in \mathcal{S}^{|\mathcal{A}|\times(t+1)}, x_0(t+1), x_v(t+1) \in \mathcal{S}$ and μ is defined as constructed in Lemma 4.4. As before, the probabilities on the right depend only on the law of $\xi_v(t+1)$, which is given. Finally let

$$\mathcal{J}^{t+1}_{\mathcal{A}}(x_{\mathcal{A}}[t+1]) := \left(\prod_{v\in\mathcal{A}} P(x_v(t+1) \mid x_{\mathcal{A}}[t])\right) \cdot \mathcal{J}^t_{\mathcal{A}}(x_{\mathcal{A}}[t]).$$

In the following proposition, we show that $\overline{X}_{\mathcal{A}} \overset{(d)}{=} X_{\mathcal{A}}$, and thereby also verify that the construction of $\mathcal{J}_{\mathcal{A}}$ indeed yields a valid probability distribution.

Proposition 4.10. *Fix $\alpha, F_1, F_2, F_3, \{\lambda_v\}_{v\in V}, \{\xi_v(t)\}_{v\in V, t\in\mathbb{N}}$ as in Definition 4.2 and let X be as defined in Construction 4.5 and $\overline{X}_{\mathcal{A}}$ have law $\mathcal{J}_{\mathcal{A}}$, as given by Construction 4.9. Then $\overline{X}_{\mathcal{A}} \overset{(d)}{=} X_{\mathcal{A}}$.*

Proof. We prove the result by induction on t. For $t = 0$, by the i.i.d. assumption, it is clear that $\overline{X}_{\mathcal{A}}(0) \overset{(d)}{=} X_{\mathcal{A}}(0)$. Now fix $t > 0$ and assume that $\overline{X}_{\mathcal{A}}[t] \overset{(d)}{=} X_{\mathcal{A}}[t]$. Thus, we need only show that:

$$\mathbb{P}(X_{\mathcal{A}}(t+1) = x_{\mathcal{A}}(t+1) \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]) = \mathbb{P}\big(\overline{X}_{\mathcal{A}}(t+1) = x_{\mathcal{A}}(t+1) \mid \overline{X}_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big)$$

By the independence of the $\xi_v(t+1)$ over $v \in \mathcal{A}$, we have that $X_v(t+1)$ are independent given $X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]$ for $v \in \mathcal{A} \setminus \{-1, \alpha-1\} = \{0, 1, \ldots, \alpha-2\}$, so

$$
\begin{aligned}
&\mathbb{P}(X_{\mathcal{A}}(t+1) = x_{\mathcal{A}}(t+1) \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]) \\
&\quad = \mathbb{P}\big(X_{-1,\alpha-1}(t+1) = x_{-1,\alpha-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\qquad\quad \cdot \prod_{v \in \mathcal{A}\setminus\{-1,\alpha-1\}} \mathbb{P}(X_v(t+1) = x_v(t+1) \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t])\,.
\end{aligned}
$$

By the Markovian transition dynamics of X, we have that for every $v \in \mathcal{A} \setminus \{-1, \alpha-1\}$,

$$
\mathbb{P}(X_v(t+1) = x_v(t+1) \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]) = P(x_v(t+1) \mid x_{\mathcal{A}}[t]),
$$

where P is as defined in Equations (4.10) and (4.11). Applying the conditional independence property 3.7, we have that

$$
\begin{aligned}
&\mathbb{P}\big(X_{-1,\alpha-1}(t+1) = x_{-1,\alpha-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \mathbb{P}\big(X_{-1}(t+1) = x_{-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\qquad \cdot \mathbb{P}\big(X_{\alpha-1}(t+1) = x_{\alpha-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big)
\end{aligned}
\tag{4.12}
$$

Now, taking the first term in the product,

$$
\begin{aligned}
&\mathbb{P}\big(X_{-1}(t+1) = x_{-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x \in \mathcal{S}} \mathbb{P}\big(X_{-1}(t+1) = x_{-1}(t+1), X_{-2}(t) = x \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x \in \mathcal{S}} \mathbb{P}\big(X_{-1}(t+1) = x_{-1}(t+1) \mid X_{-2}(t) = x, X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\qquad\qquad \cdot \mathbb{P}\big(X_{-2}(t) = x \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x \in \mathcal{S}} \mathbb{P}(X_{-1}(t+1) = x_{-1}(t+1) \mid X_{-2}(t) = x, X_{-1,0}(t) = x_{-1,0}(t)) \cdot \mathbb{P}(X_{-2}(t) = x \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]) \\
&\quad = \sum_{x \in \mathcal{S}} \mathbb{P}(F_3(x_{-1}(t), \{x, x_{-1}(t), x_0(t)\}, \xi_{-1}(t+1)) = x_{-1}(t+1))) \cdot c^t_{-2}(x, x_{-1,0}[t]) \\
&\quad = P(x_{-1}(t+1) \mid x_{\mathcal{A}}[t])
\end{aligned}
$$

as defined in Construction 4.9, where the third equality follows from the Markovian transition dynamic of X and the fourth from the conditional independence property 3.7.

Similarly, we can rewrite the second term of the right-hand side of equation (4.12) as

$$
\begin{aligned}
&\mathbb{P}\big(X_{\alpha-1}(t+1) = x_{\alpha-1}(t+1) \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x\in\mathcal{S}} \mathbb{P}\big(X_{\alpha-1}(t+1) = x_{\alpha-1}(t+1), X_{\alpha}(t) = x \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x\in\mathcal{S}} \mathbb{P}\big(X_{\alpha-1}(t+1) = x_{\alpha-1}(t+1) \mid X_{\alpha}(t) = x, X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\qquad\qquad \cdot \mathbb{P}\big(X_{\alpha}(t) = x \mid X_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1) = x_{\mathcal{A}\setminus\{-1,\alpha-1\}}(t+1), X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]\big) \\
&\quad = \sum_{x\in\mathcal{S}} \mathbb{P}(X_{\alpha-1}(t+1) = x_{\alpha-1}(t+1) \mid X_{\alpha}(t) = x, X_{-1,0}(t) = x_{-1,0}(t)) \cdot \mathbb{P}(X_{\alpha}(t) = x \mid X_{\mathcal{A}}[t] = x_{\mathcal{A}}[t]) \\
&\quad = \sum_{x\in\mathcal{S}} \mathbb{P}(F_3(x_{\alpha-1}(t), \{x, x_{\alpha-1}(t), x_{\alpha-2}(t)\}, \xi_{\alpha-1}(t+1)) = x_{-1}(t+1))) \cdot c^t_{\alpha}(x, x_{\alpha-1,\alpha-2}[t]) \\
&\quad = P(x_{\alpha-1}(t+1) \mid x_{\mathcal{A}}[t])
\end{aligned}
$$

Thus, we conclude that

$$\mathbb{P}(X_{\mathcal{A}}[t+1] = x_{\mathcal{A}}[t+1]) = \mathbb{P}\big(\overline{X}_{\mathcal{A}}[t+1] = x_{\mathcal{A}}[t+1]\big)\,.$$

$\square$

5 Opinion Dynamics on Deterministic and Random Graph Structures

In this section, we consider LIMC with two particular transition dynamics, both representing opinion formation: the voter model and majority dynamics. For simplicity, we restrict ourselves to a binary space $\{0,1\}$, where the two values could represent, for instance, two different political parties or yes and no votes for a referendum. In the voter model, at every time t, each person in the system decides their opinion at $t+1$ based on the average opinion of their neighbors within the interaction graph. In the majority dynamics, the decision is based on the majority opinion of the neighbors, with ties broken by a fair coin flip.

In Sections 5.1 and 5.2 we give numerical results for the voter model and majority dynamics, respectively, on a variety of sparse and dense, deterministic and random graph sequences.

5.1 Voter Model on Deterministic and Random Graphs

The voter model is a particular family of locally interacting Markov dynamics which represents opinion formation. Let the state space $\mathcal{S} = \{0,1\}$ denote two possible opinions, and let $X_v(t)$ represent the opinion of person v at time t. Then the voter model describes the dynamics in which at each time $t+1$, person v decides their opinion by flipping a biased coin, with the probability of choosing opinion 1 given by the proportion of v's neighbors holding opinion 1. More formally, in the notation of Definition 2.2, the voter model dynamics is are specified as follows.

Definition 5.1 (Voter model)**.** Let $\{\xi_v(t)\}_{v\in V, t\in\mathbb{N}}$ be i.i.d. with the Unif($[0,1]$) distribution. Then, fixing a parameter $\lambda \in [0,1]$, for any $v \in V$, $t \in \mathbb{N}_0$, in the notation of Definition 2.2, the voter model transition functions are given by

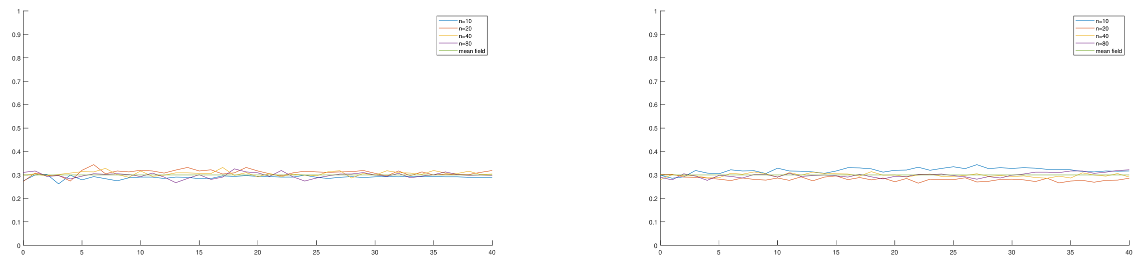
$$\begin{aligned} F^0_{v,\lambda}(\xi) &:= \mathbb{1}_{\{\xi\le\lambda\}} \\ F(x_v(t), x_{\overline{\partial v}(t)}, \xi) &:= \mathbb{1}_{\{\xi \le \frac{1}{|\overline{\partial v}|}\sum_{u\in\overline{\partial v}} x_u(t)\}} \text{ for } t \ge 0. \end{aligned}$$

Thus, for $t \ge 0$:

$$X_v(t+1) = \begin{cases} 1 & \text{w.p. } \frac{1}{|\overline{\partial v}|}\sum_{u\in\overline{\partial v}} X_u(t) \\ 0 & \text{w.p. } 1 - \frac{1}{|\overline{\partial v}|}\sum_{u\in\overline{\partial v}} X_u(t) \end{cases}$$

5.1.1 Dynamics of a single particle

A question of interest is the marginal evolution of the opinion of a single person within the system, that is, for a fixed $v \in V$ we want to characterize the evolution of the marginal distribution of $X_v(t)$, $t \in \mathbb{N}_0$. We can investigate this question computationally



complete graph ring graph

Figure 6: Computational estimates of the marginal probability (5.1) for simulated voter model dynamics for time $t = 0, \ldots, 40$ on deterministic graph structures with initial parameter $\lambda = 0.3$ for vertex set sizes $n = 10, 20, 40, 80$. ($I = 10^3$.) The corresponding mean field approximation is indicated by the green line.

by simulating the voter model dynamics and computing the following numerical estimator for the marginal probability

$$P_v^{(I)}(t) := \sum_{i=1}^{I} \mathbb{1}_{\{X_v^{(i)}(t)=1\}}, \tag{5.1}$$

where I denotes the number of iterations of the simulation, and $X^{(i)}$ is the realization of the simulated dynamics on the ith iteration. Figure 6 visualizes the numerically computed results for the marginal probability, which motivates us to formulated the following conjecture.

From the computational results shown in Figures 6 and 7, we see that on average the empirical marginal probability stays constant and equal to the initial probability of the node taking value 1 at time 0, regardless of the graph structure. This motivates us to formulate the following conjecture.

Conjecture 5.2. *For the voter model dynamics on any graph, as given in Definition 5.1,*

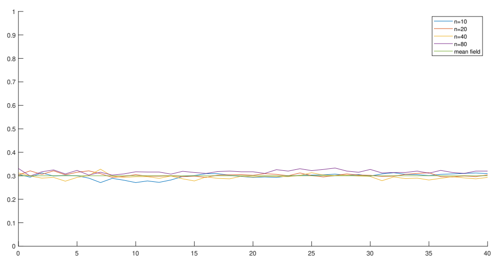
$$X_v(t) = \begin{cases} 1 & w.p.\ \lambda, \\ 0 & w.p.\ 1-\lambda \end{cases}$$

for any $v \in V, t \in \mathbb{N}$.

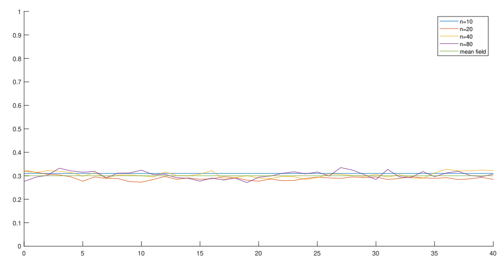
Lemma 5.3. *Conjecture 5.2 holds for $t = 1$.*

Proof. Let $N := |\overline{\partial v}|$.

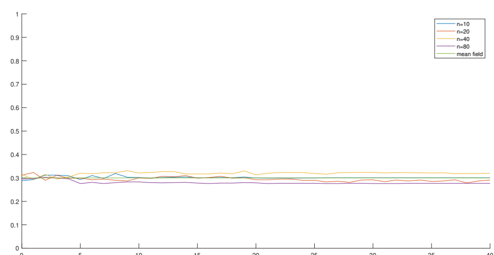
$$\begin{aligned}
\mathbb{P}(X_v(1) = 1) &= \sum_{k=0}^{N} \mathbb{P}\left(X_v(1) = 1 \,\middle|\, \sum_{i \in \overline{\partial v}} X_i(0) = k \right) \mathbb{P}\left(\sum_{i \in \overline{\partial v}} X_i(0) = k \right) \\
&= \sum_{k=0}^{N} \frac{k}{N} \binom{N}{k} \lambda^k (1-\lambda)^{N-k} = \sum_{k=1}^{N} \frac{k}{N} \frac{N!}{k!(N-k)!} \lambda^k (1-\lambda)^{N-k}
\end{aligned}$$



$p_n = 0.3$



$p_n = \log(n)/n$



$p_n = 1/n$

Figure 7: Computational estimates of the marginal probability (5.1) for simulated voter model dynamics for time $t = 0, \ldots, 40$ on ER graph sequences (dense above, sparse below) with initial parameter $\lambda = 0.3$ for vertex set sizes $n = 10, 20, 40, 80$. ($I = 10^3$.) The corresponding mean field approximation is indicated by the green line.

Rearranging terms, we obtain:

$$\begin{aligned}
\mathbb{P}(X_v(1) = 1) &= \sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!(N-k)!} \lambda^k (1-\lambda)^{N-k} \\
&= \sum_{k=0}^{N-1} \frac{(N-1)!}{k!(N-1-k)!} \lambda^{k+1} (1-\lambda)^{N-1-k} \\
&= \lambda \sum_{k=0}^{N-1} \binom{N-1}{k} \lambda^k (1-\lambda)^{(N-1)-k} \\
&= \lambda.
\end{aligned}$$

□

It is easy to show that this single-node marginal distribution corresponds exactly to the mean field limit for the voter model for i.i.d. Bernoulli(λ) initial conditions. To apply Theorem 3.1, we identify the following forms for the memory and transition probability of the voter model:

$$\begin{aligned}
R^n(t) &= \mu_1^n(t), \quad g(R^n(t), \mu^n(t+1)) = \mu_1^n(t+1), \\
K_{a,b}^n(r) &= (r)^b (1-r)^{1-b}.
\end{aligned}$$

By the strong law of large numbers, $\mu_i(0) = \lambda \mathbb{1}_{\{i=1\}} + (1-\lambda)\mathbb{1}_{\{i=0\}}, \quad \rho(0) = \lambda$ By induction, one can prove that $\mu(t) = \mu(0)$ and $\rho(t) = \lambda$ for all $t \geq 0$, so:

$$\lim_{n\to\infty} \mu^n(t) = \lambda \mathbb{1}_{\{i=1\}} + (1-\lambda)\mathbb{1}_{\{i=0\}} \text{ and } \lim_{n\to\infty} R^n(t) = \lambda.$$

In the case of random graphs, we consider the quenched marginal probability for a "typical" node in a random graph. For a random graph $G = (V, E)$, we define the marginal quenched probability to be given by

$$\mathbb{E}_{G,\rho}[\mathbb{P}(X_\rho(t) \mid G, \rho)]$$

where the randomly chosen root ρ is distributed uniformly on V. The numerical estimator for the quenched probability is computed by:

$$\frac{1}{I} \sum_{i=1}^{I} \mathbb{1}_{\{X_{\rho(i)}^{(i)}(t)=1\}}$$

Assuming that the conjecture is true, we see that the one-node marginal is invariant under the graph structure. In particular, the behavior between sparse and dense grpahs sequences does not appear to be different. Therefore, we consider the pairwise marginal of two neighboring nodes within the graph.

5.1.2 Simulations for pairwise marginals

In the following lemma, we verify that the pairwise marginal distribution for the voter model differs from the mean field approximation.

Lemma 5.4. *Let G^n be the ring graph on n vertices and X the voter model dynamics on G^n. Then the joint distribution of $X_j(1), X_{j+1}(1)$ is given by:*

$$\mathbb{P}(X_j(1) = x, X_{j+1}(1) = y) = \begin{cases} \frac{7\lambda^2}{9} + \frac{2\lambda}{9} & \text{if } (x,y) = (1,1) \\ 1 - \frac{16\lambda}{9} + \frac{7\lambda^2}{9} & \text{if } (x,y) = (0,0) \\ \frac{14\lambda}{9} - \frac{14\lambda^2}{9} & \text{if } (x,y) \in \{(0,1),(1,0)\} \\ 0 & \text{if } x, y \notin \{0,1\} \end{cases}$$

Proof.

$$\begin{aligned}
\mathbb{P}(X_j(1) = 1, X_{j+1}(1) = 1) &= \lambda^2 \mathbb{P}(X_j(1) = 1, X_{j+1}(1) = 1 \mid X_j(0) = 1, X_{j+1}(0) = 1) \\
&\quad + 2\lambda(1-\lambda)\mathbb{P}(X_j(1) = 1, X_{j+1}(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 1) \\
&\quad + (1-\lambda)^2 \mathbb{P}(X_j(1) = 1, X_{j+1}(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 0) \\
&= \lambda^2 \mathbb{P}(X_j(1) = 1 \mid X_j(0) = 1, X_{j+1}(0) = 1) \\
&\qquad \cdot \mathbb{P}(X_{j+1}(1) = 1 \mid X_j(0) = 1, X_{j+1}(0) = 1) \\
&\quad + 2\lambda(1-\lambda)\mathbb{P}(X_j(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 1) \\
&\qquad \cdot \mathbb{P}(X_{j+1}(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 1) \\
&\quad + (1-\lambda)^2 \mathbb{P}(X_j(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 0) \\
&\qquad \cdot \mathbb{P}(X_{j+1}(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 0) \\
&= \lambda^2 \mathbb{P}(X_j(1) = 1 \mid X_j(0) = 1, X_{j+1}(0) = 1)^2 \\
&\quad + 2\lambda(1-\lambda)\mathbb{P}(X_j(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 1)^2 \\
&\quad + (1-\lambda)^2 \mathbb{P}(X_j(1) = 1 \mid X_j(0) = 0, X_{j+1}(0) = 0)^2 \\
&= \lambda^2(\lambda + \tfrac{2}{3}(1-\lambda))^2 + 2\lambda(1-\lambda)(\tfrac{2}{3}\lambda + \tfrac{1}{3}(1-\lambda))^2 + (1-\lambda)^2(\tfrac{1}{3}\lambda)^2 \\
&= \frac{7\lambda^2}{9} + \frac{2\lambda}{9}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(X_j(1) = 0, X_{j+1}(1) = 0) &= \frac{7(1-\lambda)^2}{9} + \frac{2(1-\lambda)}{9} \\
&= 1 - \frac{16\lambda}{9} + \frac{7\lambda^2}{9}
\end{aligned}$$

□

From Figures 8 and 9, we see that both for deterministic and random graph sequences, the dynamics for dense sequences (the complete graph and dense ER sequences) converge to the mean field limit, while the dynamcis in sparse sequences are clearly far away

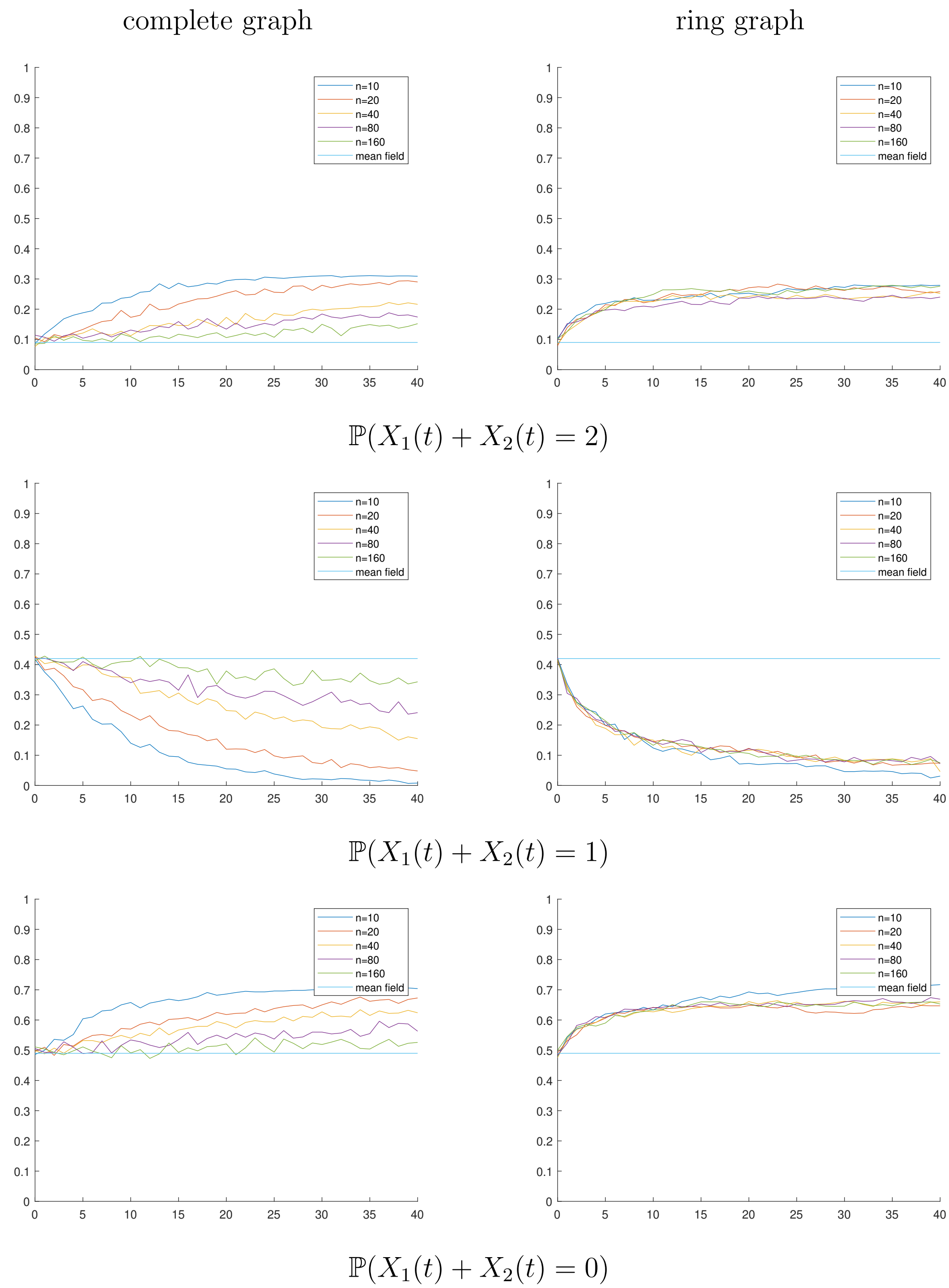
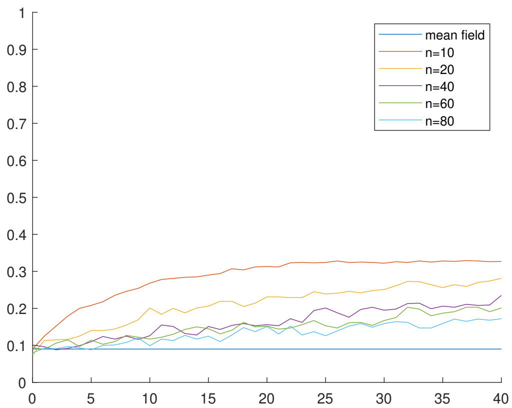
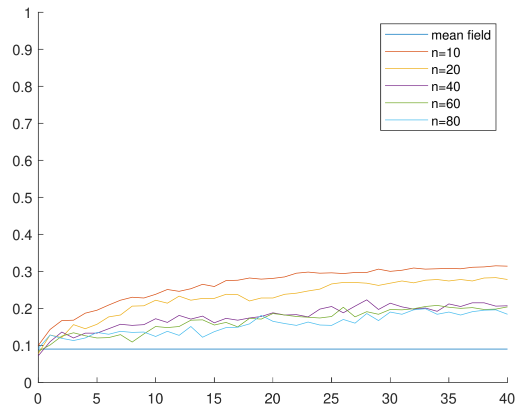


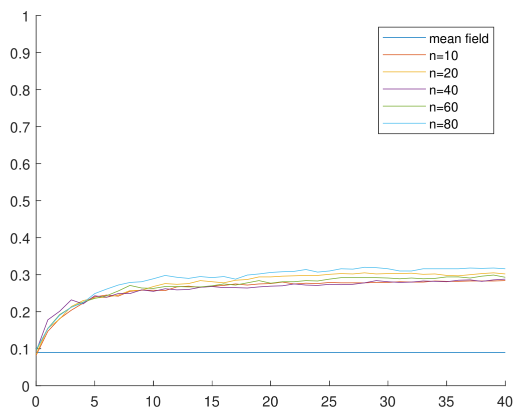
Figure 8: Empirical pairwise marginal probability (5.1) for simulated voter model dynamics for time $t = 0, \ldots, 40$ on complete and ring graphs with initial parameter $\lambda = 0.3$ for vertex set sizes $n = 10, 20, 40, 80, 160$. ($I = 10^3$.) The corresponding mean field approximation is indicated.



$p_n = 0.3$



$p_n = \log(n)/n$



$p_n = 1/n$

Figure 9: Empirical pairwise marginal probability $\mathbb{P}(X_1(t) + X_2(t) = 2)$ (5.1) for simulated voter model dynamics for time $t = 0, \ldots, 40$ on ER graph sequences (dense above, sparse below) with initial parameter $\lambda = 0.3$ for vertex set sizes $n = 10, 20, 40, 80, 160$. ($I = 10^3$.) The corresponding mean field approximation is indicated by the green line.

from the mean field limit, even for relatively small t. In subsection 5.3, we will discuss simulations for the local approximation of sparse LIMC sequences, based on the results of [2] (see Section 3.2).

5.2 Majority Dynamics

For a graph $G^n = (V_n, E_n)$ on n vertices, where $V_n = \{1, \dots, n\}$ and $E_n \subset V_n \times V_n$. Define the process $X(t) = \{X_j(t)\}_{j \in V_n}$ for every $t \in \mathbb{N}$ by the transition

$$X_j(t+1) = \begin{cases} 1 & \text{if } \frac{1}{|\overline{\partial j}|} \sum_{i \in \overline{\partial j}} X_i(t) > \frac{1}{2} \\ 0 & \text{if } \frac{1}{|\overline{\partial j}|} \sum_{i \in \overline{\partial j}} X_i(t) < \frac{1}{2} \\ \xi_t & \text{if } \frac{1}{|\overline{\partial j}|} \sum_{i \in \overline{\partial j}} X_i(t) = \frac{1}{2} \end{cases}$$

where $\{\xi_t\}_{t \in \mathbb{N}_0}$ are i.i.d. with Bernoulli$\left(\frac{1}{2}\right)$ distribution. and initial distribution

$$X_j(0) = \begin{cases} 1 & \text{w.p. } \lambda \\ 0 & \text{w.p. } 1 - \lambda \end{cases}$$

for some fixed parameter $\lambda \in [0, 1]$.

The majority dynamics are of interest in comparison to the voter model, since the marginal dynamics of a single node depend on the graph structure, as demonstrated by the simulation results in Figures 10, 11, and 12 for complete and ring graphs, dense and sparse ER random graphs, and d-regular trees, respectively.

Computations for $\mathbb{P}(X_j(1) = 1)$

For a ring graph on n vertices:

$$\mathbb{P}(X_j(1) = 1) = p^3 + 3p^2(1-p) = -2p^3 + 3p^2$$

For a complete graph n vertices:

$$\mathbb{P}(X_j(1) = 1) = \begin{cases} \frac{1}{2}\binom{n}{\frac{n}{2}} p^{\frac{n}{2}}(1-p)^{\frac{n}{2}} + \sum_{k=\frac{n}{2}+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } n \text{ even} \\ \sum_{k=\frac{n-1}{2}+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } n \text{ odd} \end{cases}$$

	$\mathbb{P}(\cdot)$	MF
$X_1(1) + X_2(1) = 2$	0.14	0.09
$X_1(1) + X_2(1) = 1$	0.54	0.42
$X_1(1) + X_2(1) = 0$	0.33	0.49

Table 1: Simulated values representing Lemma 5.4 as compared with the corresponding mean field approximation of each probability. $\lambda = 0.3$, $I = 10^3$

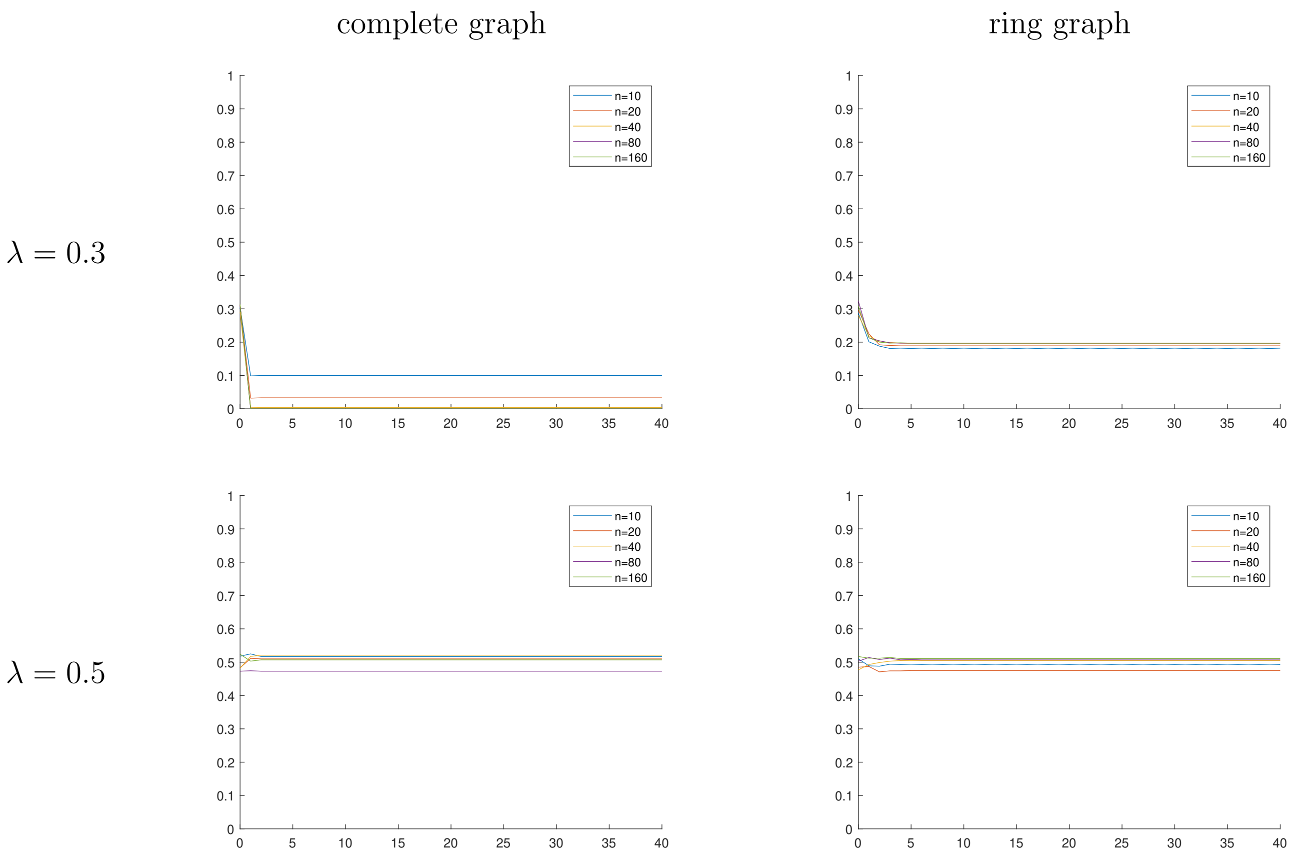
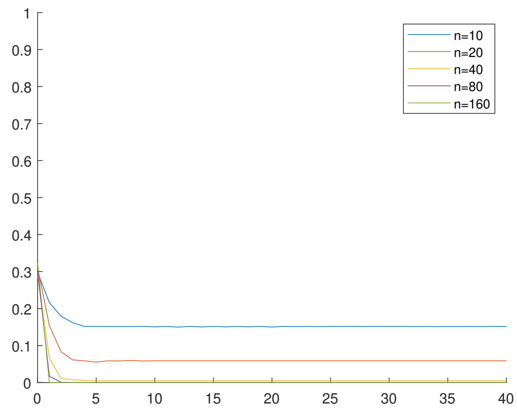
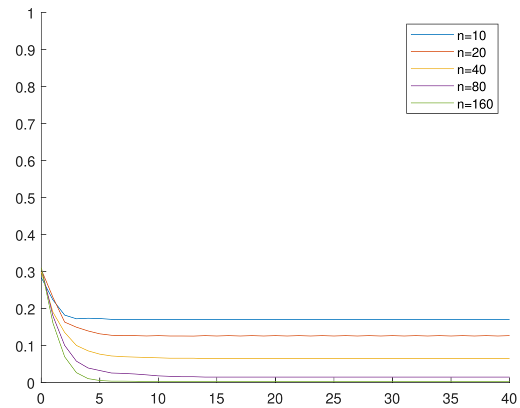


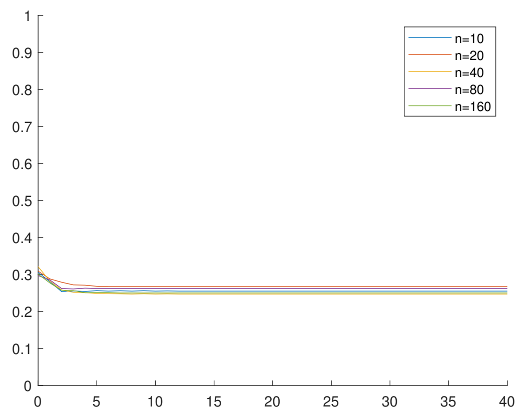
Figure 10: Empirical marginal probability $\mathbb{P}(X_1(t) = 1)$ (5.1) for simulated majority dynamics for time $t = 0, \dots, 40$ on complete and ring graph sequences with initial parameters $\lambda = 0.3, 0.5$ for vertex set sizes $n = 10, 20, 40, 80, 160$. ($I = 10^3$.)



$p_n = 0.3$



$p_n = \log(n)/n$



$p_n = 1/n$

Figure 11: Empirical marginal probability $\mathbb{P}(X_1(t) = 1)$ (5.1) for simulated majority dynamics for time $t = 0, \ldots, 40$ on ER graph sequences (dense above, sparse below) with initial parameter $\lambda = 0.3$ for vertex set sizes $n = 10, 20, 40, 80, 160$. ($I = 10^3$.)

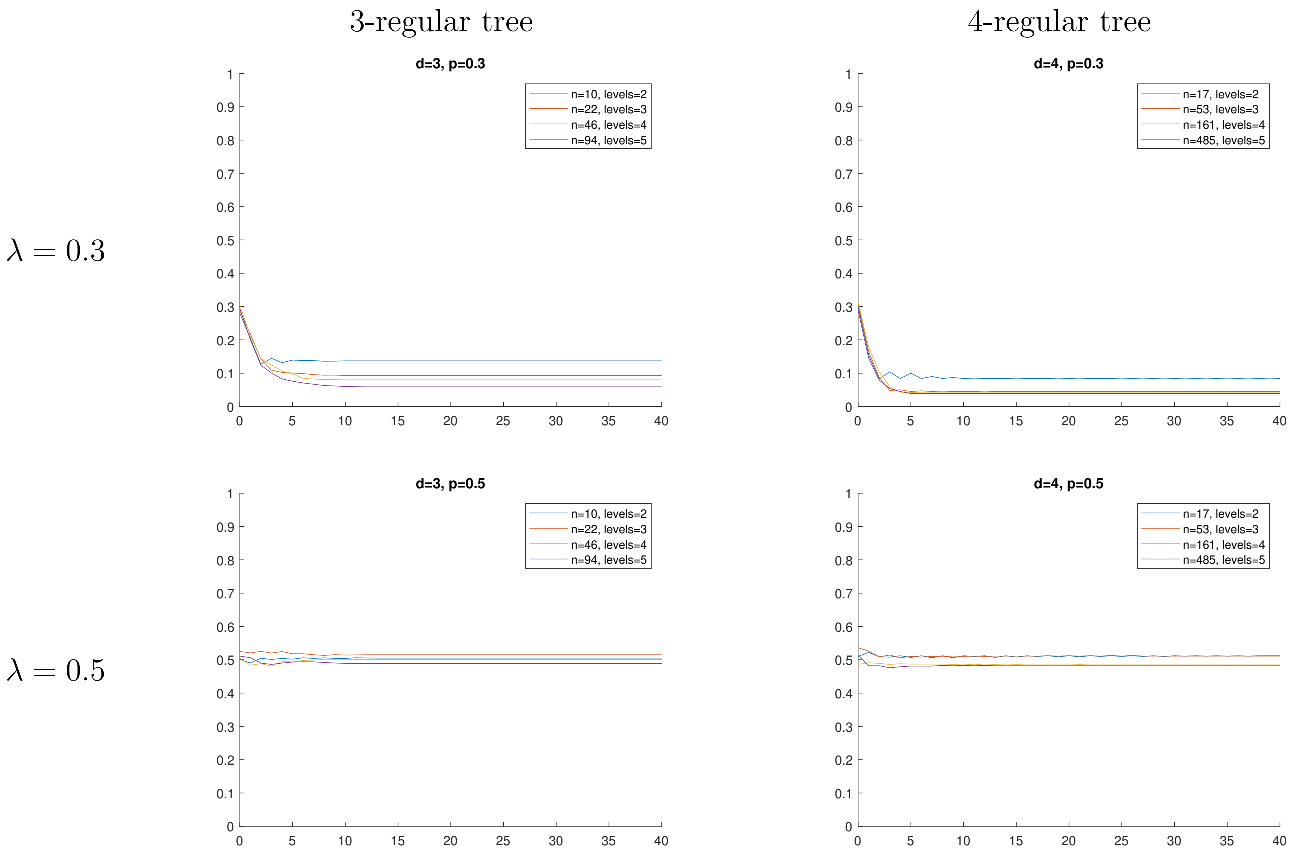


Figure 12: Empirical marginal probability $\mathbb{P}(X_1(t) = 1)$ (5.1) for simulated majority dynamics for time $t = 0, \ldots, 40$ on 3- and 4-regular tree sequences with initial parameters $\lambda = 0.3, 0.5$ for total tree levels $2, 3, 4, 5$. ($I = 10^3$.)

More generally, for a vertex j with $|\overline{\partial j}| = N$:

$$P_j^N := \mathbb{P}(X_j(1)=1) = \begin{cases} \frac{1}{2}\binom{N}{\frac{N}{2}} p^{\frac{N}{2}}(1-p)^{\frac{N}{2}} + \sum_{k=\frac{N}{2}+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} & \text{if } N \text{ even} \\ \sum_{k=\frac{N-1}{2}+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} & \text{if } N \text{ odd} \end{cases}$$

Now let $G^n = ER(n, p_n)$ for some fixed n. Then:

$$\begin{aligned} \mathbb{E}_{G^n}[\mathbb{P}(X_j(1)=1)] &= \mathbb{E}_N\left[\sum_{k=0}^{N} \mathbb{P}\left(X_j(1)=1 \,\middle|\, \sum_{i\in\overline{\partial j}} X_i(0)=k\right)\mathbb{P}\left(\sum_{i\in\overline{\partial j}} X_i(0)=k\right)\right] \\ &= \sum_{N=1}^{n} P_j^N \binom{n}{N} p_n^N (1-p)^{N-n} \end{aligned}$$

Remark. When $p = \frac{1}{2}$ and $|\overline{\partial j}| = N$,

$$P_j^N = \mathbb{P}(X_j(1)=1) = \begin{cases} \frac{1}{2^N}\left(\frac{1}{2}\binom{N}{\frac{N}{2}} + \sum_{k=\frac{N}{2}+1}^{N}\binom{N}{k}\right) & \text{if } N \text{ even} \\ \frac{1}{2^N}\left(\sum_{k=\frac{N-1}{2}+1}^{N}\binom{N}{k}\right) & \text{if } N \text{ odd} \end{cases}$$

$$= \begin{cases} \frac{1}{2^N}\cdot 2^{N-1} & \text{if } N \text{ even} \\ \frac{1}{2^N}\cdot 2^{N-1} & \text{if } N \text{ odd} \end{cases} = \frac{1}{2}$$

5.3 Local Approximation for Sparse Graphs

5.3.1 Motivation

From the simulations above, it is clear that even after a short time, the mean field approximation performs well for dense graph sequences, but performs poorly for sparse graph sequences, for those dynamics which exhibit different effects for marginal dynamics depending on the graph structure. We can quantify these observations succinctly by summarizing the total variation distance between the marginal distributions for simulated full dynamics and the corresponding mean field approximation. Recall that in general the total variation distance between two probability distributions P, Q on a discrete state space $\mathcal{X}$ is given by

$$\delta(P,Q) := \sum_{x\in\mathcal{X}} |P(x) - Q(x)|.$$

In some cases, the total variation distance is normalized by a factor of $\frac{1}{2}$ to take values in [0,1].

graph	n	$t=0$	$t=10$	$t=40$	
complete	10	0.0043	0.2732	0.4135	$\delta(P^t_G, P^t_{\text{MF}})$
	160	0.0052	0.0259	0.0900	
$ER(n, 0.4)$	10	0.0130	0.2750	0.4150	
	160	0.0220	0.0420	0.0900	
$ER\Big(n, \frac{\log(n)}{n}\Big)$	10	0.0230	0.2990	0.4140	
	160	0.0220	0.0700	0.1250	
$ER(n, \frac{1}{n})$	10	0.0380	0.3870	0.4190	
	160	0.0320	0.3500	0.3980	

Table 2: Total variation between the empirical neighbor marginal distribution P^t_G (averaged over 10^3 trials) for the voter model on dense graphs and the corresponding mean field approximation P^t_{MF}.

		local approximation $\delta(P^t_{R_n}, P^t_{\text{loc}})$			mean field $\delta(P^t_{R_n}, P^t_{\text{MF}})$		
graph	n	$t=0$	$t=4$	$t=8$	$t=0$	$t=4$	$t=8$
ring	10	0.0069	0.0070	0.0054	0.0085	0.1979	0.2573
	160	0.0117	0.0101	0.0044	0.0092	0.2062	0.2507

Table 3: Total variation between the empirical neighbor marginal distribution P^t_G (averaged over 10^3 trials) for the voter model on and the corresponding local approximation P^t_{loc} on the sequence of ring graphs on n vertices, $\{R_n\}$.

In Table 2, we give values for the total variation distance between the numerically simulated empirical distribution of a pair of neighboring particles and the corresponding mean field approximation for a variety of graph sequences. For small n, the empirical distribution differs from the mean field for all graph sequences. For larger n, the empirical distributions associated with dense sequences, like complete graphs, $ER(n, 0.4)$, and $ER\Big(n, \frac{\log(n)}{n}\Big)$, is close to the corresponding mean field approximation; however, the distance is large for sparse sequences, like $ER(n, \frac{1}{n})$.

By comparison, in Table 3, we show the corresponding comparison of total variation distance between the empirical neighbor marginal distribution and the local approximation developed by Lacker, Ramanan, and Wu [2] (see Section also 3.2) on ring graphs. In this case, as in the case of $ER(n, \frac{1}{n})$, the mean field performs poorly for large n, while the local approximation closely matches the dynamics of the finite system. One can see heuristically in Figure 13 that the local approximation matches the dynamics on the ring closely even for small t and n.

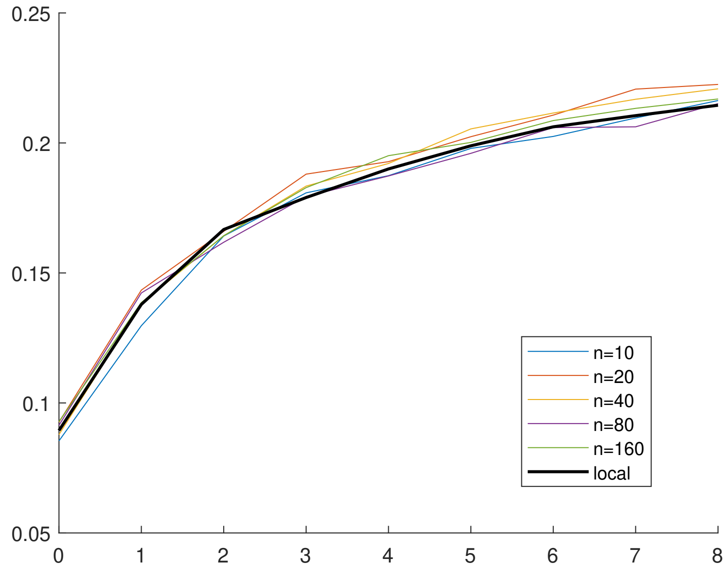


Figure 13: Empirical marginal probability $\mathbb{P}(X_1(t) = 1, X_2(t) = 1)$ (5.1) for simulated voter model for time $t = 0, \ldots, 8$ on the ring graph sequence with initial parameters $\lambda = 0.3$ compared with the corresponding local approximation. ($I = 10^3$.)

5.3.2 Local Equations on Unimodular Galton Watson Trees

In [2], Lacker, Ramanan, and Wu develop a local recursion for the marginal dynamics of UGW trees. (Recall Definition 2.12.) To simulate the local dynamics of a LIMC on a UGW tree, we compute explicitly the local equations for the law of the root node and its (random) neighborhood of nodes in the first generation of the tree. More formally, we wish to characterize the distribution of the root node and its neighborhood of LIMC dynamics on a UGW tree with infinite number of levels.

For any time $t \in \mathbb{N}_0$, we compute the joint distribution of $(N; X_{\varnothing}[t]; X_{1,\ldots N}[t])$, where N is the size of the root neighborhood ($N \sim \hat{\rho}$), and $X_{1,\ldots N}$ are the dynamics of the children of the root node, whose index is denoted $\varnothing$.

See the appendix for the computations of the law as well as pseudocode for recursively obtaining the distribution of the root node and its neighborhood numerically. Here we discuss a few of the main questions arising from simulating the local dynamics of LIMC on UGW trees. The key difficulty is that to compute the joint distribution even for small t requires storing a joint probability distribution over a large state space, which becomes intractable to compute both in terms of memory and computation time. Since the local dynamics are non-Markovian, in principle, the value of the particles at a time t in principle depends on the entire history of the local system up to that time. To mitigate the memory concerns, one could cut off the history after some finite κ number of steps, an approximation which has been shown numerically to be relatively accurate for other dynamics on deterministic graphs [9], [10]. However, there remain many compelling questions regarding numerical computation of the local equations for UGW trees which could be investigated further.

6 Discussion and Further Questions

In this thesis, we introduced the ring cluster model and showed that by applying and extending techniques for dense and sparse LIMC sequences, one can characterize the limit of the ring cluster model and compute its marginal dynamics. In particular, since the model is heterogeneous, both the prelimit and limit systems have nodes of different types: nodes on the ring fall under the sparse regime, while nodes within the cluster fall under the dense regime, with some nodes sitting at the interface of the sparse ring and the dense cluster, and they all interact with each other. We showed in Section 4 that nodes in the dense subset of the graph converged to (Markovian) mean field limits, while nodes in the interface and ring subsets converge locally to a limiting system whose marginal dynamics were equal in law to a certain recursively constructed non-Markovian discrete-time stochastic process.

As a first step, the setting of the ring cluster model is useful in showing how the techniques for dense and sparse LIMC sequences may be combined. In further work, we would like to extend our understanding to a broader class of heterogeneous graph sequences, deterministic and random. Using a similar approach to ours, one could generalize the model to a greater variety of sparse structures for which local limits are known (e.g. d-regular trees, Galton-Watson trees) and dense clusters (e.g. dense Erdős-Rényi random graph sequences). Ultimately, we would like to further our theory of limits and understand random systems with less structured heterogeneity, such as the stochastic block model or generalized random graph.

Furthermore, we aim to use the characterization of the limit system and marginal dynamics to study properties of heterogeneous systems. Applications in neuroscience are of special interest, since cortical tissue in the brain has been shown to have dense clusters of neurons between which there are only sparse neural connections [7]. The discrete Kuramoto model is a particular model of coupled oscillators which exhibits synchronization patterns that mirror oscillatory phenomena in the cortex [8]. Our results concerning the marginal dynamics of a within large collection of particles could give insight on various phenomena within heterogeneous neural structures.

References

[1] Thomas Milton Liggett. *Interacting particle systems*. Vol. 276. Springer Science & Business Media, 2012.

[2] Daniel Lacker, Kavita Ramanan, and Ruoyu Wu. *Interacting Markov Chains on Graphs*. Working paper. 2020.

[3] Daniel Lacker, Kavita Ramanan, and Ruoyu Wu. *Large sparse networks of interacting diffusions*. 2019. arXiv: 1904.02585 [math.PR].

[4] Henry McKean. "A class of Markov processes associated with nonlinear parabolic equations". In: *Proceedings of the National Academy of Sciences of the United States of America* 56 (Jan. 1967), pp. 1907–11. DOI: 10.1073/pnas.56.6.1907.

[5] Jean-Yves Le Boudec, David McDonald, and Jochen Mundinger. "A Generic Mean Field Convergence Result for Systems of Interacting Objects". In: *Proceedings of the Fourth International Conference on Quantitative Evaluation of Systems*. QEST '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 3–18. ISBN: 0-7695-2883-X. DOI: 10.1109/QEST.2007.3. URL: https://doi.org/10.1109/QEST.2007.3.

[6] Daniel Lacker, Kavita Ramanan, and Ruoyu Wu. *Local Weak convergence and propagation of ergodicity for large networks of interacting processes.* Preprint. 2020.

[7] Vladimir V Klinshov et al. "Dense neuron clustering explains connectivity statistics in cortical microcircuits". In: *PloS one* 9.4 (2014).

[8] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. "Generative models of cortical oscillations: neurobiological implications of the Kuramoto model". In: *Frontiers in human neuroscience* 4 (2010), p. 190.

[9] Mitchell Wortsman. *Interacting Particles Systems and Effcient Approximations for Large Sparse Graphs*. Honors thesis. 2019.

[10] Timothy Sudijono. *Stationarity and Ergodicity of Local Dynamics of Interacting Markov Chains on Large Sparse Graphs*. Honors thesis. 2019.

[11] Leticia F Cugliandolo. *Advanced Statistical Physics: Quenched disordered systems.* 2017.

[12] Richard Durrett. *Random graph dynamics*. Vol. 200. 7. Cambridge University Press Cambridge, 2007.

Appendix: Local Equations on Unimodular Galton-Watson Trees

Here we provide the explicit computation for the local equations on UGW, as developed in [2].

$$\gamma_k(m, z_{2,\dots,m}(k) \mid x_\varnothing[k], x_1[k]) = \frac{\mathbb{E}\Big[\frac{m}{N_1}\mathbb{1}_{\{N_0=m, \overline{X}_{2,\dots,m}(k)=z_{2,\dots,m}(k)\}} \,\Big|\, \overline{X}_{\varnothing,1}[k] = x_{\varnothing,1}[k], \mathbb{1}_{\{N_0\geq 1\}} = a\Big]}{\mathbb{E}\Big[\frac{N_0}{N_1} \,\Big|\, \overline{X}_{\varnothing,1}[k] = x_{\varnothing,1}[k], N_0 \geq 1\Big]}$$

$$\text{numerator } = m\sum_{n_1=1}^{\infty}\frac{1}{n_1}\cdot\mathbb{P}\big(\overline{X}_{2,\dots,m}(k) = z_{2,\dots,m}(k), N_0 = m, N_1 = n_1 \mid \overline{X}_{\varnothing,1}[k] = x_{\varnothing,1}[k], \mathbb{1}_{\{N_0\geq 1\}} = a\big)$$

$$\text{denominator } = \sum_{n_0,n_1=1}^{\infty}\frac{n_0}{n_1}\cdot\mathbb{P}\big(N_0 = n_0, N_1 = n_1 \mid \overline{X}_{\varnothing,1}[k] = x_{\varnothing,1}[k], N_0 \geq 1\big)$$

$$\begin{aligned}
J_t(n_0, n_1, x_{0,1,\dots,n_0}[t]) &= \mathbb{P}(N_0 = n_0, N_1 = n_1, X_{0,1,\dots,n_0}[t] = x_{0,1,\dots,n_0}[t]) \\
\alpha_t(z_{2,\dots,m}(t), m, n_1 \mid x_{\varnothing,1}[t]) &:= \mathbb{P}\big(\overline{X}_{2,\dots,m}(t) = z_{2,\dots,m}(t), N_0 = m, N_1 = n_1 \mid \overline{X}_{\varnothing,1}[t] = x_{\varnothing,1}[t]\big) \\
\beta_t(n_0, n_1 \mid x_{\varnothing,1}[t]) &:= \mathbb{P}\big(N_0 = n_0, N_1 = n_1 \mid \overline{X}_{\varnothing,1}[t] = x_{\varnothing,1}[t], N_0 \geq 1\big) \\
&= \sum_{\{y_{\{2,\dots,n_0\}}(t)\}} \alpha_t(y_{2,\dots,n_0}(t), n_0, n_1 \mid x_{\varnothing,1}[t]) \\
\text{numerator } &= m\sum_{n_1=1}^{\infty}\frac{1}{n_1}\alpha_t(z_{2,\dots,m}(t), m, n_1 \mid x_{\varnothing,1}[t]) \\
\text{denominator } &= \sum_{n_0,n_1=1}^{\infty}\frac{n_0}{n_1}\cdot\beta_t(n_0, n_1 \mid x_{\varnothing,1}[t])
\end{aligned}$$

Suppose that for some time $t \geq 1$ we know $J_t(n_0, n_1, x_{0,1,\dots,n_0}[t])$ for all $n_0, n_1 \in \mathbb{N}$ and $x_{0,1,\dots,n_0}[t] \in \mathcal{S}^{(t+1)(n_0+1)}$.

We wish to compute J_{t+1}. In fact,

$$\begin{aligned}
&J_{t+1}(n_0, n_1, x_{0,1,\dots,n_0}[t+1]) \\
&\quad = \overbrace{\mathbb{P}(X_{0,1,\dots,n_0}(t+1) = x_{0,1,\dots,n_0}(t+1) \mid N_0 = n_0, N_1 = n_1, X_{0,1,\dots,n_0}[t] = x_{0,1,\dots,n_0}[t])}^{(*)} \\
&\qquad\cdot J_t(n_0, n_1, x_{0,1,\dots,n_0}[t])
\end{aligned}$$

We can further rewrite $(*)$ as follows:

$$(*) = \frac{\mathbb{P}(N_1 = n_1, X_{0,1,\dots,n_0}(t+1) = x_{0,1,\dots,n_0}(t+1) \mid N_0 = n_0, X_{0,1,\dots,n_0}[t] = x_{0,1,\dots,n_0}[t])}{\mathbb{P}(N_1 = n_1 \mid N_0 = n_0, X_{0,1,\dots,n_0}[t] = x_{0,1,\dots,n_0}[t])}$$

The denominator of this expression can be computed from J_t:

$$\mathbb{P}(N_1 = n_1 \mid N_0 = n_0, X_{0,1,\dots,n_0}[t] = x_{0,1,\dots,n_0}[t]) = \frac{J_t(n_0, n_1, x_{0,1,\dots,n_0}[t])}{\sum_{n=1}^{\infty} J_t(n_0, n, x_{0,1,\dots,n_0}[t])}$$

While the numerator can be rewritten in terms of the transition probability function and γ_t:

$$\begin{aligned}\text{numerator of } (*) = \sum_{n_2,\dots,n_{n_0}=1}^{\infty} \sum_{\substack{\{y_{j1,\dots,jc_j}\}\\ j=2,\dots,n_0}} & P(x_0(t+1)|n_0; x_{0,1,\dots,n_0}(t)) \\ & \cdot \prod_{j=1}^{n_0} (P(x_j(t+1)|n_j; x_{0,j}(t), y_{j1,\dots,jc_j}(t)) \\ & \qquad \cdot \gamma_t(n_j, y_{j1,\dots,jc_j}(t)|x_j[t], x_0[t]))\end{aligned}$$

Now that we have computed J_{t+1}, we marginalize and condition to compute α_{t+1} and β_{t+1}.

Define:

$$\begin{aligned} & a_{t+1}(z_{2,\dots,m}(t+1), m, n_1, x_{\varnothing,1}[t+1]) \\ & \quad := \mathbb{P}\big(\overline{X}_{2,\dots,m}(t+1) = z_{2,\dots,m}(t+1), N_0 = m, N_1 = n_1, \overline{X}_{\varnothing,1}[t+1] = x_{\varnothing,1}[t+1]\big) \\ & \quad = \sum_{\{x_{2,\dots,m}[t]\}} J_{t+1}(n_0, n_1, x_{0,1,\dots,n_0}[t+1]) \end{aligned}$$

Then:

$$\alpha_{t+1}(x_{2,\dots,m}(t+1), m, n_1 \mid x_{\varnothing,1}[t+1]) = \frac{a_{t+1}(x_{2,\dots,m}(t+1), m, n_1, x_{\varnothing,1}[t+1])}{\sum_{\{y_{2,\dots,m}(t+1), N_0, N_1\}} a_{t+1}(y_{2,\dots,m}(t+1), N_0, N_1, x_{\varnothing,1}[t+1])}$$

and

$$\beta_{t+1}(m, n_1 \mid x_{\varnothing,1}[t+1]) = \sum_{\{y_{2,\dots,m}(t+1)\}} \alpha_{t+1}(y_{2,\dots,m}(t+1), m, n_1 \mid x_{\varnothing,1}[t+1])$$

$$\begin{aligned}\gamma_k(m, z_{1,\dots,m}[k] \mid x_{\varnothing,1}[k]) &= \text{Law}\,(N_0, (X_{1,\dots,N_0}) \mid X_{\varnothing,1}[k] = x_{\varnothing,1}[k])\Big|_{N_0=m, X_{1,\dots,N_0}[k]=z_{1,\dots,m}[k]} \\ \alpha_k(x_\varnothing[k], z_{1,\dots,m}[k] \mid m, n_1) &= \mathbb{P}\big(\overline{X}_{1,\dots,m}[k] = z_{1,\dots,m}[k], \overline{X}_\varnothing[k] = x_\varnothing[k] \mid N_0 = m, N_1 = n_1, N_0 \geq 1\big) \\ \beta_k(x_{\varnothing,1}[k] \mid n_0, n_1) &= \mathbb{P}\big(\overline{X}_{\varnothing,1}[k] = x_{\varnothing,1}[k] \mid N_0 = n_0, N_1 = n_1, N_0 \geq 1\big)\end{aligned}$$

Algorithm 2 Simulate dynamics on UGW

$n_0 \sim \text{Poisson}(c)$
$\hat{X}_{\varnothing,1,\ldots,n0}(0) \sim \lambda^{n_0+1}$
for $t = 1, \ldots, T$ **do**
  $\hat{X}_{\varnothing}(t+1) \leftarrow \textsc{transition}\Big(\hat{X}_{\varnothing}[t], n_0, \hat{X}_{\varnothing,1,\ldots,n0}[t]\Big)$
  for $j = 1, \ldots, n_0$ **do**
    $(n_j, x_{j1,\ldots,jn_j}[t]) \sim \gamma_t(\,\cdot\,,\,\cdot\,\mid \hat{X}_j[t], \hat{X}_{\varnothing}[t])$
    $\hat{X}_j(t+1) \leftarrow \textsc{transition}\Big(\hat{X}_j[t], n_j, x_{j1,\ldots,jn_j}[t]\Big)$
  end for
end for
function TRANSITION(parent, N, children)
  $x \sim$ Bernoulli((parent + children)/(N+1))
  return x
end function

Algorithm 1 Calculate γ

$c \leftarrow$ Poisson children distribution parameter
$M \leftarrow$ neighbor maximum
$T \leftarrow$ time maximum
$\lambda \leftarrow$ initial distribution of node
$f([\,\cdot\,]) \leftarrow$ transition probability function
$P \leftarrow \mathbb{P}(N_0 = \,\cdot\,, N_1 = \,\cdot\, \mid N_0 > 0)$
for $m = 0, \ldots, M$ **do**
  $\gamma_0(m, [\,\cdot\,] \mid \cdot\,) = \mathbb{P}(N_0 = m) \times \lambda^m$
  $\alpha_0([\,\cdot\,], \cdot \mid m, \cdot\,) = \lambda^{m+1}$
  $\beta_0(\,\cdot\,, \cdot \mid \cdot\,, \cdot\,) = \lambda^2$
end for
for $t = 1, \ldots, T$ **do**
  $\alpha_{t+1} = \textsc{compute alpha}(\alpha_t, \gamma_t)$
  $\beta_{t+1} = \sum_{*} \alpha_{t+1}([\,\cdot\,, *], \cdot \mid \cdot\,, \cdot\,)$
  for $m = 0, \ldots, M$ **do**

$$\gamma_{t+1}(m, [\,\cdot\,] \mid \cdot\,) = \sum_{n_1=1}^{M} \left(P(m, n_1) \cdot \alpha_{t+1}([\,\cdot\,], \cdot \mid m, n_1) \cdot \left(\sum_{n_0=1}^{M} \frac{1}{n_0} P(n_0, n_1) \cdot \beta_{t+1}(\,\cdot\,, \cdot \mid n_0, n_1) \right)^{-1} \right)$$

  end for
end for

function $\textsc{compute alpha}(\alpha_t, \gamma_t)$
  for $m = 0, \ldots, M$ **do**
    for all $x \in \mathcal{S}^{(t+2)(m+1)}$, **do**

$$\begin{aligned} \alpha_{t+1}(x \mid m, \cdot\,) &= f(x_{\varnothing,1,\ldots,m}[t+1]) \cdot \alpha_t(x \mid m, \cdot\,) \\ &\cdot \left(\sum_{*} f(x_1[t+1], *) \cdot \gamma_t(n_1, * \mid x_{\varnothing,1}[t+1]) \cdot (\mathbb{P}(N_1 = n_1 \mid N_0 \geq 0))^{-1} \right) \\ &\cdot \prod_{j=2}^{m} \left(* * * * * + \sum_{n_j=2}^{M} \left(\sum_{*} f(x_1[t+1], *) \cdot \gamma_t(n_j - 1, * \mid x_{\varnothing,1}[t+1]) \right) \right) \end{aligned}$$

    end for
  end for
  return α_{t+1}
end function