Community detection in the hypergraph stochastic block model and reconstruction on hypertrees

Yuzhou Gu   YUZHOUGU@IAS.EDU
Institute for Advanced Study

Aaradhya Pandey   AP9898@PRINCETON.EDU
Princeton University

Editors: Shipra Agrawal and Aaron Roth

Abstract

We study the weak recovery problem on the r-uniform hypergraph stochastic block model (r-HSBM) with two balanced communities. In this model, n vertices are randomly divided into two communities, and size-r hyperedges are added randomly depending on whether all vertices in the hyperedge are in the same community. The goal of weak recovery is to recover a non-trivial fraction of the communities given the hypergraph. Pal and Zhu (2021); Stephan and Zhu (2022) established that weak recovery is always possible above a natural threshold called the Kesten-Stigum (KS) threshold. For assortative models (i.e., monochromatic hyperedges are preferred), Gu and Polyanskiy (2023b) proved that the KS threshold is tight if $r \leq 4$ or the expected degree d is small. For other cases, the tightness of the KS threshold remained open.

In this paper we determine the tightness of the KS threshold for a wide range of parameters. We prove that for $r \leq 6$ and d large enough, the KS threshold is tight. This shows that there is no information-computation gap in this regime and partially confirms a conjecture of Angelini et al. (2015). On the other hand, we show that for $r \geq 5$, there exist parameters for which the KS threshold is not tight. In particular, for $r \geq 7$, the KS threshold is not tight if the model is disassortative (i.e., polychromatic hyperedges are preferred) or d is large enough. This provides more evidence supporting the existence of an information-computation gap in these cases.

Furthermore, we establish asymptotic bounds on the weak recovery threshold for fixed r and large d. We also obtain a number of results regarding the broadcasting on hypertrees (BOHT) model, including the asymptotics of the reconstruction threshold for $r \geq 7$ and impossibility of robust reconstruction at criticality.

Keywords: hypergraph stochastic block model, weak recovery, reconstruction on hypertrees, information-computation gap

1. Introduction

1.1. Hypergraph stochastic block model

The stochastic block model (SBM) is a random graph model with community structures. It is the canonical model for studying community detection, and has applications in statistics, machine learning, and network science. The theoretical study of community detection in the SBM has received a lot of attention in the past decade since Decelle et al. (2011); see Abbe (2017) for a survey.

The hypergraph stochastic block model (HSBM) is a generalization of the SBM to hypergraphs. In this paper, we study the r-uniform HSBM with two symmetric communities, defined as follows. Let $n \in \mathbb{Z}_{\geq 0}$ denote the number of vertices and $V = [n]$ be the set of vertices[1]. Let $r \in \mathbb{Z}_{\geq 2}$ be

---

1. We use the notation $[n] = \{1, \ldots, n\}$.

the hyperedge size and $a, b \in \mathbb{R}_{\geq 0}$ be two parameters[2]. Generate a random label $X_u \in \{\pm\}$ for all vertices $u \in V$ i.i.d. $\sim \mathrm{Unif}(\{\pm\})$. For every subset $S \subseteq V$ of size r, the hyperedge S is added with probability $\frac{a}{\binom{n}{r-1}}$ if all vertices in S have the same label; otherwise add the hyperedge with probability $\frac{b}{\binom{n}{r-1}}$. The model outputs (X, G), where G is the resulting hypergraph. We denote the model as $\mathrm{HSBM}(n, r, a, b)$.

We define two useful derived parameters d and λ.

$$d = \frac{a - b + 2^{r-1}b}{2^{r-1}}, \qquad \lambda = \frac{a - b}{a - b + 2^{r-1}b}. \tag{1}$$

The expected degree of every vertex is $d \pm o(1)$, and λ is a strength parameter with range $\left[-\frac{1}{2^{r-1}-1}, 1 \right]$.

For the constant degree regime we consider, the most relevant community detection question is the weak recovery problem, which asks to recover a non-trivial fraction of the community labels (up to a global sign flip) given the hypergraph G. Formally, the goal is to design a (potentially inefficient) estimator $\widehat{X} = \widehat{X}(G)$ such that for a constant $c < \frac{1}{2}$,

$$\lim_{n \to \infty} \mathbb{P}\left[d_H(\widehat{X}, X) \leq (c + o(1))n \right] = 1, \tag{2}$$

$$\text{where} \qquad d_H(X, Y) = \min_{s \in \{\pm\}} \sum_{u \in V} \mathbb{1}\{X_u \neq s Y_u\}. \tag{3}$$

For $r = 2$, the model $\mathrm{HSBM}(n, 2, a, b)$ becomes SBM with two symmetric communities, where the weak recovery threshold was established by Massoulié (2014); Mossel et al. (2015, 2018): weak recovery is possible if and only if $d\lambda^2 > 1$. The threshold $d\lambda^2 = 1$ is called the Kesten-Stigum (KS) threshold and originates from the study of multitype Galton-Watson processes (Kesten and Stigum (1966)). For the HSBM, the natural generalization of the KS threshold is $(r - 1)d\lambda^2 = 1$. Angelini et al. (2015) conjectured that a phase transition for weak recovery occurs at the KS threshold. Pal and Zhu (2021); Stephan and Zhu (2022) proved the positive part of the conjecture, that is, when $(r-1)d\lambda^2 > 1$, weak recovery is possible and can be achieved using an efficient (polynomial time) algorithm. On the negative part, Gu and Polyanskiy (2023b) proved that for assortative models (i.e., $a > b$), when $r = 3, 4$ or $\lambda \geq \frac{1}{5}$, if $(r - 1)d\lambda^2 \leq 1$, then weak recovery is impossible. They also provided evidence (via studying the reconstruction on hypertrees problem, see Section 1.2) suggesting that for $r \geq 7$ and d large enough, weak recovery might be information-theoretically possible below the KS threshold. Nevertheless, the tightness of the KS threshold for disassortative models (i.e., $a < b$) or assortative models with $r \geq 5$ and large d remained open before the current work.

Our first contribution is to show that the KS threshold is tight for $r \leq 6$ and d large enough.

Theorem 1 (KS is tight for $r \leq 6$ and large degree, HSBM) *Consider the model $\mathrm{HSBM}(n, r, a, b)$ where $r \leq 6$ and $a, b \in \mathbb{R}_{\geq 0}$. There exists $d_0 = d_0(r)$ such that for all $d \geq d_0$, if $(r - 1)d\lambda^2 \leq 1$, then weak recovery is impossible.*

---

2. While several previous works (e.g., Pal and Zhu (2021); Zhang et al. (2022); Gu and Polyanskiy (2023b)) assume that $a \geq b$, we do not make this assumption.

Our second contribution is to provide an upper bound for HSBM weak recovery. For $x, y \in [0, 1]$, define the binary Kullback-Leibler (KL) divergence function $d_{\mathrm{KL}}(\cdot \| \cdot)$ as[3]

$$d_{\mathrm{KL}}(x \| y) = x \log \frac{x}{y} + (1 - x) \log \frac{1 - x}{1 - y}. \tag{4}$$

Theorem 2 (HSBM weak recovery upper bound) *Consider the model* $\mathrm{HSBM}(n, r, a, b)$ *with* $a, b \in \mathbb{R}_{\geq 0}$. *If*

$$\frac{d}{r} \cdot d_{\mathrm{KL}} \left(\lambda + \frac{1 - \lambda}{2^{r-1}} \, \middle\| \, \frac{1}{2^{r-1}} \right) > \log 2, \tag{5}$$

then weak recovery is possible.

Comparing Eq. (5) with the KS threshold $(r - 1)d\lambda^2 = 1$, we obtain that the KS threshold is not tight for a wide range of parameters.

Corollary 3 (Non-tightness of the KS threshold for $r \geq 5$, HSBM) *Work in the setting of Theorem 2. If $r \geq 5$, then there exists $\lambda_0 = \lambda_0(r) \in \left(-\frac{1}{2^{r-1}-1}, 1 \right]$ such that for any non-zero $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \lambda_0 \right)$, there exists d satisfying $(r - 1)d\lambda^2 < 1$ such that weak recovery is possible. Furthermore, for $r \geq 7$ we can take $\lambda_0(r) > 0$.*

In Corollary 3, we can take $\lambda_0(r)$ to be the minimum solution in $\left[-\frac{1}{2^{r-1}-1}, 1 \right]$ to the equation

$$d_{\mathrm{KL}} \left(\lambda + \frac{1 - \lambda}{2^{r-1}} \, \middle\| \, \frac{1}{2^{r-1}} \right) = r(r - 1)\lambda^2 \log 2. \tag{6}$$

Table 1 presents numerical values of $\lambda_0(r)$ for small r.

r	5	6	7	8	9	10
λ_0	-0.06575	-0.02154	0.00413	0.01920	0.02807	0.03320

Table 1: Minimum solution to Eq. (6) for $5 \leq r \leq 10$, rounded down to the nearest multiple of 10^{-5}. For non-zero $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \lambda_0 \right)$, the KS threshold is not tight. See Corollary 3 for the complete statement.

Furthermore, we obtain the following asymptotic lower and upper bounds for the HSBM weak recovery threshold.

Theorem 4 (Asymptotic HSBM weak recovery bounds) *Consider the model* $\mathrm{HSBM}(n, r, a, b)$ *with $r \geq 2$ and $a, b \in \mathbb{R}_{\geq 0}$.*

(i) *(Upper bound) For any $\epsilon > 0$, there exists $d_0 = d_0(r, \epsilon)$, such that for all $d \geq d_0$, if $d\lambda^2 \geq \frac{2r \log 2}{2^{r-1}-1} + \epsilon$, then weak recovery is possible.*

---

3. Throughout the paper, log denotes natural logarithm.

(ii) (Lower bound) For any $\epsilon > 0$, there exists $d_0 = d_0(r, \epsilon)$, such that for all $d \geq d_0$, if $d\lambda^2 \leq \frac{1}{2^{r-2}} - \epsilon$, then weak recovery is impossible.

Our current knowledge about the tightness of the KS threshold for HSBM weak recovery is summarized in Table 2. Our contribution is the "large d" column (except for the $r = 3, 4, \lambda > 0$ case which was established in Gu and Polyanskiy (2023b)) and all non-tightness results.

	any d	small d	large d
$r = 3, 4, \lambda > 0$	always tight	(tight)	(tight)
$r = 3, 4, \lambda < 0$	unknown	unknown	tight
$r = 5, 6, \lambda > 0$	unknown	tight	tight
$r = 5, 6, \lambda < 0$	unknown	not tight	tight
$r \geq 7, \lambda > 0$	unknown	tight	not tight
$r \geq 7, \lambda < 0$	never tight	(not tight)	(not tight)

Table 2: Tightness of the KS threshold for HSBM weak recovery. An entry in parentheses means it is implied by another entry in the same row. References: Possibility above KS: Pal and Zhu (2021) ($\lambda > 0$), Stephan and Zhu (2022) (general HSBM). Tightness for $r = 3, 4$, $\lambda > 0$, and $r \geq 5$, $\lambda > 0$, small d: Gu and Polyanskiy (2023b). Tightness for $r \leq 6$, large d: Theorem 1. Non-tightness results: Corollary 3. This table is also our current knowledge about tightness of the KS threshold for BOHT reconstruction.

Proofs of Theorem 1 and Theorem 4(ii) are based on the reconstruction on hypertrees problem, which we will discuss in Section 1.2. Proof of Theorem 2 uses a simple estimator which returns a balanced partition with expected in-community hyperedge density. We show that when Eq. (5) is satisfied, any balanced partition that achieves expected in-community hyperedge density is correlated with the true partition. This method was previously used in Banks et al. (2016) in the case of q-community SBM. Corollary 3 and Theorem 4(i) follow from Theorem 2.

Another interesting community detection problem for the HSBM is the testing problem, which asks to distinguish an HSBM hypergraph from an Erdős-Rényi hypergraph with the same hyperedge density. Yuan et al. (2022) studied the testing problem for a class of symmetric HSBMs (which includes our two-community case), proving that testing (with $o(1)$ error probability) is possible above the KS threshold. They showed that the KS threshold is tight for the two-community three-uniform HSBM, and that the KS threshold is not tight in general. In fact, they proved that when Eq. (5) holds, testing is possible.[4] Their results do not imply our Theorem 2 because a direct relationship between testing and weak recovery is not known. Nevertheless, this provides evidence suggesting that there might be a close relationship between the two problems.

1.2. Reconstruction on hypertrees

Theorem 1 and Theorem 4(ii) are proved by studying the reconstruction problem for the broadcasting on hypertrees (BOHT) model, defined as follows. Let $r \in \mathbb{Z}_{\geq 2}$ be the hyperedge size,

---

4. We remark that (Yuan et al., 2022, Theorem 2.6, Part 1, second half) is off by a factor of two because of a computation error. In particular, in the second-to-last display in the proof of Theorem 2.6, the third step performs Taylor expansion to order ϵ. However, because the final result is of order ϵ^2, to get the correct coefficient in the final result, one needs to perform Taylor expansion to order ϵ^2.

$\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$ be a strength parameter, D be a distribution on $\mathbb{Z}_{\geq 0}$, which is usually taken to be the point distribution at $d \in \mathbb{Z}_{\geq 0}$ or the Poisson distribution $\mathrm{Pois}(d)$ with expectation $d \in \mathbb{R}_{\geq 0}$. Let T be the r-uniform Galton-Watson linear[5] hypertree, where every vertex independently has $b \sim D$ downward hyperedges. The BOHT model assigns a $\pm$ label to every vertex according to the following process. The root ρ has label σ_ρ generated from the uniform distribution $\pi = \mathrm{Unif}(\{\pm\})$. Let $B_{r,\lambda} : \{\pm\} \to \{\pm\}^{r-1}$ be the channel

$$B_{r,\lambda}(y_1, \ldots, y_{r-1}|x) = \begin{cases} \lambda + \frac{1-\lambda}{2^{r-1}}, & \text{if } y_1 = \cdots = y_{r-1} = x, \\ \frac{1-\lambda}{2^{r-1}}, & \text{otherwise.} \end{cases} \tag{7}$$

Suppose we have generated label σ_u of a vertex u. Then for every downward hyperedge $\{u, v_1, \ldots, v_{r-1}\}$, children labels $\sigma_{v_1}, \ldots, \sigma_{v_{r-1}}$ are generated according to $B_{r,\lambda}$, that is, $(\sigma_{v_1}, \ldots, \sigma_{v_{r-1}}) \sim B_{r,\lambda}(\cdot|\sigma_u)$. In this way every vertex has received a label. The model outputs (T, σ) and we denote the model as $\mathrm{BOHT}(r, \lambda, D)$. When D is a point distribution at $d \in \mathbb{Z}_{\geq 0}$ we denote the model as $\mathrm{BOHT}(r, \lambda, d)$.

The reconstruction problem is an important problem for the BOHT model. Let $(T, \sigma) \sim \mathrm{BOHT}(r, \lambda, D)$. For $k \in \mathbb{Z}_{\geq 0}$, let T_k denote the tree structure up to the k-th level (the root being at the 0-th level), and L_k denote the set of vertices at the k-th level. We say reconstruction is possible if

$$\lim_{k \to \infty} I(\sigma_\rho; T_k, \sigma_{L_k}) \tag{8}$$

is positive, and impossible otherwise. Here $I(\cdot; \cdot)$ denotes mutual information. In other words, reconstruction is possible if and only if we can reconstruct the root label σ_ρ with $\frac{1}{2} - \Omega(1)$ error probability given the leaf labels σ_{L_k}, as $k \to \infty$.

The reconstruction problem is useful for the study of HSBM weak recovery. Consider the model $\mathrm{HSBM}(n, r, a, b)$. Let d, λ be defined as in Eq. (1). Consider the model $\mathrm{BOHT}(r, \lambda, \mathrm{Pois}(d))$, which we call the BOHT model corresponding to $\mathrm{HSBM}(n, r, a, b)$. If reconstruction is impossible for $\mathrm{BOHT}(r, \lambda, \mathrm{Pois}(d))$, then weak recovery is impossible for $\mathrm{HSBM}(n, r, a, b)$. This relationship was first established by Mossel et al. (2015, 2018) for the $r = 2$ case, and their proof can be easily generalized. See (Gu, 2023, Theorem 5.15) for a proof for general HSBMs. We note that the converse is not true in general: for the q-community SBM, there exist parameters where weak recovery is impossible for the SBM but reconstruction is possible for the corresponding broadcasting on trees (BOT) model.

In the $r = 2$ case, the model $\mathrm{BOHT}(2, \lambda, D)$ is called the Ising model, where the reconstruction threshold was established by Bleher et al. (1995); Evans et al. (2000): reconstruction is possible if and only if $d\lambda^2 > 1$. The threshold $d\lambda^2 = 1$ is also called the Kesten-Stigum (KS) threshold. For $r \geq 3$, the natural generalization of the KS threshold is $(r-1)d\lambda^2 = 1$. In general, it is known that reconstruction is possible whenever $(r-1)d\lambda^2 > 1$, and the tightness of the KS threshold is an interesting question. Gu and Polyanskiy (2023b) proved that reconstruction for $\mathrm{BOHT}(r, \lambda, D)$ is impossible for $(r-1)d\lambda^2 \leq 1$ if $r \leq 4$, $\lambda > 0$, or $\lambda \geq \frac{1}{5}$. Furthermore, they proved that the KS threshold is not tight for $r \geq 7$ and d large enough. The remaining cases were left open.

Our main contribution on the BOHT model is tightness of the KS threshold for $r \leq 6$ and large d.

---

5. Linear means the intersection of two different hyperedges has size at most one.

Theorem 5 (KS is tight for $r \leq 6$ and large degree, BOHT) *Consider the model* $\mathrm{BOHT}(r, \lambda, d)$ *or* $\mathrm{BOHT}(r, \lambda, \mathrm{Pois}(d))$ *where* $r \leq 6$ *and* $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$. *There exists* $d_0 = d_0(r)$ *such that for all* $d \geq d_0$, *if* $(r-1)d\lambda^2 \leq 1$, *then reconstruction is impossible.*

Not coincidentally, Table 2 is also our current knowledge about tightness of the KS threshold for the reconstruction problem on $\mathrm{BOHT}(r, \lambda, D)$ (for $D = d$ or $D = \mathrm{Pois}(d)$). Possibility of reconstruction above KS is folklore by a counting estimator. Tightness for $r = 3, 4$, $\lambda > 0$, and $r \geq 5$, $\lambda > 0$, small d, and non-tightness for $r \geq 7$, large d are by Gu and Polyanskiy (2023b). Tightness for $r \leq 6$, large d (except for $r = 3, 4$, $\lambda > 0$) is by Theorem 5. Other non-tightness results are by Corollary 3.

Theorem 1 follows directly from Theorem 5 using the relationship between HSBM weak recovery and BOHT reconstruction ((Gu, 2023, Theorem 5.15)).

Let us now discuss the proof of Theorem 5. For $k \in \mathbb{Z}_{\geq 0}$, let M_k denote the channel $\sigma_\rho \mapsto (T_k, \sigma_{L_k})$. Then $(M_k)_{k \geq 0}$ satisfies the belief propagation (BP) recursion

$$M_{k+1} = \mathrm{BP}(M_k) \tag{9}$$

where BP is the belief propagation operator defined as

$$\mathrm{BP}(P) = \mathbb{E}_{b \sim D} \left(P^{\times(r-1)} \circ B_{r,\lambda} \right)^{\star b}. \tag{10}$$

(See Section 2.1 for the relevant information theory background.) In other words, $M_k = \mathrm{BP}^k(\mathrm{Id})$, where $\mathrm{Id} : \{\pm\} \to \{\pm\}$ denotes the identity channel. Then non-reconstruction is equivalent to

$$\lim_{k \to \infty} C(\mathrm{BP}^k(\mathrm{Id})) = 0, \tag{11}$$

where C denotes channel capacity. Note that the channels M_k are binary memoryless symmetric (BMS) channels (Section 2.1) and the BP operator sends BMS channels to BMS channels.

Our proof is a hypertree version of Sly's method. This method was introduced in Sly (2009, 2011) for reconstruction on the Potts model. It has been applied to broadcasting with binary asymmetric channels (Liu and Ning (2019)) and is recently refined by Mossel et al. (2023) for the Potts channel. The method has two parts: large degree asymptotics and robust reconstruction.

In the first part (large degree asymptotics), we prove the following result.

Proposition 6 (Large degree asymptotics) *Fix $r \leq 6$. For any $\epsilon > 0$, there exists $d_0 = d_0(r, \epsilon)$ such that for any $d \geq d_0$ and $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$ with $(r-1)d\lambda^2 \leq 1$, we have*

$$\lim_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(\mathrm{Id})) < \epsilon, \tag{12}$$

where C_{χ^2} denotes χ^2-capacity (Section 2.1).

The idea is that $C_{\chi^2}(P^{\star d})$ for a BMS channel P is the expectation of a function involving d i.i.d. variables. When d is large, the result can be approximated using the expectation of a function involving a Gaussian variable. For the BOHT model, Gu and Polyanskiy (2023b) established the Gaussian approximation. The resulting function is a Gaussian integral (Eq. (22)), and we prove the numerical observation that it has no fixed points on $(0, 1]$.

The second part (robust reconstruction) is more difficult. We show that for any $r \geq 2$, there exists $\epsilon = \epsilon(r) > 0$ such that if $(r-1)d\lambda^2 \leq 1$, then for any BMS channel P with $C_{\chi^2}(P) \leq \epsilon$, we have

$$\lim_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(P)) = 0. \tag{13}$$

This is called robust reconstruction because we are considering reconstruction of the root label σ_ρ given observations of the leaf labels σ_{L_k} through a very noisy channel P, instead of the identity channel Id. Janson and Mossel (2004) studied the robust reconstruction problem on trees, and proved that the robust reconstruction threshold matches the KS threshold in most cases. However, their result does not apply to the hypertree model, and does not work for the critical case (i.e., at the KS threshold). The robust reconstruction result we prove is very precise: it works at criticality and the robustness parameter ϵ does not depend on d.

Let $x_k = C_{\chi^2}(\mathrm{BP}^k(\mathrm{Id}))$. We prove that if x_k is small enough, then

$$x_{k+1} = (r-1)d\lambda^2 x_k - \left((r-1)(r-2)d\lambda^3 + (r-1)^2 \left(\mathbb{E}_{b \sim D} b(b-1)\right) \lambda^4\right) x_k^2 + O_r(x_k^3). \tag{14}$$

This is the main expansion in the proof of robust reconstruction. The proof of Eq. (14) is by expanding $C_{\chi^2}(\mathrm{BP}(P))$ when $C_{\chi^2}(P)$ is close to 0.

There are several technical difficulties in the this approach. The first difficulty is that $C_{\chi^2}(\mathrm{BP}(P))$ is not a function of $C_{\chi^2}(P)$. Any BMS P is described by a random variable θ_P on $[-1, 1]$ (see Section 2.1). The χ^2-information $C_{\chi^2}(P)$ satisfies

$$C_{\chi^2}(P) = \mathbb{E}\theta_P = \mathbb{E}\theta_P^2. \tag{15}$$

Then $C_{\chi^2}(\mathrm{BP}(P))$ is the expectation of a function of $(r-1)d$ i.i.d. θ_P variables. In the expansion of $C_{\chi^2}(\mathrm{BP}(P))$, there will be higher order moments of θ_P which a priori may affect the first few orders in Eq. (14). Our solution is to also perform expansion on the third moment. We define

$$y_k = \mathbb{E}\theta_{M_k}^3 = \mathbb{E}\theta_{M_k}^4. \tag{16}$$

Because $\|\theta_P\| \leq 1$ almost surely for any P, we have $y_k \leq x_k$. We first perform a crude expansion of x_{k+1} and y_{k+1}, which roughly speaking allows us to conclude that $y_k = O(x_k^2)$. In this way we can control the influence of y_k and perform a finer expansion of x_{k+1}, which give us Eq. (14).

The second difficulty is that, unlike previous works which work with trees, we work with hypertrees. As we noted, $C_{\chi^2}(\mathrm{BP}(P))$ is the expectation of a function with $(r-1)d$ variables, while previous works have d variables. When performing the expansion, there will be a large number of terms, even after considering symmetry. Our solution is to perform the expansion in two steps. Let M'_{k+1} be the channel $M_k^{\times(r-1)} \circ B_{r,\lambda}$. We decompose the BP recursion $M_{k+1} = \mathrm{BP}(M_k)$ into two conceptually smaller steps $M_k \mapsto M'_{k+1}$ and $M'_{k+1} \mapsto M_{k+1}$. Equivalently, we decompose the expansion step $(x_k, y_k) \mapsto (x_{k+1}, y_{k+1})$ into two steps $(x_k, y_k) \mapsto (x'_{k+1}, y'_{k+1})$ and $(x'_{k+1}, y'_{k+1}) \mapsto (x_{k+1}, y_{k+1})$, where

$$x'_k = \mathbb{E}\theta_{M'_k} = \mathbb{E}\theta_{M'_k}^2, \qquad y'_k = \mathbb{E}\theta_{M'_k}^3 = \mathbb{E}\theta_{M'_k}^4. \tag{17}$$

The fine expansions we get for x and x' are as follows.

$$x'_{k+1} = (r-1)\lambda^2 x_k - (r-1)(r-2)\lambda^3 x_k^2 + O_r(\lambda^3 x_k^3), \tag{18}$$

$$x_{k+1} = dx'_{k+1} - \mathbb{E}_{b \sim D} b(b-1) x'^2_{k+1} + O_r(d^3 x'^3_{k+1}). \tag{19}$$

7

Combining these smaller steps leads to the main expansion Eq. (14).

After establishing Eq. (14), we observe that for any parameters $r \geq 3$, $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1\right]$, $d \geq 0$, if $(r-1)d\lambda^2 = 1$, then the coefficient of x_k^2 is negative. This means that x_k contracts if it is small enough. In this way we get the robust reconstruction result.

Theorem 7 (Robust reconstruction is impossible below and at KS) *Fix $r \geq 3$. Consider the model* $\mathrm{BOHT}(r, \lambda, d)$ *or* $\mathrm{BOHT}(r, \lambda, \mathrm{Pois}(d))$. *There exists $\epsilon = \epsilon(r) > 0$ such that if $(r-1)d\lambda^2 \leq 1$, then for any BMS channel P with $C_{\chi^2}(P) \leq \epsilon$, we have*

$$\lim_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(P)) = 0. \tag{20}$$

In other words, for fixed r, robust reconstruction is impossible below and at the KS threshold in a way that is uniform in d and λ.

Finally, Theorem 5 is proved by combining Prop. 6 and Theorem 7. We first take $\epsilon > 0$ in the robust reconstruction part such that Theorem 7 holds. Then we take d large enough such that large degree asymptotics works and Eq. (12) holds.

Our proof of Theorem 5 also determines the asymptotic reconstruction threshold for BOHT with $r \geq 7$.

Theorem 8 (Asymptotic BOHT reconstruction threshold) *Consider the model* $\mathrm{BOHT}(r, \lambda, d)$ *or* $\mathrm{BOHT}(r, \lambda, \mathrm{Pois}(d))$ *where $r \geq 7$ and $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1\right]$. Let $w^* = w^*(r)$ be defined as*

$$w^*(r) = \sup\{w > 0 : g_{r,w}(x) \leq x \forall 0 \leq x \leq 1\}, \tag{21}$$

$$g_{r,w}(x) = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \tanh\left(s_{r,w}(x) + \sqrt{s_{r,w}(x)}Z\right), \tag{22}$$

$$s_{r,w}(x) = \frac{w}{2}\left((1+x)^{r-1} - (1-x)^{r-1}\right). \tag{23}$$

Then for any $\epsilon > 0$, there exists $d_0 = d_0(r, \epsilon)$ such that for $d \geq d_0$, if $d\lambda^2 \geq w^* + \epsilon$ then reconstruction is possible; if $d\lambda^2 \leq w^* - \epsilon$ then reconstruction is impossible.

1.3. Organization

In Section 2 we present the proof structure for Theorem 5 (tightness of the KS threshold for $r \leq 6$ and large d), which includes the two main parts Prop. 6 (large degree asymptotics) and Theorem 7 (robust reconstruction below and at KS). In Section 3 we prove Theorem 2 (non-asymptotic HSBM weak recovery upper bound).

In Section A we provide missing proofs of lemmas used in Section 2. In Section B we provide proofs of our remaining main results, including Corollary 3 (non-tightness of the KS threshold), Theorem 4 (asymptotic HSBM weak recovery bounds), and Theorem 8 (asymptotic BOHT reconstruction threshold). In Section C we make discussions on related works and further directions.

2. Reconstruction on hypertrees

2.1. Preliminaries

A binary memoryless symmetric (BMS) channel is a channel $P : \{\pm\} \to \mathcal{Y}$ together with a measurable involution $\sigma : \mathcal{Y} \to \mathcal{Y}$ such that $P(E|+) = P(\sigma(E)|-)$ for all measurable $E \subseteq \mathcal{Y}$. Examples

of BMS channels include binary erasure channels (BECs) and binary symmetric channels (BSCs). The broadcast channel $B_{r,\lambda}$ (Eq. (7)) and the channels M_k (defined as $\sigma_\rho \mapsto (T_k, \sigma_{L_k})$) are all BMS channels (with the involution being the global sign flip).

Every BMS channel is equivalent to a mixture of BSC channels. That is, every BMS channel P is equivalent to a channel $X \mapsto (\Delta, Z)$ where $\Delta \in \left[0, \frac{1}{2}\right]$ is independent of X and $P_{Z|\Delta,X} = \mathrm{BSC}_\Delta(\cdot|X)$. We use Δ_P to denote the Δ-component of P. It is a random variable taking values in the interval $\left[0, \frac{1}{2}\right]$. In this way, BMS channels (up to channel equivalence) are in one-to-one correspondence with distributions on $\left[0, \frac{1}{2}\right]$.

In computations, it is often easier to work with the signed θ-component, rather than the Δ-component. Let P be a BMS channel and Δ_P be its Δ-component. Define $\theta_P \in [-1, 1]$ to be the random variable such that conditioned on Δ_P,

$$\theta_P = \begin{cases} 1 - 2\Delta_P, & \text{w.p. } 1 - \Delta_P, \\ 2\Delta_P - 1, & \text{w.p. } \Delta_P. \end{cases} \tag{24}$$

We call θ_P the signed θ-component of P.

The χ^2-capacity of a BMS channel P is defined as

$$C_{\chi^2}(P) = I_{\chi^2}(\pi, P) = \mathbb{E}(1 - 2\Delta_P)^2 = \mathbb{E}\theta_P = \mathbb{E}\theta_P^2, \tag{25}$$

where $\pi = \mathrm{Unif}(\{\pm\})$. For BMS channels, the χ^2-capacity is within a constant factor of the capacity. That is,

$$\frac{1}{2}C_{\chi^2}(P) \le C(P) \le (\log 2)C_{\chi^2}(P). \tag{26}$$

Therefore, Eq. (11) holds if and only if

$$\lim_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(\mathrm{Id})) = 0. \tag{27}$$

Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X} \to \mathcal{Z}$ be two channels with the same input alphabet. We define the $\star$-convolution channel $P \star Q : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ as $(P \star Q)(E \times F|x) = P(E|x)Q(F|x)$ for measurable subsets $E \subseteq \mathcal{Y}$, $F \subseteq \mathcal{Z}$. For $n \in \mathbb{Z}_{\ge 0}$, we use $P^{\star n}$ to denote the n-th $\star$-convolution power.

Let $P : \mathcal{X} \to \mathcal{Y}$ and $P' : \mathcal{X}' \to \mathcal{Y}'$ be two channels. We define the tensor product channel $P \times P' : \mathcal{X} \times \mathcal{X}' \to \mathcal{Y} \times \mathcal{Y}'$ as $(P \times P')(E \times E'|(x, x')) = P(E|x)P(E'|x')$ for measurable subsets $E \subseteq \mathcal{X}$, $E' \subseteq \mathcal{X}'$. For $n \in \mathbb{Z}_{\ge 0}$, we use $P^{\times n}$ to denote the n-th tensor power.

Let $\mathcal{A}$ be a collection of information channels with the same input alphabet $\mathcal{X}$ and μ be a distribution on $\mathcal{A}$. We define $\mathbb{E}_{P \sim \mu} P$ to be the channel that maps $x \in \mathcal{X}$ to (P, Y), where $P \sim \mu$ and $Y \sim P(\cdot|x)$. This channel is called a mixture of $\mathcal{A}$.

With the above definitions, the BP operator (Eq. (10)) is well-defined and maps BMS channels to BMS channels.

2.2. Large degree asymptotics

Proposition 9 (Large degree asymptotics, (Gu and Polyanskiy, 2023b, Prop. 30)) *Fix $r \ge 2$. For any $\epsilon > 0$ and $C > 0$, there exists $d_0 = d_0(r, \epsilon)$ such that for any $d \ge d_0$ and $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1\right]$ with $(r-1)d\lambda^2 \le C$, for any BMS channel P we have*

$$\left|C_{\chi^2}(\mathrm{BP}(P)) - g_{r,d\lambda^2}(C_{\chi^2}(P))\right| \le \epsilon, \tag{28}$$

where $g_{r,w}(x)$ is defined in Eq. (22).

(Gu and Polyanskiy, 2023b, Prop. 30) stated the result only for $\lambda > 0$ and $C = 1$, but their proof works without any change for $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1\right]$ and constant C.

Lemma 10 (Properties of g) *The function $g_{r,w}$ and $s_{r,w}$ defined in Eq. (22) satisfies the following properties.*

(i) *$g_{r,w}(x)$ is non-decreasing in $r \in \mathbb{Z}_{\geq 2}$ (strict when $w \neq 0$ and $x \neq 0$), $w \in \mathbb{R}_{\geq 0}$ (strict when $x \neq 0$), and $x \in [0, 1]$ (strict when $w \neq 0$).*

(ii) *$g_{r,w}(x) \leq s_{r,w}(x)$. The inequality is strict when $w \neq 0$ and $x \neq 0$.*

Lemma 11 *Fix $r \leq 6$. Then $g_{r,\frac{1}{r-1}}(x) < x$ for all $0 < x \leq 1$.*

Proofs of Lemma 10 and Lemma 11 are deferred to Section A.2.

Proof [Proof of Prop. 6] By Lemma 11, for any $\epsilon > 0$ there exists $\delta > 0$ such that $g_{r,\frac{1}{r-1}}(x) \leq x - \delta$ for all $x \in [\epsilon, 1]$. Take $d_0 = d_0(r, \delta/2)$ in Prop. 9. Then for any BMS channel P with $C_{\chi^2}(P) \geq \epsilon$, we have

$$C_{\chi^2}(\mathrm{BP}(P)) \leq g_{r,d\lambda^2}(C_{\chi^2}(P)) + \delta/2 \leq C_{\chi^2}(P) - \delta/2. \tag{29}$$

If Eq. (12) does not hold, then

$$\lim_{k \to \infty} C_{\chi^2}(M_k) = \lim_{k \to \infty} C_{\chi^2}(M_{k+1}) \leq \lim_{k \to \infty} (C_{\chi^2}(M_k) - \delta/2), \tag{30}$$

which is absurd. ∎

2.3. Robust reconstruction

In this section we prove Theorem 7. We note that the case $\lambda \geq \frac{1}{5}$ follows from (Gu and Polyanskiy, 2023b, Theorem 2(iii)). In the following, assume that $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \frac{1}{5}\right]$.

Fix d, λ, and P as in the statement of Theorem 7. Our goal is to prove that

$$\lim_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(P)) = 0 \tag{31}$$

when $C_{\chi^2}(P)$ is sufficiently small.

For $k \in \mathbb{Z}_{\geq 0}$, define

$$M_{P,k} = \mathrm{BP}^k(P), \qquad M'_{P,k+1} = M_{P,k}^{\times(r-1)} \circ B_{r,\lambda}. \tag{32}$$

Define

$$x_k = C_{\chi^2}(M_{P,k}) = \mathbb{E}\theta_{M_{P,k}} = \mathbb{E}\theta^2_{M_{P,k}}, \tag{33}$$

$$y_k = \mathbb{E}\theta^3_{M_{P,k}} = \mathbb{E}\theta^4_{M_{P,k}}, \tag{34}$$

$$x'_{k+1} = C_{\chi^2}(M'_{P,k+1}) = \mathbb{E}\theta_{M'_{P,k+1}} = \mathbb{E}\theta^2_{M'_{P,k+1}}, \tag{35}$$

$$y'_{k+1} = \mathbb{E}\theta^3_{M'_{P,k+1}} = \mathbb{E}\theta^4_{M'_{P,k+1}}. \tag{36}$$

Because $|\theta| \leq 1$, we have $y_k \leq x_k$, $y'_{k+1} \leq x'_{k+1}$ for all $k \in \mathbb{Z}_{\geq 0}$.

Lemma 12 (A priori estimates) *We have*

$$x'_{k+1} = (r-1)\lambda^2 x_k + O_r(\lambda^2 x_k^2), \tag{37}$$

$$y'_{k+1} = (r-1)\lambda^4 y_k + O_r(\lambda^2 x_k^2), \tag{38}$$

$$x_{k+1} = dx'_{k+1} + O_r(d^2 x'^2_{k+1}), \tag{39}$$

$$y_{k+1} = dy'_{k+1} + O_r(d^2 x'^2_{k+1}). \tag{40}$$

In particular, we have

$$x_{k+1} = (r-1)d\lambda^2 x_k + O_r(x_k^2), \tag{41}$$

$$y_{k+1} = (r-1)d\lambda^4 y_k + O_r(x_k^2). \tag{42}$$

Proof of Lemma 12 is deferred to Section A.3.

Lemma 12 already implies that robust reconstruction is impossible for $(r-1)d\lambda^2 < 1$. However the robustness parameter ϵ goes to 0 as $(r-1)d\lambda^2 \to 1$. Nevertheless, we can obtain that there exists $\epsilon = \epsilon(r) > 0$ such that for $(r-1)d\lambda^2 \le 0.99$, robust reconstruction is impossible with respect to any BMS channel P with $C_{\chi^2}(P) \le \epsilon$. In the following, assume that $0.99 \le (r-1)d\lambda^2 \le 1$.

Let $\gamma = \gamma(r) > 0$ be a large enough constant. We say a BMS channel P is γ-normal if

$$\mathbb{E}\theta_P^4 \le \gamma \left(\mathbb{E}\theta_P^2\right)^2. \tag{43}$$

Using Lemma 12, we can show that $M_{P,k}$ is γ-normal for some k not too large.

Lemma 13 *For large enough $\gamma = \gamma(r)$ the following holds. For any $\epsilon > 0$, there exists $\delta = \delta(r,\epsilon) > 0$, such that for any BMS channel P with $C_{\chi^2}(P) \le \delta$, for some $k_0 \in \mathbb{Z}_{\ge 0}$, we have*

$$x_{k_0} \le \epsilon, \qquad y_{k_0} \le \gamma x_{k_0}^2. \tag{44}$$

Proof of Lemma 13 is deferred to Section A.3.

Using the γ-normal property, we can obtain finer estimates.

Lemma 14 (Fine estimates) *For large enough $\gamma = \gamma(r)$ the following holds. If $M_{P,k}$ is γ-normal, then*

$$x'_{k+1} = (r-1)\lambda^2 x_k - (r-1)(r-2)\lambda^3 x_k^2 + O_r(\lambda^2 x_k^3), \tag{45}$$

$$x_{k+1} = dx'_{k+1} - \left(\mathbb{E}_{b\sim d}b(b-1)\right) x'^2_{k+1} + O_r(d^3 x'^3_{k+1}). \tag{46}$$

In particular,

$$x_{k+1} = (r-1)d\lambda^2 x_k - ((r-1)(r-2)d\lambda^3 + (r-1)^2 \left(\mathbb{E}_{b\sim D}b(b-1)\right)\lambda^4)x_k^2 + O_r(x_k^3). \tag{47}$$

Furthermore, $M_{P,k+1}$ is γ-normal.

Proof of Lemma 14 is deferred to Section A.4.

Proof [Proof of Theorem 7]

If $\lambda \ge \frac{1}{5}$, then the result follows from the reconstruction threshold for the BOHT model ((Gu and Polyanskiy, 2023b, Theorem 2(iii))). In the following assume $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \frac{1}{5}\right]$.

If $(r-1)d\lambda^2 \le 0.99$, then the result follows from Eq. (41) by taking $\epsilon > 0$ small enough. In the following assume $0.99 \le (r-1)d\lambda^2 \le 1$.

In Eq. (47), we have

$$(r-1)(r-2)d\lambda^3 + (r-1)^2 \left(\mathbb{E}_{b\sim D} b(b-1) \right) \lambda^4 \tag{48}$$
$$\ge (r-1)(r-2)d\lambda^3 + (r-1)^2 d(d-1)\lambda^4$$
$$= (r-1)d\lambda^2 ((r-1)d\lambda^2 + (r-2)\lambda - (r-1)\lambda^2).$$

For $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 0 \right]$, we have

$$(48) \ge 0.99 \cdot \left(0.99 - \frac{r-2}{2^{r-1}-1} - \frac{r-1}{(2^{r-1}-1)^2} \right) \ge 0.4. \tag{49}$$

For $\lambda \in \left[0, \frac{1}{5} \right]$, we have

$$(48) \ge 0.99 \cdot (0.99 + \lambda((r-2) - (r-1)\lambda)) \ge 0.99^2. \tag{50}$$

Therefore $(48) \ge 0.4$ always holds.

Take γ large enough such that Lemma 13 and Lemma 14 hold. Take ϵ such that Eq. (47) has no fixed point for all $x \le \epsilon$. Such ϵ exists because the coefficient of x^2 is upper bounded by -0.4. Then for any γ-normal BMS channel P with $C_{\chi^2}(P) \le \epsilon$, robust reconstruction with respect to P is impossible.

Take δ such that Lemma 13 holds. Then for any BMS channel P with $C_{\chi^2}(P) \le \delta$, by Lemma 13, for some $k_0 \in \mathbb{Z}_{\ge 0}$, $M_{k_0,P}$ is γ-normal and $C_{\chi^2}(M_{k_0,P}) \le \epsilon$, so robust reconstruction with respect to $M_{k_0,P}$, thus P, is impossible. This finishes the proof. ∎

2.4. Finishing the proof

We are now ready to prove Theorem 5.

Proof [Proof of Theorem 5] Take $\epsilon = \epsilon(r)$ in Theorem 7. Take $d_0 = d_0(r,\epsilon)$ in Prop. 6. Fix $d \ge d_0$, $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$ such that $(r-1)d\lambda^2 \le 1$. By Prop. 6, there exists k_0 such that $C_{\chi^2}(M_{k_0}) \le \epsilon$. By Theorem 7, we have

$$\lim_{k\to\infty} C_{\chi^2}(M_k) = \lim_{k\to\infty} C_{\chi^2}(M_{k+k_0}) = \lim_{k\to\infty} C_{\chi^2}(\mathrm{BP}^k(M_{k_0})) = 0. \tag{51}$$

∎

3. HSBM weak recovery upper bound

In this section we prove Theorem 2. We say an event happens a.a.s. (asymptotically almost surely) if it happens with probability $1 - o(1)$ as $n \to \infty$.

Proof [Proof of Theorem 2] We first describe our algorithm. Suppose we are given a HSBM instance $G = (V, E)$. We say a partition $Y : V \to \{\pm\}$ is almost balanced if $\left| \#Y^{-1}(+) - \frac{n}{2} \right| \le n^{0.6}$. We say a hyperedge $e \in E$ is in-community with respect to Y if Y_vs are equal for all $v \in e$.

We say a partition $Y : V \to \{\pm\}$ is good if the number of in-community hyperedges is in $\left[\frac{an}{2^{r-1}r} - n^{0.6}, \frac{an}{2^{r-1}r} + n^{0.6}\right]$, and bad otherwise. The algorithm looks for a good almost balanced partition. If such a partition exists, return any of them; otherwise report failure.

We say a partition Y is correlated with X (the true partition) if $d_H(Y, X) \le \left(\frac{1}{2} - \epsilon\right) n$, and uncorrelated with X otherwise. We prove that for some $\epsilon > 0$, a.a.s. the following are true.

(a) There exists a good almost balanced partition.

(b) All almost balanced partitions that are uncorrelated with X are bad.

If (a) holds, our algorithm returns a good partition. If in addition (b) holds, the returned partition is correlated with the true partition X, and the algorithm succeeds. It remains to prove (a) and (b).

Proof of (a) We show that X is a.a.s. a good almost balanced partition. By the central limit theorem, X is a.a.s. almost balanced. For such X, the number of in-community sets $S \in \binom{V}{r}$ is $\binom{|X|}{r} + \binom{n-|X|}{r} = (1 \pm O(n^{-1}))\frac{n^r}{2^{r-1} \cdot r!}$. Each such set is an hyperedge independently with probability $\frac{a}{\binom{n}{r-1}} = (1 \pm O(n^{-1}))\frac{a(r-1)!}{n^{r-1}}$. By Chebyshev's inequality, a.a.s. the number of in-community hyperedges is in $\left[\frac{an}{2^{r-1}r} - n^{0.6}, \frac{an}{2^{r-1}r} + n^{0.6}\right]$.

Proof of (b) By ignoring events that happen with probability $o(1)$, we make the following assumptions.

1. X is almost balanced.

2. The total number of hyperedges m is in $\left[\frac{dn}{r} - n^{0.6}, \frac{dn}{r} - n^{0.6}\right]$.

We show that (b) is true conditioned on m. We consider the following random hypergraph model $G(n, m, X)$. For each of the m hyperedge, generate a sequence $(x_1, \dots, x_r)$ of r bits according to the following distribution

$$p_{x_1,\dots,x_r} = \begin{cases} \frac{1}{2} \cdot \frac{a}{a-b+2^{r-1}b}, & \text{if } x_1 = \cdots = x_r, \\ \frac{1}{2} \cdot \frac{b}{a-b+2^{r-1}b}, & \text{o.w.} \end{cases} \tag{52}$$

Then choose r vertices $v_1, \dots, v_r$ independently according to $v_i \sim \text{Unif}(X^{-1}(x_i))$, and add hyperedge $e = (v_1, \dots, v_r)$ to the hypergraph.

The hypergraph $G(n, m, X)$ is not necessarily simple and may have degenerate hyperedges (i.e., hyperedges with less than r distinct vertices). However, with constant probability, this does not happen. Furthermore, conditioned on that $G(n, m, X)$ is simple, it has the same distribution as $\text{HSBM}(n, r, a, b)$ conditioned on X and m. Therefore, any event that happens a.a.s. for $G(n, m, X)$ also happens a.a.s. for $\text{HSBM}(n, r, a, b)$ conditioned on X and m.

Now consider an almost balanced partition Y uncorrelated with X. Say $d_H(Y, X) = \left(\frac{1}{2} - \delta\right) n$ for some $\delta \le \epsilon$. By possibly replacing Y with $-Y$, WLOG assume that $\#\{v \in V : X_v \ne Y_v\} = \left(\frac{1}{2} - \delta\right) n$. For $x, y \in \{\pm\}$, let $c_{x,y}$ denote the number of vertices $v \in V$ such that $X_v = x$ and $Y_v = y$. Then

$$c_{++} = \left(\frac{1}{4} + \frac{\delta}{2}\right) n \pm O(n^{0.6}), \qquad c_{+-} = \left(\frac{1}{4} - \frac{\delta}{2}\right) n \pm O(n^{0.6}), \tag{53}$$

$$c_{-+} = \left(\frac{1}{4} - \frac{\delta}{2}\right) n \pm O(n^{0.6}), \qquad c_{--} = \left(\frac{1}{4} + \frac{\delta}{2}\right) n \pm O(n^{0.6}).$$

Consider the process of generating one hyperedge under $G(n, m, X)$. Let $y_i = Y_{v_i}$. Then we have

$$\mathbb{P}[y_i = +|x_i = +] = \frac{1}{2} + \delta \pm O(n^{-0.4}), \qquad \mathbb{P}[y_i = +|x_i = -] = \frac{1}{2} - \delta \pm O(n^{-0.4}). \quad (54)$$

Therefore, under $G(n, m, X)$, each hyperedge is in-community with respect to Y with probability

$$\sum_{x_1, \ldots, x_r \in \{\pm\}} p_{x_1, \ldots, x_r} \left(\prod_{i \in [r]} \mathbb{P}[y_i = +|x_i] + \prod_{i \in [r]} \mathbb{P}[y_i = -|x_i] \right) \quad (55)$$

$$= \sum_{0 \le j \le r} p_{+j - r - j} \binom{r}{j} \left(\left(\frac{1}{2} + \delta \right)^j \left(\frac{1}{2} - \delta \right)^{r-j} + \left(\frac{1}{2} - \delta \right)^j \left(\frac{1}{2} + \delta \right)^{r-j} \right) \pm O(n^{-0.4})$$

$$= \frac{1 - \lambda}{2^{r-1}} + \lambda \left(\left(\frac{1}{2} + \delta \right)^r + \left(\frac{1}{2} - \delta \right)^r \right) \pm O(n^{-0.4}).$$

The last value can be arbitrarily close to $\frac{1}{2^{r-1}}$ for small enough $\delta > 0$.

By Sanov's theorem, the probability that there are $\left[\frac{an}{2^{r-1}r} - n^{0.6}, \frac{an}{2^{r-1}r} + n^{0.6} \right]$ in-community hyperedges with respect to Y is

$$\exp \left(-\frac{nd}{r} \cdot d_{\mathrm{KL}} \left(\lambda + \frac{1 - \lambda}{2^{r-1}} \middle\| \frac{1 - \lambda}{2^{r-1}} + \lambda \left(\left(\frac{1}{2} + \delta \right)^r + \left(\frac{1}{2} - \delta \right)^r \right) \right) \pm o(n) \right). \quad (56)$$

By Eq. (5), for $\epsilon > 0$ close enough to 0, the last expression is $\exp\left(-n(\log 2 + \epsilon') \pm o(n)\right)$ for some $\epsilon' > 0$. Therefore every uncorrelated balanced partition Y is good with probability $\exp\left(-n(\log 2 + \epsilon') \pm o(n)\right)$. On the other hand, the total number of partitions is 2^n. By union bound, a.a.s. there are no good uncorrelated balanced partitions. This finishes the proof of (b). ∎

Acknowledgments

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.

Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 66–73. IEEE, 2015.

Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*, pages 383–416. PMLR, 2016.

Pavel M. Bleher, Jean Ruiz, and Valentin A. Zagrebnov. On the purity of the limiting gibbs state for the ising model on the bethe lattice. *Journal of Statistical Physics*, 79:473–482, 1995.

Christian Borgs, Jennifer Chayes, Elchanan Mossel, and Sébastien Roch. The kesten-stigum reconstruction bound is tight for roughly symmetric binary channels. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 518–530. IEEE, 2006.

Philip Chodrow, Nicole Eikmeier, and Jamie Haddock. Nonbacktracking spectral clustering of nonuniform hypergraphs. *SIAM Journal on Mathematics of Data Science*, 5(2):251–279, 2023.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

Ioana Dumitriu, Haixiao Wang, and Yizhe Zhu. Partial recovery and weak consistency in the non-uniform hypergraph stochastic block model. *arXiv preprint arXiv:2112.11671*, 2021.

William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000.

Yuzhou Gu. *Channel Comparison Methods and Statistical Problems on Graphs*. PhD thesis, Massachusetts Institute of Technology, 2023.

Yuzhou Gu and Yury Polyanskiy. Non-linear log-Sobolev inequalities for the Potts semigroup and applications to reconstruction problems. *Communications in Mathematical Physics*, 404(2):769–831, 2023a.

Yuzhou Gu and Yury Polyanskiy. Weak recovery threshold for the hypergraph stochastic block model. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 885–920. PMLR, 12–15 Jul 2023b. URL https://proceedings.mlr.press/v195/gu23b.html.

Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *The Annals of Probability*, 32(3B):2630–2649, 2004.

Harry Kesten and Bernt P. Stigum. Additional limit theorems for indecomposable multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, 37(6):1463–1481, 1966.

Christof Külske and Marco Formentin. A symmetric entropy bound on the non-reconstruction regime of Markov chains on Galton-Watson trees. *Electronic Communications in Probability*, 14:587–596, 2009.

Wenjian Liu and Ning Ning. Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels. *Journal of Statistical Physics*, 174:1161–1188, 2019.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing*, pages 694–703, 2014.

Elchanan Mossel. Survey: Information flow on trees. *arXiv preprint math/0406446*, 2004.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461, 2015.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Elchanan Mossel, Allan Sly, and Youngtak Sohn. Exact phase transitions for stochastic block models and reconstruction on trees. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 96–102, 2023.

Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, 59(3):407–463, 2021.

Allan Sly. Reconstruction of random colourings. *Communications in Mathematical Physics*, 288 (3):943–961, 2009.

Allan Sly. Reconstruction for the Potts model. *The Annals of Probability*, 39(4):1365 – 1406, 2011. doi: 10.1214/10-AOP584. URL https://doi.org/10.1214/10-AOP584.

Allan Sly and Yumeng Zhang. Reconstruction of colourings without freezing. *arXiv preprint arXiv:1610.02770*, 2016.

Ludovic Stephan and Yizhe Zhu. Sparse random hypergraphs: Non-backtracking spectra and community detection. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 567–575. IEEE, 2022.

Mingao Yuan, Ruiqi Liu, Yang Feng, and Zuofeng Shang. Testing community structure for hypergraphs. *The Annals of Statistics*, 50(1):147–169, 2022.

Erchuan Zhang, David Suter, Giang Truong, and Syed Zulqarnain Gilani. Sparse hypergraph community detection thresholds in stochastic block model. In *Thirty-Sixth Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

Yumeng Zhang. *Phase Transitions of Random Constraints Satisfaction Problem*. University of California, Berkeley, 2017.

Appendix A. Reconstruction on hypertrees

A.1. More preliminaries

Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X} \to \mathcal{Z}$ be two channels. We say P is more degraded than Q, denoted $P \leq_{\deg} Q$, if there exists a channel $R : \mathcal{Z} \to \mathcal{Y}$ such that $P = R \circ Q$. If P and Q are BMSs, then it is possible to take R such that it respects the involution structure of BMSs.

For two BMS channels P and Q, $P \leq_{\deg} Q$ if and only if there exists a coupling between Δ_P and Δ_Q such that $\mathbb{E}[\Delta_Q|\Delta_P] \leq \Delta_P$ almost surely.

For two BMS channels P and Q, if $P \leq_{\deg} Q$, then

$$C_{\chi^2}(P) \leq C_{\chi^2}(Q). \tag{57}$$

This is known as the data processing inequality (DPI).

Degradation relationship is preserved under pre-composition, $\star$-convolution, tensor product, and mixtures.

Lemma 15 *Consider two BOHT models* $\mathrm{BOHT}(r, \lambda, D)$ *and* $\mathrm{BOHT}(r, c\lambda, D)$ *for some* $c \in [0, 1]$. *Let* BP *and* $\widetilde{\mathrm{BP}}$ *be the corresponding* BP *operators. Then for any BMS channel P, we have* $\widetilde{\mathrm{BP}}(P) \leq_{\deg} \mathrm{BP}(P)$. *In particular, we have*

$$\limsup_{k \to \infty} C_{\chi^2}(\mathrm{BP}^k(P)) \geq \limsup_{k \to \infty} C_{\chi^2}(\widetilde{\mathrm{BP}}^k(P)). \tag{58}$$

In other words, if robust reconstruction with respect to P is impossible for $\mathrm{BOHT}(r, \lambda, D)$*, then robust reconstruction with respect to P is impossible for* $\mathrm{BOHT}(r, c\lambda, D)$.

Proof Recall that

$$\mathrm{BP}(P) = \mathbb{E}_{b \sim D}(P^{\times(r-1)} \circ B_{r,\lambda})^{\star b}, \qquad \widetilde{\mathrm{BP}}(P) = \mathbb{E}_{b \sim D}(P^{\times(r-1)} \circ B_{r,c\lambda})^{\star b}. \tag{59}$$

Because $\leq_{\deg}$ is preserved under mixtures and $\star$-convolution, it suffices to prove that

$$P^{\times(r-1)} \circ B_{r,c\lambda} \leq_{\deg} P^{\times(r-1)} \circ B_{r,\lambda}. \tag{60}$$

Let $\mathcal{Y}$ be the output alphabet of P. Then the output alphabets of $P^{\times(r-1)} \circ B_{r,\lambda}$ and $P^{\times(r-1)} \circ B_{r,c\lambda}$ are $\mathcal{Y}^{r-1}$. We construct a channel $R : \mathcal{Y}^{r-1} \to \mathcal{Y}^{r-1}$ such that

$$P^{\times(r-1)} \circ B_{r,c\lambda} = R \circ P^{\times(r-1)} \circ B_{r,\lambda}. \tag{61}$$

Let R be the following channel: on input $y^{r-1} \in \mathcal{Y}^{r-1}$, it outputs y^{r-1} with probability c, and outputs a sample from $(P \circ \mathrm{Unif}(\{\pm\}))^{\times(r-1)}$ with probability $1 - c$. It is easy to check that Eq. (61) holds. ∎

Let $P : \{\pm\} \to \mathcal{Y}$ be a BMS channel with a discrete target space $\mathcal{Y}$. For $x \in \{\pm\}$, we use P_x to denote the distribution $P(\cdot|x)$. The θ-component θ_P is directly related with the distribution P_+. For $y \in \mathcal{Y}$, define

$$\theta_P(y) = \frac{P_+(y) - P_-(y)}{P_+(y) + P_-(y)}. \tag{62}$$

Then the distribution of $\theta_P(y)$ for $y \sim P_+$ is exactly the same as the distribution of θ_P. In particular,

$$C_{\chi^2}(P) = \mathbb{E}_{y \sim P_+} \theta_P(y). \tag{63}$$

A.2. Large degree asymptotics

Proof [Proof of Lemma 10] Let $g : [0, 1] \to [0, 1]$ be the function

$$g(s) = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \tanh(s + \sqrt{s}Z). \tag{64}$$

By (Sly, 2011, Lemma 4.4), g is strictly increasing and continuously differentiable on $[0, 1]$.

Proof of (i) Because g is strictly increasing, the result follows from that $s_{r,w}(x)$ (Eq. (23)) is non-decreasing in $r \in \mathbb{Z}_{\geq 2}$ (strict when $w \neq 0$ and $x \neq 0$), $w \in \mathbb{R}_{\geq 0}$ (strict when $x \neq 0$), and $x \in [0,1]$ (strict when $w \neq 0$).

Proof of (ii) It suffices to prove that $g(s) \leq s$ for all $s \in [0,1]$, and the inequality is strict when $s \neq 0$. We have

$$g(s) = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \tanh(s + \sqrt{s}Z) \tag{65}$$
$$= \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \frac{1}{2} \left(\tanh(s + \sqrt{s}Z) + \tanh(s - \sqrt{s}Z)\right)$$
$$\leq \tanh(s) \leq s.$$

The third step is by Lemma 16. The last step is strict when $s \neq 0$. ∎

Lemma 16 *For $x, y \in \mathbb{R}$, if $x + y \geq 0$, then $\tanh x + \tanh y \leq 2 \tanh \frac{x+y}{2}$.*

Proof We have

$$\tanh x + \tanh y = 2 \tanh \frac{x+y}{2} \cdot \frac{1 + \tanh x \tanh y}{1 + \tanh^2 \frac{x+y}{2}}. \tag{66}$$

It suffices to prove that $\tanh x \tanh y \leq \tanh^2 \frac{x+y}{2}$. If $xy < 0$ then LHS is < 0 and the inequality holds. If $x, y \geq 0$, then the inequality follows because $\log \tanh x$ is concave on $\mathbb{R}_{>0}$, ∎

Proof [Proof of Lemma 11] Recall function g defined in Eq. (64). Because g is strictly increasing on $[0,1]$ and $s_{r,\frac{1}{r-1}}(x)$ is non-decreasing in $r \in \{2,3,4,5,6\}$ for $0 \leq x \leq 1$, $g_{r,\frac{1}{r-1}}(x) = g\left(s_{r,\frac{1}{r-1}}(x)\right)$ is non-decreasing in the same range. Therefore it suffices to prove the statement for $r = 6$.

Our goal is to show that

$$g\left(x + 2x^3 + \frac{x^5}{5}\right) < x \tag{67}$$

for all $x \in (0,1]$.

Let $\delta = 0.1$. We prove the result for $x \in (0, \delta]$ and $x \in [\delta, 1]$ separately.

Case $x \in (0, \delta]$. Write $s = x + 2x^3 + \frac{x^5}{5}$. For $Z \sim \mathcal{N}(0,1)$, let $Z_+ = \exp(s + \sqrt{s}Z)$, $Z_- = \exp(-s - \sqrt{s}Z)$. Then

$$g(s) = \mathbb{E}_Z \frac{Z_+ - Z_-}{Z_+ + Z_-} = \frac{1}{2} \mathbb{E}_Z \frac{Z_+ - Z_-}{1 + \frac{1}{2}(Z_+ + Z_- - 2)}. \tag{68}$$

Expanding the last term using

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \frac{x^4}{1+x}, \tag{69}$$

we get

$$g(s) = \frac{L_0}{2} - \frac{L_1}{4} + \frac{L_2}{8} - \frac{L_3}{16} + \frac{E_4}{16}, \tag{70}$$

where

$$L_0 = \mathbb{E}_Z[Z_+ - Z_-], \tag{71}$$
$$L_1 = \mathbb{E}_Z[(Z_+ - Z_-)(Z_+ + Z_- - 2)], \tag{72}$$
$$L_2 = \mathbb{E}_Z[(Z_+ - Z_-)(Z_+ + Z_- - 2)^2], \tag{73}$$
$$L_3 = \mathbb{E}_Z[(Z_+ - Z_-)(Z_+ + Z_- - 2)^3], \tag{74}$$
$$E_4 = \mathbb{E}_Z\left[\frac{(Z_+ - Z_-)(Z_+ + Z_- - 2)^4}{Z_+ + Z_-}\right] \le \mathbb{E}_Z(Z_+ + Z_- - 2)^4. \tag{75}$$

Computing the terms using

$$\mathbb{E}_Z \exp(\mu + \sigma Z) = e^{\mu + \frac{\sigma^2}{2}}, \tag{76}$$

we get

$$L_0 = e^{3s/2} - e^{-s/2}, \tag{77}$$
$$L_1 = e^{4s} - 2e^{3s/2} + 2e^{-s/2} - 1, \tag{78}$$
$$L_2 = e^{15s/2} - 4e^{4s} + 4e^{3s/2} - 5e^{-s/2} + 4, \tag{79}$$
$$L_3 = e^{12s} - 6e^{15s/2} + 13e^{4s} - 8e^{3s/2} + 14e^{-s/2} - 14, \tag{80}$$
$$E_4 \le e^{12s} - 8e^{15s/2} + 29e^{4s} - 64e^{3s/2} - 56e^{-s/2} + 98. \tag{81}$$

Collecting terms, we get

$$g(s) \le \frac{1}{4}\left(e^{4s} - 8e^{3s/2} - 24e^{-s/2} + 31\right). \tag{82}$$

When $x \le 0.1$, we have $s \le 0.105$. So the largest exponent is $0.105 \cdot 4 \le 0.5$. By Taylor's theorem, for $y \le 0.5$, we have

$$\left|e^y - \sum_{0 \le i \le 3} \frac{y^i}{i!}\right| \le e^{0.5}\frac{y^4}{4!} \le \frac{y^4}{12}. \tag{83}$$

Applying this to Eq. (82), we get

$$g(s) \le s - s^2 + \frac{5}{3}s^3 + \frac{135}{16}s^4 \le s - \frac{2}{3}s^2, \tag{84}$$

where the second step uses $s \le 0.105$. Note that $s - \frac{2}{3}s^2$ is increasing for $s \in [0, 0.105]$. Because $x \le 0.1$, we have

$$s = x + 2x^3 + \frac{x^5}{5} \le x + 0.2002x^2 \le x + \frac{x^2}{4}. \tag{85}$$

19

Therefore

$$g(s) \leq \left(x + \frac{x^2}{4}\right) - \frac{2}{3}\left(x + \frac{x^2}{4}\right)^2 = x - \frac{5x^2}{12} - \frac{x^3}{3} - \frac{x^4}{24} < x. \tag{86}$$

Case $x \in [\delta, 1]$. Because $\frac{ds}{dx} = 1 + 6x^2 + x^4$, s is 8-Lipschitz for $x \in [\delta, 1]$. For $x \geq \delta$, $\frac{d\sqrt{s}}{dx} = \frac{1 + 6x^2 + x^4}{2\sqrt{x + 2x^3 + \frac{x^5}{5}}} \leq 2.5$. So $\sqrt{s}$ is 2.5-Lipschitz for $x \in [\delta, 1]$.

Let $x, x' \in [\delta, 1]$, $s = x + 2x^3 + \frac{x^5}{5}$, $s = x' + 2x'^3 + \frac{x'^5}{5}$. Then

$$|g(s) - g(s')| \leq |s - s'| + \mathbb{E}_{Z \sim \mathcal{N}(0,1)} |Z| \left|\sqrt{s} - \sqrt{s'}\right| \tag{87}$$

$$= |s - s'| + \sqrt{\frac{2}{\pi}} \left|\sqrt{s} - \sqrt{s'}\right|$$

$$\leq 10|x - x'|.$$

Consider the set $S = \{\delta, \delta + 0.001, \ldots, 1\}$. Suppose we know that $g\left(x + 2x^3 + \frac{x^5}{5}\right) \leq x - 0.006$ for all $x \in S$. Then for any $x \in [\delta, 1]$, there exists $x^* \in S$ such that $|x - x^*| \leq 0.0005$, and

$$g\left(x + 2x^3 + \frac{x^5}{5}\right) \leq g\left(x^* + 2x^{*3} + \frac{x^{*5}}{5}\right) + 0.005 \leq x^* - 0.001 \leq x - 0.0005. \tag{88}$$

Therefore it remains to prove that $g\left(x + 2x^3 + \frac{x^5}{5}\right) \leq x - 0.006$ for all $x \in S$. We do this using a rigorous numerical integration. Suppose we would like to evaluate $g(s)$ where $s = x + 2x^3 + \frac{x^5}{5}$. Let n be a positive integer to determined later.

$$g(s) = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \tanh(s + \sqrt{s}Z) \tag{89}$$

$$= \mathbb{P}[|Z| \geq 5] + \int_{-5}^{5} \tanh(s + \sqrt{s}t) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt$$

$$\leq 10^{-6} + \int_{-5}^{5} \tanh(s + \sqrt{s}t) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt$$

$$\leq 10^{-6} + \sum_{0 \leq i \leq 10n-1} \frac{1}{n} \tanh\left(s + \sqrt{s}\left(-5 + \frac{i+1}{n}\right)\right) \frac{e^{-\frac{1}{2}\max\{(-5+\frac{i}{n})^2, (-5+\frac{i+1}{n})^2\}}}{\sqrt{2\pi}}.$$

The last expression can be evaluated to arbitrarily high precision using e.g., Mathematica. Computation shows that taking $n = 1000$ suffices. ∎

A.3. A priori estimates

The goal of this section is to prove Lemma 12 and its corollary Lemma 13.

Proof [Proof of Lemma 12, Eq. (37)] Given $x_k = C_{\chi^2}(M_{P,k})$, we would like to compute $x'_{k+1} = C_{\chi^2}(M'_{P,k+1})$ where $M'_{P,k+1} = M_{P,k}^{\times(r-1)} \circ B_{r,\lambda}$. Let $\theta_1, \ldots, \theta_{r-1}$ be i.i.d. copies of $\theta_{M_{P,k}}$. Then

$$x'_{k+1} = C_{\chi^2}\left(M_{P,k}^{\times(r-1)} \circ B_{r,\lambda}\right) \tag{90}$$

$$= \mathbb{E}_{\theta_1, \ldots, \theta_{r-1}} C_{\chi^2}((B_{\theta_1} \times \cdots \times B_{\theta_{r-1}}) \circ B_{r,\lambda}),$$

where B_θ denotes the channel $\mathrm{BSC}_{\frac{1-\theta}{2}}$.

For fixed $\theta_1, \ldots, \theta_{r-1} \in [-1, 1]$, write $R = (B_{\theta_1} \times \cdots \times B_{\theta_{r-1}}) \circ B_{r,\lambda} : \{\pm\} \to \{\pm\}^{r-1}$. For $y \in \{\pm\}^{r-1}$, we have

$$R_+(y) = \sum_{z \in \{\pm\}^{r-1}} (B_{r,\lambda})_+(z) \prod_{i \in [r-1]} B_{\theta_i}(y_i | z_i) \tag{91}$$

$$= \sum_{z \in \{\pm\}^{r-1}} \left(\frac{1-\lambda}{2^{r-1}} + \lambda \mathbb{1}\{z = +^{r-1}\} \right) \prod_{i \in [r-1]} \frac{1}{2}(1 + \theta_i y_i z_i)$$

$$= \frac{1-\lambda}{2^{r-1}} + \frac{\lambda}{2^{r-1}} \prod_{i \in [r-1]} (1 + \theta_i y_i).$$

For simplicity, for two sequences $\theta \in [-1, 1]^{r-1}$, $y \in \{\pm\}^{r-1}$, define

$$\alpha(\theta, y) = \prod_{i \in [r-1]} (1 + \theta_i y_i). \tag{92}$$

Then we have

$$R_+(y) = \frac{1-\lambda}{2^{r-1}} + \frac{\lambda}{2^{r-1}} \alpha(\theta, y). \tag{93}$$

Then $\theta_R(y)$, the θ-component corresponding to $y \in \{\pm\}^{r-1}$, is described as

$$\theta_R(y) = \frac{R_+(y) - R_-(y)}{R_+(y) + R_-(y)} = \frac{\lambda(\alpha(\theta, y) - \alpha(\theta, -y))}{2(1 - \lambda) + \lambda(\alpha(\theta, y) + \alpha(\theta, -y))}. \tag{94}$$

Combining Eq. (93) and Eq. (94), we get

$$C_{\chi^2}(R) = \sum_{y \in \{\pm\}^{r-1}} R_+(y) \theta_R(y) = \frac{\lambda^2}{2^r} \sum_{y \in \{\pm\}^{r-1}} \frac{\alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))}{1 + \frac{\lambda}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)}. \tag{95}$$

Expanding the last term using

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}, \tag{96}$$

we get

$$C_{\chi^2}(R) = \frac{\lambda^2}{2^r} \left(T_0(\theta) - \frac{\lambda}{2} T_1(\theta) + \frac{\lambda^2}{4} R_2(\theta) \right), \tag{97}$$

where

$$T_0(\theta) = \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y)), \tag{98}$$

$$T_1(\theta) = \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))(\alpha(\theta, y) + \alpha(\theta, -y) - 2), \tag{99}$$

$$R_2(\theta) = \sum_{y \in \{\pm\}^{r-1}} \frac{\alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))(\alpha(\theta, y) + \alpha(\theta, -y) - 2)^2}{1 + \frac{\lambda}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)}. \tag{100}$$

21

Note that

$$\frac{\lambda}{2}|R_2(\theta)| \leq \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) + \alpha(\theta, -y) - 2)^2 \left| \frac{R_+(y) - R_-(y)}{R_+(y) + R_-(y)} \right| \qquad (101)$$

$$\leq \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) + \alpha(\theta, -y) - 2)^2.$$

By Lemma 17, we have

$$\mathbb{E}T_0(\theta) = 2^{r-1} \left((1 + x_k)^{r-1} - (1 - x_k)^{r-1} \right), \qquad (102)$$

$$\mathbb{E}T_1(\theta) = 2^{r-1} \left((1 + 3x_k)^{r-1} - 2(1 + x_k)^{r-1} + (1 - x_k)^{r-1} \right), \qquad (103)$$

$$\frac{\lambda}{2}\mathbb{E}|R_2(\theta)| \leq 2^{r-1} \left((1 + 3x_k)^{r-1} - (1 - x_k)^{r-1} - 4(1 + x_k)^{r-1} + 4 \right). \qquad (104)$$

Expanding with respect to x_k, we get

$$\mathbb{E}T_0(\theta) = 2^r(r - 1)x_k + O_r(x_k^2), \qquad (105)$$

$$\mathbb{E}T_1(\theta) = O_r(x_k^2), \qquad (106)$$

$$\frac{\lambda}{2}\mathbb{E}|R_2(\theta)| = O_r(x_k^2). \qquad (107)$$

Finally,

$$x'_{k+1} = \mathbb{E}_{\theta_1, \ldots, \theta_{r-1}} C_{\chi^2}((B_{\theta_1} \times \cdots \times B_{\theta_{r-1}}) \circ B_{r,\lambda}) \qquad (108)$$

$$= \mathbb{E}_{\theta_1, \ldots, \theta_{r-1}} \frac{\lambda^2}{2^r} \left(T_0(\theta) - \frac{\lambda}{2}T_1(\theta) + \frac{\lambda^2}{4}R_2(\theta) \right)$$

$$= (r - 1)\lambda^2 x_k + O_r(\lambda^2 x_k^2).$$

This finishes the proof. ■

Proof [Proof of Lemma 12, Eq. (38)] Use the same notation as in the proof of Lemma 12, Eq. (37). We need to compute

$$\mathbb{E}\theta_R^3 = \sum_{y \in \{\pm\}^{r-1}} R_+(y)\theta_R^3(y) = \frac{\lambda^4}{2^{r+2}} \sum_{y \in \{\pm\}^{r-1}} \frac{\alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))^3}{\left(1 + \frac{\lambda}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)\right)^3}. \qquad (109)$$

Expanding using

$$\frac{1}{(1 + x)^3} = 1 - 3x + \frac{6x^2 + 8x^3 + 3x^4}{(1 + x)^3}, \qquad (110)$$

we get

$$\mathbb{E}\theta_R^3 = \sum_{y \in \{\pm\}^{r-1}} R_+(y)\theta_R^3(y) = \frac{\lambda^4}{2^{r+2}} \left(S_0(\theta) - \frac{3\lambda}{2}S_1(\theta) + E_2(\theta) \right), \qquad (111)$$

where

$$S_0(\theta) = \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))^3, \tag{112}$$

$$S_1(\theta) = \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))^3(\alpha(\theta, y) + \alpha(\theta, -y) - 2), \tag{113}$$

$$E_2(\theta) = \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))^3 \cdot \frac{6x^2 + 8x^3 + 3x^4}{(1+x)^3} \tag{114}$$

where $x = \frac{\lambda}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)$.

Note that

$$\frac{\lambda^3}{8}|E_2(\theta)| \leq \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(6x^2 + 8x^3 + 3x^4) \left| \frac{R_+(y) - R_-(y)}{R_+(y) + R_-(y)} \right|^3 \tag{115}$$

$$\leq \sum_{y \in \{\pm\}^{r-1}} \alpha(\theta, y)(6x^2 + 8x^3 + 3x^4).$$

By Lemma 17, we have

$$\mathbb{E}S_0(\theta) = 2^{r-1}\left((1 + 6x_k + y_k)^{r-1} - 4(1 - y_k)^{r-1} + 3(1 - 2x_k + y_k)^{r-1}\right), \tag{116}$$

$$\mathbb{E}S_1(\theta) = 2^{r-1}\left((1 + 10x_k + 5y_k)^{r-1} - 3(1 + 2x_k - 3y_k)^{r-1} - 4(1 - 2x_k + y_k)^{r-1}\right. \tag{117}$$

$$\left. -2(1 + 6x_k + y_k)^{r-1} + 8(1 - y_k)^{r-1}\right),$$

$$\frac{\lambda^3}{8}\mathbb{E}|E_2(\theta)| \leq \frac{3\lambda^2}{2} \cdot 2^{r-1}\left((1 + 3x_k)^{r-1} - (1 - x_k)^{r-1} - 4(1 + x_k)^{r-1} + 4\right) \tag{118}$$

$$+ \lambda^3 \cdot 2^{r-1}\left((1 + 6x_k + y_k)^{r-1} + 4(1 - y_k)^{r-1} - 6(1 + 3x_k)^{r-1}\right.$$

$$\left. +3(1 - 2x_k + y_k)^{r-1} - 6(1 - x_k)^{r-1} + 12(1 + x_k)^{r-1} - 8\right)$$

$$+ \frac{3\lambda^4}{16} \cdot 2^{r-1}\left((1 + 10x_k + 5y_k)^{r-1} + 5(1 + 2x_k - 3y_k)^{r-1} - 8(1 + 6x_k + y_k)^{r-1}\right.$$

$$+ 10(1 - 2x_k + y_k)^{r-1} - 32(1 - y_k)^{r-1} + 24(1 + 3x_k)^{r-1} - 24(1 - 2x_k + y_k)^{r-1}$$

$$\left. +40(1 - x_k)^{r-1} - 32(1 + x_k)^{r-1} + 16\right).$$

Expanding with respect to x_k (and using that $y_k \leq x_k$) we get

$$\mathbb{E}S_0(\theta) = 2^{r+2}(r - 1)y_k + O_r(x_k^2), \tag{119}$$

$$\mathbb{E}S_1(\theta) = O_r(x_k^2), \tag{120}$$

$$\frac{\lambda^3}{8}\mathbb{E}|E_2(\theta)| = O_r(\lambda^2 x_k^2). \tag{121}$$

Finally,

$$y'_{k+1} = \mathbb{E}_{\theta_1,\dots,\theta_{r-1}}\mathbb{E}\theta_R^3 \tag{122}$$

$$= \mathbb{E}_{\theta_1,\dots,\theta_{r-1}}\frac{\lambda^4}{2^{r+2}}\left(S_0(\theta) - \frac{3\lambda}{2}S_1(\theta) + E_2(\theta)\right)$$

$$= (r - 1)\lambda^4 y_k + O_r(\lambda^2 x_k^2).$$

This finishes the proof. ■

Proof [Proof of Lemma 12, Eq. (39)] Given $x'_{k+1} = C_{\chi^2}(M'_{P,k+1})$, we would like to compute $x_{k+1} = C_{\chi^2}(M_{P,k+1})$, where $M_{P,k+1} = \mathbb{E}_{b\sim D}(M'_{P,k+1})^{\star b}$. Let $(\theta_i)_{i\in\mathbb{Z}_{\geq 0}}$ be i.i.d. copies of $\theta_{M'_{P,k+1}}$. Then

$$x_{k+1} = C_{\chi^2}\left(\mathbb{E}_{b\sim D}(M'_{P,k+1})^{\star b}\right) \tag{123}$$

$$= \mathbb{E}_{b\sim D}\mathbb{E}_\theta C_{\chi^2}(B_{\theta_1} \star \cdots \star B_{\theta_b}). \tag{124}$$

For fixed b and $(\theta_i)_{i\in\mathbb{Z}_{\geq 0}}$, write $R = B_{\theta_1} \star \cdots \star B_{\theta_b} : \{\pm\} \to \{\pm\}^b$. For $y \in \{\pm\}^b$, we have

$$R_+(y) = \prod_{i\in[b]}(B_{\theta_i})_+(y_i) = \frac{1}{2^b}\alpha(\theta, y). \tag{125}$$

Then $\theta_R(y)$, the θ-component corresponding to $y \in \{\pm\}^b$, is described as

$$\theta_R(y) = \frac{R_+(y) - R_-(y)}{R_+(y) + R_-(y)} = \frac{\alpha(\theta, y) - \alpha(\theta, -y)}{\alpha(\theta, y) + \alpha(\theta, -y)}. \tag{126}$$

Combining Eq. (125) and Eq. (126), we get

$$C_{\chi^2}(R) = \sum_{y\in\{\pm\}^b} R_+(y)\theta_R(y) = \frac{1}{2^{b+1}} \sum_{y\in\{\pm\}^b} \frac{\alpha(\theta, y) - \alpha(\theta, -y)}{1 + \frac{1}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)}. \tag{127}$$

Expanding the last term using Eq. (96), we get

$$C_{\chi^2}(R) = \frac{1}{2^{b+1}}\left(T_0(\theta) - \frac{1}{2}T_1(\theta) + \frac{1}{4}R_2(\theta)\right), \tag{128}$$

where

$$T_0(\theta) = \sum_{y\in\{\pm\}^b} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y)), \tag{129}$$

$$T_1(\theta) = \sum_{y\in\{\pm\}^b} \alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))(\alpha(\theta, y) + \alpha(\theta, -y) - 2), \tag{130}$$

$$R_2(\theta) = \sum_{y\in\{\pm\}^b} \frac{\alpha(\theta, y)(\alpha(\theta, y) - \alpha(\theta, -y))(\alpha(\theta, y) + \alpha(\theta, -y) - 2)^2}{1 + \frac{1}{2}(\alpha(\theta, y) + \alpha(\theta, -y) - 2)}. \tag{131}$$

Using the same computation as the proof of Lemma 12, Eq. (37), we get

$$\mathbb{E}_\theta T_0(\theta) = 2^b\left((1 + x'_{k+1})^b - (1 - x'_{k+1})^b\right), \tag{132}$$

$$\mathbb{E}_\theta T_1(\theta) = 2^b\left((1 + 3x'_{k+1})^b - 2(1 + x'_{k+1})^b + (1 - x'_{k+1})^b\right), \tag{133}$$

$$\frac{1}{2}\mathbb{E}_\theta R_2(\theta) \leq 2^b\left((1 + 3x'_{k+1})^b - (1 - x'_{k+1})^b - 4(1 + x'_{k+1})^b + 4\right). \tag{134}$$

By Eq. (37), $x'_{k+1} = O_r(\lambda^2 x_k)$. Using Lemma 18, we have

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_0(\theta) = dx'_{k+1} + O_r(d^2 x'^2_{k+1}), \tag{135}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_1(\theta) = O_r(d^2 x'^2_{k+1}), \tag{136}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta R_2(\theta) = O_r(d^2 x'^2_{k+1}). \tag{137}$$

Finally,

$$\begin{aligned} x_{k+1} &= \mathbb{E}_b \mathbb{E}_{\theta_1,\ldots,\theta_b} C_{\chi^2}(B_{\theta_1} \star \cdots \star B_{\theta_b}) \\ &= \mathbb{E}_b \mathbb{E}_{\theta_1,\ldots,\theta_b} \frac{1}{2^{b+1}} \left(T_0(\theta) - \frac{1}{2} T_1(\theta) + \frac{1}{4} R_2(\theta) \right) \\ &= dx'_{k+1} + O_r(d^2 x'^2_{k+1}). \end{aligned} \tag{138}$$

This finishes the proof. $\blacksquare$

Proof [Proof of Lemma 12, Eq. (40)] Use the same notation as in the proof of Lemma 12, Eq. (39). We need to compute

$$\mathbb{E}\theta_R^3 = \sum_{y\in\{\pm\}^b} R_+(y)\theta_R^3(y) = \frac{1}{2^{b+3}} \sum_{y\in\{\pm\}^b} \frac{\alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))^3}{\left(1 + \frac{1}{2}(\alpha(\theta,y) + \alpha(\theta,-y) - 2)\right)^3}. \tag{139}$$

Expanding using Eq. (110), we get

$$\mathbb{E}\theta_R^3 = \sum_{y\in\{\pm\}^b} R_+(y)\theta_R^3(y) = \frac{1}{2^{b+3}} \left(S_0(\theta) - \frac{3}{2} S_1(\theta) + E_2(\theta) \right), \tag{140}$$

where

$$S_0(\theta) = \sum_{y\in\{\pm\}^b} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))^3, \tag{141}$$

$$S_1(\theta) = \sum_{y\in\{\pm\}^b} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))^3(\alpha(\theta,y) + \alpha(\theta,-y) - 2), \tag{142}$$

$$E_2(\theta) = \sum_{y\in\{\pm\}^b} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))^3 \cdot \frac{6x^2 + 8x^3 + 3x^4}{(1+x)^3} \tag{143}$$

where $x = \frac{1}{2}(\alpha(\theta,y) + \alpha(\theta,-y) - 2)$.

Using the same computation as the proof of Lemma 12, Eq. (38), we get

$$\mathbb{E}_\theta S_0(\theta) = 2^b \left((1 + 6x'_{k+1} + y'_{k+1})^b - 4(1 - y'_{k+1})^b + 3(1 - 2x'_{k+1} + y'_{k+1})^b \right), \quad (144)$$

$$\mathbb{E}_\theta S_1(\theta) = 2^b \left((1 + 10x'_{k+1} + 5y'_{k+1})^b - 3(1 + 2x'_{k+1} - 3y'_{k+1})^b \right. \quad (145)$$

$$\left. -4(1 - 2x'_{k+1} + y'_{k+1})^b - 2(1 + 6x'_{k+1} + y'_{k+1})^b + 8(1 - y'_{k+1})^b \right),$$

$$\frac{1}{8}\mathbb{E}_\theta|E_2(\theta)| \le \frac{3}{2} \cdot 2^b \left((1 + 3x'_{k+1})^b - (1 - x'_{k+1})^b - 4(1 + x'_{k+1})^b + 4 \right) \quad (146)$$

$$+ 2^b \left((1 + 6x'_{k+1} + y'_{k+1})^b + 4(1 - y'_{k+1})^b - 6(1 + 3x'_{k+1})^b \right.$$

$$+ 3(1 - 2x'_{k+1} + y'_{k+1})^b - 6(1 - x'_{k+1})^b + 12(1 + x'_{k+1})^b - 8 \Big)$$

$$+ \frac{3}{16} \cdot 2^b \left((1 + 10x'_{k+1} + 5y'_{k+1})^b + 5(1 + 2x'_{k+1} - 3y'_{k+1})^b \right.$$

$$- 8(1 + 6x'_{k+1} + y'_{k+1})^b + 10(1 - 2x'_{k+1} + y'_{k+1})^b - 32(1 - y'_{k+1})^b$$

$$+ 24(1 + 3x'_{k+1})^b - 24(1 - 2x'_{k+1} + y'_{k+1})^b + 40(1 - x'_{k+1})^b$$

$$- 32(1 + x'_{k+1})^b + 16 \Big).$$

Using $y'_{k+1} \le x'_{k+1} = O_r(\lambda^2 x_k)$ and Lemma 18, we have

$$\mathbb{E}_b \frac{1}{2^{b+3}} \mathbb{E}_\theta S_0(\theta) = dy'_{k+1} + O_r(d^2 x'^2_{k+1}), \quad (147)$$

$$\mathbb{E}_b \frac{1}{2^{b+3}} \mathbb{E}_\theta S_1(\theta) = O_r(d^2 x'^2_{k+1}), \quad (148)$$

$$\mathbb{E}_b \frac{1}{2^{b+3}} \mathbb{E}_\theta |E_2(\theta)| = O_r(d^2 x'^2_{k+1}). \quad (149)$$

Finally,

$$y_{k+1} = \mathbb{E}_b \mathbb{E}_{\theta_1,\ldots,\theta_b} \mathbb{E}\theta_R^3 \quad (150)$$

$$= \mathbb{E}_b \mathbb{E}_{\theta_1,\ldots,\theta_b} \frac{1}{2^{b+3}} \left(S_0(\theta) - \frac{3}{2} S_1(\theta) + E_2(\theta) \right)$$

$$= dy'_{k+1} + O_r(d^2 x'^2_{k+1}).$$

This finishes the proof. ∎

Lemma 17 (Moment calculations) *Let P be a BMS channel and $\ell \in \mathbb{Z}_{\ge 0}$. Let $x = \mathbb{E}\theta_P$ and $z = \mathbb{E}\theta_P^3$. Let $\theta_1, \ldots, \theta_\ell$ be i.i.d. copies of θ_P, the θ-component of P. For $y \in \{\pm\}^\ell$, define*

$\alpha(\theta, y) = \prod_{i \in [\ell]} (1 + \theta_i y_i)$. *Then we have*

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha(\theta, y) = 2^\ell, \tag{151}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^2(\theta, y) = 2^\ell (1 + x)^\ell, \tag{152}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha(\theta, y) \alpha(\theta, -y) = 2^\ell (1 - x)^\ell, \tag{153}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^3(\theta, y) = 2^\ell (1 + 3x)^\ell, \tag{154}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^2(\theta, y) \alpha(\theta, -y) = 2^\ell (1 - x)^\ell, \tag{155}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^4(\theta, y) = 2^\ell (1 + 6x + z)^\ell, \tag{156}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^3(\theta, y) \alpha(\theta, -y) = 2^\ell (1 - z)^\ell, \tag{157}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^2(\theta, y) \alpha^2(\theta, -y) = 2^\ell (1 - 2x + z)^\ell, \tag{158}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^5(\theta, y) = 2^\ell (1 + 10x + 5z)^\ell, \tag{159}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^4(\theta, y) \alpha(\theta, -y) = 2^\ell (1 + 2x - 3z)^\ell, \tag{160}$$

$$\mathbb{E} \sum_{y \in \{\pm\}^\ell} \alpha^3(\theta, y) \alpha^2(\theta, -y) = 2^\ell (1 - 2x + z)^\ell. \tag{161}$$

Proof The proof is by direct computations.

First moment

$$\sum_{y \in \{\pm\}^\ell} \alpha(\theta, y) = \sum_{y \in \{\pm\}^\ell} \prod_{i \in [\ell]} (1 + \theta_i y_i) = 2^\ell. \tag{162}$$

Second moments

$$\sum_{y \in \{\pm\}^\ell} \alpha^2(\theta, y) = \sum_{y \in \{\pm\}^\ell} \prod_{i \in [\ell]} (1 + 2\theta_i y_i + \theta_i^2) = 2^\ell \prod_{i \in [\ell]} (1 + \theta_i^2), \tag{163}$$

$$\sum_{y \in \{\pm\}^\ell} \alpha(\theta, y) \alpha(\theta, -y) = \sum_{y \in \{\pm\}^\ell} \prod_{i \in [\ell]} (1 - \theta_i^2) = 2^\ell \prod_{i \in [\ell]} (1 - \theta_i^2). \tag{164}$$

27

Third moments

$$\sum_{y\in\{\pm\}^\ell} \alpha^3(\theta,y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+3\theta_i y_i+3\theta_i^2+\theta_i^3 y_i) = 2^\ell \prod_{i\in[\ell]}(1+3\theta_i^2), \qquad (165)$$

$$\sum_{y\in\{\pm\}^\ell} \alpha^2(\theta,y)\alpha(\theta,-y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+\theta_i y_i-\theta_i^2-\theta_i^3 y_i) = 2^\ell \prod_{i\in[\ell]}(1-\theta_i^2). \qquad (166)$$

Fourth moments

$$\sum_{y\in\{\pm\}^\ell} \alpha^4(\theta,y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+4\theta_i y_i+6\theta_i^2+4\theta_i^3 y_i+\theta_i^4) \qquad (167)$$
$$= 2^\ell \prod_{i\in[\ell]}(1+6\theta_i^2+\theta_i^4),$$

$$\sum_{y\in\{\pm\}^\ell} \alpha^3(\theta,y)\alpha(\theta,-y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+2\theta_i y_i-2\theta_i^3 y_i-\theta_i^4) = 2^\ell \prod_{i\in[\ell]}(1-\theta_i^4), \qquad (168)$$

$$\sum_{y\in\{\pm\}^\ell} \alpha^2(\theta,y)\alpha^2(\theta,-y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1-2\theta_i^2+\theta_i^4) = 2^\ell \prod_{i\in[\ell]}(1-2\theta_i^2+\theta_i^4). \qquad (169)$$

Fifth moments

$$\sum_{y\in\{\pm\}^\ell} \alpha^5(\theta,y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+5\theta_i y_i+10\theta_i^2+10\theta_i^3 y_i+5\theta_i^4+\theta_i^5) \qquad (170)$$
$$= 2^\ell \prod_{i\in[\ell]}(1+10\theta_i^2+5\theta_i^4),$$

$$\sum_{y\in\{\pm\}^\ell} \alpha^4(\theta,y)\alpha(\theta,-y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+3\theta_i y_i+2\theta_i^2-2\theta_i^3 y_i-3\theta_i^4-\theta_i^5) \qquad (171)$$
$$= 2^\ell \prod_{i\in[\ell]}(1+2\theta_i^2-3\theta_i^4),$$

$$\sum_{y\in\{\pm\}^\ell} \alpha^3(\theta,y)\alpha^2(\theta,-y) = \sum_{y\in\{\pm\}^\ell} \prod_{i\in[\ell]}(1+\theta_i y_i-2\theta_i^2-2\theta_i^3 y_i+\theta_i^4+\theta_i^5) \qquad (172)$$
$$= 2^\ell \prod_{i\in[\ell]}(1-2\theta_i^2+\theta_i^4).$$

∎

Lemma 18 *There exists $\epsilon > 0$ and $C > 0$ such that for all $d \geq 0$ and $\lambda \in [-1,1]$ satisfying $(r-1)d\lambda^2 \leq 1$, for all $x \leq \lambda^2\epsilon$ and $m \in \mathbb{Z}_{\geq 0}$, we have*

$$\left| \mathbb{E}_{b\sim D}(1+x)^b - \sum_{0\leq i\leq m} \mathbb{E}_{b\sim D}\binom{b}{i}x^i \right| \leq C(dx)^{m+1}. \qquad (173)$$

Here D denotes either the point distribution at d, or $\mathrm{Pois}(d)$.

Proof Follows from the proof of (Mossel et al., 2023, Lemma 3.7). ∎

Lemma 18 is the only place in the proof of Theorem 7 where we use that D is a point distribution or a Poisson distribution. We remark that (Mossel et al., 2023, Lemma 3.7) works for a large class of offspring distributions D (see (Mossel et al., 2023, Assumption 1.10 and 1.11)). Therefore Theorem 7 and Theorem 5 holds for more general offspring distributions by replacing Lemma 18 with a more general version.

Lemma 12, Eq. (41) and Eq. (42) directly follow from Eq. (37), Eq. (38), Eq. (39), Eq. (40). Thus we have finished the proof of Lemma 12.

Proof [Proof of Lemma 13] Because the statement is monotone in δ, WLOG assume that $C_{\chi^2}(P) = \delta$. By Lemma 12, there exists $C = C(r) > 0$ such that

$$\left| x_{k+1} - (r-1)d\lambda^2 x_k \right| \leq C(r-1)d\lambda^2 x_k^2, \tag{174}$$

$$\left| y_{k+1} - (r-1)d\lambda^4 y_k \right| \leq C x_k^2. \tag{175}$$

Take $k_0 = \Theta_r(1/\delta)$ such that $\frac{1}{2} \leq (1 - 2C\delta)^{k_0}$ and $(1 + 2C\delta)^{k_0} \leq 2$. Then for all $0 \leq k \leq k_0$, we have

$$\frac{1}{2}((r-1)d\lambda^2)^k\delta \leq x_k \leq 2((r-1)d\lambda^2)^k\delta. \tag{176}$$

On the other hand,

$$y_{k_0} \leq \sum_{0 \leq k \leq k_0 - 1} C((r-1)d\lambda^4)^{k_0-k-1}x_k^2 + ((r-1)d\lambda^4)^{k_0}y_0 \tag{177}$$

$$\leq 4C \sum_{0 \leq k \leq k_0 - 1} \lambda^{2(k_0-k-1)}((r-1)d\lambda^2)^{k_0+k-1}\delta^2 + \lambda^{2k_0}((r-1)d\lambda^2)^{k_0}\delta.$$

Because $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \frac{1}{5} \right]$ and $(r-1)d\lambda^2 \geq 0.99$, we have $|\lambda| \leq (r-1)d\lambda^2$. So

$$\lambda^{2(k_0-k-1)}((r-1)d\lambda^2)^{k_0+k-1} \leq |\lambda|^{k_0-k-1}((r-1)d\lambda^2)^{2k_0-2} \tag{178}$$

$$\leq 1.1|\lambda|^{k_0-k-1}((r-1)d\lambda^2)^{2k_0}.$$

Furthermore, because $k_0 = \Theta(1/\delta) = \omega(\log(1/\delta))$, we have $|\lambda|^{k_0} \leq \delta$ (for small enough δ). So

$$\lambda^{2k_0}((r-1)d\lambda^2)^{k_0}\delta \leq ((r-1)d\lambda^2)^{2k_0}\delta^2. \tag{179}$$

Combining the above, we get

$$y_{k_0} \leq \left(4.4C \sum_{0 \leq k \leq k_0 - 1} |\lambda|^{k_0-k-1} + 1 \right) \left(((r-1)d\lambda^2)^{k_0}\delta \right)^2 \tag{180}$$

$$\leq (9C + 1) \left(((r-1)d\lambda^2)^{k_0}\delta \right)^2.$$

This finishes the proof. ∎

A.4. Fine estimates

The goal of this section is to prove Lemma 14.

Proof [Proof of Lemma 14, Eq. (45)] Use the same notation as in the proof of Lemma 12, Eq. (37). Expanding Eq. (95) using

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \frac{x^4}{1+x},\tag{181}$$

we get

$$C_{\chi^2}(R) = \frac{\lambda^2}{2^r}\left(T_0(\theta) - \frac{\lambda}{2}T_1(\theta) + \frac{\lambda^2}{4}T_2(\theta) - \frac{\lambda^3}{8}T_3(\theta) + \frac{\lambda^4}{16}R_4(\theta)\right),\tag{182}$$

where $T_0(\theta)$ is defined in Eq. (98), $T_1(\theta)$ is defined in Eq. (99), and the remaining are defined as

$$T_2(\theta) = \sum_{y\in\{\pm\}^{r-1}} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^2,\tag{183}$$

$$T_3(\theta) = \sum_{y\in\{\pm\}^{r-1}} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^3,\tag{184}$$

$$R_4(\theta) = \sum_{y\in\{\pm\}^{r-1}} \frac{\alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^4}{1 + \frac{\lambda}{2}(\alpha(\theta,y) + \alpha(\theta,-y) - 2)}.\tag{185}$$

Note that

$$\frac{\lambda}{2}|R_4(\theta)| \leq \sum_{y\in\{\pm\}^{r-1}} \alpha(\theta,y)(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^4\left|\frac{R_+(y) - R_-(y)}{R_+(y) + R_-(y)}\right|\tag{186}$$

$$\leq \sum_{y\in\{\pm\}^{r-1}} \alpha(\theta,y)(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^4.$$

By Lemma 17, we have

$$\mathbb{E}T_2(\theta) = 2^{r-1}\left((1 + 6x_k + y_k)^{r-1} - 4(1 + 3x_k)^{r-1} - (1 - 2x_k + y_k)^{r-1} + 4(1 + x_k)^{r-1}\right),\tag{187}$$

$$\mathbb{E}T_3(\theta) = 2^{r-1}\left((1 + 10x_k + 5y_k)^{r-1} + (1 + 2x_k - 3y_k)^{r-1} - 6(1 + 6x_k + y_k)^{r-1}\right.\tag{188}$$
$$\left.+ 12(1 + 3x_k)^{r-1} + 4(1 - 2x_k + y_k)^{r-1} - 8(1 + x_k)^{r-1} - 4(1 - x_k)^{r-1}\right),$$

$$\frac{\lambda}{2}\mathbb{E}|R_4(\theta)| \leq 2^{r-1}\left((1 + 10x_k + 5y_k)^{r-1} + 5(1 + 2x_k - 3y_k)^{r-1} - 8(1 + 6x_k + y_k)^{r-1}\right.\tag{189}$$
$$+ 10(1 - 2x_k + y_k)^{r-1} - 32(1 - y_k)^{r-1} + 24(1 + 3x_k)^{r-1} - 24(1 - 2x_k + y_k)^{r-1}$$
$$\left.+ 40(1 - x_k)^{r-1} - 32(1 + x_k)^{r-1} + 16\right).$$

Expanding with respect to x_k and using $y_k \leq \gamma x_k^2$, we get

$$\mathbb{E}T_0(\theta) = 2^r(r-1)x_k + O_r(x_k^3),\tag{190}$$

$$\mathbb{E}T_1(\theta) = 2^{r+1}(r-1)(r-2)x_k^2 + O_r(x_k^3),\tag{191}$$

$$\mathbb{E}T_2(\theta) = O_r(x_k^3),\tag{192}$$

$$\mathbb{E}T_3(\theta) = O_r(x_k^3),\tag{193}$$

$$\frac{\lambda}{2}\mathbb{E}|R_4(\theta)| = O_r(x_k^3).\tag{194}$$

Finally,

$$
\begin{aligned}
x'_{k+1} &= \mathbb{E}_{\theta_1,\ldots,\theta_{r-1}} C_{\chi^2}((B_{\theta_1} \times \cdots \times B_{\theta_{r-1}}) \circ B_{r,\lambda}) \\
&= \mathbb{E}_{\theta_1,\ldots,\theta_{r-1}} \frac{\lambda^2}{2r} \left(T_0(\theta) - \frac{\lambda}{2}T_1(\theta) + \frac{\lambda^2}{4}T_2(\theta) - \frac{\lambda^3}{8}T_3(\theta) + \frac{\lambda^4}{16}R_4(\theta) \right), \\
&= (r-1)\lambda^2 x_k - (r-1)(r-2)\lambda^3 x_k^2 + O_r(\lambda^2 x_k^3).
\end{aligned} \tag{195}
$$

This finishes the proof. ∎

Proof [Proof of Lemma 14, Eq. (46)] Use the same notation as in the proof of Lemma 12, Eq. (39). Expanding Eq. (127) using Eq. (181), we get

$$
C_{\chi^2}(R) = \frac{1}{2^{b+1}} \left(T_0(\theta) - \frac{1}{2}T_1(\theta) + \frac{1}{4}T_2(\theta) - \frac{1}{8}T_3(\theta) + \frac{1}{16}R_4(\theta) \right), \tag{196}
$$

where $T_0(\theta)$ is defined in Eq. (129), $T_1(\theta)$ is defined in Eq. (130), and the remaining are defined as

$$
T_2(\theta) = \sum_{y \in \{\pm\}^b} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^2, \tag{197}
$$

$$
T_3(\theta) = \sum_{y \in \{\pm\}^b} \alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^3, \tag{198}
$$

$$
R_4(\theta) = \sum_{y \in \{\pm\}^b} \frac{\alpha(\theta,y)(\alpha(\theta,y) - \alpha(\theta,-y))(\alpha(\theta,y) + \alpha(\theta,-y) - 2)^4}{1 + \frac{1}{2}(\alpha(\theta,y) + \alpha(\theta,-y) - 2)}. \tag{199}
$$

Using the same computation as the proof of Lemma 14, Eq. (45), we get

$$
\mathbb{E}_\theta T_2(\theta) = 2^b \left((1 + 6x'_{k+1} + y'_{k+1})^b - 4(1 + 3x'_{k+1})^b - (1 - 2x'_{k+1} + y'_{k+1})^b \right. \tag{200}
$$

$$
\left. + 4(1 + x'_{k+1})^b \right), \tag{201}
$$

$$
\mathbb{E}_\theta T_3(\theta) = 2^b \left((1 + 10x'_{k+1} + 5y'_{k+1})^b + (1 + 2x'_{k+1} - 3y'_{k+1})^b - 6(1 + 6x'_{k+1} + y'_{k+1})^b \right. \tag{202}
$$

$$
\left. + 12(1 + 3x'_{k+1})^b + 4(1 - 2x'_{k+1} + y'_{k+1})^b - 8(1 + x'_{k+1})^b - 4(1 - x'_{k+1})^b \right),
$$

$$
\frac{1}{2}\mathbb{E}_\theta|R_4(\theta)| \leq 2^b \left((1 + 10x'_{k+1} + 5y'_{k+1})^b + 5(1 + 2x'_{k+1} - 3y'_{k+1})^b - 8(1 + 6x'_{k+1} + y'_{k+1})^b \right. \tag{203}
$$

$$
+ 10(1 - 2x'_{k+1} + y'_{k+1})^b - 32(1 - y'_{k+1})^b + 24(1 + 3x'_{k+1})^b
$$

$$
\left. - 24(1 - 2x'_{k+1} + y'_{k+1})^b + 40(1 - x'_{k+1})^b - 32(1 + x'_{k+1})^b + 16 \right).
$$

Because $y_k \leq \gamma x_k^2$, using Eq. (37) and Eq. (38) we see that $y'_{k+1} \leq \gamma x'^2_{k+1}$. Using Lemma 18, we get

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_0(\theta) = dx'_{k+1} + O_r(d^3 x'^3_{k+1}), \tag{204}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_1(\theta) = 2 \left(\mathbb{E}_{b \sim D} b(b-1)\right) x'^2_{k+1} + O_r(d^3 x'^3_{k+1}), \tag{205}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_2(\theta) = O_r(d^3 x'^3_{k+1}), \tag{206}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta T_3(\theta) = O_r(d^3 x'^3_{k+1}), \tag{207}$$

$$\mathbb{E}_b \frac{1}{2^{b+1}} \mathbb{E}_\theta |R_4(\theta)| = O_r(d^3 x'^3_{k+1}). \tag{208}$$

Finally,

$$x_{k+1} = \mathbb{E}_{\theta_1,\dots,\theta_b} C_{\chi^2}(B_{\theta_1} \star \cdots \star B_{\theta_b}) \tag{209}$$

$$= \mathbb{E}_{\theta_1,\dots,\theta_b} \frac{1}{2^{b+1}} \left(T_0(\theta) - \frac{1}{2}T_1(\theta) + \frac{1}{4}T_2(\theta) - \frac{1}{8}T_3(\theta) + \frac{1}{16}R_4(\theta)\right),$$

$$= dx'_{k+1} - \left(\mathbb{E}_{b \sim D} b(b-1)\right) x'^2_{k+1} + O_r(d^3 x'^3_{k+1}).$$

This finishes the proof. ∎

Proof [Proof of Lemma 14] Eq. (47) follows directly from Eq. (45) and Eq. (46). It remains to prove that $M_{P,k+1}$ is γ-normal.

Recall our assumption that $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \frac{1}{5}\right]$ and $(r-1)d\lambda^2 \geq 0.99$. By Eq. (41) and Eq. (42), there exists $C = C(r) > 0$ such that

$$|x_{k+1} - (r-1)d\lambda^2 x_k| \leq C(r-1)d\lambda^2 x_k^2, \tag{210}$$

$$|y_{k+1} - (r-1)d\lambda^4 y_k| \leq C x_k^2. \tag{211}$$

If x_k is small enough, then

$$x_{k+1} \geq ((r-1)d\lambda^2 - 0.01)x_k \geq 0.98 x_k. \tag{212}$$

On the other hand,

$$\begin{aligned} y_{k+1} &\leq (r-1)d\lambda^4 y_k + C x_k^2 \tag{213} \\ &\leq (r-1)d\lambda^4 \gamma x_k^2 + C x_k^2 \\ &\leq \lambda^2 \gamma x_k^2 + C x_k^2 \\ &\leq 0.5\gamma x_k^2 + C x_k^2, \end{aligned}$$

where the last step uses the assumption that $|\lambda| \leq \max\left\{\frac{1}{2^{r-1}-1}, \frac{1}{5}\right\}$. Note that C does not depend on the choice of γ. Therefore we can take γ large enough such that $0.5\gamma + C \leq \gamma \cdot 0.98^2$. In this way, $y_{k+1} \leq \gamma x_{k+1}^2$. ∎

Appendix B. Proofs of remaining results

In this section we prove our remaining main results, Corollary 3 (non-tightness of the KS threshold), Theorem 8 (asymptotic BOHT reconstruction threshold) and Theorem 4 (asymptotic HSBM weak recovery bounds).

B.1. Non-tightness of the KS threshold

Proof [Proof of Corollary 3] For $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$, define

$$f_r(\lambda) = d_{\mathrm{KL}} \left(\lambda + \frac{1-\lambda}{2^{r-1}} \, \middle\| \, \frac{1}{2^{r-1}} \right). \tag{214}$$

By Theorem 2, if for some (r, λ) we have

$$f_r(\lambda) > r(r-1)\lambda^2 \log 2, \tag{215}$$

then for this (r, λ) the KS threshold is not tight for HSBM weak recovery.

For $r = 5, 6$, we take λ_0 as in Table 1 and verify that Eq. (215) holds for all $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \lambda_0 \right)$. Therefore the KS threshold is not tight for such (r, λ).

It remains to prove that for all $r \geq 7$, there exists $\lambda_0(r) > 0$ such that Eq. (215) holds for all $\lambda \in \left[-\frac{1}{2^{r-1}-1}, \lambda_0 \right)$. For $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 0 \right)$, we apply Lemma 19 with $x = \lambda + \frac{1-\lambda}{2^{r-1}}$, $y = \frac{1}{2^{r-1}}$ and get

$$f_r(\lambda) \geq \frac{2^{r-1}-1}{2}\lambda^2 > r(r-1)\lambda^2 \log 2 \tag{216}$$

where the last step holds for all $r \geq 7$.

For $\lambda > 0$, we observe that

$$\lim_{\lambda \to 0} \frac{1}{\lambda^2} f_r(\lambda) = f_r''(0) = \frac{2^{r-1}-1}{2}. \tag{217}$$

So for $r \geq 7$, there exists $\lambda_0 = \lambda_0(r) > 0$ such that Eq. (215) holds for all $\lambda \in (0, \lambda_0)$. Combining both cases we finish the proof. ∎

Lemma 19 *For $0 \leq x \leq y \leq \frac{1}{2}$, we have*

$$d_{\mathrm{KL}}(x\|y) \geq \frac{(x-y)^2}{2y(1-y)}. \tag{218}$$

Proof We first consider the case $x > 0$. Let $t = y - x$. Then

$$d_{\mathrm{KL}}(x\|y) = x \log\left(1 - \frac{t}{y}\right) + (1 - x)\log\left(1 + \frac{t}{1-y}\right) \tag{219}$$

$$= (-y + t)\sum_{k\geq 1}\frac{1}{k}\left(\frac{t}{y}\right)^k + (1 - y + t)\sum_{k\geq 1}(-1)^{k-1}\frac{1}{k}\left(\frac{t}{1-y}\right)^k$$

$$= \left(-t + \sum_{k\geq 2}\frac{t^k y^{-k+1}}{k(k-1)}\right) + \left(t + \sum_{k\geq 2}(-1)^k\frac{t^k(1-y)^{-k+1}}{k(k-1)}\right)$$

$$= \sum_{k\geq 2}\frac{t^k}{k(k-1)}\left(y^{-k+1} + (-1)^k(1-y)^{-k+1}\right).$$

Because $y \leq \frac{1}{2}$, the summand is non-negative for all $k \geq 2$. Taking $k = 2$, we get

$$d_{\mathrm{KL}}(x\|y) \geq \frac{1}{2}t^2\left(y^{-1} + (1-y)^{-1}\right) = \frac{(y-x)^2}{2y(1-y)}. \tag{220}$$

The case $x = 0$ follows by continuity. ∎

B.2. Asymptotic BOHT reconstruction threshold

Proof [Proof of Theorem 8] **Upper bound.** Let $w = w^* + \epsilon$. By Lemma 15, it suffices to prove for $(r-1)d\lambda^2 = w$. By Lemma 10(i) and definition of w^*, there exists $x_0 \in (0, 1]$ such that $g_{r,w}(x_0) > x_0$. Let $\epsilon = g_{r,w}(x_0) - x_0$. Take $d_0 = d_0(r, \epsilon)$ in Prop. 9. By Prop. 9, for $d \geq d_0$ and $(r-1)d\lambda^2 = w$, for any BMS channel P with $C_{\chi^2}(P) \geq x_0$, we have

$$C_{\chi^2}(\mathrm{BP}(P)) \geq g_{r,w}(C_{\chi^2}(P)) - \epsilon \geq g_{r,w}(x_0) - \epsilon = x_0, \tag{221}$$

where the second step is by Lemma 10(i). Therefore $\lim_{k\to\infty} C_{\chi^2}(M_k) \geq x_0 > 0$ and reconstruction is possible.

Lower bound. Let $w = w^* - \epsilon$. By Lemma 10(i) and definition of w^*, we have $g_{r,w}(x) < x$ for all $x \in (0, 1]$. By the same proof as Prop. 6, for any $\epsilon' > 0$ there exists $d_0 = d_0(r, \epsilon')$ such that $\lim_{k\to\infty} C_{\chi^2}(M_k) < \epsilon'$. The rest of the proof is the same as the proof of Theorem 5. ∎

B.3. Asymptotic HSBM weak recovery bounds

Proof [Proof of Theorem 4(ii)] We prove that $w^*(r) \geq \frac{1}{2^{r-2}}$, where $w^*(r)$ is defined in Eq. (21). That is, for $w = \frac{1}{2^{r-2}}$, $g_{r,w}(x) \leq x$ for all $0 \leq x \leq 1$. By Lemma 10(ii), $g_{r,w}(x) \leq s_{r,w}(x)$. Therefore it suffices to prove that $s_{r,w}(x) \leq x$. Note That

$$\frac{s_{r,w}(x)}{x} = w \cdot \sum_{0 \leq i \leq \left\lfloor\frac{r-1}{2}\right\rfloor}\binom{r-1}{2i+1}x^{2i} \tag{222}$$

is a polynomial with non-negative coefficients. Therefore for $0 \leq x \leq 1$, $\frac{s_{r,w}(x)}{x}$ is maximized at $x = 1$, taking value $w2^{r-2} = 1$. So $w^*(r) \geq w$. Applying Theorem 8 and (Gu, 2023, Theorem 5.15) we finish the proof. ∎

Proof [Proof of Theorem 4(i)] Recall function f_r defined in Eq. (214). By Eq. (217), for any $\delta > 0$, there exists $\lambda_0 = \lambda_0(r, \delta) > 0$ such that for all $|\lambda| \leq \lambda_0$, we have $f_r(\lambda) > \left(\frac{2^{r-1}-1}{2} - \delta \right) \lambda^2$. Furthermore, because $f_r(\lambda)$ is monotone increasing for $\lambda > 0$ and monotone decreasing for $\lambda < 0$, we can take λ_0 such that $f_r(\lambda) > \left(\frac{2^{r-1}-1}{2} - \delta \right) \min\{\lambda^2, \lambda_0^2\}$ for all $\lambda \in \left[-\frac{1}{2^{r-1}-1}, 1 \right]$.

Now fix $\epsilon > 0$. Take $\delta > 0$ such that $\left(\frac{2r \log 2}{2^{r-1}-1} + \epsilon \right) \cdot r^{-1} \left(\frac{2^{r-1}-1}{2} - \delta \right) = \log 2$. Take $\lambda_0 = \lambda_0(r, \delta)$. Take $d_0 = \lambda_0^{-2} \left(\frac{2r \log 2}{2^{r-1}-1} + \epsilon \right)$. Then for any $d \geq d_0$ and $d\lambda^2 \geq \frac{2r \log 2}{2^{r-1}-1} + \epsilon$, we have

$$\frac{d}{r} \cdot f_r(\lambda) > \frac{d}{r} \cdot \left(\frac{2^{r-1}-1}{2} - \delta \right) \min\{\lambda^2, \lambda_0^2\} = \frac{d \min\{\lambda^2, \lambda_0^2\} \log 2}{\frac{2r \log 2}{2^{r-1}-1} + \epsilon} \geq \log 2. \qquad (223)$$

By Theorem 2, we finish the proof. ∎

Appendix C. Discussions

C.1. Comparison with the q-community SBM

The HSBM is a generalization of the two-community SBM to hypergraphs. Another natural generalization is the q-community SBM (q-SBM). In this model, n vertices are given i.i.d. $\mathrm{Unif}([q])$ labels X_v. An edge (u, v) is added with probability $\frac{a}{n}$ if $X_u = X_v$, and with probability $\frac{b}{n}$ if $X_u \neq X_v$. The goal of the weak recovery problem is to recover a non-trivial fraction of the labels given the generated graph.

For the q-SBM, define

$$d = \frac{a + (q-1)b}{q}, \qquad \lambda = \frac{a - b}{a + (q-1)b}. \qquad (224)$$

The Kesten-Stigum threshold is at $d\lambda^2 = 1$. It plays a similar role in the q-SBM as in the HSBM: weak recovery is always possible above the KS threshold, and this is sometimes, but not always, tight.

Our current knowledge about tightness of the KS threshold for q-SBM weak recovery is summarized in Table 3. Comparing Table 2 and Table 3, we make a few interesting observations. In both models, in the large d regime, as r and q goes larger, the KS threshold goes from tight to non-tight. However, the HSBM has a small d regime ($\lambda \geq \frac{1}{5}$) where KS is tight for any r, but the q-SBM is not known to have such a regime.

For the 4-SBM, it is conjectured that tightness of the KS threshold is different for $\lambda > 0$ and $\lambda < 0$ cases. It is known that for $\lambda < 0$ and small d, KS is not tight, and conjectured that for $\lambda > 0$, KS is always tight. We expect similar behavior to hold for the HSBM with $r = 5, 6$.

Unlike the HSBM case, we know more about tightness of the KS threshold for q-ary Potts model on a tree than the q-SBM. For $q \geq 5$, it is known that the KS threshold is never tight (Sly (2011); Mossel et al. (2023)) for the Potts model. However, as shown in Table 3, there are several regimes with $q \geq 5$ where the KS threshold is not known to be not tight for the q-SBM.

	any d	small d	large d
$q = 3, \lambda > 0$	unknown	unknown	tight
$q = 3, \lambda < 0$	unknown	unknown	tight
$q = 4, \lambda > 0$	unknown	unknown	tight
$q = 4, \lambda < 0$	unknown	not tight	tight
$5 \le q \le 10, \lambda > 0$	unknown	unknown	unknown
$5 \le q \le 10, \lambda < 0$	unknown	not tight	unknown
$q \ge 11, \lambda > 0$	unknown	unknown	not tight
$q \ge 11, \lambda < 0$	never tight	(not tight)	(not tight)

Table 3: Tightness of the KS threshold for symmetric q-SBM weak recovery. An entry in parentheses means it is implied by another entry in the same row. References: Possibility above KS: Abbe and Sandon (2015). Tightness results: Mossel et al. (2023). Non-tightness results: Abbe and Sandon (2015) (all non-tightness results) and Banks et al. (2016) (for $q \ge 5$).

C.2. Methods for reconstruction on trees and hypertrees

As noted in Mossel (2004), there are a number of different methods for establishing the reconstruction threshold for the Ising model, but few, if any, generalize to other BOT problems. After that, new methods are developed for proving bounds on the reconstruction threshold.

For the upper bound (reconstruction), the Kesten-Stigum bound is simple and works universally. When it is not tight, upper bounds are proved by establishing non-contraction properties. Suppose for a certain non-trivial channel P, we have $\mathrm{BP}(P) \ge P$, where $\ge$ is a certain channel preorder that respects the BP operator (which is often the degradation preorder), then by iterating we get $\mathrm{BP}^k(\mathrm{Id}) \ge P$ for all $k \ge 0$, and reconstruction is possible. Sly (2009) used freezing probability as the information measure to prove reconstruction results for the random coloring model, which can be viewed as taking $\ge$ to be the degradation preorder and P to be an erasure channel. Sly (2011); Mossel et al. (2023) used χ^2-capacity (also called magnetization) to prove reconstruction results for the Potts model in the large degree regime (via the large degree asymptotics method). This was also used in Liu and Ning (2019) for the asymmetric Ising model and in Gu and Polyanskiy (2023b) for the BOHT model. The current work uses this method in Theorem 8. Sometimes the channel P is designed explicitly, such as Sly and Zhang (2016); Zhang (2017) which established lower order terms of the reconstruction threshold for the random coloring model and the NAE-k-SAT model (in other words, the BOHT model with $\lambda = -\frac{1}{2^{r-1}-1}$).

For the lower bound (non-reconstruction), there are two particularly successful methods. One method proves contraction of certain information measures using strong data processing inequalities (SDPIs) and subadditivity. Borgs et al. (2006) proved tightness of the KS threshold for roughly symmetric Ising models using χ^2-information. Külske and Formentin (2009) proved non-reconstruction results for Potts model using this method with symmetric KL (SKL) information, and Gu and Polyanskiy (2023a) obtained improved results using mutual information. Gu and Polyanskiy (2023b) proved non-reconstruction results for the BOHT model using a multi-terminal generalization of the method with SKL information and χ^2-information. The advantage of this method is that it is often simple to obtain, works for any degree, and can be easily generalized to different tree structures

(non-Galton-Watson trees). However, the choice of the information measure is mysterious and it is unclear what is the best information measure to use.

The other method is Sly's method, which uses large degree asymptotics and robust reconstruction. This method was introduced in Sly (2009, 2011) for the random coloring model and the Potts model. Liu and Ning (2019) used the method for the asymmetric Ising model. Mossel et al. (2023) improved this method by analyzing more orders in the Taylor expansion, and removing the need for a certain concentration result. The current work uses a hypertree verion of Sly's method in Theorem 5. The advantage of this method is that it often gives tight results. The disadvantage is that it needs a large degree assumption and requires heavy computation.

C.3. Further directions

Tightness of the KS threshold The obvious open question is to fill in the blanks of Table 2 and Table 3. It could be possible that using a clever information measure in the SDPI method gives some tight non-reconstruction results. However, it is unclear what could be such an information measure, other than mutual information, χ^2-information, and SKL information.

For $r = 5, 6$ and $\lambda > 0$, Table 2 shows that the KS threshold is tight for small d and large d. If there is a certain kind of convexity property, then we could conclude that KS is tight for all d.

For $r \geq 7$ and $\lambda > 0$, Table 2 shows that the KS threshold is tight for small d and not tight for large d. Therefore, it seems possible that there exists $d^* = d^*(r)$ for which KS is tight for $d < d^*$ and not tight for $d > d^*$.

Asymptotic HSBM weak recovery bounds Theorem 4 gives asymptotic bounds on the weak recovery threshold, where there is a factor $\Theta(r)$ gap. In Banks et al. (2016), similar asymptotic bounds were obtained for the q-SBM, and the gap there is of a factor of two. Reducing the gaps in the asymptotic bounds for HSBM is an interesting question. Evidences suggest that the upper bound (Theorem 4(i)) might be closer to the true threshold as $r \to \infty$. The lower bound (Theorem 4(i)) could be not tight for two reasons. One is that the HSBM weak recovery threshold may be different from the BOHT reconstruction threhsold; the other is that the constant $\frac{1}{2^{r-2}}$ is not an asymptotically tight bound of $w^*(r)$ in Eq. (21). We choose $\frac{1}{2^{r-2}}$ so that the statement of Theorem 4(i) works for any $r \geq 3$ rather than as $r \to \infty$.

Non-uniform HSBM An interesting direction of generalization is the non-uniform HSBMs, where hyperedges may have different sizes. This model has a sequence of non-negative real parameters $(a_r, b_r)_{r \geq 2}$ with finitely many non-zero entries (not depending on n, the number of vertices). First, the vertices are given i.i.d. Unif($\{\pm\}$) labels. Then each set S of vertices with $|S| \geq 2$ is added as an hyperedge with probability $\frac{a_{|S|}}{\binom{n}{|S|-1}}$ or $\frac{b_{|S|}}{\binom{n}{|S|-1}}$ depending on whether the set S is monochromatic. The KS threshold is defined as $\sum_{r \geq 2}(r-1)d_r\lambda_r^2 = 1$ where d_r and λ_r are defined from a_r and b_r as in Eq. (1). Chodrow et al. (2023) conjectured that weak recovery is possible above the KS threshold, and partial progress was made in Dumitriu et al. (2021). By analyzing the corresponding non-uniform BOHT model, one could show that weak recovery is impossible below the KS threshold under certain conditions. For example, by adapting the multi-terminal SDPI method from Gu and Polyanskiy (2023b), one can show that if $a_r \geq b_r$ for $r \leq 4$ and $a_r = b_r = 0$ for $r \geq 5$, then weak recovery is impossible below the KS threshold. Furthermore, our robust reconstruction result (Theorem 7) seems to still hold in the non-uniform case. On the other hand, by adapting our proof

of Theorem 2 one could prove that weak recovery is possible whenever

$$\sum_{r \geq 2} \frac{d_r}{r} \cdot d_{\mathrm{KL}}\left(\lambda_r + \frac{1 - \lambda_r}{2^{r-1}} \,\middle\|\, \frac{1}{2^{r-1}}\right) > \log 2. \tag{225}$$