Received: 14 July 2020 | Revised: 5 December 2020 | Accepted: 21 December 2020 | Published on: 14 March 2021

DOI: 10.1002/rsa.21006

RESEARCH ARTICLE

WILEY

Community detection in the sparse hypergraph stochastic block model

Soumik Pal[1] | Yizhe Zhu[2]

[1]Department of Mathematics, University of Washington, Seattle, Washington, USA
[2]Department of Mathematics, University of California, San Diego, La Jolla, 92093, California, USA

Correspondence
Yizhe Zhu, Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA.
Email: yiz084@ucsd.edu

Funding information
NSF grant, Grant/Award Numbers: DMS-1612483, DMS-1949617

Abstract

We consider the community detection problem in sparse random hypergraphs. Angelini et al. in [6] conjectured the existence of a sharp threshold on model parameters for community detection in sparse hypergraphs generated by a hypergraph stochastic block model. We solve the positive part of the conjecture for the case of two blocks: above the threshold, there is a spectral algorithm which asymptotically almost surely constructs a partition of the hypergraph correlated with the true partition. Our method is a generalization to random hypergraphs of the method developed by Massoulié (2014) for sparse random graphs.

KEYWORDS

community detection, random tensor, self-avoiding walk, sparse hypergraph, stochastic block model

1 | INTRODUCTION

Clustering is an important topic in network analysis, machine learning, and computer vision [24]. Many clustering algorithms are based on graphs, which represent pairwise relationships among data. Hypergraphs can be used to represent higher-order relationships among objects, including co-authorship and citation networks, and they have been shown empirically to have advantages over graphs [40]. Recently hypergraphs have been used as the data model in machine learning, including recommender system [38], image retrieval [5, 30] and bioinformatics [39]. The stochastic block model (SBM) is a generative model for random graphs with community structures, which serves as a useful benchmark for clustering algorithms on graph data. It is natural to have an analogous model for random hypergraphs to model higher-order relations. In this paper, we consider a higher-order SBM called the hypergraph stochastic block model (HSBM). Before describing HSBMs, let's recall clustering on graph SBMs.

1.1 | The stochastic block model for graphs

In this section, we summarize the state-of-the-art results for graph SBM with two blocks of roughly equal size. Let Σ_n be the set of all pairs (G, σ), where $G = ([n], E)$ is a graph with vertex set $[n]$ and edge set E, $\sigma = (\sigma_1, \dots, \sigma_n) \in \{+1, -1\}^n$ are spins on $[n]$, i.e., each vertex $i \in [n]$ is assigned with a spin $\sigma_i \in \{-1, +1\}$. From this finite set Σ_n, one can generate a random element (G, σ) in two steps.

1. First generate i.i.d random variables $\sigma_i \in \{-1, +1\}$ equally likely for all $i \in [n]$.
2. Then given $\sigma = (\sigma_1, \dots, \sigma_n)$, we generate a random graph G where each edge $\{i, j\}$ is included independently with probability p if $\sigma_i = \sigma_j$ and with probability q if $\sigma_i \neq \sigma_j$.

The law of this pair (G, σ) will be denoted by $\mathcal{G}(n, p, q)$. In particular, we are interested in the model $\mathcal{G}(n, p_n, q_n)$ where p_n, q_n are parameters depending on n. We use the shorthand notation $\mathbb{P}_{\mathcal{G}_n}$ to emphasize that the integration is taken under the law $\mathcal{G}(n, p_n, q_n)$.

Imagine $C_1 = \{i : \sigma_i = +1\}$ and $C_2 = \{i : \sigma_i = -1\}$ as two communities in the graph G. Observing only G from a sample (G, σ) from the distribution $\mathcal{G}(n, p_n, q_n)$, the goal of community detection is to estimate the unknown vector σ up to a sign flip. Namely, we construct label estimators $\hat{\sigma}_i \in \{\pm 1\}$ for each i and consider the empirical overlap between $\hat{\sigma}$ and unknown σ defined by

$$ov_n(\hat{\sigma}, \sigma) := \frac{1}{n} \sum_{i \in [n]} \sigma_i \hat{\sigma}_i. \tag{1.1}$$

We may ask the following questions about the estimation as n tends to infinity:

1. Exact recovery (strong consistency):

$$\lim_{n \to \infty} \mathbb{P}_{\mathcal{G}_n} \left(\{ov_n(\hat{\sigma}, \sigma) = 1\} \cup \{ov_n(\hat{\sigma}, \sigma) = -1\}\right) = 1.$$

2. Almost exact recovery (weak consistency): for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}_{\mathcal{G}_n} \left(\{|ov_n(\hat{\sigma}, \sigma) - 1| > \varepsilon\} \cap \{|ov_n(\hat{\sigma}, \sigma) + 1| > \varepsilon\}\right) = 0.$$

3. Detection: Find a partition which is correlated with the true partition. More precisely, there exists a constant $r > 0$ such that it satisfies the following: for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}_{\mathcal{G}_n}(\{|ov_n(\hat{\sigma}, \sigma) - r| > \varepsilon\} \cap \{|ov_n(\hat{\sigma}, \sigma) + r| > \varepsilon\}) = 0. \tag{1.2}$$

There are many works on these questions using different tools, we list some of them. A conjecture of [14] based on nonrigorous ideas from statistical physics predicts a threshold of detection in the SBM, which is called the Kesten–Stigum threshold. In particular, if $p_n = \frac{a}{n}$ and $q_n = \frac{b}{n}$ where a, b are positive constants independent of n, then the detection is possible if and only if $(a - b)^2 > 2(a + b)$. This conjecture was confirmed in [8, 32, 33, 35] where [8, 32, 35] provided efficient algorithms to achieve the threshold. Very recently, two alternative spectral algorithms were proposed based on distance matrices [36] and a graph powering method in [3], and they both achieved the detection threshold.

Suppose $p_n = \frac{a \log n}{n}, q_n = \frac{b \log n}{n}$ where a, b are constant independent of n. Then the exact recovery is possible if and only if $\left(\sqrt{a} - \sqrt{b}\right)^2 > 2$, which was solved in [2, 23] with efficient algorithms achieving the threshold. Besides the phase transition behavior, various algorithms were proposed and analyzed in different regimes and more general settings beyond the 2-block SBM [4, 7, 10, 11, 13, 22, 28, 34, 37], including spectral methods, semidefinite programming, belief-propagation, and approximate message-passing algorithms. We recommend [1] for further details.

1.2 | Hypergraph stochastic block models

The hypergraph stochastic block model (HSBM) is a generalization of the SBM for graphs, which was first studied in [18], where the authors consider hypergraphs generated by the stochastic block models that are dense and uniform. A faithful representation of a hypergraph is its adjacency tensor (see Definition 2.2). However, most of the computations involving tensors are NP-hard [25]. Instead, they considered spectral algorithms for exact recovery using hypergraph Laplacians. Subsequently, they extended their results to sparse, nonuniform hypergraphs [19–21]. For exact recovery, it was shown that the phase transition occurs in the regime of logarithmic average degrees in [11, 12, 29] and the exact threshold was given in [27], by a generalization of the techniques in [2]. Almost exact recovery for HSBMs was studied in [11, 12, 21].

For detection of the HSBM with two blocks, the authors of [6] proposed a conjecture that the phase transition occurs in the regime of constant average degree, based on the performance of the belief-propagation algorithm. Also, they conjectured a spectral algorithm based on nonbacktracking operators on hypergraphs could reach the threshold. In [17], the authors showed an algorithm for detection when the average degree is bigger than some constant by reducing it to a bipartite stochastic block model. They also mentioned a barrier to further improvement. We confirm the positive part of the conjecture in [6] for the case of two blocks: above the threshold, there is a spectral algorithm which asymptotically almost surely constructs a partition of the hypergraph correlated with the true partition.

Now we specify our d-uniform hypergraph stochastic block model with two clusters. Analogous to $\mathcal{G}(n, p_n, q_n)$, we define $\mathcal{H}(n, d, p_n, q_n)$ for d-uniform hypergraphs. Let Σ_n be the set of all pair (H, σ), where $H = ([n], E)$ is a d-uniform hypergraph (see Definition 2.1) with vertex set $[n]$ and hyperedge set E, $\sigma = (\sigma_1, \dots, \sigma_n) \in \{+1, -1\}^n$ are the spins on $[n]$. From this finite set Σ_n, one can generate a random element (H, σ) in two steps.

(1) First generate i.i.d random variables $\sigma_i \in \{-1, +1\}$ equally likely for all $i \in [n]$.
(2) Then given $\sigma = (\sigma_1, \dots, \sigma_n)$, we generate a random hypergraph H where each hyperedge $\{i_1, \dots i_d\}$ is included independently with probability p_n if $\sigma_{i_1} = \cdots = \sigma_{i_d}$ and with probability q_n if the spins $\sigma_{i_1}, \dots \sigma_{i_d}$ are not the same.

The law of this pair (H, σ) will be denoted by $\mathcal{H}(n, d, p_n, q_n)$. We use the shorthand notation $\mathbb{P}_{\mathcal{H}_n}$ and $\mathbb{E}_{\mathcal{H}_n}$ to emphasize that integration is taken under the law $\mathcal{H}(n, d, p_n, q_n)$. Often we drop the index n from our notation, but it will be clear from $\mathbb{P}_{\mathcal{H}_n}$.

1.3 | Main results

We consider the detection problem of the HSBM in the constant expected degree regime. Let

$$p_n := \frac{a}{\binom{n}{d-1}}, \quad q_n := \frac{b}{\binom{n}{d-1}}$$

for some constants $a \geq b > 0$ and a constant integer $d \geq 3$. Let

$$\alpha := (d-1)\frac{a + (2^{d-1} - 1)b}{2^{d-1}}, \quad \beta := (d-1)\frac{a - b}{2^{d-1}}. \tag{1.3}$$

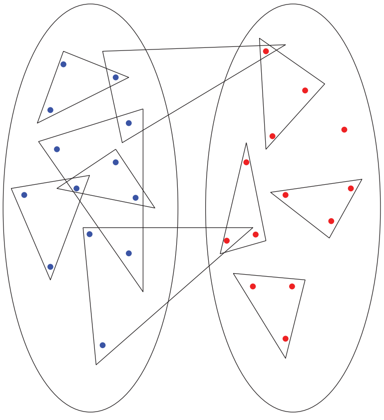

FIGURE 1 An HSBM with $d=3$. Vertices in blue and red have spin $+$ and $-$, respectively

Here α is a constant which measures the expected degree of any vertex, and β measures the discrepancy between the number of neighbors with $+$ sign and $-$ sign of any vertex. For $d=2$, α, β are the same parameters for the graph case in [32]. Now we are able to state our main result which is an extension of the result of for graph SBMs in [32]. Note that with the definition of α, β, we have $\alpha > \beta$. The condition $\beta^2 > \alpha$ in the statement of Theorem (1.1) below implies $\alpha, \beta > 1$, which will be assumed for the rest of the paper.

Theorem 1.1. *Assume $\beta^2 > \alpha$. Let (H, σ) be a random labeled hypergraph sampled from $\mathcal{H}(n, d, p_n, q_n)$ and $B^{(l)}$ be its lth self-avoiding matrix (see Definition 2.6). Set $l = c\log(n)$ for a constant c such that $c\log(\alpha) < 1/8$. Let x be a l_2-normalized eigenvector corresponding to the second largest eigenvalue of $B^{(l)}$. There exists a constant t such that, if we define the label estimator $\hat{\sigma}_i$ as*

$$\hat{\sigma}_i = \begin{cases} +1 & \text{if } x_i \geq t/\sqrt{n}, \\ -1 & \text{otherwise,} \end{cases}$$

then detection is possible. More precisely, there exists a constant $r > 0$ such that the empirical overlap between $\hat{\sigma}$ and σ defined similar to (1.1) satisfies the following: for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\{|ov_n(\hat{\sigma}, \sigma) - r| > \epsilon\} \bigcap \{|ov_n(\hat{\sigma}, \sigma) + r| > \epsilon\} \right) = 0.$$

Remark 1.2. If we take $d=2$, the condition $\beta^2 > \alpha$ is the threshold for detection in graph SBMs proved in [32, 33, 35]. When $d \geq 3$, the conjectured detection threshold for HSBMs is given in Equation (48) of [6]. With our notations, in the 2-block case, Equation (48) in [6] can be written as $\frac{\alpha-\beta}{\alpha+\beta} = \frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$, which says $\beta^2 = \alpha$ is the conjectured detection threshold for HSBMs. This is an analog of the Kesten-Stigum threshold proved in the graph case [8, 14, 32, 33, 35]. Our Theorem 1.1 proves the positive part of the conjecture.

Our algorithm can be summarized in two steps. The first step is a dimension reduction: $B^{(l)}$ has n^2 many entries from the original adjacency tensor T (see Definition 2.2) of n^d many entries. Since the l-neighborhood of any vertex contains at most one cycle with high probability (see Lemma 4.4), by breadth-first search, the matrix $B^{(l)}$ can be constructed in polynomial time. The second step is a simple spectral clustering according to leading eigenvectors as the common clustering algorithm in the graph case.

Unlike graph SBMs, in the HSBMs, the random hypergraph H we observe is essentially a random tensor. Getting the spectral information of a tensor is NP-hard [25] in general, making the corresponding problems in HSBMs very different from graph SBMs. It is not immediately clear which operator to associate to H that encodes the community structure in the bounded expected degree regime. The novelty of our method is a way to project the random tensor into matrix forms (the self-avoiding matrix $B^{(l)}$ and the adjacency matrix A) that give us the community structure from their leading eigenvectors. In practice, the hypergraphs we observed are usually not d-uniform, which cannot be represented as a tensor. However, we can still construct the matrix $B^{(l)}$ since the definition of self-avoiding walks does not depend on the uniformity assumption. In this paper, we focus on the d-uniform case to simplify the presentation, but our proof techniques can be applied to the nonuniform case.

The analysis of HSBMs is harder than the original graph SBMs due to the extra dependency in the hypergraph structure and the lack of linear algebra tools for tensors. To overcome these difficulties, new techniques are developed in this paper to establish the desired results.

There are multiple ways to define self-avoiding walks on hypergraphs, and our definition (see Definition 2.4) is the only one that works for us when applying the moment method. We develop a moment method suitable for sparse random hypergraphs in Section 7 that controls the spectral norms by counting concatenations of self-avoiding walks on hypergraphs. The combinatorial counting argument in the proof of Lemma 7.1 is more involved as we need to consider labeled vertices and labeled hyperedges. The moment method for hypergraphs developed here could be of independent interest for other random hypergraph problems.

The growth control of the size of the local neighborhood (Section 4) for HSBMs turns out to be more challenging compared to graph SBMs in [32] due to the dependency between the number of vertices with spin + and −, and overlaps between different hyperedges. We use a new second-moment estimate to obtain a matching lower bound and upper bound for the size of the neighborhoods in the proof of Theorem 8.4. The issues mentioned above do not appear in the sparse random graph case.

To analyze the local structure of HSBMs, we prove a new coupling result between a typical neighborhood of a vertex in the sparse random hypergraph H and a multi-type Galton–Watson hypertree described in Section 10, which is a stronger version of local weak convergence of sparse random hypergraphs (local weak convergence for hypergraphs was recently introduced in [15]). Compared to the classical 2-type Galton–Watson tree in the graph case, the vertex $\pm$ labels in a hyperedge is not assigned independently. We carefully designed the probability of different types of hyperedges that appear in the hypertree to match the local structure of the HSBM.

Combining all the new ingredients, we obtain the weak Ramanujan property of $B^{(l)}$ for sparse HSBMs in Theorem 6.1 as a generalization of the results in [32]. We conclude the proof of our Theorem 1.1 in Section 6.

Our Theorem 1.1 deals with the positive part of the phase transition conjecture in [6]. To have a complete characterization of the phase transition, one needs to show an impossibility result when $\beta^2 < \alpha$. Namely, below this threshold, no algorithms (even with exponential running time) will solve the detection problem with high probability. For graph SBMs, the impossibility result was proved in [33] based on a reduction to the broadcasting problem on Galton–Watson trees analyzed in [16]. To answer the corresponding problem in the HSBMs, one needs to establish a similar information-theoretical lower bound for the broadcasting problem on hypertrees and relate the problem to the detection problem on HSBMs. To the best of our knowledge, even for the very first step, the broadcasting problem on hypertrees has not been studied yet. The multi-type Galton–Watson hypertrees described in Section 10 can be used as a model to study this type of problem on hypergraphs. We leave it as a future direction.

2 | PRELIMINARIES

Definition 2.1 (Hypergraph)**.** A *hypergraph* H is a pair $H=(V,E)$ where V is a set of vertices and E is the set of nonempty subsets of V called *hyperedges*. If any hyperedge $e\in E$ is a set of d elements of V, we call H d-uniform. In particular, 2-uniform hypergraph is an ordinary graph. A d-uniform hypergraph is complete if any set of d vertices is a hyperedge and we denote a complete d-uniform hypergraph on $[n]$ by $K_{n,d}$. The *degree* of a vertex $i\in V$ is the number of hyperedges in H that contains i.

Definition 2.2 (Adjacency tensor)**.** Let $H=(V,E)$ be a d-uniform hypergraph with $V=[n]$. We define T to be the adjacency tensor of H such that for any set of vertices $\{i_1,i_2,\dots,i_d\}$,

$$T_{i_1,\dots,i_d}=\begin{cases}1 & \text{if } \{i_1,\dots,i_d\}\in E,\\ 0 & \text{otherwise.}\end{cases}$$

We set $T_{\sigma(i_1),\sigma(i_2),\dots,\sigma(i_d)}=T_{i_1,\dots,i_d}$ for any permutation σ. We may write T_e in place of $T_{i_1,\dots,i_d}$ where $e=\{i_1,\dots,i_d\}$.

Definition 2.3 (Adjacency matrix)**.** The adjacency matrix A of a d-uniform hypergraph $H=(V,E)$ with vertex set $[n]$ is a $n\times n$ symmetric matrix such that for any $i\neq j$, A_{ij} is the number of hyperedges in E which contains i,j and $A_{ii}=0$ for $i\in[n]$. Equivalently, we have

$$A_{ij}=\begin{cases}\sum_{e:\{i,j\}\in e}T_e & \text{if } i\neq j,\\ 0 & \text{if } i=j.\end{cases}$$

Definition 2.4 (Walk)**.** A *walk* of length l on a hypergraph H is a sequence $(i_0,e_1,i_1,\dots,e_l,i_l)$ such that $i_{j-1}\neq i_j$ and $\{i_{j-1},i_j\}\subset e_j$ for all $1\le j\le l$. A walk is closed if $i_0=i_l$ and we call it a *circuit*. A *self-avoiding walk* of length l is a walk $(i_0,e_1,i_1,\dots,e_l,i_l)$ such that

(1) $|\{i_0,i_1,\dots,i_l\}|=l+1$.
(2) Any consecutive hyperedges e_{j-1},e_j satisfy $e_{j-1}\cap e_j=\{i_{j-1}\}$ for $2\le j\le l$.
(3) Any two hyperedges e_j,e_k with $1\le j<k\le l, k\neq j+1$ satisfy $e_j\cap e_k=\emptyset$.

See Figure 2 for an example of a self-avoiding walk in a 3-uniform hypergraph. Recall that a self-avoiding walk of length l on a graph is a walk $(i_0,\dots,i_l)$ without repeated vertices. Our definition is a generalization of the self-avoiding walk to hypergraphs.

Definition 2.5 (Cycle and hypertree)**.** A *cycle* of length l with $l\ge 2$ in a hypergraph H is a walk $(i_0,e_1,\dots,i_{l-1},e_l,i_0)$ such that $i_0,\dots i_{l-1}$ are distinct vertices and $e_1\dots e_l$ are distinct hyperedges. A *hypertree* is a hypergraph which contains no cycles.

Let $\binom{[n]}{d}$ be the collection of all subsets of $[n]$ with size d. For any subset $e\in\binom{[n]}{d}$ and $i\neq j\in[n]$, we define

$$A^e_{ij}=\begin{cases}1 & \text{if } \{i,j\}\in e \text{ and } e\in E,\\ 0 & \text{otherwise,}\end{cases}$$

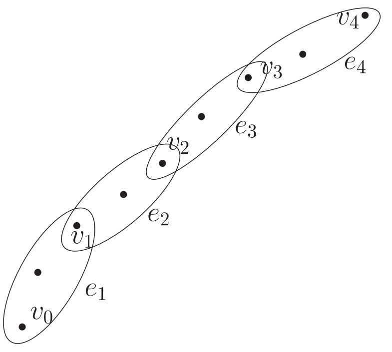

FIGURE 2 A self-avoiding walk of length 4 denoted by $(v_0, e_1, v_1, e_2, v_2, e_3, v_3, e_4, v_4)$

and we define $A_{ii}^e = 0$ for all $i \in [n]$. With our notation above, $A_{ij} = \sum_{e \in \binom{[n]}{d}} A_{ij}^e$. We have the following expansion of the trace of A^k for any integer $k \geq 0$:

$$\mathrm{tr} A^k = \sum_{i_0, i_2, \dots, i_{k-1} \in [n]} A_{i_0 i_1} A_{i_2 i_3} \dots A_{i_{k-1} i_0} = \sum_{\substack{i_0, i_1, \dots, i_{k-1} \in [n] \\ e_1, \dots, e_k \in \binom{[n]}{d}}} A_{i_0 i_1}^{e_1} \dots A_{i_{k-2} i_{k-1}}^{e_{k-1}} A_{i_{k-1} i_0}^{e_k}.$$

Therefore, $\mathrm{tr}A^k$ counts the number of circuits $(i_0, e_1, i_1, \dots, i_{k-1}, e_k, i_0)$ in the hypergraph H of length k. This connection was used in [31] to study the spectra of the Laplacian of random hypergraphs.

From our definition of self-avoiding walks on hypergraphs, we associate a self-avoiding adjacency matrix to the hypergraph.

Definition 2.6 (Self-avoiding matrix)**.** Let $H = (V, E)$ be a hypergraph with $V = [n]$. For any $l \geq 1$, a lth *self-avoiding matrix* $B^{(l)}$ is a $n \times n$ matrix where for $i \neq j \in [n]$, $B_{ij}^{(l)}$ counts the number of self avoiding walks of length l from i to j and $B_{ii}^{(l)} = 0$ for $i \in [n]$.

$B^{(l)}$ is a symmetric matrix since a time-reversing self avoiding walk from i to j is a self avoiding walk from j to i. Let SAW_{ij} be the set of all self-avoiding walks of length l connecting i and j in the complete d-uniform hypergraph on vertex set $[n]$. We denote a walk of length l by $w = (i_0, e_{i_1}, \dots, i_{l-1}, e_{i_l}, i_l)$. Then for any $i, j \in [n]$,

$$B_{ij}^{(l)} = \sum_{w \in \mathrm{SAW}_{ij}} \prod_{t=1}^{l} A_{i_{t-1} i_t}^{e_{i_t}}. \tag{2.1}$$

3 | MATRIX EXPANSION AND SPECTRAL NORM BOUNDS

Consider a random labeled d-uniform hypergraph H sampled from $\mathcal{H}(n, d, p_n, q_n)$ with adjacency matrix A and self-avoiding matrix $B^{(l)}$. Let $\overline{A} := \mathbb{E}_{\mathcal{H}_n}[A|\sigma]$. Let $\rho(A) := \sup_{x: \|x\|_2 = 1} \|Ax\|_2$ be the spectral norm of a matrix A. Recall (2.1), define

$$\Delta_{ij}^{(l)} := \sum_{w \in \mathrm{SAW}_{ij}} \prod_{t=1}^{l} (A_{i_{t-1} i_t}^{e_{i_t}} - \overline{A}_{i_{t-1} i_t}^{e_{i_t}}), \tag{3.1}$$

where $\overline{A}_{i_{t-1} i_t}^{e_{i_t}} = \mathbb{E}_{\mathcal{H}_n}[A_{i_{t-1} i_t}^{e_{i_t}} | \sigma]$. $\Delta^{(l)}$ can be regarded as a centered version of $B^{(l)}$.

We will apply the classical moment method to estimate the spectral norm of $\Delta^{(l)}$, since this method works well for centered random variables. Then we can relate the spectrum of $\Delta^{(l)}$ to the spectrum of $B^{(l)}$ through a matrix expansion formula which connects $\overline{A}$, $B^{(l)}$ and $\Delta^{(l)}$ in the following theorem. Recall the definition of α in (1.3).

Theorem 3.1. *Let H be a random hypergraph sampled from $\mathcal{H}(n, d, p_n, q_n)$ and $B^{(l)}$ be its lth self avoiding matrix. Then the following holds.*

1. *There exist some matrices $\{\Gamma^{(l,m)}\}_{m=1}^{l}$ such that for any $l \geq 1$, $B^{(l)}$ satisfies the identity*

$$B^{(l)} = \Delta^{(l)} + \sum_{m=1}^{l} (\Delta^{(l-m)} \overline{A} B^{(m-1)}) - \sum_{m=1}^{l} \Gamma^{(l,m)}. \tag{3.2}$$

2. *For any sequence $l_n = O(\log n)$ and any fixed $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n} \left(\rho(\Delta^{(l_n)}) \leq n^{\varepsilon} \alpha^{l_n/2} \right) = 1, \tag{3.3}$$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n} \left(\bigcap_{m=1}^{l_n} \left\{ \rho(\Gamma^{(l_n,m)}) \leq n^{\varepsilon-1} \alpha^{(l_n+m)/2} \right\} \right) = 1. \tag{3.4}$$

Theorem 3.1 is one of the main ingredients to show $B^{(l)}$ has a spectral gap. Together with the local analysis in Section 4, we will show in Theorem 6.1 that the bulk eigenvalues of $B^{(l)}$ are separated from the first and second eigenvalues. The proof of Theorem 3.1 is deferred to Section 7. The matrices $\{\Gamma^{(l,m)}\}_{m=1}^{l}$ in Theorem 3.1 record concatenations of self-avoiding walks with different weights, which will be carefully analyzed in Lemma 7.2 of Section 7.

4 | LOCAL ANALYSIS

In this section, we study the structure of the local neighborhoods in the HSBM. Namely, what the neighborhood of a typical vertex in the random hypergraph looks like.

Definition 4.1. In a hypergraph H, we define the *distance* $d(i,j)$ between two vertices i, j to be the minimal length of walks between i and j. Define the *t-neighborhood* $V_t(i)$ of a fixed vertex i to be the set of vertices which have distance t from i. Define $V_{\leq t}(i) := \bigcup_{k\leq t} V_k(i)$ to be the set all of vertices which have distance at most t from i and $V_{>t} = [n] \setminus V_{\leq t}$. Let $V_t^{\pm}(i)$ be the vertices in $V_t(i)$ with spin $\pm$ and define it similarly for $V_{\leq t}^{\pm}(i)$.

For $i \in [n]$, define

$$S_t(i) := |V_t(i)|, \quad D_t(i) := \sum_{j:d(i,j)=t} \sigma_j.$$

Let $\mathbf{1} = (1 \dots, 1) \in \mathbb{R}^n$ and recall $\sigma \in \{-1, 1\}^n$. We will show that when $l = c \log n$ with $c \log \alpha < 1/8$, $S_l(i), D_l(i)$ are close to the corresponding quantities $(B^{(l)}\mathbf{1})_i$, $(B^{(l)}\sigma)_i$ (see Lemma 11.1). In particular, the vector $(D_l(i))_{1\leq i\leq n}$ is asymptotically aligned with the second eigenvector of $B^{(l)}$, from which we get the information on the partitions.

We give the following growth estimates of $S_t(i)$ and $D_t(i)$. The proof of Theorem 4.2 is given in Section 8.

Theorem 4.2. *Assume $\beta^2 > \alpha > 1$ and $l = c\log n$, for a constant c such that $c\log\alpha < 1/4$. There exists constants $C, \gamma > 0$ such that for sufficiently large n, with probability at least $1 - O(n^{-\gamma})$ the following holds for all $i \in [n]$ and $1 \le t \le l$:*

$$S_t(i) \le C\log(n)\alpha^t, \tag{4.1}$$

$$|D_t(i)| \le C\log(n)\beta^t, \tag{4.2}$$

$$S_t(i) = \alpha^{t-l}S_l(i) + O(\log(n)\alpha^{t/2}), \tag{4.3}$$

$$D_t(i) = \beta^{t-l}D_l(i) + O(\log(n)\alpha^{t/2}). \tag{4.4}$$

The approximate independence of neighborhoods of distinct vertices is given in the following lemma. It will be used later to analyze the martingales constructed on the Galton–Watson hypertree defined in Section 10. The proof of Lemma 4.3 is given in Appendix A.1.

Lemma 4.3. *For any two fixed vertices $i \neq j$, let $l = c\log(n)$ with constant $c\log(\alpha) < 1/4$. Then the total variation distance between the joint law $\mathcal{L}((U_k^{\pm}(i))_{k\le l}, (U_k^{\pm}(j))_{k\le l})$ and the law with the same marginals and independence between them, denoted by $\mathcal{L}((U_k^{\pm}(i))_{k\le l} \otimes (U_k^{\pm}(j))_{k\le l})$, is $O(n^{-\gamma})$ for some $\gamma > 0$.*

Now we consider number of cycles in $V_{\le l}(i)$ of any vertex $i \in [n]$. We say H is *l-tangle-free* if for any $i \in [n]$, there is no more than one cycle in $V_{\le l}(i)$.

Lemma 4.4. *Assume $l = c\log n$ with $c\log(\alpha) < 1/4$. Let (H, σ) be a sample from $\mathcal{H}(n, d, p_n, q_n)$. Then*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(|\{i \in [n] : V_{\le l}(i) \text{ contains at least one cycle}\}| \le \log^4(n)\alpha^{2l}\right) = 1,$$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}(H \text{ is } l\text{-tangle-free}) = 1.$$

The proof of Lemma 4.4 is given in Appendix A.2.

In the next lemma, we translate the local analysis of the neighborhoods to the control of vectors $B^{(m)}\mathbf{1}, B^{(m)}\sigma$. The proof is similar to the proof of Lemma 4.3 in [32], and we include it in Appendix A0.3. For any event A_n, we say A_n happens *asymptotically almost surely* if $\lim_{n\to\infty}\mathbb{P}_{\mathcal{H}_n}(A_n) = 1$.

Lemma 4.5. *Let $\mathcal{B}$ be the set of vertices i whose $l-$neighborhood contains a cycle. For $l = c\log n$ with $c\log(\alpha) < 1/4$, asymptotically almost surely the following holds:*

(1) for all $m \le l$ and all $i \notin \mathcal{B}$ the following holds

$$(B^{(m-1)}\mathbf{1})_i = \alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{(m-1)/2}\log n), \tag{4.5}$$

$$(B^{(m-1)}\sigma)_i = \beta^{m-1-l}(B^{(l)}\sigma)_i + O(\alpha^{(m-1)/2}\log n). \tag{4.6}$$

(2) For all $i \in \mathcal{B}$:

$$|(B^{(m)}\sigma)_i| \le |(B^{(m)}\mathbf{1})_i| \le 2\sum_{t=0}^{m} S_t(i) = O(\alpha^m \log n). \tag{4.7}$$

Combining Theorems 3.1,4.2, and Lemma 4.5, we are able to prove the following theorem.

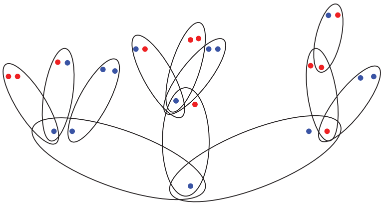

FIGURE 3 A Galton–Watson hypertree with $d=3$. The vertices with spin + are in blue and vertices with spin − are in red

Theorem 4.6. *Assume $\beta^2 > \alpha > 1$ and $l = c\log n$ with $c\log(\alpha) < 1/8$. Then the following holds: for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\sup_{\|x\|_2=1, x^\top(B^{(l)}\mathbf{1})=x^\top(B^{(l)}\sigma)=0} \|B^{(l)}x\|_2 \leq n^{\varepsilon}\alpha^{l/2}\right) = 1.$$

Theorem 4.6 is a key ingredient to prove the bulk eigenvalues of $B^{(l)}$ are $O(n^{\varepsilon}\alpha^{l/2})$ in Theorem 6.1. The proof of Theorem 4.6 is given in in Section 9.

5 | COUPLING WITH MULTI-TYPE POISSON HYPERTREES

Recall the definition of a hypertree from Definition 2.5. We construct a hypertree growth process in the following way. The hypertree is designed to obtain a coupling with the local neighborhoods of the random hypergraph H.

- Generate a root ρ with spin $\tau(\rho) = +$, then generate Pois$\left(\frac{\alpha}{d-1}\right)$ many hyperedges that only intersects at ρ. Call the vertices in these hyperedges except ρ to be the *children* of ρ and of generation 1. Call ρ to be their *parent*.
- For $0 \leq r \leq d-1$, we define a hyperedge is of type r if r many children in the hyperedge has spin $\tau(\rho)$ and $(d-1-r)$ many children has spin $-\tau(\rho)$. We first assign a type for each hyperedge independently. Each hyperedge will be of type $(d-1)$ with probability $\frac{(d-1)a}{\alpha 2^{d-1}}$ and of type r with probability $\frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}}$ for $0 \leq r \leq d-2$. Since $\frac{(d-1)a}{\alpha 2^{d-1}} + \sum_{r=0}^{d-2} \frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}} = 1$, the probabilities of being various types of hyperedges add up to 1. Because the type is chosen i.i.d for each hyperedge, by Poisson thinning, the number of hyperedges of different types are independent and Poisson.
- We draw the hypertree in a plane and label each child from left to right. For each type r hyperedge, we uniformly randomly pick r vertices among $d-1$ vertices in the first generation to put spins $\tau(\rho)$, and the rest $d-1-r$ many vertices are assigned with spins $-\tau(\rho)$.
- After defining the first generation, we keep constructing subsequent generations by induction. For each children v with spin $\tau(v)$ in the previous generation, we generate Pois$\left(\frac{\alpha}{d-1}\right)$ many hyperedges that pairwise intersects at v and assign a type to each hyperedge by the same rule with $\tau(\rho)$ replaced by $\tau(v)$. We call such random hypergraphs with spins a *multi-type Galton–Watson hypertree*, denoted by (T, ρ, τ) (see Figure 3).

Let $W_t^{\pm}$ be the number of vertices with $\pm$ spins at the tth generation and $W_t^{(r)}$ be the number of hyperedges which contains exactly r children with spin + in the tth generation. Let $\mathcal{G}_{t-1} := \sigma(W_k^{\pm}, 1 \leq$

$k \le t-1$) be the σ-algebra generated by $W_k^{\pm}, 1 \le k \le t-1$. From our definition, $W_0^+ = 1, W_0^- = 0$ and $\{W_t^{(r)}\}_{0 \le r \le d-1}$ are independent conditioned on $\mathcal{G}_{t-1}$, and the conditioned laws of $W_t^{(r)}$ are given by

$$\mathcal{L}(W_t^{(d-1)}|\mathcal{G}_{t-1}) = \mathrm{Pois}\left(\frac{a}{2^{d-1}}W_{t-1}^+ + \frac{b}{2^{d-1}}W_{t-1}^-\right), \tag{5.1}$$

$$\mathcal{L}(W_t^{(0)}|\mathcal{G}_{t-1}) = \mathrm{Pois}\left(\frac{a}{2^{d-1}}W_{t-1}^- + \frac{b}{2^{d-1}}W_{t-1}^+\right), \tag{5.2}$$

$$\mathcal{L}(W_t^{(r)}|\mathcal{G}_{t-1}) = \mathrm{Pois}\left(\frac{b\binom{d-1}{r}}{2^{d-1}}(W_{t-1}^- + W_{t-1}^+)\right), \quad 1 \le r \le d-2. \tag{5.3}$$

We also have

$$W_t^+ = \sum_{r=0}^{d-1} r W_t^{(r)}, \quad W_t^- = \sum_{r=0}^{d-1}(d-1-r)W_t^{(r)}. \tag{5.4}$$

Definition 5.1. A *rooted hypergraph* is a hypergraph H with a distinguished vertex $i \in V(H)$, denoted by (H, i). We say two rooted hypergraphs (H, i) and (H', i') are *isomorphic* and if and only if there is a bijection $\phi : V(H) \to V(H')$ such that $\phi(i) = i'$ and $e \in E(H)$ if and only if $\phi(e) := \{\phi(j) : j \in e\} \in E(H')$.

Let (H, i, σ) be a rooted hypergraph with root i and each vertex j is given a spin $\sigma(j) \in \{-1, +1\}$. Let (H', i', σ') be a rooted hypergraph with root i' where for each vertex $j \in V(H')$, a spin $\sigma'(j) \in \{-1, +1\}$ is given. We say (H, i, σ) and (H', i', σ') are *spin-preserving isomorphic* and denoted by $(H, i, \sigma) \equiv (H', i', \sigma')$ if and only if there is an isomorphism $\phi : (H, i) \to (H', i')$ with $\sigma(v) = \sigma'(\phi(v))$ for each $v \in V(H)$.

Let $(H, i, \sigma)_t, (T, \rho, \tau)_t$ be the rooted hypergraphs $(H, i, \sigma), (T, \rho, \tau)$ truncated at distance t from i, ρ respectively, and let $(T, \rho, -\tau)$ be the corresponding hypertree growth process where the root ρ has spin -1. We prove a local weak convergence of a typical neighborhood of a vertex in the hypergraph H to the hypertree process T we described above. In fact, we prove the following stronger statement. The proof of Theorem 5.2 is given in Section 10.

Theorem 5.2. *Let (H, σ) be a random hypergraph H with spin σ sampled from $\mathcal{H}_n$. Let $i \in [n]$ be fixed with spin σ_i. Let $l = c\log(n)$ with $c\log(\alpha) < 1/4$, the following holds for sufficiently large n.*

1. *If $\sigma_i = +1$, there exists a coupling between (H, i, σ) and (T, ρ, τ) such that $(H, i, \sigma)_l \equiv (T, \rho, \tau)_l$ with probability at least $1 - n^{-1/5}$.*
2. *If $\sigma_i = -1$, there exists a coupling between (H, i, σ) and $(T, \rho, -\tau)$ such that $(H, i, \sigma)_l \equiv (T, \rho, -\tau)_l$ with probability at least $1 - n^{-1/5}$.*

Now we construct two martingales from the Poisson hypertree growth process. Define two processes

$$M_t := \alpha^{-t}(W_t^+ + W_t^-), \quad \Delta_t := \beta^{-t}(W_t^+ - W_t^-).$$

Lemma 5.3. *The two processes $\{M_t\}, \{\Delta_t\}$ are $\mathcal{G}_t$-martingales. If $\beta^2 > \alpha > 1$, $\{M_t\}$ and $\{\Delta_t\}$ are uniformly integrable. The martingale $\{\Delta_t\}$ converges almost surely and in L^2 to a unit mean random* *variable Δ_∞. Moreover, Δ_∞ has a finite variance and*

$$\lim_{t\to\infty} \mathbb{E}|\Delta_t^2 - \Delta_\infty^2| = 0. \tag{5.5}$$

The following Lemma will be used in the proof of Theorem 1.1 to analyze the correlation between the estimator we construct and the correct labels of vertices based on the random variable Δ_∞. The proof is similar to the proof of Theorem 4.2 in [32], and we include it in Appendix A.5.

Lemma 5.4. *Let $l = c\log n$ with $c\log\alpha < 1/8$. For any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left| \frac{1}{n}\sum_{i=1}^{n} \beta^{-2l} D_l^2(i) - \mathbb{E}[\Delta_\infty^2] \right| > \varepsilon \right) = 0. \tag{5.6}$$

Let $y^{(n)} \in \mathbb{R}^n$ be a random sequence of l_2-normalized vectors defined by

$$y_i^{(n)} := \frac{D_l(i)}{\sqrt{\sum_{j=1}^{n} D_l(j)^2}}, 1 \le i \le n.$$

Let $x^{(n)}$ be any sequence of random vectors in $\mathbb{R}^n$ such that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}(\|x^{(n)} - y^{(n)}\|_2 > \varepsilon) = 0.$$

For all $\tau \in \mathbb{R}$ that is a point of continuity of the distribution of both Δ_∞ and $-\Delta_\infty$, for any $\varepsilon > 0$, one has the following

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left| \frac{1}{n}\sum_{i\in[n]:\sigma_i=+} \mathbf{1}\left\{ x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]} \right\} - \frac{1}{2}\mathbb{P}(\Delta_\infty \ge \tau) \right| > \varepsilon \right) = 0,$$
$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left| \frac{1}{n}\sum_{i\in[n]:\sigma_i=-} \mathbf{1}\left\{ x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]} \right\} - \frac{1}{2}\mathbb{P}(-\Delta_\infty \ge \tau) \right| > \varepsilon \right) = 0. \tag{5.7}$$

6 | PROOF OF THE MAIN RESULT

Let $\vec{S}_l := (S_l(1), \dots, S_l(n))$ and $\vec{D}_l := (D_l(1), \dots, D_l(n))$. We say the the sequence of vectors $\{v_n\}_{\ge 1}$ is *asymptotically aligned* with the sequence of vectors $\{w_n\}_{n\ge 1}$ if

$$\lim_{n\to\infty} \frac{|\langle v_n, w_n\rangle|}{\|v_n\|_2 \cdot \|w_n\|_2} = 1.$$

With all the ingredients in Sections 3–10, we establish the following weak Ramanujan property of $B^{(l)}$. The proof of Theorem 6.1 is given in Section 11.

Theorem 6.1. *For $l = c\log(n)$ with $c\log(\alpha) < 1/8$, asymptotically almost surely the two leading eigenvectors of $B^{(l)}$ are asymptotically aligned with vectors $\vec{S}_l, \vec{D}_l$, where the first eigenvalue is of order $\Theta(\alpha^l)$ up to some logarithmic factor and the second eigenvalue is of order $\Omega(\beta^l)$. All other eigenvalues are of order $O(n^\varepsilon \alpha^{l/2})$ for any $\varepsilon > 0$.*

Theorem 6.1 connects the leading eigenvectors of $B^{(l)}$ with the local structures of the random hypergraph H and shows that the bulk eigenvalues of $B^{(l)}$ are separated from the two top eigenvalues. Equipped with Theorem 6.1 and Lemma 5.4, we are ready to prove our main result.

Proof of Theorem 1.1. Let $x^{(n)}$ be the l_2-normalized second eigenvector of $B^{(l)}$, by Theorem 6.1, $x^{(n)}$ is asymptotically aligned with the l_2-normalized vector

$$y_i^{(n)} = \frac{D_l(i)}{\sqrt{\sum_{j=1}^{n} D_l(j)^2}}, 1 \le i \le n$$

asymptotically almost surely. So we have $\|x^{(n)} - y^{(n)}\|_2 \to 0$ or $\|x^{(n)} + y^{(n)}\|_2 \to 0$ asymptotically almost surely.

We first assume $\|x^{(n)} - y^{(n)}\|_2 \to 0$. Since $\mathbb{E}\Delta_\infty = 1$, from the proof of Theorem 2.1 in [32], there exists a point $\tau \in \mathbb{R}$, in the set of continuity points of both Δ_∞ and $-\Delta_\infty$, that satisfies $r := \mathbb{P}(\Delta_\infty \ge \tau) - \mathbb{P}(-\Delta_\infty \ge \tau) > 0$. Take $t = \tau/\sqrt{\mathbb{E}(\Delta_\infty^2)}$ and let $\mathcal{N}^+, \mathcal{N}^-$ be the set of vertices with spin $+$ and $-$, respectively.

From the definition of $\hat{\sigma}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{i\in[n]}\sigma_i\hat{\sigma}_i =& \frac{1}{n}\sum_{i\in[n]}\sigma_i\left(\mathbf{1}_{\left\{x_i^{(n)}\ge t/\sqrt{n}\right\}} - \mathbf{1}_{\left\{x_i^{(n)}<t/\sqrt{n}\right\}}\right)\\ =& -\frac{1}{n}\sum_{i\in[n]}\sigma_i + \frac{2}{n}\sum_{i\in\mathcal{N}^+}\mathbf{1}_{\left\{x_i^{(n)}\ge\tau/\sqrt{n\mathbb{E}\Delta_\infty^2}\right\}} - \frac{2}{n}\sum_{i\in\mathcal{N}^-}\mathbf{1}_{\left\{x_i^{(n)}\ge\tau/\sqrt{n\mathbb{E}\Delta_\infty^2}\right\}}.\end{aligned} \tag{6.1}$$

Note that $\frac{1}{n}\sum_{i\in[n]}\sigma_i \to 0$ in probability by the law of large numbers. From (5.7) in Lemma 5.4, we have (6.1) converges in probability to $\mathbb{P}(\Delta_\infty \ge \tau) - \mathbb{P}(-\Delta_\infty \ge \tau) = r$. If $\|x^{(n)} + y^{(n)}\|_2 \to 0$, similarly we have $\frac{1}{n}\sum_{i\in[n]}\sigma_i\hat{\sigma}_i$ converges to $-r$ in probability. From these two cases, for any $\varepsilon > 0$,

$$\lim_{n\to\infty}\mathbb{P}_{\mathcal{H}_n}\left(\{|ov_n(\hat{\sigma},\sigma) - r| > \varepsilon\}\bigcap\{|ov_n(\hat{\sigma},\sigma) + r| > \varepsilon\}\right) = 0.$$

This concludes the proof of Theorem 1.1. ■

7 | PROOF OF THEOREM 3.1

7.1 | Proof of (3.2) in Theorem 3.1

For ease of notation, we drop the index n from l_n in the proof, and it will be clear from the law $\mathcal{H}_n$. For any sequences of real numbers $\{a_t\}_{t=1}^l, \{b_t\}_{t=1}^l$, we have the following expansion identity for $l \ge 2$ (see for example, Equation (15) in [32] and Equation (27) in [8]):

$$\prod_{t=1}^{l}(a_t - b_t) = \prod_{t=1}^{l}a_t - \sum_{m=1}^{l}\left(\prod_{t=1}^{l-m}(a_t - b_t)\right)b_{l-m+1}\prod_{t=l-m+2}^{l}a_t.$$

Therefore the following identity holds.

$$\prod_{t=1}^{l}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) = \prod_{t=1}^{l}A_{i_{t-1}i_t}^{e_{i_t}} - \sum_{m=1}^{l}\left(\prod_{t=1}^{l-m}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right)\overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}}\prod_{t=l-m+2}^{l}A_{i_{t-1}i_t}^{e_{i_t}}.$$

Summing over all $w \in \mathrm{SAW}_{ij}$, $\Delta_{ij}^{(l)}$ can be written as

$$B_{ij}^{(l)} - \sum_{m=1}^{l} \sum_{w \in \mathrm{SAW}_{ij}} \left(\prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \right) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}. \tag{7.1}$$

Introduce the set Q_{ij}^m of walks w defined by concatenations of two self-avoiding walks w_1, w_2 such that w_1 is a self-avoiding walk of length $l-m$ from i to some vertex k, and w_2 is a self-avoiding walk of length m from k to j for all possible $1 \le m \le l$ and $k \in [n]$. Then $\mathrm{SAW}_{ij} \subset Q_{ij}^m$ for all $1 \le m \le l$. Let $R_{ij}^m = Q_{ij}^m \setminus \mathrm{SAW}_{ij}$. Define the matrix $\Gamma^{(l,m)}$ as

$$\Gamma_{ij}^{(l,m)} := \sum_{w \in R_{ij}^m} \prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}. \tag{7.2}$$

From (7.1), $\Delta_{ij}^{(l)}$ can be expanded as

$$B_{ij}^{(l)} - \sum_{m=1}^{l} \sum_{w \in Q_{ij}^m \setminus R_{ij}^m} \left(\prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \right) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}.$$

It can be further written as

$$B_{ij}^{(l)} - \sum_{m=1}^{l} \sum_{w \in Q_{ij}^m} \prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}} + \sum_{m=1}^{l} \Gamma_{ij}^{(l,m)}.$$

From the definition of matrix multiplication, we have

$$\begin{aligned} &\sum_{w \in Q_{ij}^m} \prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}} \\ &= \sum_{1 \le u,v \le n} \Delta_{iu}^{(l-m)} \overline{A}_{uv} B_{vj}^{(m-1)} = \left(\Delta^{(l-m)} \overline{A} B^{(m-1)} \right)_{ij}. \end{aligned} \tag{7.3}$$

Combining the expansion of $\Delta_{ij}^{(l)}$ above and (7.3), we obtain

$$\Delta_{ij}^{(l)} = B_{ij}^{(l)} - \sum_{m=1}^{l} (\Delta^{(l-m)} \overline{A} B^{(m-1)})_{ij} + \sum_{m=1}^{l} \Gamma_{ij}^{(l,m)}. \tag{7.4}$$

Since (7.4) is true for any $i, j \in [n]$, it implies (3.2).

7.2 | Proof of (3.3) in Theorem 3.1

We first prove the following spectral norm bound on $\Delta^{(l)}$.

Lemma 7.1. *For $l = O(\log n)$ and fixed k, we have*

$$\mathbb{E}_{\mathcal{H}_n}[\rho(\Delta^{(l)})^{2k}] = O(n\alpha^{kl} \log^{6k} n). \tag{7.5}$$

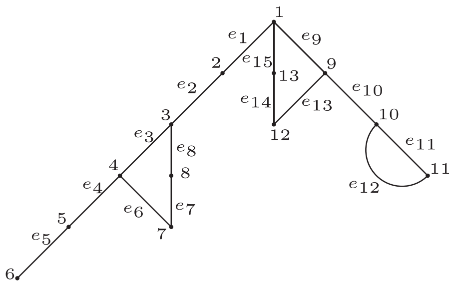

FIGURE 4 A multigraph $G(w)$ associated to a circuit $w = (w_1, \dots, w_4)$ of length $2kl$ with $k = 2, l = 5$. $w_1 = (1, e_1, 2, e_2, 3, e_3, 4, e_4, 5, e_5, 6)$, $w_2 = (6, e_5, 5, e_4, 4, e_6, 7, e_7, 8, e_8, 3)$, $w_3 = (3, e_2, 2, e_1, 1, e_9, 9, e_{10}, 10, e_{11}, 11)$, $w_4 = (11, e_{12}, 10, e_{10}, 9, e_{13}, 12, e_{14}, 13, e_{15}, 1)$. Edges that are not included in $T(w)$ are $\{e_8, e_{12}, e_{15}\}$. The triplet sequences associated to the 4 self-avoiding walks $\{w_i\}_{i=1}^4$ are given by $(0, 6, 0)$; $(4, 2, 3), (0, 0, 0)$; $(1, 3, 0)$; $(0, 0, 10), (9, 2, 1), (0, 0, 0)$, respectively.

Proof. Note that $\mathbb{E}_{\mathcal{H}_n}[\rho(\Delta^{(l)})^{2k}] \leq \mathbb{E}_{\mathcal{H}_n}[\mathrm{tr}(\Delta^{(l)})^{2k}]$. The estimation is based on a coding argument, and we modify the proof in [32] to count circuits in hypergraphs. Let $W_{2k,l}$ be the set of all circuits of length $2kl$ in the complete hypergraph $K_{n,d}$ which are concatenations of $2k$ many self-avoiding walks of length l. For any circuits $w \in W_{2k,l}$, we denote it by $w = (i_0, e_{i_1}, i_1, \dots e_{i_{2kl}}, i_{2kl})$, with $i_{2kl} = i_0$. From (3.1), we have

$$\mathbb{E}_{\mathcal{H}_n}\left[\mathrm{tr}(\Delta^{(l)})^{2k}\right] = \sum_{j_1, \dots j_{2k} \in [n]} \mathbb{E}_{\mathcal{H}_n}\left[\Delta^{(l)}_{j_1 j_2} \Delta^{(l)}_{j_2 j_3} \dots \Delta^{(l)}_{j_{2k} j_1}\right] = \sum_{w \in W_{2k,l}} \mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl} (A^{e_{i_t}}_{i_{t-1} i_t} - \overline{A}^{e_{i_t}}_{i_{t-1} i_t})\right]. \tag{7.6}$$

For each circuit, the weight it contributes to the sum is the product of $(A^e_{ij} - \overline{A^e_{ij}})$ over all the hyperedges e traversed in the circuits. In order to have an upper bound on $\mathbb{E}_{\mathcal{H}_n}[\mathrm{tr}(\Delta^{(l)})^{2k}]$, we need to estimate how many such circuits are included in the sum and what are the weights they contribute.

We also write $w = (w_1, w_2, \dots w_{2k})$, where each w_i is a self-avoiding walk of length l. Let v and h be the number of distinct vertices and hyperedges traversed by the circuit, respectively. The idea is to bound the number of all possible circuits w in (7.6) with given v and h, and then sum over all possible (v, h) pairs.

Fix v and h, for any circuit w we form a labeled multigraph $G(w)$ with labeled vertices $\{1, \dots, v\}$ and labeled multiple edges $\{e_1, \dots, e_h\}$ by the following rules:

- Label the vertices in $G(w)$ by the order they first appear in w, starting from 1. For any pair vertices $i, j \in [v]$, we add an edge between i, j in $G(w)$ whenever a hyperedge appears between the ith and jth distinct vertices in the circuit w. $G(w)$ is a multigraph since it is possible that for some i, j, there exists two distinct hyperedges connecting the ith and jth distinct vertices in w, which corresponds to two distinct edges in $G(w)$ connecting i, j.
- Label the edges in $G(w)$ by the order in which the corresponding hyperedge first appears in w from e_1 to e_h. Note that the number of edges in $G(w)$ is at least h since distinct edges in $G(w)$ can get the same hyperedge labels. At the end we obtain a multigraph $G(w) = (V(w), E(w))$ with vertex set $\{1, \dots, v\}$ and edge set $E(w)$ with hyperedge labels in $\{e_1, \dots e_h\}$.

It is crucial to see that the labeling of vertices and edges in $G(w)$ is in order, and it tells us how the circuit w is traversed. Consider any edge in $G(w)$ such that its right endpoint (in the order of the traversal of w) is a new vertex that has not been traversed by w. We call it a *tree edge*. Denote by $T(w)$ the tree spanned by those edges. It is clear for the construction that $T(w)$ includes all vertices in $G(w)$, so $T(w)$ is a spanning tree of $G(w)$. Since the labels of vertices and edges are given in $G(w)$, $T(w)$ is uniquely defined. See Figure 4 for an example.

For a given $w \in W_{2k,l}$ with distinct hyperedges $e_1, \dots, e_h$, define $\mathrm{end}(e_i)$ to be the set of vertices in $V(w)$ such that they are the endpoints of edges with label e_i in $G(w)$. For example, consider a hyperedge $e_1 = \{1,2,3,4\}$ such that $\{1,2\}, \{1,3\}$ are all the edges in $G(w)$ with labels e_1, then $\mathrm{end}(e_1) = \{1,2,3\}$. We consider circuits w in three different cases and estimate their contribution to (7.6) separately.

Case (1). We first consider $w \in W_{2k,l}$ such that

- each hyperedge label in $\{e_i\}_{1 \le i \le h}$ appears exactly once on the edges of $G(w)$;
- vertices in $e_i \setminus \mathrm{end}(e_i)$ are all distinct for $1 \le i \le h$, and they are not vertices with labels in $V(w)$.

The first condition implies the number of edges in $G(w)$ is h. The second condition implies that there are exactly $(d-2)h + v$ many distinct vertices in w. We will break each self-avoiding walk w_i into three types of successive sub-walks where each sub-walk is exactly one of the following 3 types, and we encode these sub-walks as follows.

- Type 1: hyperedges with corresponding edges in $G(w) \setminus T(w)$. Given our position in the circuit w, we can encode a hyperedge of this type by its right-end vertex. Hyperedges of Type 1 breaks the walk w_i into disjoint sub-walks, and we partition these sub-walks into Type 2 and 3 below.
- Type 2: sub-walks such that all their hyperedges correspond to edges of $T(w)$ and have been traversed already by $w_1, \dots, w_{i-1}$. Each sub-walk is a part of a self-avoiding walk, and it is a path contained in the tree $T(w)$. Given its initial and its end vertices, there will be exactly one such path in $T(w)$. Therefore these walks can be encoded by the end vertices.
- Type 3: sub-walks such that their hyperedges correspond to edges of $T(w)$ and they are being traversed for the first time. Given the initial vertex of a sub-walk of this type, since it is traversing new edges and knowing in what order the vertices are discovered, we can encode these walks by their length, and from the given length, we know at which vertex the sub-walk ends.

We encode any Type 1, Type 2, or Type 3 sub-walk by 0 if the sub-walk is empty. Now we can decompose each w_i into sequences characterizing by its sub-walks:

$$(p_1, q_1, r_1), (p_2, q_2, r_2), \dots, (p_t, q_t, r_t). \tag{7.7}$$

Here $r_1, \dots r_{t-1}$ are codes from sub-walks of Type 1. From the way we encode such hyperedges, we have $r_i \in \{1, \dots v\}$ for $1 \le i \le t-1$. Type 2 and Type 3 sub-walks are encoded by $p_1, \dots, p_t$ and $q_1, \dots, q_t$, respectively. Since Type 1 hyperedges break w into disjoint pieces, we use (p_t, q_t, r_t) to represent the last piece of the sub-walk and make $r_t = 0$. Each p_i represents the right-end vertex of the Type 2 sub-walk, and $p_i = 0$ if it the sub-walk is empty, hence $p_i \in \{0, \dots v\}$ for $1 \le i \le t$. Each q_i represents the length of Type 3 sub-walks, so $q_i \in \{0, \dots l\}$ for $1 \le i \le t$. From the way we encode these sub-walks, there are at most $(v+1)^2(l+1)$ many possibilities for each triplet (p_j, q_j, r_j).

We now consider how many ways we can concatenate sub-walks encoded by the triplets to form a circuit w. All triples with $r_j \in [v]$ for $1 \le j \le t-1$ indicate the traversal of an edge not in $T(w)$. Since we know the number of edges in $G(w) \setminus T(w)$ is $(h - v + 1)$, and within a self-avoiding walk w_i, edges on $G(w)$ can be traversed at most once, the length of the triples in (7.7) satisfies $t - 1 \le h - v + 1$, which implies $t \le h - v + 2$. Since each hyperedge can be traversed at most $2k$ many times by w due to the constraint that the circuits w of length $2kl$ are formed by self-avoiding walks, so the number of triple sequences for fixed v, h is at most $[(v+1)^2(l+1)]^{2k(2+h-v)}$.

There are multiple w with the same code sequence. However, they must all have the same number of vertices and edges, and the positions where vertices and hyperedges are repeated must be the same.

The number of ordered sequences of v distinct vertices is at most n^v. Given the vertex sequence, the number of ordered sequences of h distinct hyperedges in $K_{n,d}$ is at most $\binom{n}{d-2}^h$. Therefore, given v, h, the number of circuits that share the same triple sequence (7.7) is at most $n^v\binom{n}{d-2}^h$.

Combining the two estimates, the number of all possible circuits w with fixed v, h in Case (1) is at most

$$n^v\binom{n}{d-2}^h[(v+1)^2(l+1)]^{2k(2+h-v)}. \tag{7.8}$$

Now we consider the expected weight of each circuit in the sum (7.6). Given σ, if $i,j\in e$, we have $A_{ij}^e \sim \text{Ber}\left(p_{\sigma(e)}\right)$, where $p_{\sigma(e)} = \frac{a}{\binom{n}{d-1}}$ if vertices in e have the same $\pm$ spins and $p_{\sigma(e)} = \frac{b}{\binom{n}{d-1}}$ otherwise. For a given hyperedge appearing in w with multiplicity $m\in\{1, \dots, 2k\}$, the corresponding expectation $\mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e - \overline{A_{ij}^e})^m\right]$ is 0 if $m=1$. Since $0\le A_{ij}^e\le 1$, for $m\ge 2$, we have

$$\mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e - \overline{A_{ij}^e})^m|\sigma\right] \le \mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e - \overline{A_{ij}^e})^2|\sigma\right] \le p_{\sigma(e)}. \tag{7.9}$$

For any hyperedge e corresponding to an edge in $G(w)\setminus T(w)$ we have the upper bound

$$p_{\sigma(e)} \le \frac{a\vee b}{\binom{n}{d-1}}. \tag{7.10}$$

Taking the expectation over σ we have

$$\mathbb{E}_\sigma[p_{\sigma(e)}] = \frac{a+(2^{d-1}-1)b}{2^{d-1}\binom{n}{d-1}} = \frac{\alpha}{(d-1)\binom{n}{d-1}}. \tag{7.11}$$

Recall the weight of each circuit in the sum (7.6) is given by

$$\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right].$$

Conditioned on σ, $(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})$ are independent random variables for distinct hyperedges. Denote these distinct hyperedges by $e_1, \dots e_h$ with multiplicity $m_1, \dots m_h$ and we temporarily order them such that $e_1, \dots e_{v-1}$ are the hyperedges corresponding to edges on $T(w)$. Introduce the random variables $A^{e_i}\sim \text{Ber}\left(p_{\sigma(e_i)}\right)$ for $1\le i\le h$ and denote $\overline{A^{e_i}} = \mathbb{E}_{\mathcal{H}_n}[A^{e_i}|\sigma]$. Therefore from (7.9) we have

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right] &= \mathbb{E}_\sigma\left[\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})|\sigma\right]\right]\\ &= \mathbb{E}_\sigma\left[\prod_{i=1}^{h}E_{\mathcal{H}_n}\left[(A^{e_i}-\overline{A^{e_i}})^{m_i}|\sigma\right]\right] \le \mathbb{E}_\sigma\left[\prod_{i=1}^{h}p_{\sigma(e_i)}\right].\end{aligned}$$

We use the bound (7.10) for $p_{\sigma(e_v)}, \dots, p_{\sigma(e_h)}$, which implies

$$\mathbb{E}_\sigma\left[\prod_{i=1}^{h}p_{\sigma(e_i)}\right] \le \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1}\mathbb{E}_\sigma\left[\prod_{i=1}^{v-1}p_{\sigma(e_i)}\right]. \tag{7.12}$$

From the second condition for w in Case (1), any two hyperedges among $\{e_1, \dots e_{v-1}\}$ share at most 1 vertex, and $p_{\sigma(e_i)}, p_{\sigma(e_j)}$ are pairwise independent for all $1 \le i < j \le v-1$. Moreover, since the corresponding edges of $e_1, \dots e_{v-1}$ form the spanning tree $T(w)$, taking any e_j such that the corresponding edge in $T(w)$ is attached to some leaf, we know e_j and $\bigcup_{i \neq j, 1 \le i \le v} e_i$ share exactly one common vertex, therefore $p_{\sigma(e_j)}$ is independent of $\prod_{1 \le i \le v-1, i \neq j} p_{\sigma(e_i)}$. We then have

$$\mathbb{E}_\sigma \left[\prod_{i=1}^{v-1} p_{\sigma(e_i)} \right] = \mathbb{E}_\sigma [p_{\sigma(e_j)}] \cdot \mathbb{E}_\sigma \left[\prod_{1 \le i \le v-1, i \neq j} p_{\sigma(e_i)} \right]. \tag{7.13}$$

Now the corresponding edges of all hyperedges $\{e_1, \dots e_{v-1}\} \setminus \{e_j\}$ form a tree in $G(w)$ again and the factorization of expectation in (7.13) can proceed as long as we have some edge attached to leaves. Repeating (7.13) recursively, with (7.11), we have

$$\mathbb{E}_\sigma \left[\prod_{i=1}^{v-1} p_{\sigma(e_i)} \right] = \prod_{i=1}^{v-1} \mathbb{E}_\sigma [p_{\sigma(e_i)}] = \left(\frac{\alpha}{(d-1)\binom{n}{d-1}} \right)^{v-1}. \tag{7.14}$$

Since every hyperedge in w must be visited at least twice to make its expected weight nonzero, and w is of length $2kl$, we must have $h \le kl$. In the multigraph $G(w)$, we have the constraint $v \le h+1 \le kl+1$. Since the first self-avoiding walk in w of length l takes $l+1$ distinct vertices, we also have $v \ge l+1$. So the possible range of v is $l+1 \le v \le kl+1$ and h satisfies $v-1 \le h \le kl$.

Putting all the estimates above together, for fixed v, h, the total contribution of self-avoiding walks from Case (1) to the sum is bounded by

$$n^v \binom{n}{d-2}^h [(v+1)^2(l+1)]^{2k(2+h-v)} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}} \right)^{v-1} \left(\frac{a \vee b}{\binom{n}{d-1}} \right)^{h-v+1}.$$

Denote S_1 to be the sum of all contributions from self-avoiding walks in Case (1). Then

$$S_1 \le \sum_{v=l+1}^{kl+1} \sum_{h=v-1}^{kl} n^v \left(\frac{d-1}{n-d+2} \right)^h \left(\frac{\alpha}{d-1} \right)^{v-1} [(v+1)^2(l+1)]^{2k(2+h-v)} (a \vee b)^{h-v+1}. \tag{7.15}$$

When $l = O(\log n)$ and d, k are fixed, for sufficiently large n, $\left(\frac{n}{n-d+2}\right)^h \le 2$. Then from (7.15),

$$\begin{aligned} S_1 &\le \sum_{v=l+1}^{kl+1} \sum_{h=v-1}^{kl} 2n^{v-h}(d-1)^{h-v+1} [(v+1)^2(l+1)]^{2k(2+h-v)} \alpha^{v-1} (a \vee b)^{h-v+1} \\ &\le 2 \sum_{v=l+1}^{kl+1} \sum_{h=v-1}^{kl} n \left[\frac{(a \vee b)(d-1)}{n} \right]^{h-v+1} [(kl+2)^2(l+1)]^{2k(2+h-v)} \alpha^{v-1}. \end{aligned}$$

Hence

$$\frac{S_1}{n\alpha^{kl}[(kl+2)^2(l+1)]^{2k}} \le 2 \sum_{v=l+1}^{kl+1} \alpha^{v-1-kl} \sum_{h=v-1}^{kl} \left[n^{-1}(a \vee b)(d-1)((kl+2)^2(l+1))^{2k} \right]^{h-v+1}. \tag{7.16}$$

Since for fixed d, k and $l = O(\log n)$, $n^{-1}(a \vee b)(d-1)((kl+2)^2(l+1))^{2k} = o(1)$ for n sufficiently large, the leading term in (7.16) is the term with $h = v - 1$. For sufficiently large n, we have

$$\frac{S_1}{n\alpha^{kl}[(kl+2)^2(l+1)]^{2k}} \le 3\sum_{v=l+1}^{kl+1} \alpha^{v-1-kl} = 3 \cdot \frac{\alpha - \alpha^{(1-k)l}}{\alpha - 1} \le \frac{3\alpha}{\alpha - 1}.$$

It implies that $S_1 = O(n\alpha^{kl}\log^{6k} n)$.

Case (2). We now consider $w \in W_{2k,l}$ such that

- the number of edges in $G(w)$ is greater than h;
- vertices in $e_i \setminus \text{end}(e_i)$ are all distinct for $1 \le i \le h$, and they are not vertices with labels in $V(w)$.

Let $\tilde{h}$ be the number of edges in $G(w)$ with $\tilde{h} \ge h + 1$. Same as in Case (1), the number of triple sequence is at most $[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}$. Let s_i, $1 \le i \le h$ be the size of end(e_i). We have $\sum_{i=1}^{h} s_i = 2\tilde{h}$. Note that when $s_i > 3$, there are more than 2 vertices in e_i contained in $V(w)$, therefore given the choices of vertices with labels in $V(w)$, we have fewer possibilities to choose the rest of vertices in e_i. Compared with (7.8), the number of all possible circuits in Case (2) with fixed $v, h, \tilde{h}$ is now bounded by

$$[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)} n^v \binom{n}{d - s_1} \cdots \binom{n}{d - s_h}.$$

When k is fixed and $l = O(\log n)$, for large n, the quantity above is bounded by

$$2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)} n^v \left(\frac{d-1}{n}\right)^{2\tilde{h}-h} \binom{n}{d-1}^h.$$

Now we consider the expected weight of each circuit in Case (2). In the spanning tree $T(w)$, we keep edges with distinct hyperedge labels that appear first in the circuit w and remove other edges. This gives us a forest denoted $F(w)$ inside $T(w)$, with at least $v - 1 - \tilde{h} + h$ many edges. We temporarily label those edges in the forest as $e_1, \dots, e_q$ with $q \ge v - 1 - \tilde{h} + h$. Then similar to the analysis of (7.14) in Case (1), we have

$$\mathbb{E}_\sigma\left[\prod_{i=1}^{q} p_{\sigma(e_q)}\right] = \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^q,$$

and

$$\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t} - \overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right] \le \mathbb{E}_\sigma\left[\prod_{i=1}^{h} p_{\sigma(e_i)}\right] \le \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h}.$$

Since every hyperedge in w must be visited at least twice to make its expected weight nonzero, we must have $l \le h \le kl$. In the multigraph $G(w)$, we have the constraint $v \le \tilde{h} + 1$. Since the first self-avoiding walk in w of length l takes $l + 1$ distinct vertices, we also have $v \ge l + 1$. So the possible range of v is $l + 1 \le v \le \tilde{h} + 1$ and h satisfies $l \le h \le kl$. Therefore we have

$$
\begin{aligned}
S_2 \le & 2\sum_{h=l}^{kl}\sum_{\tilde{h}=h+1}^{2kl}\sum_{v=l+1}^{\tilde{h}+1}[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\left(\frac{d-1}{n}\right)^{2\tilde{h}-h}\binom{n}{d-1}^h \\
& \cdot\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h} \\
= & O(\alpha^{kl}\log^{6k}n).
\end{aligned}
$$

Case (3). We now consider $w\in W_{2k,l}$ not included in Cases (1) or Case (2), which satisfies that

- for some $i\neq j$, there are common vertices in $e_i\setminus\mathrm{end}(e_i)$ and $e_j\setminus\mathrm{end}(e_j)$;
- or there are vertices in $e_i\setminus\mathrm{end}(e_i)$ with labels in $V(w)$.

Let $v,h,\tilde{h}$ be defined in the same way as in Case (2). The number of triple sequence is at most $[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}$. Consider the forest $F(w)$ introduced in Case (2) as a subgraph of $T(w)$, which has at least $(v-1-\tilde{h}+h)$ many edges with distinct hyperedge labels. We temporarily denote the edges by $e_1,\dots,e_q$, and the ordering is chosen such that e_1 is adjacent to a leaf in $F(w)$, and each $e_i, i\le 2$ is adjacent to a leaf in $F(w)\setminus\{e_1,\dots,e_{i-1}\}$. For $1\le i\le q$, we call e_i a *bad* hyperedge if the set $e_i\setminus\mathrm{end}(e_i)$ share a vertex with some set $e_j\setminus\mathrm{end}(e_j)$ for $j>i$, or there are vertices in $e_i\setminus\mathrm{end}(e_i)$ with labels in $V(w)$. In both cases, we have fewer choices for the vertices in e_i.

Suppose among $e_i, 1\le i\le q$, there are t bad hyperedges. Let $s_i, 1\le i\le h$ be the size of $\mathrm{end}(e_i)$ in $G(w)$. Then the number of all possible circuits in Case (3) with fixed $v,h,\tilde{h}$, and t, is bounded by

$$
[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\binom{n}{d-s_1-\delta_1}\cdots\binom{n}{d-s_h-\delta_h}, \tag{7.17}
$$

where $\delta_i\in\{0,1\}$ and $\delta_i=1$ if e_i is a bad hyperedge. Note that $\sum_{i=1}^h s_h=2\tilde{h}$ and $\sum_{i=1}^h\delta_i=t$. For large n, the number in (7.17) is at most

$$
2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\left(\frac{d-1}{n}\right)^{2\tilde{h}-h+t}\binom{n}{d-1}^h.
$$

After removing the t edges with bad hyperedge labels from the forest $F(w)$, we can do the same analysis as in Case (2). The expected weight of each circuit in Case (3) with given $v,h,\tilde{h},t$ now satisfies

$$
\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right]\le\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1+t}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h-t}.
$$

Let S_3 be the total contribution from circuits in Case (3) to (7.6) . Then

$$
\begin{aligned}
S_3 \le & \sum_{h=l}^{kl}\sum_{\tilde{h}=h}^{2kl}\sum_{v=l+1}^{\tilde{h}+1}\sum_{t=0}^{v-1}2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\left(\frac{d-1}{n}\right)^{2\tilde{h}-h+t}\binom{n}{d-1}^h \\
& \cdot\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1+t}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h-t} \\
= & O(n\alpha^{kl}\log^{6k}n).
\end{aligned}
$$

From the estimates on S_1, S_2 and S_3, Lemma 7.1 holds. ■

With Lemma 7.1, we are able to derive (3.3). For any fixed $\varepsilon > 0$, choose k such that $1 - 2k\varepsilon < 0$, using Markov inequality, we have

$$\mathbb{P}_{\mathcal{H}_n}(\rho(\Delta^{(l)}) \geq n^{\varepsilon}\alpha^{l/2}) \leq \frac{\mathbb{E}_{\mathcal{H}_n}(\rho(\Delta^{(l)})^{2k})}{n^{2k\varepsilon}\alpha^{kl}} = O(n^{1-2k\varepsilon}\log^{6k} n).$$

This implies (3.3) in the statement of Theorem 3.1.

7.3 | Proof of (3.4) in Theorem 3.1

Using a similar argument as in the proof of Lemma 7.1, we can prove the following estimate of $\rho(\Gamma^{(l,m)})$. The proof is given in Appendix A.6.

Lemma 7.2. *For $l = O(\log n)$, fixed k, and any $1 \leq m \leq l$, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mathcal{H}_n}[\rho(\Gamma^{(l,m)})^{2k}] \leq Cn^{1-2k}\alpha^{k(l+m-2)}\log^{14k} n. \tag{7.18}$$

With Lemma 7.2, we can apply the union bound and Markov inequality. For any $\varepsilon > 0$, choose $k > 0$ such that $1 - 2k\varepsilon < 0$, we have

$$\mathbb{P}_{\mathcal{H}_n}\left(\cup_{m=1}^{l}\left\{\rho(\Gamma^{(l,m)}) \geq n^{\varepsilon-1}\alpha^{(l+m)/2}\right\}\right) \leq \sum_{m=1}^{l}\mathbb{P}_{\mathcal{H}_n}\left(\rho(\Gamma^{(l,m)}) \geq n^{\varepsilon-1}\alpha^{(l+m)/2}\right)$$
$$\leq \sum_{m=1}^{l}\frac{\mathbb{E}_{\mathcal{H}_n}\rho(\Gamma^{(l,m)})^{2k}}{n^{2k(\varepsilon-1)}\alpha^{k(l+m)}} \leq \sum_{m=1}^{l}\frac{C\log^{14k}(n)\cdot n^{1-2k}\alpha^{k(l+m-2)}}{n^{2k(\varepsilon-1)}\alpha^{k(l+m)}} = O\left((\log^{14k+1}(n)\cdot n^{1-2k\varepsilon}\alpha^{-2k}\right).$$

This proves (3.4) in Theorem 3.1.

8 | PROOF OF THEOREM 4.2

Let $n^{\pm}$ be the number of vertices with spin $\pm$ respectively. Consider the event

$$\tilde{\Omega} := \left\{\left|n^{\pm} - \frac{n}{2}\right| \leq \log(n)\sqrt{n}\right\}. \tag{8.1}$$

By Hoeffding's inequality,

$$\mathbb{P}_{\sigma}\left(|n^{\pm} - \frac{n}{2}| \geq \log(n)\sqrt{n}\right) \leq 2\exp(-2\log^2(n)), \tag{8.2}$$

which implies $\mathbb{P}_{\sigma}(\tilde{\Omega}) \geq 1 - 2\exp(-2\log^2(n))$. In the rest of this section we will condition on the event $\tilde{\Omega}$, which will not effect our conclusion and probability bounds, since for any event A, if $\mathbb{P}_{\mathcal{H}_n}(A|\tilde{\Omega}) = 1 - O(n^{-\gamma})$ for some $\gamma > 0$, we have

$$\mathbb{P}_{\mathcal{H}_n}(A) = \mathbb{P}_{\mathcal{H}_n}(A|\tilde{\Omega})\mathbb{P}_{\mathcal{H}_n}(\tilde{\Omega}) + \mathbb{P}_{\mathcal{H}_n}(A|\tilde{\Omega}^c)\mathbb{P}_{\mathcal{H}_n}(\tilde{\Omega}^c) = 1 - O(n^{-\gamma}).$$

The following identity from Equation (38) in [32] will be helpful in the proof.

Lemma 8.1. *For any nonnegative integers i, j, n and nonnegative numbers a, b such that $a/n, b/n < 1$, we have*

$$\frac{ai+bj}{n} - \frac{1}{2}\left(\frac{ai+bj}{n}\right)^2 \leq 1-(1-a/n)^i(1-b/n)^j \leq \frac{ai+bj}{n}. \tag{8.3}$$

We will also use the following version of Chernoff bound (see [9]):

Lemma 8.2. *Let X be a sum of independent random variables taking values in $\{0,1\}$. Let $\mu = \mathbb{E}[X]$. Then for any $\delta > 0$, we have*

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp(-\mu h(1+\delta)), \tag{8.4}$$

$$\mathbb{P}(|X-\mu| \leq \delta\mu) \geq 1 - 2\exp(-\mu\tilde{h}(\delta)), \tag{8.5}$$

where

$$h(x) := x\log(x) - x + 1, \quad \tilde{h}(x) := \min\{(1+x)\log(1+x) - x, (1-x)\log(1-x) + x\}.$$

For any $t \geq 0$, the number of vertices with spin $\pm$ at distance t (respectively $\leq$) of vertices i is denoted $U_t^{\pm}(i)$ (respectively, $U_{\leq t}^{\pm}(i)$) and we know $S_t(i) = U_t^+(i) + U_t^-(i)$. We will omit index i when considering quantities related to a fixed vertex i. Let $n^{\pm}$ be the number of vertices with spin $\pm$ and $\mathcal{N}^{\pm}$ be the set of vertices with spin $\pm$. For a fixed vertex i. Let

$$\mathcal{F}_t := \sigma(U_k^+, U_k^-, k \leq t, \sigma_i, 1 \leq i \leq n) \tag{8.6}$$

be the σ-algebra generated by $\{U_k^+, U_k^-, 0 \leq k \leq t\}$ and $\{\sigma_i, 1 \leq i \leq n\}$. In the remainder of the section we condition on the spins σ_i of all $i \in [n]$ and assume $\tilde{\Omega}$ holds. We denote $\mathbb{P}(\cdot) := \mathbb{P}_{\mathcal{H}_n}(\cdot|\tilde{\Omega})$.

A main difficulty to analyze U_t^+, U_t^- compared to the graph SBM in [32] is that $U_k^{\pm}$ are no longer independent conditioned on $\mathcal{F}_{k-1}$. Instead, we can only approximate $U_k^{\pm}$ by counting subsets connected to V_{k-1}. To make it more precise, we have the following definition for connected-subsets.

Definition 8.3. A *connected s-subset* in V_k for $1 \leq s \leq d-1$ is a subset of size s which is contained in some hyperedge e in H and the rest $d-s$ vertices in e are from V_{k-1} (see Figure 5 for an example). Define $U_{k,s}^{(r)}, 0 \leq r \leq s$ to be the number of *connected s-subsets* in V_k where exactly r many vertices have $+$ spins. For convenience, we write $U_k^{(r)} := U_{k,d-1}^{(r)}$ for $0 \leq r \leq d-1$. Let $U_{k,s} = \sum_{r=0}^{s} U_{k,s}^{(r)}$ be the number of all connected s-subsets in V_k.

We will show that $\sum_{r=0}^{d-1} rU_k^{(r)}$ is a good approximation of U_k^+ and $\sum_{r=0}^{d-1}(d-1-r)U_k^{(r)}$ is a good approximation of U_k^-, then the concentration of $U_k^{(r)}, 0 \leq r \leq d-1$ implies the concentration of $U_k^{\pm}$.

Since each hyperedge appears independently, conditioned on $\mathcal{F}_{k-1}$, we know $\{U_k^{(r)}, 0 \leq r \leq d-1\}$ are independent binomial random variables. For $U_k^{(d-1)}$, the number of all possible connected $(d-1)$-subsets with $d-1$ many $+$ signs is $\binom{n^+ - U_{\leq k-1}^+}{d-1}$, and each such subset is included in the hypergraph if and only if it forms a hyperedge with any vertex in V_{k-1}. Therefore each such subset is included

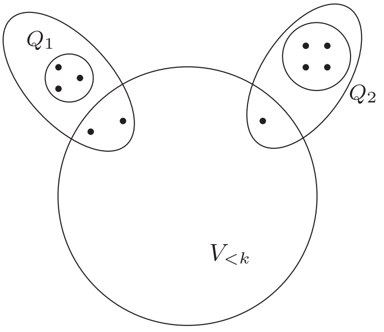

FIGURE 5 $d = 5$, Q_1 is a connected 3-subsets in V_k and Q_2 is a connected 4-subsets in V_k

independently with probability

$$1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^+_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^-_{k-1}}.$$

Similarly, we have the following distributions for $U_k^{(r)}, 1 \le r \le d-1$:

$$U_k^{(d-1)} \sim \mathrm{Bin}\left(\binom{n^+ - U^+_{\le k-1}}{d-1}, 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^+_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^-_{k-1}}\right), \tag{8.7}$$

$$U_k^{(0)} \sim \mathrm{Bin}\left(\binom{n^- - U^-_{\le k-1}}{d-1}, 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^-_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^+_{k-1}}\right), \tag{8.8}$$

and for $1 \le r \le d-2$,

$$U_k^{(r)} \sim \mathrm{Bin}\left(\binom{n^+ - U^+_{\le k-1}}{r}\binom{n^- - U^-_{\le k-1}}{d-1-r}, 1 - \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{S_{k-1}}\right). \tag{8.9}$$

For two random variable X, Y, we denote $X \preccurlyeq Y$ if X is stochastically dominant by Y, i.e., $\mathbb{P}(X \le x) \ge \mathbb{P}(Y \le x)$ for any $x \in \mathbb{R}$. We denote $U_k^* := \sum_{s=1}^{d-2} U_{k,s}$ to be the number of all connected s-subsets in V_k for $1 \le s \le d-2$.

For each $1 \le s \le d-2$, conditioned on $\mathcal{F}_{k-1}$, the number of possible s-subsets is at most $\binom{n}{s}$, and each subset is included in the hypergraph independently with probability at most $\left(\frac{a \vee b}{\binom{n}{d-1}}\binom{S_{k-1}}{d-s}\right) \wedge 1$, so we have

$$U_{k,s} \preccurlyeq \mathrm{Bin}\left(\binom{n}{s}, \frac{a \vee b}{\binom{n}{d-1}}\binom{S_{k-1}}{d-s} \wedge 1\right). \tag{8.10}$$

With the definitions above, we have the following inequality for $U_k^{\pm}$ by counting the number of $\pm$ signs from each type of subsets:

$$U_k^+ \le \sum_{r=0}^{d-1} r U_k^{(r)} + (d-2)U_k^*, \tag{8.11}$$

$$U_k^- \le \sum_{r=0}^{d-1} (d-1-r) U_k^{(r)} + (d-2)U_k^*. \tag{8.12}$$

To obtain the upper bound of $U_k^{\pm}$, we will show that U_k^* is negligible compared to the number of $\pm$ signs from $U_k^{(r)}$. Since $U_k^{(r)}, 1 \le r \le d-1$ are independent binomial random variables, we can prove concentration results of these random variables. For the lower bound of $U_k^{\pm}$, we need to show that only a negligible portion of $(d-1)$ connected subsets are overlapped, therefore U_k^+ is lower bounded by $\sum_{r=0}^{d-1} rU_k^{(r)}$ minus some small term, and we can do it similarly for U_k^-. We will extensively use Chernoff bounds in Lemma 8.2 to prove the concentration of $U_k^{\pm}$ in the following theorem.

Theorem 8.4. *Let $\varepsilon \in (0,1)$, and $l = c\log(n)$ with $c\log(\alpha) < 1/4$. For any $\gamma \in (0, 3/8)$, there exists some constant $K > 0$ and such that the following holds with probability at least $1 - O(n^{-\gamma})$ for all $i \in [n]$.*

1. *Let $T := \inf\{t \le l : S_t \ge K\log n\}$, then $S_T = \Theta(\log n)$.*
2. *Let $\varepsilon_t := \varepsilon\alpha^{-(t-T)/2}$ for some $\varepsilon > 0$ and*

$$M := \frac{1}{2}\begin{bmatrix} \alpha+\beta & \alpha-\beta \\ \alpha-\beta & \alpha+\beta \end{bmatrix}. \tag{8.13}$$

Then for all $t, t' \in \{T, \dots l\}$, $t > t'$, the vector $\vec{U}_t := (U_t^+, U_t^-)^\top$ satisfies the coordinate-wise bounds:

$$U_t^+ \in \left[\prod_{s=t'}^{t-1}(1-\varepsilon_s), \prod_{s=t'}^{t-1}(1+\varepsilon_s)\right](M^{t-t'}\vec{U}_{t'})_1, \tag{8.14}$$

$$U_t^- \in \left[\prod_{s=t'}^{t-1}(1-\varepsilon_s), \prod_{s=t'}^{t-1}(1+\varepsilon_s)\right](M^{t-t'}\vec{U}_{t'})_2, \tag{8.15}$$

where $(M^{t-t'}\vec{U}_{t'})_j$ is the jth coordinate of the vector $M^{t-t'}\vec{U}_{t'}$ for $j = 1, 2$.

Proof. In this proof, all constants C_i's, C, C' are distinct for different inequalities unless stated otherwise. By the definition of T, $S_{T-1} \le K\log(n)$. Let Z_T be the number of all hyperedges in H that are incident to at least one vertices in V_{T-1}. We have $S_T \le (d-1)Z_T$, and since the number of all possible hyperedges including a vertex in V_{T-1} is at most $S_{T-1}\binom{n}{d-1}$, Z_T is stochastically dominated by

$$\mathrm{Bin}\left(K\log(n)\binom{n}{d-1}, \frac{a \vee b}{\binom{n}{d-1}}\right),$$

which has mean $(a \vee b)K\log(n)$. Let $K_1 = (a \vee b)K$. By (8.4) in Lemma 8.2, we have for any constant $K_2 > 0$,

$$\mathbb{P}(Z_T \geq K_2 \log(n)|\mathcal{F}_{T-1}) \leq \exp(-K_1 \log(n) h(K_2/K_1)) \tag{8.16}$$

Taking $K_2 > K_1$ large enough such that $K_1 h(K_2/K_1) \geq 2 + \gamma$, we then have

$$\mathbb{P}(Z_T \geq K_2 \log(n)|\mathcal{F}_{T-1}) \leq n^{-2-\gamma}. \tag{8.17}$$

So with probability at least $1-n^{-2-\gamma}$, for a fixed $i \in [n]$, $S_T \leq K_3 \log(n)$ with $K_3 = (d-2)K_2$. Taking a union bound over $i \in [n]$, part (1) in Lemma 8.4 holds. We continue to prove (8.14) and (8.15) in several steps.

Step 1: base case. For the first step, we prove (8.14) and (8.15) for $t = T+1, t' = T$, which is

$$U_{T+1}^{\pm} \in [1-\varepsilon, 1+\varepsilon]\left(\frac{\alpha+\beta}{2}U_T^{\pm} + \frac{\alpha-\beta}{2}U_T^{\mp}\right). \tag{8.18}$$

This involves a two-sided estimate of $U_{T+1}^{\pm}$. The idea is to show the expectation of $U_{T+1}^{\pm}$ conditioned on $\mathcal{F}_T$ is closed to $\frac{\alpha+\beta}{2}U_T^{\pm} + \frac{\alpha-\beta}{2}U_T^{\mp}$, and $U_{T+1}^{\pm}$ is concentrated around its mean.

(i) Upper bound. Define the event $\mathcal{A}_T := \{S_T \leq K_3 \log n\}$. We have just shown for a fixed i,

$$\mathbb{P}(\mathcal{A}_T) \geq 1 - n^{-2-\gamma}. \tag{8.19}$$

Recall $|n^{\pm} - n/2| \leq \sqrt{n}\log n$ and conditioned on $\mathcal{A}_T$, for some constant $C > 0$,

$$U_{\leq T}^{+} \leq \sum_{t=0}^{T} S_t \leq 1 + TK_3 \log n \leq 1 + lK_3 \log n \leq CK_3 \log^2 n.$$

Conditioned on $\mathcal{F}_T$ and $\mathcal{A}_T$, for sufficiently large n, there exists constants $C_1 > 0$ such that

$$\binom{n^+ - U_{\leq T}^+}{d-1} \geq C_1 \binom{\frac{n}{2}}{d-1}.$$

From inequality (8.3), there exists constant $C_2 > 0$ such that

$$\begin{aligned} 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_T^+}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_T^-} &\geq \frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}} - \frac{1}{2}\left(\frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}}\right)^2 \\ &\geq \frac{C_2(aU_T^+ + bU_T^-)}{\binom{n}{d-1}} \geq \frac{C_2(a \wedge b)K \log n}{\binom{n}{d-1}}. \end{aligned}$$

Then from (8.7), for some constant $C_3 > 0$,

$$\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T, \mathcal{A}_T] = \binom{n^+ - U_{\leq T}^+}{d-1}\left(1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_T^+}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_T^-}\right)$$

$$\geq C_1 \binom{\frac{n}{2}}{d-1} \cdot \frac{C_2(a\wedge b)K\log n}{\binom{n}{d-1}} \geq C_3 K \log n.$$

We can choose K large enough such that $C_3K\tilde{h}(\varepsilon/(2d)) \geq 2+\gamma$, then from (8.5) in Lemma 8.2, for any given $\varepsilon > 0$ and $\gamma \in (0,1)$,

$$\begin{aligned}
&\mathbb{P}\left(|U_{T+1}^{(d-1)} - \mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]| \leq \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]\Big|\mathcal{F}_T\right) \\
&\geq \mathbb{P}\left(|U_{T+1}^{(d-1)} - \mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]| \leq \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]\Big|\mathcal{F}_T, \mathcal{A}_T\right)\mathbb{P}(\mathcal{A}_T) \\
&\geq \left[1-\exp\left(-\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T,\mathcal{A}_T]\tilde{h}(\varepsilon/2d)\right)\right](1-n^{-2-\gamma}) \geq (1-n^{-2-\gamma})^2 \geq 1-2n^{-2-\gamma}.
\end{aligned}$$

From the symmetry of $\pm$ labels, the concentration of $U_{T+1}^{(0)}$ works in the same way. Similarly, there exists a constant $C_1>0$ such that $\mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T]$, $1\leq r\leq d-2$:

$$\mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T] = \binom{n^+ - U_{\leq T}^+}{r}\binom{n^- - U_{\leq T}^-}{d-1-r}\left(1-\left(1-\frac{b}{\binom{n}{d-1}}\right)^{S_T}\right) \geq C_1 K\log n.$$

We can choose K large enough such that for all $0\leq r\leq d-1$,

$$\mathbb{P}\left(\left|U_{T+1}^{(r)} - \mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T]\right| \leq \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T]|\mathcal{F}_T\right) \geq 1-2n^{-2-\gamma}.$$

Next, we estimate $U_{T+1}^{*} = \sum_{s=1}^{d-2} U_{T+1,s}$. Recall from (8.10), we have $U_{T+1,s} \preccurlyeq Z_{T+1,s}$ where

$$Z_{T+1,s} \sim \mathrm{Bin}\left(\binom{n}{s}, \frac{a\vee b}{\binom{n}{d-1}}\binom{S_T}{d-s}\right).$$

Conditioned on $\mathcal{A}_T$ we know $K\log n \leq S_T \leq K_3\log n$, and

$$\mathbb{E}[Z_{T+1,s}|\mathcal{A}_T,\mathcal{F}_T] = \binom{n}{s}\frac{a\vee b}{\binom{n}{d-1}}\binom{S_T}{d-s} \leq C_2 \log^{d-s}(n) n^{1+s-d}$$

for some constant $C_2>0$. Using the fact that $h(x)\geq \frac{1}{2}x\log(x)$ for x large enough, from (8.4), we have for any constant $\lambda>0$, $1\leq s\leq d-2$, there exists a constant $C_3>0$ such that for large n,

$$\begin{aligned}
\mathbb{P}(U_{T+1,s} \geq \lambda S_T|\mathcal{F}_T,\mathcal{A}_T) &\leq \mathbb{P}(Z_{T+1,s}\geq \lambda S_T|\mathcal{F}_T,\mathcal{A}_T) \\
&\leq \exp\left(-\mathbb{E}[Z_{T+1,s}|\mathcal{A}_T,\mathcal{F}_T]h\left(\frac{\lambda S_T}{\mathbb{E}[Z_{T+1,s}|\mathcal{A}_T,\mathcal{F}_T]}\right)\right) \\
&\leq \exp\left(-\frac{1}{2}\lambda S_T\log\left(\frac{\lambda S_T}{\mathbb{E}[Z_{T+1,s}|\mathcal{A}_T,\mathcal{F}_T]}\right)\right) \leq \exp(-\lambda C_3\log^2 n) \leq n^{-2-\gamma}.
\end{aligned} \tag{8.20}$$

Therefore with (8.19) and (8.20),

$$\mathbb{P}(U_{T+1,s} < \lambda S_T|\mathcal{F}_T) \geq \mathbb{P}(U_{T+1,s} < \lambda S_T|\mathcal{F}_T,\mathcal{A}_T)\mathbb{P}(\mathcal{A}_T) \geq (1-n^{-2-\gamma})^2 \geq 1-2n^{-2-\gamma}.$$

Taking $\lambda = \frac{(\alpha-\beta)\varepsilon}{4d^2}$, we have $U_{T+1,s} \leq \frac{(\alpha-\beta)\varepsilon}{4d^2} S_T$ with probability at least $1 - 2n^{-2-\gamma}$ for any $\gamma \in (0, 1)$.

Taking a union bound over $2 \leq r \leq d - 1$, it implies

$$U^*_{T+1} \leq \frac{(\alpha - \beta)\varepsilon}{4d} S_T \tag{8.21}$$

with probability $1 - O(n^{-2-\gamma})$ for any $\gamma \in (0, 1)$.

Note that $n^{\pm} = \frac{n}{2} + O(\sqrt{n}\log n)$ and $U^{\pm}_{\leq T} = \sum_{k=1}^{T} S_k = O(\log^2(n))$. From (8.3),

$$\left(1 - \frac{aU_T^+ + bU_T^-}{2\binom{n}{d-1}}\right)\frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}} \leq 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_T^+}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_T^-} \leq \frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}}.$$

It implies that

$$\begin{aligned}
\mathbb{E}[U^{(d-1)}_{T+1}|\mathcal{F}_T, \mathcal{A}_T] &= \binom{\frac{n}{2} + O(\sqrt{n}\log n)}{d-1}\left(1 + O\left(\frac{\log(n)}{n^{d-1}}\right)\right)\frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}} \\
&= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)(aU_T^+ + bU_T^-).
\end{aligned} \tag{8.22}$$

Similarly, for $1 \leq r \leq d - 2$.

$$\begin{aligned}
\mathbb{E}[U^{(0)}_{T+1}|\mathcal{F}_T, \mathcal{A}_T] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)(bU_T^+ + aU_T^-), \\
\mathbb{E}[U^{(r)}_{T+1}|\mathcal{F}_T, \mathcal{A}_T] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\binom{d-1}{r}(bU_T^+ + bU_T^-).
\end{aligned}$$

Therefore from the estimations above, with the definition of α, β from (1.3),

$$\begin{aligned}
&\mathbb{E}\left[\sum_{r=0}^{d-1} rU^{(r)}_{T+1}|\mathcal{F}_T, \mathcal{A}_T\right] \\
&= \left(1 + O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\frac{1}{2^{d-1}}\left((d-1)(aU_T^+ + bU_T^-) + \sum_{r=1}^{d-2} r\binom{d-1}{r} b(U_T^+ + U_T^-)\right) \\
&= \left(1 + O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\left(\frac{\alpha + \beta}{2}U_T^+ + \frac{\alpha - \beta}{2}U_T^-\right).
\end{aligned} \tag{8.23}$$

Since we have shown $\sum_{r=0}^{d-1} U^{(r)}_{T+1}$ concentrated around its mean by $\frac{\varepsilon}{2d}$ with probability at least $1 - O(n^{-2-\gamma})$, conditioned on $\mathcal{A}_T$, we obtain

$$\left|\sum_{r=0}^{d-1} rU^{(r)}_{T+1} - \mathbb{E}[\sum_{r=0}^{d-1} rU^{(r)}_{T+1}|\mathcal{F}_T]\right| \leq \sum_{r=0}^{d-1} r\left|U^{(r)}_{T+1} - \mathbb{E}[U^{(r)}_{T+1}|\mathcal{F}_T]\right| \leq \frac{\varepsilon}{2d}\sum_{r=1}^{d-1} r\mathbb{E}[U^{(r)}_{T+1}|\mathcal{F}_T]$$

$$\leq \frac{\varepsilon}{4}\left(1+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right) \tag{8.24}$$

with probability $1-O(n^{-2-\gamma})$. Therefore from (8.23), conditioned on $\mathcal{A}_T$, for large n, with probability $1-O(n^{-2-\gamma})$,

$$\sum_{r=0}^{d-1} rU_{T+1}^{(r)} \in \left[1-\frac{\varepsilon}{3}, 1+\frac{\varepsilon}{3}\right]\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right). \tag{8.25}$$

From (8.11), (8.21), and (8.25), conditioned on $\mathcal{A}_T$ and $\mathcal{F}_T$, with probability $1-O(n^{-2-\gamma})$,

$$\begin{aligned} U_{T+1}^+ &\leq \sum_{r=0}^{d-1} rU_{T+1}^{(r)} + (d-2)U_{T+1}^* \leq \sum_{r=0}^{d-1} rU_{T+1}^{(r)} + (d-2)\frac{(\alpha-\beta)\varepsilon S_T}{4d} \\ &\leq (1+\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right). \end{aligned}$$

Since $\mathbb{P}(\mathcal{A}_T) = 1-n^{-2-\gamma}$, and by symmetry of $\pm$ labels, with probability $1-O(n^{-2-\gamma})$,

$$U_{T+1}^{\pm} \leq (1+\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^{\pm} + \frac{\alpha-\beta}{2}U_T^{\mp}\right). \tag{8.26}$$

(ii) Lower bound. To show (8.14), (8.15) for $t'=T+1, t=T$, we cannot directly bound $U_{T+1}^{\pm}$ from below by $U_{T+1}^{(r)}, 1 \leq r \leq d-1$ since from our definition of the connected $(d-1)$-subsets, they can overlap with each other, which leads to over-counting of the number vertices with $\pm$ labels. In the following we show the overlaps between different connected $(d-1)$-sets are small, which gives us the desired lower bound.

Let $W_{t+1,i}^{\pm}$ be the set of vertices in $V_{>t}$ with spin $\pm$ and appear in at least i distinct connected $(d-1)$-subsets in $V_{>t}$ for $i \geq 1$. Let $W_{t+1,i} = W_{t+1,i}^+ \cup W_{t+1,i}^-$. From our definition, $W_{T+1,1}^+$ are the vertices with spin + that appear in at least one connected $(d-1)$-subsets, so $|W_{T+1,1}^+| \leq U_{T+1}^+$. By counting the multiplicity of vertices with spin +, we have the following relation

$$\sum_{r=1}^{d-1} rU_{T+1}^{(r)} = |W_{T+1,1}^+| + \sum_{i\geq 2}|W_{T+1,i}^+| \leq U_{T+1}^+ + \sum_{i\geq 2}|W_{T+1,i}|. \tag{8.27}$$

This implies a lower bound on U_{T+1}^+:

$$U_{T+1}^+ \geq \sum_{r=1}^{d-1} rU_{T+1}^{(r)} - \sum_{i\geq 2}|W_{T+1,i}|. \tag{8.28}$$

Next we control $|W_{T+1,2}|$. Let $m = n - |V_{\leq T}|$. We enumerate all vertices in $V_{>T}$ from 1 to m temporarily for the proof of the lower bound. Let $X_i, 1 \leq i \leq m$ be the random variables that $X_i = 1$ if $i \in W_{T+1,2}$ and 0 otherwise, we then have $|W_{T+1,2}| = \sum_{i=1}^m X_i$. A simple calculation yields

$$|W_{T+1,2}|^2 - |W_{T+1,2}| = \left(\sum_{i=1}^m X_i\right)^2 - \sum_{i=1}^m X_i = 2\sum_{1\leq i<j\leq m} X_iX_j. \tag{8.29}$$

The product X_iX_j is 1 if $i,j \in W_{T+1,2}$ and 0 otherwise.

We further consider 3 events, E_{ij}^s for $s=0,1,2$, where E_{ij}^0 is the event that all $(d-1)$-subsets in $V_{>T}$ containing i,j are not connected to V_T, E_{ij}^1 is the event that there is only one $(d-1)$-subset in $V_{>T}$ containing i,j connected to V_T and E_{ij}^2 is the event that there are at least two $(d-1)$-subsets in $V_{>T}$ containing i,j connected to V_T. Now we have

$$\begin{aligned}\mathbb{E}[X_iX_j|\mathcal{F}_T,\mathcal{A}_T] &= \mathbb{P}\left(i,j\in W_{T+1,2}|\mathcal{F}_T,\mathcal{A}_T\right)\\ &= \sum_{r=0}^{2}\mathbb{P}\left(i,j\in W_{T+1,2}|E_{ij}^r,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}(E_{ij}^r|\mathcal{F}_T,\mathcal{A}_T).\end{aligned} \tag{8.30}$$

We estimate the three terms in the sum separately. Conditioned on E_{ij}^0, $\mathcal{F}_T$, and $\mathcal{A}_T$, the two events that $i\in W_{T+1,2}$ and $j\in W_{T+1,2}$ are independent. And the probability that $i\in W_{T+1,2}$ is bounded by

$$\binom{n}{d-2}^2\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^2 S_T^2\le \frac{C_1\log^2(n)}{n^2}$$

for some constant $C_1>0$. So we have

$$\begin{aligned}&\mathbb{P}\left(i,j\in W_{T+1,2}|E_{ij}^0,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}(E_{ij}^0|\mathcal{F}_T,\mathcal{A}_T)\le \mathbb{P}\left(i,j\in W_{T+1,2}|E_{ij}^0,\mathcal{F}_T,\mathcal{A}_T\right)\\ &\quad=\mathbb{P}\left(i\in W_{T+1,2}|E_{ij}^0,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}\left(j\in W_{T+1,2}|E_{ij}^0,\mathcal{F}_T,\mathcal{A}_T\right)\le \frac{C_1^2\log^4 n}{n^4}.\end{aligned} \tag{8.31}$$

For the term that involves E_{ij}^1, we know for some $C_2>0$,

$$\mathbb{P}(E_{ij}^1|\mathcal{F}_T,\mathcal{A}_T)\le \binom{n}{d-3}\frac{a\vee b}{\binom{n}{d-1}}S_T\le \frac{C_2\log n}{n^2},$$

and conditioned on E_{ij}^1 and $\mathcal{F}_T,\mathcal{A}_T$, the two events that $i\in W_{T+1,2}$ and $j\in W_{T+1,2}$ are independent again, since we require i,j to be contained in at least 2 connected-subsets. We have

$$\mathbb{P}\left(i\in W_{T+1,2}|E_{ij}^1,\mathcal{F}_T,\mathcal{A}_T\right)\le \binom{n}{d-2}S_T\frac{a\vee b}{\binom{n}{d-1}}\le \frac{C_2\log n}{n}.$$

Therefore we have

$$\begin{aligned}&\mathbb{P}\left(i,j\in W_{T+1,2}|E_{ij}^1,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}(E_{ij}^1|\mathcal{F}_T,\mathcal{A}_T)\\ &\quad=\mathbb{P}\left(i\in W_{T+1,2}|E_{ij}^1,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}\left(j\in W_{T+1,2}|E_{ij}^1,\mathcal{F}_T,\mathcal{A}_T\right)\mathbb{P}(E_{ij}^1|\mathcal{F}_T,\mathcal{A}_T)\\ &\quad\le \frac{C_2^2\log^2 n}{n^2}\cdot\frac{C_2\log n}{n^2}=\frac{C_2^3\log^3 n}{n^4}.\end{aligned} \tag{8.32}$$

Conditioned on E_{ij}^2, i,j have already been included in 2 connected $(d-1)$ subsets, so

$$\mathbb{P}\left(i,j\in W_{T+1,2}|E_{ij}^2,\mathcal{F}_T,\mathcal{A}_T\right)=1.$$

We then have for some $C_3 > 0$,

$$\begin{aligned}&\mathbb{P}\left(i,j \in W_{T+1,2}|E_{ij}^2, \mathcal{F}_T, \mathcal{A}_T\right)\mathbb{P}(E_{ij}^2|\mathcal{F}_T, \mathcal{A}_T)\\&= \mathbb{P}(E_{ij}^2|\mathcal{F}_T, \mathcal{A}_T) \le \binom{n}{d-3}^2 S_T^2\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^2 \le \frac{C_3\log^2 n}{n^4}.\end{aligned} \tag{8.33}$$

Combining (8.31)–(8.33), we have for some constant $C' > 0$,

$$\mathbb{E}[X_iX_j|\mathcal{F}_T, \mathcal{A}_T] \le \frac{C'\log^4 n}{n^4}. \tag{8.34}$$

Taking conditional expectation in (8.29), we have

$$\mathbb{E}\left[|W_{T+1,2}|^2 - |W_{T+1,2}||\mathcal{F}_T, \mathcal{A}_T\right] = 2\sum_{1\le i<j\le m}\mathbb{E}[X_iX_j|\mathcal{F}_T, \mathcal{A}_T] \le \frac{C'\log^4 n}{n^2}.$$

By Markov's inequality, there exists a constant $C > 0$ such that for any constant $\lambda > 0$ and sufficiently large n,

$$\begin{aligned}&\mathbb{P}\left(|W_{T+1,2}| > \lambda S_T|\mathcal{F}_T, \mathcal{A}_T\right) \le \mathbb{P}\left(|W_{T+1,2}|(|W_{T+1,2}|-1) > \lambda S_T(\lambda S_T - 1)|\mathcal{F}_T, \mathcal{A}_T\right)\\&\le \frac{\mathbb{E}[|W_{T+1,2}|(|W_{T+1,2}|-1)|\mathcal{F}_T, \mathcal{A}_T]}{\lambda S_T(\lambda S_T-1)} \le \frac{C\log^2 n}{\lambda^2 n^2},\end{aligned} \tag{8.35}$$

where in the last inequality we use the fact that $S_T \ge K\log n$. Taking $\lambda = \frac{(\alpha-\beta)\varepsilon}{4}$, we have for all large n and for any $\gamma \in (0,1)$,

$$\mathbb{P}\left(|W_{T+1,2}| > \frac{(\alpha-\beta)\varepsilon}{4}S_T|\mathcal{F}_T, \mathcal{A}_T\right) = O\left(\frac{\log^2 n}{n^2}\right) \le n^{-1-\gamma}. \tag{8.36}$$

For a fixed vertex $j \in V_{>T}$, the probability that $j \in W_{T+1,i}$ is at most $\binom{n}{d-2}^i S_T^i\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^i$, then we have for sufficiently large n,

$$\mathbb{E}[|W_{T+1,i}||\mathcal{F}_T, \mathcal{A}_T] \le n\binom{n}{d-2}^i S_T^i\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^i \le n\left(\frac{C_4\log n}{n}\right)^i$$

for some $C_4 > 0$. For the rest of the terms in (8.27), we have for some constant $C > 0$,

$$\mathbb{E}\left[\sum_{i\ge 3}|W_{T+1,i}|\Big|\mathcal{F}_T, \mathcal{A}_T\right] \le n\sum_{i=3}^{\infty}\left(\frac{C_4\log n}{n}\right)^i \le \frac{C\log^3(n)}{n^2}.$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{i\ge 3}|W_{T+1,i}| \ge \frac{(\alpha-\beta)\varepsilon}{4}S_T|\mathcal{F}_T, \mathcal{A}_T\right) \le \frac{C\log^2(n)}{n^2} \le n^{-1-\gamma}.$$

Together with (8.36), we have conditioned on $\mathcal{A}_T$, $\sum_{i\geq 2}|W^+_{T+1,2}| \leq \frac{(\alpha-\beta)\varepsilon}{2}S_T$ with probability at least $1-2n^{-1-\gamma}$ for any $\gamma\in(0,1)$ and all large n.

Note that

$$\frac{(\alpha-\beta)\varepsilon}{2}S_T \leq \frac{\varepsilon}{2}\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right).$$

With (8.25), (8.28), and (8.19), we have

$$U^+_{T+1} \geq \sum_{r=1}^{d-1} rU^{(r)}_{T+1} - \frac{\varepsilon}{2}\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right) \geq (1-\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right)$$

with probability $1-O(n^{-1-\gamma})$. By symmetry, the argument works for U^-_{T+1}, therefore with probability $1-O(n^{-1-\gamma})$ for any $\gamma\in(0,1)$, we have

$$U^\pm_{T+1} \geq (1-\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^\pm + \frac{\alpha-\beta}{2}U_T^\mp\right). \tag{8.37}$$

From (8.26) and (8.37), we have with probability $1-O(n^{-1-\gamma})$ for any $\gamma\in(0,1)$, (8.18) holds.

Step 2: Induction. It remains to extend this estimate in Step 1 for all $T\leq t'<t\leq l$. We now define the event

$$\mathcal{A}_t := \left\{U_t^\pm \in [1-\varepsilon_{t-1}, 1+\varepsilon_{t-1}]\left(\frac{\alpha+\beta}{2}U^\pm_{t-1} + \frac{\alpha-\beta}{2}U^\pm_{t-1}\right)\right\} \tag{8.38}$$

for $T+1\leq t\leq l$, and recall $\varepsilon_t = \varepsilon\alpha^{-(t-T)/2}$, $\mathcal{A}_T = \{S_T \leq K_3\log n\}$.

From the proof above, we have shown $\mathcal{A}_{T+1}$ holds with probability $1-O(n^{-1-\gamma})$. Conditioned on $\mathcal{A}_T, \mathcal{A}_{T+1}, \dots, \mathcal{A}_t$ for some fix t with $T+2\leq t\leq l$, the vector $\vec{U}_t = (U_t^+, U_t^-)$ satisfies (8.14), (8.15) for any $T\leq t'<t$.

Set $t'=T+1$. From [32], for any integer $k>0$, $M^k = \frac{1}{2}\begin{bmatrix}\alpha^k+\beta^k & \alpha^k-\beta^k\\ \alpha^k-\beta^k & \alpha^k+\beta^k\end{bmatrix}$. (8.14) implies that

$$\begin{aligned}
U_t^\pm &\geq \left(\prod_{s=T+1}^{t-1}(1-\varepsilon_s)\right)\left(\frac{\alpha^{t-T-1}+\beta^{t-T-1}}{2}U^\pm_{T+1} + \frac{\alpha^{t-T-1}-\beta^{t-T-1}}{2}U^\mp_{T+1}\right)\\
&\geq (1-O(\varepsilon))\frac{\alpha^{t-T-1}}{2}(1-\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^\pm + \frac{\alpha-\beta}{2}U_T^\mp\right)\\
&\geq (1-O(\varepsilon))\alpha^{t-T}\frac{(1-\varepsilon)(\alpha-\beta)}{4\alpha}S_T \geq C_1\alpha^{t-T}\log(n),
\end{aligned} \tag{8.39}$$

for some constant $C_1>0$. For any t with $T\leq t$, conditioned on $\mathcal{A}_T, \mathcal{A}_{T+1}, \dots, \mathcal{A}_t$, since $\beta<\alpha$,

$$\begin{aligned}
U_t^\pm &\leq \left(\prod_{s=T}^{t-1}(1+\varepsilon_s)\right)\left(\frac{\alpha^{t-T}+\beta^{t-T}}{2}U_T^\pm + \frac{\alpha^{t-T}-\beta^{t-T}}{2}U_T^\mp\right)\\
&\leq (1+O(\varepsilon))\frac{\alpha^{t-T}+\beta^{t-T}}{2}S_T \leq (1+O(\varepsilon))\alpha^{t-T}K_3\log(n) \leq C_2\alpha^{t-T}\log n
\end{aligned} \tag{8.40}$$

for some $C_2>0$. Combining lower and upper bounds on $U_t^\pm$, we obtain

$$S_t = U_t^+ + U_t^- = \Theta(\alpha^{t-T}\log n). \tag{8.41}$$

We now show by induction that $\mathcal{A}_{t+1}$ holds with high enough probability conditioned on $\{\mathcal{A}_j, T \le j \le t\}$.

(i) Upper bound. Note that $\alpha^l = o(n^{1/4})$, for some constant $C > 0$

$$U^+_{\le t} \le \sum_{i=1}^{t} S_i \le C\alpha^{t-T}\log^2 n \le C\alpha^l \log n = o(n^{1/4}\log n).$$

Recall $|n^{\pm} - \frac{n}{2}| \le \sqrt{n}\log n$. From (8.7)–(8.9), similar to the case for $t = T$, we have

$$\begin{aligned}\mathbb{E}[U^{(d-1)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] &= \binom{n^+ - U^+_{\le t}}{d-1}\left(1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_t^+}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_t^-}\right)\\ &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)(aU_t^+ + bU_t^-),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[U^{(0)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)(bU_t^+ + aU_t^-),\\ \mathbb{E}[U^{(r)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)\binom{d-1}{r}(bU_t^+ + bU_t^-),\end{aligned}$$

for $1 \le r \le d-2$. Hence there exists a constant $C_0 > 0$ such that for all $0 \le r \le d-1$,

$$\mathbb{E}[U^{(r)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] \ge C_0 S_t.$$

From (8.5) in Lemma 8.2, for any $0 \le r \le d-1$, to show

$$\mathbb{P}\left(\left|U^{(r)}_{t+1} - \mathbb{E}[U^{(r)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t]\right| \le \frac{\varepsilon}{2d}\mathbb{E}[U^{(r)}_{t+1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] \,\middle|\, \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \ge 1 - n^{-2-\gamma}, \tag{8.42}$$

it suffices to have

$$C_0 S_t \tilde{h}\left(\frac{\varepsilon_t}{2d}\right) \ge (2+\gamma)\log n. \tag{8.43}$$

From (8.5), by a second-order expansion of $\tilde{h}$ around 0, $\tilde{h}(x) \ge x^2/3$ when $x > 0$ is small. For $\gamma \in (0,1)$, the left hand side in (8.43) is lower bounded by

$$C_1 K\alpha^{t-T}\log(n)\tilde{h}\left(\frac{\varepsilon_t}{2d}\right) \ge C_2\alpha^{t-T}K\log(n)\varepsilon_t^2 = C_2 K\log n \ge (2+\gamma)\log n,$$

by taking K large enough. Therefore (8.42) holds.

We also have

$$U_{t+1,s} \preccurlyeq Z_{t+1,s}, \quad Z_{t+1,s} \sim \mathrm{Bin}\left(\binom{n}{s}, \frac{a \vee b}{\binom{n}{d-1}}\binom{S_t}{d-s}\right),$$

and $Z_{t+1,s}$ has mean $\binom{n}{s}\frac{a\vee b}{\binom{n}{d-1}}\binom{S_t}{d-s} = \Theta\left(\frac{\alpha^{(d-s)(t-T)}\log^{d-s}(n)}{n^{d-1-s}}\right)$. For $1 \le s \le d-2$, using the fact that $h(x) \ge \frac{1}{2}x\log(x)$ for x large enough, similar to (8.20), there are constants $C_1, C_2, C_3, C_4 > 0$ such that for any $\lambda > 0$,

$$\begin{aligned}&\mathbb{P}(U_{t+1,s} \ge \lambda S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) \le \mathbb{P}(Z_{t+1,s} \ge \lambda S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t)\\ &\quad\le \exp\left(-C_1\lambda\alpha^{t-T}\log(n)\log\left(\frac{C_2\lambda\alpha^{t-T}\log(n)}{C_3\alpha^{(d-s)(t-T)}\log^{d-s}(n)n^{1+s-d}}\right)\right).\end{aligned}$$

Taking $\lambda = \frac{(\alpha-\beta)\varepsilon_t}{4d^2} = \frac{(\alpha-\beta)\varepsilon\alpha^{-(t-T)/2}}{4d^2}$, we have

$$\begin{aligned}&\mathbb{P}\left(U_{t+1,s} \ge \frac{(\alpha-\beta)\varepsilon_t}{4d^2}S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right)\\ &\quad\le \exp\left(-C_1'\alpha^{(t-T)/2}\log(n)\cdot\log(C_2'\alpha^{(s-d+\frac{1}{2})(t-T)}\log^{1+s-d}(n)n^{d-1-s})\right).\end{aligned}$$

Since for some constants $C_4, C_5, C_6 > 0$,

$$\begin{aligned}&\log(C_2'\alpha^{(s-d+\frac{1}{2})(t-T)}\log^{1+s-d}(n)n^{d-1-s})\\ &\quad\ge C_4 - C_5(t-T)\log(\alpha) + \log(\log^{1+s-d}(n)) + (d-1-s)\log n \ge C_6\log n,\end{aligned}$$

we have for all $1 \le s \le d-2$,

$$\mathbb{P}(U_{t+1,s} \ge \frac{(\alpha-\beta)\varepsilon_t}{4d^2}S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) \le \exp\left(-C_1'C_6\log^2 n\right) \le n^{-2-\gamma} \tag{8.44}$$

for any $\gamma \in (0,1)$. Recall for sufficiently large n,

$$\varepsilon_t = \varepsilon\alpha^{-(t-T)/2} \ge \varepsilon\alpha^{-l/2} > n^{-1/8}.$$

Therefore $\frac{\log n}{\sqrt{n}} = o(\varepsilon_t)$. From (8.44), conditioned on $\mathcal{A}_T, \dots, A_t$ and $\mathcal{F}_t$,

$$U_{t+1}^+ \le \sum_{r=1}^{d-1} rU_{t+1}^{(r)} + (d-2)U_{t+1}^* \le (1+\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^+ + \frac{\alpha-\beta}{2}U_t^-\right)$$

with probability at least $1 - O(n^{-2-\gamma})$. A similar bound works for U_{T+1}^-, which implies conditioned on $\mathcal{A}_T, \dots, A_t$,

$$U_{t+1}^{\pm} \le (1+\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\mp}\right) \tag{8.45}$$

with probability $1 - O(n^{-2-\gamma})$ for any $\gamma \in (0,1)$.

(ii) Lower bound. We need to show that conditioned on $\mathcal{A}_T, \dots, \mathcal{A}_t$, $U_{t+1}^{\pm} \ge (1 - \varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\mp}\right)$ with probability $1 - O(n^{-1-\gamma})$ for some $\gamma \in (0,1)$. This part of the proof is very similar to the case for $t = T$. Same as (8.28), we have the following lower bound on U_{T+1}^+:

$$U_{t+1}^+ \ge \sum_{r=1}^{d-1} rU_{t+1}^{(r)} - \sum_{i\ge 2}|W_{t+1,i}|.$$

Next we control $|W_{t+1,2}|$. Let $m = n - |V_{\leq t}|$ and we enumerate all vertices in $V_{>t}$ from 1 to m. Let $X_1, \dots X_m$ be the random variable that $X_i = 1$ if $i \in W_{t+1,2}$ and 0 otherwise. Same as (8.29),

$$|W_{t+1,2}|^2 - |W_{t+1,2}| = 2 \sum_{1 \leq i < j \leq m} X_i X_j. \tag{8.46}$$

Let E_{ij}^s for $s = 0, 1, 2$, be the similar events as in (8.30) before, now we have

$$\begin{aligned}
\mathbb{E}[X_i X_j | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t] &= \mathbb{P}\left(i, j \in W_{t+1,2} | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right) \\
&= \sum_{r=0}^{2} \mathbb{P}\left(i, j \in W_{t+1,2} | E_{ij}^r, \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^r | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t).
\end{aligned}$$

The three terms in the sum can be estimated separately in the same way as before. By using the upper bound $C\alpha^{t-T} \log n \leq S_t \leq C_0 \alpha^{t-T} \log n$ for some $C, C_0 > 0$, and use the same argument for the case when $t = T$, we have the following three inequalities for some constants $C_1, C_2, C_3 > 0$:

$$\begin{aligned}
\mathbb{P}\left(i, j \in W_{t+1,2} | E_{ij}^0, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^0 | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_1^2 \alpha^{4(t-T)} \log^4 n}{n^4}, \\
\mathbb{P}\left(i, j \in W_{t+1,2} | E_{ij}^1, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^1 | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_2^3 \alpha^{3(t-T)} \log^3 n}{n^4}, \\
\mathbb{P}\left(i, j \in W_{t+1,2} | E_{ij}^2, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^2 | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_3 \alpha^{2(t-T)} \log^2 n}{n^4}.
\end{aligned}$$

This implies $\mathbb{E}[X_i X_j | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t] \leq \frac{C' \alpha^{4(t-T)} \log^4 n}{n^4}$ for some $C' > 0$. Taking conditional expectation in (8.46), we have

$$\mathbb{E}\left[|W_{t+1,2}|^2 - |W_{t+1,2}| | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right] \leq \frac{C' \alpha^{4(t-T)} \log^4 n}{n^2}.$$

Then by Markov inequality and (8.41), similar to (8.35), there exists a constant $C > 0$ such that for any $\lambda = \Omega(\alpha^{-(t-T)})$,

$$\mathbb{P}\left(|W_{t+1,2}| > \lambda S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right) \leq \frac{C\alpha^{2(t-T)} \log^2 n}{\lambda^2 n^2}.$$

Take $\lambda = \frac{(\alpha-\beta)\epsilon_t}{4}$. Since $c \log(\alpha) < 1/4$, we have $\alpha^l < n^{1/4}$, and

$$\mathbb{P}\left(|W_{t+1,2}| > \frac{(\alpha - \beta)\epsilon_t}{4} S_t | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right) \leq \frac{C\alpha^{2(t-T)} \log^2 n}{n^2} \leq n^{-1-\gamma}$$

for any $\gamma \in (0, 1/2)$.

For each $|W_{t+1,i}|$ for $i \geq 3$, we have for sufficiently large n, there exists a constant $C_4 > 0$

$$\mathbb{E}[|W_{t+1,i}| | \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t] \leq n \binom{n}{d-2}^i S_t^i \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^i \leq n \left(\frac{C_4 \alpha^{t-T} \log n}{n}\right)^i.$$

For the rest of the terms, we have for some constant $C_4' > 0$,

$$\mathbb{E}\left[\sum_{i\geq 3}|W_i|\mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right] \leq n\sum_{i=3}^{\infty}\left(\frac{C_4\alpha^{t-T}\log n}{n}\right)^i \leq \frac{C_4'\alpha^{3(t-T)}\log^3(n)}{n^2}.$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{i\geq 3}|W_i| \geq \frac{(\alpha-\beta)\epsilon_t}{4}S_t\mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \leq \frac{C_5\alpha^{2.5(t-T)}\log^2(n)}{n^2} \leq n^{-1-\gamma}$$

for any $\gamma \in (0, 3/8)$. Together with the estimate on $W_{t+1,2}$, we have

$$\sum_{i\geq 2}|W_{t+1,2}^+| \leq \frac{(\alpha-\beta)\epsilon_t}{2}S_t \leq \frac{\epsilon_t}{2}\left(\frac{\alpha+\beta}{2}U_t^+ + \frac{\alpha-\beta}{2}U_t^-\right)$$

with probability $1 - 2n^{-1-\gamma}$ for any $\gamma \in (0, 3/8)$.

With (8.28) and (8.25), $U_{t+1}^+ \geq (1-\epsilon_t)\left(\frac{\alpha+\beta}{2}U_t^+ + \frac{\alpha-\beta}{2}U_t^-\right)$ with probability $1 - O(n^{-1-\gamma})$. By symmetry, the argument works for U_{T+1}^-. Therefore conditioned on $\mathcal{A}_T, \dots, \mathcal{A}_t$, with probability $1 - O(n^{-1-\gamma})$ for any $\gamma \in (0, 3/8)$,

$$U_{t+1}^{\pm} \geq (1-\epsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\mp}\right). \tag{8.47}$$

This finishes the proof the lower bound part of Step 2.

Recall (8.38). With (8.47) and (8.45), we have shown that conditioned on $\mathcal{A}_T, \dots, \mathcal{A}_t$, with probability $1 - O(n^{-1-\gamma})$, $\mathcal{A}_{t+1}$ holds. This finishes the induction step. Finally, for fixed $i \in [n]$ and $\gamma \in (0, 3/8)$,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{t=T}^{l}\mathcal{A}_t\right) &= \mathbb{P}(\mathcal{A}_T)\prod_{t=T+1}^{l}\mathbb{P}(\mathcal{A}_t|\mathcal{A}_{t-1}, \dots, \mathcal{A}_T)\\ &\geq (1 - Cn^{-2-\gamma})(1 - Cn^{-1-\gamma})^l \geq 1 - C_6\log(n)n^{-1-\gamma},\end{aligned}$$

for some constant $C_6 > 0$. Taking a union bound over $i \in [n]$, we have shown $\mathcal{A}_t$ holds for all $T \leq t \leq l$ and all $i \in [n]$ with probability $1 - O(n^{-\gamma})$ for any $\gamma \in (0, 3/8)$. This completes the proof of Theorem 8.4. ∎

With Theorem 8.4, the rest of the proof of Theorem 4.2 follows similarly from the proof of Theorem 2.3 in [32]. We include it for completeness.

Proof of Theorem 4.2. Assume all the estimates in statement of Theorem 8.4 hold. For $t \leq l$, if $t \leq T$, from the definition of T, we have $S_t, |D_t| = O(\log n)$. For $t > T$, from [32], M satisfies

$$M^k = \frac{1}{2}\begin{bmatrix}\alpha^k + \beta^k & \alpha^k - \beta^k\\ \alpha^k - \beta^k & \alpha^k + \beta^k\end{bmatrix}.$$

Using (8.14) and (8.15), we have for $t > t' \geq T$,

$$S_t \leq \left(\prod_{s=t'}^{t-1}(1+\varepsilon_s)\right)(1,1)M^{t-t'}\vec{U}_{t'} \leq \left(\prod_{s=t'}^{t-1}(1+\varepsilon_s)\right)\alpha^{t-t'}S_{t'}, \tag{8.48}$$

$$S_t \geq \left(\prod_{s=t'}^{t-1}(1-\varepsilon_s)\right)(1,1)M^{t-t'}\vec{U}_{t'} \geq \left(\prod_{s=t'}^{t-1}(1-\varepsilon_s)\right)\alpha^{t-t'}S_{t'}. \tag{8.49}$$

Setting $t' = T$ in (8.48), we obtain

$$S_t \leq \left(\prod_{s=T}^{t-1}(1+\varepsilon_s)\right)\alpha^{t-T}S_T = O(\alpha^{t-T}\log n) = O(\alpha^t \log n).$$

Therefore (4.1) holds. Let $t = l$ in (8.48) and (8.49), we have for all $T \leq t' < l$,

$$\left(\prod_{s=t'}^{l-1}(1-\varepsilon_s)\right)\alpha^{l-t'}S_{t'} \leq S_l \leq \left(\prod_{s=t'}^{l-1}(1+\varepsilon_s)\right)\alpha^{l-t'}S_{t'}.$$

And it implies

$$\left(\prod_{s=t'}^{l-1}(1-\varepsilon_s)\right)S_{t'} \leq \alpha^{t'-l}S_l \leq \left(\prod_{s=t'}^{l-1}(1+\varepsilon_s)\right)S_{t'}. \tag{8.50}$$

Note that

$$\max\left\{\prod_{s=t'}^{l-1}(1+\varepsilon_s) - 1, 1 - \prod_{s=t'}^{l-1}(1-\varepsilon_s)\right\} = O(\varepsilon_{t'}) = O(\alpha^{-t'/2}).$$

Together with (8.50), we have for all $T \leq t' < l$,

$$|S_{t'} - \alpha^{t'-l}S_l| \leq O(\alpha^{-t'/2})S_{t'} = O(\alpha^{t'/2}\log n). \tag{8.51}$$

On the other hand, for $t \leq T$, we know $S_t = O(\log n)$. Let $t' = T$ in (8.51), we have

$$|S_T - \alpha^{T-l}S_l| = O(\alpha^{T/2}\log n). \tag{8.52}$$

So for $1 \leq t \leq T$,

$$\begin{aligned}|S_t - \alpha^{t-l}S_l| &= O(\log n) + \alpha^{t-T}(S_T + O(\log(n)\alpha^{T/2})) \\ &= O(\log n) + O(\alpha^{t-T/2}\log n) = O(\alpha^{t/2}\log n).\end{aligned} \tag{8.53}$$

The last inequality comes from the inequality $t - T/2 \leq t/2$. Combining (8.51) and (8.53), we have proved (4.3) holds for all $1 \leq t \leq l$.

Using (8.14) and (8.15), we have

$$D_{t+1} = U^+_{t+1} - U^-_{t+1} \leq \beta(U^+_t - U^-_t) + \alpha\varepsilon_t(U^+_t + U^-_t) = \beta D_t + \alpha\varepsilon_t S_t.$$

Similarly, $\beta D_t - \alpha\varepsilon_t S_t \le D_{t+1} \le \beta D_t + \alpha\varepsilon_t S_t$. By iterating, we have for $l \ge t > t' \ge T$,

$$|D_t - \beta^{t-t'} D_{t'}| \le \sum_{s=t'}^{t-1} \alpha\beta^{t-1-s}\varepsilon_s S_s. \tag{8.54}$$

Recall $S_s = O(\log(n)\alpha^{s-T})$, $|D_T| = O(\log n)$, and $\varepsilon_s = \alpha^{-(s-T)/2}$. Taking $t' = T$ in (8.54), for $t > T$,

$$|D_t| = O\left(\log(n)\beta^t\right) + O\left(\sum_{s=T}^{t-1} \alpha\beta^{t-1-s}\log(n)\alpha^{(s-T)/2}\right).$$

Since $1 < \alpha < \beta^2$, it follows that

$$\begin{aligned}\sum_{s=T}^{t-1} \alpha\beta^{t-1-s}\log(n)\alpha^{(s-T)/2} &= \beta^{t-1}\alpha^{1-T/2}\log(n)\sum_{s=T}^{t-1}\left(\frac{\alpha}{\beta^2}\right)^{s/2}\\ &= \beta^{t-1}\alpha^{1-T/2}\log(n)O(\alpha^{T/2}\beta^{-T}) = O(\log(n)\beta^t).\end{aligned}$$

So we have $|D_t| = O(\log n\beta^t)$. The right side of (8.54) is of order

$$\sum_{s=t'}^{t-1} \alpha\beta^{t-1-s}\alpha^{(s-T)/2}\log(n) = O(\log(n)\beta^{t-t'}\alpha^{t'/2}).$$

Thus setting $t = l$ in (8.54), for $l > t' \ge T$, we obtain $D_l - \beta^{l-t'}D_{t'} = O(\log(n)\beta^{l-t'}\alpha^{t'/2})$. Therefore $D_{t'} = \beta^{t'-l}D_l + O(\log(n)\alpha^{t'/2})$ holds for all $T \le t' < l$. For $t' < T$, we have $D_{t'} = O(\log n)$ and

$$\begin{aligned}|D_{t'} - \beta^{t'-l}D_l| &\le O(\log n) + \beta^{t'-T}(|D_T| + O(\log(n)\alpha^{T/2}))\\ &= O(\log n) + O(\beta^{t'-T}\alpha^{T/2}\log n) = O(\alpha^{t'/2}\log n),\end{aligned}$$

where the last estimate is because $\beta^{t'-T} < \alpha^{(t'-T)/2}$ under the condition that $t' < T$. Altogether we have shown (4.4) holds for all $1 \le t' \le l$. This completes the proof of Theorem 4.2. ∎

9 | PROOF OF THEOREM 4.6

We first state the following lemma before proving Theorem 4.6. The proof is included in Appendix A.7.

Lemma 9.1. *For all $m \in \{1, \dots, l\}$ with $l = c\log n$, $c\log\alpha < 1/4$, it holds asymptotically almost surely that*

$$\sup_{\|x\|_2=1, x^\top B^{(l)}\mathbf{1}=x^\top B^{(l)}\sigma=0} \|\mathbf{1}^\top B^{(m-1)}x\|_2 = O(\sqrt{n}\alpha^{(m-1)/2}\log n), \tag{9.1}$$

$$\sup_{\|x\|_2=1, x^\top B^{(l)}\mathbf{1}=x^\top B^{(l)}\sigma=0} \|\sigma^\top B^{(m-1)}x\|_2 = O(\sqrt{n}\alpha^{(m-1)/2}\log n). \tag{9.2}$$

Proof of Theorem 4.6. Using matrix expansion identity (3.2) and the estimates in Theorem 3.1, for any l_2-normalized vector x with $x^\top B^{(l)}\mathbf{1} = x^\top B^{(l)}\sigma = 0$, we have for sufficiently large n, asymptotically almost surely

$$\begin{aligned}
\|B^{(l)}x\|_2 &= \left\| \Delta^{(l)}x + \sum_{m=1}^{l}(\Delta^{(l-m)}\overline{A}B^{(m-1)})x - \sum_{m=1}^{l}\Gamma^{(l,m)}x \right\|_2 \\
&\le \rho(\Delta^{(l)}) + \sum_{m=1}^{l}\rho(\Delta^{(l-m)})\|\overline{A}B^{(m-1)}x\|_2 + \sum_{m=1}^{l}\rho(\Gamma^{(l,m)}) \\
&\le 2n^{\varepsilon}\alpha^{l/2} + \sum_{m=1}^{l}n^{\varepsilon}\alpha^{(l-m)/2}\|\overline{A}B^{(m-1)}x\|_2,
\end{aligned} \tag{9.3}$$

where $\overline{A} = \mathbb{E}_{\mathcal{H}_n}[A|\sigma]$. We have the following expression for entries of $\overline{A}$. If $i \neq j$ and $\sigma_i = \sigma_j = +1$,

$$\overline{A}_{ij} = \frac{a}{\binom{n}{d-1}}\binom{n^+-2}{d-2} + \frac{b}{\binom{n}{d-1}}\left(\binom{n-2}{d-2} - \binom{n^+-2}{d-2}\right) =: \tilde{a}_n^+.$$

If $i \neq j$ and $\sigma_i = \sigma_j = -1$,

$$\overline{A}_{ij} = \frac{a}{\binom{n}{d-1}}\binom{n^--2}{d-2} + \frac{b}{\binom{n}{d-1}}\left(\binom{n-2}{d-2} - \binom{n^--2}{d-2}\right) =: \tilde{a}_n^-.$$

If $\sigma_i \neq \sigma_j$,

$$\overline{A}_{ij} = \frac{b}{\binom{n}{d-1}}\binom{n-2}{d-2} := \tilde{b}_n.$$

We then have $\tilde{a}_n^+, \tilde{a}_n^-, \tilde{b}_n = O(1/n)$. Conditioned on the event $\{|n^\pm - n/2| \le \log(n)\sqrt{n}\}$, we obtain

$$\tilde{a}_n^- - \tilde{a}_n^+ = \frac{a-b}{\binom{n}{d-1}}\left(\binom{n^--2}{d-2} - \binom{n^+-2}{d-2}\right) = O\left(\frac{\log n}{n^{3/2}}\right).$$

Let R be a $n \times n$ matrix such that

$$R_{ij} = \begin{cases} 1 & \sigma_i = \sigma_j = -1 \ \text{ and } \ i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

We then have $\|R\|_2 \le \sqrt{\sum_{ij}R_{ij}^2} \le n$. The following decomposition of $\overline{A}$ holds.

$$\overline{A} = \tilde{a}_n^+\left[\frac{1}{2}(\mathbf{1}\cdot\mathbf{1}^\top + \sigma\sigma^\top) - I\right] + \frac{\tilde{b}_n}{2}(\mathbf{1}\cdot\mathbf{1}^\top - \sigma\sigma^\top) + (\tilde{a}_n^- - \tilde{a}_n^+)R \tag{9.4}$$

$$= \frac{\tilde{a}_n^+ + \tilde{b}_n}{2}\mathbf{1}\cdot\mathbf{1}^\top + \frac{\tilde{a}_n^+ - \tilde{b}_n}{2}\sigma\sigma^\top + \left((\tilde{a}_n^- - \tilde{a}_n^+)R - \tilde{a}_n^+ I\right). \tag{9.5}$$

Since

$$\|(\tilde{a}_n^- - \tilde{a}_n^+)R - \tilde{a}_n^+ I\|_2 \le |\tilde{a}_n^- - \tilde{a}_n^+| \cdot \|R\|_2 + |\tilde{a}_n^+| = O(\log n/\sqrt{n}),$$

by (9.5), we have

$$\|\overline{A}B^{(m-1)}x\|_2 = O\left(\frac{1}{n}\right) \|\mathbf{1}\cdot\mathbf{1}^\top B^{(m-1)}x\|_2 + O\left(\frac{1}{n}\right) \|\sigma\sigma^\top B^{(m-1)}x\|_2 + O\left(\frac{\log n}{\sqrt{n}}\right) \|B^{(m-1)}x\|_2.$$

By Cauchy inequality,

$$\|\mathbf{1}\cdot\mathbf{1}^\top B^{(m-1)}x\|_2 \le \sqrt{n}\|\mathbf{1}^\top B^{(m-1)}x\|_2, \quad \|\sigma\sigma^\top B^{(m-1)}x\|_2 \le \sqrt{n}\|\sigma^\top B^{(m-1)}x\|_2.$$

Therefore,

$$\|\overline{A}B^{(m-1)}x\|_2 = O(n^{-1/2})(\|\sigma^\top B^{(m-1)}x\|_2 + \|\mathbf{1}^\top B^{(m-1)}x\|_2) + O(\log n/\sqrt{n})\|B^{(m-1)}x\|_2.$$

Using (9.1) and (9.2), the right hand side in the expression above is upper bounded by

$$O(\alpha^{(m-1)/2}\log n) + O(\|B^{(m-1)}x\|_2 \cdot \log n/\sqrt{n}). \tag{9.6}$$

Since $B^{(m-1)}$ is a nonnegative matrix, the spectral norm is bounded by the maximum row sum (see Theorem 8.1.22 in [26]), we have that

$$\|B^{(m-1)}x\|_2 \le \rho(B^{(m-1)}) \le \max_i \sum_{j=1}^n B_{ij}^{(m-1)}.$$

By (4.1), (4.5) and (4.7), the right hand side above is $O(\alpha^{m-1}\log n)$. Combing (9.6) and noting that $\alpha^{m-1}/\sqrt{n} = o(n^{-1/4})$, it implies

$$\|\overline{A}B^{(m-1)}x\|_2 = O(\alpha^{(m-1)/2}\log n) + O(\alpha^{m-1}\log^2 n/\sqrt{n}) = O(\alpha^{(m-1)/2}\log n). \tag{9.7}$$

Taking (9.7) into (9.3), we have for any $\varepsilon > 0$, with high probability, $\|B^{(l)}x\|_2 = O(n^{\varepsilon}\alpha^{l/2}\log^2 n) \le n^{2\varepsilon}\alpha^{l/2}$ for n sufficiently large. This completes the proof. ■

10 | PROOF OF THEOREM 5.2

The proof in this section is a generalization of the method in [33] for sparse random graphs. We now prove the case where $\sigma_i = +1$, and the case for $\sigma_i = -1$ can be treated in the same way. Recall the definition of V_t from Definition 4.1. Let A_t be the event that no vertex in V_t is connected by two distinct hyperedges to V_{t-1}. Let B_t be the event that there does not exist two vertices in V_t that are contained in a hyperedge $e \subset \binom{V_t}{d}$.

We can construct the multi-type Poisson hypertree (T,ρ,τ) in the following way. For a vertex $v\in T$, Let $Y_v^{(r)}, 0\leq r\leq d-1$ be the number of hyperedges incident to v which among the remaining $d-1$ vertices, r of them have the same spin with $\tau(v)$. We have

$$Y_v^{(d-1)}\sim \mathrm{Pois}\left(\frac{a}{2^{d-1}}\right),\quad Y_v^{(r)}\sim \mathrm{Pois}\left(\frac{\binom{d-1}{r}b}{2^{d-1}}\right), 0\leq r\leq d-2.$$

Note that (T,ρ,τ) can be entirely reconstructed from the label of the root and the sequence $\{Y_v^{(r)}\}$ for $v\in V(T), 0\leq r\leq d-1$.

We define similar random variables for (H,i,σ). For a vertex $v\in V_t$, let $X_v^{(r)}$ be the number of hyperedges incident to v, where all the remaining $d-1$ vertices are in V_{t+1} such that r of them have spin $\sigma(v)$. Then we have

$$X_v^{(d-1)}\sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{d-1}, \frac{a}{\binom{n}{d-1}}\right),$$
$$X_v^{(r)}\sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{r}\binom{|V_{>t}^{-\sigma(v)}|}{d-1-r}, \frac{b}{\binom{n}{d-1}}\right),\quad 0\leq r\leq d-2$$

and conditioned on $\mathcal{F}_t$ (recall the definition of $\mathcal{F}_t$ from (8.6)) they are independent. Recall Definition 5.1. We have the following lemma on the spin-preserving isomorphism. The proof of Lemma 10.1 is given in Appendix A.8.

Lemma 10.1. *Let $(H,i,\sigma)_t, (T,\rho,\tau)_t$ be the rooted hypergraph truncated at distance t from i,ρ respectively. If*

(1) there is a spin-preserving isomorphism ϕ such that $(H,i,\sigma)_{t-1}\equiv(T,\rho,\tau)_{t-1}$,
(2) for every $v\in V_{t-1}$, $X_v^{(r)}=Y_{\phi(v)}^{(r)}$ for $0\leq r\leq d-1$,
(3) A_t, B_t hold,

then $(H,i,\sigma)_t\equiv(T,\rho,\tau)_t$.

To make our notation simpler, for the rest of this section, we will identify v with $\phi(v)$. Recall the event $\Omega_t(i)=\{S_t(i)\leq C\log(n)\alpha^t\}$ where the constant C is the same one as in Theorem 4.2. Now define a new event

$$C_t := \bigcap_{s\leq t}\Omega_s(i). \tag{10.1}$$

From the proof of Theorem 4.2, for all $t\leq l$, $\mathbb{P}_{\mathcal{H}_n}(C_t)=1-O(n^{-1-\gamma})$ for any $\gamma\in(0,3/8)$. Note that conditioned on C_t, there exists $C'>0$ such that

$$|V_{\leq t}|\leq \sum_{s\leq t}C\log(n)\alpha^t\leq C'\log^2(n)\alpha^t. \tag{10.2}$$

We now estimate the probability of event A_t, B_t conditioned on C_t. The proof is included in Appendix A.9.

Lemma 10.2. *For any $t \geq 1$,*

$$\mathbb{P}(A_t|C_t) \geq 1 - o(n^{-1/2}), \quad \mathbb{P}(B_t|C_t) \geq 1 - o(n^{-1/2}).$$

Before proving Theorem 5.2, we also need the following bound on the total variation distance between binomial and Poisson random variables, see for example Lemma 4.6 in [33].

Lemma 10.3. *Let m, n be integers and c be a positive constant. The following holds:*

$$\left\| \mathrm{Bin}\left(m, \frac{c}{n}\right) - \mathrm{Pois}(c) \right\|_{\mathrm{TV}} = O\left(\frac{1 \vee |m-n|}{n}\right).$$

Proof of Theorem 5.2. Fix t and suppose that C_t holds, and $(T, \rho)_t \equiv (H, i)_t$. Then for each $v \in V_t$, recall

$$X_v^{(d-1)} \sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{d-1}, \frac{a}{\binom{n}{d-1}} \right), \quad X_v^{(r)} \sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{r} \binom{|V_{>t}^{-\sigma(v)}|}{d-1-r}, \frac{b}{\binom{n}{d-1}} \right)$$

and

$$Y_v^{(d-1)} \sim \mathrm{Pois}\left(\frac{a}{2^{d-1}}\right), \quad Y_v^{(r)} \sim \mathrm{Pois}\left(\frac{\binom{d-1}{r} b}{2^{d-1}} \right), \quad 0 \leq r \leq d-2.$$

Recall $|n^{\pm} - n/2| \leq \sqrt{n}\log n$. We have the following bound for $V_{>t}^{\pm}$:

$$\begin{aligned}
|V_{>t}^{\pm}| &\geq n^{\pm} - |V_{\leq t}| \geq \frac{n}{2} - \sqrt{n}\log(n) - O(\log^2(n)\alpha^{2t}) \geq \frac{n}{2} - 2\sqrt{n}\log(n), \\
|V_{>t}^{\pm}| &\leq n^{\pm} \leq \frac{n}{2} + \sqrt{n}\log(n).
\end{aligned}$$

Therefore $|V_{>t}^{\pm} - \frac{n}{2}| \leq 2\sqrt{n}\log n$. Then from Lemma 10.3,

$$\begin{aligned}
\|X_v^{(d-1)} - Y_v^{(d-1)}\|_{\mathrm{TV}} &\leq C \frac{\left| \binom{|V_{>t}^{\sigma(v)}|}{d-1} - \frac{1}{2^{d-1}}\binom{n}{d-1} \right|}{\frac{1}{2^{d-1}}\binom{n}{d-1}} = O(n^{-1/2}\log n), \\
\|X_v^{(r)} - Y_v^{(r)}\|_{\mathrm{TV}} &= O(n^{-1/2}\log n), \quad 0 \leq r \leq d-2.
\end{aligned}$$

We can couple $X_v^{(r)}$ with $Y_v^{(r)}, 0 \leq r \leq d-1$ such that $\mathbb{P}\left(X_v^{(r)} \neq Y_v^{(r)}\right) = O(n^{-1/2}\log n)$. Taking a union bound over all $v \in V_t$, and $0 \leq r \leq d-1$ and recall (10.2), we can find a coupling such that with probability at least

$$1 - O(\log^3(n)\alpha^l n^{-1/2}) \geq 1 - o(n^{-1/4}),$$

$X_v^{(r)} = Y_v^{(r)}$ for every $v \in V_t$ and $0 \leq r \leq d-1$.

Lemma 10.2 implies A_t, B_t, C_t hold simultaneously with probability at least $1 - o(n^{-1/4})$. Altogether we have that assumptions (2),(3) in Lemma 10.1 hold with probability $1 - o(n^{-1/4})$, which can be written as

$$\mathbb{P}\left((H, i, \sigma)_{t+1} \equiv (T, \rho, \tau)_{t+1}, C_{t+1} \middle| (H, i, \sigma)_t \equiv (T, \rho, \tau)_t, C_t\right) \geq 1 - o(n^{-1/4}).$$

Since we can certainly couple i with ρ from our construction, we have $\mathbb{P}\left((H, i, \sigma)_0 \equiv (T, \rho, \tau)_0, C_0\right) = 1$. Therefore for large n,

$$\begin{aligned}
&\mathbb{P}((H, i, \sigma)_l \equiv (T, \rho, \tau)_l) \\
&= \prod_{t=1}^{l} \mathbb{P}\left((H, i, \sigma)_t \equiv (T, \rho, \tau)_t, C_t \middle| (H, i, \sigma)_{t-1} \equiv (T, \rho, \tau)_{t-1}, C_{t-1}\right) \cdot \mathbb{P}\left((H, i, \sigma)_0 \equiv (T, \rho, \tau)_0, C_0\right) \\
&\geq (1 - o(n^{-1/4}))^l \geq 1 - n^{-1/5}.
\end{aligned}$$

This completes the proof. ∎

11 | PROOF OF THEOREM 6.1

The proof of the following Lemma 11.1 follows in a similar way as Lemma 4.4 in [32], and we include it in Appendix A.10.

Lemma 11.1. *For $l = c\log(n), c\log(\alpha) < 1/4$, the following hold asymptotically almost surely*

$$\|B^{(l)}\mathbf{1} - \vec{S}_l\|_2 = o(\|B^{(l)}\mathbf{1}\|_2), \tag{11.1}$$

$$\|B^{(l)}\sigma - \vec{D}_l\|_2 = o(\|B^{(l)}\sigma\|_2), \tag{11.2}$$

$$\langle B^{(l)}\mathbf{1}, B^{(l)}\sigma\rangle = o\left(\|B^{(l)}\mathbf{1}\|_2 \cdot \|B^{(l)}\sigma\|_2\right). \tag{11.3}$$

The next lemma estimate $\|B^{(l)}x\|_2$ when $x = B^{(l)}\sigma$ and $B^{(l)}\mathbf{1}$. The proof of Lemma 11.2 is provided in Appendix A.11.

Lemma 11.2. *Assume $\beta^2 > \alpha > 1$ and $l = c\log(n)$ with $c\log(\alpha) < 1/8$. Then for some fixed $\gamma > 0$ asymptotically almost surely one has*

$$\Omega(\alpha^l)\|B^{(l)}\mathbf{1}\|_2 \leq \|B^{(l)}B^{(l)}\mathbf{1}\|_2 \leq O(\alpha^l \log n)\|B^{(l)}\mathbf{1}\|_2, \tag{11.4}$$

$$\Omega(\beta^l)\|B^{(l)}\sigma\|_2 \leq \|B^{(l)}B^{(l)}\sigma\|_2 \leq O(n^{-\gamma}\alpha^l)\|B^{(l)}\sigma\|_2. \tag{11.5}$$

Together with Lemmas 11.1 and 11.2, we are ready to prove Theorem 6.1.

Proof of Theorem 6.1. From Theorem 4.6 and Lemma 11.2, the top two eigenvalues of $B^{(l)}$ will be asymptotically in the span of $B^{(l)}\mathbf{1}$ and $B^{(l)}\sigma$. By the lower bound in (11.4) and the upper bound in (11.5), the largest eigenvalue of $B^{(l)}$ will be $\Theta(\alpha^l)$ up to a logarithmic factor, and the first eigenvector is asymptotically aligned with $B^{(l)}\mathbf{1}$.

From (11.1), $B^{(l)}\mathbf{1}$ is also asymptotically aligned with $\vec{S}_l$, therefore our statement for the first eigenvalue and eigenvector holds. Since $B^{(l)}\mathbf{1}$ and $B^{(l)}\sigma$ are asymptotically orthogonal from (11.3), together with (11.5), the second eigenvalue of $B^{(l)}$ is $\Omega(\beta^l)$ and the second eigenvector is asymptotically aligned with $B^{(l)}\sigma$.

From (11.2), $B^{(l)}\sigma$ is asymptotically aligned with $\vec{D}_l$. So the statement for the second eigenvalue and eigenvector holds. The order of other eigenvalues follows from Theorem 4.6 and the Courant minimax principle (see [26]). ■

REFERENCES

1. E. Abbe, *Community detection and stochastic block models: Recent developments*, J. Mach. Learn. Res. **18** (2018), 1–86.
2. E. Abbe, A. S. Bandeira, and G. Hall, *Exact recovery in the stochastic block model*, IEEE Trans. Inform. Theory **62** (2016), 471–487.
3. E. Abbe et al., *Graph powering and spectral robustness*, SIAM J. Math. Data Sci. **2** (2020), 132–157.
4. E. Abbe and C. Sandon, *Proof of the achievability conjectures for the general stochastic block model*, Comm. Pure Appl. Math. **71** (2018), 1334–1406.
5. S. Agarwal et al., *Beyond pairwise clustering*, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol **2**, IEEE, New York, 2005, 838–845.
6. M. Chiara Angelini et al., *Spectral detection on sparse hypergraphs*, in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, New York, 2015, 66–73.
7. C. Bordenave, S. Coste, and R. R. Nadakuditi, *Detection thresholds in very sparse matrix completion*. arXiv preprint arXiv:2005.06062, 2020.
8. C. Bordenave, M. Lelarge, and L. Massoulié, *Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs*, Ann. Probab. **46** (2018), 1–71.
9. S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, Oxford, 2013.
10. G. Brito et al., *Recovery and rigidity in a regular stochastic block model*, in *Proceedings of the Twenty-seventh Annual ACM–SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, New York, 2016, 1589–1601.
11. I. Chien, C.-Y. Lin, and I.-H. Wang, *Community detection in hypergraphs: Optimal statistical limit and efficient algorithms*, in *International Conference on Artificial Intelligence and Statistics*, Microtome Publishing, Brookline, MA, 2018, 871–879.
12. I. Eli Chien, C.-Y. Lin, and I.-H. Wang, *On the minimax misclassification ratio of hypergraph community detection*, IEEE Trans. Inform. Theory **65** (2019), 8095–8118.
13. S. Cole and Y. Zhu, *Exact recovery in the hypergraph stochastic block model: A spectral algorithm*, Linear Algebra Appl. **593** (2020), 45–73.
14. A. Decelle et al., *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*. Phys. Rev. E (3), **84**:066106, 2011.
15. P. Delgosha and V. Anantharam, *Load balancing in hypergraphs*, J. Statist. Phys. **173** (2018), 546–625.
16. W. Evans et al., *Broadcasting on trees and the ising model*, Ann. Appl. Probab. **10** (2000), 410–433.
17. L. Florescu and W. Perkins, *Spectral thresholds in the bipartite stochastic block model*, in *Conference on Learning Theory*, Microtome Publishing, Brookline, MA, 2016, 943–959.
18. D. Ghoshdastidar and A. Dukkipati, *Consistency of spectral partitioning of uniform hypergraphs under planted partition model*, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2014, 397–405.
19. D. Ghoshdastidar and A. Dukkipati, *A provable generalized tensor spectral method for uniform hypergraph partitioning*, in *International Conference on Machine Learning*, Microtome Publishing, Brookline, MA, 2015, 400–409.
20. D. Ghoshdastidar and A. Dukkipati, *Spectral clustering using multilinear SVD: Analysis, approximations and applications*, AAAI **29** (2015), 2610–2616.
21. D. Ghoshdastidar and A. Dukkipati, *Consistency of spectral hypergraph partitioning under planted partition model*, Ann. Statist. **45** (2017), 289–315.
22. O. Guédon and R. Vershynin, *Community detection in sparse networks via Grothendieck's inequality*, Probab. Theory Related Fields **165** (2016), 1025–1049.
23. B. Hajek, Y. Wu, and J. Xu, *Achieving exact cluster recovery threshold via semidefinite programming*, IEEE Trans. Inform. Theory **62** (2016), 2788–2797.
24. C. Hennig et al., *Handbook of cluster analysis*, CRC Press, New York, NY, 2015.
25. C. J. Hillar and L.-H. Lim, *Most tensor problems are np-hard*, J. ACM **60** (2013), 45.
26. R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, Cambridge, UK, 2012.

27. C. Kim, A. S. Bandeira, and M. X. Goemans. *Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach*. arXiv preprint arXiv:1807.02884, 2018.
28. F. Krzakala et al., *Spectral redemption in clustering sparse networks*, Proc. Natl. Acad. Sci. U.S.A. **110** (2013), 20935–20940.
29. C.-Y. Lin, I. Eli Chien, and I-H. Wang, *On the fundamental statistical limit of community detection in random hypergraphs*, in *2017 IEEE International Symposium on Information Theory (ISIT)*, IEEE, New York, 2017, 2178–2182.
30. Q. Liu, Y. Huang, and D. N. Metaxas, *Hypergraph with sampling for image retrieval*, Pattern Recogn. **44** (2011), 2255–2262.
31. L. Lu and X. Peng, *Loose laplacian spectra of random hypergraphs*, Random Structures Algorithms **41** (2012), 521–545.
32. L. Massoulié, *Community detection thresholds and the weak Ramanujan property*, in *Proceedings of the Forty-Sixth Annual ACM symposium on Theory of Computing*, ACM, New York, NY, 2014, 694–703.
33. E. Mossel, J. Neeman, and A. Sly, *Reconstruction and estimation in the planted partition model*, Probab. Theory Related Fields **162** (2015), 431–461.
34. E. Mossel, J. Neeman, and A. Sly, *Belief propagation, robust reconstruction and optimal recovery of block models*, Ann. Appl. Probab. **26** (2016), 2211–2256.
35. E. Mossel, J. Neeman, and A. Sly, *A proof of the block model threshold conjecture*, Combinatorica **38** (2018), 665–708.
36. L. Stephan and L. Massoulié, *Robustness of spectral methods for community detection*, in *Conference on Learning Theory*, PMLR, Brookline, MA, 2019, 2831–2860.
37. L. Stephan and L. Massoulié. *Non-backtracking spectra of weighted inhomogeneous random graphs*. arXiv preprint arXiv:2004.07408, 2020.
38. S. Tan et al., *Using rich social media information for music recommendation via hypergraph model*, ACM Trans. Multimedia Comput. Commun. Appl. **7** (2011), 22.
39. Z. Tian, T. H. Hwang, and R. Kuang, *A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge*, Bioinformatics **25** (2009), 2831–2838.
40. D. Zhou, J. Huang, and B. Schölkopf, *Learning with hypergraphs: Clustering, classification, and embedding*, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2007, 1601–1608.

How to cite this article: Pal S, Zhu Y. Community detection in the sparse hypergraph stochastic block model. *Random Struct Alg* 2021;59:407–463. https://doi.org/10.1002/rsa.21006

APPENDIX A

A.1 | Proof of Lemma 4.3

Proof. The two sequences $(U_k^{\pm}(i))_{k\leq l}$, $(U_k^{\pm}(j))_{k\leq l}$ are independent conditioned on the event $\{V_{\leq l}(i)\cap V_{\leq l}(j)=\emptyset\}$. It remains to estimate $\mathbb{P}_{\mathcal{H}_n}\left(\{V_{\leq l}(i)\cap V_{\leq l}(j)=\emptyset\}\right)$. Introduce the events

$$\mathcal{J}_k := \bigcap_{t\leq k}\{S_t(i)\vee S_t(j)\leq C\log(n)\alpha^t\},\quad \mathcal{L}_k := \{V_{\leq k}(i)\bigcap V_{\leq k}(j)=\emptyset\},$$

where the constant C is the same one as in the statement of Theorem 4.2. For any vertex $v\in[n]\setminus(V_{\leq k}(i)\cup V_{\leq k}(j))$, Conditioned on $\mathcal{L}_k$ and $\mathcal{J}_k$, there are two possible situations where v is included in $V_{k+1}(i)\cap V_{k+1}(j)$:

1. There is a hyperedge containing v and a vertex in $V_k(i)$, and a different hyperedge containing v and a vertex in $V_k(j)$.
2. There is a hyperedge containing v, one vertex in $V_k(i)$, and another vertex in $V_k(j)$.

There exists a constant $C_1 > 0$ such that Case (1) happens with probability at most

$$S_k(i)S_k(j)\binom{n}{d-2}^2\left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2 \leq C_1 \log^2(n)\alpha^{2k}/n^2,$$

and Case (2) happens with probability at most

$$S_k(i)S_k(j)\binom{n}{d-3}\frac{a \vee b}{\binom{n}{d-1}} = C_1 \log^2(n)\alpha^{2k}/n^2.$$

Since $\alpha^{2l} = n^{2c\log\alpha} = o(n^{1/2})$, we have for large n,

$$\mathbb{P}_{\mathcal{H}_n}(v \in V_{k+1}(i) \cap V_{k+1}(j) | \mathcal{J}_k, \mathcal{L}_k) \leq 2C_1 \log^2(n)\alpha^{2l}/n^2 < n^{-1.5}.$$

Taking a union bound over all possible v, we have for some constant $C_3 > 0$,

$$\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset | \mathcal{J}_k, \mathcal{L}_k) \geq 1 - C_3 n^{-1/2}.$$

From the proof of Theorem 4.2, for all $0 \leq k \leq l$, $\mathbb{P}_{H_n}(\mathcal{J}_k) = 1 - O(n^{-1-\gamma})$ for any $\gamma \in (0, 3/8)$. We then have

$$\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset | \mathcal{L}_k) \geq \mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset | \mathcal{J}_k, \mathcal{L}_k)\ \mathbb{P}_{\mathcal{H}_n}(\mathcal{J}_k) \geq 1 - O(n^{-1/2}).$$

Finally, for large n,

$$\begin{aligned}
\mathbb{P}_{\mathcal{H}_n}\left(\{V_{\leq l}(i) \cap V_{\leq l}(j) = \emptyset\}\right) &= \mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_l) \geq \mathbb{P}_{\mathcal{H}_n}(V_l(i) \cap V_l(j) = \emptyset | \mathcal{L}_{l-1})\mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_{l-1}) \\
&\geq \mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_0)\prod_{k=0}^{l-1}\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset | \mathcal{L}_k) \\
&\geq (1 - O(n^{-1/2}))^l \geq 1 - n^{-1/3}.
\end{aligned}$$

This completes the proof. ■

A.2 | Proof of Lemma 4.4

Proof. Consider the exploration process of the neighborhood of a fixed vertex i. Conditioned on $\mathcal{F}_{k-1}$, there are two ways to create new cycles in $V_{\geq k-1}(i)$:

(1) Type 1: a new hyperedge $e \subset V_{\geq k-1}(i)$ containing two vertices in $V_{k-1}(i)$ may appear, which creates a cycle including two vertices in $V_{k-1}(i)$.
(2) Type 2: two vertices in $V_{k-1}(i)$ may be connected to the same vertex in $V_{\geq k}(i)$ by two new distinct hyperedges.

Define the event

$$\Omega_{k-1}(i) := \{S_{k-1}(i) \le C\log(n)\alpha^{k-1}\}, \tag{A1}$$

where the constant C is the same one as in Theorem 4.2. From the proof of Theorem 4.2, $\mathbb{P}_{\mathcal{H}_n}(\Omega_k(i)) = 1 - O(n^{-1-\gamma})$ for some $\gamma \in (0, 3/8)$.

Let $E_k^{(1)}(i)$ be the number of hyperedges of type 1. Conditioned on $\mathcal{F}_{k-1}$, $E_k^{(1)}(i)$ is stochastically dominated by Bin $\left(\binom{S_{k-1}(i)}{2}\binom{n}{d-2}, \frac{a\vee b}{\binom{n}{d-1}}\right)$. Then for some constant $C_1 > 0$,

$$\mathbb{E}_{\mathcal{H}_n}[E_k^{(1)}(i)|\Omega_{k-1}(i)] \le C_1\log^2(n)\alpha^{2k-2}/n \le C_1\log^2(n)\alpha^{2l}/n.$$

By Markov's inequality,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \ge 1\}) &\le \mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \ge 1\}|\Omega_{k-1}(i)) + \mathbb{P}_{\mathcal{H}_n}(\Omega_{k-1}^c(i)) \\ &\le \mathbb{E}_{\mathcal{H}_n}[E_k^{(1)}(i)|\Omega_{k-1}(i)] + O(n^{-1-\gamma}) = O(\log^2(n)\alpha^{2l}/n).\end{aligned}$$

Taking the union bound, the probability that there is a type 1 hyperedge in the l-neighborhood of i is

$$\mathbb{P}_{\mathcal{H}_n}\left(\bigcup_{k=1}^{l}\{E_k^{(1)}(i) \ge 1\}\right) \le \sum_{k=1}^{l}\mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \ge 1\}) = O(\log^3(n)\alpha^{2l}/n).$$

The number of hyperedge pair (e_1, e_2) of Type 2 is stochastically dominated by

$$\text{Bin}\left(nS_{k-1}^2\binom{n}{d-2}^2, \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^2\right),$$

which conditioned on $\Omega_{k-1}(i)$ has expectation $O(\log^2(n)\alpha^{2l}/n)$. By a Markov's inequality and a union bound, in the same way as the proof for Type 1, we have the probability there is a type 2 hyperedge pair in the l-neighborhood of i is $O(\log^2(n)\alpha^{2l}/n)$. Altogether the probability that there are at least one cycles within the $l-$neighborhood of i is $O(\log^3(n)\alpha^{2l}/n)$.

Let Z_i be the random variable such that $Z_i = 1$ if l-neighborhood of i contains one cycle and $Z_i = 0$ otherwise. From the analysis above, we have $\mathbb{E}[Z_i] = O(\log^3(n)\alpha^{2l}/n)$. By Markov's inequality,

$$\mathbb{P}_{\mathcal{H}_n}\left(\sum_{i\in[n]}Z_i \ge \alpha^{2l}\log^4(n)\right) \le \frac{\sum_i\mathbb{E}[Z_i]}{\log^4(n)\alpha^{2l}} = \frac{O(\log^3(n)\alpha^{2l})}{\alpha^{2l}\log^4(n)} = O(\log^{-1}(n)).$$

Then asymptotically almost surely the number of vertices whose l-neighborhood contains one cycle at most $\log^4(n)\alpha^{2l}$.

It remains to show H is l-tangle free asymptotically almost surely. For a fixed vertex $i \in [n]$, there are several possible cases where there can be two cycles in $V_{\le l}(i)$.

(1) There is one hyperedge of Type 1 or a hyperedge pair of Type 2 which creates more than one cycles. We discuss in the following cases conditioned on the event $\cap_{t=1}^{l}\Omega_t(i)$.

(a) The number of hyperedge of the first type which connects to more than two vertices in V_{k-1} is stochastically dominated by Bin $\left(\binom{S_{k-1}}{3}\binom{n}{d-3}, \frac{a\vee b}{\binom{n}{d-1}}\right)$. The expectation is at most $O(\alpha^{3l}\log^3(n)/n^2)$.

(b) If the intersection of the hyperedge pair of Type 2 contains 2 vertices in $V_{\geq k}$, it will create two cycles. The number of such hyperedge pairs is stochastically dominated by $\mathrm{Bin}\left(\binom{n}{2} S_{k-1}^2 \binom{n}{d-3}^2, \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2\right)$ with mean $O(\log^2(n)\alpha^{2l}/n^2)$.

Then by Markov's inequality and a union bound, asymptotically almost surely, there is no $V_{\leq l}(i)$ such that its neighborhood contains Type 1 hyperedges or Type 2 hyperedge pairs which create more than one cycles.

(2) The remaining case is that there is a $V_{\leq l}(i)$ where two cycles are created by two Type 1 hyperedges or two Type 2 hyperedge pairs or one Type 1 hyperedge and another hyperedge pairs. By the same argument, under the event $\cap_{t=1}^{l}\Omega_t(i)$, the probability that such event happens is $O(\log^6(n)\alpha^{4l}/n^2)$. Since $\alpha^{4l} = o(n)$, by taking a union bound over $i \in [n]$, we have H is l-tangle-free asymptotically almost surely. ■

A.3 | Proof of Lemma 4.5

Proof. Let $i \notin \mathcal{B}$ whose l-neighborhood contains no cycles. For any $k \in [n]$ and any $m \leq l$, there is a unique self-avoiding walk of length m from i to k if and only if $d(i,k) = m$, so we have $B_{ik}^{(m)} = \mathbf{1}_{d(i,k)=m}$. For such i we have

$$(B^{(m)}\mathbf{1})_i = S_m(i), \quad (B^{(m)}\sigma)_i = D_m(i).$$

Then (4.5), (4.6) follows from Theorem 4.2.

By Lemma 4.4, asymptotically almost surely all vertices in $\mathcal{B}$ have only one cycle in l-neighborhood. For any $m \leq l, i \in \mathcal{B}$, since $(B^{(m)}\mathbf{1})_i = \sum_{k\in[n]} B_{ik}^{(m)}$, and only vertices at distance at most m from i can be reached by a self-avoiding walk of length m from i, which will be counted in $(B^{(m)}\mathbf{1})_i$. Moreover, for any $k \in [n]$ with $B_{ik}^{(m)} \neq 0$, since the l-neighborhood of i contains at most one cycle, there are at most 2 self-avoiding walks of length m between i and k. Altogether we know

$$\sum_{k\in[n]} B_{ik}^{(m)} \leq 2\sum_{t=0}^{m} S_t(i) = O(\alpha^m \log n)$$

asymptotically almost surely. Then (4.7) follows. ■

A.4 | Proof of Lemma 5.3

Proof. Recall the definitions of α, β from (1.3). From (5.1)–(5.3),

$$\begin{aligned}\mathbb{E}(W_{t+1}^+|\mathcal{G}_t) &= \sum_{r=0}^{d-1} r\mathbb{E}(W_{t+1}^{(r)}|\mathcal{G}_t) = \sum_{r=1}^{d-2} r\left(\frac{b\binom{d-1}{r}}{2^{d-1}}(W_t^- + W_t^+)\right) + (d-1)\left(\frac{a}{2^{d-1}}W_t^+ + \frac{b}{2^{d-1}}W_t^-\right)\\ &= \frac{\alpha+\beta}{2}W_t^+ + \frac{\alpha-\beta}{2}W_t^- = \frac{\alpha^{t+1}}{2}M_t + \frac{\beta^{t+1}}{2}\Delta_t.\end{aligned}$$

Similarly, $\mathbb{E}[W_{t+1}^-|\mathcal{G}_t] = \frac{\alpha^{t+1}}{2}M_t - \frac{\beta^{t+1}}{2}\Delta_t$. Therefore

$$\mathbb{E}[M_{t+1}|\mathcal{G}_t] = \alpha^{-t-1}\mathbb{E}[W_{t+1}^+ + W_{t+1}^-|\mathcal{G}_t] = M_t,$$
$$\mathbb{E}[\Delta_{t+1}|\mathcal{G}_t] = \beta^{-t-1}\mathbb{E}[W_{t+1}^+ - W_{t+1}^-|\mathcal{G}_t] = \Delta_t.$$

It follows that $\{M_t\}, \{\Delta_t\}$ are martingales with respect to $\mathcal{G}_t$. From (5.1)–(5.4),

$$\mathrm{Var}(M_t|\mathcal{G}_{t-1}) = \mathrm{Var}(\alpha^{-t}(W_t^+ + W_t^-)|\mathcal{G}_{t-1}) = \alpha^{-2t}\mathrm{Var}\left((d-1)\sum_{r=0}^{d-1} W_t^{(r)}|\mathcal{G}_{t-1}\right)$$

$$= (d-1)^2\alpha^{-2t} \cdot \frac{\alpha}{d-1}(W_{t-1}^{+} + W_{t-1}^{-}) = (d-1)\alpha^{-t}M_{t-1}.$$

Sine $\mathbb{E}M_0 = 1$, by conditional variance formula,

$$\mathrm{Var}(M_t) = \mathrm{Var}(\mathbb{E}[M_t|\mathcal{G}_{t-1}]) + \mathbb{E}\mathrm{Var}(M_t|\mathcal{G}_{t-1}) = \mathrm{Var}(M_{t-1}) + (d-1)\alpha^{-t}.$$

Since $\mathrm{Var}(M_0)=0$, we have for $t\geq 0$, $\mathrm{Var}(M_t) = (d-1)\frac{1-\alpha^{-t}}{\alpha-1}$. So $\{M_t\}$ is uniformly integrable for $\alpha > 1$.

Similarly,

$$\mathrm{Var}(\Delta_t|\mathcal{G}_{t-1}) = \mathrm{Var}(\beta^{-t}(W_t^{+} - W_t^{-})|\mathcal{G}_{t-1}) = \beta^{-2t}\sum_{r=0}^{d-1}(2r-d+1)^2\mathrm{Var}(W_t^{(r)}|\mathcal{G}_{t-1})$$
$$= (\alpha/\beta^2)^t M_{t-1}(d-1)\alpha^{-1} \cdot \frac{(d-1)a + (2^{d-1}+1-d)b}{2^{d-1}} =: \kappa(\alpha/\beta^2)^t M_{t-1},$$

where $\kappa := \frac{(d-1)(a-b)+2^{d-1}b}{a+(2^{d-1}-1)b}$.

And we also have the following recursion:

$$\mathrm{Var}(\Delta_t) = \mathrm{Var}(\mathbb{E}[\Delta_t|\mathcal{G}_{t-1}]) + \mathbb{E}\mathrm{Var}(\Delta_t|\mathcal{G}_{t-1}) = \mathrm{Var}(\Delta_{t-1}) + \kappa\beta^{-2t}\alpha^t.$$

Since $\mathrm{Var}(\Delta_0) = 0$, we have for $t > 0$,

$$\mathrm{Var}(\Delta_t) = \kappa \cdot \frac{1-(\beta^2/\alpha)^{-t}}{\beta^2/\alpha - 1}. \tag{A2}$$

So $\{\Delta_t\}$ is uniformly integrable if $\beta^2 > \alpha$. From the martingale convergence theorem, $\mathbb{E}\Delta_\infty = \Delta_0 = 1$, $\mathrm{Var}(\Delta_\infty) = \frac{\kappa}{\beta^2/\alpha-1}$, and (5.5) holds. This finishes the proof. ■

A.5 | Proof of Lemma 5.4

Proof. From Theorem 5.2, For each $i \in [n]$, there exists a coupling such that with probability $1 - O(n^{-\varepsilon})$ for some positive ε, $\beta^{-l}\sigma(i)D_l(i) = \Delta_l$ and we denote this event by C. When the coupling fails, by Theorem 4.2, $\beta^{-l}\sigma(i)D_l(i) = O(\log(n))$ with probability $1 - O(n^{-\gamma})$ for some $\gamma > 0$.

Recall the event

$$\Omega_{k-1}(i) := \{S_{k-1}(i) \leq C\log(n)\alpha^{k-1}\}. \tag{A3}$$

We define $\Omega := \bigcap_{i=1}^n \Omega(i), \Omega(i) := \bigcap_{k\leq l}\Omega_k(i)$. We have

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \beta^{-2l}D_l^2(i)|\Omega\right) = O(\log^2(n))n^{-\varepsilon} + \mathbb{E}(\Delta_l^2\mathbf{1}_C|\Omega). \tag{A4}$$

Moreover,

$$|\mathbb{E}(\Delta_l^2\mathbf{1}_C|\Omega) - \mathbb{E}(\Delta_\infty^2)| = \left|\frac{\mathbb{E}(\Delta_l^2\mathbf{1}_C - \mathbb{E}(\Delta_l^2\mathbf{1}_C\mathbf{1}_{\overline{\Omega}}) - \mathbb{P}(\Omega)\mathbb{E}(\Delta_\infty^2)}{\mathbb{P}(\Omega)}\right|$$
$$\leq \frac{|\mathbb{E}(\Delta_l^2 - \Delta_\infty^2)|}{\mathbb{P}(\Omega)} + \frac{1-\mathbb{P}(\Omega)}{\mathbb{P}(\Omega)}\mathbb{E}(\Delta_\infty^2) + \frac{|\mathbb{E}(\Delta_l^2\mathbf{1}_{\overline{C}}) - \mathbb{E}(\Delta_l^2\mathbf{1}_{C\cap\overline{\Omega}})|}{\mathbb{P}(\Omega)}. \tag{A5}$$

Since we know $\mathbb{P}(\Omega \cap C) \to 1$ and (5.5), the first two terms in (A5) converges to 0. The third term also converges to 0 by dominated convergence theorem. So we have

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i)|\Omega\right) \to \mathbb{E}(\Delta_\infty^2).$$

We then estimate the second moment. Note that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i)|\Omega\right)^2 = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\beta^{-4l}D_l^4(i)|\Omega\right) + \frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(D_l(i)^2D_l^2(j)|\Omega), \tag{A6}$$

and from Theorem 4.2, the first term is $O(\log^4(n)/n) = o(1)$. Next, we show the second term satisfies

$$\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(D_l(i)^2D_l^2(j)|\Omega) = \frac{2}{n^2}\sum_{i<j}\beta^{-4l}\frac{1}{\mathbb{P}(\Omega)}\mathbb{E}(\mathbf{1}_\Omega D_l(i)^2D_l^2(j)) = o(1). \tag{A7}$$

Since $\mathbb{P}(\Omega) = 1 - O(n^{-\gamma})$, it suffices to show

$$\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_\Omega D_l(i)^2D_l^2(j)) = o(1).$$

Consider $\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l^2(i)D_l^2(j))$. From Lemma 4.3, when $i \neq j$, $D_l(i), D_l(j)$ are asymptotically independent. On the event that the coupling with independent copies fails (recall the failure probability is $O(n^{-\gamma})$), we bound $D_l^2(i)D_l^2(j)$ by $O(\beta^{4l}\log^4(n))$. When the coupling succeeds,

$$\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l(i)^2D_l^2(j)) = \beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)}D_l(i)^2)\mathbb{E}(\mathbf{1}_{\Omega(j)}D_l(j)^2).$$

Then from (5.6),

$$\begin{aligned}\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l(i)^2D_l^2(j)) &= O\left(\frac{1}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)}D_l(i)^2)\mathbb{E}(\mathbf{1}_{\Omega(j)}D_l(j)^2) + O(n^{-2\gamma}\log^4 n)\right)\\ &= O\left((\mathbb{E}(\Delta_\infty^2))^2\right) = O(1).\end{aligned} \tag{A8}$$

Therefore from (A6), (A7), and (A8),

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i)|\Omega\right)^2 = O(1).$$

With (A4), by Chebyshev's inequality, conditioned on Ω, in probability we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i) = \mathbb{E}(\Delta_\infty^2).$$

Since $\mathbb{P}(\Omega) \to 1$, (5.6) follows.

We now establish (5.7). Without loss of generality, we discuss the case of + sign. Since τ is a continuous point of the distribution of Δ_∞, for any fixed $\delta > 0$, we can find two bounded K-Lipschitz function f, g for some constant $K > 0$ such that

$$f(x) \leq (\mathbf{1}_{x \geq \tau}) \leq g(x), x \in \mathbb{R}, \quad 0 \leq \mathbb{E}(g(\Delta_\infty) - f(\Delta_\infty)) \leq \delta.$$

Consider the empirical sum $\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(x_i^{(n)}\sqrt{n\mathbb{E}(\Delta_\infty^2})$, we have

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(x_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2}) - \frac{1}{n}\sum_{i\in\mathcal{N}^+} f(\beta^{-l}D_l(i))\right| \\ &\leq \frac{K}{n}\sum_{i\in\mathcal{N}^+} |(x_i^{(n)} - y_i^{(n)})\sqrt{n\mathbb{E}\Delta_\infty^2}| + \frac{K}{n}\sum_{i\in\mathcal{N}^+} |y_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2} - \beta^{-l}D_l(i)|.\end{aligned}$$

The first term converges to 0 by the assumption that $\|x - y\|_2 \to 0$ in probability. The second term converges to 0 in probability from (5.6). Moreover, $\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(\beta^{-l}D_l(i))$ converges in probability to $\frac{1}{2}\mathbb{E}f(\Delta_\infty)$. So we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(x_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2}) = \frac{1}{2}\mathbb{E}f(\Delta_\infty),$$

and the same holds for g. If follows that

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i\in[n]:\sigma_i=+} \mathbf{1}_{\left\{x_i^{(n)} \geq \tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]}\right\}} - \frac{1}{2}\mathbb{P}(\Delta_\infty \geq \tau)\right| \leq \delta$$

for any $\delta > 0$. Therefore (5.7) holds. ∎

A.6 | Proof of Lemma 7.2

Proof. For any $n \times n$ real matrix M, we have $\rho(M)^{2k} \leq \mathrm{tr}[(MM^\top)^k]$, therefore

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_n}[\rho(\Gamma^{(l,m)})^{2k}] &\leq \mathbb{E}_{\mathcal{H}_n}\left[\mathrm{tr}\left(\Gamma^{(l,m)}\Gamma^{(l,m)\top}\right)^k\right] \\ &= \sum_{i_1,\dots,i_{2k}\in[n]} \mathbb{E}_{\mathcal{H}_n}\left[\Gamma^{(l,m)}_{i_1 i_2}\Gamma^{(l,m)}_{i_3 i_2}\dots\Gamma^{(l,m)}_{i_{2k-1} i_{2k}}\Gamma^{(l,m)}_{i_1 i_{2k}}\right].\end{aligned} \tag{A9}$$

Recall the definition of $\Gamma_{ij}^{(l,m)}$ from (7.2), the sum in (A9) can be expanded to be the sum over all circuits $w = (w_1, \dots w_{2k})$ of length $2kl$ which are obtained by concatenation of $2k$ walks of length l, and each $w_i, 1 \leq i \leq 2k$ is a concatenation of two self-avoiding walks of length $l - m$ and $m - 1$. The weight that each hyperedge in the circuit contributes can be either $A_{ij}^e - \overline{A_{ij}^e}, \overline{A_{ij}^e}$ or A_{ij}^e. For all circuits w in (A9) with nonzero expected weights, there is an extra constraint that each w_i intersects with some other w_j, otherwise the expected weight that w_i contributes to the sum (A9) will be 0. We want to bound the number of such circuits with nonzero expectation.

Let v, h denoted the number of distinct vertices and hyperedges traversed by the circuit. Here we don't count the hyperedges that are weighted by $\overline{A_{ij}^e}$. We associate a multigraph $G(w)$ for each w as before, but the hyperedges with weight $\overline{A_{ij}^e}$ are not included. Since $\mathbb{E}_{\mathcal{H}_n}[\Gamma_{ij}^{(l,m)}] = 0$ for any $i, j \in [n]$, if the expected weight of w is nonzero, the corresponding graph $G(w)$ must be connected.

We detail the proof for circuits in Case (1), where

- each hyperedge label in $\{e_i\}_{1\le i\le h}$ appears exactly once on $G(w)$;
- vertices in $e_i \setminus \mathrm{end}(e_i)$ are all distinct for $1 \le i \le h$, and they are not vertices with labels in $V(w)$,

and the cases for other circuits follow similarly from the proof of Lemma 7.1.

Let m be fixed. For each circuit w, there are $4k$ self-avoiding walks, and each w_i is broken into two self-avoiding walks of length $m-1$ and $l-m$, respectively. We adopt the way of encoding each self-avoiding walk as before, except that we must also include the labels of the endpoint j after the traversal of an edge e with weight from $\overline{A_{ij}^e}$, which gives us the initial vertex of the self-avoiding walk of length $l-m$ within each w_i. These extra labels tell us how to concatenate the two self-avoiding walks of length $m-1$ and $l-m$ into the walk w_i of length l. For each w_i, label is encoded by a number from $\{1, \dots, v\}$. So all possible such labels can be bounded by v^{2k}. Then the upper bound on the number of valid triplet sequences with extra labels for fixed v, h is now given by $v^{2k}[(v+1)^2(l+1)]^{4k(2+h-v)}$.

The total number of circuits that have the same triplet sequences with extra labels is at most $n^v \binom{n}{d-2}^{h+2k}$ where $h+2k$ is the total number of distinct hyperedges we can have in w, including the hyperedges with weights from $\overline{A_{ij}^e}$.

We also need to bound the possible range of v, h. There are overall $2k(l-1)$ hyperedges traversed in w (remember we don't count the edges with weights from $\overline{A_{ij}^e}$). Out of these, $2k(l-m)$ hyperedges (with multiplicity) with weights coming from $A_{ij}^e - \overline{A_{ij}^e}$ must be at least doubled for the expectation not to vanish. Then the number of distinct hyperedges in w excluding the hyperedge weighted by some $\overline{A_{ij}^e}$, satisfies $h \le k(l-m) + (2k(l-1) - 2k(l-m)) = k(l+m-2)$. We have $v \ge \max\{m, l-m+1\}$ since each self-avoiding walk of length $m-1$ or $l-m$ has distinct vertices. Moreover, since $G(w)$ is connected, $h \ge v-1$, so we have $v-1 \le h \le k(l+m-2)$. And the range of v is then given by $\max\{m, l-m+1\} \le v \le k(l+m-2)+1$.

The expected weight that a circuit contributes can be estimated similarly as before. From (7.14), the expected weights from $v-1$ many hyperedges that corresponds to edges on $T(w)$ is bounded by $\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1}$. Similar to (7.10), the expected weights from $h-v+1+2k$ many hyperedges that corresponds to edges on $G(w) \setminus T(w)$ together with hyperedges whose weights are from $\overline{A_{ij}^e}$ is bounded by $\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1+2k}$.

Putting all estimates together, for fixed v, h, the total contribution to the sum is bounded by

$$
\begin{aligned}
& n^v \binom{n}{d-2}^{h+2k} v^{2k}[(v+1)^2(l+1)]^{4k(2+h-v)} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1} \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1+2k} \\
=& n^v \left(\frac{\alpha}{d-1}\right)^{v-1} \left(\frac{d-1}{n-d+2}\right)^{h+2k} v^{2k}\, Q(k,l,v,h),
\end{aligned}
$$

where $Q(k,l,v,h) := [(v+1)^2(l+1)]^{4k(2+h-v)}(a\vee b)^{h-v+1+2k}$.

Let S_1 be the contribution of circuits in Case (1) to the sum in (A9). We have

$$
S_1 \le \sum_{v=m\vee(l-m+1)}^{k(l+m-2)+1} \sum_{h=v-1}^{k(l+m-2)} n^v \left(\frac{\alpha}{d-1}\right)^{v-1} \left(\frac{d-1}{n-d+2}\right)^{h+2k} v^{2k}\, Q(k,l,v,h). \tag{A10}
$$

Taking $l = O(\log n)$, similar to the discussion in (7.16), the leading term in (A10) is given by the term with $h = v - 1$. So for any $1 \le m \le l$, and sufficiently large n, there are constants $C_1, C_2 > 0$ such that

$$
\begin{aligned}
S_1 \le & \sum_{v=m\vee(l-m+1)}^{k(l+m-2)+1} 2n^{1-2k}((d-1)v)^{2k}[(v+1)^2(l+1)]^{4k}\alpha^{v-1}(a\vee b)^{2k} \\
\le & C_2 \log^{14k}(n) \cdot n^{1-2k}\alpha^{k(l+m-2)}.
\end{aligned}
$$

For circuits not in Case (1), similar to the proof of Lemma 7.1, their total contribution is bounded by $C_2' n^{1-2k}\alpha^{k(l+m-2)}\log^{14k} n$ for a constant $C_2' > 0$. This completes the proof of Lemma 7.2. ■

A.7 | Proof of Lemma 9.1

Proof. Let $\mathcal{B}$ be the set of vertices such that their l-neighborhood contains a cycle. Let x be a normed vector such that $x^\top B^{(l)}\mathbf{1} = 0$. We then have

$$
\begin{aligned}
\mathbf{1}^\top B^{(m-1)}x =& \sum_{i\in[n]} x_i (B^{(m-1)}\mathbf{1})_i = \sum_{i\notin\mathcal{B}} x_i S_{m-1}(i) + \sum_{i\in\mathcal{B}} x_i (B^{m-1}\mathbf{1})_i \\
=& \sum_{i\in[n]} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n)) \\
& - \sum_{i\in\mathcal{B}} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n)) + \sum_{i\in\mathcal{B}} x_i (B^{(m-1)}\mathbf{1})_i.
\end{aligned} \tag{A10}
$$

Since we have $\mathbf{1}^\top B^{(l)}x = 0$, the first term in (A10) satisfies

$$
\left|\sum_{i\in[n]} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n))\right| = \left|\sum_{i\in[n]} x_i O(\alpha^{\frac{m-1}{2}}\log n)\right| = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n),
$$

where the last inequality above is from Cauchy inequality.

From Lemma 4.4, $|\mathcal{B}| = O(\alpha^{2l}\log^4 n)$. For the second term in (A10), recall from (4.7), for $m \le l$, $|(B^{(m)}\mathbf{1})_i| = O(\alpha^m \log n)$, then by Cauchy inequality

$$
\left|\sum_{i\in\mathcal{B}} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n))\right| \le \sqrt{|\mathcal{B}|}O(\alpha^{m-1}\log n) = O(\alpha^{l+m-1}\log^3 n).
$$

Similarly, the third term satisfies

$$
\left|\sum_{i\in\mathcal{B}} x_i (B^{(m-1)}\mathbf{1})_i\right| = O(\alpha^{l+m-1}\log^3 n).
$$

Note that $\alpha^{l+m-1} = o(n^{1/2})$, altogether we have

$$
|\mathbf{1}^\top B^{(m-1)}x| = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n + \alpha^{l+m-1}\log^3 n) = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n). \tag{A11}
$$

(9.1) then follows. Using the property $x^\top B^{(l)}\sigma = 0$ instead of $x^\top B^{(l)}\mathbf{1} = 0$ and following the same argument, (9.2) holds. ■

A.8 | Proof of Lemma 10.1

Proof. Conditioned on $(H, i, \sigma)_{t-1} \equiv (T, \rho, \tau)_{t-1}$, if A_t holds, it implies that hyperedges generated from vertices in V_{t-1} do not overlap (except for the parent vertices in V_{t-1}). If B_t holds, vertices in V_t that are in different hyperedges generated from H_{t-1} do not connect to each other. If both A_t B_t holds, $(H, i, \sigma)_t$ is still a hypertree. Since $X_v^{(r)} = Y_{\phi(v)}^{(r)}$ for $v \in V_{t-1}$, we can extend the hypergraph isomorphism ϕ by mapping the children of $v \in V_t$ to the corresponding vertices in the tth generation of children of ρ in T, which keeps the hypertree structure and the spin of each vertex. ■

A.9 | Proof of Lemma 10.2

Proof. First we fix $u, v \in V_t$. For any $w \in V_{>t}$, the probability that $(u, w), (v, w)$ are both connected is $O(n^{-2})$. We know $|V_{>t}| \leq n$ and $|V_{\leq t}| = O(\log^2(n)\alpha^t)$ conditioned on C_t. Since $\alpha^{2t} \leq \alpha^{2l} = o(n^{1/2})$, taking a union bound over all u, v, w we have

$$\mathbb{P}(A_t | C_t) \geq 1 - O(\log^4(n)\alpha^{2t}n^{-1}) = 1 - o(n^{-1/2}). \tag{A12}$$

For the second claim, the probability of having an edge between $u, v \in V_t$ is $O(n^{-1})$. Taking a union bound over all pairs of $u, v \in V_t$ implies

$$\mathbb{P}(B_t | C_t) \geq 1 - O(\log^4(n)\alpha^{2t}n^{-1}) = 1 - o(n^{-1/2}). \tag{A13}$$

■

A.10 | Proof of Lemma 11.1

Proof. In (11.1), the coordinates of two vectors on the left hand side agree at i if the l-neighborhood of l contains no cycle. Recall $\mathcal{B}$ is the set of vertices whose l-neighborhood contains a cycle, from Lemma 4.4, and (4.7), we have asymptotically almost surely,

$$\|B^{(l)}\mathbf{1} - \vec{S}_l\|_2 \leq \sqrt{|\mathcal{B}|}O(\log(n)\alpha^l) = O(\log^3(n)\alpha^{2l}) = o(\sqrt{n}). \tag{A14}$$

From (5.6) we have

$$\|\vec{D}_l\|_2 = \Theta(\sqrt{n}\beta^l) \tag{A15}$$

asymptotically almost surely, and $\|B^{(l)}\mathbf{1}\|_2 \geq \|\vec{D}_l\|_2$, therefore (11.1) follows.

Similar to (A14), we have

$$\|B^{(l)}\sigma - \vec{D}_l\|_2 = o(\sqrt{n}), \quad \|B^{(l)}\sigma\|_2 = \|\vec{D}_l\|_2 + o(\sqrt{n}) = \Theta(\sqrt{n}\beta^l). \tag{A16}$$

Then (11.2) follows.

It remains to show (11.3). Using the same argument as in Theorem 5.4, we have the following convergence in probability

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} \alpha^{-2l} S_l^2(i) = \mathbb{E}M_\infty^2, \tag{A17}$$

where M_∞ is the limit of the martingale M_t. Similarly, the following convergences in probability hold

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} \alpha^{-l}\beta^{-l}S_l(i)D_l(i) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{N}^+} \alpha^{-l}\beta^{-l}S_l(i)D_l(i) + \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{N}^-} \alpha^{-l}\beta^{-l}S_l(i)D_l(i) \\ &= \frac{1}{2}\mathbb{E}M_\infty D_\infty - \frac{1}{2}\mathbb{E}M_\infty D_\infty = 0.\end{aligned}$$

Thus $\langle \vec{S}_l, \vec{D}_l \rangle = o(n\alpha^l \beta^l)$ asymptotically almost surely.

From (A17) we have

$$\|\vec{S}_l\|_2 = \Theta(\sqrt{n}\alpha^l), \tag{A18}$$

therefore together with (A15), we have $\|\vec{S}_l\|_2 \cdot \|\vec{D}_l\|_2 = \Theta(n\alpha^l\beta^l)$. With (11.1) and (11.2), (11.3) holds. ∎

A.11 | Proof of Lemma 11.2

Proof. For the lower bound in (11.4), note that $B^{(l)}$ is symmetric, we have

$$\|B^{(l)}\mathbf{1}\|_2^2 = \langle B^{(l)}\mathbf{1}, B^{(l)}\mathbf{1}\rangle = \langle \mathbf{1}, B^{(l)}B^{(l)}\mathbf{1}\rangle \le \|\mathbf{1}\|_2 \|B^{(l)}B^{(l)}\mathbf{1}\|_2. \tag{A19}$$

Therefore from (A18) and (11.1),

$$\|B^{(l)}B^{(l)}\mathbf{1}\|_2 \ge \frac{\|B^{(l)}\mathbf{1}\|_2^2}{\|\mathbf{1}\|_2} = \Theta(\alpha^l)\|B^{(l)}\mathbf{1}\|_2. \tag{A20}$$

For the upper bound in (11.4), from (4.1) and (4.7), the maximum row sum of $B^{(l)}$ is $O(\alpha^l \log n)$. Since $B^{(l)}$ is nonnegative, the spectral norm $\rho(B^{(l)})$ is bounded by the maximal row sum, hence (11.4) holds.

The lower bound in (11.5) can be proved similarly as in (11.4), from the inequality $\|B^{(l)}\sigma\|_2^2 \le \|\sigma\|_2\|B^{(l)}B^{(l)}\sigma\|_2$ together with (A15) and (11.2).

Recall $\mathcal{B}$ is the set of vertices whose l-neighborhood contains cycles. Let $\overline{\mathcal{B}} = [n]\backslash\mathcal{B}$. Since

$$\left(B^{(l)}B^{(l)}\sigma\right)_i = \sum_{j\in[n]} B_{ij}^{(l)}(B^{(l)}\sigma)_j,$$

we can decompose the vector $B^{(l)}B^{(l)}\sigma$ as a sum of three vectors $z + z' + z''$, where

$$\begin{aligned}
z_i &:= \mathbf{1}_{\overline{\mathcal{B}}}(i) \sum_{j: d(i,j)=l} D_l(j)\mathbf{1}_{\overline{\mathcal{B}}}(j), \quad z_i' := \mathbf{1}_{\overline{\mathcal{B}}}(i)\sum_{j: d(i,j)=l} O(\alpha^l \log n)\mathbf{1}_{\mathcal{B}}(j),\\
z_i'' &:= \mathbf{1}_{\mathcal{B}}(i) O(\alpha^{2l}\log^2 n).
\end{aligned}$$

The decomposition above depends on whether $i, j \in \mathcal{B}$ and the estimation follows from (4.7). From Lemma 4.4, $\mathcal{B} = O(\alpha^{2l}\log^4(n))$ asymptotically almost surely, so one has

$$\begin{aligned}
\|z'\|_2^2 = \sum_{i=1}^n (z_i')^2 &= \sum_{i\in\overline{\mathcal{B}}}\sum_{j: d(i,j)=l}\sum_{j': d(i,j')=l} O(\alpha^{2l}\log^2 n)\mathbf{1}_{\mathcal{B}}(j)\mathbf{1}_{\mathcal{B}}(j')\\
&= \sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}}\sum_{\substack{i\in\overline{\mathcal{B}}\\ d(i,j)=d(i,j')=l}} O(\alpha^{2l}\log^2 n) = \sum_{j,j'\in\mathcal{B}} O(\alpha^{3l}\log^3 n) = O(\alpha^{7l}\log^{11} n),
\end{aligned}$$

which implies $\|z'\|_2 = O(\alpha^{7l/2}\log^{11/2} n)$. And similarly $\|z''\|_2 = O(\alpha^{3l}\log^2 n)$.

We know from (A16), $\|B^{(l)}\sigma\|_2 = \Theta(\beta^l\sqrt{n})$, and since $c\log\alpha < 1/8$, we have $\alpha^{5l/2} = n^{-\gamma'}\sqrt{n}$ for some $\gamma' > 0$, therefore

$$\|z' + z''\|_2 = O(\alpha^{7l/2}\log^{11/2} n) = o(\alpha^{5l/2}\beta^{2l}) = O(n^{-\gamma'}\beta^l\|B^{(l)}\sigma\|_2). \tag{A21}$$

It remains to upper bound $\|z\|_2$. Assume the $2l$-neighborhood of i is cycle-free, then the ith entry of $B^{(l)}B^{(l)}\sigma$, denoted by X_i, can be written as

$$X_i := (B^{(l)}B^{(l)}\sigma)_i = \sum_{k=1}^{n} B_{ik}^{(l)}(B^{(l)}\sigma)_k = \sum_{k=1}^{n} \mathbf{1}_{d(i,k)=l} \sum_{j=1}^{n} \mathbf{1}_{d(j,k)=l}\sigma_j$$
$$= \sum_{h=0}^{l} \sum_{j:d(i,j)=2h} \sigma_j |\{k : d(i,k) = d(j,k) = l\}|. \tag{A22}$$

We control the magnitude of X_i in the corresponding hypertree growth process. Since $2l = 2c\log n$ and $2c\log(\alpha) < 1/4$, the coupling result in Theorem 5.2 can apply.

Let C_i be the event that coupling between $2l$-neighborhood of i with the Poisson Galton–Watson hypertree has succeeded and $n^{-\epsilon}$ be the failure probability of the coupling. When the coupling succeeds, $z_i = X_i$, therefore

$$\mathbb{E}(\|z\|_2^2|\Omega) = \sum_{i\in[n]} n^{-\epsilon} O(\alpha^{2l}\beta^{2l}\log^2 n) + \sum_{i\in[n]} \mathbb{E}(X_i^2 \mathbf{1}_{C_i}|\Omega)$$
$$= n^{1-\epsilon} O(\alpha^{2l}\beta^{2l}\log^2 n) + \sum_{i\in[n]} \mathbb{E}(X_i^2 \mathbf{1}_{C_i}|\Omega). \tag{A23}$$

For any $i, j \in [n], t \in [l]$, define $D_{i,j}^{(t)} := |\{k : d(i,k) = d(j,k) = t\}|$. From (A22), we have

$$X_i^2 = \sum_{h,h'=0}^{l} \sum_{j:d(i,j)=2h} \sum_{j':d(i,j')=2h'} \sigma_j\sigma_{j'} D_{i,j}^{(l)} D_{i,j'}^{(l)}. \tag{A24}$$

We further classify the pair j, j' in (A24) according to their distance. Let $d(j,j') = 2(h+h'-\tau)$ for $\tau = 0, \dots, 2(h \wedge h')$. This yields

$$X_i^2 = \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} \sum_{j:d(i,j)=2h} \sum_{j':d(i,j')=2h'} \mathbf{1}_{d(j,j')=2(h+h'-\tau)} \sigma_j\sigma_j' D_{i,j}^{(l)} D_{i,j'}^{(l)}.$$

Conditioned on Ω and C_i, similar to the analysis in Appendix H in [32], we have the following holds

$$|\{k : d(i,k) = d(j,k) = l\}| = O(\alpha^{l-h}\log n), \tag{A25}$$
$$|\{k\prime : d(i,k\prime) = d(j\prime,k\prime) = l\}| = O(\alpha^{l-h\prime}\log n), \tag{A26}$$
$$|\{j : d(i,j) = 2h\}| = O(\alpha^{2h}\log n), \tag{A27}$$
$$|\{j\prime : d(i,j\prime) = 2h\prime, d(j,j\prime) = 2(h+h\prime-\tau)\}| = O(\alpha^{2h\prime-\tau}\log n). \tag{A28}$$

We claim that

$$\mathbb{E}[\sigma_j\sigma_{j\prime}|C_i] \le \left(\frac{\beta}{\alpha}\right)^{d(j,j\prime)-1}, \tag{A29}$$

and prove (A29) in Cases (a)–(d).

(a) Assume j is the parent of j' in the hypertree growth process. Then $d(j,j') = 1$. Let $\mathcal{T}_r$ be the event that the hyperedge containing j' is of type r. Given $\mathcal{T}_r$, by our construction of the hypertree process, the spin of j' is assigned to be σ_j with probability $\frac{r}{d-1}$ and $-\sigma_j$ with probability $\frac{d-1-r}{d-1}$, so we have

$$\mathbb{E}[\sigma_j\sigma_{j'}|C_i] = \sum_{r=0}^{d-1} \mathbb{E}[\sigma_j\sigma_j'|\mathcal{T}_r, C_i]\mathbb{P}[\mathcal{T}_r|C_i] = \sum_{r=0}^{d-1} \left(\frac{r}{d-1} - \frac{d-1-r}{d-1}\right)\mathbb{P}[\mathcal{T}_r|C_i].$$

Recall $\mathbb{P}[\mathcal{T}_{d-1}|C_i] = \frac{(d-1)a}{\alpha 2^{d-1}}$ and $\mathbb{P}[\mathcal{T}_r|C_i] = \frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}}$ for $0 \le r \le d-2$. A simple calculation implies $\mathbb{E}[\sigma_j\sigma_{j'}|C_i] = \frac{\beta}{\alpha} \le 1$.

(b) Suppose $d(j,j')=t$ and there is a sequence of vertices $j, j_1, \dots, j_{t-1}, j'$ such that j_1 is a child of j, j_i is a child of j_{i-1} for $1 \le i \le t$, and j' is a child of j_{t-1}. We show by induction that for $t \ge 1$,

$$\mathbb{E}[\sigma_j\sigma_{j'}|C_i] = \left(\frac{\beta}{\alpha}\right)^t.$$

When $t=1$ this has been approved in part (a). Assume it is true for all j, j' with distance $\le t-1$. Then when $d(j,j')=t$, we have

$$\begin{aligned}\mathbb{E}[\sigma_j\sigma_{j'}|C_i] &= \mathbb{E}[\sigma_j\sigma_{j'}|\sigma_{j_1}=\sigma_j, C_i]\mathbb{P}(\sigma_{j_1}=\sigma_j|C_i) + \mathbb{E}[\sigma_j\sigma_{j'}|\sigma_{j_1}=-\sigma_j, C_i]\mathbb{P}(\sigma_{j_1}=-\sigma_j|C_i)\\ &= \left(\frac{\beta}{\alpha}\right)^{t-1}\mathbb{P}(\sigma_{j_1}=\sigma_j|C_i) - \left(\frac{\beta}{\alpha}\right)^{t-1}\mathbb{P}(\sigma_{j_1}=-\sigma_j|C_i)\\ &= \left(\frac{\beta}{\alpha}\right)^{t-1}\frac{\alpha+\beta}{2\alpha} - \left(\frac{\beta}{\alpha}\right)^{t-1}\frac{\alpha-\beta}{2\alpha} = \left(\frac{\beta}{\alpha}\right)^t.\end{aligned}$$

Therefore $\mathbb{E}[\sigma_j\sigma_{j\prime}|C_i] \le \left(\frac{\beta}{\alpha}\right)^{d(j,j\prime)} \le \left(\frac{\beta}{\alpha}\right)^{d(j,j\prime)-1}$. This completes the proof for part (b).

(c) Suppose j, j' are not in the same hyperedge and there exists a vertex k such that j, k satisfies the assumption in Case (b) with $d(j,k)=t_1$, and j', k satisfy the assumption in Case (b) with $d(j',k)=t_2$. Conditioned on σ_k, we know σ_j and σ_j' are independent. Then we have

$$\begin{aligned}\mathbb{E}[\sigma_j\sigma_{j'}|C_i] &= \mathbb{E}[\mathbb{E}[\sigma_j\sigma_{j'}\sigma_k^2|\sigma_k, C_i]|C_i] = \mathbb{E}\left[\mathbb{E}[\sigma_j\sigma_k|\sigma_k, C_i]\cdot\mathbb{E}[\sigma_{j'}\sigma_k|\sigma_k, C_i]|C_i\right]\\ &= \left(\frac{\beta}{\alpha}\right)^{t_1+t_2} \le \left(\frac{\beta}{\alpha}\right)^{d(j,j')-1},\end{aligned}$$

where the last line follows from the triangle inequality $d(j,k)+d(j',k) \ge d(j,j')$ and the condition $\beta < \alpha$.

(d) If j, j' are in the same hyperedge, then $d(j,j')=1$ and (A29) holds trivially.

Combining Cases (a)–(d), (A29) holds. From (A29) and (A25)–(A28), we have

$$\begin{aligned}\mathbb{E}[X_i^2\mathbf{1}_\Omega|C_i] &\le \sum_{h,h'=0}^{l}\sum_{\tau=0}^{2(h\wedge h')}\sum_{j:d(i,j)=2h}\sum_{j':d(i,j')=2h'}\mathbf{1}_{d(j,j')=2(h+h'-\tau)}\mathbb{E}[\sigma_j\sigma_j'|C_i]R_{i,j}^{(l)}R_{i,j'}^{(l)}\\ &\le \sum_{h,h'=0}^{l}\sum_{\tau=0}^{2(h\wedge h')}\sum_{j:d(i,j)=2h}O(\alpha^{2h'-\tau}\log n)\left(\frac{\beta}{\alpha}\right)^{2(h+h'-\tau)-1}\cdot O(\alpha^{2l-h-h'}\log^2 n)\\ &= \sum_{h,h'=0}^{l}\sum_{\tau=0}^{2(h\wedge h')}O(\alpha^{2l+h+h'-\tau}\log^4 n)\left(\frac{\beta}{\alpha}\right)^{2(h+h'-\tau)-1}\\ &= \sum_{h,h'=0}^{l}\sum_{\tau=0}^{2(h\wedge h')}O(\alpha^{2l}\log^4 n)\cdot(\beta^2/\alpha)^{h+h'-\tau} = O(\beta^{4l}\log^4 n).\end{aligned}\tag{A30}$$

From (A23) and (A30), we have for some $\epsilon > 0$,

$$\mathbb{E}(\|z\|_2^2|\Omega) = n^{1-\epsilon}O(\alpha^{2l}\beta^{2l}\log^2 n) + O(n\beta^{4l}\log^2 n).$$

Then by Chebyshev's inequality, asymptotically almost surely,

$$\|z\|_2 = O(n^{1/2-\epsilon/2}\alpha^l\beta^l\log^2 n) + O(n^{1/2}\beta^{2l}\log^2 n) = (\sqrt{n}\beta^l\log^2 n)\cdot O(\beta^l \vee \alpha^l n^{-\epsilon/2}).$$

Recall $l = c\log n$. We have $\beta^l = n^{c\log\beta}, \alpha^l = n^{c\log\alpha}$. So $\beta^l = n^{-\epsilon'}\alpha^l$ for some constant $\epsilon' > 0$. Since from (A16), $\|B^{(l)}\sigma\|_2 = \Theta(\sqrt{n}\beta^l)$, we have

$$\|z\|_2 = O(n^{-\gamma''}\alpha^l\|B^{(l)}\sigma\|_2) \tag{A31}$$

for some constant $\gamma'' > 0$. Combining (A21) with (A31), it implies for some constant $\gamma > 0$,

$$\|B^{(l)}B^{(l)}\sigma\|_2 = \|z + z' + z''\|_2 = O(n^{-\gamma}\alpha^l)\|B^{(l)}\sigma\|_2.$$

Then the upper bound on $\|B^{(l)}B^{(l)}\sigma\|_2$ in (11.5) holds. ■