Uniqueness of BP fixed point for the Potts model and applications to community detection

Yuzhou Gu YUZHOUGU@MIT.EDU
Masaschusetts Institute of Technology

Yury Polyanskiy YP@MIT.EDU
Masaschusetts Institute of Technology

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

In the study of sparse stochastic block models (SBMs) one often needs to analyze a distributional recursion, known as the belief propagation (BP) recursion. Uniqueness of the fixed point of this recursion implies several results about the SBM, including optimal recovery algorithms for SBM Mossel et al. (2016) and SBM with side information Mossel and Xu (2016), and a formula for SBM mutual information Abbe et al. (2021). The 2-community case corresponds to an Ising model, for which Yu and Polyanskiy (2022) established uniqueness for all cases.

In this paper we analyze the q-ary Potts model, i.e., broadcasting of q-ary spins on a Galton-Watson tree with expected offspring degree d through Potts channels with second-largest eigenvalue λ. We allow the intermediate vertices to be observed through noisy channels (side information). We prove that BP uniqueness holds with and without side information when $d\lambda^2 \ge 1 + C \max\{\lambda, q^{-1}\} \log q$ for some absolute constant $C > 0$ independent of q, λ, d. For large q and $\lambda = o(1/\log q)$, this is asymptotically achieving the Kesten-Stigum threshold $d\lambda^2 = 1$. These results imply mutual information formulas and optimal recovery algorithms for the q-community SBM in the corresponding ranges.

For $q \ge 4$, Sly (2011); Mossel et al. (2022) showed that there exist choices of q, λ, d below Kesten-Stigum (i.e. $d\lambda^2 < 1$) but reconstruction is possible. Somewhat surprisingly, we show that in such regimes BP uniqueness does not hold at least in the presence of weak side information.

Our technical tool is a theory of q-ary symmetric channels, that we initiate here, generalizing the classical and widely-utilized information-theoretic characterization of BMS (binary memoryless symmetric) channels.

Keywords: stochastic block model, Potts model, broadcasting on trees, cavity equation, belief propagation, boundary irrelevance, q-ary symmetric channels

1. Introduction

Stochastic block model. The stochastic block model (SBM) is a random graph model with community structures. It has a rich set of results and phenomena, investigated in the last decade (see Abbe (2017) for a survey). In this paper, we focus on the sparse symmetric multi-community case. The model has four parameters: $n \in \mathbb{Z}_{\ge 1}$, the number of vertices; $q \in \mathbb{Z}_{\ge 2}$, the number of communities; $a, b \in \mathbb{R}_{\ge 0}$, parameters controlling edge probabilities. The model is defined as follows. First, we assign a random label (community) $X_i \sim \mathrm{Unif}([q])$ i.i.d for $i \in V = [n]$. Then a random graph $G = (V, E)$ is generated, where $(i, j) \in E$ with probability $\frac{a}{n}$ if $X_i = X_j$, and with probability $\frac{b}{n}$ if $X_i \ne X_j$, independently for all $(i, j) \in \binom{V}{2}$. When $a > b$, we say the model is assortative. When $a < b$, we say the model is disassortative.

For the SBM, an important problem is weak recovery. We say the model admits weak recovery if there exists an estimator $\widehat{X}(G) \in [q]^V$ such that

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E} d_H(X, \widehat{X}(G)) < 1 - \frac{1}{q}, \tag{1}$$

$$\text{where} \quad d_H(X,Y) := \min_{\tau\in \mathrm{Aut}([q])} \sum_{i\in[n]} \mathbb{1}\{X_i \ne \tau(Y_i)\}. \tag{2}$$

Decelle et al. (2011) conjectured that the (algorithmic) weak recovery threshold is at the Kesten-Stigum threshold, and there is an information-computation gap for $q \ge 5$. The positive (algorithmic) part of their conjecture has been established by a series of works Massoulié (2014); Mossel et al. (2018); Abbe and Sandon (2016, 2015a,b, 2018); Bordenave et al. (2015); Stephan and Massoulié (2019), in a very general sense allowing asymmetric communities. Abbe and Sandon (2015a,b, 2018) gave inefficient reconstruction algorithms below the Kesten-Stigum threshold, for disassortative models with $q \ge 4$ and assortative models with $q \ge 5$, giving evidence for the information-computation gap. The informational weak recovery threshold has been established in special cases by Mossel et al. (2015, 2018) ($q = 2$), Mossel et al. (2022) ($q = 3, 4$ assuming large enough degree), Coja-Oghlan et al. (2017) (general q, disassortative). For the assortative case, the informational weak recovery threshold for $q = 3, 4$ with small degree or $q \ge 5$ is still open, despite some partial progress Banks et al. (2016); Gu and Polyanskiy (2020). The information-computation gap is also wide open.

When weak recovery is possible, the natural follow-up question is to determine the optimal recovery accuracy

$$\sup_{\widehat{X}=\widehat{X}(G)} \lim_{n\to\infty} \mathbb{E}\left[1 - \frac{1}{n} d_H(X, \widehat{X}(G))\right]. \tag{3}$$

Decelle et al. (2011) conjectured that the belief propagation algorithm is optimal. The conjecture is not proved yet but significant progress has been made: a series of works Mossel et al. (2016); Abbe et al. (2021); Yu and Polyanskiy (2022) established an optimal recovery algorithm for the symmetric two-community SBM; Chin and Sly (2020, 2021) gave optimal recovery algorithms for the general case (general q, not necessarily symmetric) under the condition that SNR is large enough.

A fundamental quantity of the stochastic block model is its (normalized) mutual information

$$\lim_{n\to\infty} \frac{1}{n} I(X; G). \tag{4}$$

Coja-Oghlan et al. (2017) proved a mutual information formula for the disassortative model (general q). Abbe et al. (2021); Yu and Polyanskiy (2022) proved a mutual information formula for the $q = 2$ case. Dominguez and Mourrat (2022) conjectured another formula for the $q = 2$ case and proved a matching lower bound. It is not known whether their conjectured formula is equivalent to the one proved by Abbe et al. (2021); Yu and Polyanskiy (2022).

Broadcasting on trees. The stochastic block model has a close relationship with the broadcasting on trees (BOT) model. The reason is that in SBM, the local neighborhood of a random vertex converges (in the sense of local weak convergence) to a Galton-Watson tree with Poisson offspring distribution. Therefore, properties of BOT can often imply corresponding results on SBM.

For the symmetric q-SBM, the corresponding model is the Potts model, i.e., the BOT model whose broadcasting channels are Potts channels. This model has three parameters: $q \in \mathbb{Z}_{\geq 2}$, the number of colors; $\lambda \in [-\frac{1}{q-1}, 1]$, edge correlation strength; $d \in \mathbb{Z}_{\geq 0}$, expected offspring. The Potts model is defined as follows. Let T be a regular tree of offspring d or a Galton-Watson tree with offspring distribution $\mathrm{Pois}(d)$ (Poisson distribution with expectation d). Let ρ be the root of T. We assign to every vertex v a color $\sigma_v \in [q]$ according to the following process: (1) $\sigma_\rho \sim \mathrm{Unif}([q])$; (2) given σ_u, colors of children of u are $\stackrel{\text{i.i.d.}}{\sim} P_\lambda(\cdot|\sigma_u)$, where P_λ is the Potts channel defined as $P_\lambda(j|i) = \lambda \mathbb{1}\{i = j\} + \frac{1-\lambda}{q}$.

An important problem on BOT is the reconstruction problem, asking whether we can gain any non-trivial information about the root given observation of far away vertices. We say the model admits reconstruction if

$$\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}|T_k) > 0, \tag{5}$$

where L_k stands for the set of vertices at distance k to the root ρ. We say the model admits non-reconstruction if the limit is zero. It is known that non-reconstruction results for the Potts model imply impossibility of weak recovery for the corresponding SBM, but the other direction does not hold: in the case $a = 0$, there is a gap of factor 2 (as $q \to \infty$) between the BOT reconstruction threshold and the SBM weak recovery threshold.

The reconstruction problem on trees has been studied a lot under many different settings, e.g., Bleher et al. (1995); Evans et al. (2000); Mossel (2001); Mossel and Peres (2003); Mézard and Montanari (2006); Borgs et al. (2006); Bhatnagar et al. (2010); Sly (2009); Külske and Formentin (2009); Liu and Ning (2019); Gu and Polyanskiy (2020); Mossel et al. (2022). For the Potts model, it is known Sly (2011); Mossel et al. (2022) that the Kesten-Stigum threshold Kesten and Stigum (1966) $d\lambda^2 = 1$ is tight (i.e., equal to the reconstruction threshold) for $q = 3, 4$ when d is large enough, and is not tight when $q \geq 5$ or when $q = 4$, $\lambda < 0$ and d is small enough.

Belief propagation. Belief propagation is a powerful tool for studying the BOT model. It is usually described as an algorithm for computing posterior distribution of vertex colors given observation. Here we take an information-theoretic point of view and describe BP in terms of constructing communication channels.

We view the BOT model as an information channel from the root color to the observation. Let M_k denote the channel $\sigma_\rho \mapsto (T_k, \sigma_{L_k})$.Then $(M_k)_{k\in\mathbb{Z}_{\geq 0}}$ satisfies the following recursion, which we call belief propagation recursion:

$$M_{k+1} = \mathbb{E}_b (M_k \circ P_\lambda)^{\star b} \tag{6}$$

where b follows the branching number distribution (constant in the regular tree case, $\mathrm{Pois}(d)$ in the Poisson tree case), and $(\cdot)^{\star b}$ denotes $\star$-convolution power. Let BP be the operator

$$\mathrm{BP}(M) := \mathbb{E}_b (M \circ P_\lambda)^{\star b} \tag{7}$$

defined on the space of information channels with input alphabet $[q]$. Due to symmetry in colors, we can regard BP as an operator on the space of FMS channels (see Section 2). In terms of the BP operator, the reconstruction problem can be rephrased as whether the limit channel $\mathrm{BP}^\infty(\mathrm{Id}) := \lim_{n\to\infty} \mathrm{BP}^n(\mathrm{Id})$ (where Id stands for the identity channel $\mathrm{Id}(y|x) = \mathbb{1}\{x = y\}$) is trivial or not. The problem of optimal recovery for SBM can be reduced to the following problem on trees: whether the limit

$$\lim_{n\to\infty} I(\sigma_\rho; \omega_{L_k} | T_k) \tag{8}$$

where ω is the observation of σ through a non-trivial channel W, stays the same for any non-trivial FMS channel W. Therefore, it is important to study the non-trivial fixed points of the BP operator (the trivial channel is always a fixed point).

Mossel et al. (2016) proved uniqueness of BP fixed point for $q = 2$ and large enough SNR. Abbe et al. (2021) improved to $q = 2$ and SNR $\not\in [1, 3.513]$. Yu and Polyanskiy (2022) proved uniqueness of BP fixed point for $q = 2$ and any parameter λ, d, closing the question for binary symmetric models. For $q \ge 3$, Chin and Sly (2020) proved a local version of BP uniqueness, i.e., when the initial channel U is close enough to Id, and $d\lambda^2 > C_q$, where C_q is a constant depending on q, then $\mathrm{BP}^\infty(U) = \mathrm{BP}^\infty(\mathrm{Id})$. They did not give asymptotics for C_q, but it seems like it is at least polynomial in q. Chin and Sly (2021) generalized Chin and Sly (2020) to asymmetric models.

Boundary irrelevance. Abbe et al. (2021) reduced the SBM mutual information problem to the boundary irrelevance problem, on a tree model called the broadcasting on trees with survey (BOTS) model. In the BOTS model, we observe label of every vertex through a noisy FMS channel W (called the survey). We say the model admits boundary irrelevance with respect to W if

$$\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k} | T_k, \omega_{T_k}) = 0, \tag{9}$$

where T_k is the set of all vertices within distance at most k to the root, and ω is the observation of σ through W. We say the model admits boundary irrelevance if the model admits boundary irrelevance with respect to all erasure channels EC_ϵ with $0 \le \epsilon < 1$. Boundary irrelevance is equivalent to the condition that the operator

$$\mathrm{BP}_W(M) := \left(\mathbb{E}_b (M \circ P_\lambda)^{\star b}\right) \star W \tag{10}$$

has a unique fixed point in the space of FMS channels. Because BP and BP_W have very similar forms, the boundary irrelevance problem has a close relationship with the problem of uniqueness of BP fixed point. Indeed, these two problems can be solved using the same method.

Our results. Our first main result is uniqueness of BP fixed point and boundary irrelevance for a wide range of parameters. In fact, we prove the stronger result that the BP fixed point satisfies global stability.

Theorem 1 (Uniqueness of BP fixed point and boundary irrelevance) *There exists an absolute constant $C > 0$ such that the following statement holds. Consider the q-ary Potts model with broadcasting channel P_λ on a regular tree or a Poisson tree with expected offspring d. If either $d\lambda^2 < q^{-2}$ or $d\lambda^2 > 1 + C\max\{\lambda, q^{-1}\}\log q$, then boundary irrelevance holds. That is, for any non-trivial q-FMS survey channel W, we have*

$$\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k} | T_k, \omega_{T_k}) = 0. \tag{11}$$

Furthermore, under the same conditions, stability of BP fixed point holds, i.e., for any non-trivial q-FMS channel P, $\mathrm{BP}^k(P)$ and $\mathrm{BP}^k(\mathrm{Id})$ converge weakly to the same limit as $k \to \infty$.

For a more precise statement, see Theorem 14 and Theorem 15. We generalize this result to asymmetric initial channels in Prop. 38 and Prop. 40. Compared with Chin and Sly (2020), our result has a much weaker assumption on the initial channel, and works for a much larger region of (λ, d) (at least when $q \to \infty$).

Our second main result is that boundary irrelevance does not hold between the reconstruction threshold and the Kesten-Stigum threshold.

Theorem 2 (Boundary irrelevance does not always hold) *Consider the q-ary Potts model with broadcasting channel P_λ on a regular tree or a Poisson tree with expected offspring d. If $d\lambda^2 < 1$ and reconstruction is possible, then boundary irrelevance does not hold for weak enough survey channel. That is, there exists $\epsilon > 0$ such that for any FMS survey channel W with $C_{\chi^2}(W) \le \epsilon$, we have $\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k} | T_k, \omega_{T_k}) > 0$.*

By Sly (2011); Mossel et al. (2022), for any $q \ge 4$, there exist choices of λ, d satisfying the assumption in Theorem 2. Therefore for any fixed $q \ge 4$, boundary irrelevance does not always hold, in contrast to the binary case where boundary irrelevance always holds Yu and Polyanskiy (2022).

Applications. Main applications of uniqueness of BP fixed point and boundary irrelevance include a mutual information formula and an optimal recovery algorithm.

Theorem 3 (Mutual information formula) *Let $(X, G) \sim \mathrm{SBM}(n, q, \frac{a}{n}, \frac{b}{n})$. Let $d = \frac{a+(q-1)b}{q}$ and $\lambda = \frac{a-b}{a+(q-1)b}$. Let (T, σ) be the Potts model with broadcasting channel P_λ on a Poisson tree with expected offspring d. Let ρ be the root of T, L_k be the set of vertices at distance k to ρ, T_k be the set of vertices at distance $\le k$ to ρ. If (q, λ, d) satisfies (27) or (28), then we have*

$$\lim_{n\to\infty} \frac{1}{n} I(X; G) = \int_0^1 \lim_{k\to\infty} I(\sigma_\rho; \omega^\epsilon_{T_k \backslash \rho} | T_k) d\epsilon, \tag{12}$$

where ω^ϵ denotes observation through survey channel EC_ϵ, the erasure channel with erasing probability ϵ.

For SBM with side information, boundary irrelevance immediately implies that the local belief propagation algorithm is optimal.

Theorem 4 (Optimal recovery for SBM with side information) *Work under the same setting as Theorem 3. Suppose that in addition to G, we observe side information $Y_v \sim W(\cdot | X_v)$ for all $v \in V$, where W is some non-trivial FMS channel. If (q, λ, d, W) satisfies (27) or (29), then belief propagation (Algorithm 1) achieves the optimal recovery accuracy of*

$$1 - \lim_{k\to\infty} P_e(\sigma_\rho | T_k, \omega_{T_k}). \tag{13}$$

For vanilla SBM, uniqueness of BP fixed point implications optimal recovery, given an initial recovery algorithm with nice accuracy guarantees.

Theorem 5 (Optimal recovery for SBM) *Work under the same setting as Theorem 3. Suppose $d\lambda^2 > 1$ and (q, λ, d) satisfies (28). Suppose there is an algorithm $\mathcal{A}$ and a constant $\epsilon > 0$ (not depending on n) such that with probability $1 - o(1)$, the empirical transition matrix $F \in \mathbb{R}^{q\times q}$ defined as*

$$F_{i,j} := \frac{\#\{v \in V : X_v = i, \widehat{X}_v = j\}}{\#\{v \in V : X_v = i\}}, \qquad \widehat{X} := \mathcal{A}(G) \tag{14}$$

satisfies

(1) $\|F^\top \mathbb{1} - \mathbb{1}\|_\infty = o(1)$*;*

(2) $\sigma_{\min}(F) > \epsilon$*, where* $\sigma_{\min}$ *denotes the smallest singular value;*

(3) there exists a permutation $\tau \in \mathrm{Aut}([q])$ *such that* $F_{\tau(i),i} > F_{\tau(i),j} + \epsilon$ *for all* $i \neq j \in [q]$.

(Note that we do not asssume F *stays the same for different calls to* $\mathcal{A}$*.)*

Then there is an algorithm (Algorithm 2) achieving the optimal recovery accuracy of

$$1 - \lim_{k\to\infty} P_e(\sigma_\rho | T_k, \sigma_{L_k}). \tag{15}$$

Our assumption on the initial recovery algorithm is a generalization of the one used in the $q = 2$ case by Mossel et al. (2016). Our assumption is much weaker than the one of Chin and Sly (2020), which requires the initial point to be close enough to Id and seems unlikely to hold near the KS threshold. In comparison, our initial point assumption seems more likely to hold near the Kesten-Stigum threshold. For example, it is plausible that a balanced algorithm would achieve the empirical transition matrix F to be close to P_λ for some $|\lambda| = \Omega(1)$.

Our technique. For the positive result (Theorem 1), we generalize the degradation method of Abbe et al. (2021) to q-ary symmetric channels. In this method, we find suitable potential functions Φ on the space of FMS channels, such that for two channels $M, \widetilde{M}$ related by degradation ($\widetilde{M} \le_{\deg} M$), we have (1) $\Phi(M) - \Phi(\widetilde{M})$ contracts to 0 under iterations of BP (2) if $\Phi(M) = \Phi(\widetilde{M})$, then $M = \widetilde{M}$. This shows that the limit channels $\mathrm{BP}^\infty(M)$ and $\mathrm{BP}^\infty(\widetilde{M})$ are equal.

To carry out this method, we develop a theory of q-ary symmetric channels, generalizing the classical theory of binary memoryless symmetric (BMS) channels. We show that q-ary symmetric channels are equivalent to symmetric distributions on the probability simplex $\mathcal{P}([q])$, and degradation relationship has a coupling characterization under the distribution interpretation.

We remark that Yu and Polyanskiy (2022) also uses channel degradation but in a very different way. For more discussions see Section H.

For the negative result (Theorem 2), we show that when the survey channel W is weak enough, the limit channel BP_W^∞ can be arbitrarily weak. We prove this by studying behavior of χ^2-capacity under BP recursion. One difficulty of working with χ^2-capacity is that in general they are not subadditive under $\star$-convolution. Subadditivity is a very desirable property when studying BP recursion, and holds for KL capacity (folklore) and symmetric KL (SKL) capacity Külske and Formentin (2009). For χ^2-capacity, subadditivity is true for BMS channels Abbe and Boix-Adserà (2019), but there are counterexamples when $q \ge 3$. Nevertheless, we establish a local subadditivity result, i.e., χ^2-capacity is almost subadditive under $\star$-convolution for weak enough channels.

Mutual information formula (Theorem 3) and optimal recovery for SBM with side information (Theorem 4) are direct consequences of boundary irrelevance (Theorem 1), generalizing Abbe et al. (2021) and Mossel and Xu (2016) respectively. For optimal recovery for the vanilla SBM (Theorem 5), we need to handle asymmetric initial points (Section E), and suitably generalize the algorithm in Mossel et al. (2016).

Structure of the paper. In Section 2, we establish a theory of FMS channels. In Section 3, we prove Theorem 1, boundary irrelevance and uniqueness of BP fixed point for a wide range of parameters. In Section 4, we prove Theorem 2, that boundary irrelevance does not hold between the reconstruction threshold and the Kesten-Stigum threshold.

In Section A, we discuss limits of information channels. In Section B, we give missing proofs in Section 2. In Section C, we give missing proofs in Section 3. In Section D, we give missing proofs in Section 4, including the local subadditivity of χ^2-information (Lemma 20). In Section E, we discuss asymmetric fixed points of the BP operator. In Section F, we prove Theorem 3, SBM mutual information formula. In Section G, we prove Theorem 4 and 5, optimal recovery algorithms. In Section H, we discuss several further directions.

2. FMS Channels

In this section we introduce q-ary fully memoryless symmetric (q-FMS) channels.[1] They are a generalization of binary memoryless symmetric (BMS) channels to q-ary input alphabets. For background on BMS channels, see e.g., Richardson and Urbanke (2008).

Definition 6 (Fully memoryless symmetric (FMS) channels) *A q-ary fully memoryless symmetric (q-FMS) channel (or an FMS channel when q is obvious from context) is a channel $P : \mathcal{X} \to \mathcal{Y}$ with input alphabet $\mathcal{X} = [q]$ such that there exists a group homomorphism $\iota : \operatorname{Aut}(\mathcal{X}) \to \operatorname{Aut}(\mathcal{Y})$ such that for any measurable $E \subseteq \mathcal{Y}$, we have*

$$P(E|x) = P(\iota(\tau)E|\tau(x)) \tag{16}$$

for all $x \in \mathcal{X}$, $\tau \in \operatorname{Aut}(\mathcal{X})$. Here $\operatorname{Aut}(\mathcal{X})$ denotes the symmetry group (also known as the automorphism group) of $\mathcal{X}$.

By definition, 2-FMS channels are exactly BMS channels.

We remark that q-FMS channels are a special case of input-symmetric channels (see e.g., (Polyanskiy and Wu, 2023+, Chapter 19)) whose group of symmetries is the whole $\operatorname{Aut}(\mathcal{X})$. Makur and Polyanskiy (2018) studied comparison between q-ary symmetric channels, but their definition of symmetric channels is quite different from ours.

The BSC mixture representation of BMS channels has been useful in proving results about BMS channels. Therefore it is desirable to generalize this theory to FMS channels. We define fully symmetric channels (FSCs), which generalize BSCs, and will serve as basic building blocks for FMS channels.

Definition 7 *Let $\mathcal{X} = [q]$, $\mathcal{Y} = \operatorname{Aut}(\mathcal{X})$. For $\pi \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$, define channel $\mathrm{FSC}_\pi : \mathcal{X} \to \mathcal{Y}$ as*

$$\mathrm{FSC}_\pi(\tau|i) = \frac{1}{(q-1)!}\pi_{\tau^{-1}(i)} \qquad \forall i \in \mathcal{X}, \tau \in \operatorname{Aut}(\mathcal{X}), \tag{17}$$

where $\operatorname{Aut}(\mathcal{X})$ acts on $\mathcal{Y}$ by left multiplication.

We can verify that

$$\mathrm{FSC}_\pi(\eta\tau|\eta(i)) = \frac{1}{(q-1)!}\pi_{(\eta\tau)^{-1}(\eta(i))} = \frac{1}{(q-1)!}\pi_{\tau^{-1}(i)} = \mathrm{FSC}_\pi(\tau|i) \tag{18}$$

for $i \in \mathcal{X}$, $\eta, \tau \in \operatorname{Aut}(\mathcal{X})$. So FSCs are examples of FMS channels.

One of the most basic relationships between two channels is degradation.

1. Here “fully” modifies “symmetric”, and indicates that the symmetry group of the channel is the full symmetric group $\operatorname{Aut}(\mathcal{X})$ as opposed to a subgroup.

Definition 8 *Let $P : \mathcal{X} \to \mathcal{Y}$ ad $Q : \mathcal{X} \to \mathcal{Z}$ be two channels with the same input alphabet. We say P is a degradation of Q (denoted $P \le_{\deg} Q$) if there exists a channel $R : \mathcal{Z} \to \mathcal{Y}$ respecting $\operatorname{Aut}(\mathcal{X})$ action such that $P = R \circ Q$. We say P and Q are equivalent if $P \le_{\deg} Q$ and $Q \le_{\deg} P$.*

In other words, P is a degradation of Q if we can simulate P by postprocessing the output of Q.

In the binary case, all BMS channels are equivalent to mixtures of BSCs (see e.g., Richardson and Urbanke (2008)). We generalize this result to q-ary FMS channels.

Proposition 9 (Structure of FMS channels) *Every FMS channel is equivalent to a mixture of FSCs, i.e., every FMS channel $P : \mathcal{X} \to \mathcal{Y}$ is equivalent to a channel $X \to (\pi, Z)$ where $\pi \sim P_\pi \in \mathcal{P}(\mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X}))$ is independent of X, and $Z \sim \mathrm{FSC}_\pi(\cdot|X)$ conditioned on π and X. Furthermore, P_π is uniquely deteremined by P.*

Proof is deferred to Section B. In the setting of the above proposition, we call π the π-component of P, and P_π the π-distribution of P. Prop. 9 establishes an equivalence between an FMS channel and a probability distribution on $\mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$, which maps an FMS channel to its π-distribution. In particular, the π-distribution is an invariant property of an FMS channel under equivalence. For BMS channels we usually denote the π-component as a single number $\Delta \in [0, \frac{1}{2}]$, and call it the Δ-component.

Degradation has a nice characterization in terms of the π-components.

Proposition 10 *Let P, Q be two FMS channels, and π_P and π_Q be their π-components. Then $P \le_{\deg} Q$ if and only if there exists a coupling between π_P and π_Q such that*

$$\pi \le_m \mathbb{E}[\pi_Q|\pi_P = \pi] \qquad \forall \pi \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X}), \tag{19}$$

where $\le_m$ denotes majorization (see e.g., (Hardy et al., 1934, 2.18)). We use the convention that elements $\pi \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$ are non-increasing so that the expectation is well-defined.

Proof is deferred to Section B.

There are different ways to construct new FMS channels from given FMS channels. In this paper we focus on two ways: composition with Potts channels and $\star$-convolution.

Fix $q \ge 2$. For $\lambda \in [-\frac{1}{q-1}, 1]$, define Potts channel $P_\lambda : [q] \to [q]$ as $P_\lambda(y|x) = \lambda \mathbb{1}\{x = y\} + \frac{1-\lambda}{q}$ for $x, y \in [q]$. Then given any q-FMS channel P, $P \circ P_\lambda$ is also a q-FMS channel. Furthermore, the π-distribution of $P \circ P_\lambda$ is $f_*(P_\pi)$, where P_π is the π-distrbution of P, $f(\pi) = \lambda\pi + \frac{1-\lambda}{q}$, and f_* is the induced pushforward map.

Given two channels $P : \mathcal{X} \to \mathcal{Y}$, $Q : \mathcal{Y} \to \mathcal{Z}$, their $\star$-convolution $P \star Q : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ is defined by letting P and Q acting on the same input independently. When P and Q are q-FMS channels, $P \star Q$ has a natural q-FMS structure. If the π-component of P (resp. Q) has distribution P_π (resp. Q_π), then the π-component of $P \star Q$ has distribution

$$\mathbb{E}_{\substack{\pi \sim P_\pi \\ \pi' \sim Q_\pi}} \left[\sum_{\tau \in \operatorname{Aut}([q])} \left(\frac{1}{(q-1)!} \sum_{i \in [q]} \pi_i \pi'_{\tau(i)} \right) \mathbb{1}_{\pi \star_\tau \pi'} \right] \tag{20}$$

$$\text{where} \quad \pi \star_\tau \pi' := \left(\frac{\pi_i \pi'_{\tau(i)}}{\sum_{j \in [q]} \pi_j \pi'_{\tau(j)}} \right)_{i \in [q]} \in \mathcal{P}([q])/\operatorname{Aut}([q]) \tag{21}$$

and $\mathbb{1}_\theta \in \mathcal{P}(\mathcal{P}([q])/\operatorname{Aut}([q]))$ denotes the point distribution at $\theta \in \mathcal{P}([q])/\operatorname{Aut}([q])$. We use $M^{\star b}$ to denote the b-th $\star$-power: $M^{\star 0} = \mathrm{Id}$ and $M^{\star b} = M^{\star(b-1)} \star M$.

Given any q-FMS channel P, we can restrict the input alphabet to get a q'-FMS for $q' \le q$. (Because of symmetry, the restricted channel is unique up to channel equivalence no matter what size-q' subset we choose.) In this paper we only use the case $q' = 2$, i.e., restrict to a BMS channel. We use P^R to denote the restricted BMS channel.[2]

We study the behavior of information measures under belief propagation. The following information measures are particularly useful.

Definition 11 *Let P be a q-FMS channel and π be its π-component. We define the following quantities.*

$$P_e(P) = \mathbb{E}\min\{1-\pi_i : i \in [q]\}, \qquad \text{(probability of error)}$$

$$C(P) = \log q - \mathbb{E}\sum_{i\in[q]} \pi_i \log \frac{1}{\pi_i}, \qquad \text{(capacity)}$$

$$C_{\chi^2}(P) = \mathbb{E}\left[q\sum_{i\in[q]} \pi_i^2 - 1\right], \qquad \text{(χ^2-capacity)}$$

$$C_{\mathrm{SKL}}(P) = \mathbb{E}\left[\sum_{i\in[q]} \left(\pi_i - \frac{1}{q}\right)\log(\pi_i)\right]. \qquad \text{(SKL capacity)}$$

For BMS channels we also define the Bhattacharyya coefficient

$$Z(P) = \mathbb{E}\left[2\sqrt{\Delta(1-\Delta)}\right] \qquad \text{(Bhattacharyya coefficient)}$$

where Δ is the Δ-component of P.

These information measures respect degradation, as summarized in the next lemma.

Lemma 12 *Let P and Q be two q-FMS channels with $P \le_{\deg} Q$. Then $P_e(P) \ge P_e(Q), C(P) \le C(Q), C_{\chi^2}(P) \le C_{\chi^2}(Q), C_{\mathrm{SKL}}(P) \le C_{\mathrm{SKL}}(Q)$. If $q = 2$ then we also have $Z(P) \ge Z(Q)$.*

Proof By definition of degradation, and data processing inequality for f-divergences. ■

3. Uniqueness and boundary irrelevance results

In this section we prove uniqueness of BP fixed point and boundary irrelevance results for the Potts model for a wide range of parameters. We consider the q-ary Potts model with broadcasting channel P_λ on a regular tree or a Poisson tree with expected offspring d.

We state two results, one for the low SNR regime and one for the high SNR regime. We define the following constants used in the results.

2. Here "R" stands for "restriction".

Definition 13 *For $q \in \mathbb{Z}_{\geq 2}$, $\lambda \in \left[-\frac{1}{q-1}, 1\right]$, $d \geq 0$, we define*

$$C^L(q,\lambda) := \sup_{\substack{\pi \in \mathcal{P}([q]) \\ v \in \mathbb{1}^\perp \subseteq \mathbb{R}^q}} \frac{f^L\left(\lambda\pi + \frac{1-\lambda}{q}, v\right)}{f^L(\pi, v)}, \tag{22}$$

$$\textit{where } f^L(\pi, v) := \left\langle \pi^{-1} + \frac{1}{q}\pi^{-2}, v^2 \right\rangle, \tag{23}$$

$$C^H(q,\lambda) := \sup_{\substack{\pi \in \mathcal{P}([q]) \\ v \in \mathbb{1}^\perp \subseteq \mathbb{R}^q}} \frac{f^H\left(\lambda\pi + \frac{1-\lambda}{q}, v\right)}{f^H(\pi, v)}, \tag{24}$$

$$\textit{where } f^H(\pi, v) := \|\pi^{1/4}\|_2^2 \|\pi^{-3/4} v\|_2^2 - \left\langle \pi^{1/4}, \pi^{-3/4} v \right\rangle^2, \tag{25}$$

$$c^H(q,\lambda,d) := \left(\frac{2}{q} + \frac{q-2}{q} \cdot \frac{d\lambda^2 - 1}{d\lambda - 1}\right)^{-1}. \tag{26}$$

We have the following bounds on these constants: $C^L(q,\lambda) \leq q^2$ (Prop. 31), $C^H(q,\lambda) \leq q^{5/2}$ (Prop. 32), $c^H(q,\lambda,d) \geq 1$ (obvious).

Theorem 14 (Low SNR) *If*

$$d\lambda^2 C^L(q,\lambda) < 1, \tag{27}$$

where C^L is defined in (22), then boundary irrelevance and stability of BP fixed point hold.

Theorem 15 (High SNR) *If $d\lambda^2 > 1$ and*

$$d\lambda^2 \exp\left(-c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{2}\right) C^H(q,\lambda) < 1, \tag{28}$$

where c^H is defined in (26), C^H is defined in (24), then boundary irrelevance and stability of BP fixed point hold.

Let W be a q-FMS channel. If $d\lambda^2 > 1$ and

$$d\lambda^2 \exp\left(-c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{2}\right) C^H(q,\lambda) Z(W^R) < 1, \tag{29}$$

where W^R denotes the restriction of W to a BMS channel, and Z denotes the Bhattacharyya coefficient, then boundary irrelevance holds with repsect to W.

Proof [Proof of Theorem 1 given Theorem 14 and 15] We prove the low SNR case and the high SNR case separately.

Low SNR: By Prop. 31, $C^L(q,\lambda) \leq q^2$. If $d\lambda^2 < q^{-2}$, then (27) holds and Theorem 14 applies.

High SNR: We prove that (28) holds whenever $d\lambda^2 > 1 + 56 \max\{\lambda, q^{-1}\} \log q$.

By Prop. 32, $C^H(q,\lambda) \leq q^{5/2}$. For $d\lambda^2 > 1$, we have

$$c^H(q,\lambda,d) \geq \left(\frac{2}{q} + \frac{q-2}{q} \cdot \max\{\lambda, 0\}\right)^{-1} \geq \left(\frac{2}{q} + \max\{\lambda, 0\}\right)^{-1} \geq \frac{1}{4}\max\{\lambda, q^{-1}\}^{-1}. \tag{30}$$

Therefore

$$d\lambda^2 \exp(-c^H(q,\lambda,d) \cdot \frac{d\lambda^2-1}{2})C^H(q,\lambda) \le d\lambda^2 \exp\left(-\frac{d\lambda^2-1}{8\max\{\lambda, q^{-1}\}}\right) q^{5/2} =: g_{q,\lambda}(d). \tag{31}$$

Computing $g'_{q,\lambda}(d)$, we see that $g_{q,\lambda}(d)$ is monotone decreasing in d when $d\lambda^2 > 8\max\{\lambda, q^{-1}\}$. Therefore it suffices to prove $g_{q,\lambda}(d_0) < 1$ where $d_0\lambda^2 = 1 + 56\max\{\lambda, q^{-1}\}\log q$. We have

$$g_{q,\lambda}(d_0) = (1 + 56\max\{\lambda, q^{-1}\}\log q)\exp(-7\log q)q^{5/2} \le (1 + 56\log q)q^{-9/2}. \tag{32}$$

The last expression is < 1 for all $q \ge 3$. This finishes the proof. ■

3.1. The degradation method

Let $(M_k)_{k\ge 0}$ and $(\widetilde{M}_k)_{k\ge 0}$ be two sequences of q-FMS channels satisfying the belief propagation recursion, i.e.,

$$M_{k+1} = \mathrm{BP}(M_k), \qquad \widetilde{M}_{k+1} = \mathrm{BP}(\widetilde{M}_k), \tag{33}$$

$$\mathrm{BP}(M) := \mathbb{E}_b[(M_k \circ P_\lambda)^{\star b} \star W], \tag{34}$$

where b follows the branching number distribution (constant if working with regular trees), and W is the survey FMS channel (trivial if there is no survey).

For the boundary irrelevance problem, we take $\widetilde{M}_0 = 0$, $M_0 = \mathrm{Id}$. For stability of BP fixed point, we take $M_0 = \mathrm{Id}$ and $\widetilde{M}_0$ be a given non-trivial FMS channel. Our goal is to show the the limit $\lim_{k\to\infty} M_k$ and $\lim_{k\to\infty} \widetilde{M}_k$ both exist in the sense of weak convergence (Section A) and are equal.

From now on, we assume that $M_0 = \mathrm{Id}$, and either (1) W is non-trivial and $\widetilde{M}_0 = 0$, or (2) $W = 0$ and $\widetilde{M}_0$ is non-trivial. Note that in both cases, the initial channels satisfy $\widetilde{M}_0 \le_{\deg} M_0$. Because the BP operator preserves degradation preorder, we have $\widetilde{M}_k \le_{\deg} M_k$ for all $k \ge 0$. So the two channel sequences are naturally related to each other by degradation.

Because $M_0 = \mathrm{Id}$, we have $M_k \ge_{\deg} M_{k+1}$ for all $k \ge 0$. Therefore by Lemma 21, $M_\infty := \lim_{k\to\infty} M_k$ exists. For the boundary irrelevance problem, we also have $M_k \le_{\deg} M_{k+1}$ for all $k \ge 0$, and by Lemma 22, $\widetilde{M}_\infty := \lim_{k\to\infty} \widetilde{M}_k$ exists. However, for the stability of BP fixed point problem, it is a priori unclear whether the limit $\lim_{k\to\infty} \widetilde{M}_k$ exists.

We prove the limit $\lim_{k\to\infty} \widetilde{M}_k$ exists and is equal to M_∞ by generalizing the degradation method from Abbe et al. (2021). Let $\phi : \mathcal{P}([q]) \to \mathbb{R}$ be a strongly convex function invariant under $\mathrm{Aut}([q])$ action. Extend it to a function Φ : {FMS channels} $\to \mathbb{R}$ as $\Phi(P) = \mathbb{E}\phi(\pi_P)$. By degradation, we have $\Phi(M_k) \ge \Phi(\widetilde{M}_k)$ for all $k \ge 0$. The following proposition shows that it suffices to prove contraction of potential function Φ.

Proposition 16 *Assume that $\phi : \mathcal{P}([q]) \to \mathbb{R}$ is α-strongly convex for some $\alpha > 0$, and that*

$$\lim_{k\to\infty} (\Phi(M_k) - \Phi(\widetilde{M}_k)) = 0. \tag{35}$$

Then under the canonical coupling, we have

$$\lim_{k\to\infty} \mathbb{E}\|\pi_k - \widetilde{\pi}_k\|_2^2 = 0, \tag{36}$$

where π_k (resp. $\widetilde{\pi}_k$) is the π-component of M_k (resp. $\widetilde{M}_k$). In particular, if $M_0 = \mathrm{Id}$, then both limits $\lim_{k\to\infty} M_k$ and $\lim_{k\to\infty} \widetilde{M}_k$ exist in the sense of weak convergence, and the two limits are equal.

Proof

$$\begin{aligned}\Phi(M_k) - \Phi(\widetilde{M}_k) &= \mathbb{E}_{\widetilde{\pi}_k}\mathbb{E}[\phi(\pi_k) - \phi(\widetilde{\pi}_k)|\widetilde{\pi}_k] \\ &\ge \mathbb{E}_{\widetilde{\pi}_k}\mathbb{E}[\langle \nabla\phi(\widetilde{\pi}_k), \pi_k - \widetilde{\pi}_k\rangle + \frac{\alpha}{2}\|\pi_k - \widetilde{\pi}_k\|_2^2|\widetilde{\pi}_k] \\ &\ge \mathbb{E}_{\widetilde{\pi}_k}\langle \nabla\phi(\widetilde{\pi}_k), \mathbb{E}[\pi_k|\widetilde{\pi}_k] - \widetilde{\pi}_k\rangle + \frac{\alpha}{2}\mathbb{E}\|\pi_k - \widetilde{\pi}_k\|_2^2 \\ &\ge \frac{\alpha}{2}\mathbb{E}\|\pi_k - \widetilde{\pi}_k\|_2^2,\end{aligned} \tag{37}$$

where the second step is by α-strongly convexity, and the third step is because $\widetilde{\pi}_k \le_m \mathbb{E}[\pi_k|\widetilde{\pi}_k]$ and ϕ is convex (thus Schur-convex). Taking the limit $k \to \infty$, we see that the Wasserstein W_2 distance between the π-distributions of M_k and $\widetilde{M}_k$ goes to 0. Because $\lim_{k\to\infty} M_k$ converges weakly to a limit M_∞, $\lim_{k\to\infty} \widetilde{M}_k$ also converges to the same limit. ■

Proposition 17 *Assume that $\phi : \mathcal{P}([q]) \to \mathbb{R}$ is α-strongly convex for some $\alpha > 0$, and that*

$$\lim_{k\to\infty} (\Phi(M_k) - \Phi(\widetilde{M}_k)) = 0. \tag{38}$$

whenever $M_0 = \mathrm{Id}$ and (1) W is non-trivial and $\widetilde{M}_0 = 0$, or (2) $W = 0$ and $\widetilde{M}_0$ is non-trivial. Then boundary irrelevance and stability of BP fixed point hold.

Proof **Boundary irrelevance:** Let $\widetilde{M}_0 = 0$, $M_0 = \mathrm{Id}$. By Prop. 16, we have

$$\lim_{k\to\infty} M_k = \lim_{k\to\infty} \widetilde{M}_k. \tag{39}$$

In particular,

$$\lim_{k\to\infty} C(M_k) = \lim_{k\to\infty} C(\widetilde{M}_k) \tag{40}$$

where C denote capacity (Definition 11). Note that

$$\lim_{k\to\infty} C(M_k) = \lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}, \omega_{T_k}|T_k), \tag{41}$$

$$\lim_{k\to\infty} C(\widetilde{M}_k) = \lim_{k\to\infty} I(\sigma_\rho; \omega_{T_k}|T_k). \tag{42}$$

So this proves boundary irrelevance.

Stability of BP fixed point: Suppose there is a non-trivial fixed point FMS channel U. Let $\widetilde{M}_0 = U$, $M_0 = \mathrm{Id}$. By Prop. 16, we have

$$\lim_{k\to\infty} M_k = \lim_{k\to\infty} \widetilde{M}_k. \tag{43}$$

Because U is a fixed point, LHS is equal to U. On the other hand, RHS does not depend on U. Therefore there is a unique non-trivial FMS fixed point. ■

We remark that it might be possible to extend the degradation method to asymmetric models. See Section H for more discussions.

3.2. Low SNR

For the low SNR case, we use SKL capacity as the potential function. We define

$$\phi^L(\pi) = C_{\mathrm{SKL}}(\mathrm{FSC}_\pi) = \sum_{i\in[q]} (\pi_i - \frac{1}{q}) \log \pi_i. \tag{44}$$

It is useful for our purpose because it is strongly convex (Lemma 23) and additive under $\star$-convolution (Lemma 24). Using this properties we show that desired contraction holds (Prop. 25). We defer the proof of Theorem 14 to Section C.1.

3.3. High SNR

For the high SNR case, we use Bhattacharyya coefficient as the potential function. We define

$$\phi^H(\pi) = Z(\mathrm{FSC}^R_\pi) = \frac{1}{q-1}\left(\left(\sum_{i\in[q]}\sqrt{\pi_i}\right)^2 - 1\right). \tag{45}$$

It is useful for our purpose because it is strongly concave (Lemma 26) and multiplicative under $\star$-convolution (Lemma 27). Using this properties we show that desired contraction holds (Prop. 29). We defer the proof of Theorem 15 to Section C.2.

4. Boundary irrelevance does not always hold

In this section we prove that boundary irrelevance does not hold for the Potts model between the reconstruction threshold and the Kesten-Stigum threshold.

Proposition 18 *In the setting of Theorem 2, for all $\delta > 0$, there exists $\epsilon > 0$ such that for any FMS survey channel W with $C_{\chi^2}(W) \le \epsilon$, we have*

$$\lim_{k\to\infty} I_{\chi^2}(\sigma_\rho; \omega_{T_k}|T_k) \le \delta. \tag{46}$$

Proof [Proof of Theorem 2 given Prop. 18] In the reconstruction regime,

$$\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}, \omega_{T_k}|T_k) \ge \lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}|T_k) > 0. \tag{47}$$

Take $\delta > 0$ such that $\delta \log 2 < \lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}|T_k)$. Because $I \le I_{\chi^2} \log 2$, and by Prop. 18, for weak enough survey channel W we have

$$\lim_{k\to\infty} I(\sigma_\rho; \omega_{T_k}|T_k) \le \lim_{k\to\infty} I_{\chi^2}(\sigma_\rho; \omega_{T_k}|T_k)\log 2 \le \delta \log 2 < \lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}, \omega_{T_k}|T_k). \tag{48}$$

Therefore

$$
\begin{aligned}
\lim_{k\to\infty} I(\sigma_\rho;\sigma_{L_k}|T_k,\omega_{T_k}) &= \lim_{k\to\infty}(I(\sigma_\rho;\sigma_{L_k},\omega_{T_k}|T_k) - I(\sigma_\rho;\omega_{T_k}|T_k)) \\
&= \lim_{k\to\infty} I(\sigma_\rho;\sigma_{L_k},\omega_{T_k}|T_k) - \lim_{k\to\infty} I(\sigma_\rho;\omega_{T_k}|T_k) > 0.
\end{aligned} \tag{49}
$$

■

Proof of Prop. 18 is deferred to Section D.2. The proof uses contraction and local subadditivity properties of the information measure C_{χ^2}.

Lemma 19 (Contraction) *For any FMS channel P, we have*

$$C_{\chi^2}(P \circ P_\lambda) \le \lambda^2 C_{\chi^2}(P). \tag{50}$$

Proof By reversibility and χ^2-contraction coefficient: $\eta_{\chi^2}(\mathrm{Unif}([q]), P_\lambda) = \lambda^2$. ■

Lemma 20 (Local subadditivity) *Fix $q \in \mathbb{Z}_{\ge 2}$. For any $\epsilon > 0$ and q-FMS channels P, Q with $C_{\chi^2}(P) \le \epsilon$, we have*

$$C_{\chi^2}(P \star Q) \le (1 + O_q(\epsilon^{1/5}))(C_{\chi^2}(P) + C_{\chi^2}(Q)), \tag{51}$$

where O_q hides a constant depending on q.

Proof is deferred to Section D.1.

Acknowledgments

We thank Emmanuel Abbe, Elisabetta Cornacchia, Jingbo Liu, Youngtak Sohn for helpful discussions. We thank the anonymous reviewers for helpful comments.

Research was sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

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 Enric Boix-Adserà. Subadditivity beyond trees and the chi-squared mutual information. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 697–701. IEEE, 2019.

Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015a.

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*, 2015b.

Emmanuel Abbe and Colin Sandon. Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. *Advances in Neural Information Processing Systems*, 29, 2016.

Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

Emmanuel Abbe, Elisabetta Cornacchia, Yuzhou Gu, and Yury Polyanskiy. Stochastic block model entropy and broadcasting on trees with survey. In *Conference on Learning Theory*, pages 1–25. PMLR, 2021.

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.

Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Reconstruction threshold for the hardcore model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, pages 434–447. Springer, 2010.

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.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015.

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.

Byron Chin and Allan Sly. Optimal recovery of block models with q communities. *arXiv preprint arXiv:2010.10672*, 2020.

Byron Chin and Allan Sly. Optimal reconstruction of general sparse stochastic block models. *arXiv preprint arXiv:2111.00697*, 2021.

Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 146–157, 2017.

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.

Tomas Dominguez and Jean-Christophe Mourrat. Mutual information for the sparse stochastic block model. *arXiv preprint arXiv:2209.04513*, 2022.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

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. *arXiv preprint arXiv:2005.05444*, 2020.

Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, 1934.

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.

Anuran Makur and Yury Polyanskiy. Comparison of channels: Criteria for domination by a symmetric channel. *IEEE Transactions on Information Theory*, 64(8):5704–5725, 2018.

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.

Marc Mézard and Andrea Montanari. Reconstruction on trees and spin glass transition. *Journal of statistical physics*, 124:1317–1350, 2006.

Elchanan Mossel. Reconstruction on trees: beating the second eigenvalue. *The Annals of Applied Probability*, 11(1):285–300, 2001.

Elchanan Mossel and Yuval Peres. Information flow on trees. *The Annals of Applied Probability*, 13(3):817–844, 2003.

Elchanan Mossel and Jiaming Xu. Local algorithms for block models with side information. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 71–80, 2016.

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. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016.

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. *arXiv preprint arXiv:2212.03362*, 2022.

Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2023+.

Tom Richardson and Rüdiger Urbanke. *Modern coding theory*. Cambridge university press, 2008.

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.

Ludovic Stephan and Laurent Massoulié. Robustness of spectral methods for community detection. In *Conference on Learning Theory*, pages 2831–2860. PMLR, 2019.

Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *arXiv preprint arXiv:2211.15242*, 2022.

Appendix A. Limit of channels

In this section we build the foundation for discussing limits of information channels. We view a channel $P : \mathcal{X} \to \mathcal{Y}$ as a distribution of posterior distributions under uniform prior, i.e., the distribution of $P_{X|Y}$ where $P_X = \mathrm{Unif}(\mathcal{X})$, $Y \sim P(\cdot|X)$. Let μ denote the posterior distribution variable and $P_\mu \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$ be its distribution (called P's posterior distribution's distrbution). Note that P_μ is invariant under channel equivalence.

We often work with sequences $(P_k)_{k\ge 0}$ of channels with the same input alphabet $\mathcal{X}$. Let $P_{\pi,k}$ denote the distrbution of posterior distributions of P_k under uniform prior. Let P_∞ be a channel with input alphabet $\mathcal{X}$ and posterior distribution's distrbution $P_{\pi,\infty}$. We say $(P_k)_{k\ge 0}$ converges weakly to P_∞ if $(P_{\pi,k})_{k\ge 0}$ converges weakly to $P_{\pi,\infty}$ as distributions on $\mathcal{P}(\mathcal{X})$.

In general, given such a sequence, a limit does not necessarily exist. Nevertheless, when the channels are related to each other via degradation, a limit channel exists.

Lemma 21 *Let $(P_k : \mathcal{X} \to \mathcal{Y}_k)_{k\ge 0}$ be a sequence of channels with the same finite input alphabet. If $P_k \ge_{\deg} P_{k+1}$ for all k, then $(P_k)_{k\ge 0}$ converges weakly to some channel P_∞.*

Proof By definition of degradation, there exists channel $R_k : \mathcal{Y}_k \to \mathcal{Y}_{k+1}$ such that $P_{k+1} = R_k \circ P_k$. This gives rise to an infinite Markov chain

$$X - Y_0 - Y_1 - Y_2 - \cdots . \tag{52}$$

Let μ_k denote the posterior distribution variable $P_{X|Y_k}$. Then we have

$$\mathbb{E}[\mu_{k-1}|Y_k] = \mu_k. \tag{53}$$

Let $\mathcal{F}_k$ denote the σ-algebra generated by $(Y_i)_{i\ge k}$. Then $(\mathcal{F}_k)_{k\ge 0}$ is a reverse filtration and $(\mu_k)_{k\ge 0}$ is a reverse martingale with respect to $(\mathcal{F}_k)_{k\ge 0}$. By reverse martingale convergence theorem (e.g., (Durrett, 2019, Theorem 4.7.1)), $\lim_{k\to\infty} \mu_k$ converges almost surely. Define $\mu_\infty := \lim_{k\to\infty} \mu_k$. Let P_∞ be a channel with input alphabet $\mathcal{X}$ whose posterior distribution's distribution is μ_∞. Then $(P_k)_{k\ge 0}$ converges weakly to P_∞. ■

Lemma 22 *Let $(P_k : \mathcal{X} \to \mathcal{Y}_k)_{k\ge 0}$ be a sequence of channels with the same finite input alphabet. If $P_k \le_{\deg} P_{k+1}$ for all k, then $(P_k)_{k\ge 0}$ converges weakly to some channel P_∞.*

Proof By definition of degradation, there exists channel $R_k : \mathcal{Y}_k \to \mathcal{Y}_{k-1}$ such that $P_{k-1} = R_k \circ P_k$. This gives rise to an infinite Markov chain

$$X - Y_0 - Y_1 - Y_2 - \cdots . \tag{54}$$

Let μ_k denote the posterior distribution variable $P_{X|Y_k}$. Then we have

$$\mathbb{E}[\mu_{k+1}|Y_k] = \mu_k. \tag{55}$$

Let $\mathcal{F}_k$ denote the σ-algebra generated by $(Y_i)_{i\le k}$. Then $(\mathcal{F}_k)_{k\ge 0}$ is a filtration and $(\mu_k)_{k\ge 0}$ is a martingale with respect to $(\mathcal{F}_k)_{k\ge 0}$. Note that the variables μ_k take values in $\mathcal{P}(\mathcal{X})$, so are uniformly bounded. By martingale convergence theorem (e.g., (Durrett, 2019, Theorem 4.2.11)), $\lim_{k\to\infty} \mu_k$ converges almost surely. Define $\mu_\infty := \lim_{k\to\infty} \mu_k$. Let P_∞ be a channel with input alphabet $\mathcal{X}$ whose posterior distribution's distribution is μ_∞. Then $(P_k)_{k\ge 0}$ converges weakly to P_∞. ■

By symmetry, in Lemma 21 and Lemma 22, if the sequence $(P_k)_{k\ge 0}$ consists of FMS channels, then the limit P_∞ is an FMS channel.

Appendix B. Proofs in Section 2

Proof [Proof of Prop. 9] **Existence:** The proof strategy is to partition $\mathcal{Y}$ into $\mathrm{Aut}(\mathcal{X})$-orbits and show that the channel P restricted to each orbit is equivalent to an FSC.

Step 1. We first prove that we can replace P with an equivalent FMS channel whose $\mathrm{Aut}(\mathcal{X})$ action is free, so that in later steps each orbit is easier to handle. Define channel $\widetilde{P} : \mathcal{X} \to \mathcal{Y} \times \widetilde{\mathcal{Y}}$, where $\widetilde{\mathcal{Y}} = \mathrm{Aut}(\mathcal{X})$, sending X to $(Y, \widetilde{Y})$ where $Y \sim P(\cdot|X)$ and $\widetilde{Y} \sim \mathrm{Unif}(\mathrm{Aut}(\mathcal{X}))$ is independent of X. We give $\widetilde{P}$ an FMS structure where $\mathrm{Aut}(\mathcal{X})$ acts on $\widetilde{\mathcal{Y}}$ by left multiplication. It is easy to see that P is equivalent to $\widetilde{P}$. Therefore we can replace P with $\widetilde{P}$ and wlog assume that $\mathrm{Aut}(\mathcal{X})$ action is free.

Step 2. Let $\mathcal{O} = \mathcal{Y}/\operatorname{Aut}(\mathcal{X})$ be the space of orbits of the $\operatorname{Aut}(\mathcal{X})$ action on $\mathcal{Y}$. For an orbit $o \in \mathcal{O}$, for any two elements $y_1, y_2 \in o$, the posterior distributions $\pi_1 = P_{X|Y=y_1}$ and $\pi_2 = P_{X|Y=y_2}$ (with uniform priors) differ by a permutation, by the assumption that P is FMS. In particular, π_1 and π_2 map to the same element in $\mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$. Therefore we can uniquely assign an element $\pi_o \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$ for any $o \in \mathcal{O}$.

Note that by symmetry, the distribution of o does not depend on the input distribution. Let $P_o \in \mathcal{P}(\mathcal{O})$ be this distribution. Then P is equivalent to the channel $X \to (o, Z)$ where $o \sim P_o$ is independent of X, and $Z \sim \mathrm{FSC}_{\pi_o}(\cdot|X)$. (Because $\operatorname{Aut}(\mathcal{X})$ action on $\mathcal{Y}$ is free, this equivalence is in fact just renaming the output space.)

Step 3. Finally we prove that the FMS channel $X \to (o, Z)$ is equivalent to $X \to (\pi_o, Z)$. One side is easy: given (o, Z), we can generate (π_o, Z). For the other side, given (π, Z), we can generate $o' \sim P_{o|\pi_o=\pi}$. Then (o', Z) has the same distribution as (o, Z), conditioned on any input distribution. This finishes the existence proof.

Uniqueness: For any FMS channel $X \xrightarrow{Q} Y$, we can associate it with a distribution Q_π on $\mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$, defined as the distribution of the posterior distribution $Q_{X|Y}$, where $Y \sim Q_{Y|X} \circ \operatorname{Unif}(\mathcal{X})$ is generated with uniform prior distribution. (By definition Q_π is a distribution on $\mathcal{P}(\mathcal{X})$. However, by symmetry property of FMS, Q_π is invariant under $\operatorname{Aut}(\mathcal{X})$ action.) It is easy to see that Q_π distribution is preserved under equivalence between FMS channels. Furthermore, for an FMS channel of form $X \to (\pi, Z)$ as described in the proposition statement, this distribution of posterior distribution is equal to P_π. Therefore P_π is uniquely deteremined by P. ■

Proof [Proof of Prop. 10] **Degradation ⇒ Coupling:** Say P maps X to Y, and Q maps X to Z. Let $\pi'_P \in \mathcal{P}(\mathcal{X})$ be the posterior distribution of input X given output Y, where $Y \sim P_{Y|X} \circ \operatorname{Unif}(\mathcal{X})$ is generated with uniform prior distribution. Similarly define π'_Q. Then π_P (resp. π_Q) is the orbit of π'_P (resp. π'_Q) under permutation.

Degradation relationship $P = R \circ Q$ induces a coupling on the posterior distributions π'_P and π'_Q. One can check that this coupling is invariant under $\operatorname{Aut}(\mathcal{X})$ action and satisfies

$$\pi' = \mathbb{E}[\pi'_Q | \pi'_P = \pi'] \qquad \forall \pi' \in \mathcal{P}(\mathcal{X}). \tag{56}$$

For any $\pi' \in \mathcal{P}(\mathcal{X})$, let $p(\pi') \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$ denotes its projection. Then we have

$$p(\pi') \le_m \mathbb{E}[p(\pi'_Q) | \pi'_P = \pi']. \tag{57}$$

Taking expectation over the orbit, we get

$$\pi \le_m \mathbb{E}[\pi_Q | \pi_P = \pi] \qquad \forall \pi \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X}). \tag{58}$$

Coupling ⇒ Degradation: Step 1. We prove that for $\pi, \pi' \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X})$, if $\pi \le_m \pi'$, then $\mathrm{FSC}_\pi \le_{\deg} \mathrm{FSC}_{\pi'}$. Because $\pi \le_m \pi'$, there exists $a \in \mathcal{P}(\operatorname{Aut}(\mathcal{X}))$ such that (see e.g., (Hardy et al., 1934, 2.20))

$$\pi_i = \sum_{\sigma \in \operatorname{Aut}(\mathcal{X})} a_\sigma \pi'_{\sigma^{-1}(i)} \qquad \forall i \in \mathcal{X}. \tag{59}$$

For $\rho \in \operatorname{Aut}(\mathcal{X})$, we have

$$\mathrm{FSC}_\pi(\rho|i) = \frac{1}{(q-1)!}\pi_{\rho^{-1}(i)} = \sum_{\sigma \in \operatorname{Aut}(\mathcal{X})} a_\sigma \frac{1}{(q-1)!}\pi'_{\sigma^{-1}\rho^{-1}(i)} = \sum_{\sigma \in \operatorname{Aut}(\mathcal{X})} a_\sigma \,\mathrm{FSC}_{\pi'}(\rho\sigma|i). \tag{60}$$

Therefore we can let R map $\rho\sigma$ to ρ with probability a_σ, for all $\sigma \in \mathrm{Aut}(\mathcal{X})$. This gives the desired degradation map R.

Step 2. We use the FSC mixture representation (Prop. 9). Suppose P maps X to (π_P, Z_P), and Q maps X to (π_Q, Z_Q). If

$$\pi = \mathbb{E}[\pi_Q|\pi_P = \pi] \qquad \forall \pi \in \mathcal{P}(\mathcal{X})/\operatorname{Aut}(\mathcal{X}), \tag{61}$$

then we can construct R by mapping π_Q to coupled π_P (randomly), and keeping the Z component.

Now define an FMS channel $\widetilde{P}$ whose π-component is $f(\pi_P)$, where

$$f(\pi) := \mathbb{E}[\pi_Q|\pi_P = \pi]. \tag{62}$$

Then by Step 1, $P \le_{\deg} \widetilde{P}$. By Step 2, $\widetilde{P} \le_{\deg} Q$. Therefore $P \le_{\deg} Q$. ∎

Appendix C. Proofs in Section 3

C.1. Proofs for low SNR case

We state a few properties of the function ϕ^L.

Lemma 23 *ϕ^L is 1-strongly convex on $\mathcal{P}([q])$.*

Proof

$$\nabla^2 \phi^L(\pi) = \operatorname{diag}\left(\pi^{-1} + \frac{1}{q}\pi^{-2}\right) \succeq I. \tag{63}$$

∎

Lemma 24 *$\Phi^L(\cdot) = C_{\mathrm{SKL}}(\cdot)$ is additive under $\star$-convolution.*

Proof The statement follows from additivity of SKL divergence under $\star$-convolution Külske and Formentin (2009). For completeness, we present a direct proof using (20).

By FSC mixture decomposition (Prop. 9), it suffices to prove that

$$\Phi^L(\mathrm{FSC}_\pi \star \mathrm{FSC}_{\pi'}) = \Phi^L(\mathrm{FSC}_\pi) + \Phi^L(\mathrm{FSC}_{\pi'}). \tag{64}$$

We have

$$\begin{aligned}
&\Phi^L(\mathrm{FSC}_\pi \star \mathrm{FSC}_{\pi'}) \\
&= \sum_{\tau\in\mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i\pi'_{\tau(i)}\right) \phi^L(\pi \star_\tau \pi') \\
&= \sum_{\tau\in\mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i\pi'_{\tau(i)}\right) \sum_{j\in[q]} \left((\pi \star_\tau \pi')_j - \frac{1}{q}\right) \log(\pi \star_\tau \pi')_j \\
&= \sum_{\tau\in\mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i\pi'_{\tau(i)}\right) \sum_{j\in[q]} \left(\frac{\pi_j\pi'_{\tau(j)}}{\sum_{k\in[q]}\pi_k\pi'_{\tau(k)}} - \frac{1}{q}\right) \log \frac{\pi_j\pi'_{\tau(j)}}{\sum_{k\in[q]}\pi_k\pi'_{\tau(k)}} \\
&= \sum_{\tau\in\mathrm{Aut}([q])} \frac{1}{(q-1)!} \sum_{j\in[q]} \left(\pi_j\pi'_{\tau(j)} - \frac{1}{q}\sum_{i\in[q]}\pi_i\pi'_{\tau(i)}\right) \log \frac{\pi_j\pi'_{\tau(j)}}{\sum_{k\in[q]}\pi_k\pi'_{\tau(k)}} \\
&= \sum_{\tau\in\mathrm{Aut}([q])} \frac{1}{(q-1)!} \sum_{j\in[q]} \left(\pi_j\pi'_{\tau(j)} - \frac{1}{q}\sum_{i\in[q]}\pi_i\pi'_{\tau(i)}\right) \log(\pi_j\pi'_{\tau(j)}) \\
&= \sum_{j\in[q]} (\pi_j - \frac{1}{q}) \log \pi_j + \sum_{j\in[q]} (\pi'_j - \frac{1}{q}) \log \pi'_j \\
&= \Phi^L(\mathrm{FSC}_\pi) + \Phi^L(\mathrm{FSC}_{\pi'}).
\end{aligned}$$

■

Condition (27) implies the desired contraction.

Proposition 25 *If* (27) *holds, then* $\lim_{k\to\infty}\left(\Phi^L(M_k) - \Phi^L(\widetilde{M}_k)\right) = 0.$

Proof Using BP equation and Lemma 24, we get

$$\Phi^L(M_{k+1}) = \mathbb{E}_b\left[b\Phi^L(M_k \circ P_\lambda) + \Phi^L(W)\right] = d\Phi^L(M_k \circ P_\lambda) + \Phi^L(W), \tag{65}$$

and the same holds with M replaced with $\widetilde{M}$.

To prove that

$$\Phi^L(M_{k+1}) - \Phi^L(\widetilde{M}_{k+1}) \le c\left(\Phi^L(M_k) - \Phi^L(\widetilde{M}_k)\right), \tag{66}$$

for some $c < 1$, it suffices to prove that

$$d\Phi^L(\mathrm{FSC}_\pi \circ P_\lambda) - c\Phi^L(\mathrm{FSC}_\pi) = d\phi^L\left(\lambda\pi + \frac{1-\lambda}{q}\right) - c\phi^L(\pi) \tag{67}$$

is concave in π.

Let $c = d\lambda^2 C^L(q,\lambda)$. Then for all $v \in \mathbb{1}^\perp \in \mathbb{R}^q$, we have

$$v^\top \nabla^2 \left(d\phi^L\left(\lambda\pi + \frac{1-\lambda}{q}\right) - c\phi^L(\pi)\right) v = d\lambda^2 f^L\left(\lambda\pi + \frac{1-\lambda}{q}, v\right) - cf^L(\pi, v) \le 0 \tag{68}$$

where the first step is because $v^\top \nabla^2 \phi^L(\pi) v = f^L(\pi, v)$ and the second step is by definition of C^L. Therefore contraction holds. ■

Theorem 14 follows from combining everything.

Proof [Proof of Theorem 14] By combining Prop. 25, Lemma 23, and Prop. 17. ■

C.2. Proofs for high SNR case

We state a few properties of the function ϕ^H.

Lemma 26 *ϕ^H is α-strongly concave on $\mathcal{P}([q])$ for some $\alpha > 0$.*

Proof For any $\pi \in \mathcal{P}([q])$ we have

$$\nabla^2 \phi^H(\pi) = \frac{1}{q-1}\left(\frac{1}{2}\left(\pi^{-1/2}\right)\left(\pi^{-1/2}\right)^\top - \frac{1}{2}\left(\sum_{i\in[q]} \pi_i^{1/2}\right) \operatorname{diag}\left(\pi^{-3/2}\right)\right). \tag{69}$$

So for any $v \in \mathbb{1}^\perp \subseteq \mathbb{R}^q$,

$$v^\top \nabla^2 \phi^H(\pi) v = \frac{1}{2(q-1)}\left(\left\langle \pi^{-1/2}, v\right\rangle^2 - \left(\sum_{i\in[q]} \pi_i^{1/2}\right)\left\langle \pi^{-3/2}, v^2\right\rangle\right). \tag{70}$$

Let us prove that

$$\frac{\left\langle \pi^{-1/2}, v\right\rangle^2}{\left(\sum_{i\in[q]} \pi_i^{1/2}\right)\left\langle \pi^{-3/2}, v^2\right\rangle} \le 1 - \frac{1}{\sqrt{q}}. \tag{71}$$

Performing change of variable $u = \pi^{-3/4} v$, LHS of (71) becomes

$$\frac{\left\langle \pi^{1/4}, u\right\rangle^2}{\left(\sum_{i\in[q]} \pi_i^{1/2}\right)\|u\|_2^2}. \tag{72}$$

We would like to maximize this expression over the hyperplane $\left\langle \pi^{3/4}, u\right\rangle = 0$. By geometric interpretation, maximum value is achieved at projection of $\pi^{1/4}$ onto the hyperplane, i.e.,

$$u = \pi^{1/4} - \frac{\left\langle \pi^{1/4}, \pi^{3/4}\right\rangle}{\|\pi^{3/4}\|_2^2}\pi^{3/4} = \pi^{1/4} - \frac{\pi^{3/4}}{\|\pi^{3/4}\|_2^2}, \tag{73}$$

at which (72) achieves value

$$\begin{aligned}
&\frac{\left(\sum_{i\in[q]}\pi_i^{1/2}-\frac{1}{\sum_{i\in[q]}\pi_i^{3/2}}\right)^2}{\left(\sum_{i\in[q]}\pi_i^{1/2}\right)\left(\sum_{i\in[q]}\pi_i^{1/2}-\frac{1}{\sum_{i\in[q]}\pi_i^{3/2}}\right)} \qquad (74)\\
&=\frac{\sum_{i\in[q]}\pi_i^{1/2}-\frac{1}{\sum_{i\in[q]}\pi_i^{3/2}}}{\sum_{i\in[q]}\pi_i^{1/2}}\\
&=1-\frac{1}{\left(\sum_{i\in[q]}\pi_i^{1/2}\right)\left(\sum_{i\in[q]}\pi_i^{3/2}\right)}\\
&\le 1-\frac{1}{\sqrt{q}}
\end{aligned}$$

where the last step is because

$$\sum_{i\in[q]}\pi_i^{1/2}\le\sqrt{q},\qquad \sum_{i\in[q]}\pi_i^{3/2}\le 1. \tag{75}$$

This finishes the proof of (71).

Therefore

$$\begin{aligned}
v^\top\nabla^2\phi^H(\pi)v &= \frac{1}{2(q-1)}\left(\left\langle\pi^{-1/2},v\right\rangle^2-\left(\sum_{i\in[q]}\pi_i^{1/2}\right)\left\langle\pi^{-3/2},v^2\right\rangle\right) \qquad (76)\\
&\le -\frac{1}{2(q-1)\sqrt{q}}\left(\sum_{i\in[q]}\pi_i^{1/2}\right)\left\langle\pi^{-3/2},v^2\right\rangle\\
&\le -\frac{1}{2(q-1)\sqrt{q}}\|v\|_2^2
\end{aligned}$$

where the second step is by (71), and the third step is because $\sum_{i\in[q]}\pi_i^{1/2}\ge 1$ and $\pi^{-3/2}\ge 1$. ■

Lemma 27 $\Phi^H(\cdot)=Z(\cdot)$ *is multiplicative under* $\star$*-convolution.*

Proof The statement follows from tensorization property of Hellinger distance (e.g., Polyanskiy and Wu (2023+)). For completeness, we present a direct proof using (20).

By FSC mixture decomposition (Prop. 9), it suffices to prove that

$$\Phi^H(\mathrm{FSC}_\pi\star\mathrm{FSC}_{\pi'})=\Phi^H(\mathrm{FSC}_\pi)+\Phi^H(\mathrm{FSC}_{\pi'}). \tag{77}$$

We have

$$
\begin{aligned}
&\Phi^H(\mathrm{FSC}_\pi \star \mathrm{FSC}_{\pi'}) \\
&= \sum_{\tau\in \mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i \pi'_{\tau(i)}\right) \phi^H(\pi \star_\tau \pi') \\
&= \sum_{\tau\in \mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i \pi'_{\tau(i)}\right) \frac{1}{q-1}\left(\left(\sum_{j\in[q]} \sqrt{(\pi\star_\tau \pi')_j}\right)^2 - 1\right) \\
&= \sum_{\tau\in \mathrm{Aut}([q])} \left(\frac{1}{(q-1)!}\sum_{i\in[q]} \pi_i \pi'_{\tau(i)}\right) \frac{1}{q-1}\left(\left(\sum_{j\in[q]} \sqrt{\frac{\pi_j\pi'_{\tau(j)}}{\sum_{k\in[q]}\pi_k\pi'_{\tau(k)}}}\right)^2 - 1\right) \\
&= \sum_{\tau\in \mathrm{Aut}([q])} \frac{1}{(q-1)!}\cdot\frac{1}{q-1}\left(\left(\sum_{j\in[q]}\sqrt{\pi_j\pi'_{\tau(j)}}\right)^2 - \sum_{i\in[q]}\pi_i\pi'_{\tau(i)}\right) \\
&= \sum_{\tau\in \mathrm{Aut}([q])} \frac{1}{(q-1)!}\cdot\frac{1}{q-1}\sum_{j\neq k\in[q]}\sqrt{\pi_j\pi'_{\tau(j)}\pi_k\pi'_{\tau(k)}} \\
&= \frac{1}{(q-1)^2}\left(\sum_{j\neq k\in[q]}\sqrt{\pi_j\pi_k}\right)\left(\sum_{j'\neq k'\in[q]}\sqrt{\pi'_{\tau(j')}\pi'_{\tau(k')}}\right) \\
&= \Phi^H(\mathrm{FSC}_\pi)\Phi^H(\mathrm{FSC}_{\pi'}).
\end{aligned}
$$

■

Lemma 28 *For any BMS channel P, we have*

$$Z(P) \le \sqrt{1 - C_{\chi^2}(P)}. \tag{78}$$

Proof Let Δ be the Δ-component of P. Then

$$Z(P) = \mathbb{E}[2\sqrt{\Delta(1-\Delta)}] \le \sqrt{1-\mathbb{E}[(1-2\Delta)^2]} = \sqrt{1-C_{\chi^2}(P)}. \tag{79}$$

The inequality step is by concavity of $\sqrt{\cdot}$. ■

Condition (28) implies the desired contraction.

Proposition 29 *If (28) or (29) holds, then* $\lim_{k\to\infty}\left(\Phi^H(M_k) - \Phi^H(\widetilde{M}_k)\right) = 0$.

Proof We treat the regular tree case and the Poisson tree case (almost) uniformly. For simplicity, in this proof, we use the following notation. Let $\mathbb{1}_R$ be 1 if we are working with regular trees, and 0 otherwise. Let $\mathbb{1}_P$ be 1 if we are working with Poisson trees, and 0 otherwise.

Using BP equation and Lemma 27, we have

$$\Phi^H(M_{k+1}) = \mathbb{E}_b\left[\left(\Phi^H(M_k\circ P_\lambda)\right)^b \Phi^H(W)\right] \tag{80}$$

and the same holds with M replaced with $\widetilde{M}$.

For $i \geq 0$, define

$$\Phi_i = \mathbb{E}_b \left[\prod_{j\in[b]} \left(\left(\Phi^H(\widetilde{M}_k \circ P_\lambda) \right)^{\mathbb{1}\{j\le i\}} \left(\Phi^H(M_k \circ P_\lambda) \right)^{\mathbb{1}\{j>i\}} \right) \Phi^H(W) \right]. \tag{81}$$

Fix $i \geq 1$. Let us prove that

$$\Phi_i - \Phi_{i-1} \le c_i \left(\Phi(\widetilde{M}_k) - \Phi(M_k) \right) \tag{82}$$

for some constant c_i to be determined later.

Note that

$$\begin{aligned}\Phi_i - \Phi_{i-1} = & \left(\Phi^H(\widetilde{M}_k \circ P_\lambda) - \Phi^H(M_k \circ P_\lambda) \right) \\ & \cdot \mathbb{E}_b \left[\mathbb{1}\{b \ge i\} \left(\Phi^H(\widetilde{M}_k \circ P_\lambda) \right)^{i-1} \left(\Phi^H(M_k \circ P_\lambda) \right)^{b-i} \right] \Phi^H(W).\end{aligned} \tag{83}$$

Note that $f^H(\pi, v) = -2(q-1)v^\top \nabla^2 \phi^H(\pi) v$. Therefore by definition of $C^H(q,\lambda)$, we have

$$\nabla^2 \left(\Phi^H(\mathrm{FSC}_\pi \circ P_\lambda) - \lambda^2 C^H(q,\lambda) \Phi^H(\mathrm{FSC}_\pi) \right) \succeq 0. \tag{84}$$

So by degradation,

$$\Phi^H(\widetilde{M}_k \circ P_\lambda) - \Phi^H(M_k \circ P_\lambda) \le \lambda^2 C^H(q,\lambda) \left(\Phi^H(\widetilde{M}_k) - \Phi^H(M_k) \right). \tag{85}$$

By Prop. 30 and Lemma 28, for any $\epsilon > 0$, for k large enough, we have

$$\Phi^H(M_k \circ P_\lambda) \le \left(1 - c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{d - \mathbb{1}_R} + \epsilon \right)_+^{1/2}. \tag{86}$$

Let

$$c_i = \lambda^2 \Phi^H(W) C^H(q,\lambda) \mathbb{E}_b \left[\mathbb{1}\{b \ge i\} \left(1 - c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{d - \mathbb{1}_R} + \epsilon \right)_+^{\frac{b-1}{2}} \right]. \tag{87}$$

Combining (83)(85)(86)(87), we get

$$\Phi_i - \Phi_{i-1} \le c_i \left(\Phi^H(\widetilde{M}_k) - \Phi^H(M_k) \right). \tag{88}$$

Let us compute sum of c_i. In the regular tree case, we have

$$\sum_{i\ge 1} c_i = d\lambda^2 \Phi^H(W) C^H(q,\lambda) \left(1 - c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{d-1} + \epsilon \right)_+^{\frac{d-1}{2}}. \tag{89}$$

For the Poisson tree case, we have

$$\sum_{i\ge 1} c_i = d\lambda^2 \Phi^H(W) C^H(q,\lambda) \exp\left(-d \left(1 - \left(1 - c^H(q,\lambda,d) \cdot \frac{d\lambda^2 - 1}{d} + \epsilon \right)_+^{1/2} \right) \right). \tag{90}$$

Note that $\epsilon > 0$ can be chosen to be arbitrarily small. Therefore in both cases, for any $\epsilon' > 0$, for k large enough, we can choose c_i such that (88) holds and

$$\sum_{i\ge 1} c_i \le d\lambda^2 \Phi^H(W) C^H(q,\lambda) \exp\left(-c^H(q,\lambda,d)\cdot \frac{d\lambda^2-1}{2} + \epsilon'\right). \tag{91}$$

Then

$$\begin{aligned}
&\Phi^H(\widetilde{M}_{k+1}) - \Phi^H(M_{k+1}) \\
&\le \left(\sum_{i\ge 1} c_i\right)\left(\Phi^H(\widetilde{M}_k) - \Phi^H(M_k)\right) \\
&\le d\lambda^2 \Phi^H(W) C^H(q,\lambda) \exp\left(-c^H(q,\lambda,d)\cdot \frac{d\lambda^2-1}{2} + \epsilon'\right)\left(\Phi^H(\widetilde{M}_k) - \Phi^H(M_k)\right).
\end{aligned} \tag{92}$$

Because (28) or (29) holds, we can choose $\epsilon' > 0$ small enough so that

$$d\lambda^2 \Phi^H(W) C^H(q,\lambda) \exp\left(-c^H(q,\lambda,d)\cdot \frac{d\lambda^2-1}{2} + \epsilon'\right) < 1. \tag{93}$$

This leads to the desired contraction. ∎

Theorem 15 follows from combining everything.

Proof [Proof of Theorem 15] Combine Prop. 29, Lemma 26, and Prop. 17. ∎

C.3. Majority decider

Proposition 30 *Consider the Potts model* $\mathrm{BOT}(q,\lambda,d)$ *or* $\mathrm{BOT}(q,\lambda,\mathrm{Pois}(d))$ *with leaf observations through a non-trivial FMS channel* U*. Let* M_k^U *denote the channel* $\sigma_\rho \to \nu_{L_k}$ *where* $\nu_v \sim U(\cdot|\sigma_v)$*. Assume that* $d\lambda^2 > 1$*. Then*

$$\lim_{k\to\infty} C_{\chi^2}\left((M_k^U \circ P_\lambda)^R\right) \ge \begin{cases} c(q,\lambda,d)\cdot \frac{d\lambda^2-1}{d-1} & \textit{Regular tree case,} \\ c(q,\lambda,d)\cdot \frac{d\lambda^2-1}{d} & \textit{Poisson tree case.}\end{cases} \tag{94}$$

where

$$c(q,\lambda,d) := \left(\frac{2}{q} + \frac{q-2}{q}\cdot \frac{d\lambda^2-1}{d\lambda-1}\right)^{-1}. \tag{95}$$

Furthermore, $c(q,\lambda,d) \ge 1$ *for all* $\lambda \in \left[-\frac{1}{q-1}, 1\right]$ *and* $d\lambda^2 > 1$*.*

Proof Let U^* be the reverse channel of U. Then the composition $U^* \circ U$ is a non-trivial ferromagnetic Potts channel. So there exists $\eta > 0$ such that $P_\eta \le_{\deg} U$. By replacing U with P_η (and using Lemma 12), we can wlog assume that $U = P_\eta$ for some $\eta > 0$.

Fix any embedding $\{\pm\} \subseteq [q]$. Let $e \in \mathbb{R}^q$ denote the vector with $e_+ = 1, e_- = -1, e_i = 0$ for $i \notin \{\pm\}$. Let

$$S_k = \sum_{v \in L_k} e_{\nu_v}. \tag{96}$$

We view S_k as a channel $[q] \to \mathbb{Z}$. By variational characterization of χ^2-divergence, we have

$$C_{\chi^2}((M_k^U \circ P_\lambda)^R) \geq \frac{(\mathbb{E}^+[S_k \circ P_\lambda])^2}{\mathbb{E}^+[S_k^2 \circ P_\lambda]} \tag{97}$$

where $\mathbb{E}^+$ denotes expectation conditioned on root label being $+$. Similarly, we use $\mathbb{E}^-$ to denote expectation conditioned on root label being $-$, and use $\mathbb{E}^0$ to denote expectation conditioned on root label being any label not $\pm$. Same for Var^+, Var^-, Var^0.

For simplicity, in this proof, we use the following notation. Let $\mathbb{1}_R$ be 1 if we are working with regular trees, and 0 otherwise. Let $\mathbb{1}_P$ be 1 if we are working with Poisson trees, and 0 otherwise. Clearly $\mathbb{1}_P + \mathbb{1}_R = 1$.

It is easy to see that

$$\mathbb{E}^i S_k = e_i \eta (d\lambda)^k. \tag{98}$$

Using variance decomposition formula, we have

$$\begin{aligned}\operatorname{Var}^i(S_{k+1}) &= \operatorname{Var}^i(\mathbb{E}[S_{k+1}|b]) + \mathbb{E}_b \operatorname{Var}^i(\mathbb{E}[S_{k+1}|b, \sigma_1, \ldots, \sigma_b]) \\ &\quad + \mathbb{E} \operatorname{Var}^i(S_{k+1}|b, \sigma_1, \ldots, \sigma_b)\end{aligned} \tag{99}$$

where $\sigma_1, \ldots, \sigma_b$ are labels of the children.

Let us compute each summand.

$$\operatorname{Var}^i(\mathbb{E}[S_{k+1}|b]) = \operatorname{Var}^i(b\lambda e_i \eta (d\lambda)^k) = e_i^2 d\lambda^2 \eta^2 (d\lambda)^{2k} \mathbb{1}_P, \tag{100}$$

$$\mathbb{E}_b \operatorname{Var}^i(\mathbb{E}[S_{k+1}|b, \sigma_1, \ldots, \sigma_b]) = d\eta^2 (d\lambda)^{2k} \operatorname{Var}_{j \sim P_\lambda(\cdot|i)}(e_j), \tag{101}$$

$$\mathbb{E} \operatorname{Var}^i(S_{k+1}|b, \sigma_1, \ldots, \sigma_b) = d\mathbb{E}_{j \sim P_\lambda(\cdot|i)}[\operatorname{Var}^j(S_k)]. \tag{102}$$

We have $\operatorname{Var}^-(S_k) = \operatorname{Var}^+(S_k)$ and

$$\begin{aligned}\operatorname{Var}^+(S_{k+1}) &= d\eta^2 (d\lambda)^{2k} \left(\left(\lambda + \frac{1-\lambda}{q} \cdot 2 \right) - \lambda^2 \mathbb{1}_R \right) \\ &\quad + d \left(\left(\lambda + \frac{1-\lambda}{q} \cdot 2 \right) \operatorname{Var}^+(S_k) + \frac{1-\lambda}{q} \cdot (q-2) \operatorname{Var}^0(S_k) \right),\end{aligned} \tag{103}$$

$$\begin{aligned}\operatorname{Var}^0(S_{k+1}) &= d\eta^2 (d\lambda)^{2k} \left(\frac{1-\lambda}{q} \cdot 2 \right) \\ &\quad + d \left(\frac{1-\lambda}{q} \cdot 2 \operatorname{Var}^+(S_k) + \left(\lambda + \frac{1-\lambda}{q} \cdot (q-2) \right) \operatorname{Var}^0(S_k) \right).\end{aligned} \tag{104}$$

By computing linear combinations of (103)(104), we get

$$\begin{aligned}\operatorname{Var}^+(S_{k+1}) - \operatorname{Var}^0(S_{k+1}) &= d\lambda\left(\operatorname{Var}^+(S_k) - \operatorname{Var}^0(S_k)\right) \\ &\quad + d\eta^2(d\lambda)^{2k}(\lambda - \lambda^2\mathbb{1}_R).\end{aligned} \tag{105}$$

$$\begin{aligned}\operatorname{Var}^+(S_{k+1}) + \frac{q-2}{2}\operatorname{Var}^0(S_{k+1}) &= d\left(\operatorname{Var}^+(S_k) + \frac{q-2}{2}\operatorname{Var}^0(S_{k+1})\right) \\ &\quad + d\eta^2(d\lambda)^{2k}(1 - \lambda^2\mathbb{1}_R).\end{aligned} \tag{106}$$

Solving (105)(106) we get

$$\begin{aligned}&\operatorname{Var}^+(S_k) - \operatorname{Var}^0(S_k) \\ &= \left(\operatorname{Var}^+(S_0) - \operatorname{Var}^0(S_0)\right)(d\lambda)^k + \sum_{1\le i\le k} d\eta^2(d\lambda)^{2i-2}(d\lambda)^{k-i}(\lambda - \lambda^2\mathbb{1}_R) \\ &= O\left((d\lambda)^k\right) + d\eta^2(d\lambda)^{k-1}\frac{(d\lambda)^k - 1}{d\lambda - 1}(\lambda - \lambda^2\mathbb{1}_R) \\ &= (1+o(1))\frac{1-\lambda\mathbb{1}_R}{d\lambda - 1}\eta^2(d\lambda)^{2k}\end{aligned} \tag{107}$$

and

$$\begin{aligned}&\operatorname{Var}^+(S_k) + \frac{q-2}{2}\operatorname{Var}^0(S_k) \\ &= \left(\operatorname{Var}^+(S_0) + \frac{q-2}{2}\operatorname{Var}^0(S_0)\right)d^k + \sum_{1\le i\le k} d\eta^2(d\lambda)^{2i-2}d^{k-i}(1 - \lambda^2\mathbb{1}_R) \\ &= O\left(d^k\right) + \eta^2 d^k\frac{(d\lambda^2)^k - 1}{d\lambda^2 - 1}(1 - \lambda^2\mathbb{1}_R) \\ &= (1+o(1))\frac{1-\lambda^2\mathbb{1}_R}{d\lambda^2 - 1}\eta^2(d\lambda)^{2k}.\end{aligned} \tag{108}$$

Combining (107)(108) we have

$$\operatorname{Var}^+(S_k) = (1+o(1))\frac{2}{q}\left(\frac{1-\lambda^2\mathbb{1}_R}{d\lambda^2 - 1} + \frac{1-\lambda\mathbb{1}_R}{d\lambda - 1}\cdot\frac{q-2}{2}\right)\eta^2(d\lambda)^{2k}. \tag{109}$$

Now we compute moments of $S_k \circ P_\lambda$.

$$\mathbb{E}^+[S_k \circ P_\lambda] = \lambda\eta(d\lambda)^k. \tag{110}$$

$$
\begin{aligned}
\mathbb{E}^+[S_k^2 \circ P_\lambda] &= \left(\lambda + \frac{1-\lambda}{q}\cdot 2\right)\mathbb{E}^+[S_k^2] + \frac{1-\lambda}{q}\cdot (q-2)\mathbb{E}^0[S_k^2] \\
&= \left(\lambda + \frac{1-\lambda}{q}\cdot 2\right)\left(\mathrm{Var}^+(S_k) + (\mathbb{E}^+ S_k)^2\right) + \frac{1-\lambda}{q}\cdot (q-2)\,\mathrm{Var}^0(S_k) \\
&= \lambda(1+o(1))\frac{2}{q}\left(\frac{1-\lambda^2\mathbb{1}_R}{d\lambda^2-1} + \frac{1-\lambda\mathbb{1}_R}{d\lambda-1}\cdot\frac{q-2}{2}\right)\eta^2(d\lambda)^{2k} \\
&+ \left(\lambda + \frac{1-\lambda}{q}\cdot 2\right)\eta^2(d\lambda)^{2k} + \frac{1-\lambda}{q}\cdot 2\cdot (1+o(1))\frac{1-\lambda^2\mathbb{1}_R}{d\lambda^2-1}\eta^2(d\lambda)^{2k} \\
&= (1+o(1))\left(\frac{2}{q}\cdot\frac{d\lambda^2-\lambda^2\mathbb{1}_R}{d\lambda^2-1} + \lambda\cdot\frac{q-2}{q}\cdot\frac{d\lambda-\lambda\mathbb{1}_R}{d\lambda-1}\right)\eta^2(d\lambda)^{2k} \\
&= (1+o(1))c(q,\lambda,d)^{-1}\frac{d-\mathbb{1}_R}{d\lambda^2-1}\lambda^2\eta^2(d\lambda)^{2k}.
\end{aligned}
\tag{111}
$$

Finally,

$$
C_{\chi^2}((M_k^U \circ P_\lambda)^R) \ge \frac{\left(\mathbb{E}^+[S_k \circ P_\lambda]\right)^2}{\mathbb{E}^+[S_k^2 \circ P_\lambda]} = (1+o(1))c(q,\lambda,d)\cdot\frac{d\lambda^2-1}{d-\mathbb{1}_R}. \tag{112}
$$

■

C.4. Bounds on key constants

In this section we prove bounds on key constants used in Theorem 14 and 15.

Proposition 31 *For $q \in \mathbb{Z}_{\ge 2}$, $\lambda \in \left[-\frac{1}{q-1}, 1\right]$, we have $C^L(q,\lambda) \le q^2$, where $C^L(q,\lambda)$ is defined in (22).*

Proof We have

$$
\begin{aligned}
&\frac{\left\langle \left(\lambda\pi + \frac{1-\lambda}{q}\right)^{-1} + \frac{1}{q}\left(\lambda\pi + \frac{1-\lambda}{q}\right)^{-2}, v^2\right\rangle}{\left\langle \pi^{-1} + \frac{1}{q}\pi^{-2}, v^2\right\rangle} \\
\le\ & \max\left\{\frac{\left\langle \left(\lambda\pi + \frac{1-\lambda}{q}\right)^{-1}, v^2\right\rangle}{\langle \pi^{-1}, v^2\rangle}, \frac{\left\langle \left(\lambda\pi + \frac{1-\lambda}{q}\right)^{-2}, v^2\right\rangle}{\langle \pi^{-2}, v^2\rangle}\right\} \\
\le\ & \max\{q, q^2\} = q^2.
\end{aligned}
\tag{113}
$$

where the second step is by Lemma 33. ■

Proposition 32 *For $q \in \mathbb{Z}_{\ge 2}$, $\lambda \in \left[-\frac{1}{q-1}, 1\right]$, we have $C^H(q,\lambda) \le q^{5/2}$, where $C^H(q,\lambda)$ is defined in (24).*

Proof By (71),

$$f(\pi, v) \geq \frac{1}{\sqrt{q}} \left(\sum_{i \in [q]} \pi_i^{1/2} \right) \left\langle \pi^{-3/2}, v^2 \right\rangle \geq \frac{1}{\sqrt{q}} \left\langle \pi^{-3/2}, v^2 \right\rangle . \tag{114}$$

On the other hand,

$$\begin{aligned} f\left(\lambda\pi + \frac{1-\lambda}{q}, v\right) &\leq \left(\sum_{i \in [q]} \left(\lambda \pi_i + \frac{1-\lambda}{q} \right)^{1/2} \right) \left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-3/2}, v^2 \right\rangle \\ &\leq \sqrt{q} \cdot \left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-3/2}, v^2 \right\rangle . \end{aligned} \tag{115}$$

Combining (114)(115) we get

$$\frac{f\left(\lambda\pi + \frac{1-\lambda}{q}, v\right)}{f(\pi, v)} \leq q \cdot \frac{\left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-3/2}, v^2 \right\rangle}{\left\langle \pi^{-3/2}, v^2 \right\rangle} \leq q^{5/2} \tag{116}$$

where the last step is by Lemma 33. ■

The following lemma is the crucial step in the proof of Prop. 31 and 32.

Lemma 33 *For $q \in \mathbb{Z}_{\geq 2}$, $\alpha \in \mathbb{R}_{\geq 1}$, $\lambda \in \left[-\frac{1}{q-1}, 1\right]$, we have*

$$\sup_{\substack{\pi \in \mathcal{P}([q]) \\ v \in \mathbb{1}^\perp \subseteq \mathbb{R}^q}} \frac{\left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-\alpha}, v^2 \right\rangle}{\left\langle \pi^{-\alpha}, v^2 \right\rangle} \leq q^\alpha . \tag{117}$$

Proof We prove the ferromagnetic case ($\lambda \in [0, 1]$) and antiferromagnetic case ($\lambda \in \left[-\frac{1}{q-1}, 0\right]$) separately.

Ferromagnetic case ($\lambda \in [0, 1]$)**.** In this case, we have

$$\lambda x + \frac{1-\lambda}{q} \geq \frac{x}{q} \tag{118}$$

for all $x \in [0, 1]$. Therefore

$$\left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-\alpha}, v^2 \right\rangle \leq \left\langle \left(\frac{\pi}{q} \right)^{-\alpha}, v^2 \right\rangle = q^\alpha \left\langle \pi^{-\alpha}, v^2 \right\rangle . \tag{119}$$

Note that we did not use the assumption that $v \in \mathbb{1}^\perp$.

Antiferromagnetic case $\left(\lambda \in \left[-\frac{1}{q-1}, 0\right]\right)$**.** We would like to prove that

$$q^\alpha \left\langle \pi^{-\alpha}, v^2 \right\rangle - \left\langle \left(\lambda\pi + \frac{1-\lambda}{q} \right)^{-\alpha}, v^2 \right\rangle =: \left\langle b, v^2 \right\rangle \tag{120}$$

is non-negative for all $\pi \in \mathcal{P}([q])$, $v \in \mathbb{1}^\perp$, where

$$b := \left(\frac{\pi}{q}\right)^{-\alpha} - \left(\lambda\pi + \frac{1-\lambda}{q}\right)^{-\alpha}. \tag{121}$$

Step 1. We fix $\pi \in \mathcal{P}([q])$ and determine the optimal $v \in \mathbb{1}^\perp$ to plug in (120), reducing the statement to one involving π only.

If $\lambda x + \frac{1-\lambda}{q} \le \frac{x}{q}$ for some $x \in [0,1]$, then $x \ge \frac{1-\lambda}{1-\lambda q} \ge \frac{q}{2q-1} > \frac{1}{2}$. So there exists at most one i such that $\lambda\pi_i + \frac{1-\lambda}{q} \le \frac{\pi_i}{q}$ (equivalently, $b_i \le 0$). We can wlog assume that $\pi_1 \ge \pi_2 \ge \cdots \ge \pi_q$. Then we know $\lambda\pi_i + \frac{1-\lambda}{q} > \frac{\pi_i}{q}$ (equivalently, $b_i > 0$) for all $2 \le i \le q$. If $b_1 \ge 0$, then $\langle b, v^2\rangle$ is non-negative for all v and we are done. Therefore, it remains to consider the case $b_1 < 0$.

If $v_1 = 0$, then $\langle b, v^2\rangle$ is non-negative. Therefore we can assume $v_1 \ne 0$. By rescaling, we can assume that $v_1 = 1$. So $v_2 + \cdots + v_q = -1$. Because $b_2, \ldots, b_q$ are all positive, to minimize $\sum_{2\le i\le q} b_i v_i^2$ under linear constraint $v_2 + \cdots + v_q = -1$, the optimal choice is $v_i = -b_i^{-1}Z^{-1}$ for $2 \le i \le q$ where $Z := \sum_{2\le i\le q} b_i^{-1}$. For this choice of v, we have

$$\langle b, v^2\rangle = b_1 + \sum_{2\le i\le q} b_i \cdot (-b_i^{-1}Z^{-1})^2 = b_1 + Z^{-1}. \tag{122}$$

Therefore, it remains to prove

$$Z \le (-b_1)^{-1} \tag{123}$$

where $\pi_1 \ge \cdots \ge \pi_q$, $b_1 < 0$, and $b_2, \ldots, b_q > 0$.

Step 2. We reduce to the case where $\pi_3 = \cdots = \pi_q = 0$. Note that

$$Z = \sum_{2\le i\le q} b_i^{-1} = \sum_{2\le i\le q} \left(\left(\frac{\pi_i}{q}\right)^{-\alpha} - \left(\lambda\pi_i + \frac{1-\lambda}{q}\right)^{-\alpha}\right)^{-1}. \tag{124}$$

By Lemma 34, for fixed π_1, the optimal choice (for maximizing Z) of $\pi_2, \ldots, \pi_q$ is $\pi_3 = \cdots = \pi_q = 0$.

Write $\pi_1 = 1 - x$, $\pi_2 = x$ where $x \in \left[0, \frac{\lambda - \lambda q}{1-\lambda q}\right]$. Then

$$b_1 = \left(\frac{1-x}{q}\right)^{-\alpha} - \left(\lambda(1-x) + \frac{1-\lambda}{q}\right)^{-\alpha}, \tag{125}$$

$$Z = \left(\left(\frac{x}{q}\right)^{-\alpha} - \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha}\right)^{-1}. \tag{126}$$

By rearranging terms in (123), we reduce to proving

$$\left(\frac{x}{q}\right)^{-\alpha} + \left(\frac{1-x}{q}\right)^{-\alpha} \ge \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha} + \left(\lambda(1-x) + \frac{1-\lambda}{q}\right)^{-\alpha} \tag{127}$$

for $q \in \mathbb{Z}_{\ge 2}$, $\alpha \in \mathbb{R}_{\ge 1}$, $\lambda \in \left[-\frac{1}{q-1}, 0\right]$, $x \in \left[0, \frac{\lambda-\lambda q}{1-\lambda q}\right]$.

Step 3. Let $g_{q,\alpha,x}(\lambda) := \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha} + \left(\lambda(1-x) + \frac{1-\lambda}{q}\right)^{-\alpha}$ be the RHS of (127). Then

$$\begin{aligned} g''_{q,\alpha,x}(\lambda) &= \alpha(\alpha+1)\left(x - \frac{1}{q}\right)^2 \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha-2} \\ &+ \alpha(\alpha+1)\left(1 - x - \frac{1}{q}\right)^2 \left(\lambda(1-x) + \frac{1-\lambda}{q}\right)^{-\alpha-2} \\ &> 0. \end{aligned}$$

So $g_{q,\alpha,x}$ is convex in λ. Therefore it suffices to verify (127) for $\lambda = 0$ and $\lambda = -\frac{1}{q-1}$. When $\lambda = 0$, we have

$$g_{q,\alpha,x}(\lambda) = \left(\frac{1}{q}\right)^{-\alpha} + \left(\frac{1}{q}\right)^{-\alpha} \le \left(\frac{x}{q}\right)^{-\alpha} + \left(\frac{1-x}{q}\right)^{-\alpha}. \tag{128}$$

When $\lambda = -\frac{1}{q-1}$, we have

$$g_{q,\alpha,x}(\lambda) = \left(\frac{1-x}{q-1}\right)^{-\alpha} + \left(\frac{x}{q-1}\right)^{-\alpha} \le \left(\frac{x}{q}\right)^{-\alpha} + \left(\frac{1-x}{q}\right)^{-\alpha}. \tag{129}$$

This finishes the proof. ■

Lemma 34 *For $q \in \mathbb{Z}_{\ge 2}$, $\alpha \in \mathbb{R}_{\ge 1}$, $\lambda \in \left[-\frac{1}{q-1}, 0\right]$, the function*

$$f(x) := \left(\left(\frac{x}{q}\right)^{-\alpha} - \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha}\right)^{-1}. \tag{130}$$

is convex in $x \in [0, \frac{1-\lambda}{1-\lambda q}]$.

Proof Let $g(x) = \frac{1}{f(x)}$. Then

$$f''(x) = \frac{2g'(x)^2 - g(x)g''(x)}{g(x)^3}. \tag{131}$$

It suffices to prove that

$$2g'(x)^2 - g(x)g''(x) \ge 0. \tag{132}$$

We have

$$\begin{aligned} 2g'(x)^2 - g(x)g''(x) = 2\alpha^2 &\left(q^{-1}\left(\frac{x}{q}\right)^{-\alpha-1} - \lambda\left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha-1}\right)^2 \\ &- \left(\left(\frac{x}{q}\right)^{-\alpha} - \left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha}\right) \\ &\cdot \alpha(\alpha+1)\left(q^{-2}\left(\frac{x}{q}\right)^{-\alpha-2} - \lambda^2\left(\lambda x + \frac{1-\lambda}{q}\right)^{-\alpha-2}\right). \end{aligned} \tag{133}$$

Write $u = \frac{x}{q}$, $v = \lambda x + \frac{1-\lambda}{q}$, $c = q\lambda$. Then we have $0 \le u \le v \le 1$ and $-\frac{q}{q-1} \le c \le 0$. It suffices to prove

$$2\alpha^2(u^{-\alpha-1} - cv^{-\alpha-1})^2 - \alpha(\alpha+1)(u^{-\alpha} - v^{-\alpha})(u^{-\alpha-2} - c^2 v^{-\alpha-2}) \ge 0. \tag{134}$$

We have

$$\begin{aligned}
&2\alpha^2(u^{-\alpha-1} - cv^{-\alpha-1})^2 - \alpha(\alpha+1)(u^{-\alpha} - v^{-\alpha})(u^{-\alpha-2} - c^2 v^{-\alpha-2}) \\
&\ge \alpha(\alpha+1)((u^{-\alpha-1} - cv^{-\alpha-1})^2 - (u^{-\alpha} - v^{-\alpha})(u^{-\alpha-2} - c^2 v^{-\alpha-2})) \\
&= \alpha(\alpha+1)u^{-\alpha}v^{-\alpha}(u^{-1} - cv^{-1})^2 \\
&\ge 0.
\end{aligned} \tag{135}$$

This finishes the proof. ■

Appendix D. Proofs in Section 4

D.1. Local subadditivity

In this section we prove Lemma 20. We first prove the special case of FSCs.

Lemma 35 *Fix $q \in \mathbb{Z}_{\ge 2}$. For any $\epsilon > 0$ and $\pi, \pi' \in \mathcal{P}([q])/\operatorname{Aut}([q])$ with $C_{\chi^2}(\mathrm{FSC}_{\pi'}) \le \epsilon$, we have*

$$C_{\chi^2}(\mathrm{FSC}_\pi \star \mathrm{FSC}_{\pi'}) \le (1 + O_q(\epsilon^{1/2}))(C_{\chi^2}(\mathrm{FSC}_\pi) + C_{\chi^2}(\mathrm{FSC}_{\pi'})). \tag{136}$$

Proof [Proof of Lemma 20 given Lemma 35] Let π_P (resp. π_Q) be the π-component of P (resp. Q).

Because the constant does not depend on Q, it suffices to prove the case where Q is an FSC, i.e., π_Q is fixed.

If $C_{\chi^2}(Q) \le \epsilon^{2/5}$, then by Lemma 35, we have

$$\begin{aligned}
C_{\chi^2}(P \star Q) &= \mathbb{E}_{\pi_P} C_{\chi^2}(\mathrm{FSC}_{\pi_P} \star \mathrm{FSC}_{\pi_Q}) \\
&\le \mathbb{E}_{\pi_P}\left[(1 + O_q(\epsilon^{1/5}))(C_{\chi^2}(\mathrm{FSC}_{\pi_P}) + C_{\chi^2}(\mathrm{FSC}_{\pi_Q}))\right] \\
&= (1 + O_q(\epsilon^{1/5}))(C_{\chi^2}(P) + C_{\chi^2}(Q)).
\end{aligned} \tag{137}$$

In the following we assume that $C_{\chi^2}(Q) > \epsilon^{2/5}$. By Markov's inequality, we have

$$\mathbb{P}\left[C_{\chi^2}(\mathrm{FSC}_{\pi_P}) \ge \epsilon^{2/5}\right] \le \epsilon^{3/5}. \tag{138}$$

Write

$$\begin{aligned}
C_{\chi^2}(P \star Q) = \mathbb{E}_{\pi_P}&\left[C_{\chi^2}(\mathrm{FSC}_{\pi_P} \star \mathrm{FSC}_{\pi_Q}) \mathbb{1}\{C_{\chi^2}(\mathrm{FSC}_{\pi_P}) \le \epsilon^{2/5}\}\right] \\
&+ \mathbb{E}_{\pi_P}\left[C_{\chi^2}(\mathrm{FSC}_{\pi_P} \star \mathrm{FSC}_{\pi_Q}) \mathbb{1}\{C_{\chi^2}(\mathrm{FSC}_{\pi_P}) > \epsilon^{2/5}\}\right] \\
=:&\ L + R.
\end{aligned} \tag{139}$$

For L, by Lemma 35, we have

$$L \le (1+O_q(\epsilon^{1/5}))\mathbb{E}_{\pi_P}\left[(C_{\chi^2}(\mathrm{FSC}_{\pi_P})+C_{\chi^2}(\mathrm{FSC}_{\pi_Q}))\mathbb{1}\{C_{\chi^2}(\mathrm{FSC}_{\pi_P})\le \epsilon^{2/5}\}\right] \\ \le (1+O_q(\epsilon^{1/5}))(C_{\chi^2}(P)+C_{\chi^2}(Q)). \tag{140}$$

For R, by (138) and the assumption that $C_{\chi^2}(Q) > \epsilon^{2/5}$, we have

$$\mathbb{E}_{\pi_P}\left[C_{\chi^2}(\mathrm{FSC}_{\pi_P}\star\mathrm{FSC}_{\pi_Q})\mathbb{1}\{C_{\chi^2}(\mathrm{FSC}_{\pi_P})>\epsilon^{2/5}\}\right] \le O_q(\epsilon^{3/5}) \le O_q(\epsilon^{1/5})C_{\chi^2}(Q). \tag{141}$$

Combining (140) and (141) we finish the proof. ■

Proof [Proof of Lemma 35] Because the statement is monotone in ϵ, we can wlog assume that $C_{\chi^2}(\mathrm{FSC}_{\pi'}) = \epsilon$.

Let $\pi'_i = \frac{1+\epsilon_i}{q}$. Then

$$\sum_i \epsilon_i = 0, \quad C_{\chi^2}(\mathrm{FSC}_{\pi'}) = \frac{1}{q}\sum_i \epsilon_i^2 = \epsilon. \tag{142}$$

By (20),

$$C_{\chi^2}(\mathrm{FSC}_\pi\star\mathrm{FSC}_{\pi'}) = q\sum_{\tau\in S_q}\frac{1}{(q-1)!}\cdot\frac{\sum_i \pi_i^2\pi'^2_{\tau(i)}}{\sum_i \pi_i\pi'_{\tau(i)}} - 1 \\ = q\sum_{\tau\in S_q}\frac{1}{q!}\cdot\frac{\sum_i \pi_i^2(1+\epsilon_{\tau(i)})^2}{1+\sum_i \pi_i\epsilon_{\tau(i)}} - 1. \tag{143}$$

Recall the following basic equality.

$$\frac{1}{1+x} = 1 - x + x^2 - \frac{x^3}{1+x}. \tag{144}$$

We apply (144) with $x = \sum_i \pi_i\epsilon_{\tau(i)}$. Because $|x| = O_q(\epsilon^{1/2})$, we have

$$\left|\frac{x^3}{1+x}\right| = O_q(\epsilon^{3/2}). \tag{145}$$

So

$$\frac{\sum_i \pi_i^2(1+\epsilon_{\tau(i)})^2}{1+\sum_i \pi_i\epsilon_{\tau(i)}} \\ = \left(\sum_i \pi_i^2(1+\epsilon_{\tau(i)})^2\right)\left(1-x+x^2-\frac{x^3}{1+x}\right) \\ \le \left(\sum_i \pi_i^2(1+\epsilon_{\tau(i)})^2\right)\left(1-\sum_i \pi_i\epsilon_{\tau(i)}+\left(\sum_i \pi_i\epsilon_{\tau(i)}\right)^2+O_q(\epsilon^{3/2})\right) \\ \le \left(\sum_i \pi_i^2(1+\epsilon_{\tau(i)})^2\right)\left(1-\sum_i \pi_i\epsilon_{\tau(i)}+\left(\sum_i \pi_i\epsilon_{\tau(i)}\right)^2\right)+O_q(\epsilon^{3/2}), \tag{146}$$

where the last step is by

$$\sum_i \pi_i^2 (1+\epsilon_{\tau(i)})^2 = O(1). \tag{147}$$

Let us expand the first summand in (146).

$$\begin{aligned}
&\left(\sum_i \pi_i^2 (1+\epsilon_{\tau(i)})^2\right)\left(1-\sum_i \pi_i\epsilon_{\tau(i)} + \left(\sum_i \pi_i \epsilon_{\tau(i)}\right)^2\right) \\
&= \left(\sum_i \pi_i^2 + 2\sum_i \pi_i^2 \epsilon_{\tau(i)} + \sum_i \pi_i^2 \epsilon_{\tau(i)}^2\right)\left(1-\sum_i \pi_i\epsilon_{\tau(i)} + \left(\sum_i \pi_i \epsilon_{\tau(i)}\right)^2\right) \\
&=: (① + ② + ③)(1 - ④ + ⑤) \\
&= ① - ①④ + ①⑤ + ② - ②④ + ②⑤ + ③ - ③④ + ③⑤.
\end{aligned} \tag{148}$$

Note that we have the following loose bounds:

$$① = O_q(1), \quad |②| = O_q(\epsilon^{1/2}), \quad ③ \le O_q(\epsilon), \quad |④| \le O_q(\epsilon^{1/2}), \quad ⑤ \le O_q(\epsilon). \tag{149}$$

Let us study every term under $\sum_{\tau\in S_q} \frac{1}{q!}$. For simplicity, write

$$A = \sum_i \pi_i^2 = \frac{1}{q}\left(1 + C_{\chi^2}(\mathrm{FSC}_\pi)\right). \tag{150}$$

①:

$$\sum_{\tau\in S_q} \frac{1}{q!} \cdot ① = \sum_{\tau\in S_q} \frac{1}{q!} \sum_i \pi_i^2 = A. \tag{151}$$

①④:

$$\sum_{\tau\in S_q} \frac{1}{q!} \cdot ①④ = \sum_{\tau\in S_q} \frac{1}{q!} \left(\sum_i \pi_i^2\right)\left(\sum_i \pi_i \epsilon_{\tau(i)}\right) = 0. \tag{152}$$

①⑤:

$$
\begin{aligned}
\sum_{\tau \in S_q} \frac{1}{q!} \cdot ①⑤ &= \sum_{\tau \in S_q} \frac{1}{q!} \left(\sum_i \pi_i^2 \right) \left(\sum_i \pi_i \epsilon_{\tau(i)} \right)^2 \\
&= A \sum_{i,j} \sum_{\tau \in S_q} \frac{1}{q!} \cdot \pi_i \pi_j \epsilon_{\tau(i)} \epsilon_{\tau(j)} \\
&= A \sum_{i,j} \epsilon_i \epsilon_j \sum_{\tau \in S_q} \frac{1}{q!} \cdot \pi_{\tau(i)} \pi_{\tau(j)} \\
&= A \sum_{i,j} \epsilon_i \epsilon_j \begin{cases} \frac{1}{q} \sum_k \pi_k^2 & i = j, \\ \frac{1}{q(q-1)} \left(1 - \sum_k \pi_k^2\right) & i \neq j, \end{cases} \\
&= A \left(\sum_i \epsilon_i^2 \cdot \frac{1}{q} \sum_k \pi_k^2 + \sum_i \epsilon_i (-\epsilon_i) \cdot \frac{1}{q(q-1)} \left(1 - \sum_k \pi_k^2 \right) \right) \\
&= A \cdot \frac{1}{q} \sum_i \epsilon_i^2 \left(\sum_k \pi_k^2 - \frac{1}{q-1} \left(1 - \sum_k \pi_k^2 \right) \right) \\
&= \epsilon A \cdot \frac{qA - 1}{q - 1}.
\end{aligned} \tag{153}
$$

②:

$$
\sum_{\tau \in S_q} \frac{1}{q!} \cdot ② = \sum_{\tau \in S_q} \frac{1}{q!} \cdot 2 \sum_i \pi_i^2 \epsilon_{\tau(i)} = 0. \tag{154}
$$

②④:

$$
\begin{aligned}
\sum_{\tau \in S_q} \frac{1}{q!} \cdot ②④ &= \sum_{\tau \in S_q} \frac{1}{q!} \left(2 \sum_i \pi_i^2 \epsilon_{\tau(i)} \right) \left(\sum_i \pi_i \epsilon_{\tau(i)} \right) \\
&= \sum_{i,j} \sum_{\tau \in S_q} \frac{1}{q!} \cdot 2 \pi_i^2 \pi_j \epsilon_{\tau(i)} \epsilon_{\tau(j)} \\
&= \sum_{i,j} 2 \epsilon_i \epsilon_j \sum_{\tau \in S_q} \frac{1}{q!} \cdot \pi_{\tau(i)}^2 \pi_{\tau(j)} \\
&= \sum_{i,j} 2 \epsilon_i \epsilon_j \begin{cases} \frac{1}{q} \sum_k \pi_k^3 & i = j, \\ \frac{1}{q(q-1)} \sum_k \pi_k^2 (1 - \pi_k) & i \neq j, \end{cases} \\
&= 2 \sum_i \epsilon_i^2 \cdot \frac{1}{q} \sum_k \pi_k^3 + 2 \sum_i \epsilon_i (-\epsilon_i) \cdot \frac{1}{q(q-1)} \sum_k \pi_k^2 (1 - \pi_k) \\
&= 2 \cdot \frac{1}{q} \sum_i \epsilon_i^2 \left(\sum_k \pi_k^3 - \frac{1}{q-1} \sum_k \pi_k^2 (1 - \pi_k) \right) \\
&= 2\epsilon \cdot \frac{1}{q-1} \left(q \sum_k \pi_k^3 - \sum_k \pi_k^2 \right) \geq 0,
\end{aligned} \tag{155}
$$

where the last step is by

$$\sum_k \pi_k^3 = \left(\sum_k \pi_k^3\right)\left(\sum_k \pi_k\right) \ge \left(\sum_k \pi_k^2\right)^2 \ge \frac{1}{q}\left(\sum_k \pi_k^2\right). \tag{156}$$

②⑤: By (149),

$$\left|\sum_{\tau\in S_q} \frac{1}{q!}\cdot ②⑤\right| = O_q(\epsilon^{3/2}). \tag{157}$$

③:

$$\sum_{\tau\in S_q} \frac{1}{q!}\cdot ③ = \sum_{\tau\in S_q} \frac{1}{q!}\left(\sum_i \pi_i^2 \epsilon_{\tau(i)}^2\right) = \sum_i \pi_i^2 \cdot \frac{1}{q}\sum_j \epsilon_j^2 = \epsilon A. \tag{158}$$

③④: By (149),

$$\left|\sum_{\tau\in S_q} \frac{1}{q!}\cdot ③④\right| = O_q(\epsilon^{3/2}). \tag{159}$$

③⑤: By (149),

$$\left|\sum_{\tau\in S_q} \frac{1}{q!}\cdot ③⑤\right| = O_q(\epsilon^{2}). \tag{160}$$

Plugging (151) - (160) into (148)(146)(143), we get

$$\begin{aligned} C_{\chi^2}(\mathrm{FSC}_\pi \star \mathrm{FSC}_{\pi'}) &\le q\left(A + \epsilon A\cdot \frac{qA-1}{q-1} + \epsilon A + O_q(\epsilon^{3/2})\right) - 1 \\ &= (qA-1)\left(1 + \frac{q\epsilon A}{q-1} + \epsilon\right) + \epsilon + O_q(\epsilon^{3/2}) \\ &= \left(1 + \frac{q\epsilon A}{q-1} + \epsilon\right) C_{\chi^2}(\mathrm{FSC}_\pi) + (1 + O_q(\epsilon^{1/2}))C_{\chi^2}(\mathrm{FSC}_{\pi'}) \\ &= (1 + O_q(\epsilon^{1/2}))(C_{\chi^2}(\mathrm{FSC}_\pi) + C_{\chi^2}(\mathrm{FSC}_{\pi'})), \end{aligned} \tag{161}$$

where the last step is by $A \le 1$. ■

D.2. Proof of Prop. 18

Proof [Proof of Prop. 18] We treat the regular tree case and the Poisson tree case uniformly. Let D be the offspring distribution, i.e., $D = \mathbb{1}_d$ for the regular case and $D = \mathrm{Pois}(d)$ for the Poisson case.

Let $C_1 > 0$ be the constant in Lemma 20, i.e., for all $\epsilon > 0$, FMS channels $P : \mathcal{X} \to \mathcal{Y}$, $Q : \mathcal{X} \to \mathcal{Z}$ with $C_{\chi^2}(P) \le \epsilon$, we have

$$C_{\chi^2}(P \star Q) \le \left(1 + C_1 \epsilon^{1/5}\right) (C_{\chi^2}(P) + C_{\chi^2}(Q)). \tag{162}$$

Let $C_2 > 0$ be such that $C_{\chi^2}(P) \le C_2$ for all FMS channels P.

Take $c_1, c_2 > 0$ such that

$$\exp(c_1) d\lambda^2 + c_2 < 1. \tag{163}$$

Take b_0 such that for all $t \ge b_0$, we have

$$t^5 \mathbb{P}_{b\sim D}[b > t] < c_2 C_2^{-1} (c_1 C_1^{-1})^5. \tag{164}$$

For the regular case we can take $b_0 = d$. For the Poisson case the existence of such b_0 follows from Poisson tail behavior.

For $\epsilon > 0$, define

$$b(\epsilon) := c_1 C_1^{-1} \epsilon^{-1/5}. \tag{165}$$

Take $\epsilon_0 > 0$ such that $b(\epsilon_0) > b_0$.

Take $\epsilon_1 > 0$ such that $\epsilon_1 < \min\{\epsilon_0, \delta\}$. Take $\epsilon > 0$ such that $\epsilon < \epsilon_1$ and

$$(\exp(c_1) d\lambda^2 + c_2)\epsilon_1 + \exp(c_1)\epsilon < \epsilon_1. \tag{166}$$

Let W (resp. P) be a q-FMS channel satisfying $C_{\chi^2}(W) \le \epsilon$ (resp. $C_{\chi^2}(P) \le \epsilon_1$). We have

$$\begin{aligned} C_{\chi^2}(\mathrm{BP}_W(P)) &= \mathbb{E}_{b\sim D} C_{\chi^2}((P \circ P_\lambda)^{\star b} \star W) \\ &= \mathbb{E}_{b\sim D} \left[C_{\chi^2}((P \circ P_\lambda)^{\star b} \star W) \mathbb{1}\{b \le b(\epsilon_1)\}\right] \\ &\quad + \mathbb{E}_{b\sim D} \left[C_{\chi^2}((P \circ P_\lambda)^{\star b} \star W) \mathbb{1}\{b > b(\epsilon_1)\}\right] \\ &=: L + R. \end{aligned} \tag{167}$$

For L, by induction on b we have

$$C_{\chi^2}((P \circ P_\lambda)^{\star b} \star W) \le (1 + C_1 \epsilon_1^{1/5})^b (b C_{\chi^2}(P \circ P_\lambda) + \epsilon) \le (1 + C_1 \epsilon_1^{1/5})^b (b\lambda^2 \epsilon_1 + \epsilon). \tag{168}$$

Then

$$\begin{aligned} L &\le \mathbb{E}_{b\sim D} \left[\exp(C_1 \epsilon_1^{1/5} b)(b\lambda^2 \epsilon_1 + \epsilon) \mathbb{1}\{b \le b(\epsilon_1)\}\right] \\ &\le \mathbb{E}_{b\sim D} \left[\exp(C_1 \epsilon_1^{1/5} b(\epsilon_1))(b\lambda^2 \epsilon_1 + \epsilon)\right] \\ &\le \exp(C_1 \epsilon_1^{1/5} b(\epsilon_1))(d\lambda^2 \epsilon_1 + \epsilon) \\ &= \exp(c_1)(d\lambda^2 \epsilon_1 + \epsilon). \end{aligned} \tag{169}$$

For R we have

$$R \le C_2 \cdot c_2 C_2^{-1} \left(c_1 C_1^{-1}\right)^5 \cdot b(\epsilon_1)^{-5} = c_2 \epsilon_1. \tag{170}$$

Combining (167)(169)(170) we get

$$C_{\chi^2}(\mathrm{BP}_W(P)) \le (\exp(c_1)d\lambda^2 + c_2)\epsilon_1 + \exp(c_1)\epsilon < \epsilon_1. \tag{171}$$

Let M_k be the channel $\sigma_\rho \mapsto (T_k, \omega_{T_k})$. Then $M_{k+1} = \mathrm{BP}_W(M_k)$. By (171) and $\epsilon \le \epsilon_1$ we see that

$$C_{\chi^2}(M_k) \le \epsilon_1 < \delta \tag{172}$$

for all k. This finishes the proof. ■

Appendix E. Asymmetric fixed points

Up to now we have focused on symmetric fixed points of the BP operator. If we view the BP operator as an operator on the space of q-ary input (possibly asymmetric) channels, then a natural question to determine the (possibly asymmetric) fixed points. In the case $q = 2$, Yu and Polyanskiy (2022) showed that there is only one non-trivial fixed point, and the fixed point is symmetric. For $q \ge 3$, it is no longer the case.

Proposition 36 *Work under the setting of Theorem 1. If $q \ge 3$ and $d\lambda^2 > 1$, then the* BP *operator (7) has at least one non-trivial asymmetric fixed point.*

Proof Consider the channel $U : [q] \to \{\pm\}$, which maps 1 to $+$ and $2, \dots, q$ to $-$. Because $\mathrm{BP}(U) \le_{\deg} U$, the sequence $(\mathrm{BP}^k(U))_{k\ge 0}$ is non-increasing in degradation preorder. Therefore a limit channel $\mathrm{BP}^\infty(U)$ exists by Lemma 21.

We would like to show that $\mathrm{BP}^\infty(U)$ is a non-FMS non-trivial fixed point. When $d\lambda^2 > 1$, count-reconstruction is possible (see e.g. Mossel (2001)). So it is possible to gain non-trivial information about whether the input is 1 by counting the number of $+$. So $\mathrm{BP}^\infty(U)$ is non-trivial.

On the other hand, $\mathrm{BP}^\infty(U)(\cdot|i)$ are the same for $i = 2, \dots, q$. This cannot happen for any non-trivial FMS channel. Therefore $\mathrm{BP}^\infty(U)$ is not an FMS channel. ■

Nevertheless, when the condition in Theorem 1 holds, for an open set of initial channels, it will converge to the unique FMS fixed point under BP iterations. We make the following definition.

Definition 37 *Let $U : \mathcal{X} \to \mathcal{Y}$ be a channel where $\mathcal{X} = [q]$. We say U has full rank if there exists a partition of $\mathcal{Y}$ into q measurable subsets $\mathcal{Y} = E_1 \cup \cdots \cup E_q$ such that the $q \times q$ matrix*

$$(U(E_j|i))_{i\in[q], j\in[q]} \tag{173}$$

is invertible.

Proposition 38 *Work under the setting of Theorem 1. If (q, λ, d) satisfies (27) or (28), then for any q-ary input (possibly asymmetric) channel U of full rank, we have*

$$\mathrm{BP}^\infty(U) = \mathrm{BP}^\infty(\mathrm{Id}). \tag{174}$$

Proof We prove that under the condition that U has full rank, there exists a channel R such that $R \circ U$ is a non-trivial Potts channel.

Because U has full rank, there exists a partition $\mathcal{Y} = E_1 \cup \cdots \cup E_q$ such that

$$(U(E_j|i))_{i\in[q],j\in[q]} \tag{175}$$

is invertible. Define $Q : \mathcal{Y} \to [q]$ by mapping $y \in E_i$ to i for all $i \in [q]$. Then we can replace U with $Q \circ U$ and wlog assume that $\mathcal{Y} = [q]$.

By Lemma 39, there exists $\lambda > 0$ such that $P_\lambda \le_{\deg} U$. Therefore $P_\lambda \le_{\deg} U \le_{\deg} \mathrm{Id}$. Degradation of q-ary input (possibly asymmetric) channels is preserved under BP operator. So by iterating the BP operator, we get

$$\mathrm{BP}^\infty(P_\lambda) \le_{\deg} \mathrm{BP}^\infty(U) \le_{\deg} \mathrm{BP}^\infty(\mathrm{Id}). \tag{176}$$

The first and third channels are equal by Theorem 1. Therefore $\mathrm{BP}^\infty(U) = \mathrm{BP}^\infty(\mathrm{Id})$. ∎

Lemma 39 *Fix $q \in \mathbb{Z}_{\ge 2}$ and $\epsilon > 0$. Then there exists $\lambda > 0$ such that for any probability kernel $U : [q] \to [q]$ with $\sigma_{\min}(U) > \epsilon$, we have $P_\lambda \le_{\deg} U$.*

Proof Because $\sigma_{\min}(U) > \epsilon$, we have

$$\max_{i,j\in[q]} \left| \left(U^{-1}\right)_{i,j} \right| \le \|U^{-1}\|_2 \le \sqrt{q}\epsilon^{-1}. \tag{177}$$

Let $J \in \mathbb{R}^{q\times q}$ be the all ones matrix. Because U is a stochastic matrix, we have $U^{-1}J = J$. Because the maximum (in absolute value) entry of U^{-1} is bounded by an constant, for some constant $\lambda > 0$ the matrix $U^{-1}P_\lambda$ has non-negative entries. Note that $U^{-1}P_\lambda \mathbb{1} = U^{-1}\mathbb{1} = \mathbb{1}$. So $U^{-1}P_\lambda$ is a stochastic matrix. Let $R = U^{-1}P_\lambda$. Then $R \circ U = UR = P_\lambda$, thus $P_\lambda \le_{\deg} U$. ∎

For the boundary irrelevance operator BP_W (10), the situation is simpler: when the survey FMS channel W is non-trivial, there is no asymmetric fixed point.

Proposition 40 *Work under the setting of Theorem 1. If (q, λ, d, W) satisfies (27) or (29), then BP_W has only one fixed point.*

Proof Suppose U is a fixed point of BP_W. We have

$$0 \le_{\deg} U \le_{\deg} \mathrm{Id}\,. \tag{178}$$

Degradation for q-ary input (possibly asymmetric) channels is preserved under BP_W operator. So by iterating the BP_W operator, we get

$$\mathrm{BP}_W^\infty(0) \le_{\deg} \mathrm{BP}_W^\infty(U) \le_{\deg} \mathrm{BP}_W^\infty(\mathrm{Id}). \tag{179}$$

The first and third channels are equal by Theorem 1. Therefore $U = \mathrm{BP}_W^\infty(U) = \mathrm{BP}_W^\infty(\mathrm{Id})$ is equal to the unique FMS fixed point. ∎

Appendix F. SBM mutual information

In this section we prove Theorem 3. The proof is a direct generalization of (Abbe et al., 2021, Theorem 1).

Proof [Proof of Theorem 3] Let $Y_v^\epsilon \sim \mathrm{EC}_\epsilon(\cdot|X_v)$ for $v \in V$ and $\epsilon \in [0,1]$. Let $u \in V$ be a fixed vertex. Define $f(\epsilon) := \frac{1}{n} I(X; G, Y^\epsilon)$. Then $f(0) = H(X_u)$ and $f(1) = \frac{1}{n} I(X; G)$. Furthermore, calculation shows that

$$f'(\epsilon) = -H(X_u|G, Y^\epsilon_{V\backslash u}). \tag{180}$$

Let $k \in \mathbb{Z}_{\geq 1}$ be a constant, $B(u,k)$ be the set of vertices with distance $\leq k$ to u, and $\partial B(u,k)$ be the set of vertices at distance k to u. By the data processing inequality and Lemma 42, we have

$$I(X_u; G, Y^\epsilon_{B(u,k)\backslash u}) \leq I(X_u; G, Y^\epsilon_{V\backslash u}) \leq I(X_u; G, Y^\epsilon_{B(u,k)\backslash u}, X_{\partial B(u,k)}) + o(1). \tag{181}$$

By Lemma 41, we have

$$I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}|T_k) - o(1) \leq I(X_u; G, Y^\epsilon_{V\backslash u}) \leq I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}, \sigma_{L_k}|T_k) + o(1). \tag{182}$$

Taking limit $n \to \infty$, then taking limit $k \to \infty$, we get

$$\lim_{k\to\infty} I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}|T_k) \leq \lim_{n\to\infty} I(X_u; G, Y^\epsilon_{V\backslash u}) \leq \lim_{k\to\infty} I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}, \sigma_{L_k}|T_k). \tag{183}$$

The first and third terms are equal by the boundary irrelevance assumption. Therefore

$$\lim_{n\to\infty} I(X_u; G, Y^\epsilon_{V\backslash u}) = \lim_{k\to\infty} I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}|T_k) \tag{184}$$

So

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} I(X; G) &= H(X_u) - \int_0^1 \lim_{n\to\infty} H(X_u|G, Y^\epsilon_{V\backslash u}) d\epsilon \\ &= \int_0^1 \lim_{n\to\infty} I(X_u; G, Y^\epsilon_{V\backslash u}) d\epsilon \\ &= \int_0^1 \lim_{k\to\infty} I(\sigma_\rho; \omega^\epsilon_{T_k\backslash\rho}|T_k) d\epsilon.\end{aligned} \tag{185}$$

■

Lemma 41 (SBM-BOT coupling) *Let $(X,G) \sim \mathrm{SBM}(n, q, \frac{a}{n}, \frac{b}{n})$. Let $v \in V$ and $k = o(\log n)$. Let $B(v,k)$ be the set of vertices with distance $\leq k$ to v, and $\partial B(v,k)$ be the set of vertices at distance k to v. Let $d = \frac{a+(q-1)b}{q}$ and $\lambda = \frac{a-b}{a+(q-1)b}$. Let (T,σ) be the Potts model with broadcasting channel P_λ on a Poisson tree with expected offspring d. Let ρ be the root of T, L_k be the set of vertices at distance k to ρ, T_k be the set of vertices at distance $\leq k$ to ρ.*

Then $(G|_{B(v,k)}, X_{B(v,k)})$ can be coupled (with $o(1)$ TV distance) to (T_k, σ_{T_k}).

Proof This is a well-known result and has appeared in many places. For a proof, see e.g., (Mossel et al., 2022, Lemma 6.2). ■

In the following, we use a.a.s. (asymptotically almost surely) to denote that a event happens with probability $1 - o(1)$.

Lemma 42 (No long range correlations) *Let $(X, G = (V, E)) \sim \mathrm{SBM}(n, q, \frac{a}{n}, \frac{b}{n})$. Let $A = A(G), B = B(G), C = C(G) \subseteq V$ be a (random) partition of V such that B separates A and C in G. If $|A \cup B| = o(\sqrt{n})$ a.a.s., then*

$$\mathbb{P}(X_A | X_{B\cup C}, G) = (1 \pm o(1))\mathbb{P}(X_A | X_B, G) \text{ a.a.s.} \tag{186}$$

Proof This is a special case of (Gu, 2023, Prop. 5.6). For completeness we present a self-contained proof here. The proof is a generalization of (Mossel et al., 2015, Lemma 6).

For $e = \{u, v\} \in \binom{V}{2}$, define

$$\psi_e(G, X) := \begin{cases} \frac{w_{X_u,X_v}}{n}, & \text{if } e \in E, \\ 1 - \frac{w_{X_u,X_v}}{n}, & \text{if } e \not\in E, \end{cases} \tag{187}$$

$$\text{where} \quad w_{i,j} := \begin{cases} a, & \text{if } i = j, \\ b, & \text{if } i \neq j. \end{cases} \tag{188}$$

Then

$$\mathbb{P}(G, X) := \mathbb{P}(X)\mathbb{P}(G|X) = q^{-n} \prod_{e\in\binom{V}{2}} \psi_e(G, X). \tag{189}$$

We partition $\binom{V}{2}$ into four parts. Define

$$E_1 := \left\{ e \in \binom{V}{2} : |e \cap A| = |e \cap C| = 1 \right\}, \tag{190}$$

$$E_2 := \left\{ e \in \binom{V}{2} : |e \cap C| = 0 \right\}, \tag{191}$$

$$E_3 := \left\{ e \in \binom{V}{2} : |e \cap A| = 0, |e \cap C| \geq 1 \right\}. \tag{192}$$

Then $E_1 \cup E_2 \cup E_3 = \binom{V}{2}$ is a partition of $\binom{V}{2}$. Define

$$Q_i := Q_i(G, X) := \prod_{e\in E_i} \psi_e(G, X) \quad \forall i \in [3]. \tag{193}$$

Then

$$\mathbb{P}(G, X) = q^{-n} Q_1 Q_2 Q_3. \tag{194}$$

We prove that Q_1 is approximately independent of $X_{B\cup C}$ a.a.s. Let $(\alpha_n)_{n\geq 0}$ be a deterministic sequence with $\alpha_n = \omega(\sqrt{n})$ and $\alpha_n |A| = o(n)$ a.a.s. Define

$$\Omega := \{Y \in [q]^V : \left| N_i(Y) - \frac{n}{q} \right| \leq \alpha_n \forall i \in [q]\}, \tag{195}$$

$$\Omega_U := \{Y \in \Omega : Y_U = X_U\}, \tag{196}$$

$$\text{where} \quad N_i(Y) := \#\{v \in V : Y_v = i\}. \tag{197}$$

By concentration of $N_i(X)$, we have $X \in \Omega$ a.a.s.

Note that $E_1 \cap E = \emptyset$. So

$$\begin{aligned} Q_1 &= \prod_{u\in A, v\in C} \psi_{uv}(G,X) \\ &= \prod_{u\in A, v\in C} \left(1 - \frac{w_{X_u,X_v}}{n}\right) \\ &= (1+o(1)) \prod_{u\in A, v\in C} \exp\left(-\frac{w_{X_u,X_v}}{n}\right) \text{ a.a.s.} \end{aligned} \tag{198}$$

where the third step is because $|A| = o(\sqrt{n})$ a.a.s. and

$$\exp\left(-\frac{w_{X_u,X_v}}{n}\right) = \left(1 + O\left(n^{-2}\right)\right)\left(1 - \frac{w_{X_u,X_v}}{n}\right). \tag{199}$$

For every $u \in A$ and $X \in \Omega$, we have

$$\begin{aligned} &\prod_{v\in C} \exp\left(-\frac{w_{X_u,X_v}}{n}\right) \\ &= \exp\left(-\sum_{v\in C} \frac{w_{X_u,X_v}}{n}\right) \\ &= \exp\left(-\sum_{i\in[q]} \frac{w_{X_u,i}}{n}\cdot\left(\frac{n}{q} \pm O(\alpha_n)\right) \pm O(n^{-1})\right) \\ &= \exp\left(-\frac{1}{q}\sum_{i\in[q]} w_{X_u,i} \pm O(\alpha_n n^{-1})\right) \\ &= \exp\left(-d \pm O(\alpha_n n^{-1})\right) \text{ a.a.s.} \end{aligned} \tag{200}$$

Combining (198) and (200) we get

$$\begin{aligned} Q_1 &= (1+o(1))\exp\left(-d|A| \pm O(\alpha_n |A| n^{-1})\right) \\ &= (1\pm o(1))\exp\left(-d|A|\right) =: (1\pm o(1))K(G) \text{ a.a.s.} \end{aligned} \tag{201}$$

By (201), we have

$$\mathbb{P}(G,X) = (1\pm o(1)) q^{-n} K(G) Q_2 Q_3 \text{ a.a.s.} \tag{202}$$

Furthermore, for any $U = U(G) \subseteq V$ we have

$$\begin{aligned} \mathbb{P}(G,X_U) &= (1\pm o(1))\mathbb{P}(G,X_U, X\in\Omega) \\ &= (1\pm o(1)) \sum_{Y\in\Omega_U} q^{-n} K(G) Q_2(G,Y) Q_3(G,Y) \text{ a.a.s.} \end{aligned} \tag{203}$$

Therefore

$$\mathbb{P}(X_A|X_B,G) = \frac{\mathbb{P}(X_{A\cup B},G)}{\mathbb{P}(X_B,G)} = (1\pm o(1)) \frac{\sum_{Y\in\Omega_{A\cup B}} Q_2(G,Y)Q_3(G,Y)}{\sum_{Y\in\Omega_B} Q_2(G,Y)Q_3(G,Y)} \text{ a.a.s.} \tag{204}$$

Note that $Q_2(G,Y)$ is a function of $(G, Y_{A\cup B})$ and $Q_3(G,Y)$ is a function of $(G, Y_{B\cup C})$. So the numerator of (204) is a.a.s. equal to

$$Q_2(G,X) \sum_{Y\in\Omega_{A\cup B}} Q_3(G,Y) \tag{205}$$

and the denominator of (204) is a.a.s. equal to

$$\left(\sum_{Y\in\Omega_{B\cup C}} Q_2(G,Y)\right)\left(\sum_{Y\in\Omega_{A\cup B}} Q_3(G,Y)\right). \tag{206}$$

Combining (204)(205)(206), we get

$$\mathbb{P}(X_A|X_B,G) = (1\pm o(1))\frac{Q_2(G,X)}{\sum_{Y\in\Omega_{B\cup C}} Q_2(G,Y)} \text{ a.a.s.} \tag{207}$$

Similarly, we have

$$\begin{aligned}\mathbb{P}(X_A|X_{B\cup C},G) &= \frac{\mathbb{P}(X,G)}{\mathbb{P}(X_{B\cup C},G)}\\ &= (1\pm o(1))\frac{Q_2(G,X)Q_3(G,X)}{\sum_{Y\in\Omega_{B\cup C}} Q_2(G,Y)Q_3(G,Y)}\\ &= (1\pm o(1))\frac{Q_2(G,X)}{\sum_{Y\in\Omega_{B\cup C}} Q_2(G,Y)} \text{ a.a.s.}\end{aligned} \tag{208}$$

Comparing (207) and (208) we finish the proof. ■

Appendix G. Optimal recovery algorithm

In this section we prove Theorem 4 and 5.

Proof [Proof of Theorem 4] We run Algorithm 1. Let $\rho\in V$ be a fixed vertex. For $k\in\mathbb{Z}_{\geq 1}$, define $B(\rho,k)$, $\partial B(\rho,k)$ as in Lemma 41. By Lemma 41 and induction on t, we see that $m^{(t)}_{u\to v}$ has the same distribution (up to $o(1)$ TV distance) as the posterior distribution of σ_ρ conditioned ω_{T_t}. Therefore $m^{(r)}_{u\to v}$ has the same distribution (up to $o(1)$ TV distance) as the posterior distribution of σ_ρ conditioned ω_{T_r}. So as $n\to\infty$, Algorithm 1 achieves accuracy

$$1-\lim_{k\to\infty} P_e(\sigma_\rho|T_k,\omega_{T_k}). \tag{209}$$

On the other hand, we have

$$\begin{aligned}P_e(X_\rho|G,Y) &\geq P_e(X_\rho|G,Y,X_{\partial B(\rho,k)})\\ &= P_e(X_\rho|G,Y_{B(\rho,k)},X_{\partial B(\rho,k)})\pm o(1)\\ &\geq P_e(\sigma_\rho|T_k,\omega_{T_k},\sigma_{L_k})-o(1).\\ &= P_e(\sigma_\rho|T_k,\omega_{T_k})-o(1).\end{aligned} \tag{210}$$

Algorithm 1 Belief propagation algorithm for SBM with side information

Input: SBM graph $G = (V, E)$, side information $Y \in \mathcal{Y}^V$
Output: $\widehat{X} \in [q]^V$
$m^{(0)}_{u\to v} \leftarrow \left(\frac{W(Y_u|i)}{\sum_{j\in[q]} W(Y_u|j)}\right)_{i\in[q]} \quad \forall (u,v) \in E$
$r \leftarrow \lfloor \log^{0.9} n \rfloor$
for $t = 0 \to r-1$ **do**
    for $(u,v) \in E$ **do**

$$Z^{(t+1)}_{u\to v} \leftarrow \sum_{j\in[q]} W(Y_u|j) \prod_{(u,w)\in E, w\neq v} \sum_{k\in[q]} m^{(t)}_{w\to u}(k) P_\lambda(k|j)$$

$$m^{(t+1)}_{u\to v} \leftarrow \left(\left(Z^{(t+1)}_{u\to v}\right)^{-1} \cdot W(Y_u|i) \prod_{(u,w)\in E, w\neq v} \sum_{k\in[q]} m^{(t)}_{w\to u}(k) P_\lambda(k|i)\right)_{i\in[q]}$$

    end for
end for
for $u \in V$ **do**

$$Z_u \leftarrow \sum_{j\in[q]} W(Y_u|j) \prod_{(u,w)\in E} \sum_{k\in[q]} m^{(r)}_{w\to u}(k) P_\lambda(k|j)$$

$$m_u \leftarrow \left(Z_u^{-1} \cdot W(Y_u|i) \prod_{(u,w)\in E} \sum_{k\in[q]} m^{(r)}_{w\to u}(k) P_\lambda(k|i)\right)_{i\in[q]}$$

    $\widehat{X}_u \leftarrow \arg\max_{i\in[q]} m_u(i)$
end for
return $\widehat{X}$

where the first step is by data processing inequality, the second step is by Lemma 42, the third step is by Lemma 41, the fourth step is by boundary irrelevance with respect to W. Taking limit $n \to \infty$, then $k \to \infty$, we see that

$$P_e(X_\rho|G, Y) \geq \lim_{k\to\infty} P_e(\sigma_\rho|T_k, \omega_{T_k}). \tag{211}$$

This shows that Algorithm 1 is optimal. ■

Proof [Proof of Theorem 5] We run Algorithm 2. The proof is a variation of the proof in Mossel et al. (2016).

Choice of u_i. For every $i \in [q]$, the set $\{u \in U : X_u = i\}$ has size $\frac{n}{q} \pm o(n)$. Therefore with high probability, there exists $u \in U$ with $X_u = i$ that satisfies (a). Furthermore, because Y is independent of U, we can equivalently first generate the graph $G \backslash U$, then compute Y, then generate the edges adjacent to U. In this way, we see that with high probability, for all $u \in U$ satisfying (a), the empirical distribution of $\{Y_v : v \in V, (u,v) \in E\}$ has $o(1)$ total variation distance to $P_\lambda F$. By

Algorithm 2 Belief propagation algorithm for SBM

1: **Input:** SBM graph $G = (V, E)$, initial recovery algorithm $\mathcal{A}$
2: **Output:** $\widehat{X} \in [q]^V$
3: $s \leftarrow 1$ if model is assortative; $s \leftarrow -1$ if model is disassortative
4: $r \leftarrow \lfloor \log^{0.9} n \rfloor$
5: $U \leftarrow$ random subset of V of size $\lfloor \sqrt{n} \rfloor$
6: $Y \leftarrow \mathcal{A}(G \backslash U)$
7: For $i \in [q]$, $u \in U$, compute $N_Y(u, i) \leftarrow \#\{Y_v = i : v \in V, (u, v) \in E\}$
8: For $i \in [q]$, choose $u_i \in U$ such that
   (a) u_i has at least $\sqrt{\log n}$ neighbors in $V \backslash U$, and
   (b) $sN_Y(u_i, i) > sN_Y(u_i, j)$ for $j \in [q] \backslash i$.
9: **for** $v \in V \backslash U$ **do**
10: $Y^v \leftarrow \mathcal{A}(G \backslash B(v, r-1) \backslash U)$
11: Relabel Y^v by performing a permutation $\tau \in \mathrm{Aut}([q])$, so that $sN_{Y^v}(u_i, i) > sN_{Y^v}(u_i, j)$ for $i \in [q]$, $j \in [q] \backslash i$. Report failure if this cannot be achieved.
12: $M^v_{i,j} \leftarrow \frac{N_{Y^v}(u_i, j)}{\sum_{j \in [q]} N_{Y^v}(u_i, j)}$
13: Run belief propagation on $B(v, r-1)$ with boundary condition $Y^v_{\partial B(v,r)}$, assuming the channel from $\partial B(v, r-1)$ to $\partial B(v, r)$ is H^v
14: $\widehat{X}_v \leftarrow$ maximum likelihood label according to belief propagation
15: **end for**
16: $\widehat{X}_v \leftarrow 1$ for all $v \in U$
17: **return** $\widehat{X}$

assumption (1)(3) in Theorem 5, we have $s(P_\lambda F)_{i,\tau(i)} > s(P_\lambda F)_{i,\tau(j)} + |\lambda|\epsilon$ for $i \in [q]$, $j \in [q] \backslash i$. Therefore with high probability, for all $u \in U$ satisfying (a), we can identify X_u up to a permutation $\tau \in \mathrm{Aut}([q])$ by computing $\arg\max_{j \in [q]} sN_Y(u, j)$. Therefore with high probability we are able to choose the u_is in Line 8.

Alignment of Y with Y^v. The above discussion still holds with Y replaced by Y^v. One thing to note is that by Lemma 41, $|B(v, r-1)| = n^{o(1)}$ with high probability. So removing $B(v, r-1)$ from G has negligible influence to the the empirical distribution of labels of neighbors of u_is. Therefore, with high probability, we are able to permute the labels Y^v so that the empirical distributions align with that of Y. (Note that we do not assume the empirical distributions for Y and Y^v are the same; we only use that they both satisfy condition (3).) Furthermore, we can compute the transition matrix

$$M^v = P_\lambda F^v \pm o(1). \tag{212}$$

Boundary condition of BP. Because Y^v is independent of edges between $\partial B(v, r-1)$ and $\partial B(v, r)$, we can equivalently first generate the graph $G \backslash B(v, r-1) \backslash U$, then compute Y^v, then generate $E(\partial B(v, r-1), \partial B(v, r))$. In this way, it is clear that Y^v_w for one $(u, w) \in E(\partial B(v, r-1), \partial B(v, r))$ is equivalent to one observation of X_u through channel M^v.

Property of M^v. Note that $M^v \geq_{\deg} P_{\lambda - o(1)} F^v$. By Lemma 39 and condition (2), we have $F^v \geq_{\deg} P_{\lambda'}$ for some constant $\lambda' > 0$ not depending on n. Therefore $M^v \geq_{\deg} P_{\lambda''}$ for some $\lambda'' > 0$ not depending on n.

Convergence of BP recursion. Because $\lambda'' > 0$ is a constant, by the stability of BP fixed point assumption, for any $\kappa > 0$, there exists some integer k_0 not depending on n such that

$$P_e(\mathrm{BP}^{k_0}(M^v)) \geq P_e(\mathrm{BP}^{k_0}(P_{\lambda''})) > \lim_{k\to\infty} P_e(\mathrm{BP}^k(\mathrm{Id})) - \kappa. \tag{213}$$

Because $r = \omega(1)$, belief propagation in Line 13 converges to $o(1)$ in TV distance to the fixed point. Therefore we achieve desired accuracy in Line 14. ■

Appendix H. Discussions

Phase transition for BP uniqueness and boundary irrelevance. We summarize the phase transition for BP uniqueness and boundary irrelevance as follows. The important thresholds are the reconstruction threshold and the Kesten-Stigum threshold ($d\lambda^2 = 1$).

- Below the reconstruction threshold: The BP operator (without survey) has no non-trivial fixed points. We conjecture that BI always holds below the reconstruction threshold. Low SNR part in Theorem 1 shows that BI holds whenever $d\lambda^2 < q^{-2}$.
- Between the reconstruction and the KS threshold: Theorem 2 shows that boundary irrelevance does not hold in this regime. The BP operator has a non-trivial fixed point. This fixed point is not globally stable, i.e., there exists non-trivial channel U such that $\mathrm{BP}^\infty(U)$ is trivial Janson and Mossel (2004). We do not know whether the BP operator has a unique non-trivial fixed point.
- Above the KS threshold: We conjecture that BP uniqueness and BI always hold in this regime, and the unique non-trivial fixed point is globally stable. High SNR part in Theorem 1 is tight within a factor of $1+\log q$ for general (λ, d), and asymptotically tight within a factor of $1+o_q(1)$ for $q \to \infty$ and $\lambda = o(\log q)$.

Asymmetric models. The Potts model is a very symmetric model and can be studied using FMS channels. For more general SBM and BOT models, the class of FMS channels is no longer suitable. Nevertheless, it might be possible to extend our degradation method to the asymmetric case to achieve better results than Chin and Sly (2021), which uses a generalization of the method of Mossel et al. (2016).

Fix a finite alphabet $\mathcal{X}$ and a distribution π on $\mathcal{X}$ with full support. Then a channel $\mathcal{P} : \mathcal{X} \to \mathcal{Y}$ can be viewed as a distribution of posterior distributions under prior π, i.e., the distribution of $P_{X|Y}$ where $P_X = \pi$, $Y \sim P(\cdot|X)$. Note that this is a distribution on $\mathcal{P}(\mathcal{X})$. Similarly to Prop. 10, degradation between channels with input alphabet $\mathcal{X}$ can be equivalently characterized as a coupling between the two distributions of posterior distributions. Let ϕ be a strongly convex function on $\mathcal{P}(\mathcal{X})$. Using the above distributional characterization of channels, we can extend ϕ to a potential function Φ on the space of channels with input alphabet $\mathcal{X}$. For two sequences $(M_k)_{k\geq 0}$, $(\widetilde{M_k})_{k\geq 0}$ satisfying the BP recursion and related by degradation $\widetilde{M_k} \leq_{\deg} M_k$, if $\lim_{k\to\infty}(\Phi(M_k) - \Phi(\widetilde{M_k})) = 0$, then we should expect $M_\infty = \widetilde{M}_\infty$.

The difficulty in carrying out the above plan is to analyze behavior of Φ under BP recursion. One also needs to keep in mind that the Potts model admits more than one non-trivial fixed points in the space of all q-ary input channels (Section E). So extra assumptions are needed when dealing with the general case.

Yu and Polyanskiy (2022)'s degradation method. Yu and Polyanskiy (2022) proved uniqueness of BP fixed point and boundary irrelevance for all symmetric Ising models. Therefore a natural idea is to generalize their proof to the Potts model. However, it seems difficult to make this work.

While Yu and Polyanskiy (2022)'s proof is also based on degradation of BMS channels, their method is quite different from Abbe et al. (2021). Specifically, Yu and Polyanskiy (2022) uses a stronger version of the stringy tree lemma of Evans et al. (2000), which says that $(P^{\star d}) \circ \mathrm{BSC}_\delta \le_{\deg} (P \circ \mathrm{BSC}_\delta)^{\star d}$ for all BMS channels P, and under additional assumptions one even has $(P^{\star d}) \circ \mathrm{BSC}_\delta \le_{\deg} (P \circ \mathrm{BSC}_\delta)^{\star d} \circ \mathrm{BSC}_\epsilon$ for some $\epsilon > 0$ depending on P, δ, d. The natural generalization of the original stringy tree lemma to the Potts model (with BSC_δ replaced by P_λ and BMS channels replaced by FMS channels) is not true, even for $d\lambda^2 > 1$. Therefore it is unclear how to generalize Yu and Polyanskiy (2022)'s proof to the Potts model.

Robust reconstruction. The boundary irrelevance problem is related to the robust reconstruction problem, which asks whether the trivial fixed point of the BP operator is locally stable, i.e., whether for weak enough initial channel U, we have $\mathrm{BP}^\infty(U) = 0$. In fact, our proof of Theorem 2 can be slightly modified to show that the Potts model does not admit robust reconstruction below the Kesten-Stigum threshold.

The robust reconstruction problem for BOT models has been extensively studied in Janson and Mossel (2004), which showed that the robust reconstruction threshold is at the KS threshold, i.e., the model admits robust reconstruction when $d\lambda^2 > 1$ and does not when $d\lambda^2 < 1$. Their proof uses the contraction of a χ^2-like information measure, which can take value ∞ in some corner cases, e.g., for the coloring model with erasure leaf observations, or for the boundary irrelevance problem. Furthermore, they only considered trees with bounded maximum degree, which does not include the Poisson tree. Therefore to prove Theorem 2 we cannot directly use Janson and Mossel (2004).