

OPINION DYNAMICS ON NON-SPARSE NETWORKS WITH COMMUNITY STRUCTURE

PANAGIOTIS ANDREOU AND MARIANA OLVERA-CRAVIOTO

ABSTRACT. We study the evolution of opinions on a directed network with community structure. Individuals update their opinions synchronously based on a weighted average of their neighbors' opinions, their own previous opinions, and external media signals. Our model is akin to the popular Friedkin-Johnsen model, and is able to incorporate factors such as stubbornness, confirmation bias, selective exposure, and multiple topics, which are believed to play an important role in the formation of opinions. Our main result shows that, in the large graph limit, the opinion process concentrates around its mean-field approximation for any level of edge density, provided the average degree grows to infinity. Moreover, we show that the opinion process exhibits propagation of chaos. We also give results for the trajectories of individual vertices and the stationary version of the opinion process, and prove that the limits in time and in the size of the network commute. The mean-field approximation is explicit and can be used to quantify consensus and polarization.

1. INTRODUCTION

As the consumption of information through virtual social networks has grown, the number of news outlets has multiplied, and individuals increasingly find ways to disseminate their opinions through online forums and video sharing platforms, populations around the world have become heavily polarized on a number of cultural and political issues [48, 9, 44, 46]. The asymmetry of the information from one source to another has solidified positions into ideological bubbles with little in common with each other. The formation of such bubbles occurs through a combination of cognitive biases that allow individuals to choose to listen only to opinions that conform to their preconceptions, as well as reject people who espouse views contrary to their own.

Our collective need to understand the mechanisms behind polarization and propose ideas to reduce its negative effects has inspired much of the current literature in the social sciences [48, 9, 44, 46, 1, 27, 29]. Simultaneously, researchers in the mathematical and physical sciences have proposed models that can capture and explain phenomena such as consensus and polarization, often focusing on one or two modeling features at a time. In this work, we analyze a model based on the classical DeGroot [19] and Friedkin-Johnsen [25] models, that can simultaneously capture features such as:

- a.) *Confirmation bias*: the tendency to search for, interpret, favor, and recall information in a way that confirms or supports one's prior beliefs or values;
- b.) *Selective exposure*: individuals' tendency to favor information which reinforces their pre-existing views while avoiding contradictory information;

1991 *Mathematics Subject Classification.* 60G10, 05C80, 60G17, 60J20, 60J85, 60K35.

Key words and phrases. opinion dynamics, inhomogeneous random graphs, interacting particle systems, mean-field approximation, propagation of chaos, strong couplings.

- c.) *Community structure*: when a network divides naturally into groups of nodes with dense connections internally and sparser connections between groups;
- d.) *Bots*: software applications that run automated tasks over the Internet on a large scale, usually with the intent of imitating human activity;
- e.) *Multiple topics*: the analysis of the evolution of opinions on multiple (possibly correlated) topics at a time.

A single-topic version of this model was recently analyzed in [22] on general complex networks without community structure. This work focused on sparse (i.e., bounded degree) graphs, where the graph's local behavior determines the evolution of the opinion process. Its main results show the power that the media has in framing people's opinions, and how selective exposure alone can explain much of the polarization that we observe in many societies today. One of the main features of the model in [22], as well as of the model studied here, is that, due to its linearity, it produces computable formulas that can be used to quantify the effect of various model parameters on the evolution of opinions, and potentially, of any intervention one may want to propose.

The current work studies a more general opinion model than the one analyzed in [22] on sparse graphs, but rather than focus on its explanatory/predictive power, it computes its limiting behavior on denser graphs with community structure. From a modeling point of view, it is often hard to decide whether a real-world network is sparse with a large average degree, or, say, semi-sparse, with an average degree that grows slowly with the size of the graph. Given the lack of precise boundaries, we believe it is interesting to understand the difference between sparse approximations and dense, or mean-field, approximations. At a high level, sparse approximations are determined by the entire local neighborhood of each vertex, while mean-field approximations are obtained by replacing the individual contribution of each neighbor with their average behavior, which is justified only when vertices in the graph have a sufficiently large number of neighbors.

Our main interest in this paper is to show that there is a single mean-field approximation for our opinion model that works for graphs with any density level other than sparse. However, the precision level of this approximation does exhibit a phase transition when the average degree crosses the familiar connectivity threshold $\log n$ (with n the number of vertices in the graph), and seamlessly explains how the model's behavior transitions into the sparse regime, where local approximations such as those found in [22] take over. In fact, our analysis of the semi-sparse regime is based on the same local approximation that is used for sparse graphs, which when the average degree grows with n further simplifies to the mean-field approximation obtained in the dense case. The transition into the sparse regime can also be seen from the point of view of the trajectories for different individuals, which in the setting of this paper exhibit propagation of chaos (i.e., asymptotically independent trajectories). In contrast, on sparse graphs, only a weaker form of propagation of chaos holds, which states that only finite sets of uniformly chosen trajectories are asymptotically independent. The in-depth analysis of the model's explanatory and predictive power will be covered in a separate set of work.

This paper is organized as follows. In Section 2 we introduce our model for the evolution of opinions on a large complex network with community structure, as well as the random graph family used in our analysis. In Section 3 we present our main results, which give an approximation for the stationary opinion process that is valid for any density level of the underlying graph other than sparse. In Section 4 we discuss related models for opinion dynamics, as well as the technical aspects of our work, such as mean-field

approximations and local approximations for random graphs. Finally, in Section 5 we give all the proofs for the theorems in Section 3.

2. THE MODEL

Our opinion model consists of two parts: 1) the opinion process defined on a fixed directed graph, and 2) the random graph model used to represent the underlying social network.

2.1. The opinion process on a fixed graph. Consider a directed graph $G = (V, E)$, where individuals in the social network are represented by vertices in the set V , and a directed edge $(i, j) \in E$ means that individual j listens to the opinion of individual i ; $|V|$ denotes the cardinality of the set. Each individual in the set V is assumed to belong to one of K communities, and has a set of attributes that will remain fixed throughout the evolution of the opinion process. We assume we have ℓ topics, with ℓ a finite number, and individuals' opinions on each topic take values on the interval $[-1, 1]$. The main object of study is the opinion process $\{R^{(k)} : k \geq 0\}$, taking values on $[-1, 1]^{|V| \times \ell}$, where $R_{ij}^{(k)}$ denotes the expressed opinion of individual $i \in V$ on topic $j \in \{1, \dots, \ell\}$ at time $k \in \mathbb{N} = \{0, 1, 2, \dots\}$. We will use $\mathbf{R}_i^{(k)}$ to denote the i th row of $R^{(k)}$, which corresponds to the multi-topic opinion vector of individual i at time k .

Before we describe the updating rule for the process $\{R^{(k)} : k \geq 0\}$, we need to describe the vertex attributes that determine how each individual engages with its neighbors and incorporates their own prejudices. To start, each individual $i \in V$ has an attribute vector and an extended attribute vector of the form:

$$\mathbf{a}_i = (J_i, \mathbf{q}_i) \in \{1, \dots, K\} \times [-1, 1]^\ell \quad \text{and} \quad \mathbf{y}_i = (J_i, \mathbf{q}_i, \mathbf{b}_i) \in \{1, \dots, K\} \times [-1, 1]^\ell \times \mathbb{R}_+^{|V|},$$

where J_i denotes the community to which individual i belongs, $\mathbf{q}_i = (q_{i1}, \dots, q_{i\ell})$ is a row vector of pre-conceived ideas or internal beliefs on each of the ℓ topics, and $\mathbf{b}_i = (b_{i1}, \dots, b_{i|V|})$ is a nonnegative row vector of ‘‘unnormalized’’ weights, where b_{ij} is the weight that individual i gives to individual j 's expressed opinions (these weights will later be normalized to add up to one). The extended attribute vectors $\{\mathbf{y}_i : i \in V\}$ do not change throughout the evolution of the process, reflecting the reasonable modeling assumption that they change at a much longer time-scale than the expressed opinions process. To make them easy to identify, all vectors are written in boldface.

Given the set of edges E , we now construct the (normalized) weights according to:

$$c_{ij} = \frac{b_{ij}1(j \rightarrow i)}{\sum_{l \in V} b_{il}1(l \rightarrow i)} 1 \left(\sum_{l \in V} b_{il}1(l \rightarrow i) > 0, i \neq j \right), \quad i, j \in V, \quad (2.1)$$

where $j \rightarrow i$ indicates that $(j, i) \in E$. The model also requires two parameters, $c, d \in [0, 1]$, such that $0 < c + d \leq 1$ and $d > 0$. The parameter c controls the amount of value that individuals give to the social network, and d controls the amount of weight they give to external influences, e.g., media and political leaders. Although one can easily make the parameters c and d depend on each individual, for simplicity we interpret them as average weights and keep them the same across all individuals. Our assumption that $d > 0$ is important both to the explanatory/predictive power of the model, and to its mathematical properties, as we will see later.

Having described the attributes of each individual, we can now describe the evolution of the opinion process. At time zero, we initialize the process with a matrix $R^{(0)} \in [-1, 1]^{|V| \times \ell}$ chosen according to some initial

distribution μ_0 such that $\{\mathbf{R}_i^{(0)} : J_i = r\}$ are identically distributed for each $1 \leq r \leq K$. At each time step k , each individual i listens to an external signal $\mathbf{W}_i^{(k)} \in [-1, 1]^\ell$ (row vector), and updates their opinion according to:

$$\mathbf{R}_i^{(k+1)} = c \sum_{j \in V} c_{ij} \mathbf{R}_j^{(k)} + \mathbf{W}_i^{(k+1)} + (1 - c - d) \mathbf{R}_i^{(k)}, \quad k \geq 0, i \in V. \quad (2.2)$$

The external signals $\{\mathbf{W}_i^{(k)} : i \in V, k \geq 0\}$ are assumed to be independent of each other, and of the initial opinion matrix $R^{(0)}$. In order to include the possibility of having vertices with no inbound neighbors, it is convenient to assume that the external signals take the form:

$$\mathbf{W}_i^{(k)} = d \mathbf{Z}_i^{(k)} + c \mathbf{q}_i 1 (d_i^- = 0), \quad (2.3)$$

where $d_i^- = \sum_{j \in V} 1(j \rightarrow i)$ is the number of inbound neighbors of vertex i and $\{\mathbf{Z}_i^{(k)} : k \geq 0\}$ can be interpreted as media signals. This form of the external signals replaces the contribution of the neighbors with the vertex's internal belief when it has no inbound neighbors. For a vertex $i \in V$ having attribute vector \mathbf{a}_i , the media signals $\{\mathbf{Z}_i^{(k)} : k \geq 0\}$ are assumed to be i.i.d. with some distribution $\nu(\mathbf{a}_i)$. If needed, one could make ν depend on the extended attribute vector of each vertex, and even add more vertex properties to the extended attribute vector, but for simplicity we omit this extension here. We use $W^{(k)}$ to denote the $|V| \times \ell$ matrix whose i th row is $\mathbf{W}_i^{(k)}$.

Under the assumption that $d > 0$, we will see that the process $\{R^{(k)} : k \geq 0\}$ has a stationary distribution, i.e., $R^{(k)} \Rightarrow R$ as $k \rightarrow \infty$, where \Rightarrow denotes weak convergence on $[-1, 1]^{|V| \times \ell}$. This paper focuses on the behavior of both the trajectories $\{\mathbf{R}_i^{(k)} : k \geq 0\}$ and the stationary opinion matrix R .

2.2. The network. Although the opinion process itself is defined on a fixed directed graph, our results are based on analyzing the process on a random graph sequence $\{G_n : n \geq 1\}$, where $G_n = (V_n, E_n)$. From a modeling point of view, one can think of the fixed social network of interest as simply one realization from the random graph model being used. For simplicity, we assume $|V_n| = n$.

In order to model the community structure and individual-level confirmation bias, we use as the basis for our model a directed stochastic block model (dSBM) with K communities. Each individual in $i \in V_n$ has a community label J_i , and given the community labels $\mathcal{J}_n = \{J_i : i \in V_n\}$, edge (i, j) is present in the graph with probability:

$$p_{ij}^{(n)} = \frac{\kappa(J_i, J_j) \theta_n}{n} \wedge 1, \quad i, j \in V_n, \quad (2.4)$$

independently of everything else, where κ is a nonnegative kernel (equivalently, a nonnegative matrix in $\mathbb{R}^{K \times K}$), and θ_n is a density parameter.

Since our interest in this paper is to obtain approximations for semi-sparse to dense graphs, it is not important to incorporate degree-corrections into our model other than those produced by the kernel κ . In the sparse setting, where one can closely model the inhomogeneity of the network (e.g., scale-free degree distributions), using a degree-corrected dSBM would be more appropriate.

The proportion of vertices belonging to each community is controlled by the probability vector $(\pi_1^{(n)}, \dots, \pi_K^{(n)})$, so that a proportion $\pi_r^{(n)}$ of the vertices in V_n belong to community r . The main assumption on the distribution of the community labels is given below.

Assumption A. There exists a probability vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ with $\pi_r > 0$ for all $1 \leq r \leq K$, such that

$$\pi_r^{(n)} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(J_i = r) \xrightarrow{P} \pi_r,$$

as $n \rightarrow \infty$, for each $r \in \{1, \dots, K\}$.

Next, for each $r \in \{1, \dots, K\}$ let F_r be a distribution on $[-1, 1]^\ell$. For each vertex $i \in V_n$ sample its internal belief vector \mathbf{Q}_i according to F_{J_i} , where J_i is the community label of vertex i .

Finally, to complete the description of the model on a random graph, we sample each of the unnormalized weights $\{B_{ij} : i, j \in V_n\}$ conditionally independent of each other given the community labels $\{J_i : i \in V_n\}$, with B_{ij} sampled from a distribution G_{J_i, J_j} on the interval $[0, H]$, with H a finite constant. Note that the distribution of the unnormalized weights is allowed to depend on the communities to which each vertex belongs, which means one can give less weight to opinions coming from members of a different community. This is a form of *confirmation bias* [38, 39, 20], applied to the individual rather than the opinion itself.

Remarks 2.1. We end this section by explaining how one can choose the various parameters to model many of the features that social scientists believe influence the evolution of opinions. Since the full explanatory and predictive power of the model will be discussed in a future companion paper, we give here only a high-level description of what our model can be used for.

- a.) The kernel κ and the unnormalized weights $\{B_{ij} : i, j \in V_n\}$ can be used to model a form of confirmation bias that affects the individuals directly, i.e., by both controlling the number of neighbors that an individual has who belong to communities different from their own, and by discounting their opinions. Note that this form of confirmation bias depends on neither the topic nor the specific expressed opinions being shared at any point. Other ways of modeling confirmation bias directly applied to the expressed opinions being shared are discussed in Section 4.
- b.) A direct way in which our model incorporates selective exposure is through the choice of the distribution $\nu(\mathbf{A}_i)$ for the media signals, where $\mathbf{A}_i = (J_i, \mathbf{Q}_i)$. For example, if we believe that each community has its own preferred media outlets, e.g., news and political commentators, then we can choose the distribution ν to depend on the community label J_i . Moreover, if we want to model the human tendency to reject information that contradicts our internal beliefs or pre-existing views, we can do so by allowing ν to depend on the internal belief vector \mathbf{Q}_i (e.g., a person who listens to an outlet that has extreme views on topics 1 and 2, but only agrees with the outlet on topic 1, may choose to ignore the media signal for topic 2).
- c.) The dSBM can also be used to model the presence of bots, targeted or not, by creating additional communities. A targeted bot chooses who to send its messages to based on their community label. Since bots are not real people, they are not influenced by what they hear from their inbound neighbors, so we can model them to have no inbound neighbors by adjusting the kernel κ accordingly.

3. MAIN RESULTS

We are now ready to present the main results of the paper. We assume from here onwards that we are working on a probability space (Ω, \mathcal{F}, P) where we can construct the entire random graph sequence $\{G_n : n \geq 1\}$ along with their vertex marks, as well as all the external signals needed to define the opinion process on each of the graphs in the sequence. We use $E[\cdot]$ to denote the corresponding expectation. The opinion process $\{R^{(k)} : k \geq 0\}$ is constructed on a fixed random graph $G_n = (V_n, E_n)$ from the sequence, but for simplicity the dependence on n is not explicit in the notation.

As mentioned earlier, our work shows that there is a unique approximation for the opinion process $\{R^{(k)} : k \geq 0\}$ under the large graph limit $n \rightarrow \infty$, that works for all choices of the density parameter θ_n . The phase transition occurs only in the precision level of the approximation, around the threshold $\theta_n = \log n$. Interestingly, the proof technique is also significantly different above and below the threshold $\log n$, reflecting the fact that as the density parameter θ_n drops below the threshold, a vertex's entire neighborhood becomes increasingly important. To be able to drop the minimum in (2.4) and simplify our results, we assume throughout the paper that

$$\limsup_{n \rightarrow \infty} \max_{1 \leq r, s \leq K} \frac{\kappa(r, s)\theta_n}{n} \leq 1.$$

The process $\{R^{(k)} : k \geq 0\}$ will be approximated by another process $\{\mathcal{R}^{(k)} : k \geq 0\}$ whose main characteristic is that its rows $\{\mathcal{R}_i^{(k)} : k \geq 0, i \in V_n\}$ are conditionally independent of each other given the community labels. We will now describe the approximating process, leaving an intuitive derivation of how this process is obtained to Section 5.

To start, define the matrix $M \in [0, 1]^{K \times K}$ whose (r, s) th component is given by

$$m_{rs} = \frac{\pi_s \beta_{r,s} \kappa(s, r)}{\pi_1 \beta_{r,1} \kappa(1, r) + \dots + \pi_K \beta_{r,K} \kappa(K, r)} \cdot 1(\pi_1 \beta_{r,1} \kappa(1, r) + \dots + \pi_K \beta_{r,K} \kappa(K, r) > 0), \quad (3.1)$$

where $\beta_{r,s} = E[B_{ij} | J_i = r, J_j = s]$. Note that since one may want bot communities to have no inbound neighbors, M may have zero rows. Define also $v_{r,s} = E[B_{ij}^2 | J_i = r, J_j = s]$. Recall that the media signals, given by $\{\mathbf{Z}_i^{(k)} : k \geq 0, i \in V_n\}$, are independent of each other, with $\{\mathbf{Z}_i^{(k)} : k \geq 0\}$ conditionally i.i.d. given \mathbf{A}_i , with distribution $\nu(\mathbf{A}_i)$. The external signals $\{\mathbf{W}_i^{(k)} : k \geq 0, i \in V_n\}$ follow the form in (2.3). We will use $\bar{W} \in [-1, 1]^{K \times \ell}$ to denote the matrix whose r th row is $E[\mathbf{W}_i^{(0)} | J_i = r]$, and $\bar{R} \in [-1, 1]^{K \times \ell}$ to denote the matrix whose r th row is $E[\mathbf{R}_i^{(0)} | J_i = r]$.

The approximating process $\{\mathcal{R}^{(k)} : k \geq 0\}$ is given by: $\mathcal{R}^{(0)} = R^{(0)}$ and

$$\mathcal{R}_i^{(k)} = \sum_{t=0}^{k-1} (1-c-d)^t \mathbf{W}_i^{(k-t)} + 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} (M^s \bar{W})_{J_i \bullet} + \sum_{s=1}^k a_{s,k} (M^s \bar{R})_{J_i \bullet} + (1-c-d)^k \mathbf{R}_i^{(0)}, \quad (3.2)$$

for $k \geq 1$ and $i \in V_n$, where $a_{s,t} = \binom{t}{s} (1-c-d)^{t-s} c^s$ and $H_{j \bullet}$ denotes the j th row of matrix H ($H_{\bullet j}$ denotes the j th column). Note that conditionally on the vertex attributes $\{\mathbf{A}_i : i \in V_n\}$, (3.2) takes the form of an autoregressive process of order one, without any interactions among different vertices. Before stating the main theorems, we will need a few additional definitions.

For a vector $\mathbf{x} \in \mathbb{R}^n$, we use the standard l_p -norm $\|\cdot\|_p$, $p \geq 1$, namely,

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad \text{and} \quad \|\mathbf{x}\|_\infty := \max_{1 \leq i \leq n} |x_i|.$$

Applied to a matrix $A \in \mathbb{R}^{n \times \ell}$, $\|\cdot\|_p$ is the induced norm

$$\|A\|_p := \sup_{\|\mathbf{x}\| \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

We will also use the notation

$$\mu_r^{(n)} = \sum_{s=1}^K \beta_{r,s} \pi_s^{(n)} \kappa(s, r) \quad \text{and} \quad \nu_r^{(n)} = \sum_{s=1}^K \nu_{r,s} \pi_s^{(n)} \kappa(s, r),$$

as well as

$$\Delta_n = \max_{r \in \mathcal{I}} \frac{\nu_r^{(n)}}{(\mu_r^{(n)})^2} \quad \text{and} \quad \Lambda_n = \max_{r \in \mathcal{I}} \frac{\mu_r^{(n)}}{\nu_r^{(n)}},$$

where $\mathcal{I} = \{r \in \{1, \dots, K\} : \sum_{s=1}^K m_{rs} > 0\}$ is the set of non-zero rows of M . Note that both Δ_n and Λ_n converge to positive constants as $n \rightarrow \infty$ under Assumption A. We use $\mathbb{E}_n[\cdot] = E[\cdot | \mathcal{I}_n]$ to denote the conditional expectation given the community labels, and \mathbb{P}_n to denote its corresponding conditional probability. The first main result of the paper is given below.

Theorem 3.1. *Define the process $\{\mathcal{R}^{(k)} : k \geq 0\}$ according to (3.2). Suppose $\theta_n \geq (6H\Lambda_n)^2 \Delta_n \log n$. Then, there exists a constant $\Gamma < \infty$ such that*

$$\sup_{k \geq 0} \mathbb{E}_n \left[\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right] \leq \Gamma \left(\sqrt{\frac{\log n}{\theta_n}} + \mathcal{E}_n \right), \quad (3.3)$$

where $\mathcal{E}_n := \max_{1 \leq r, s \leq K} \left| \frac{\pi_s^{(n)} \pi_r - \pi_s \pi_r^{(n)}}{\pi_r^{(n)} \pi_s} \right|$. Moreover, for any sequence θ_n satisfying $\theta_n \rightarrow \infty$ as $n \rightarrow \infty$,

$$\sup_{k \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \xrightarrow{P} 0, \quad (3.4)$$

as $n \rightarrow \infty$.

Remarks 3.1. a.) Since $\max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \leq \mathbb{E}_n \left[\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right]$, Theorem 3.1 shows that the approximation is stronger when $\theta_n / \log n \rightarrow \infty$, and it gradually weakens as the rate at which θ_n grows drops below the critical rate $\log n$. Intuitively, this can be explained by noting that the average number of neighbors that any vertex has grows with the density parameter θ_n . The larger the number of neighbors, the more their aggregate contributions behave as the average opinion. The weakest result is valid for any $\theta_n \rightarrow \infty$, regardless of how slow the growth is. However, it is worth pointing out that when θ_n is bounded, the approximation fails, and the correct approximation is the *sparse* approximation, which is determined by the local neighborhood of each vertex (see Theorem 2 in [22]). The threshold $\log n$ determining the strength of the approximation also appears in [11, 10] and [3], where the authors study other processes on graphs with various edge density levels.

- b.) For each $i \in V_n$, the sequence $\{\mathbf{R}_i^{(k)} : k \geq 0\}$ is called the *trajectory* of vertex i . Since the rows in the limiting process $\{\mathcal{R}^{(k)} : k \geq 1\}$ are independent of each other, Theorem 3.1 yields that the trajectories of the process $\{R^{(k)} : k \geq 0\}$ are asymptotically independent; in other words, the process exhibits *propagation of chaos* (note that only particles in the same community are exchangeable). Since

$$\frac{1}{n} \sum_{i=1}^n \sup_{k \geq 0} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \leq \sup_{k \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \xrightarrow{P} 0, \quad n \rightarrow \infty,$$

the second statement in Theorem 3.1 implies pointwise propagation of chaos in the sense of Definition 4.1 in [13], throughout the entire range of k . We point out that this is not the case in the sparse regime, where the local approximation given by Theorem 2 in [22] corresponds to a process whose trajectories are not independent of each other, and the trajectories of vertices close to each other may in fact have significant dependence.

- c.) Note that since $\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty$ is bounded, Theorem 3.1 implies that for $\log n/\theta_n \rightarrow 0$, we have that for each $k \geq 0$,

$$\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

An alternative way of understanding the approximation is to look at the empirical measure

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{R}_i^{(0)}, \dots, \mathbf{R}_i^{(k)} \in A),$$

for any measurable set $A \subseteq ([-1, 1]^\ell)^{k+1}$. Written in this way, one can identify this measure as the conditional law of a *typical opinion*, understood as the trajectory of a uniformly chosen vertex $I_n \in V_n$, given the underlying graph $G_n = (V_n, E_n)$ and the entire process $\{\mathcal{R}^{(t)} : 0 \leq t \leq k\}$. In our model, the distribution of the typical trajectory converges as $n \rightarrow \infty$ to a law that depends only on the community label. Since in the random graph literature the typical vertex converges to the distribution of the root of a rooted graph (usually a branching tree), the notation we use in the following theorem is consistent with that used in the sparse regime, where \emptyset denotes the root of the limiting graph. We point out that in this paper, the limiting graph is not locally finite, so the notion of local weak convergence (see [49]) does not apply.

Theorem 3.2. Fix $k \geq 0$ and define the random variables $(\{\mathcal{R}_\emptyset^{(m)} : 0 \leq m \leq k\}, \mathcal{J}_\emptyset)$ according to:

$$P(\mathcal{J}_\emptyset = s) = \pi_s, \quad 1 \leq s \leq K,$$

$$P\left((\mathcal{R}_\emptyset^{(0)}, \mathcal{R}_\emptyset^{(1)}, \dots, \mathcal{R}_\emptyset^{(k)}) \in A \mid \mathcal{J}_\emptyset = s\right) = \mathbb{P}_n\left((\mathcal{R}_i^{(0)}, \mathcal{R}_i^{(1)}, \dots, \mathcal{R}_i^{(k)}) \in A\right), \quad A \subseteq ([-1, 1]^\ell)^{k+1},$$

where $\{\mathcal{R}_i^{(k)} : k \geq 0\}$ is given by (3.2) and i is any vertex belonging to community s . Define $V_{k,i} \in [-1, 1]^{\ell \times (k+1)}$ to be the matrix having j th column $(V_{k,i})_{\bullet j} = (\mathbf{R}_i^{(j-1)})^\top$, and let $\mathcal{V}_k \in [-1, 1]^{\ell \times (k+1)}$ be the matrix having j th column $(\mathcal{V}_k)_{\bullet j} = (\mathcal{R}_\emptyset^{(j-1)})^\top$. Then, for any $1 \leq r \leq K$ and any bounded and continuous

(with respect to the $\|\cdot\|_1$ operator norm) function $f : [-1, 1]^{\ell \times (k+1)} \rightarrow \mathbb{R}$, we have

$$\frac{1}{n} \sum_{i=1}^n f(V_{k,i}) 1(J_i = r) \xrightarrow{P} E[f(\mathcal{V}_k) 1(\mathcal{J}_\emptyset = r)]$$

as $n \rightarrow \infty$. Furthermore, for any arbitrary collection of vertices $\{i_1, \dots, i_m\} \subseteq V_n$ having community labels $\{r_1, \dots, r_m\}$, $m \geq 1$, and any set of continuous (with respect to $\|\cdot\|_1$) bounded functions $\{f_1, \dots, f_m\}$ on $[-1, 1]^{\ell \times (k+1)}$, $k \geq 0$, we have

$$\mathbb{E}_n \left[\prod_{j=1}^m f_j(V_{k,i_j}) \right] \xrightarrow{P} \prod_{j=1}^m E[f_j(\mathcal{V}_k) | \mathcal{J}_\emptyset = r_j], \quad n \rightarrow \infty.$$

Remark 3.1. The last statement in Theorem 3.2 is closely related to Theorem 2.6.B in [23], which states that trajectories of a finite set of vertices uniformly chosen at random are asymptotically independent of each other, a limited form of propagation of chaos that holds only for vertices that are likely to be far from each other. However, the last statement in Theorem 3.2 allows the vertices to be chosen arbitrarily.

The last set of results in the paper refers to the stationary behavior of the Markov chain $\{R^{(k)} : k \geq 0\}$. Recall that R denotes a matrix in $[-1, 1]^{n \times \ell}$ having the stationary distribution of the chain, and \mathbf{R}_{I_n} denotes a uniformly chosen row of the matrix R . Note that \mathbf{R}_{I_n} represents the *typical stationary opinion* on the graph, across all ℓ topics. Our last theorem provides a characterization for \mathbf{R}_{I_n} when n is large, by establishing the existence of a random variable $\mathcal{R}_\emptyset \in [-1, 1]^\ell$ such that

$$\mathbf{R}_{I_n} \Rightarrow \mathcal{R}_\emptyset, \quad n \rightarrow \infty,$$

where \Rightarrow denotes weak convergence on $[-1, 1]^\ell$. Since \mathcal{R}_\emptyset also turns out to be the weak limit of $\{\mathcal{R}_\emptyset^{(k)} : k \geq 0\}$ as $k \rightarrow \infty$, our results show that we can exchange the limits in k (time) and n (number of vertices).

Theorem 3.3. Define the random variables $(\mathcal{R}_\emptyset, \mathcal{J}_\emptyset)$ according to:

$$P(\mathcal{R}_\emptyset \in A | \mathcal{J}_\emptyset = r) = \mathbb{P}_n \left(\sum_{t=0}^{\infty} (1-c-d)^t \mathbf{W}_i^{(t)} + \sum_{t=1}^{\infty} \sum_{s=1}^t a_{s,t} (M^s \bar{W})_{J_i} \in A \mid J_i = r \right), \quad A \subseteq [-1, 1]^\ell,$$

$$P(\mathcal{J}_\emptyset = r) = \pi_r, \quad 1 \leq r \leq K.$$

Then, there exists a coupling $(\mathbf{R}_{I_n}, J_{I_n}, \mathcal{R}_\emptyset, \mathcal{J}_\emptyset)$, such that for any continuous (with respect to $\|\cdot\|_1$) and bounded function $f : [-1, 1]^\ell \rightarrow \mathbb{R}$, we have for any $1 \leq r \leq K$,

$$\mathbb{E}_n [\|\mathbf{R}_{I_n} - \mathcal{R}_\emptyset\|_1] \xrightarrow{P} 0 \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n f(\mathbf{R}_i) 1(J_i = r) \xrightarrow{P} E[f(\mathcal{R}_\emptyset) 1(\mathcal{J}_\emptyset = r)]$$

as $n \rightarrow \infty$. Furthermore, for any arbitrary collection of vertices $\{i_1, \dots, i_m\} \subseteq V_n$ having community labels $\{r_1, \dots, r_m\}$, $m \geq 1$, and any set of continuous (with respect to $\|\cdot\|_1$) bounded functions $\{f_1, \dots, f_m\}$ on $[-1, 1]^\ell$, we have

$$\mathbb{E}_n \left[\prod_{j=1}^m f_j(\mathbf{R}_{i_j}) \right] \xrightarrow{P} \prod_{j=1}^m E[f_j(\mathcal{R}_\emptyset) | \mathcal{J}_\emptyset = r_j], \quad n \rightarrow \infty.$$

Remark 3.2. Note that the shape of \mathcal{R}_\emptyset is simply the limit as $k \rightarrow \infty$ of $\mathcal{R}_{I_n}^{(k)}$ after a time reversal of the external signals, i.e., $\mathbf{W}_{I_n}^{(k-t)} \stackrel{\mathcal{D}}{=} \mathbf{W}_{I_n}^{(t)}$.

The proofs of all the theorems are contained in Section 5.

4. RELATED LITERATURE

Interest in the evolution of opinions on social networks dates back to the 50's, with the work of French [24] and DeGroot [19]. The model in [19] corresponds in our notation to the recursion $\mathbf{R}^{(k+1)} = A\mathbf{R}^{(k)}$, for a given $\mathbf{R}^{(0)}$ and $A = cC + (1-c)I$, with $C = (c_{ij})$ as in (2.1) (the model is for a single topic, hence the vectors instead of matrices). The main question studied in [19, 8] was the existence of consensus, seen as a limit $\mathbf{R} = \lim_{k \rightarrow \infty} \mathbf{R}^{(k)}$ whose components are all equal to each other (a consequence of the Perron-Frobenius theorem when the matrix A is irreducible and aperiodic). In [25], Friedkin and Johnsen extended the model to allow disagreement by considering the recursion $\mathbf{R}^{(k+1)} = A\mathbf{R}^{(k)} + \mathbf{Q}$, with \mathbf{Q} a deterministic vector, which is a special case of our model. The models in [51] replace the deterministic \mathbf{Q} with a sequence of random signals $\{\mathbf{W}^{(k)} : k \geq 0\}$, as we do, although they are independent of the underlying network, since they are used to model miscommunication. A more recent generalization of the DeGroot model was given in [16], where the authors introduced the concept of *biased assimilation*, namely the tendency of individuals to pull their opinion towards their initial opinion once presented with inconclusive evidence, in order to make the repeated averaging scheme of DeGroot lead to polarization. Another closely related model that is capable of introducing confirmation bias by rejecting an opinion from a neighbor when it is too different from their own expressed opinion is the Hegselmann-Krause model [42], which due to its non-linear updating rule is not covered by our analysis. All these models are defined on a fixed graph, so the stochastic matrix A is deterministic, unlike ours, where we use both the updating rules of the opinion process and the properties of the underlying graph to explain complex phenomena. Other variants of the DeGroot and Friedkin-Johnsen models that incorporate multiple topics and time-varying self-weights (e.g., d in our model is allowed to be different for each vertex and change with time) are given in [31, 41, 47].

While all the models mentioned above are defined on discrete time with synchronous updating rules, there is a different family of popular opinion models where interactions occur one at a time, at either random or deterministic times. Among those are the bounded confidence models, such as the Deffuant-Weisbuch model [50, 18, 17]. In this model, at each time step a pair of neighboring individuals is randomly selected. Then, the two neighbors update their opinions according to a simple linear equation if, and only if, the difference of their current opinions is less than a specified threshold. This model, like the Hegselmann-Krause model, is meant to explain fragmentation and polarization through a direct confirmation bias. Its non-linearity and lack of an external signal means that its analysis is usually done using the methods of interacting particle systems or numerical stochastic simulation [30, 6, 34, 15, 28, 37]. A more recent analysis of the Deffuant-Weisbuch model that takes into account the network's topology is [21]. Our model is not meant to tackle confirmation bias at the level of the opinions being shared, but rather at the level of the individuals who share them, and the main appeal of our model is that it accounts for the combined effects of the network topology, the media framing effect, and people's selective exposure, while still producing computable formulas for means and variances for the opinion process.

As mentioned in the introduction, the main focus of this paper is not the explanatory and predictive power of our model, but rather its behavior on graphs of various levels of edge density. So, before we discuss that aspect of the related literature, we just want to mention that there is yet another family of classical opinion models taking values on finite sets (rather than continuous intervals), like the voter model [36], where opinions can be either 0 or 1, or the Axelrod model [4], where each vertex has a fixed set of features and traits, leading to a notion of “similarity” between vertices. The former can be used to model consensus through asynchronous updates, whereas the latter achieves a diversity of opinions by using a trait updating rule that depends on how similar the features of the two interacting vertices are, a concept known as *homophily*. The underlying graph in both of these models is usually a lattice, although the voter model has also been analyzed in heterogeneous graphs in [45]. The work in [43] contains a review of interesting variants of the voter model.

Our opinion model falls in the category of interacting particle systems in discrete time with synchronous updates and continuous state space, although only vertices in the same community are exchangeable. The approximating process $\{\mathcal{R}^{(k)} : k \geq 0\}$ is a classical mean-field approximation, where the complex interactions between neighboring vertices are replaced with their average behavior. Our main theorem proves propagation of chaos by coupling the trajectories, as described in the surveys [13, 12]. The literature on mean-field approximations is so vast, and covers such a wide range of fields, that it would be impossible to mention all the related models for which results similar to ours have been derived. The closest work to the techniques used in this paper is the analysis of the personalized PageRank algorithm done in [3], where a stochastic block model is the only source of randomness in what otherwise would be a deterministic linear recursion. The main results in [3] are closely related to our Theorem 3.1 for density parameters $\theta_n / \log n \rightarrow \infty$, albeit for a different recursion and under an l_2 -norm (rather than the l_∞ -norm we use). Our work shows that the approximation itself continues to hold beyond the $\log n$ threshold, for any $\theta_n \rightarrow \infty$, but with lesser precision (a weaker norm). Relative to the majority of mean-field results in the interacting particles literature, our work is different in the explicit role that the underlying random graph plays, its ability to seamlessly cover the entire range of edge densities (other than the sparse case), the explicit characterization of the limiting process as the number of particles goes to infinity, and the exchange of limits with the stationary version of the limiting process. Other works that analyze stochastic processes on graphs with various levels of edge density are [11], which studies the supermarket model on graphs, and [10], which analyzes weakly interacting diffusions on time varying random graphs. It is also worth pointing out that mean-field approximations for bounded confidence models such as the Deffuant-Weisbuch model and some of its variants have been given in [34, 15, 28].

Finally, we would like to point out that the analysis of our model in the semi-sparse regime, i.e., θ_n below the $\log n$ threshold, is based on a coupling of the underlying random graph with a marked multi-type branching process. This technique is based on the notion of local weak convergence of random graph sequences [7, 2] (see also [49]). Although for non-sparse random graphs the local weak limit is not a locally bounded graph, the dSBM can still be coupled with a marked multi-type Galton-Watson tree (with average degree that grows with n). In fact, a strong coupling, in the sense of the intermediate coupling in [40], still holds, which is what enables the work in this paper. For sparse graphs, local weak convergence or strong couplings have been used to study the PageRank algorithm in [14, 26, 5], general discrete-time stochastic recursions on sparse graphs in [23], and interacting diffusions in [32, 33]. The techniques in this paper

will likely extend to more general classes of stochastic recursions on random graphs with growing edge densities.

5. PROOFS

This section contains the proofs for all the theorems in Section 3. The proof of our main result, Theorem 3.1, requires two different approaches depending on the speed at which $\theta_n \rightarrow \infty$, however, both approaches start with the same intermediate approximation. We will start by presenting the intermediate approximation in this section, and then complete the analysis for the two different density regimes; in Section 5.1 for the case when $\theta_n \geq \gamma \log n$ for some $\gamma > 0$, and in Section 5.2 for the case when $\limsup_{n \rightarrow \infty} \log n / \theta_n > 0$. Finally, as mentioned earlier, Theorems 3.2 and 3.3 are consequences of Theorem 3.1, so we give their proofs in Section 5.3.

Recall that the opinion process will be constructed on a fixed directed graph G_n sampled according to Section 2.2, where each vertex $i \in V_n$ has an attribute $\mathbf{A}_i = (J_i, \mathbf{Q}_i)$ and an extended attribute $\mathbf{Y}_i = (J_i, \mathbf{Q}_i, \mathbf{B}_i)$, with $\mathbf{B}_i = (B_{ij})_{1 \leq j \leq n}$ the unnormalized weights.

In order for us to derive the intermediate approximation, we first note that the opinion recursion (2.2) can be rewritten in matrix notation as

$$R^{(k)} = AR^{(k-1)} + W^{(k)}, \quad (5.1)$$

where $R^{(k)} \in [-1, 1]^{n \times \ell}$ and $W^{(k)} \in [-1, 1]^{n \times \ell}$ are the matrices having (i, j) th components $R_{ij}^{(k)}$ and $W_{ij}^{(k)}$, respectively, and $A \in [0, 1]^{n \times n}$ the matrix satisfying

$$A_{ij} = c \cdot C_{ij}, \quad i \neq j, \quad A_{ii} = 1 - c - d,$$

with C_{ij} constructed according to (2.1) using the random unnormalized weights $\{B_{ij} : i, j \in V_n\}$ (the use of upper case is to remind the reader that these are random). Iterating (5.1) gives the explicit solution

$$R^{(k)} = \sum_{t=0}^{k-1} A^t W^{(k-t)} + A^k R^{(0)}.$$

Using the observation that $A = cC + (1 - c - d)I$, where C is the matrix having (i, j) th component C_{ij} , and I is the identity matrix in $\mathbb{R}^{n \times n}$, we get

$$A^t = (cC + (1 - c - d)I)^t = \sum_{s=0}^t \binom{t}{s} (1 - c - d)^{t-s} c^s C^s = \sum_{s=0}^t a_{s,t} C^s, \quad t \geq 0,$$

and, therefore,

$$R^{(k)} = \sum_{t=0}^{k-1} \sum_{s=0}^t a_{s,t} C^s W^{(k-t)} + \sum_{s=0}^k a_{s,k} C^s R^{(0)}, \quad k \geq 1. \quad (5.2)$$

Now, construct the matrix $\tilde{M} \in \mathbb{R}^{n \times n}$ according to

$$\tilde{M}_{ij} = \frac{\beta_{J_i, J_j} \kappa(J_j, J_i)}{n \left(\beta_{J_i, 1} \pi_1^{(n)} \kappa(1, J_i) + \cdots + \beta_{J_i, K} \pi_K^{(n)} \kappa(K, J_i) \right)} \mathbf{1}(i \neq j),$$

and recall that $\beta_{r,s} = E[B_{ij}|J_i = r, J_j = s]$, with $\tilde{M}_{ij} \equiv 0$ if the denominator is zero. The intuition behind this definition of \tilde{M} is that it is the ‘‘approximate’’ conditional mean of C given the community labels \mathcal{J}_n , meaning that for $i, j \in V_n$, $i \neq j$, and n sufficiently large (so that the minimum in $p_{ij}^{(n)}$ can be dropped), we have

$$\tilde{M}_{ij} = \frac{\beta_{J_i, J_j} p_{ij}^{(n)}}{n \left(\pi_1^{(n)} \beta_{J_i, 1} p_{1i}^{(n)} + \cdots + \pi_K^{(n)} \beta_{J_i, K} p_{Ki}^{(n)} \right)} = \frac{\mathbb{E}_n[B_{ij} 1(j \rightarrow i)]}{\mathbb{E}_n \left[\sum_{r=1}^n B_{ir} 1(r \rightarrow i) \right]}.$$

Define $\check{W} = \mathbb{E}_n [W^{(0)}]$, $\check{R} = \mathbb{E}_n [R^{(0)}]$, and construct the new process $\{\tilde{R}^{(k)} : k \geq 0\}$ defined by

$$\tilde{R}^{(0)} = R^{(0)},$$

$$\tilde{R}^{(k)} = \sum_{t=0}^{k-1} (1-c-d)^t W^{(k-t)} + 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \tilde{M}^s \check{W} + \sum_{s=1}^k a_{s,k} \tilde{M}^s \check{R} + (1-c-d)^k R^{(0)}, \quad k \geq 1. \quad (5.3)$$

Intuitively, $\tilde{R}^{(k)}$ replaces all neighbor contributions with their approximate means, i.e., every term of the form $C^s X$ with $s \geq 1$ and X a random matrix is replaced with $\tilde{M}^s \mathbb{E}_n[X]$. It is important to point out that the rows of $\tilde{R}^{(k)}$, which we will denote $\{\tilde{\mathbf{R}}_i^{(k)} : i \in V_n\}$, become independent as a consequence. Our first result computes a bound for $\mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right]$ and $\max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \tilde{\mathbf{R}}_i^{(k)} \right\|_1 \right]$.

Theorem 5.1. *For any $k \geq 1$, we have*

$$\begin{aligned} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \tilde{\mathbf{R}}_i^{(k)} \right\|_1 \right] &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s W^{(0)})_{i\bullet} - (\tilde{M}^s \check{W})_{i\bullet} \right\|_1 \right] \\ &\quad + \sum_{s=1}^k a_{s,k} \max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s R^{(0)})_{i\bullet} - (\tilde{M}^s \check{R})_{i\bullet} \right\|_1 \right]. \end{aligned}$$

Furthermore, for $h(X) = \mathbb{E}_n \left[\left\| (C - \tilde{M})X \right\|_{\infty} \right]$, we have

$$\begin{aligned} \mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right] &\leq \max_{r \geq 1} h(\tilde{M}^r \check{W}) \sum_{t=0}^{k-1} (1-d)^{t-1} t + \mathbb{E}_n \left[\left\| C W^{(0)} - \tilde{M} \check{W} \right\|_{\infty} \right] \sum_{t=0}^{k-1} (1-d)^t \\ &\quad + \max_{r \geq 1} h(\tilde{M}^r \check{R}) k(1-d)^{k-1} + \mathbb{E}_n \left[\left\| C R^{(0)} - \tilde{M} \check{R} \right\|_{\infty} \right] (1-d)^k. \end{aligned}$$

Proof. Recall that $a_{s,t} = \binom{t}{s} (1-c-d)^{t-s} c^s$ and note that for any $i \in V_n$ we have

$$\begin{aligned} \left\| \mathbf{R}_i^{(k)} - \tilde{\mathbf{R}}_i^{(k)} \right\|_1 &= \left\| 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \left(C^s W^{(k-t)} - \tilde{M}^s \check{W} \right)_{i\bullet} + \sum_{s=1}^k a_{s,k} \left(C^s R^{(0)} - \tilde{M}^s \check{R} \right)_{i\bullet} \right\|_1 \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \left\| \left(C^s W^{(k-t)} - \tilde{M}^s \check{W} \right)_{i\bullet} \right\|_1 + \sum_{s=1}^k a_{s,k} \left\| \left(C^s R^{(0)} - \tilde{M}^s \check{R} \right)_{i\bullet} \right\|_1. \end{aligned}$$

To obtain the first statement of the theorem simply take expectations on both sides to get

$$\begin{aligned} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \tilde{\mathbf{R}}_i^{(k)} \right\|_1 \right] &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \left(C^s W^{(k-t)} - \tilde{M}^s \check{W} \right)_{i \bullet} \right\|_1 \right] \\ &\quad + \sum_{s=1}^k a_{s,k} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \left(C^s R^{(0)} - \tilde{M}^s \check{R} \right)_{i \bullet} \right\|_1 \right]. \end{aligned}$$

To continue to the second statement, note that from the properties of the norm $\| \cdot \|_\infty$, we obtain that

$$\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_\infty \leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \left\| C^s W^{(k-t)} - \tilde{M}^s \check{W} \right\|_\infty + \sum_{s=1}^k a_{s,k} \left\| C^s R^{(0)} - \tilde{M}^s \check{R} \right\|_\infty.$$

Furthermore, since for any matrices $X, \check{X} \in \mathbb{R}^{n \times \ell}$ and any $s \geq 1$ we have

$$\begin{aligned} C^s X - \tilde{M}^s \check{X} &= C^{s-1} (C X - \tilde{M} \check{X}) + (C^{s-1} - \tilde{M}^{s-1}) \tilde{M} \check{X} \\ &= C^{s-1} (C X - \tilde{M} \check{X}) + \sum_{r=1}^{s-1} C^{s-r-1} (C - \tilde{M}) \tilde{M}^r \check{X}, \end{aligned}$$

then

$$\begin{aligned} \left\| C^s X - \tilde{M}^s \check{X} \right\|_\infty &\leq \sum_{r=1}^{s-1} \left\| C^{s-r-1} (C - \tilde{M}) \tilde{M}^r \check{X} \right\|_\infty + \left\| C^{s-1} (C X - \tilde{M} \check{X}) \right\|_\infty \\ &\leq \sum_{r=1}^{s-1} \|C\|_\infty^{s-r-1} \left\| (C - \tilde{M}) \tilde{M}^r \check{X} \right\|_\infty + \|C\|_\infty^{s-1} \left\| C X - \tilde{M} \check{X} \right\|_\infty \\ &= \sum_{r=1}^{s-1} \left\| (C - \tilde{M}) \tilde{M}^r \check{X} \right\|_\infty + \left\| C X - \tilde{M} \check{X} \right\|_\infty, \end{aligned}$$

where we used the fact that $\|C\|_\infty = 1$. We conclude that

$$\begin{aligned} &\mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_\infty \right] \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \mathbb{E}_n \left[\left\| C^s W^{(k-t)} - \tilde{M}^s \check{W} \right\|_\infty \right] + \sum_{s=1}^k a_{s,k} \mathbb{E}_n \left[\left\| C^s R^{(0)} - \tilde{M}^s \check{R} \right\|_\infty \right] \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \left(\sum_{r=1}^{s-1} \mathbb{E}_n \left[\left\| (C - \tilde{M}) \tilde{M}^r \check{W} \right\|_\infty \right] + \mathbb{E}_n \left[\left\| C W^{(k-t)} - \tilde{M} \check{W} \right\|_\infty \right] \right) \\ &\quad + \sum_{s=1}^k a_{s,k} \left(\sum_{r=1}^{s-1} \mathbb{E}_n \left[\left\| (C - \tilde{M}) \tilde{M}^r \check{R} \right\|_\infty \right] + \mathbb{E}_n \left[\left\| C R^{(0)} - \tilde{M} \check{R} \right\|_\infty \right] \right) \\ &\leq \max_{r \geq 1} h(\tilde{M}^r \check{W}) 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} (s-1) + \mathbb{E}_n \left[\left\| C W^{(0)} - \tilde{M} \check{W} \right\|_\infty \right] 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \end{aligned}$$

$$+ \max_{r \geq 1} h(\tilde{M}^r \check{R}) \sum_{s=1}^k a_{s,k}(s-1) + \mathbb{E}_n \left[\left\| CR^{(0)} - \tilde{M} \check{R} \right\|_{\infty} \right] \sum_{s=1}^k a_{s,k}.$$

To compute the double sums more explicitly, let $\text{Bin}(n, p)$ denote a binomial random variable with parameters (n, p) , and note that for $k \geq 1$,

$$\sum_{t=1}^k \sum_{s=1}^t a_{s,t} s = \sum_{t=1}^k \sum_{s=1}^t \binom{t}{s} (1-c-d)^{t-s} c^s s = \sum_{t=1}^k (1-d)^t E[\text{Bin}(t, c/(1-d))] = c \sum_{t=1}^k (1-d)^{t-1} t, \quad (5.4)$$

and

$$\begin{aligned} \sum_{t=1}^k \sum_{s=1}^t a_{s,t} &= \sum_{t=1}^k \sum_{s=1}^t \binom{t}{s} (1-c-d)^{t-s} c^s = \sum_{t=1}^k (1-d)^t P(\text{Bin}(t, c/(1-d)) \geq 1) \\ &= \sum_{t=1}^k (1-d)^t (1 - (1-c-d)^t / (1-d)^t) = \sum_{t=1}^k ((1-d)^t - (1-c-d)^t). \end{aligned} \quad (5.5)$$

Similarly, the single sums can be computed to be

$$\sum_{s=1}^k a_{s,k} s = \sum_{s=1}^k \binom{k}{s} (1-c-d)^{k-s} c^s s = (1-d)^k E[\text{Bin}(k, c/(1-d))] = ck(1-d)^{k-1} \quad (5.6)$$

and

$$\sum_{s=1}^k a_{s,k} = \sum_{s=1}^k \binom{k}{s} (1-c-d)^{k-s} c^s = (1-d)^k P(\text{Bin}(k, c/(1-d)) \geq 1) = (1-d)^k - (1-c-d)^k. \quad (5.7)$$

Consequently,

$$\begin{aligned} \mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right] &\leq \max_{r \geq 1} h(\tilde{M}^r \check{W}) 1(k \geq 2) \sum_{t=1}^{k-1} (1-d)^{t-1} t \\ &\quad + \mathbb{E}_n \left[\left\| CW^{(0)} - \tilde{M} \check{W} \right\|_{\infty} \right] 1(k \geq 2) \sum_{t=1}^{k-1} (1-d)^t + \\ &\quad + \max_{r \geq 1} h(\tilde{M}^r \check{R}) k(1-d)^{k-1} + \mathbb{E}_n \left[\left\| CR^{(0)} - \tilde{M} \check{R} \right\|_{\infty} \right] (1-d)^k. \end{aligned}$$

□

Once the intermediate approximation $\{\tilde{R}^{(k)} : k \geq 0\}$ given by (5.3) is obtained, its connection to the final approximation $\{\mathcal{R}^{(k)} : k \geq 0\}$ given by (3.2) becomes apparent, since for all $s \geq 1$ we have

$$(\tilde{M}^s \check{W})_{i\bullet} \approx (M^s \bar{W})_{J_i\bullet} \quad \text{and} \quad (\tilde{M}^s \check{R})_{i\bullet} \approx (M^s \bar{R})_{J_i\bullet},$$

for large n . The following result provides a bound for the norm between these two processes.

Theorem 5.2. For any $\theta_n \rightarrow \infty$,

$$\sup_{k \geq 0} \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} \leq \frac{\ell c}{d^2} \cdot \mathcal{E}_n,$$

$$\text{where } \mathcal{E}_n = \max_{1 \leq r, s \leq K} \left| \frac{\pi_s^{(n)} \pi_r - \pi_s \pi_r^{(n)}}{\pi_r^{(n)} \pi_s} \right|.$$

Proof. Note that for any $1 \leq j \leq K$,

$$\begin{aligned} & \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} \\ &= \max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} (\tilde{M}^s \check{W})_{ij} + \sum_{s=1}^k a_{s,k} (\tilde{M}^s \check{R})_{ij} - \right. \\ & \quad \left. - 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} (M^s \bar{W})_{J_i j} - \sum_{s=1}^k a_{s,k} (M^s \bar{R})_{J_i j} \right| \\ &= \max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \left((\tilde{M}^s \check{W})_{ij} - (M^s \bar{W})_{J_i j} \right) + \sum_{s=1}^k a_{s,k} \left((\tilde{M}^s \check{R})_{ij} - (M^s \bar{R})_{J_i j} \right) \right| \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| (\tilde{M}^s \check{W})_{ij} - (M^s \bar{W})_{J_i j} \right| + \sum_{s=1}^k a_{s,k} \max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| (\tilde{M}^s \check{R})_{ij} - (M^s \bar{R})_{J_i j} \right|. \end{aligned}$$

Note that $\tilde{M} \in \mathbb{R}^{n \times n}$ is such that all the rows $\mathcal{S}_r = \{\tilde{M}_{i \bullet} : J_i = r\}$ are equal to each other, all the columns $\{\tilde{M}_{\bullet j} : J_j = r\}$ are equal to each other, and the cardinality of the \mathcal{S}_r is $n\pi_r^{(n)}$, for $1 \leq r \leq K$. Similarly, the vector $\check{W}_{\bullet j} \in \mathbb{R}^n$ satisfies that all the components $\{\check{W}_{ij} : J_i = r\}$ are equal to each other and the order in which the vertex labels appear is the same as the columns of \tilde{M} . Define the matrix $\check{M} \in \mathbb{R}^{K \times K}$ according to

$$\check{M}_{rs} = \frac{\beta_{r,s} \pi_s^{(n)} \kappa(s, r)}{\beta_{r,1} \pi_1^{(n)} \kappa(1, r) + \cdots + \beta_{r,K} \pi_K^{(n)} \kappa(K, r)}, \quad 1 \leq r, s \leq K, \quad (5.8)$$

with $\check{M}_{rs} \equiv 0$ if the denominator is zero, and note that

$$\check{M}_{rs} = \sum_{j=1}^n \tilde{M}_{ij} 1(J_j = s) \quad \text{for any } i \text{ such that } J_i = r.$$

It follows that for any $i \in V_n$,

$$(\tilde{M} \check{W})_{ij} = \sum_{l=1}^n \tilde{M}_{il} \check{W}_{lj} = \sum_{r=1}^K \sum_{l=1}^n \tilde{M}_{il} \check{W}_{lj} 1(J_l = r) = \sum_{r=1}^K \check{M}_{J_i, r} \bar{W}_{rj} = (\check{M} \bar{W})_{J_i j}. \quad (5.9)$$

Now suppose that $(\tilde{M}^{s-1} \check{W})_{ij} = (\check{M}^{s-1} \bar{W})_{J_i j}$ for $s \geq 2$, and note that

$$(\tilde{M}^s \check{W})_{ij} = \sum_{l=1}^n \tilde{M}_{il} (\tilde{M}^{s-1} \check{W})_{lj} = \sum_{r=1}^K \sum_{l=1}^n \tilde{M}_{il} (\check{M}^{s-1} \bar{W})_{J_l j} 1(J_l = r) = \sum_{r=1}^K \check{M}_{J_i r} (\check{M}^{s-1} \bar{W})_{rj} = (\check{M}^s \bar{W})_{J_i j},$$

therefore, by induction, we conclude that

$$(\check{M}^s \check{W})_{ij} = (\check{M}^s \bar{W})_{J_{ij}} \quad \text{for all } s \geq 1 \text{ and all } i \in V_n.$$

The same arguments also yield $(\check{M}^s \check{R})_{ij} = (\check{M}^s \bar{R})_{J_{ij}}$ for all $s \geq 1$ and all $i \in V_n$. Hence,

$$\max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| (\check{M}^s \check{W})_{ij} - (M^s \bar{W})_{J_{ij}} \right| = \max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| (\check{M}^s \bar{W})_{J_{ij}} - (M^s \bar{W})_{J_{ij}} \right| = \|(\check{M}^s - M^s) \bar{W}\|_{\infty}$$

and

$$\max_{1 \leq i \leq n} \sum_{j=1}^{\ell} \left| (\check{M}^s \bar{R})_{ij} - (M^s \bar{R})_{J_{ij}} \right| = \|(\check{M}^s - M^s) \bar{R}\|_{\infty}.$$

The first inequality now becomes

$$\begin{aligned} \left\| \check{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \|(\check{M}^s - M^s) \bar{W}\|_{\infty} + \sum_{s=1}^k a_{s,k} \|(\check{M}^s - M^s) \bar{R}\|_{\infty} \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \|\check{M}^s - M^s\|_{\infty} \|\bar{W}\|_{\infty} + \sum_{s=1}^k a_{s,k} \|\check{M}^s - M^s\|_{\infty} \|\bar{R}\|_{\infty} \\ &\leq \ell \sum_{t=1}^k \sum_{s=1}^t a_{s,t} \|\check{M}^s - M^s\|_{\infty}, \end{aligned}$$

where in the last step we used the observation that $\|\bar{W}\|_{\infty} \leq \ell$ and $\|\bar{R}\|_{\infty} \leq \ell$. For the remaining norm, note that

$$\begin{aligned} x_s &:= \|\check{M}^s - M^s\|_{\infty} \leq \|\check{M}^s - \check{M}^{s-1} M\|_{\infty} + \|\check{M}^{s-1} M - M^s\|_{\infty} \\ &\leq \|\check{M}\|_{\infty}^{s-1} \|\check{M} - M\|_{\infty} + \|\check{M}^{s-1} - M^{s-1}\|_{\infty} \|M\|_{\infty} \leq \|\check{M} - M\|_{\infty} + x_{s-1} \\ &\leq s \|\check{M} - M\|_{\infty}, \end{aligned}$$

since $x_0 = 0$ and $\|M\|_{\infty} = \|\check{M}\|_{\infty} = 1$. Finally, note that for \mathcal{I} the set of non-zero rows of M (same as the non-zero rows of \check{M}), we have

$$\begin{aligned} \|\check{M} - M\|_{\infty} &= \max_{r \in \mathcal{I}} \sum_{s=1}^K \left| \frac{\beta_{r,s} \pi_s^{(n)} \kappa(s, r)}{\beta_{r,1} \pi_1^{(n)} \kappa(1, r) + \cdots + \beta_{r,K} \pi_K^{(n)} \kappa(K, r)} - \frac{\beta_{r,s} \pi_s \kappa(s, r)}{\beta_{r,1} \pi_1 \kappa(1, r) + \cdots + \beta_{r,K} \pi_K \kappa(K, r)} \right| \\ &= \max_{r \in \mathcal{I}} \frac{\sum_{s=1}^K \beta_{r,s} \kappa(s, r) \left| (\pi_s^{(n)} \pi_1 - \pi_s \pi_1^{(n)}) \beta_{r,1} \kappa(1, r) + \cdots + (\pi_s^{(n)} \pi_K - \pi_s \pi_K^{(n)}) \beta_{r,K} \kappa(K, r) \right|}{(\beta_{r,1} \pi_1^{(n)} \kappa(1, r) + \cdots + \beta_{r,K} \pi_K^{(n)} \kappa(K, r)) (\beta_{r,1} \pi_1 \kappa(1, r) + \cdots + \beta_{r,K} \pi_K \kappa(K, r))} \\ &= \max_{r \in \mathcal{I}} \frac{\sum_{s=1}^K \beta_{r,s} \pi_s \kappa(s, r) \sum_{l=1}^K \left| \frac{\pi_s^{(n)} \pi_l - \pi_s \pi_l^{(n)}}{\pi_l^{(n)} \pi_s} \right| \beta_{r,l} \pi_l^{(n)} \kappa(l, r)}{\sum_{m=1}^K \beta_{r,m} \pi_m^{(n)} \kappa(m, r) \sum_{n=1}^K \beta_{r,n} \pi_n \kappa(n, r)} \end{aligned}$$

$$\leq \max_{1 \leq r, s \leq K} \left| \frac{\pi_s^{(n)} \pi_r - \pi_s \pi_r^{(n)}}{\pi_r^{(n)} \pi_s} \right| = \mathcal{E}_n.$$

Combining our estimates, we deduce

$$\left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} \leq \ell \mathcal{E}_n \sum_{t=1}^k \sum_{s=1}^t a_{s,t} s = \ell c \mathcal{E}_n \sum_{t=1}^k (1-d)^{t-1} t \leq \ell c \mathcal{E}_n \sum_{t=1}^{\infty} (1-d)^{t-1} t \leq \frac{\ell c}{d^2} \cdot \mathcal{E}_n,$$

where the last equality comes from (5.4). Taking the supremum over k completes the proof. \square

The proof of Theorem 3.1 is based on the estimation in Theorem 5.1 of the expected norms

$$\mathbb{E}_n \left[\left\| (C - \tilde{M}) \check{X} \right\|_{\infty} \right] \quad \text{and} \quad \mathbb{E}_n \left[\left\| CX - \tilde{M} \check{X} \right\|_{\infty} \right] \quad (5.10)$$

for X conditionally independent of C given $\mathcal{F}_n = \sigma(\mathbf{Y}_i : i \in V_n)$ and having mean $\check{X} = \mathbb{E}_n[X]$, when $\theta_n \geq \gamma \log n$, and of the expected norms

$$\max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s X)_{i \bullet} - (\tilde{M}^s \check{X})_{i \bullet} \right\|_1 \right] \quad (5.11)$$

for $s \geq 1$, when $\theta_n / \log n \rightarrow 0$ as $n \rightarrow \infty$. This is done in the next two sections.

5.1. The semi-sparse to dense regimes. When the edge density in the graph is sufficiently large, the concentration of the matrix C around the matrix \tilde{M} is very strong. It is worth pointing out that \tilde{M} is not the mean of C , but rather, its more tractable asymptotic mean. For the norm involving $R^{(0)}$, the concentration occurs around its mean matrix \tilde{R} , and the strength of this concentration is determined by the matrix \tilde{M} . The main result in this section provides estimates for the two expected norms in (5.10). Its proof relies on the following concentration inequality for Binomial or Poisson random sums, and its consequence on ratios of random sums.

Proposition 5.1. *Let $\{Y_i^{(r)} : i \geq 1, 1 \leq r \leq K\}$ be independent random variables on $[-H, H]$, with $\{Y_i^{(r)} : i \geq 1\}$ i.i.d. for each $1 \leq r \leq K$, and $E[(Y_1^{(r)})^2] \leq v_r$. Let $\{N_r : 1 \leq r \leq K\}$ be independent random variables on \mathbb{N} , independent of $\{Y_i^{(r)} : i \geq 1, 1 \leq r \leq K\}$, and satisfying $E[e^{sN_r}] \leq e^{E[N_r](e^s - 1)}$ for $s \in \mathbb{R}$. Define $S_r = \sum_{i=1}^{N_r} Y_i^{(r)}$. Then, for any $\epsilon > 0$,*

$$P \left(\sum_{r=1}^K (S_r - E[S_r]) > \epsilon \mu \right) \leq \exp \left(-\frac{(\epsilon \mu)^2}{2\nu} + H \frac{(\epsilon \mu)^3}{2\nu^2} \right),$$

where $\mu = \sum_{r=1}^K E[N_r] E[|Y_1^{(r)}|]$ and $\nu = \sum_{r=1}^K E[N_r] v_r$.

Proof. Start by conditioning on $\{N_r : 1 \leq r \leq K\}$ and use Chernoff's bound to obtain that for any $\theta > 0$,

$$\begin{aligned} P \left(\sum_{r=1}^K (S_r - E[S_r]) > \epsilon \mu \right) &= E \left[P \left(\sum_{r=1}^K (S_r - E[S_r]) > \epsilon \mu \mid N_1, \dots, N_K \right) \right] \\ &\leq E \left[e^{-\theta \epsilon \mu} \prod_{r=1}^K E \left[e^{\theta (S_r - E[S_r])} \mid N_r \right] \right] = e^{-\theta \epsilon \mu} \prod_{r=1}^K e^{-\theta E[S_r]} E \left[E \left[e^{\theta Y_1^{(r)}} \right]^{N_r} \right]. \end{aligned}$$

Using the third order Taylor approximation for $e^{\theta x}$, we obtain

$$\begin{aligned} E[e^{\theta Y_1^{(r)}}] &\leq 1 + \theta E[Y_1^{(r)}] + \sum_{k=2}^{\infty} \frac{\theta^k E[|Y_1^{(r)}|^k]}{k!} \leq 1 + \theta E[Y_1^{(r)}] + \left(\sum_{k=2}^{\infty} \frac{\theta^k H^{k-2}}{k!} \right) E[(Y_1^{(r)})^2] \\ &\leq 1 + \theta E[Y_1^{(r)}] + \frac{\nu_r}{H^2} (e^{\theta H} - \theta H - 1) =: 1 + \lambda_r(\theta). \end{aligned}$$

It follows from the assumption on the moment generating function of N_r and the observation $E[S_r] = E[N_r]E[Y_1^{(r)}]$, that

$$\begin{aligned} P\left(\sum_{r=1}^K (S_r - E[S_r]) > \epsilon\mu\right) &\leq e^{-\theta\epsilon\mu} \prod_{r=1}^K e^{-\theta E[S_r]} E[(1 + \lambda_r(\theta))^{N_r}] = e^{-\theta\epsilon\mu} \prod_{r=1}^K e^{-\theta E[S_r]} E\left[e^{N_r \log(1 + \lambda_r(\theta))}\right] \\ &\leq e^{-\theta\epsilon\mu} \prod_{r=1}^K e^{-\theta E[S_r] + E[N_r](e^{\log(1 + \lambda_r(\theta))} - 1)} = e^{-\theta\epsilon\mu + \sum_{r=1}^K E[N_r](\lambda_r(\theta) - \theta E[Y_1^{(r)}])} \\ &= e^{-\theta\epsilon\mu + \frac{\nu}{H^2}(e^{\theta H} - \theta H - 1)}. \end{aligned}$$

Picking $\theta = \frac{1}{H} \log(1 + H\epsilon\mu/\nu) \geq \epsilon\mu/\nu - (H\epsilon^2\mu^2)/(2\nu^2)$, we get

$$P\left(\sum_{r=1}^K (S_r - E[S_r]) > \epsilon\mu\right) \leq e^{-\theta\epsilon\mu + \frac{\nu}{H^2}(\epsilon\mu/\nu - \theta)} \leq e^{-\left(\frac{\epsilon\mu}{\nu} - \frac{H\epsilon^2\mu^2}{2\nu^2}\right)(\epsilon\mu + \nu/H) + \frac{\epsilon\mu}{H}} = e^{-\frac{\epsilon^2\mu^2}{2\nu} + \frac{H\epsilon^3\mu^3}{2\nu^2}}.$$

□

Next, we use Proposition 5.1 to bound the event that ratios of random sums are far from the ratios of their means.

Lemma 5.1. *Let $\{B_i^{(r)}, X_i^{(r)} : i \geq 1, 1 \leq r \leq K\}$ be independent random variables, with $B_i^{(r)} \in [0, H]$, $X_i^{(r)} \in [-1, 1]$, and such that $\{(B_i^{(r)}, X_i^{(r)}) : i \geq 1\}$ are i.i.d. for each $1 \leq r \leq K$. Let $\{N_r : 1 \leq r \leq K\}$ be independent random variables on \mathbb{N} , independent of the $\{(B_i^{(r)}, X_i^{(r)}) : i \geq 1, 1 \leq r \leq K\}$, and satisfying $E[e^{sN_r}] \leq e^{E[N_r](e^s - 1)}$ for $s \in \mathbb{R}$. Define $S_r = \sum_{i=1}^{N_r} B_i^{(r)}$ and $\tilde{S}_r = \sum_{i=1}^{N_r} X_i^{(r)} B_i^{(r)}$. Then, for any $\epsilon > 0$,*

$$P\left(\left|\frac{\tilde{S}_1 + \dots + \tilde{S}_K}{S_1 + \dots + S_K} - \frac{E[\tilde{S}_1] + \dots + E[\tilde{S}_K]}{E[S_1] + \dots + E[S_K]}\right| > \epsilon\right) \leq 4 \exp\left(-\frac{(\epsilon/2)^2 \mu^2}{2\nu} + \frac{H(\epsilon/2)^3 \mu^3}{2\nu^2}\right),$$

where $\mu = \sum_{r=1}^K E[N_r]E[B_1^{(r)}]$ and $\nu = \sum_{r=1}^K E[N_r]E[(B_1^{(r)})^2]$.

Proof. To start, note that since $|\tilde{S}_r| \leq S_r$ for each $r \in \{1, \dots, K\}$, then

$$\begin{aligned} &\left|\frac{\tilde{S}_1 + \dots + \tilde{S}_K}{S_1 + \dots + S_K} - \frac{E[\tilde{S}_1] + \dots + E[\tilde{S}_K]}{E[S_1] + \dots + E[S_K]}\right| \\ &= \frac{1}{(S_1 + \dots + S_K)E[S_1 + \dots + S_K]} \left|\sum_{r=1}^K \sum_{t=1}^K E[S_r] \tilde{S}_t - \sum_{r=1}^K \sum_{t=1}^K S_r E[\tilde{S}_t]\right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{(S_1 + \dots + S_K)\mu} \left(\left| \sum_{r=1}^K \sum_{t=1}^K E[S_r] \tilde{S}_t - \sum_{r=1}^K \sum_{t=1}^K S_r \tilde{S}_t \right| + \left| \sum_{r=1}^K \sum_{t=1}^K S_r \tilde{S}_t - \sum_{r=1}^K \sum_{t=1}^K S_r E[\tilde{S}_t] \right| \right) \\
&= \frac{1}{(S_1 + \dots + S_K)\mu} \left(\left| \sum_{t=1}^K \tilde{S}_t \left| \sum_{r=1}^K (E[S_r] - S_r) \right| + \sum_{r=1}^K S_r \left| \sum_{t=1}^K (\tilde{S}_t - E[\tilde{S}_t]) \right| \right| \right) \\
&\leq \frac{1}{\mu} \left(\left| \sum_{r=1}^K (E[S_r] - S_r) \right| + \left| \sum_{t=1}^K (\tilde{S}_t - E[\tilde{S}_t]) \right| \right).
\end{aligned}$$

Next, define the events

$$F = \left\{ \left| \sum_{r=1}^K (S_r - E[S_r]) \right| > \epsilon\mu/2 \right\} \quad \text{and} \quad G = \left\{ \left| \sum_{r=1}^K (\tilde{S}_r - E[\tilde{S}_r]) \right| > \epsilon\mu/2 \right\},$$

and use the union bound to obtain that

$$\begin{aligned}
P \left(\left| \frac{\tilde{S}_1 + \dots + \tilde{S}_K}{S_1 + \dots + S_K} - \frac{E[\tilde{S}_1] + \dots + E[\tilde{S}_K]}{E[S_1] + \dots + E[S_K]} \right| > \epsilon \right) &\leq P \left(\left| \sum_{r=1}^K (E[S_r] - S_r) \right| + \left| \sum_{t=1}^K (\tilde{S}_t - E[\tilde{S}_t]) \right| > \epsilon\mu \right) \\
&\leq P(F^c) + P(G^c).
\end{aligned}$$

Finally, let $\tilde{\mu} = \sum_{r=1}^K E[N_r] E[|X_1^{(r)}|] E[B_1^{(r)}]$, and use Proposition 5.1 with $Y_i^{(r)} = B_i^{(r)}$ and $Y_i^{(r)} = -B_i^{(r)}$ to obtain that

$$P(F^c) \leq 2 \exp \left(-\frac{(\epsilon/2)^2 \mu^2}{2\nu} + \frac{H(\epsilon/2)^3 \mu^3}{2\nu^2} \right),$$

and then again with $Y_i^{(r)} = -X_i^{(r)} B_i^{(r)}$ and $Y_i^{(r)} = X_i^{(r)} B_i^{(r)}$ to get

$$P(G^c) \leq 2 \exp \left(-\frac{(\epsilon\mu/(2\tilde{\mu}))^2 \tilde{\mu}^2}{2\nu} + \frac{H(\epsilon\mu/(2\tilde{\mu}))^3 \tilde{\mu}^3}{2\nu^2} \right) = 2 \exp \left(-\frac{(\epsilon/2)^2 \mu^2}{2\nu} + \frac{H(\epsilon/2)^3 \mu^3}{2\nu^2} \right).$$

This completes the proof. \square

Now, we are able to prove the main result of this section. Recall that $\beta_{r,s} = E[B_{ij} | J_i = r, J_j = s]$, $v_{r,s} = E[B_{ij}^2 | J_i = r, J_j = s]$ for $1 \leq r, s \leq K$,

$$\mu_r^{(n)} = \sum_{s=1}^K \beta_{r,s} \pi_s^{(n)} \kappa(s, r) \quad \text{and} \quad \nu_r^{(n)} = \sum_{s=1}^K v_{r,s} \pi_s^{(n)} \kappa(s, r),$$

for $1 \leq r \leq K$ and large enough n , $\Delta_n = \max_{r \in \mathcal{I}} \nu_r^{(n)} / (\mu_r^{(n)})^2$ and $\Lambda_n = \max_{r \in \mathcal{I}} \mu_r^{(n)} / \nu_r^{(n)}$, where \mathcal{I} is the set of nonzero rows of M .

Theorem 5.3. *Let $X \in \mathbb{R}^{n \times \ell}$ be a random matrix conditionally independent of C given \mathcal{F}_n and having mean $\tilde{X} = \mathbb{E}_n[X]$. Furthermore, suppose X has conditionally independent rows $\{\mathbf{X}_i : i \in V_n\}$ given \mathcal{F}_n*

and satisfies $\|\mathbf{X}_i\|_\infty \leq 1$ a.s., and for each $1 \leq r \leq K$, its rows $\{\mathbf{X}_i : J_i = r\}$ are identically distributed. Then, for $\theta_n \geq (2H\Lambda_n/\gamma)^2 \Delta_n \log n$ and $0 < \gamma < 1/2$, there exists a constant $\Gamma' < \infty$ such that

$$\begin{aligned} \mathbb{E}_n \left[\left\| CX - \tilde{M}\check{X} \right\|_\infty \right] &\leq \Gamma' \sqrt{\frac{\log n}{\theta_n}}, \\ \mathbb{E}_n \left[\left\| (C - \tilde{M})\check{X} \right\|_\infty \right] &\leq \Gamma' \sqrt{\frac{\log n}{\theta_n}}. \end{aligned}$$

Proof. To start, note that since C and \tilde{M} have the same deterministically zero rows (if any), we can simply ignore them throughout the proof. Now let \mathcal{I}_* denote the set of non-zero rows of \tilde{M} , fix $\epsilon > 0$, and define the events

$$F_{it} = \left\{ \left| (CX)_{it} - (\tilde{M}\check{X})_{it} \right| \leq 2\epsilon \right\}, \quad i \in \mathcal{I}_*, 1 \leq t \leq \ell.$$

Next, note that

$$\begin{aligned} \mathbb{E}_n \left[\left\| CX - \tilde{M}\check{X} \right\|_\infty \right] &= \mathbb{E}_n \left[\max_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} \left| (CX)_{it} - (\tilde{M}\check{X})_{it} \right| \right] \\ &\leq \mathbb{E}_n \left[\max_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} \left| (CX)_{it} - (\tilde{M}\check{X})_{it} \right| 1(F_{it}) \right] \\ &\quad + \mathbb{E}_n \left[\max_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} \left(|(CX)_{it}| + |(\tilde{M}\check{X})_{it}| \right) 1(F_{it}^c) \right] \\ &\leq \mathbb{E}_n \left[\max_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} 2\epsilon 1(F_{it}) \right] + \mathbb{E}_n \left[\max_{i \in \mathcal{I}_*} 2 \sum_{t=1}^{\ell} 1(F_{it}^c) \right] \\ &\leq 2\ell\epsilon + 2 \sum_{i \in \mathcal{I}_*} \mathbb{E}_n \left[\sum_{t=1}^{\ell} 1(F_{it}^c) \right], \end{aligned}$$

where in the last step we used the observation that

$$\sum_{t=1}^{\ell} |(CX)_{it}| 1(F_{it}^c) \leq \sum_{t=1}^{\ell} \sum_{j=1}^n C_{ij} |X_{jt}| 1(F_{it}^c) \leq \sum_{t=1}^{\ell} \|C_{i\bullet}\|_1 1(F_{it}^c) \leq \sum_{t=1}^{\ell} 1(F_{it}^c),$$

and similarly, $\sum_{t=1}^{\ell} |(\tilde{M}\check{X})_{it}| 1(F_{it}^c) \leq \sum_{t=1}^{\ell} 1(F_{it}^c)$. It follows that

$$\mathbb{E}_n \left[\left\| CX - \tilde{M}\check{X} \right\|_\infty \right] \leq 2\ell\epsilon + 2 \sum_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} \mathbb{P}_n(F_{it}^c).$$

To bound $\mathbb{P}_n(F_{it}^c)$, start by noting that

$$\left| (CX)_{it} - (\tilde{M}\check{X})_{it} \right| = \left| \sum_{j=1}^n C_{ij} X_{jt} - \sum_{j=1}^n \tilde{M}_{ij} \check{X}_{jt} \right| = \left| \frac{\sum_{j=1}^n X_{jt} B_{ij} 1(j \rightarrow i)}{\sum_{r=1}^n B_{ir} 1(r \rightarrow i)} - \frac{\sum_{j=1}^n \check{X}_{jt} \beta_{J_i, J_j} p_{ji}^{(n)}}{\sum_{r=1}^n \beta_{J_i, J_r} p_{ri}^{(n)}} \right|.$$

Next, note that we can rewrite

$$\sum_{j=1}^n X_{jt} B_{ij} 1(j \rightarrow i) = \sum_{r=1}^K \sum_{j=1}^n X_{jt} B_{ij} 1(j \rightarrow i) 1(J_j = r) =: \sum_{r=1}^K S_{i,r}^{(t)}$$

and

$$\sum_{j=1}^n B_{ij} 1(j \rightarrow i) = \sum_{r=1}^K \sum_{j=1}^n B_{ij} 1(j \rightarrow i) 1(J_j = r) =: \sum_{r=1}^K S_{i,r}.$$

Moreover, conditionally on J_i , we have that $S_{i,r}^{(t)} \stackrel{\mathcal{D}}{=} \sum_{j=1}^{N_{i,r}} X_j^{(t,r)} B_j^{(i,r)}$ and $S_{i,r} \stackrel{\mathcal{D}}{=} \sum_{j=1}^{N_{i,r}} B_j^{(i,r)}$, with $N_{i,r}$ a binomial with parameters $(n\pi_r^{(n)}, \kappa(r, J_i)\theta_n/n)$, independent of the i.i.d. sequences $\{B_j^{(i,r)} : j \geq 1\}$ and $\{X_j^{(t,r)} : j \geq 1\}$, which satisfy $B_j^{(i,r)} \stackrel{\mathcal{D}}{=} (B_{ij}|J_i, J_j = r)$, $X_j^{(t,r)} \stackrel{\mathcal{D}}{=} (X_{jt}|J_j = r)$, and $\bar{x}_j = \check{X}_{jt}$ for $J_j = r$. It follows from Lemma 5.1 that

$$\begin{aligned} \mathbb{P}_n(F_{it}^c) &= P \left(\left| \frac{S_{i,1}^{(t)} + \dots + S_{i,K}^{(t)}}{S_{i,1} + \dots + S_{i,K}} - \frac{E[S_{i,1}^{(t)}] + \dots + E[S_{i,K}^{(t)}]}{E[S_{i,1}] + \dots + E[S_{i,K}]} \right| > 2\epsilon \right) \\ &\leq 4 \exp \left(-\frac{\epsilon^2 (\theta_n \mu_{J_i}^{(n)})^2}{2\theta_n \nu_{J_i}^{(n)}} + \frac{H\epsilon^3 (\theta_n \mu_{J_i}^{(n)})^3}{2(\theta_n \nu_{J_i}^{(n)})^2} \right) = 4 \exp \left(-\frac{\epsilon^2 \theta_n (\mu_{J_i}^{(n)})^2}{2\nu_{J_i}^{(n)}} \left(1 - \frac{H\epsilon \mu_{J_i}^{(n)}}{\nu_{J_i}^{(n)}} \right) \right) \end{aligned}$$

where

$$\begin{aligned} \mu_{J_i}^{(n)} &= \sum_{r=1}^K \pi_r^{(n)} \kappa(r, J_i) \beta_{J_i, r} = \theta_n^{-1} \sum_{r=1}^K E[N_{i,r}] E[B_1^{(i,r)}] \\ \nu_{J_i}^{(n)} &= \sum_{r=1}^K \pi_r^{(n)} \kappa(r, J_i) \nu_{J_i, r} = \theta_n^{-1} \sum_{r=1}^K E[N_{i,r}] E[(B_1^{(r)})^2]. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}_n \left[\left\| CX - \tilde{M}\check{X} \right\|_{\infty} \right] &\leq 2\ell\epsilon + 8 \sum_{i \in \mathcal{I}_*} \sum_{t=1}^{\ell} \exp \left(-\frac{\epsilon^2 \theta_n (\mu_{J_i}^{(n)})^2}{2\nu_{J_i}^{(n)}} \left(1 - \frac{H\epsilon \mu_{J_i}^{(n)}}{\nu_{J_i}^{(n)}} \right) \right) \\ &= 2\ell\epsilon + 8\ell \sum_{r \in \mathcal{I}} n\pi_r^{(n)} \exp \left(-\frac{\epsilon^2 \theta_n (\mu_r^{(n)})^2}{2\nu_r^{(n)}} \left(1 - \frac{H\epsilon \mu_r^{(n)}}{\nu_r^{(n)}} \right) \right) \\ &\leq 2\ell\epsilon + 8\ell \sum_{r \in \mathcal{I}} n\pi_r^{(n)} \exp \left(-\frac{\epsilon^2 \theta_n (\mu_r^{(n)})^2}{2\nu_r^{(n)}} (1 - H\epsilon \Lambda_n) \right) \\ &\leq 2\ell\epsilon + 8\ell n \exp \left(-\frac{\epsilon^2 \theta_n}{2\Delta_n} (1 - H\epsilon \Lambda_n) \right), \end{aligned}$$

where $\Delta_n = \max_{r \in \mathcal{I}} \nu_r^{(n)} / (\mu_r^{(n)})^2$ and $\Lambda_n = \max_{r \in \mathcal{I}} \mu_r^{(n)} / \nu_r^{(n)}$. Next, choose $\epsilon = \sqrt{4\Delta_n \log n / \theta_n}$ to obtain

$$\begin{aligned} n \exp\left(-\frac{\epsilon^2 \theta_n}{2\Delta_n} (1 - H\epsilon\Lambda_n)\right) &= n \exp\left(-2 \log n \left(1 - \frac{2H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right) \\ &= \exp\left(-\log n \left(1 - \frac{4H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right), \end{aligned}$$

which in turn yields

$$\mathbb{E}_n \left[\left\| CX - \tilde{M}\check{X} \right\|_\infty \right] \leq 4\ell \sqrt{\Delta_n \log n / \theta_n} + 8\ell \exp\left(-\log n \left(1 - \frac{4H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right).$$

Since the second norm corresponds to the case when X is deterministic, the same argument also yields

$$\mathbb{E}_n \left[\left\| (C - \tilde{M})\check{X} \right\|_\infty \right] \leq 4\ell \sqrt{\Delta_n \log n / \theta_n} + 8\ell \exp\left(-\log n \left(1 - \frac{4H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right).$$

To complete the proof, note that if $(2H\Lambda_n/\gamma)^2 \Delta_n \log n \leq \theta_n \leq (\log n)^3$, then

$$\begin{aligned} &\sqrt{\Delta_n \log n / \theta_n} + \exp\left(-\log n \left(1 - \frac{4H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right) \\ &\leq \Delta_n^{1/2} \sqrt{\frac{\log n}{\theta_n}} + \exp(-(1-2\gamma)\log n) = \Delta_n^{1/2} \sqrt{\frac{\log n}{\theta_n}} + n^{-1+2\gamma} = O\left(\sqrt{\frac{\log n}{\theta_n}}\right) \end{aligned}$$

as $n \rightarrow \infty$, while for $(\log n)^3 < \theta_n \leq n/\kappa_*$, with $\kappa_* = \max_{1 \leq r, s \leq K} \kappa(r, s)$,

$$\begin{aligned} &\sqrt{\Delta_n \log n / \theta_n} + \exp\left(-\log n \left(1 - \frac{4H\Lambda_n \sqrt{\Delta_n \log n}}{\theta_n^{1/2}}\right)\right) \\ &\leq \Delta_n^{1/2} \sqrt{\frac{\log n}{\theta_n}} + n^{-1} e^{4H\Lambda_n \sqrt{\Delta_n}} = O\left(\sqrt{\frac{\log n}{\theta_n}}\right) \end{aligned}$$

as $n \rightarrow \infty$. The statement of the theorem now follows. \square

We can now give the proof to the first part of Theorem 3.1.

Proof of Theorem 3.1. Suppose $\theta_n \geq (6H\Lambda_n)^2 \Delta_n \log n$. We will prove that

$$\sup_{k \geq 0} \mathbb{E}_n \left[\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right] = O\left(\sqrt{\frac{\log n}{\theta_n}} + \mathcal{E}_n\right)$$

as $n \rightarrow \infty$, where $\mathcal{E}_n = \max_{1 \leq r, s \leq K} \left| \frac{\pi_s^{(n)} \pi_r - \pi_s \pi_r^{(n)}}{\pi_r^{(n)} \pi_s} \right|$. To start, use the triangle inequality to obtain

$$\sup_{k \geq 0} \mathbb{E}_n \left[\left\| R^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right] \leq \sup_{k \geq 0} \mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_\infty \right] + \mathbb{E}_n \left[\sup_{k \geq 0} \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right].$$

Theorem 5.2 yields

$$\sup_{k \geq 0} \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} \leq \frac{c\ell}{d^2} \mathcal{E}_n.$$

For the remaining term, use Theorem 5.1 to obtain that

$$\begin{aligned} \mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right] &\leq \max_{r \geq 1} h(\tilde{M}^r \check{W}) \sum_{t=0}^{k-1} (1-d)^{t-1} t + \mathbb{E}_n \left[\left\| CW^{(0)} - \tilde{M} \check{W} \right\|_{\infty} \right] \sum_{t=0}^{k-1} (1-d)^t \\ &\quad + \max_{r \geq 1} h(\tilde{M}^r \check{R}) k(1-d)^{k-1} + \mathbb{E}_n \left[\left\| CR^{(0)} - \tilde{M} \check{R} \right\|_{\infty} \right] (1-d)^k \end{aligned}$$

where $h(X) = \mathbb{E}_n \left[\left\| (C - \tilde{M})X \right\|_{\infty} \right]$. Now use Theorem 5.3 to obtain that, since $\left\| \tilde{M}^r \check{W} \right\|_{\infty} \leq \ell$ and $\left\| \tilde{M}^r \check{R} \right\|_{\infty} \leq \ell$ for all $r \geq 0$, then

$$\max \left\{ \max_{r \geq 1} h(\tilde{M}^r \check{W}), \max_{r \geq 1} h(\tilde{M}^r \check{R}), \mathbb{E}_n \left[\left\| CW^{(0)} - \tilde{M} \check{W} \right\|_{\infty} \right], \mathbb{E}_n \left[\left\| CR^{(0)} - \tilde{M} \check{R} \right\|_{\infty} \right] \right\} \leq \ell \Gamma' \sqrt{\frac{\log n}{\theta_n}}.$$

Therefore,

$$\mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right] \leq \ell \Gamma' \sqrt{\frac{\log n}{\theta_n}} \left(\sum_{t=0}^{k-1} (1-d)^{t-1} t + \sum_{t=0}^{k-1} (1-d)^t + k(1-d)^{k-1} + (1-d)^k \right),$$

and since the supremum over k is finite, there exists some other constant $\Gamma < \infty$ such that

$$\sup_{k \geq 0} \left(\mathbb{E}_n \left[\left\| R^{(k)} - \tilde{R}^{(k)} \right\|_{\infty} \right] + \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_{\infty} \right) \leq \Gamma \left(\sqrt{\frac{\log n}{\theta_n}} + \mathcal{E}_n \right).$$

This completes the proof for the first part of the theorem. \square

We now move to the density regime where $\log n / \theta_n$ may not converge to zero. In that case, we can still prove concentration of the opinion matrix $R^{(k)}$ to the mean-field limit $\mathcal{R}^{(k)}$ as $n \rightarrow \infty$, but in a weaker mode of convergence.

5.2. The subcritical to sparse regimes. In this section we complete the proof of Theorem 3.1 by analyzing the expected norms (5.11). When $\limsup_{n \rightarrow \infty} (\log n) / \theta_n > 0$, the number of neighbors is not enough to provide a full concentration of the matrix C around its asymptotic mean \tilde{M} , since up to an $o(n)$ number of vertices will not be sufficiently close to their limits. Instead of using the supremum norm as in the denser case, we use an averaged l_1 -norm, which is what we typically use in sparse graphs where the local limiting structure of the dSBM is what enables the analysis. The key observation in this section is that for any matrix $X \in \mathbb{R}^{n \times \ell}$, $(C^s X)_{ij}$ is a function of the inbound neighborhood of vertex $i \in V_n$ of length s , which is approximately equal to $(\tilde{M}^s X)_{ij}$. In order for us to make this connection, we need to describe the local behavior of neighborhoods in the marked graph $G_n = (V_n, E_n; \mathcal{A}_n)$, where $\mathcal{A}_n = \{\mathbf{Y}_i : i \in V_n\}$ are the extended vertex attributes.

5.2.1. *Coupling with a multi-type Galton-Watson tree.* Pick a vertex I_n , chosen uniformly at random from the vertex set V_n . The idea behind local weak convergence is to couple the exploration of the inbound neighborhood of I_n with the construction of a rooted graph, whose root \emptyset corresponds to I_n . We point out that in our setting, the local weak limit of the dSBM is a non locally-finite graph, so the coupling we need is one between the exploration of the inbound neighborhood of I_n with a branching tree whose degree distribution grows with n . We start by introducing some terminology for trees.

We enumerate nodes in a rooted tree using the Ulam-Harris notation, which contains the ancestry path of any node all the way to the root. For that, let any node in generation $k \geq 1$ of the tree have a label of the form $\mathbf{i} = (i_1, \dots, i_k) \in \mathbb{N}_+^k$, and define the concatenation operation $(\mathbf{i}, j) = (i_1, \dots, i_k, j)$ for a node in generation $k + 1$. The root is always denoted by \emptyset , but is omitted from the label of any other node; to simplify the notation, we omit the parentheses for nodes in the first generation. We denote by $\mathcal{U} = \{\emptyset\} \cup \bigcup_{k=1}^{\infty} \mathbb{N}_+^k$ the set of all possible node labels. The generation of a node $\mathbf{i} \in \mathcal{U}$ is denoted by $|\mathbf{i}|$, for $\mathbf{i} \neq \emptyset$.

The local weak limit of a sparse dSBM with K communities is a marked K -type Galton-Watson tree. Recall that our network is not a sparse graph, so instead we are using an *intermediate* coupling (see [40] for a detailed description of the construction). In order to obtain a description of the coupled tree, we start by reminding the reader that vertices in G_n have extended attributes:

$$\mathbf{Y}_i = (J_i, \mathbf{Q}_i, \mathbf{B}_i) \in \{1, \dots, K\} \times [-1, 1]^\ell \times [0, H]^n, \quad (5.12)$$

and $\mathcal{F}_n = \sigma(\mathbf{Y}_i : i \in V_n)$. In the K -type Galton-Watson tree, each node \mathbf{i} will have an *attribute*:

$$\hat{\mathbf{A}}_{\mathbf{i}} = (\hat{J}_{\mathbf{i}}, \hat{\mathbf{Q}}_{\mathbf{i}}) \in \{1, \dots, K\} \times [-1, 1]^\ell,$$

and a *full mark* of the form:

$$\hat{\mathbf{X}}_{\mathbf{i}} = (\hat{J}_{\mathbf{i}}, \hat{\mathbf{Q}}_{\mathbf{i}}, \{\hat{N}_{\mathbf{i}}^{(r)}\}_{r=1}^K, \{\hat{B}_{(\mathbf{i},s)}\}_{s=1}^{\hat{N}_{\mathbf{i}}}) \in \{1, \dots, K\} \times [-1, 1]^\ell \times \mathbb{N}^K \times [0, H]^\infty,$$

where $\hat{J}_{\mathbf{i}}$ is the type of node \mathbf{i} , $\hat{N}_{\mathbf{i}}^{(r)}$ is the number of type r offspring that node \mathbf{i} has, and $\hat{N}_{\mathbf{i}} = \hat{N}_{\mathbf{i}}^{(1)} + \dots + \hat{N}_{\mathbf{i}}^{(K)}$.

We will construct the tree conditionally on the community labels $\mathcal{J}_n = \{J_i : i \in V_n\}$. The unmarked tree will be denoted $\mathcal{T} = \bigcup_{k=0}^{\infty} \mathcal{A}_k \subset \mathcal{U}$, where \mathcal{A}_k is the set of nodes in its k th generation, and each node $\mathbf{i} \in \mathcal{T}$ will have a type $\hat{J}_{\mathbf{i}}$. To start, set $\mathcal{A}_0 = \{\emptyset\}$ and choose the type of the root \emptyset according to:

$$\mathbb{P}_n(\hat{J}_{\emptyset} = r) = \frac{1}{n} \sum_{i=1}^n 1(J_i = r) = \pi_r^{(n)}.$$

From here on, any node $\mathbf{i} \in \mathcal{A}_k$, will have an offspring vector $(\hat{N}_{\mathbf{i}}^{(1)}, \dots, \hat{N}_{\mathbf{i}}^{(K)})$, conditionally independent of everything else given its type, and having distribution:

$$\mathbb{P}_n(\hat{N}_{\mathbf{i}}^{(1)} = n_1, \dots, \hat{N}_{\mathbf{i}}^{(K)} = n_K \mid \hat{J}_{\mathbf{i}} = r) = \prod_{s=1}^K e^{-q_{sr}^{(n)}} \frac{(q_{sr}^{(n)})^{n_s}}{n_s!},$$

where $q_{sr}^{(n)} = \kappa(s, r) \pi_s^{(n)} \theta_n$. In other words, given $\hat{J}_{\mathbf{i}} = r$, the numbers of offspring of each type s of a node of type r are independent Poisson random variables with means $q_{sr}^{(n)}$, $1 \leq s \leq K$. To prevent the labels

from conveying any information, we usually permute the \hat{N}_i offspring uniformly at random, and then set

$$\mathcal{A}_{k+1} = \left\{ (\mathbf{i}, j) : \mathbf{i} \in \mathcal{A}_k, 1 \leq j \leq \hat{N}_i \right\}.$$

Note that the conditional distribution of $\hat{J}_{(\mathbf{i},j)}$ for $|\mathbf{i}| = k$, $k \geq 0$, depends on \hat{J}_i and is given by

$$\mathbb{P}_n \left(\hat{J}_{(\mathbf{i},j)} = s \mid \hat{J}_i = r \right) = \mathbb{E}_n \left[\frac{\hat{N}_i^{(s)} 1(\hat{N}_i > 0)}{\hat{N}_i} \mid \hat{J}_i = r \right] = \frac{q_{sr}^{(n)}}{\sum_{t=1}^K q_{tr}^{(n)}} = \frac{\kappa(s, r) \pi_s^{(n)}}{\sum_{t=1}^K \kappa(t, r) \pi_t^{(n)}}.$$

Moreover, conditionally on \hat{J}_i , the $\{\hat{J}_{(\mathbf{i},j)} : j \geq 1\}$ are independent of each other.

Now that the unmarked tree $\mathcal{T} = \bigcup_{k=0}^{\infty} \mathcal{A}_k$ has been constructed, we assign to each node $\mathbf{i} \in \mathcal{T}$ a mark $(\hat{Q}_i, \{\hat{B}_{(\mathbf{i},s)}\}_{s=1}^{\hat{N}_i})$. The internal belief vector \hat{Q}_i depends only on \hat{J}_i and is distributed according to $F_{\hat{J}_i}$. Finally, the unnormalized weights satisfy for $x \in \mathbb{R}$,

$$\mathbb{P}_n \left(\hat{B}_{(\mathbf{i},j)} \leq x \mid \hat{J}_i, \hat{J}_{(\mathbf{i},j)} \right) = G_{\hat{J}_i, \hat{J}_{(\mathbf{i},j)}}(x), \quad (5.13)$$

independently of everything else.

Definition 5.1. Two simple vertex-marked directed graphs, $G = (V, E; \mathcal{A})$ and $G' = (V', E'; \mathcal{A}')$, whose vertices have marks \mathbf{a}_i , $i \in V$, and \mathbf{a}'_i , $i \in V'$, respectively, are said to be *isomorphic* if there exists a bijection $\vartheta : V \rightarrow V'$ such that $(i, j) \in E$ if, and only if, $(\vartheta(i), \vartheta(j)) \in E'$, for nodes $i, j \in V$, and for all $i \in V$ the vertex mark $\mathbf{a}_i = \mathbf{a}'_{\vartheta(i)}$. We write $G \simeq G'$ to express that the graphs G and G' are isomorphic.

For each $i \in V_n$, the intermediate coupling will construct a marked tree $\mathcal{T}_{\emptyset(i)}(\hat{J}) \stackrel{\mathcal{D}}{=} (\mathcal{T}(\hat{J}) | \hat{J}_{\emptyset} = J_i)$ using as marks in Definition 5.1 the types, which correspond to the community labels in the dSBM (we do not need the internal opinions nor the unnormalized weights for this purpose, since their distribution is determined by the types). Note that the subscript $\emptyset(i)$ emphasizes that the root of the tree corresponds to vertex i in the graph. For any fixed $s \geq 0$, let $\mathcal{G}_i^{(s)}(J)$ denote the induced subgraph of G_n determined by the vertices in V_n that have directed paths of length at most s connecting them to vertex i , including their community labels, and let $\mathcal{T}_{\emptyset(i)}^{(s)}(\hat{J})$ denote the restriction of $\mathcal{T}_{\emptyset(i)}(\hat{J})$ to its first s generations. Then, using the same type of intermediate coupling as in [35], one obtains the following theorem.

Theorem 5.4. *For the dSBM under Assumption A, with $\theta_n \leq (\log n)^b$ for some $b > 0$, and any fixed $k \geq 0$, there exists a sequence of marked K -type Galton-Watson trees $\{\mathcal{T}_{\emptyset(i)}(\hat{J}) : i \in V_n\}$ such that $\mathcal{T}_{\emptyset(i)}(\hat{J}) \stackrel{\mathcal{D}}{=} (\mathcal{T}(\hat{J}) | \hat{J}_{\emptyset} = J_i)$ and*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{P}_n \left(\mathcal{G}_i^{(k)}(J) \neq \mathcal{T}_{\emptyset(i)}^{(k)}(\hat{J}) \right) \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

Note that when the coupling holds for a vertex i up to generation s , i.e., $\mathcal{G}_i^{(s)}(J) \simeq \mathcal{T}_{\emptyset(i)}^{(s)}(\hat{J})$, we can copy the internal belief vectors and the unnormalized weights from the graph, so $\hat{Q}_i = \mathbf{Q}_{\vartheta(i)}$ and $\hat{B}_{(\mathbf{i},j)} = B_{\vartheta(\mathbf{i}), \vartheta((\mathbf{i},j))}$, where ϑ is the bijection defining the isomorphism. When the coupling fails, we simply sample them

independently according to $F_{\hat{J}_i}$ and (5.13), respectively. Once we have the unnormalized weights, we define

$$\hat{C}_{(i,j)} = \frac{\hat{B}_{(i,j)}}{\sum_{s=1}^{\hat{N}_i} \hat{B}_{(i,s)}} \cdot \mathbb{1} \left(\sum_{s=1}^{\hat{N}_i} \hat{B}_{(i,s)} > 0 \right).$$

On the tree, the weights $\{\hat{C}_i : i \in \mathcal{T}\}$ are used to construct the weights

$$\Pi_{(i,j)} = \Pi_i \hat{C}_{(i,j)}, \quad \Pi_\emptyset = 1.$$

5.2.2. *Proof of Theorem 3.1 for semi-sparse graphs.* Our goal is to bound the norm

$$\max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s X)_{i\bullet} - (\tilde{M}^s \check{X})_{i\bullet} \right\|_1 \right] = \max_{i \in V_n} \sum_{j=1}^t \mathbb{E}_n \left[\left| (C^s X)_{ij} - (\tilde{M}^s \check{X})_{ij} \right| \right]$$

for $X \in [-1, 1]^{n \times \ell}$ a random matrix, conditionally independent of C given \mathcal{F}_n and having mean $\check{X} = \mathbb{E}_n[X]$.

To simplify the notation, note that if we set $\mathbf{X} = X_{\bullet j}$ and $\bar{\mathbf{x}} = \check{X}_{\bullet j} = \mathbb{E}_n[\mathbf{X}]$, then

$$\mathbb{E}_n \left[\left| (C^s X)_{ij} - (\tilde{M}^s \check{X})_{ij} \right| \right] = \mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\tilde{M}^s \bar{\mathbf{x}})_i \right| \right],$$

so it suffices to analyze expectations of this form. We start by defining for each $i \in V_n$ and any $s \geq 1$, the event that the coupling with the K -type Galton-Watson tree holds:

$$D_{s,i} = \left\{ \mathcal{G}_i^{(s)}(J) \simeq \mathcal{T}_{\emptyset(i)}^{(s)}(\hat{J}) \right\}.$$

Lemma 5.2. *Let $\mathbf{X} = (X_1, \dots, X_n)^\top \in [-1, 1]^n$ be a random vector such that X_i has distribution $H_{\mathbf{A}_i}$, independently of everything else; let $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)^\top = \mathbb{E}_n[\mathbf{X}]$ denote its mean. Furthermore, suppose $\bar{x}_i = x_r$ for all i such that $J_i = r$, $1 \leq r \leq K$. Then, for any $i \in V_n$ and $s \geq 1$,*

$$\mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\tilde{M}^s \bar{\mathbf{x}})_i \right| \right] \leq \mathbb{E}_n \left[\left\| \sum_{|\mathbf{i}|=s} \Pi_{\mathbf{i}} \hat{X}_{\mathbf{i}} - (\tilde{M} \bar{\mathbf{x}})_{J_i} \right\| \mathbb{1}_{\hat{J}_\emptyset = J_i} \right] + 2\mathbb{P}_n(D_{s,i}^c),$$

where \tilde{M} is defined in (5.8), $\mathbf{x} = (x_1, \dots, x_K)^\top$, the $\Pi_{\mathbf{j}}$ are defined on the marked tree $\mathcal{T}(\hat{J})$, and $\hat{X}_{\mathbf{i}}$ is distributed according to $H_{\hat{\mathbf{A}}_{\mathbf{i}}}$, independently of everything else.

Proof. We start by considering one vertex $i \in V_n$ and construct its coupled tree $\mathcal{T}_{\emptyset(i)}(\hat{J}) \stackrel{\mathcal{D}}{=} (\mathcal{T}(\hat{J}) | \hat{J}_\emptyset = J_i)$. Note that on the event $D_{s,i}$ there is a unique path connecting vertex i to each of its neighbors at distance s , and these paths are the same as those in the tree, so

$$(C^s \mathbf{X})_i = \sum_{j_1=1}^n \cdots \sum_{j_s=1}^n C_{ij_1} C_{j_1 j_2} \cdots C_{j_{s-1} j_s} X_{j_s} = \sum_{|\mathbf{i}|=s} \Pi_{\mathbf{i}} \hat{X}_{\mathbf{i}}, \quad (5.14)$$

with $\hat{X}_{\mathbf{i}} = X_{\vartheta(\mathbf{i})}$, with ϑ the bijection determining the event $D_{s,i}$. Since the distribution of $X_{\vartheta(\mathbf{i})}$ depends only on $\mathbf{A}_{\vartheta(\mathbf{i})}$ and $\mathbf{A}_{\vartheta(\mathbf{i})} = \hat{\mathbf{A}}_{\mathbf{i}}$, then $(\hat{X}_{\mathbf{i}} | \hat{\mathbf{A}}_{\mathbf{i}})$ is distributed according to $H_{\hat{\mathbf{A}}_{\mathbf{i}}}$.

The induction proof given within the proof of Theorem 5.2 shows that

$$(\tilde{M}^s \bar{\mathbf{x}})_i = (\check{M} \mathbf{x})_{J_i}.$$

Consequently,

$$\begin{aligned} \mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\tilde{M}^s \bar{\mathbf{x}})_i \right| \right] &= \mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\check{M}^s \bar{\mathbf{x}})_i \right| \right] \\ &\leq \mathbb{E}_n \left[1(D_{s,i}) \left| (C^s \mathbf{X})_i - (\check{M}^s \bar{\mathbf{x}})_i \right| \right] + \mathbb{E}_n \left[1(D_{s,i}^c) \left(|(C^s \mathbf{X})_i| + |(\check{M}^s \bar{\mathbf{x}})_i| \right) \right] \\ &\leq \mathbb{E}_n \left[\left\| \sum_{|\mathbf{i}|=s} \Pi_{\mathbf{i}} \hat{X}_{\mathbf{i}} - (\check{M} \mathbf{x})_{J_i} \right\| \Bigg| \hat{J}_\emptyset = J_i \right] + 2\mathbb{P}_n(D_{s,i}^c), \end{aligned}$$

where the weights $\Pi_{\mathbf{i}}$ and the terminal values $\hat{X}_{\mathbf{i}}$ are defined on the marked tree $\mathcal{T}_{\emptyset(i)}(\hat{J})$, and we have used the observation that $\|C^s \mathbf{X}\|_\infty \leq 1$ and $\|\check{M}^s \bar{\mathbf{x}}\|_\infty \leq 1$. \square

Next, define the vector $\mathbf{a}_s := (a_s(1), \dots, a_s(K))^\top$ given by:

$$a_s(r) := \mathbb{E}_n \left[\left\| \sum_{|\mathbf{i}|=s} \hat{\Pi}_{\mathbf{i}} \hat{X}_{\mathbf{i}} - (\check{M}^s \mathbf{x})_r \right\| \Bigg| \hat{J}_\emptyset = r \right]. \quad (5.15)$$

Before proving the main theorem of this section, we give a bound for $\|\mathbf{a}_1\|_\infty$.

Lemma 5.3. *For \mathbf{a}_1 defined according to (5.15), and provided $\theta_n \rightarrow \infty$ as $n \rightarrow \infty$, there exists a constant $\Gamma' < \infty$ that does not depend on the choice of $\{\hat{X}_{\mathbf{i}}\}$ such that*

$$\|\mathbf{a}_1\|_\infty \leq \Gamma' \theta_n^{-1/2}.$$

Proof. Note that if the r th row of \check{M} is zero, then $a_1(r) = 0$, and since the non-zero rows of \check{M} coincide with those of M , it suffices to consider $r \in \mathcal{I}$. For $r \in \mathcal{I}$,

$$a_1(r) = \mathbb{E}_n \left[\left\| \sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i - (\check{M} \mathbf{x})_r \right\| \Bigg| \hat{J}_\emptyset = r \right],$$

and let

$$S_j := \sum_{i=1}^{\hat{N}_\emptyset} \hat{B}_i \hat{X}_i 1(\hat{J}_i = j) \quad \text{and} \quad \tilde{S}_j := \sum_{i=1}^{\hat{N}_\emptyset} \hat{B}_i \hat{X}_i 1(\tilde{J}_i = j),$$

so that

$$\sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i = \sum_{i=1}^{\hat{N}_\emptyset} \frac{\hat{B}_i}{\sum_{s=1}^{\hat{N}_\emptyset} \hat{B}_s} \hat{X}_i = \frac{\tilde{S}_1 + \dots + \tilde{S}_K}{S_1 + \dots + S_K}.$$

Define the event

$$F = \left\{ \left| \frac{\tilde{S}_1 + \dots + \tilde{S}_K}{S_1 + \dots + S_K} - (\check{M} \mathbf{x})_{\hat{J}_\emptyset} \right| \leq 2/(3H\Lambda_n) \right\},$$

and condition on \hat{J}_\emptyset to obtain that

$$\begin{aligned} & \mathbb{E}_n \left[\left\| \sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i - (\check{M}\mathbf{x})_r \right\| \middle| \hat{J}_\emptyset = r \right] \\ & \leq \mathbb{E}_n \left[1(F) \left\| \sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i - (\check{M}\mathbf{x})_r \right\| \middle| \hat{J}_\emptyset = r \right] + \mathbb{E}_n \left[1(F^c) \left(\sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i |\hat{X}_i| + |(\check{M}\mathbf{x})_r| \right) \middle| \hat{J}_\emptyset = r \right] \\ & = \int_0^{2/(3H\Lambda_n)} \mathbb{P}_n \left(\left\| \sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i - (\check{M}\mathbf{x})_r \right\| > t \middle| \hat{J}_\emptyset = r \right) dt + 2 \mathbb{P}_n \left(F^c \middle| \hat{J}_\emptyset = r \right), \end{aligned}$$

where we used the observation that $\sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i |\hat{X}_i| \leq 1$ and $|(\check{M}\mathbf{x})_r| \leq \|\check{M}\mathbf{x}\|_\infty \leq \|\check{M}\|_\infty \|\mathbf{x}\|_\infty \leq 1$, and therefore, the integrand is bounded by one.

Next, note that conditionally on $\hat{J}_\emptyset = r$, we have that $\tilde{S}_j \stackrel{D}{=} \sum_{i=1}^{N_j} B_i^{(r,j)} X_i^{(j)}$, where N_j is a Poisson random variable with mean $q_{jr}^{(n)} = \kappa(j, r) \pi_j^{(n)} \theta_n$, and is independent of the i.i.d. sequences $\{B_i^{(r,j)} : i \geq 1\}$ and $\{X_i^{(j)} : i \geq 1\}$, with $B_i^{(r,j)} \stackrel{D}{=} (\hat{B}_i | \hat{J}_\emptyset = r, \hat{J}_i = j)$ and $X_i^{(j)} \stackrel{D}{=} (\hat{X}_i | \hat{J}_i = j)$. Similarly, $S_j \stackrel{D}{=} \sum_{i=1}^{N_r} B_i^{(r,j)}$, and all the random variables $\{S_j, \tilde{S}_j : 1 \leq j \leq K\}$ are conditionally independent of each other given \hat{J}_\emptyset . It follows from Lemma 5.1 that for $r \in \mathcal{I}$,

$$\begin{aligned} \mathbb{P}_n \left(\left\| \sum_{i=1}^{\hat{N}_\emptyset} \hat{C}_i \hat{X}_i - (\check{M}\mathbf{x})_r \right\| > t \middle| \hat{J}_\emptyset = r \right) & \leq 4 \exp \left(-\frac{(t/2)^2 (\theta_n \mu_r^{(n)})^2}{2\theta_n \nu_r^{(n)}} + \frac{H(t/2)^3 (\theta_n \mu_r^{(n)})^3}{2(\theta_n \nu_r^{(n)})^2} \right) \\ & \leq 4 \exp \left(-\frac{(t/2)^2 \theta_n}{2\Delta_n} (1 - H\Lambda_n t/2) \right), \end{aligned}$$

as in the proof of Theorem 5.3, and

$$\mathbb{P}_n \left(F^c \middle| \hat{J}_\emptyset = r \right) \leq 4 \exp \left(-\frac{(1/(3H\Lambda_n))^2 \theta_n}{2\Delta_n} (1 - H\Lambda_n/(3H\Lambda_n)) \right) = 4 \exp \left(-\frac{\theta_n}{27\Delta_n (H\Lambda_n)^2} \right).$$

Since the bounds do not depend on \hat{J}_\emptyset , we obtain that

$$\begin{aligned} \|\mathbf{a}_1\|_\infty & \leq 4 \int_0^{2/(3H\Lambda_n)} \exp \left(-\frac{(t/2)^2 \theta_n}{2\Delta_n} (1 - H\Lambda_n t/2) \right) dt + 8 \exp \left(-\frac{\theta_n}{27\Delta_n (H\Lambda_n)^2} \right) \\ & \leq 4 \int_0^{2/(3H\Lambda_n)} \exp \left(-\frac{(t/2)^2 \theta_n}{3\Delta_n} \right) dt + 8 \exp \left(-\frac{\theta_n}{27\Delta_n (H\Lambda_n)^2} \right) \\ & = \frac{4}{\sqrt{\theta_n/(6\Delta_n)}} \int_0^{2\sqrt{\theta_n/(6\Delta_n)}/(3H\Lambda_n)} e^{-z^2/2} dz + 8 \exp \left(-\frac{\theta_n}{27\Delta_n (H\Lambda_n)^2} \right) \\ & \leq \frac{8\sqrt{3\Delta_n\pi}}{\sqrt{\theta_n}} \int_0^\infty \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz + 8 \exp \left(-\frac{\theta_n}{27\Delta_n (H\Lambda_n)^2} \right) \end{aligned}$$

$$= \frac{4\sqrt{3\Delta_n\pi}}{\sqrt{\theta_n}} + 8 \exp\left(-\frac{\theta_n}{27\Delta_n(H\Lambda_n)^2}\right) = O\left(\theta_n^{-1/2}\right)$$

as $n \rightarrow \infty$. Noting that the last estimates do not depend on the choice of the $\{\hat{X}_i\}$ completes the proof. \square

As a corollary to the previous lemma, we obtain the following bound for $\|\mathbf{a}_s\|_\infty$.

Corollary 5.1. *For $s \geq 1$ and \mathbf{a}_s defined according to (5.15), and provided $\theta_n \rightarrow \infty$ as $n \rightarrow \infty$, there exists a constant $\Gamma' < \infty$ that does not depend on the choice of $\{\hat{X}_i\}$ such that*

$$\|\mathbf{a}_s\|_\infty \leq \Gamma' s \theta_n^{-1/2}.$$

Proof. To derive a recursion for \mathbf{a}_s , $s \geq 1$, note that for each $j \in \{1, \dots, K\}$,

$$\begin{aligned} a_s(j) &= \mathbb{E}_n \left[\left\| \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \sum_{|(r,\mathbf{i})|=s} \frac{\hat{\Pi}_{(r,\mathbf{i})}}{\hat{C}_r} \hat{X}_{(r,\mathbf{i})} - \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r (\check{M}^{s-1}\mathbf{x})_{j_r} + \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r (\check{M}^{s-1}\mathbf{x})_{j_r} - (\check{M}^s\mathbf{x})_j \right\| \middle| \hat{J}_\emptyset = j \right] \\ &\leq \mathbb{E}_n \left[\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \left\| \sum_{|(r,\mathbf{i})|=s} \frac{\hat{\Pi}_{(r,\mathbf{i})}}{\hat{C}_r} \hat{X}_{(r,\mathbf{i})} - (\check{M}^{s-1}\mathbf{x})_{j_r} \right\| \middle| \hat{J}_\emptyset = j \right] \\ &\quad + \mathbb{E}_n \left[\left\| \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r (\check{M}^{s-1}\mathbf{x})_{j_r} - (\check{M}^s\mathbf{x})_j \right\| \middle| \hat{J}_\emptyset = j \right] \\ &= \mathbb{E}_n \left[\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \mathbb{E}_n \left[\left\| \sum_{|(r,\mathbf{i})|=s} \frac{\hat{\Pi}_{(r,\mathbf{i})}}{\hat{C}_r} \hat{X}_{(r,\mathbf{i})} - (\check{M}^{s-1}\mathbf{x})_{j_r} \right\| \middle| \hat{N}_\emptyset, \{\hat{C}_r, \hat{J}_r\}_{r=1}^{\hat{N}_\emptyset} \right] \middle| \hat{J}_\emptyset = j \right] \\ &\quad + \mathbb{E}_n \left[\left\| \sum_{l=1}^K \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \mathbf{1}(\hat{J}_r = l) (\check{M}^{s-1}\mathbf{x})_l - \sum_{l=1}^K \check{M}_{jl} (\check{M}^{s-1}\mathbf{x})_l \right\| \middle| \hat{J}_\emptyset = j \right] \\ &= \mathbb{E}_n \left[\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r a_{s-1}(\hat{J}_r) \middle| \hat{J}_\emptyset = j \right] + \mathbb{E}_n \left[\left\| \sum_{l=1}^K (\check{M}^{s-1}\mathbf{x})_l \left(\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \mathbf{1}(\hat{J}_r = l) - \check{M}_{jl} \right) \right\| \middle| \hat{J}_\emptyset = j \right] \\ &\leq \max_{1 \leq j \leq K} a_{s-1}(j) \mathbb{E}_n \left[\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \middle| \hat{J}_\emptyset = j \right] + \mathbb{E}_n \left[\left\| \sum_{l=1}^K (\check{M}^{s-1}\mathbf{x})_l \left(\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \mathbf{1}(\hat{J}_r = l) - \check{M}_{jl} \right) \right\| \middle| \hat{J}_\emptyset = j \right]. \end{aligned}$$

To simplify the bound, recall that $\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r = 1$. To bound the second expectation, let $\hat{Y}_r = y_l = (\check{M}^{s-1}\mathbf{x})_l$ for all r such that $\hat{J}_r = l$, define $\mathbf{y} = (y_1, \dots, y_K)^\top$, and note that

$$\mathbb{E}_n \left[\left\| \sum_{l=1}^K (\check{M}^{s-1}\mathbf{x})_l \left(\sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \mathbf{1}(\hat{J}_r = l) - \check{M}_{jl} \right) \right\| \middle| \hat{J}_\emptyset = j \right] = \mathbb{E}_n \left[\left\| \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \hat{Y}_r - (\check{M}\mathbf{y})_j \right\| \middle| \hat{J}_\emptyset = j \right],$$

so by Lemma 5.3 we get

$$\max_{1 \leq j \leq K} \mathbb{E}_n \left[\left\| \sum_{r=1}^{\hat{N}_\emptyset} \hat{C}_r \hat{Y}_r - (\check{M}\mathbf{y})_j \right\| \middle| \hat{J}_\emptyset = j \right] \leq \Gamma' s \theta_n^{-1/2}$$

for some constant Γ' that does not depend on the $\{\hat{Y}_i\}$. Therefore,

$$\|\mathbf{a}_s\|_\infty \leq \|\mathbf{a}_{s-1}\|_\infty + \Gamma' \theta_n^{-1/2}.$$

Iterating the recursion gives

$$\|\mathbf{a}_s\|_\infty \leq \Gamma' s \theta_n^{-1/2}.$$

□

We now state the counterpart of Theorem 5.3 for this regime.

Theorem 5.5. *Let $\mathbf{X} = (X_1, \dots, X_n)^\top \in [-1, 1]^n$ be a random vector such that X_i has distribution $H_{\mathbf{A}_i}$, independently of everything else; let $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)^\top = \mathbb{E}_n[\mathbf{X}]$ denote its mean. Furthermore, suppose $\bar{x}_i = x_r$ for all i such that $J_i = r$, $1 \leq r \leq K$. Then, there exists a constant Γ' that does not depend on the choice of \mathbf{X} such that for any $i \in V_n$ and $s \geq 1$,*

$$\mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\tilde{M}^s \bar{\mathbf{x}})_i \right| \right] \leq \frac{\Gamma' s}{\theta_n^{1/2}} + 2\mathbb{P}_n(D_{s,i}^c).$$

Proof. By Lemma 5.2 and Corollary 5.1, we have that

$$\begin{aligned} \mathbb{E}_n \left[\left| (C^s \mathbf{X})_i - (\tilde{M}^s \bar{\mathbf{x}})_i \right| \right] &\leq \mathbb{E}_n \left[\left\| \sum_{|\mathbf{i}|=s} \Pi_{\mathbf{i}} \hat{X}_{\mathbf{i}} - (\check{M}\mathbf{x})_{J_i} \right\| \middle| \hat{J}_\emptyset = J_i \right] + 2\mathbb{P}_n(D_{s,i}^c) \\ &\leq a_s(J_i) + 2\mathbb{P}_n(D_{s,i}^c) \leq \|\mathbf{a}_s\|_\infty + 2\mathbb{P}_n(D_{s,i}^c) \\ &\leq \Gamma' s \theta_n^{-1/2} + 2\mathbb{P}_n(D_{s,i}^c). \end{aligned}$$

□

We can now give the proof to the second part of Theorem 3.1.

Proof of Theorem 3.1. Suppose $\theta_n \rightarrow \infty$ as $n \rightarrow \infty$, and recall that we need to prove that

$$\sup_{k \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \xrightarrow{P} 0$$

as $n \rightarrow \infty$. Note that by Remark 3.1(a), it suffices to assume that $\theta_n \leq (\log n)^b$ for some $b > 1$. To start, note that by Theorem 5.1, we have that for any fixed $k \geq 0$,

$$\begin{aligned} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] &\leq \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \tilde{\mathbf{R}}_i^{(k)} \right\|_1 \right] + \max_{i \in V_n} \mathbb{E}_n \left[\left\| \tilde{\mathbf{R}}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \\ &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s W^{(0)})_{i_\bullet} - (\tilde{M}^s \check{W})_{i_\bullet} \right\|_1 \right] \end{aligned}$$

$$+ \sum_{s=1}^k a_{s,k} \max_{i \in V_n} \mathbb{E}_n \left[\left\| (C^s R^{(0)})_{i\bullet} - (\tilde{M}^s \check{R})_{i\bullet} \right\|_1 \right] + \mathbb{E}_n \left[\left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right].$$

Moreover, by Theorem 5.2,

$$\mathbb{E}_n \left[\left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right] \leq \mathbb{E}_n \left[\sup_{k \geq 0} \left\| \tilde{R}^{(k)} - \mathcal{R}^{(k)} \right\|_\infty \right] \leq \frac{\ell c}{d^2} \cdot \mathcal{E}_n.$$

Now use Theorem 5.5 to obtain that

$$\mathbb{E}_n \left[\left\| (C^s W^{(0)})_{i\bullet} - (\tilde{M}^s \check{W})_{i\bullet} \right\|_1 \right] = \sum_{j=1}^{\ell} \mathbb{E}_n \left[\left\| (C^s W_{\bullet j}^{(0)})_i - (\tilde{M}^s \check{W}_{\bullet j})_i \right\|_1 \right] \leq \ell \left(\Gamma' s \theta_n^{-1/2} + 2\mathbb{P}_n(D_{s,i}^c) \right)$$

and

$$\mathbb{E}_n \left[\left\| (C^s R^{(0)})_{i\bullet} - (\tilde{M}^s \check{R})_{i\bullet} \right\|_1 \right] \leq \ell \left(\Gamma' s \theta_n^{-1/2} + 2\mathbb{P}_n(D_{s,i}^c) \right).$$

It follows that

$$\begin{aligned} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] &\leq 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \ell \left(\Gamma' s \theta_n^{-1/2} + 2\mathbb{P}_n(D_{s,i}^c) \right) \\ &\quad + \sum_{s=1}^k a_{s,k} \max_{i \in V_n} \ell \left(\Gamma' s \theta_n^{-1/2} + 2\mathbb{P}_n(D_{s,i}^c) \right) + \frac{\ell c}{d^2} \cdot \mathcal{E}_n \\ &= \ell \Gamma' \theta_n^{-1/2} \sum_{t=1}^k \sum_{s=1}^t a_{s,t} s + 2 \sum_{t=1}^k \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) + \frac{\ell c}{d^2} \cdot \mathcal{E}_n \\ &= \ell \Gamma' \theta_n^{-1/2} c \sum_{t=1}^k (1-d)^{t-1} t + 2 \sum_{t=1}^k \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) + \frac{\ell c}{d^2} \cdot \mathcal{E}_n, \end{aligned}$$

where in the last equality we used (5.4). Since $\sum_{t=1}^k (1-d)^{t-1} t \leq d^{-2}$ for all $k \geq 0$, and

$$\begin{aligned} \sum_{t=1}^k \sum_{s=1}^t a_{s,t} \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) &= \sum_{s=1}^k \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) c^s \sum_{t=s}^k \binom{t}{s} (1-c-d)^{t-s} \\ &= \sum_{s=1}^k \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) c^s (c+d)^{-s-1} \sum_{m=0}^{k-s} \binom{m+s}{m} (1-c-d)^m (c+d)^{s+1} \\ &\leq \sum_{s=1}^{\infty} \max_{i \in V_n} \mathbb{P}_n(D_{s,i}^c) \frac{c^s}{(c+d)^{s+1}} = \frac{c}{(c+d)d} \mathbb{E}_n \left[\max_{i \in V_n} \mathbb{P}_n(D_{T,i}^c) \right], \end{aligned}$$

where T is a geometric random variable on $\{1, 2, \dots\}$ with success probability $p = d/(c+d)$, we have shown that

$$\sup_{k \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(k)} - \mathcal{R}_i^{(k)} \right\|_1 \right] \leq \frac{\ell c}{d^2} \left(\Gamma' \theta_n^{-1/2} + \mathcal{E}_n \right) + \frac{2c}{(c+d)d} \mathbb{E}_n \left[\max_{i \in V_n} \mathbb{P}_n(D_{T,i}^c) \right].$$

It only remains to show that $\max_{i \in V_n} \mathbb{P}_n(D_{T,i}^c) \xrightarrow{P} 0$ as $n \rightarrow \infty$ for any fixed T , since then the bounded convergence theorem would complete the proof. To see this is the case, note that all vertices $\{i \in V_n : J_i = r\}$, $1 \leq r \leq K$, are exchangeable, so for any $i \in V_n$,

$$\mathbb{P}_n(D_{T,i}^c) = \sum_{r=1}^K \mathbb{P}_n(D_{T,i}^c) 1(J_i = r) = \sum_{r=1}^K \frac{1}{\pi_r^{(n)} n} \sum_{m=1}^n \mathbb{P}_n(D_{T,m}^c) 1(J_m = r) = \frac{1}{n} \sum_{m=1}^n \mathbb{P}_n(D_{T,m}^c) \sum_{r=1}^K \frac{1(J_m = r)}{\pi_r^{(n)}},$$

and therefore, by Theorem 5.4 and Assumption A,

$$\max_{i \in V_n} \mathbb{P}_n(D_{T,i}^c) \leq \frac{1}{n} \sum_{m=1}^n \mathbb{P}_n(D_{T,m}^c) \cdot \sum_{r=1}^K \frac{1}{\pi_r^{(n)}} \xrightarrow{P} 0$$

as $n \rightarrow \infty$. This completes the proof. \square

The last section of the paper contains the proofs of Theorems 3.2 and 3.3, which are a consequence of Theorem 3.1.

5.3. Proofs of Theorems 3.2 and 3.3. We start by proving Theorem 3.2, which focuses on the trajectories of the process $\{R^{(k)} : k \geq 0\}$.

Proof of Theorem 3.2. Fix $k \geq 0$ and recall that $V_{k,i} \in [-1, 1]^{\ell \times (k+1)}$ is the matrix having j th column $(V_{k,i})_{\bullet j} = \mathbf{R}_i^{(j-1)}$, and $\mathcal{V}_k \in [-1, 1]^{\ell \times (k+1)}$ is the matrix having j th column $(\mathcal{V}_k)_{\bullet j} = \mathcal{R}_\emptyset^{(j-1)}$. Assume $f : [-1, 1]^{\ell \times (k+1)} \rightarrow \mathbb{R}$ is bounded and continuous with respect to $\|\cdot\|_1$ (the induced operator norm). Since f is continuous on a compact set, it is also uniformly continuous, so for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(X) - f(Y)| < \epsilon \quad \text{whenever} \quad \|X - Y\|_1 < \delta.$$

Let $\mathcal{V}_{k,i} \in [-1, 1]^{\ell \times (k+1)}$ denote the matrix having j th column $(\mathcal{V}_{k,i})_{\bullet j} = \mathcal{R}_i^{(j-1)}$, and let

$$f^* = \max_{X \in [-1, 1]^{\ell \times (k+1)}} |f(X)|.$$

It follows that for any $1 \leq r \leq K$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) \\ & \leq \epsilon \cdot \frac{1}{n} \sum_{i=1}^n 1(\|V_{k,i} - \mathcal{V}_{k,i}\|_1 < \delta) + 2f^* \cdot \frac{1}{n} \sum_{i=1}^n 1(\|V_{k,i} - \mathcal{V}_{k,i}\|_1 \geq \delta) \\ & \leq \epsilon + \frac{2f^*}{\delta} \cdot \frac{1}{n} \sum_{i=1}^n \|V_{k,i} - \mathcal{V}_{k,i}\|_1 \\ & = \epsilon + \frac{2f^*}{\delta} \cdot \frac{1}{n} \sum_{i=1}^n \max_{0 \leq m \leq k} \left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \\ & \leq \epsilon + \frac{2f^*}{\delta} \cdot \frac{1}{n} \sum_{m=0}^k \sum_{i=1}^n \left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \end{aligned}$$

Take expectations on both sides to obtain

$$\begin{aligned}
\mathbb{E}_n \left[\frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) \right] &\leq \varepsilon + \frac{2f^*}{\delta} \cdot \frac{1}{n} \sum_{m=0}^k \sum_{i=1}^n \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \\
&\leq \varepsilon + \frac{2f^*}{\delta} \cdot \sum_{m=0}^k \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \\
&\leq \varepsilon + \frac{2(k+1)f^*}{\delta} \cdot \sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right].
\end{aligned}$$

From Theorem 3.1, we have that $\sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \xrightarrow{P} 0$ as $n \rightarrow \infty$. The bounded convergence theorem then gives

$$\lim_{n \rightarrow \infty} E \left[\frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) \right] \leq \varepsilon + \frac{2(k+1)f^*}{\delta} \cdot \lim_{n \rightarrow \infty} E \left[\sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \right] = \varepsilon,$$

which in turn implies that

$$\frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

Moreover, Kolmogorov's strong law of large numbers gives

$$\left| \frac{1}{n} \sum_{i=1}^n (f(\mathcal{V}_{k,i}) 1(J_i = r) - E[f(\mathcal{V}_{k,i}) 1(J_i = r)]) \right| \rightarrow 0 \quad a.s.$$

as $n \rightarrow \infty$. Since

$$E[f(\mathcal{V}_{k,i}) 1(J_i = r)] = E[f(\mathcal{V}_{k,i}) | J_i = r] E[\pi_r^{(n)}] = E[f(\mathcal{V}_k) | \mathcal{J}_\emptyset = r] E[\pi_r^{(n)}] = E[f(\mathcal{V}_k) 1(\mathcal{J}_\emptyset = r)] E[\pi_r^{(n)}] / \pi_r,$$

we have shown that

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{i=1}^n f(V_{k,i}) 1(J_i = r) - E[f(\mathcal{V}_k) 1(\mathcal{J}_\emptyset = r)] \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) + \left| \frac{1}{n} \sum_{i=1}^n f(\mathcal{V}_{k,i}) 1(J_i = r) - E[f(\mathcal{V}_k) 1(\mathcal{J}_\emptyset = r)] \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n |f(V_{k,i}) - f(\mathcal{V}_{k,i})| 1(J_i = r) + \left| \frac{1}{n} \sum_{i=1}^n (f(\mathcal{V}_{k,i}) 1(J_i = r) - E[f(\mathcal{V}_{k,i}) 1(J_i = r)]) \right| \\
&\quad + |E[f(\mathcal{V}_k) 1(\mathcal{J}_\emptyset = r)]| \left| \frac{E[\pi_r^{(n)}]}{\pi_r} - 1 \right| \xrightarrow{P} \varepsilon
\end{aligned}$$

as $n \rightarrow \infty$. Since $\varepsilon > 0$ was arbitrary, the first statement of the theorem follows.

For the second statement, let $\{f_1, \dots, f_m\}$ be bounded and continuous (with respect to $\|\cdot\|_1$) functions on $[-1, 1]^{\ell \times (k+1)}$. Recall that we may assume that the functions are uniformly continuous. Now note that

$$E[f_j(\mathcal{V}_k) | \mathcal{J}_\emptyset = J_{i_j}] = \mathbb{E}_n[f_j(\mathcal{V}_{k, i_j})],$$

and since the $\{\mathcal{V}_{k, i} : i \in V_n\}$ are conditionally independent given the community labels \mathcal{J}_n , then

$$\begin{aligned} \left| \mathbb{E}_n \left[\prod_{j=1}^m f_j(V_{k, i_j}) \right] - \prod_{j=1}^m E[f_j(\mathcal{V}_k) | \mathcal{J}_\emptyset = J_{i_j}] \right| &= \left| \mathbb{E}_n \left[\prod_{j=1}^m f_j(V_{k, i_j}) \right] - \mathbb{E}_n \left[\prod_{j=1}^m f_j(\mathcal{V}_{k, i_j}) \right] \right| \\ &\leq \mathbb{E}_n \left[\left| \prod_{j=1}^m f_j(V_{k, i_j}) - \prod_{j=1}^m f_j(\mathcal{V}_{k, i_j}) \right| \right] \\ &\leq \prod_{j=1}^m f_j^* \sum_{j=1}^m \mathbb{E}_n[|f_j(V_{k, i_j}) - f_j(\mathcal{V}_{k, i_j})|], \end{aligned}$$

where $f_j^* = \max_{X \in [-1, 1]^{\ell \times (k+1)}} |f_j(X)|$, and we have used the inequality

$$\left| \prod_{i=1}^m x_i - \prod_{i=1}^m y_i \right| \leq \prod_{i=1}^m (|x_i| \vee |y_i|) \sum_{i=1}^m |x_i - y_i|.$$

To complete the proof, note that the random variables $\{|f_j(V_{k, t}) - f_j(\mathcal{V}_{k, t})| : J_t = r_j\}$ are exchangeable, so if $J_{i_j} = r_j$, then

$$\begin{aligned} \mathbb{E}_n[|f_j(V_{k, i_j}) - f_j(\mathcal{V}_{k, i_j})|] &= \frac{1}{n\pi_{r_j}^{(n)}} \sum_{t: J_t = r_j} \mathbb{E}_n[|f_j(V_{k, t}) - f_j(\mathcal{V}_{k, t})|] \\ &= \frac{1}{\pi_{r_j}^{(n)}} \mathbb{E}_n \left[\frac{1}{n} \sum_{t=1}^n |f_j(V_{k, t}) - f_j(\mathcal{V}_{k, t})| 1(J_t = r_j) \right]. \end{aligned}$$

Now apply the same arguments used in the first part of the theorem to obtain that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{E}_n[|f_j(V_{k, i_j}) - f_j(\mathcal{V}_{k, i_j})|] \leq \frac{1}{\pi_{r_j}^{(n)}} \left(\varepsilon + \frac{2(k+1)f_j^*}{\delta} \sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \right).$$

We conclude that

$$\begin{aligned} &\left| \mathbb{E}_n \left[\prod_{j=1}^m f_j(V_{k, i_j}) \right] - \prod_{j=1}^m E[f_j(\mathcal{V}_k) | \mathcal{J}_\emptyset = J_{i_j}] \right| \\ &\leq \prod_{j=1}^m f_j^* \sum_{j=1}^m \frac{1}{\pi_{r_j}^{(n)}} \left(\varepsilon + \frac{2(k+1)f_j^*}{\delta} \sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \right) \xrightarrow{P} \prod_{j=1}^m f_j^* \sum_{j=1}^m \frac{\varepsilon}{\pi_{r_j}} \end{aligned}$$

as $n \rightarrow \infty$. Taking $\varepsilon \downarrow 0$ completes the proof. \square

We now prove Theorem 3.3, which refers to the stationary behavior of the process $\{R^{(k)} : k \geq 0\}$.

Proof of Theorem 3.3. We will start by constructing a coupling of two matrices, $R \in [-1, 1]^{n \times \ell}$ and $\mathcal{R} \in [-1, 1]^{n \times \ell}$, the first one with rows $R_{j\bullet} = \mathbf{R}_j$ and the second one with rows $\mathcal{R}_{j\bullet} = \mathcal{R}_j$, as well as of the random variable \mathcal{J}_\emptyset , such that R has the stationary distribution of $\{R^{(k)} : k \geq 0\}$,

$$\mathcal{R}_i = \sum_{t=0}^{\infty} (1-c-d)^t \mathbf{W}_i^{(t)} + \sum_{t=1}^{\infty} \sum_{s=1}^t a_{s,t} (M^s \bar{W})_{J_i}, \quad i \in V_n,$$

and

$$\max_{i \in V_n} \mathbb{E}_n [\|\mathbf{R}_i - \mathcal{R}_i\|_1] \rightarrow 0, \quad n \rightarrow \infty.$$

Note that exactly the same arguments used in the proof of Theorem 3.2 (set $k = 0$) will then yield

$$\frac{1}{n} \sum_{i=1}^n f(\mathbf{R}_i) 1(J_i = r) \xrightarrow{P} E[f(\mathcal{R}_\emptyset) 1(\mathcal{J}_\emptyset = r)] \quad \text{and} \quad \mathbb{E}_n \left[\prod_{j=1}^m f_j(\mathbf{R}_{i_j}) \right] \xrightarrow{P} \prod_{j=1}^m E[f_j(\mathcal{R}_\emptyset) | \mathcal{J}_\emptyset = J_{i_j}],$$

as $n \rightarrow \infty$. Hence, we only need to exhibit the coupling.

Define $\{B^{(k)} : k \geq 0\} \subset [-1, 1]^{n \times \ell}$ to be the process defined according to

$$B^{(k)} = \sum_{t=0}^{k-1} A^t W^{(t)},$$

and let $B = \sum_{t=0}^{\infty} A^t W^{(t)}$. To see that B is well-defined, recall that $\|A\|_\infty = 1 - d < 1$, and therefore,

$$\|B - B^{(k)}\|_\infty = \left\| \sum_{t=k}^{\infty} A^t W^{(t)} \right\|_\infty \leq \sum_{t=k}^{\infty} \|A\|_\infty^t \|W^{(t)}\|_\infty \leq \ell \sum_{t=k}^{\infty} (1-d)^t = \ell(1-d)^k/d \rightarrow 0,$$

as $k \rightarrow \infty$, which yields $B = \lim_{k \rightarrow \infty} B^{(k)}$.

Going back to the original opinion recursion (5.1), note that it satisfies

$$R^{(k)} = \sum_{r=0}^{k-1} A^r W^{(k-r)} + A^k R^{(0)} \stackrel{D}{=} B^{(k)} + A^k R^{(0)},$$

from where we obtain

$$R^{(k)} \Rightarrow B, \quad k \rightarrow \infty.$$

It follows that $B \stackrel{D}{=} R$. The coupling we need is given by (B, \mathcal{R}) . It remains to show that

$$\max_{i \in V_n} \mathbb{E}_n [\|\mathbf{B}_i - \mathcal{R}_i\|_1] \rightarrow 0, \quad n \rightarrow \infty,$$

where \mathbf{B}_i is the i th row of B .

To this end, define the matrix $\mathcal{B}^{(k)} \in [-1, 1]^{n \times \ell}$ having rows $(\mathcal{B}^{(k)})_{i\bullet} = \mathcal{B}_i^{(k)}$ given by

$$\mathcal{B}_i^{(k)} = \sum_{t=0}^{k-1} (1-c-d)^t \mathbf{W}_i^{(t)} + 1(k \geq 2) \sum_{t=1}^{k-1} \sum_{s=1}^t a_{s,t} (M^s \bar{W})_{J_i}.$$

Next, note that for any $k \geq 1$ and any $i \in V_n$,

$$\begin{aligned} \|\mathbf{B}_i - \mathcal{R}_i\|_1 &\leq \left\| \mathbf{B}_i - \mathbf{B}_i^{(k)} \right\|_1 + \left\| \mathbf{B}_i^{(k)} - \mathcal{B}_i^{(k)} \right\|_1 + \left\| \mathcal{B}_i^{(k)} - \mathcal{R}_i \right\|_1 \\ &\leq \|B - B^{(k)}\|_\infty + \left\| \mathbf{B}_i^{(k)} - \mathcal{B}_i^{(k)} \right\|_1 + \|\mathcal{B}^{(k)} - \mathcal{R}\|_\infty. \end{aligned}$$

We have already computed $\|B - B^{(k)}\|_\infty \leq \ell(1-d)^k/d$, and the last norm can be computed as follows:

$$\begin{aligned} \left\| \mathcal{B}^{(k)} - \mathcal{R} \right\|_\infty &\leq \sum_{t=k}^{\infty} (1-c-d)^t \|W^{(t)}\|_\infty + \sum_{t=k}^{\infty} \sum_{s=1}^t a_{s,t} \|M^s \bar{W}\|_\infty \\ &\leq \ell \sum_{t=k}^{\infty} (1-c-d)^t + \ell \sum_{t=k}^{\infty} \sum_{s=1}^t a_{s,t} = \ell \sum_{t=k}^{\infty} (1-c-d)^t + \ell \sum_{t=k}^{\infty} \sum_{s=1}^t \binom{t}{s} (1-c-d)^{t-s} c^s \\ &= \ell \sum_{t=k}^{\infty} (1-c-d)^t + \ell \sum_{t=k}^{\infty} ((1-d)^t - (1-c-d)^t) = \ell \sum_{t=k}^{\infty} (1-d)^t = \ell(1-d)^k/d, \end{aligned}$$

where in the second inequality we used the observation that $\|W^{(t)}\|_\infty \leq \ell$ and $\|M^s \bar{W}\|_\infty \leq \ell$.

We have thus shown that for any $i \in V_n$ and any $k \geq 0$,

$$\|\mathbf{B}_i - \mathcal{R}_i\|_1 \leq \left\| \mathbf{B}_i^{(k)} - \mathcal{B}_i^{(k)} \right\|_1 + 2\ell(1-d)^k/d.$$

Taking expectations on both sides, followed by the maximum over i , we get

$$\max_{i \in V_n} \mathbb{E}_n [\|\mathbf{B}_i - \mathcal{R}_i\|_1] \leq 2\ell(1-d)^k/d + \sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{B}_i^{(m)} - \mathcal{B}_i^{(m)} \right\|_1 \right]. \quad (5.16)$$

Finally, note that by Theorem 3.1,

$$\sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{B}_i^{(m)} - \mathcal{B}_i^{(m)} \right\|_1 \right] = \sup_{m \geq 0} \max_{i \in V_n} \mathbb{E}_n \left[\left\| \mathbf{R}_i^{(m)} - \mathcal{R}_i^{(m)} \right\|_1 \right] \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

Taking $k \rightarrow \infty$ in (5.16) completes the proof. \square

Acknowledgments. This work was supported by NSF grant CMMI-2243261.

REFERENCES

- [1] Abramowitz, A. I. and Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2):542–555.
- [2] Aldous, D. and Steele, J. M. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on discrete structures*, pages 1–72.
- [3] Avrachenkov, K., Kadavankandy, A., and Litvak, N. (2018). Mean field analysis of personalized pagerank with implications for local graph clustering. *Journal of statistical physics*, 173:895–916.
- [4] Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226.

- [5] Banerjee, S. and Olvera-Cravioto, M. (2022). Pagerank asymptotics on directed preferential attachment networks. *The Annals of Applied Probability*, 32(4):3060–3084.
- [6] Ben-Naim, E., Krapivsky, P. L., and Redner, S. (2003). Bifurcations and patterns in compromise processes. *Physica D: nonlinear phenomena*, 183(3-4):190–204.
- [7] Benjamini, I. and Schramm, O. (2001). Recurrence of Distributional Limits of Finite Planar Graphs. *Electronic Journal of Probability*, 6(none):1 – 13.
- [8] Berger, R. L. (1981). A necessary and sufficient condition for reaching a consensus using degroot’s method. *Journal of the American Statistical Association*, 76(374):415–418.
- [9] Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., and Quattrociocchi, W. (2016). Users polarization on facebook and youtube. *PLOS ONE*, 11(8):1–24.
- [10] Bhamidi, S., Budhiraja, A., and Wu, R. (2019). Weakly interacting particle systems on inhomogeneous random graphs. *Stochastic Processes and their Applications*, 129(6):2174–2206.
- [11] Budhiraja, A., Mukherjee, D., and Wu, R. (2019). Supermarket model on graphs. *The Annals of Applied Probability*, 29(3):1740 – 1777.
- [12] Chaintron, L.-P. and Diez, A. (2021). Propagation of chaos: a review of models, methods and applications. ii. applications. *arXiv preprint arXiv:2106.14812*.
- [13] Chaintron, L.-P. and Diez, A. (2022). Propagation of chaos: a review of models, methods and applications. i. models and methods. *arXiv preprint arXiv:2203.00446*.
- [14] Chen, N., Litvak, N., and Olvera-Cravioto, M. (2014). Pagerank in scale-free random graphs. In *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings 11*, pages 120–131. Springer.
- [15] Como, G. and Fagnani, F. (2011). Scaling limits for continuous opinion dynamics systems. *The Annals of Applied Probability*, 21(4):1537–1567.
- [16] Dandekar, P., Goel, A., and Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796.
- [17] Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. (2002). How can extremism prevail? a study based on the relative agreement interaction model. *Journal of artificial societies and social simulation*, 5(4).
- [18] Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98.
- [19] DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, 69(345):118–121.
- [20] Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific reports*, 7(1):40391.
- [21] Fennell, S. C., Burke, K., Quayle, M., and Gleeson, J. P. (2021). Generalized mean-field approximation for the deffuant opinion dynamics model on networks. *Physical Review E*, 103(1):012314.

- [22] Fraiman, N., Lin, T.-C., and Olvera-Cravioto, M. (2022a). Opinion dynamics on directed complex networks. *arXiv preprint arXiv:2209.00969*.
- [23] Fraiman, N., Lin, T.-C., and Olvera-Cravioto, M. (2022b). Stochastic recursions on directed random graphs. *Stochastic Processes and their Applications*.
- [24] French, J. R. (1956). A formal theory of social power. *Psychological review*, 63(3):181–194.
- [25] Friedkin, N. E. and Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206.
- [26] Garavaglia, A., van der Hofstad, R., and Litvak, N. (2020). Local weak convergence for pagerank. *Annals of Applied Probability*.
- [27] Gentzkow, M. (2016). Polarization in 2016. *Toulouse Network for Information Technology Whitepaper*, 1.
- [28] Gómez-Serrano, J., Graham, C., and Le Boudec, J.-Y. (2012). The bounded confidence model of opinion dynamics. *Mathematical Models and Methods in Applied Sciences*, 22(02):1150007.
- [29] Hetherington, M. J. (2009). Putting polarization in perspective. *British Journal of Political Science*, 39(2):413–448.
- [30] Holyst, J. A., Kacperski, K., and Schweitzer, F. (2001). Social impact models of opinion dynamics. *Annual Reviews Of Computational Physics IX*, pages 253–273.
- [31] Jia, P., MirTabatabaei, A., Friedkin, N. E., and Bullo, F. (2015). Opinion dynamics and the evolution of social power in influence networks. *SIAM review*, 57(3):367–397.
- [32] Lacker, D., Ramanan, K., and Wu, R. (2019). Local weak convergence for sparse networks of interacting processes. *arXiv preprint arXiv:1904.02585*.
- [33] Lacker, D., Ramanan, K., and Wu, R. (2020). Marginal dynamics of interacting diffusions on unimodular galton-watson trees. *arXiv preprint arXiv:2009.11667*.
- [34] Le Boudec, J.-Y., McDonald, D., and Mundinger, J. (2007). A generic mean field convergence result for systems of interacting objects. In *Fourth international conference on the quantitative evaluation of systems (QEST 2007)*, pages 3–18. IEEE.
- [35] Lee, J. and Olvera-Cravioto, M. (2020). Pagerank on inhomogeneous random digraphs. *Stochastic Processes and their Applications*, 130(4):2312–2348.
- [36] Liggett, T. M. (1999). *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324. springer science & Business Media.
- [37] Meng, X. F., Van Gorder, R. A., and Porter, M. A. (2018). Opinion formation and distribution in a bounded-confidence model on various networks. *Physical Review E*, 97(2):022312.
- [38] Mynatt, C. R., Doherty, M. E., and Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29(1):85–95.

- [39] Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220.
- [40] Olvera-Cravioto, M. (2022). Strong couplings for static locally tree-like random graphs. *Journal of Applied Probability*, 59(4):1261–1285.
- [41] Parsegov, S. E., Proskurnikov, A. V., Tempo, R., and Friedkin, N. E. (2016). Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285.
- [42] Rainer, H. and Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3).
- [43] Redner, S. (2019). Reality-inspired voter models: A mini-review. *Comptes Rendus Physique*, 20(4):275–292.
- [44] Schmidt, A. L., Zollo, F., Vicario, M. D., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2017). Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, 114(12):3035–3039.
- [45] Sood, V. and Redner, S. (2005). Voter model on heterogeneous graphs. *Physical review letters*, 94(17):178701.
- [46] Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30(3):341–366.
- [47] Tian, Y., Jia, P., Mirtabatabaei, A., Wang, L., Friedkin, N. E., and Bullo, F. (2021). Social power evolution in influence networks with stubborn individuals. *IEEE Transactions on Automatic Control*, 67(2):574–588.
- [48] Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., and Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, 25(11):913–916.
- [49] Van Der Hofstad, R. (2024). *Random graphs and complex networks*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
- [50] Weisbuch, G. and Boudjema, G. (1999). Dynamical aspects in the adoption of agri-environmental measures. *Advances in Complex Systems*, 2(01):11–36.
- [51] Yang, M., Qu, X., Cao, Z., and Yang, X. (2017). Innovation governs everything eventually: Extensions of the degroot model. *Acta Mathematicae Applicatae Sinica, English Series*, 33:35–42.

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Email address: pandreou@email.unc.edu

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Email address: molvera@email.unc.edu