Proceedings of Machine Learning Research vol 134:1–25, 2021 34th Annual Conference on Learning Theory

Stochastic block model entropy and broadcasting on trees with survey

Emmanuel Abbe EMMANUEL.ABBE@EPFL.CH
Elisabetta Cornacchia ELISABETTA.CORNACCHIA@EPFL.CH
École polytechnique fédérale de Lausanne

Yuzhou Gu YUZHOUGU@MIT.EDU
Yury Polyanskiy YP@MIT.EDU
Massachusetts Institute of Technology

Editors: Mikhail Belkin and Samory Kpotufe

Abstract

The limit of the entropy in the stochastic block model (SBM) has been characterized in the sparse regime for the special case of disassortative communities Coja-Oghlan et al. (2017) and for the classical case of assortative communities but in the dense regime Deshpande et al. (2016). The problem has not been closed in the classical sparse and assortative case. This paper establishes the result in this case for any SNR besides for the interval $(1, 3.513)$. It further gives an approximation to the limit in this window.

The result is obtained by expressing the global SBM entropy as an integral of local tree entropies in a broadcasting on tree model with erasure side information. The main technical advancement then relies on showing the irrelevance of the boundary in such a model, also studied with variants in Kanade et al. (2016), Mossel et al. (2016) and Mossel and Xu (2015). In particular, we establish the uniqueness of the BP fixed point in the survey model for any SNR above 3.513 or below 1. This only leaves a narrow region in the plane between SNR and survey strength where the uniqueness of BP conjectured in these papers remains unproved.

Keywords: stochastic block model, broadcasting on trees with side information, belief propagation, local algorithms optimality, information-theoretic limits.

1. Introduction

Over the last decade, several works have established a precise picture for the statistical and algorithmic behavior of the stochastic block model (see an account in Abbe (2018)). In particular, the questions of weak and exact recovery, i.e., whether it is possible (or not) to recover the communities in the extremal cases of weak and exact accuracy, have been fully closed in the two-community symmetric SBM by establishing sharp threshold phenomena in terms of appropriate signal-to-noise (SNR) ratios Massoulié (2014); Mossel et al. (2015, 2018); Abbe et al. (2016). Yet, despite significant progress, the more nuanced question of proving how much information or agreement can be recovered about the communities at any given value of the SNR has remained open even in this simplest case.

More specifically, for two symmetric communities and in the sparse regime, the expression of the limiting entropy of the SBM is characterized[1] at all SNR for the special case of disassortative communities (i.e., communities that connect more outside than inside) Coja-Oghlan et al. (2017).

1. Characterizing the limit does not mean obtaining an explicit expression; it refers to an implicit n-independent expression relying on integrals and fixed point equations for the quantities of interest in all of the papers discussed here.

The problem for assortative communities is closed but in the denser regimes, where the vertex degrees diverge while maintaining a finite SNR Deshpande et al. (2016). However, for the classical case of assortative communities and in the sparse regime, a complete characterization remains open, despite significant progress Kanade et al. (2016); Mossel et al. (2016); Mossel and Xu (2015). The expression of the optimal agreement (rather than the entropy) is known in this case for SNR large enough, and is related to the problem of robust reconstruction on a tree Mossel et al. (2016). This result is conjectured to hold all the way down to the optimal threshold of 1, i.e., the threshold until which the communities can be weakly recovered. We make progress on this question by establishing the result down to 3.513. Further, we establish new results and improvements of prior results for the problem of broadcasting on a tree with side information; see Section 1.1.

The SBM entropy. Recall that in the symmetric SBM with two communities, a random variable X is drawn uniformly at random in $\{\pm 1\}^n$ and an n-vertex graph G is drawn by connecting vertices having same (resp. different) values in X with probability a/n (resp. b/n).

The SBM mutual information is defined by the limit (if it exists)

$$\mathcal{I}(a,b) := \lim_{n\to\infty} \frac{1}{n} I(X;G), \tag{1}$$

where I is the mutual information. Note that establishing the existence of this limit is nontrivial. This was proved in Abbe and Montanari (2015) for the case of $a < b$, the same case for which the value of the limit has more recently been established Coja-Oghlan et al. (2017). Note also that due to the chain rule $I(X;G) = H(X) - H(X|G)$, the SBM mutual information is the complement of the SBM conditional entropy (called simply the SBM entropy)

$$\mathcal{H}(a,b) := \lim_{n\to\infty} \frac{1}{n} H(X|G). \tag{2}$$

Informally, the SBM mutual information measures how much information can be recovered about the communities after observing the graph, and equivalently, the SBM entropy measures how much uncertainty is left about the communities after observing the graph. More formally, it quantifies the average number of bits needed to represent the communities after observing the graph; see Abbe (2016) for formal relations to graph compression.

Note that one may use other measures on the communities signal given the graph, such as the optimal (normalized) mean square error of reconstructing the $n \times n$ rank-2 block matrix (with a/n in the $n/2 \times n/2$ diagonal blocks and b/n in the off diagonal blocks), or the optimal (normalized) agreement (Hamming distance) of reconstructing X up to a community relabelling. These can be explicitly related to each other in the tree models discussed next, and require bounds in the SBM context; see for instance Deshpande et al. (2016). The conditional entropy allows however for a direct reduction from the SBM to the tree model with side information, as discussed below.

The BOTS entropy. Consider the following problem of broadcasting on a tree with side information (BOTS). This will be later defined on general trees and with general side information, but consider for simplicity the case of regular trees (where each vertex has exactly d descendants) and erasure side information. In this model, a random bit is attached to the root of the tree and broadcasted down the tree by flipping its value independently with probability δ on each edge (for convenience we call $\theta = 1 - 2\delta$). We denote by σ_ρ the root bit, by σ_{L_k} the d^k-dimensional vector of the leaf bits at generation k, and by $\omega^{\epsilon}_{T_k}$ the side information up to depth k: these are the vertices labelled that are revealed in the tree (besides the root) independently with probability $1-\epsilon$. We call this side information the "survey". Note that this is the type of side information used in our connection between BOTS and SBM entropies, but other types of side information are of independent interest. In this paper we devote attention to general (but symmetric with respect to the spin flip) observation model of the nodes, which we refer to as the BMS channel W.

We are now interested in two quantities:

1. the limiting entropy of the root bit after observing the leaf bits and the survey, i.e.[2],

$$\bar{h}(d,\theta,\epsilon) := \lim_{k\to\infty} H(\sigma_\rho|\sigma_{L_k},\omega^\epsilon_{T_k}),$$

2. the same quantity without the leaf bits being observed, i.e.,

$$h(d,\theta,\epsilon) := \lim_{k\to\infty} H(\sigma_\rho|\omega^\epsilon_{T_k}).$$

We now give a rather direct method to express the SBM entropy in terms of BOTS entropies.

SBM to BOTS entropy reduction. The relation obtained between the SBM and BOTS conditional entropy is as follows: if for some range of parameters d,θ, we can establish that

$$h=\bar{h}, \quad \forall\epsilon\in(0,1),$$

i.e., if the *boundary is irrelevant*, then we can characterize $\mathcal{H}$ as an integral of $\bar{h}$ using the parameter correspondence $d=(a+b)/2$ and $\theta=\frac{a-b}{a+b}$ (see Theorem 1).

Our starting point to such a reduction is an area-theorem or interpolation trick that is commonly used in coding theory Richardson and Urbanke (2001) and related statistical physics literature Mézard and Montanari (2009).

The idea is to express the entropy in the SBM $H(X|G)$ as the integral

$$\frac{1}{n}H(X|G)=\int_0^1 \frac{1}{n}\frac{\partial}{\partial\epsilon}H(X|G,Y^\epsilon)d\epsilon, \tag{3}$$

where, similarly as before, Y^ϵ is an erasure survey that reveals the community of each vertex in X independently with probability $1-\epsilon$. We then use the fact that $\frac{1}{n}\frac{\partial}{\partial\epsilon}H(X|G,Y_\epsilon)=H(X_1|G,Y^\epsilon_{\sim 1})$, where 1 is an arbitrary vertex in the graph and $Y^\epsilon_{\sim 1}$ denotes the erasure survey on all vertices excluding vertex 1. Since conditioning reduces entropy, one can upper bound $H(X_1|G,Y^\epsilon_{\sim 1})$ by considering only the information in the vertex 1 neighborhood, and due to the local tree-like topology of SBMs, this gives an upper bound with the BOTS entropy without leaf information. Moreover, one can add the leaf information in the conditioning to cut-off the graph beyond a local neighborhood, using the Markovianity[3] of the model, obtaining as well a lower bound from the BOTS entropy but this time with the leaf information, cf. (17).

Different kind of reductions from SBMs to tree models have long been known and leveraged in the SBM in Coja-Oghlan et al. (2017); Mossel et al. (2016); Alaoui and Montanari (2019); we refer to Section 1.1 for further discussions on these.

We now turn to the crux of the analysis, i.e., the establishment of $h=\bar{h}$.

2. Note that in these tree models, the limits can be proved to always exist.
3. Strict Markovanity does not hold in the SBM due to the weak effect of non-edges, and this requires a technical lemma; see proof of Theorem 1. This technicality can also be avoided by considering the related Censored Block Model (CBM), rather than the SBM, for which strict Markovianity holds.

Uniqueness of BP fixed point for BOTS. Our main contribution is to show that in a wide range of parameters and side information models, the BOTS associated distributional fixed point equation (known as BP fixed point) has a unique solution. This automatically has several implications.

First, this establishes the desired "Boundary Irrelevance" property for BEC survey, i.e., $h = \bar{h}$:

$$\lim_{k\to\infty} H(\sigma_\rho|\sigma_{L_k}, \omega^\epsilon_{T_k}) = \lim_{k\to\infty} H(\sigma_\rho|\omega^\epsilon_{T_k}). \tag{4}$$

This implies

$$\lim_{\epsilon\to 1}\lim_{k\to\infty} H(\sigma_\rho|\omega^\epsilon_{T_k}) = \lim_{k\to\infty} H(\sigma_\rho|\sigma_{L_k}). \tag{5}$$

Indeed, one only needs to notice that $\lim_{\epsilon\to 1}\lim_{k\to\infty} H(\sigma_\rho|\sigma_{L_k}, \omega^\epsilon_{T_k}) = \sup_{\epsilon,k} H(\sigma_\rho|\sigma_{L_k}, \omega^\epsilon_{T_k})$ and that for every k the latter quantity is continuous in $\epsilon \in [0, 1]$ including at the boundary.

Further, the presence of the survey allows to convert the absence of leaf information into the presence of noisy leaf information, thereby obtaining the robust reconstruction property in the presence and in the absence of the survey Mossel et al. (2016).

Property (5) is also known in the SBM literature as the condition for "optimality of local algorithms", and was investigated in Kanade et al. (2016); Mossel and Xu (2015). These works build on the crucial contribution of Mossel et al. (2016), which shows uniqueness of BP fixed point for BOT without survey and $d\theta^2 > C$, where C is "large enough" (see Appendix D for our estimates of how large). Note that since the conditional entropy in (5) can be sandwiched between $H(\sigma_\rho|\sigma_{L_k})$ and $H(\sigma_\rho|\omega^\epsilon_{L_k})$, the result of Mossel et al. (2016) implies (5), as indeed observed in (Kanade et al., 2016, Prop. 3). However, Kanade et al. (2016) derives result for the case where $\epsilon \to 1$, relying on Mossel et al. (2016) for large enough C. It also conjectures the more general (4) (for all d, θ, ϵ and BEC survey), and our paper validates this conjecture in a wide range of parameters (see Fig. 1), including for all values of the $d\theta^2 \notin (1, 3.513)$.

Finally, subsequent work Mossel and Xu (2015) focuses on the case of BSC_ϵ rather than BEC_ϵ survey, and also conjectured (4) for all d, θ, ϵ. They demonstrate the uniqueness of the BP fixed point in this setting for some range of parameters (which as $\epsilon \to 1/2$ reduces to $d\theta^2 > C$ for some large enough C). Although the method of Mossel and Xu (2015) is an extension of Mossel et al. (2016), the authors make the remark "We note however that the paper Mossel et al. (2016) did not consider side information and the adaptation of the proof is far from trivial." This is further expanded in the current paper.

1.1. Novelty and comparison to the literature

We believe that our proof technique offers the following improvements compared to Mossel et al. (2016); Mossel and Xu (2015): (a) it is much shorter; (b) we do not need to consider large θ, small d and small d large θ cases separately; (c) it works simultaneously for $d\theta^2 < 1$ and $d\theta^2 > 3.513$; (d) it works simultaneously with and without side information, and the side information can be any BMS, rather than specifically the BEC or BSC; (e) it closes the entire low-SNR case $d\theta^2 < 1$[4], and to the best of our knowledge it yields the state-of-the-art threshold for the high-SNR case.

Our main innovation is the information-theoretic point of view: we consider BOTS with or without leaf observations as two binary input symmetric channels (BMSs) which are related to

4. There are, however, two related low-SNR results. (Mossel and Xu, 2015, Theorem 4.2) shows uniqueness of fixed point for $d\theta < 1$ via a simple contractivity of F_θ function in the BP recursion (27).(Kanade et al., 2016, Theorem 3) shows (5) for $d\theta^2 < 1$ as an application of information contraction from Evans et al. (2000).

each other by a property known as degradation. This implies a certain inequality between the log-likelihood ratios (LLRs), cf. (31), which we exploit in the application of the potential method. These key ideas are the content of the Prop. 2. On the more technical side, another innovation is the choice of the potential function as $\phi(r) = e^{-\frac{1}{2}r}$.

Concerning the reduction from SBM to BOTs, we note first that the reduction in Mossel et al. (2016) is obtained for the agreement metric. It is easy to navigate between agreement and entropy once on the tree models, but in the SBM, the entropy allows for the chain rule and other properties that lead to the direct reduction detailed previously. On the other hand, Mossel et al. (2016), rely on a black-box algorithms that solves weak recovery in order to bring the noisy leaves. Therefore, we are trading the noisy leaves with the survey. In turn, we can exploit the survey to obtain tighter conditions for the boundary irrelevance that lead to part (ii) of Theorem 1.

Finally, Coja-Oghlan et al. (2017) uses a reduction to trees for the entropy that does also not rely on the erasure side information as described above. In particular, the computation of the SBM entropy is linked to an optimization problem (Theorem 2.2 therein), whose solution corresponds to the dominant BP fixed point on a Galton-Watson tree (Theorem 2.4).

2. Results: Boundary Irrelevance and SBM Entropy

Broadcasting on Trees with Survey (BOTS). We start with the standard broadcasting on trees (BOT) setting. Let T be an infinite tree rooted at ρ. Let $\sigma_\rho \sim \mathrm{Unif}(\{\pm 1\})$ be the root bit and assume that it is broadcast through each edge independently with flip probability $\delta \in (0, \frac{1}{2}]$. For simplicity we use notation $\theta = 1 - 2\delta$. Let L_k denote the set of nodes at level k, and T_k denote the set of nodes at level $\leq k$ (where the root is at level 0). Reconstruction on such models consists of recovering the root bit after observing the leaves bits at large depth (Evans et al. (2000)).

We consider a slightly different problem, where we have access to some node side information, or "survey". Specifically, let W be a fixed BMS channel, and for each node u we observe $\omega_u \sim W(\sigma_u)$. We call (T, ρ, θ, W) a broadcasting instance with survey. We will also denote by Δ_W the Δ-component of the BMS W (see Section 4 for background on BMS channels). This setting includes the one in Mossel and Xu (2015), where $W = \mathrm{BSC}_\alpha$, i.e., for each node u, $\mathbb{P}[\omega_u = \sigma_u] = 1 - \mathbb{P}[\omega_u = -\sigma_u] = 1 - \alpha$; and the one in Kanade et al. (2016), where $W = \mathrm{BEC}_\epsilon$, i.e., for each node the survey reveals the correct label with probability $1 - \epsilon$ and an erasure symbol otherwise. The latter is of particular interest to us because of its application to the computation of the SBM entropy (Theorem 1). For clarity, in the case of erasure survey, we denote $\omega_u^\epsilon = \mathrm{BEC}_\epsilon(\sigma_u)$.

Theorem 1 *Let $(X, G) \sim SBM(n, 2, a/n, b/n)$. Let T be a Galton-Watson tree with Pois$(\frac{a+b}{2})$ offspring distribution and let $(T, \rho, \frac{a-b}{a+b}, \mathrm{BEC}_\epsilon)$ be a broadcasting instance with erasure survey, and edge flip probability $\frac{b}{a+b}$. Let $\alpha^* \approx 3.513$ be the unique solution in $\mathbb{R}_{>1}$ to the equation $\exp(-\frac{\alpha-1}{2})\alpha = 1$. The following hold.*

(i) For a, b such that $\frac{(a-b)^2}{2(a+b)} \leq 1$ or $\frac{(a-b)^2}{2(a+b)} \geq \alpha^* \approx 3.513$

$$\mathcal{H}(a, b) = \lim_{n\to\infty} \frac{1}{n} H(X|G) = \int_0^1 \lim_{k\to\infty} H(\sigma_\rho | T, \sigma_{L_k}, \omega_{L_k}^\epsilon) d\epsilon. \tag{6}$$

(ii) For any a, b such that $\frac{(a-b)^2}{2(a+b)} \in (1, \alpha^*)$, i.e., inside the gap of part (i),

$$\liminf_{n\to\infty} \frac{1}{n} H(X|G) = \int_0^1 \lim_{k\to\infty} H(\sigma_\rho | T, \sigma_{L_k}, \omega_{L_k}^\epsilon) d\epsilon + \xi_{\inf}, \tag{7}$$

$$\limsup_{n\to\infty} \frac{1}{n} H(X|G) = \int_0^1 \lim_{k\to\infty} H(\sigma_\rho | T, \sigma_{L_k}, \omega_{L_k}^\epsilon) d\epsilon + \xi_{\sup}, \tag{8}$$

where $0 \le \xi_{\inf}, \xi_{\sup} \le 1 - \frac{\sqrt{e}}{2} \approx 0.178$.

A crucial ingredient to establish Theorem 1 is the following property for BOTS.

Definition 1 (Boundary Irrelevance (BI)) *We say that (T, ρ, θ, W) has the Boundary Irrelevance (BI) property if*

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

which is equivalent to (4).

In words, (BI) implies that if we have access to some intermediate node information, the leaves at infinite depth become irrelevant for detecting the root bit. We focus on regular and Galton-Watson trees with Poisson offspring. We prove the following Theorem in Section 5.

Theorem 2 *Let T be a d-regular tree or a Galton-Watson tree with Poisson(d) offspring distribution, with root vertex ρ. Let W be a BMS channel. If $P_e(W) \neq \frac{1}{2}$, and*

$$d\theta^2 \exp\left(-\frac{(d\theta^2 - 1)_+}{2}\right) Z(W) < 1, \tag{10}$$

where $P_e(W)$ is the probability of error, and $Z(W)$ is the Bhattacharyya coefficient (defined in Definition 4), then (BI) holds for (T, ρ, θ, W). In particular, (BI) holds for any (T, ρ, θ, W) with $d\theta^2 < 1$ or $d\theta^2 > \alpha^*$ (and with $P_e(W) \neq \frac{1}{2}$), where $\alpha^* \approx 3.513$ is the unique solution in $\mathbb{R}_{>1}$ to the equation $\exp(-\frac{\alpha-1}{2})\alpha = 1$.

We remark that (10) is a relaxation of a sharper bound in Prop. 3 (e.g., for regular trees with $d = 2$ (BI) is proven for all cases except $d\theta^2 \in (1, 1.62)$). The following corollary lists a few direct consequences of Theorem 2.

Corollary 1 *In the setting of Theorem 2, if any of the following is true, then (BI) holds for (T, ρ, θ, W): (i) $Z(W) < \frac{\sqrt{e}}{2} \approx 0.824$; (ii) $P_e(W) < \frac{1}{2} - \frac{1}{4}\sqrt{4-e} \approx 0.217$; (iii) $W = \mathrm{BEC}_\epsilon$ and with $\epsilon < \frac{\sqrt{e}}{2} \approx 0.824$.*

Proof For (i) we use $\sup_{\alpha \ge 0}(\alpha \exp(-\frac{\alpha-1}{2})) = \frac{2}{\sqrt{e}}$. For (ii) we define $p(\Delta) = 2\sqrt{\Delta(1-\Delta)}$ and notice that $Z(W) = \mathbb{E}[p(\Delta_W)] \le p(\mathbb{E}\Delta_W) = p(P_e(W))$ because the function p is concave. So when $P_e(W) < \frac{1}{2} - \frac{1}{4}\sqrt{4-e}$, we have $Z(W) < \frac{\sqrt{e}}{2}$. (iii) follows from (i). ■

Theorem 2 is a consequence of the following more general result, that we state informally here (for the full statement see Prop.6 in Appendix C).

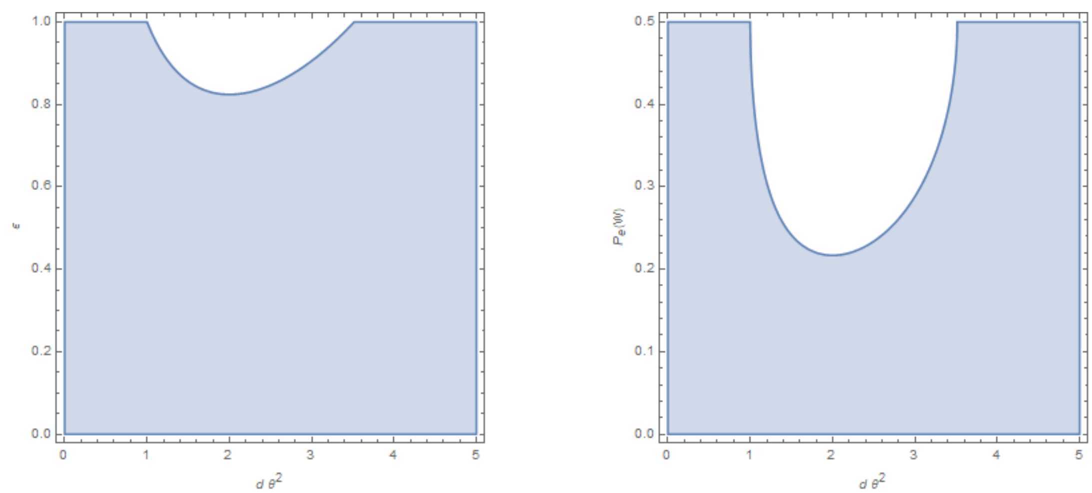

Figure 1: Left: Region of BP uniqueness for BEC survey from Corollary 1(iii).
Right: Region of BP uniqueness for BMS survey from Corollary 1(ii).

Proposition 1 (Informal, uniqueness of BP fixed point) *For BOTS, if (10) holds, then the BP distributional fixed point is unique. For BOT if $d\theta^2 < 1$ or $d\theta^2 > \alpha^*$ then the non-trivial fixed point is unique.*

We demonstrate the region of BP-uniqueness from Corollary 1 on Figure 1. We note that, taking the limit $\epsilon \to 1^-$, Theorem 2 implies that revealing an (arbitrarily) small fraction of vertex labels gives the same information about the root bit, as revealing the whole boundary labels at large distance, even in the reconstruction regime, cf. (5).

Conjecture 1 *Let T be a regular tree or a Galton-Watson tree with Poisson offspring distribution, with root vertex ρ. Then (BI) holds for (T, ρ, θ, W) for all $0 < \theta < 1$ and all W such that $P_e(W) \neq \frac{1}{2}$.*

If Conjecture 1 holds, the proof of Theorem 1 gives a precise characterisation of $\mathcal{H}(a, b)$, as in (6), in terms of BOTS entropies for the entire range of a, b.

3. Proof of Theorem 1

Let us denote $f(\epsilon) = H(X|G, Y^\epsilon)$, where similarly as before Y^ϵ is a BEC_ϵ-survey that reveals the true label of each node independently with probability $1 - \epsilon$. Note that $f(1) = H(X|G)$. Let us replace the single parameter ϵ by a set of parameters $\boldsymbol{\epsilon} = (\epsilon_u)_{u \in V(G)}$ (for each vertex u, X_u is revealed with probability $1 - \epsilon_u$), and let us denote $Y^\epsilon_{\sim u} = \{Y^\epsilon_v : v \in V(G), v \neq u\}$ and $X_{\sim u} = \{X_v : v \in V(G), v \neq u\}$. Then

$$f(\boldsymbol{\epsilon}) = (1 - \epsilon_u) H(X|G, X_u, Y^\epsilon_{\sim u}) + \epsilon_u H(X|G, Y^\epsilon_{\sim u}) \tag{11}$$

and by chain rule

$$\frac{\partial}{\partial \epsilon_u} f(\boldsymbol{\epsilon}) = H(X|G, Y^{\epsilon}_{\sim u}) - H(X|G, X_u, Y^{\epsilon}_{\sim u}) \tag{12}$$

$$= H(X_u, X_{\sim u}|G, Y^{\epsilon}_{\sim u}) - H(X_{\sim u}|G, X_u, Y^{\epsilon}_{\sim u}) \tag{13}$$

$$= H(X_u|G, Y^{\epsilon}_{\sim u}). \tag{14}$$

Then, setting $\epsilon_u = \epsilon$ for all $u \in V(G)$, we get by symmetry

$$f'(\epsilon) = \sum_{u \in V(G)} H(X_u|G, Y^{\epsilon}_{\sim u}) = nH(X_1|G, Y^{\epsilon}_{\sim 1}). \tag{15}$$

Thus, by bounded convergence

$$\lim_{n\to\infty} \frac{1}{n} H(X|G) = \int_0^1 \lim_{n\to\infty} H(X_1|G, Y^{\epsilon}_{\sim 1}) d\epsilon. \tag{16}$$

Take $k = \frac{\log n}{10 \log 2(a+b)}$ small enough compared to n, such that the neighborhood of vertex 1 at depth k is a tree with high probability (this is for instance proved as Proposition 2 in Mossel et al. (2015)), and denote such neighborhood by T_k. Specifically, w.h.p. T_k is a Galton-Watson tree with Poisson$\left(\frac{a+b}{2}\right)$ offspring distribution, rooted at 1, and the labels in X_{T_k} are distributed as BOT with flip probability $\frac{b}{a+b}$. Moreover, let X_{L_k} be the vertices at distance exactly k from 1, and let $Y^{\epsilon}_{\sim 1, T_k}$ denote the survey on nodes at distance at most k from 1 (excluding 1). We bound the integrand by the following:

$$H(X_1|T_k, Y^{\epsilon}_{\sim 1, T_k}, X_{L_k}) + o_k(1) \le H(X_1|G, Y^{\epsilon}_{\sim 1}) \le H(X_1|T_k, Y^{\epsilon}_{\sim 1, T_k}). \tag{17}$$

For the inequality on the right, we simply removed conditioning terms and thus increased the conditional entropy, specifically we ignored any information from the graph or from the survey on nodes at distance $\ge k$ to 1. The inequality on the left requires the following lemma, that is a direct consequence of Proposition 2 and Lemma 4.7 in Mossel et al. (2015).

Lemma 1 $H(X_1|G, Y^{\epsilon}_{\sim 1}, X_{L_k}) = H(X_1|T_k, Y^{\epsilon}_{\sim 1, T_k}, X_{L_k}) + o_k(1)$.

In words, Lemma 1 states that after conditioning on the leaves, the information coming from the graph outside T_k (including non-edges) becomes negligible, i.e. the model is asymptotically a Markov field. By Theorem 2, if $\frac{(a-b)^2}{2(a+b)} \le 1$ or $\frac{(a-b)^2}{2(a+b)} \ge \alpha^*$, then (BI) holds for $(T_k, 1, \frac{a-b}{a+b}, \mathrm{BEC}_\epsilon)$, for all $\epsilon < 1$, thus the leftmost and the rightmost terms in (17) are asymptotically equal. This means that the limit in the integrand in (16) exists for all $\epsilon \in (0, 1)$, thus (i) holds.

On the other hand, by Corollary 1(iii), for all $\epsilon < \epsilon^* = \frac{\sqrt{e}}{2} \approx 0.824$ and for all a, b (BI) holds for $(T_k, 1, \frac{a-b}{a+b}, \mathrm{BEC}_\epsilon)$. Thus

$$\liminf_{n\to\infty} \frac{1}{n} H(X|G) = \int_0^{\epsilon^*} \lim_{k\to\infty} H(X_1|T_k, Y^{\epsilon}_{\sim 1, T_k}, X_{L_k}) d\epsilon + \xi_{\inf}, \tag{18}$$

with

$$\xi_{\inf} = \int_{\epsilon^*}^1 \liminf_{n\to\infty} H(X_1|G, Y^{\epsilon}_{\sim 1}) - \lim_{k\to\infty} H(X_1|T_k, Y^{\epsilon}_{\sim 1, T_k}, X_{L_k}) d\epsilon \tag{19}$$

$$\le (1 - \epsilon^*) \lim_{k\to\infty} I(X_1; X_{L_k}|T_k) \le (1 - \epsilon^*) \approx 0.178. \tag{20}$$

The same holds for $\limsup$ and $\xi_{\sup}$.

4. Preliminaries on BMS channels

We give necessary preliminaries on BMS channels that are used in the proof of Theorem 2. Most material in this section can be found in e.g., (Richardson and Urbanke, 2008, Chapter 4).

Definition 2 *A channel $P : \{\pm 1\} \to \mathcal{Y}$ is called a Binary Memoryless Symmetric (BMS) channel if there exists a measurable involution $\sigma : \mathcal{Y} \to \mathcal{Y}$ such that $P(\sigma^{-1}(E)|+) = P(E|-)$ for all measurable sets $E \subseteq \mathcal{Y}$.*

Examples of BMS channels include Binary Erasure Channels (BECs) and Binary Symmetric Channels (BSCs). In fact, every BMS channel is a mixture of BSCs, in the sense of the following Lemma.

Lemma 2 *Every BMS channel P is equivalent to a channel $X \to (\Delta, Z)$, where $\Delta \in [0, \frac{1}{2}]$ is independent of X, and $P_{Z|\Delta,X} = \mathrm{BSC}_\Delta(X)$.*

In the setting of the above lemma, we call channel $X \to (\Delta, Z)$ the standard form of P, and call Δ the Δ-component of P.

Definition 3 *Let $P : \{\pm 1\} \to \mathcal{Y}$ and $Q : \{\pm 1\} \to \mathcal{Z}$ be two BMS channels. We say P is more degraded than Q (denoted $P \le_{\deg} Q$), if there exists a channel $R : \mathcal{Z} \to \mathcal{Y}$ such that $P = R \circ Q$.*

Degradation can be characterized in terms of the Δ-component.

Lemma 3 *Let P and Q be two BMS channels. Let Δ be the Δ-component of P and $\tilde{\Delta}$ be the Δ-component of Q. Then $P \ge_{\deg} Q$ if and only if there exists a coupling between Δ and $\tilde{\Delta}$ so that $\mathbb{E}[\Delta|\tilde{\Delta}] \le \tilde{\Delta}$ for all $\tilde{\Delta} \in [0, \frac{1}{2}]$ for which LHS exists.*

Definition 4 *Let P be a BMS channel and Δ be the Δ-component of P. We define the following quantities.*

$$\begin{aligned} P_e(P) &= \mathbb{E}\Delta, && \text{(probability of error)} \\ C(P) &= \mathbb{E}[\log 2 + \Delta \log \Delta + (1-\Delta)\log(1-\Delta)], && \text{(capacity)} \\ C_{\chi^2}(P) &= \mathbb{E}[(1-2\Delta)^2], && (\chi^2\text{-capacity}) \\ Z(P) &= \mathbb{E}[2\sqrt{\Delta(1-\Delta)}]. && \text{(Bhattacharyya coefficient)} \end{aligned}$$

By definition, $P_e(P) \in [0, \frac{1}{2}]$, $C(P) \in [0, \log 2]$, $C_{\chi^2}(P) \in [0,1]$, $Z(P) \in [0,1]$. These quantities behave nicely under degradation, by the coupling characterization (Lemma 3) and convexity.

Lemma 4 *If $P \le_{\deg} Q$, then the following holds:*

$$P_e(P) \ge P_e(Q), \quad C(P) \le C(Q), \quad C_{\chi^2}(P) \le C_{\chi^2}(Q), \quad Z(P) \ge Z(Q). \tag{21}$$

5. Proof of Theorem 2

Recall the BOTS model defined in Section 2. Let M_k denote the BMS channel $\sigma_\rho \to (\omega_{T_k}, \sigma_{L_k})$ and $\tilde{M}_k$ denote the BMS channel $\sigma_\rho \to \omega_{T_k}$. Let P_{Δ_k} (resp. $P_{\tilde{\Delta}_k}$) be distribution of Δ-component of BMS M_k (resp. $\tilde{M}_k$). We prove the following strengthening of Theorem 2.

Theorem 3 *In the setting of Theorem 2, P_{Δ_k} and $P_{\tilde{\Delta}_k}$ converge in distribution to the same distribution as $k \to \infty$. In particular,*

$$\lim_{k\to\infty} P_e(M_k) = \lim_{k\to\infty} P_e(\tilde{M}_k), \tag{22}$$

$$\lim_{k\to\infty} C(M_k) = \lim_{k\to\infty} C(\tilde{M}_k). \tag{23}$$

5.1. Belief propagation recursion

The maximum a posteriori probability (MAP) decoder is the optimal decoder for this reconstruction problem. It can be implemented using belief propagation (BP) as follows.

For each node u, let $L_k(u)$ denote the set of nodes in subtree rooted at u that are at distance k to u. Let $T_k(u)$ denote the set of nodes in subtree rooted at u that are at distance $\le k$ to u. Let $R_{u,k} \in \mathbb{R} \cup \{\pm\infty\}$ denote the posterior log likelihood ratio given $\omega_{T_k(u)} \cup \sigma_{L_k(u)}$:

$$R_{u,k} = \log \frac{\mathbb{P}[\sigma_u = +|\omega_{T_k(u)} \cup \sigma_{L_k(u)}]}{\mathbb{P}[\sigma_u = -|\omega_{T_k(u)} \cup \sigma_{L_k(u)}]}. \tag{24}$$

The initial value is

$$R_{u,0} = \sigma_u \cdot \infty. \tag{25}$$

Define a function $F_\theta : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R}$ as

$$F_\theta(r) = 2 \operatorname{arctanh}(\theta \tanh(\frac{1}{2} r)). \tag{26}$$

By definition of $R_{u,k}$ and Bayes rule (see e.g. Mossel and Xu (2015)), we have

$$R_{u,k+1} = \sum_{v \in L_1(u)} F_\theta(R_{v,k}) + W_u \tag{27}$$

where W_u is the log likelihood ratio induced by observation, i.e.,

$$W_u = \log \frac{\mathbb{P}[\sigma_u = +|\omega_u]}{\mathbb{P}[\sigma_u = -|\omega_u]}. \tag{28}$$

Using (25)(27) we are able to compute $R_{\rho,k}$ recursively.

For observation without leaves, let $\tilde{R}_{u,k}$ denote the posterior log likelihood ratio given $\omega_{T_k(u)}$. Then $\tilde{R}_{u,k}$ satisfies the same recursion (27), but with a different initial value

$$\tilde{R}_{u,0} = 0. \tag{29}$$

Let $M_k(u)$ denote the BMS channel $\sigma_u \to (\omega_{T_k(u)}, \sigma_{L_k(u)})$. Let $\tilde{M}_k(u)$ denote the BMS channel $\sigma_u \to \omega_{T_k(u)}$. Let $\Delta_{u,k}$ and $\tilde{\Delta}_{u,k}$ denote the corresponding Δ-components (both are random variables supported on $[0, \frac{1}{2}]$). They relate to log likelihood ratio via the following expression:

$$|R_{u,k}| = \log \frac{1 - \Delta_{u,k}}{\Delta_{u,k}}, \quad |\tilde{R}_{u,k}| = \log \frac{1 - \tilde{\Delta}_{u,k}}{\tilde{\Delta}_{u,k}}. \tag{30}$$

There exists a canonical coupling between $M_k(u)$ and $\tilde{M}_k(u)$ via forgetting $\sigma_{L_k(u)}$ (i.e., the channel $(\omega_{T_k(u)}, \sigma_{L_k(u)}) \mapsto \omega_{T_k(u)}$). So $M_k(u)$ is less degraded than $\tilde{M}_k(u)$. Furthermore, by data processing inequality for total variation, under the canonical coupling, we have

$$\mathbb{E}[\Delta_{u,k}|\tilde{\Delta}_{u,k}] \le \tilde{\Delta}_{u,k}. \tag{31}$$

As we will see, the core of our proof is the use of this degradation relationship.

Let μ_k^+ be the distribution of $R_{u,k}$ conditioned on $\sigma_u = +$ (and for Galton-Watson trees, without revealing structure of the subtree rooted at u), and $\tilde{\mu}_k^+$ be the distribution of $\tilde{R}_{u,k}$ conditioned on $\sigma_u = +$. These definitions do not depend on the choice of u. Then μ_0^+ is the point measure at $+\infty$, $\tilde{\mu}_0^+$ is the point measure at 0.

Both distributions satisfy the same recursion. Consider the equation

$$R_{u,k+1}^+ = \sum_{v \in L_1(u)} Z_v F_\theta(R_{v,k}^+) + W_u \tag{32}$$

where $\{Z_v, R_{v,k}^+, W_u : v \in L_1(u)\}$ are independent, Z_v are i.i.d. Bernoulli with $\mathbb{P}[Z_v = +1] = 1 - \mathbb{P}[Z_v = -1] = 1 - \delta$, $R_{v,k}^+ \sim \mu_k^+$, and W_u distributes as log likelihood ratio corresponding to the survey BMS. Then $R_{u,k+1}^+ \sim \mu_{k+1}^+$. The same holds if we replace $R_{v,k}^+ \sim \mu_k^+$ with $\tilde{R}_{v,k}^+ \sim \tilde{\mu}_k^+$ and $R_{u,k+1}^+ \sim \mu_{k+1}^+$ with $\tilde{R}_{u,k+1}^+ \sim \tilde{\mu}_{k+1}^+$.

BP distributional fixed point. A distribution μ on $\mathbb{R} \cup \{\pm\infty\}$ is called a BP fixed point of the BOTS (d, θ, W) if taking R_i^+ i.i.d. $\sim \mu$, $i \in [d]$, Z_i and R_W as above results in

$$R^+ = \sum_{1 \le i \le d} Z_i F_\theta(R_i^+) + R_W \tag{33}$$

having the same distribution μ. In this work we restrict our attention to symmetric distributions, i.e., distributions associated with BMS channels. We talk below about the fixed point distribution P_Δ on $[0, \frac{1}{2}]$ that is related to μ via transformation (30). Namely, a distribution P_Δ is a fixed point iff the law μ of random variable R^+ is a fixed point, where R^+ is generated via sampling $\Delta \sim P_\Delta$ and then setting

$$R^+ = \begin{cases} \log \frac{1-\Delta}{\Delta}, & \text{w.p. } 1 - \Delta, \\ -\log \frac{1-\Delta}{\Delta}, & \text{w.p. } \Delta. \end{cases} \tag{34}$$

Similarly, we define the BP fixed point for the BOTS $(\mathrm{Poi}(d), \theta, W)$ where in (33) d is replaced with $b \sim \mathrm{Poi}(d)$.

5.2. Contraction of potential function

The technical part of our proof is contraction of certain potential functions. The next proposition shows the kind of contraction result we need.

Proposition 2 *Let $\phi : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R} \cup \{\pm\infty\}$ be a function such that the function $g : [0, \frac{1}{2}] \to \mathbb{R} \cup \{\pm\infty\}$ defined as*

$$g(\Delta) = (1 - \Delta)\phi(\log \frac{1 - \Delta}{\Delta}) + \Delta\phi(-\log \frac{1 - \Delta}{\Delta}) \tag{35}$$

is decreasing and α-strongly convex for some $\alpha > 0$. If

$$\lim_{k \to \infty} \mathbb{E}[\phi(R_{\rho,k}^+) - \phi(\tilde{R}_{\rho,k}^+)] = 0, \tag{36}$$

then under the canonical coupling,

$$\lim_{k \to \infty} \mathbb{E}(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k})^2 = 0. \tag{37}$$

Proof Because g is α-strongly convex, we have

$$g(\Delta_{\rho,k}) - g(\tilde{\Delta}_{\rho,k}) \ge g'(\tilde{\Delta}_{\rho,k})(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k}) + \frac{\alpha}{2}(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k})^2. \tag{38}$$

Then

$$\begin{aligned}
\mathbb{E}[\phi(R^+_{\rho,k}) - \phi(\tilde{R}^+_{\rho,k})] &= \mathbb{E}_{\tilde{\Delta}_{\rho,k}} \mathbb{E}[\phi(R^+_{\rho,k}) - \phi(\tilde{R}^+_{\rho,k}) | \tilde{\Delta}_{\rho,k}] \\
&= \mathbb{E}_{\tilde{\Delta}_{\rho,k}} \mathbb{E}[g(\Delta_{\rho,k}) - g(\tilde{\Delta}_{\rho,k}) | \tilde{\Delta}_{\rho,k}] \\
&\ge \mathbb{E}_{\tilde{\Delta}_{\rho,k}} \mathbb{E}[g'(\tilde{\Delta}_{\rho,k})(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k}) + \frac{\alpha}{2}(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k})^2 | \tilde{\Delta}_{\rho,k}] \\
&= \mathbb{E}_{\tilde{\Delta}_{\rho,k}} [g'(\tilde{\Delta}_{\rho,k})(\mathbb{E}[\Delta_{\rho,k} | \tilde{\Delta}_{\rho,k}] - \tilde{\Delta}_{\rho,k})] + \frac{\alpha}{2}\mathbb{E}(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k})^2 \\
&\ge \frac{\alpha}{2}\mathbb{E}(\Delta_{\rho,k} - \tilde{\Delta}_{\rho,k})^2.
\end{aligned} \tag{39}$$

The second step is because $R^+_{\rho,k}$ and $\Delta_{\rho,k}$ (also $\tilde{R}^+_{\rho,k}$ and $\tilde{\Delta}_{\rho,k}$) relate via (34). By (39), we see that (36) implies (37). ■

Note that (39) also shows that $\mathbb{E}[\phi(R^+_{\rho,k}) - \phi(\tilde{R}^+_{\rho,k})]$ is non-negative as long as g is decreasing and convex.

We choose the potential function to be $\phi(r) = -\exp(-\frac{1}{2}r)$. This potential function is chosen so that the expectation of $\phi(R^+_{u,k+1})$ has a nice decomposition (49). In fact $\mathbb{E}[\exp(-\frac{1}{2}R^+)]$ is equal to the Bhattacharyya coefficient of the BMS channel, and (49) can be interpreted as multiplicativity of Bhattacharyya coefficients under *-convolution.

The function g is given by $g(\Delta) = -2\sqrt{\Delta(1-\Delta)}$. One can check that g is decreasing and 4-strongly convex on $[0, \frac{1}{2}]$.

Proposition 3 *Assume that we have a non-trivial survey channel. Let*

$$C_1 = C_1(d, \theta, W) = d\theta^2 (1 - \frac{(d\theta^2 - 1)_+}{d-1})^{\frac{d-1}{2}} Z(W). \tag{40}$$

For regular trees, under the canonical coupling, for any $\epsilon > 0$, there exists k^* such that for all $k \ge k^*$,

$$\mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{\rho,k+1}) - \exp(-\frac{1}{2}R^+_{\rho,k+1})] \le (1+\epsilon) C_1 \mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{\rho,k}) - \exp(-\frac{1}{2}R^+_{\rho,k})]. \tag{41}$$

In particular, if $C_1 < 1$, then (36) holds.

For Galton-Watson trees with Poisson offspring distribution, the same holds with C_1 replaced by

$$C_2 = C_2(d, \theta, W) = d\theta^2 \exp(-d(1 - \sqrt{1 - \frac{(d\theta^2 - 1)_+}{d}})) Z(W). \tag{42}$$

Proof of Proposition 3 is deferred to Appendix A.

Proposition 2 and 3 complete the proof of Theorem 3, because for $i = 1, 2$, we have

$$C_i \le d\theta^2 \exp(-\frac{(d\theta^2 - 1)_+}{2}) Z(W). \tag{43}$$

6. Other results

Weak spatial mixing. BOT (without survey) is an example of the Ising model. As it is typical for such models, at high temperature (i.e. $d\theta \leq 1$) it exhibits the property known as weak spatial mixing (WSM): enforcing a (far away) boundary condition does not affect the distribution of spins. This property disappears at low temperatures ($d\theta > 1$), but what is surprising is that there is a range of parameters ($d\theta > 1$ but $d\theta^2 < 1$) in which there is no WSM, but reconstruction is still impossible Bleher et al. (1995).

Now, the BOTS model can be thought of as an example of an Ising spin glass system: one first generates the survey and then, treating the survey as quenched randomness, considers an Ising model with external fields corresponding to survey. The question we ask is whether in this spin-glass type model we still have that (in the limit of vanishing survey) the threshold for WSM appears at $d\theta = 1$. Some partial results towards this are contained in Appendix E. We mention that for BEC survey we were not able to show this.

Boundary irrelevance (BI) on amenable graphs. So far we studied (BI) property (9) for trees, but it can also be defined for general graphs as follows.

Let $G = (V, E)$ be an infinite graph. Consider the Spin Synchronization model, where we have i.i.d. random variables $X_v \sim \mathrm{Unif}(\{\pm 1\})$ for $v \in V$; for each edge $uv \in E$, we observe a random variable $Y_{uv} \sim \mathrm{BSC}_\delta(X_u X_v)$, and we denote $\theta = 1 - 2\delta$. Conditioned on the X variables, the Y variables are mutually independent. In addition to the edge variables, we may observe surveys at each node: for $v \in V$, we have $\omega_v \sim W(X_v)$, with W being a fixed BMS channel. In this Section we consider BEC_ϵ survey.

Let $o \in V$ be a vertex. Let $B_n(o)$ be the set of nodes with distance $\leq n$ to o, and $\partial B_n(o)$ be the set of nodes at distance n to o. We use notation $X_{\partial B_n(o)}$ for the set $\{X_v : v \in \partial B_n(o)\}$ and notation $Y_{B_n(o)}$ for $\{Y_{uv} : uv \in E, u \in B_n(o), v \in B_n(o)\}$. We say the model (G, o, θ, W) has the (BI) property if

$$\lim_{n \to \infty} I(X_o; X_{\partial B_n(o)} | Y_{B_n(o)}, \omega_{B_n(o)}) = 0. \tag{44}$$

In Appendix F we show, by applying results of Alaoui and Montanari (2019), that (BI) holds for all amenable graphs and survey channel being BEC_ϵ. The definition of such graphs appears therein, but in a nutshell, it requires the boundary of any subset $S \subset V$ to be negligible compared to $|S|$.

Acknowledgments

This work was partially supported by the National Science Foundation under Grant No CCF-17-17842 and by the MIT-IBM Watson AI Lab. EA was partly supported by NSF Award CCF-1552131. In addition, YP thanks the hospitality of the School of Computer and Communication Sciences (IC) at EPFL, during his visit to which most of this work has been performed.

References

E. Abbe. Graph compression: The effect of clusters. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8, 2016.

E. Abbe and A. Montanari. Conditional random fields, planted constraint satisfaction, and entropy concentration. *Theory of Computing*, 11(17):413–443, 2015. doi: 10.4086/toc.2015.v011a017. URL http://www.theoryofcomputing.org/articles/v011a017.

E. Abbe, A.S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *Information Theory, IEEE Transactions on*, 62(1):471–487, Jan 2016. ISSN 0018-9448. doi: 10.1109/TIT.2015.2490670.

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. URL http://jmlr.org/papers/v18/16-480.html.

Ahmed El Alaoui and Andrea Montanari. On the computational tractability of statistical estimation on amenable graphs, 2019.

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(1):473–482, 1995.

Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborova. Information-theoretic thresholds from the cavity method. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 146–157, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450345286. doi: 10.1145/3055399.3055420. URL https://doi.org/10.1145/3055399.3055420.

Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2): 125–170, 12 2016. ISSN 2049-8764. doi: 10.1093/imaiai/iaw017. URL https://doi.org/10.1093/imaiai/iaw017.

William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the ising model. *Ann. Appl. Probab.*, 10(2):410–433, 05 2000. doi: 10.1214/aoap/1019487349. URL https://doi.org/10.1214/aoap/1019487349.

Varun Kanade, Elchanan Mossel, and Tselil Schramm. Global and local information in clustering labeled block models. *IEEE Transactions on Information Theory*, 62:5906–5917, 10 2016. doi: 10.1109/TIT.2016.2516564.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, page 694–703, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327107. doi: 10.1145/2591796.2591857. URL https://doi.org/10.1145/2591796.2591857.

M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, Oxford, 2009.

Elchanan Mossel and Jiaming Xu. Local algorithms for block models with side information, 2015.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015. doi: 10.1007/s00440-014-0576-6. URL https://doi.org/10.1007/s00440-014-0576-6.

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, Aug 2016. ISSN 1050-5164. doi: 10.1214/15-aap1145. URL http://dx.doi.org/10.1214/15-AAP1145.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018. doi: 10.1007/s00493-016-3238-8. URL https://doi.org/10.1007/s00493-016-3238-8.

T. Richardson and R. Urbanke. An introduction to the analysis of iterative coding systems. In *Codes, Systems, and Graphical Models*, IMA Volume in Mathematics and Its Applications, pages 1–37. Springer, 2001.

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

Hajir Roozbehani and Yury Polyanskiy. Low density majority codes and the problem of graceful degradation. *arXiv preprint arXiv:1911.12263*, 2019.

Appendix A. Proof of Proposition 3

Let us first deal with the regular tree case. Let u be a vertex and $v_1, \dots, v_d$ be its children. Let $R^+_{v_1,k}, \dots, R^+_{v_d,k}$ be i.i.d. $\sim \mu^+_k$, and $\tilde{R}^+_{v_1,k}, \dots, \tilde{R}^+_{v_d,k}$ be i.i.d. $\sim \tilde{\mu}^+_k$. Define $R^+_{u,k+1}$ and $\tilde{R}^+_{u,k+1}$ using (32). Furthermore, for $0 \le i \le d$, define $R^+_{u,i,k+1}$ as

$$R^+_{u,i,k+1} = \sum_{1\le j\le i} Z_j F_\theta(\tilde{R}^+_{v_j,k}) + \sum_{i+1\le j\le d} Z_j F_\theta(R^+_{v_j,k}) + W_u. \tag{45}$$

That is, $R^+_{u,0,k+1} = R^+_{u,k+1}$, and $R^+_{u,d,k+1} = \tilde{R}^+_{u,k+1}$.

For $1 \le i \le d$ and k large enough, let us prove that

$$\mathbb{E}[\exp(-\frac{1}{2}R^+_{u,i,k+1}) - \exp(-\frac{1}{2}R^+_{u,i-1,k+1})] \le (1+\epsilon)\frac{C_1}{d}\mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{v_1,k}) - \exp(-\frac{1}{2}R^+_{v_1,k})] \tag{46}$$

where C_1 is defined in (40). We prove that (46) is true even if conditioned on $\tilde{\Delta}_{v_i,k}$. For $\Delta \in [0, \frac{1}{2}]$, define

$$G(\Delta) = \mathbb{E}[\exp(-\frac{1}{2}R^+_{u,i,k+1}) - (1+\epsilon)\frac{C_1}{d}\exp(-\frac{1}{2}\tilde{R}^+_{v_i,k}) | \tilde{\Delta}_{v_i,k} = \Delta]. \tag{47}$$

Define $p(\Delta) = -g(\Delta) = 2\sqrt{\Delta(1-\Delta)}$ so that we work with non-negative numbers. So

$$\mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{v_i,k}) | \tilde{\Delta}_{v_i,k} = \Delta] = p(\Delta). \tag{48}$$

Then

$$
\begin{aligned}
&\mathbb{E}[\exp(-\frac{1}{2}R^+_{u,i,k+1})|\tilde{\Delta}_{v_i,k}=\Delta] \\
&= \prod_{1\le j\le i-1}\mathbb{E}[\exp(-\frac{1}{2}Z_jF_\theta(\tilde{R}^+_{v_j,k}))]\cdot\prod_{i+1\le j\le d}\mathbb{E}[\exp(-\frac{1}{2}Z_jF_\theta(R^+_{v_j,k}))] \\
&\cdot\mathbb{E}[\exp(-\frac{1}{2}Z_iF_\theta(\tilde{R}^+_{v_i,k}))|\tilde{\Delta}_{v_i,k}=\Delta]\cdot\mathbb{E}[\exp(-\frac{1}{2}W_u)]. \qquad (49)
\end{aligned}
$$

Let us examine $\mathbb{E}[\exp(-\frac{1}{2}Z_iF_\theta(\tilde{R}^+_{v_i,k}))|\tilde{\Delta}_{v_i,k}=\Delta]$. We can compute that

$$
\exp(-\frac{1}{2}Z_iF_\theta(\tilde{R}^+_{v_i,k})) = \begin{cases} \exp(-\frac{1}{2}\log\frac{1-\Delta*\delta}{\Delta*\delta}), & \text{w.p. } 1-\Delta*\delta, \\ \exp(+\frac{1}{2}\log\frac{1-\Delta*\delta}{\Delta*\delta}), & \text{w.p. } \Delta*\delta, \end{cases} \qquad (50)
$$

where we use notation $\delta_1*\delta_2=\delta_1(1-\delta_2)+\delta_2(1-\delta_1))$. So

$$
\mathbb{E}[\exp(-\frac{1}{2}Z_iF_\theta(\tilde{R}^+_{v_i,k}))|\tilde{\Delta}_{v_i,k}=\Delta]=\mathbb{E}[p(\Delta*\delta)]. \qquad (51)
$$

Similarly,

$$
\begin{aligned}
\mathbb{E}[\exp(-\frac{1}{2}Z_jF_\theta(\tilde{R}^+_{v_j,k}))] &= \mathbb{E}[p(\tilde{\Delta}_{v_1,k}*\delta)], \qquad (52)\\
\mathbb{E}[\exp(-\frac{1}{2}Z_jF_\theta(R^+_{v_j,k}))] &= \mathbb{E}[p(\Delta_{v_1,k}*\delta)]. \qquad (53)
\end{aligned}
$$

Finally,

$$
\mathbb{E}[\exp(-\frac{1}{2}W_u)]=\mathbb{E}[p(\Delta_W)]=Z(W). \qquad (54)
$$

So from (49) we get

$$
\mathbb{E}[\exp(-\frac{1}{2}R^+_{u,i,k+1})|\tilde{\Delta}_{v_i,k}=\Delta]=\mathbb{E}[p(\tilde{\Delta}_{v_1,k}*\delta)]^{i-1}\mathbb{E}[p(\Delta_{v_1,k}*\delta)]^{d-i}p(\Delta*\delta)Z(W). \qquad (55)
$$

So

$$
\begin{aligned}
G''(\Delta) &= \mathbb{E}[p(\tilde{\Delta}_{v_1,k}*\delta)]^{i-1}\mathbb{E}[p(\Delta_{v_1,k}*\delta)]^{d-i}Z(W)\frac{d^2}{d\Delta^2}p(\Delta*\delta) \\
&\quad -(1+\epsilon)\frac{C_1}{d}p''(\Delta). \qquad (56)
\end{aligned}
$$

Let us bound each factor.

$$
p(\Delta*\delta)=2\sqrt{(\Delta*\delta)(1-\Delta*\delta)}=\sqrt{1-\theta^2(1-2\Delta)^2}. \qquad (57)
$$

So

$$
\mathbb{E}[p(\tilde{\Delta}_{v_1,k}*\delta)]=\mathbb{E}[\sqrt{1-\theta^2(1-2\tilde{\Delta}_{v_1,k})^2}]\le\sqrt{1-\theta^2\mathbb{E}(1-2\tilde{\Delta}_{v_1,k})^2}. \qquad (58)
$$

By Proposition 4, for any $\epsilon' > 0$, for k large enough, we have

$$\mathbb{E}[(1-2\tilde{\Delta}_{v_1,k})^2] \geq (\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+. \tag{59}$$

So

$$\mathbb{E}[p(\tilde{\Delta}_{v_1,k} * \delta)] \leq \sqrt{1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+}. \tag{60}$$

Similarly,

$$\mathbb{E}[p(\Delta_{v_1,k} * \delta)] \leq \sqrt{1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+}. \tag{61}$$

Note that p is strictly concave on $[0, \frac{1}{2}]$, and $p'(\frac{1}{2}) = 0$. So

$$\frac{d^2}{d\Delta^2} p(\Delta * \delta) \geq \theta^2 p''(\Delta). \tag{62}$$

So (56) gives

$$\begin{aligned} G''(\Delta) &\geq (1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+)^{\frac{d-1}{2}} \theta^2 p''(\Delta) Z(W) - (1+\epsilon)\frac{C_1}{d} p''(\Delta) \\ &= ((1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+)^{\frac{d-1}{2}} \theta^2 Z(W) - (1+\epsilon)\frac{C_1}{d}) p''(\Delta). \end{aligned} \tag{63}$$

Note that

$$\lim_{\epsilon' \to 0} (1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+)^{\frac{d-1}{2}} = (1 - \frac{(d\theta^2-1)_+}{d-1})^{\frac{d-1}{2}}. \tag{64}$$

So we can take $\epsilon' > 0$ small enough so that

$$(1-\theta^2(\frac{d\theta^2-1}{(d-1)\theta^2} - \epsilon')_+)^{\frac{d-1}{2}} < (1+\epsilon)(1 - \frac{(d\theta^2-1)_+}{d-1})^{\frac{d-1}{2}}. \tag{65}$$

So for k large enough, $G''(\Delta) \leq 0$ for all $\Delta \in [0, \frac{1}{2}]$ and $G(\Delta)$ is convex. Also,

$$\begin{aligned} G'(\frac{1}{2}) &= \mathbb{E}[p(\tilde{\Delta}_{v_1,k} * \delta)]^{i-1} \mathbb{E}[p(\Delta_{v_1,k} * \delta)]^{d-i} Z(W) \frac{d}{d\Delta}|_{\Delta=\frac{1}{2}} p(\Delta * \delta) \\ &\quad - (1+\epsilon)\frac{C_1}{d} p'(\frac{1}{2}) \qquad (66) \\ &= 0. \qquad (67) \end{aligned}$$

So G' is non-positive, thus G is decreasing on $[0, \frac{1}{2}]$. Because $M_k(v_i)$ (BMS corresponding to $R^+_{v_i,k}$) is less degraded than $\tilde{M}_k(v_i)$ (BMS corresponding to $\tilde{R}^+_{v_i,k}$), we get (46).

For Galton-Watson trees with Poisson offspring distribution, the proof is very similar to, and slightly more involved than the regular case. Let u be a vertex. Let $R^+_{v_1,k}, R^+_{v_2,k}, \ldots$ be i.i.d. $\sim \mu^+_k$, and $\tilde{R}^+_{v_1,k}, \tilde{R}^+_{v_2,k}, \dots$ be i.i.d. $\sim \tilde{\mu}^+_k$. Let $b \sim \text{Poi}(d)$ and $v_1, \dots, v_b$ be the children of u. For $i \geq 0$, define

$$R^+_{u,i,k+1} = \sum_{1 \leq j \leq \min\{i,b\}} Z_j F_\theta(\tilde{R}^+_{v_j,k}) + \sum_{i+1 \leq j \leq b} Z_j F_\theta(R^+_{v_j,k}) + W_u. \tag{68}$$

For $i \geq 1$, let us prove that

$$\mathbb{E}[\exp(-\frac{1}{2}R^+_{u,i,k+1}) - \exp(-\frac{1}{2}R^+_{u,i-1,k+1})] \leq c_i \mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{v_1,k}) - \exp(-\frac{1}{2}R^+_{v_1,k})]. \tag{69}$$

where c_i are constants to be chosen later. Define

$$G_i(\Delta) = \mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{u,i,k+1}) - c_i \exp(-\frac{1}{2}\tilde{R}^+_{v_i,k}) | \tilde{\Delta}_{v_i,k} = \Delta]. \tag{70}$$

Let us prove that G_i is decreasing and convex on $[0, \frac{1}{2}]$. Similarly to (56), we have

$$G_i''(\Delta) = \mathbb{E}_b[\mathbb{1}_{b \geq i} \mathbb{E}[p(\tilde{\Delta}_{v_1,k} * \delta)]^{i-1} \mathbb{E}[p(\Delta_{v_1,k} * \delta)]^{b-i} Z(W) \frac{d^2}{d\Delta^2} p(\Delta * \delta)] - c_i p''(\Delta). \tag{71}$$

Let us study each term in (71). By (57) and Proposition 4, for any $\epsilon' > 0$, for k large enough, we have

$$\mathbb{E}[p(\tilde{\Delta}_{v_1,k} * \delta)] \leq \sqrt{1 - \theta^2 \mathbb{E}(1 - 2\tilde{\Delta}_{v_1,k})^2} \leq \sqrt{1 - \theta^2 (\frac{d\theta^2 - 1}{d\theta^2} - \epsilon')_+}. \tag{72}$$

Similarly,

$$\mathbb{E}[p(\Delta_{v_1,k} * \delta)] \leq \sqrt{1 - \theta^2 (\frac{d\theta^2 - 1}{d\theta^2} - \epsilon')_+}. \tag{73}$$

(62) still holds in the Poisson case. So (71) gives

$$G_i''(\Delta) \geq (\mathbb{E}_b[\mathbb{1}_{b \geq i} (1 - \theta^2 (\frac{d\theta^2 - 1}{d\theta^2} - \epsilon')_+)^{\frac{b-1}{2}}] \theta^2 Z(W) - c_i) p''(\Delta). \tag{74}$$

We can take

$$c_i = \mathbb{E}_b[\mathbb{1}_{b \geq i} (1 - \theta^2 (\frac{d\theta^2 - 1}{d\theta^2} - \epsilon')_+)^{\frac{b-1}{2}}] \theta^2 Z(W) \tag{75}$$

so that $G_i''(\Delta) \geq 0$ for all $i \geq 1$ and $\Delta \in [0, \frac{1}{2}]$. Also,

$$\begin{aligned} G_i'(\frac{1}{2}) &= \mathbb{E}_b[\mathbb{1}_{b \geq i} \mathbb{E}[p(\tilde{\Delta}_{v_1,k} * \delta)]^{i-1} \mathbb{E}[p(\Delta_{v_1,k} * \delta)]^{b-i} Z(W) \frac{d}{d\Delta}|_{\Delta = \frac{1}{2}} p(\Delta * \delta)] \\ &\quad - c_i p'(\frac{1}{2}) \qquad (76) \\ &= 0. \qquad (77) \end{aligned}$$

So G_i is decreasing.

By summing up (69) for $i \geq 1$, we get

$$\mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{u,k+1}) - \exp(-\frac{1}{2}R^+_{u,k+1})] \leq (\sum_{i\geq 1} c_i)\mathbb{E}[\exp(-\frac{1}{2}\tilde{R}^+_{v_1,k}) - \exp(-\frac{1}{2}R^+_{v_1,k})]. \tag{78}$$

By (75), we have

$$\sum_{i\geq 1} c_i = \theta^2 \mathbb{E}_b[\mathbb{1}_{b\geq i}(1-\theta^2(\frac{d\theta^2-1}{d\theta^2}-\epsilon')_+)^{\frac{b-1}{2}}]Z(W) \tag{79}$$

$$\leq d\theta^2 \exp(-d(1-\sqrt{1-\theta^2(\frac{d\theta^2-1}{d\theta^2}-\epsilon')_+}))Z(W). \tag{80}$$

We can take $\epsilon' > 0$ small enough so that

$$\exp(-d(1-\sqrt{1-\theta^2(\frac{d\theta^2-1}{d\theta^2}-\epsilon')_+})) < (1+\epsilon)\exp(-d(1-\sqrt{1-\frac{(d\theta^2-1)_+}{d}})). \tag{81}$$

This finishes the proof for the Poisson tree case.

Appendix B. χ^2-capacity of broadcasting-on-tree channels

Proposition 4 *Consider the Broadcasting on Trees model defined in Section 2, with the following observation models:*

- $M_k^1 : \sigma_\rho \to \nu_{L_k}$, *where* $\nu_v \sim \mathrm{BSC}_\eta(\sigma_v)$*;*
- $M_k^2 : \sigma_\rho \to (\sigma_{L_k}, \omega_{T_k})$.
- $M_k^3 : \sigma_\rho \to \sigma_{L_k}$*;*
- $M_k^4 : \sigma_\rho \to \omega_{T_k}$ *with non-trivial survey channel* W*;*
- $M_k^5 : \sigma_\rho \to \omega_{L_k}$ *with non-trivial survey channel* W.

For each of the above channels, we have

- *If we work with regular trees, then*

$$\lim_{k\to\infty} C_{\chi^2}(M_k) \geq \frac{(d\theta^2-1)_+}{\theta^2(d-1)}. \tag{82}$$

- *If we work with Galton-Watston trees with Poisson offspring distribution, then*

$$\lim_{k\to\infty} C_{\chi^2}(M_k) \geq \frac{(d\theta^2-1)_+}{d\theta^2}. \tag{83}$$

Proof The χ^2-capacity is always non-negative, so the $d\theta^2 \leq 1$ case is automatic. In the following we assume $d\theta^2 > 1$.

First we observe that all M_k^i's are less degraded than M_k^1 for some suitable choice of η. This is obvious for $i = 2, 3$. Clearly M_k^4 is less degraded than M_k^5. That $M_k^5 \leq_{\deg} M_k^1$ follows from

(Roozbehani and Polyanskiy, 2019, Lemma 2, 3), where we can take $\eta = P_e(W)$. So by Lemma 4, we only need to prove the result for M_k^1.

We prove the result by applying Lemma 5. To do this, we need to find a BMS channel more degraded than M_k^1 which takes value in $\mathbb{R}$. One natrual choice is the majority decoder. We define

$$S_k = \sum_{v \in L_k} \nu_v. \tag{84}$$

Then the channel $\sigma_\rho \to S_k$ is clearly more degraded than M_k^1. We apply Proposition 5 to conclude. ■

Lemma 5 (Restatement of (Evans et al., 2000, Lemma 4.2(iii))) *Let $P : X \to Y$ be a BMS channel with Y a real variable, and with involution $Y \mapsto -Y$. Then $C_{\chi^2}(P) \ge \frac{(\mathbb{E}^+ Y)^2}{\operatorname{Var}(Y)}$.*

Proof Let $X \to (\Delta, Z)$ be the equivalent standard form of P. By Cauchy-Schwarz, we have

$$\mathbb{E}^+[(1-2\Delta)^2]\mathbb{E}^+[Y^2] \ge (\mathbb{E}^+[(1-2\Delta)|Y|])^2 = (\mathbb{E}^+ Y)^2. \tag{85}$$

This is equivalent to the desired result. ■

Proposition 5 *Assume $d\theta^2 > 1$. Consider the channel $\sigma_\rho \to S_k$ defined in (84).*

For regular trees,

$$\lim_{k \to \infty} \frac{\operatorname{Var}^+ S_k}{(\mathbb{E}^+ S_k)^2} = \frac{1-\theta^2}{d\theta^2 - 1}. \tag{86}$$

For Galton-Watson trees with Poisson offspring,

$$\lim_{k \to \infty} \frac{\operatorname{Var}^+ S_k}{(\mathbb{E}^+ S_k)^2} = \frac{1}{d\theta^2 - 1}. \tag{87}$$

Proof The regular tree case is proved in (Mossel et al., 2016, Lemma 3.4, 3.5). (Note that the expression for $\lim_{k\to\infty} \frac{\operatorname{Var}^+ S_k}{(\mathbb{E}^+ S_k)^2}$ on top of (Mossel et al., 2016, pg. 2224) is incorrect.)

Let us focus on the Poisson tree case. It is easy to see that

$$\mathbb{E}^+ S_k = (1-2\eta)(d\theta)^k. \tag{88}$$

Let ρ be the root, and $v_1, \ldots, v_b$ be its children. By variance decomposition, we have

$$\begin{aligned} \operatorname{Var}^+ S_{\rho,k+1} &= \operatorname{Var}^+ \mathbb{E}[S_{\rho,k+1}|b] + \mathbb{E}_b \operatorname{Var}^+(\mathbb{E}[S_{\rho,k+1}|b, \sigma_{v_1}, \ldots, \sigma_{v_b}]|b) \\ &\quad + \mathbb{E} \operatorname{Var}^+(S_{\rho,k+1}|b, \sigma_{v_1}, \ldots, \sigma_{v_b}). \end{aligned} \tag{89}$$

Let us compute each summand.

$$\operatorname{Var}^+ \mathbb{E}[S_{\rho,k+1}|b] = \operatorname{Var}^+(b\theta(1-2\eta)(d\theta)^k) = d\theta^2(1-2\eta)^2(d\theta)^{2k}. \tag{90}$$

$$\begin{aligned}
&\mathbb{E}_b \operatorname{Var}^+(\mathbb{E}[S_{\rho,k+1}|b,\sigma_{v_1},\dots,\sigma_{v_b}]|b) \\
&= \mathbb{E}_b \operatorname{Var}^+(\sum_{i\in[b]} \sigma_{v_i}(1-2\eta)(d\theta)^k|b) \\
&= \mathbb{E}_b[b(1-\theta^2)(1-2\eta)^2(d\theta)^{2k}] \\
&= d(1-\theta^2)(1-2\eta)^2(d\theta)^{2k}.
\end{aligned} \tag{91}$$

$$\mathbb{E}\operatorname{Var}^+(S_{\rho,k+1}|b,\sigma_{v_1},\dots,\sigma_{v_b}) = \mathbb{E}_b[b\sum_{i\in[b]}\operatorname{Var}^+ S_{v_i,k}] = d\operatorname{Var}^+ S_{\rho,k}. \tag{92}$$

Plugging (90)(91)(92) into (89), we get

$$\operatorname{Var}^+ S_{\rho,k+1} = d(1-2\eta)^2(d\theta)^{2k} + d\operatorname{Var}^+ S_{\rho,k}. \tag{93}$$

Solving (93) with initial value $S_{\rho,0} = 4\eta(1-\eta)$, we get

$$\begin{aligned}
\operatorname{Var}^+ S_{\rho,k} &= 4\eta(1-\eta)d^k + \sum_{i\in[k]} d^{k-i}d(1-2\eta)^2(d\theta)^{2i-2} \\
&= 4\eta(1-\eta)d^k + (1-2\eta)^2 d^k\frac{(d\theta^2)^k-1}{d\theta^2-1}.
\end{aligned} \tag{94}$$

Putting together (88)(94), we get the desired result. ■

Appendix C. Uniqueness of BP fixed point

Proposition 6 *Fix d, δ, and a (possibly trivial) BMS W. Recall definition (33) of the BP fixed point (the P_Δ definition) for BOTS (d,θ,W).*

- *If W is non-trivial ($P_e(W) < \frac{1}{2}$) and $C_1 < 1$ (where C_1 is defined in (40)), there is exactly one BP fixed point.*
- *If W is trivial and $d\theta^2 \le 1$, there is exactly one BP fixed point, which is trivial (the point distribution at $\Delta = \frac{1}{2}$).*
- *If W is trivial and $C_1 < 1$, there are exactly two BP fixed points, one is trivial and the other is non-trivial.*

The same (statements about number of fixed points) hold for BOTS $(\mathrm{Poi}(d),\theta,W)$ with C_1 replaced by C_2 (defined in (42)).

Proof If W is trivial and $d\theta^2 \le 1$, we are in the non-reconstruction regime and there is a unique BP fixed point, and it is trivial.

If W is trivial, there is one trivial fixed point. If W is non-trivial, the trivial distribution is not a fixed point. We prove that for any (d,θ,W) satisfying $C_1 < 1$ (or $C_2 < 1$ for Poisson trees), there is exactly one non-trivial fixed point.

Suppose there are two non-trivial fixed points P_Δ and Q_Δ. Let P be a BMS corresponding to P_Δ and Q be a BMS corresponding to Q_Δ. Let $r = \max\{P_e(P), P_e(Q)\}$. Then BSC_r is non-trivial and is more degraded than both P and Q.

We consider a Broadcasting on Tree model with three different types of observations:

- M_k^a: Observe $P(\sigma_v)$ for all $v \in L_k$;
- M_k^b: Observe $Q(\sigma_v)$ for all $v \in L_k$;
- M_k^c: Observe $\mathrm{BSC}_r(\sigma_v)$ for all $v \in L_k$.

By the same proof as Theorem 3, in the limit $k \to \infty$, M_k^a and M_k^c converge to the same BMS; the same holds for M_k^b and M_k^c. Therefore in the limit $k \to \infty$, M_k^a and M_k^b converge to the same BMS.

By the assumption that P and Q are BP fixed points, M_k^a are equivalent to P for all k, and M_k^b are equivalent to Q for all k. So P and Q are equivalent BMSs. This means $P_\Delta = Q_\Delta$. ■

Appendix D. Rough estimate of C in Mossel et al. (2016)

As we mentioned, Mossel et al. (2016) proves uniqueness of BP fixed point for BOT (without survey) and $d\theta^2 > C$ for an unspecified C. Can we extract explicit C from their work? First, we point out that taken literally, the proof demands at least $C > 75$. Second, we (heuristically!) argue below that it may be difficult to reduce C below 25 without significant modifications of the proof. We remark that this section is not meant to be rigorous and it may very well be that the method therein can be tweaked in ways we did not consider.

The proof in question is divided into the large θ case and the small θ case. First, they prove that there exists a $\theta^* > 0$ so that for $\theta \le \theta^*$, uniqueness of BP fixed point holds for large enough $d\theta^2$. Then they prove that for $\theta > \theta^*$, there exists d large enough so that uniqueness of BP holds. We focus on the small θ part and analyze their proof for θ close to 0.

In (Mossel et al., 2016, middle of page 2230) authors require $d\theta^2$ larger than about 75. Let us analyze how much improvement is possible. In the following, equation and lemmas refer to the cited paper.

- In Lemma 3.6, it is impossible to achieve an RHS better than $1 - \frac{1-\theta^2}{d\theta^2}$ by using a majority estimator (which is used by both their paper and the current paper).
- In (3.8), they applied Lemma 3.9 with $p = \frac{1}{4}$. Changing this exponent would result in a big change in the proof, so we leave it as-is.
- In Lemma 3.10, by Taylor expansion, it is impossible to improve RHS to $dm^{d-1}(\mathbb{E}A^2 - \mathbb{E}B^2)^2$.
- In Lemma 3.11, by Taylor expansion

$$\sqrt{\frac{1-x}{1+x}} = 1 - x + \frac{x^2}{2} - \frac{x^3}{2} + O(x^4),$$

  their proof cannot give a RHS better than $1 - \theta^2 x_k + \frac{1}{2}\theta^2$. Combined with Lemma 3.6, their proof does not give a RHS better than $1 - \theta^2(1 - \frac{1-\theta^2}{d\theta^2}) + \frac{1}{2}\theta^2$.

- In Lemma 3.12, RHS cannot be better than $2\theta^2$, because this is less than $|\frac{d}{dx}\frac{1-\theta^2 x}{\sqrt{1-\theta^2 x^2}}|$ at $x = -1$.

- Consequently, in Lemma 3.13, the leading factor in RHS cannot be better than $2\theta^2$.
- In (3.12), RHS cannot be better than $64d^2m^{d-2}(a-b)^2$ by using (3.8) with $p = \frac{1}{4}$ and Lemma 3.10.
- Combining the above, in the expression in the middle of Page 2230, RHS cannot be better than
$$64((2\theta^2)^2d^2(1-\theta^2(1-\frac{1-\theta^2}{d\theta^2})+\frac{1}{2}\theta^2)^{d-2})z.$$
Computation shows that, for the factor before z to be smaller than 1, we need at least $d\theta^2 \geq 26$ in the limit $\theta \to 0$.

Appendix E. Weak spatial mixing

In Section 5, we studied whether BP message (with recursion (27)) converges to the same value under perfect observation or no observation of leaves. A related question is weak spatial mixing, i.e., whether BP message converges to the same value under any observation of leaves.

Fix $k \geq 0$. Let $R_{L_k,0}$ and $R'_{L_k,0}$ be two boundary conditions. Define $R_{\rho,k}$ (resp. $R'_{\rho,k}$) by using (27) recursively, with initial condition $R_{L_k,0}$ (resp. $R'_{L_k,0}$). We say the model has weak spatial mixing if

$$\lim_{k\to\infty} \mathbb{E}_{T,\omega_{T_k}} \sup_{R_{L_k,0},R'_{L_k,0}} |f(R_{\rho,k}) - f(R'_{\rho,k})| = 0 \tag{95}$$

for all bounded continuous functions $f : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R}$.

In the following we focus on regular trees. It is known Bleher et al. (1995) that in the case there is no survey, $d\theta = 1$ is the threshold for WSM, i.e., when $d\theta < 1$, WSM holds; when $d\theta > 1$, WSM does not hold. The following result shows that for WSM with survey, this is still almost the case.

Theorem 4

- *For $d\theta < 1$ and any survey, WSM holds.*
- *For $d\theta > 1$, there exists $\epsilon = \epsilon(d,\theta) > 0$ such that for* BSC *survey with $P_e(W) > \frac{1}{2} - \epsilon$, WSM does not hold.*

Proof For $d\theta < 1$: For any node u, We have

$$\begin{aligned}\mathbb{E}|R_{u,k+1} - R'_{u,k+1}| &= \mathbb{E}|\sum_{v\in L_1(u)} (F_\theta(R_{v,k}) - F_\theta(R'_{v,k}))| &(96)\\ &\leq d\theta\mathbb{E}|R_{v_1,k} - R'_{v_1,k}|. &(97)\end{aligned}$$

The second step uses the fact that F_θ is θ-Lipschitz. Then

$$\mathbb{E}|R_{\rho,k+1} - R'_{\rho,k+1}| \leq d\theta\mathbb{E}|R_{\rho,k} - R'_{\rho,k}| \tag{98}$$

and we get the desired contraction.

For $d\theta > 1$: We separate the limit BP distribution for $(+)$-boundary condition and $(-)$-boundary condition. Because $d\theta > 1$, there exists $x > 0$ such that $dF_\theta(x) > x$. Let ϵ be small enough so that for all η with $P_e(\mathrm{BSC}_\eta) > \frac{1}{2} - \epsilon$, we have

$$dF_\theta(x) - \log \frac{1-\eta}{\eta} > x. \tag{99}$$

In this case, we can prove by induction that if we start with the $(+)$-boundary condition, then $R_{u,k} > x$ for all u and k. By symmetry, if we start with the $(-)$-boundary condition, then $R_{u,k} < -x$ for all u and k. So we get the desired separation. ■

Note that for the case $d\theta > 1$ we only prove for BSC survey. Numerical computation suggests that the result should hold for any BMS survey with sufficiently large P_e. Thus we make the following conjecture.

Conjecture 2 *For $d\theta > 1$, there exists $\epsilon = \epsilon(d, \theta) > 0$ such that for any BMS survey W with $P_e(W) > \frac{1}{2} - \epsilon$, WSM does not hold.*

Appendix F. Amenable graphs

Recall definition of the spin synchronization system and the (BI) property given in (44).

Definition 6 (Amenable graph (Alaoui and Montanari (2019))) *We say a graph G is amenable if $\inf\{|\partial S|/|S| : S \subset V \text{ finite}, o \in S\} = 0$, where $\partial S = \{u \in S : \exists v \notin S, (u, v) \in E\}$.*

Theorem 5 *Let G be an amenable graph. For any $\epsilon \in [0, 1)$, the (BI) holds for $(G, o, \theta, \mathrm{BEC}_\epsilon)$.*

Proof A consistent part of this proof is inspired by Lemma 6.3 in Alaoui and Montanari (2019). We reproduce it for a self-contained exposure. As in the proof of Theorem 1, let us replace the single parameter ϵ by a set of parameters $(\epsilon_u)_{u\in V(G)}$ (for each vertex u, X_u is revealed with probability $1 - \epsilon_u$), and let us denote $X^\epsilon_{\sim u} = \{X^\epsilon_v : v \in V(G), v \neq u\}$. For brevity, we write $B_n, \partial B_n$ for $B_n(o), \partial B_n(o)$. Then,

$$\frac{\partial}{\partial \epsilon_u} H(X_{\partial B_n} \mid Y, X^\epsilon) = I(X_u; X_{\partial B_n} \mid Y, X^\epsilon_{\sim u}),$$

and setting $\epsilon_u = \epsilon$ for every $u \in B_n$ we get

$$\frac{d}{d\epsilon} H(X_{\partial B_n} \mid Y, X^\epsilon) = \sum_{u\in B_n} I(X_u; X_{\partial B_n} \mid Y, X^\epsilon_{\sim u}). \tag{100}$$

Thus, integrating with respect to ϵ we get

$$\begin{aligned}
\int_\epsilon^1 \sum_{u\in B_n} I(X_u; X_{\partial B_n} \mid Y, X^{\epsilon'}_{\sim u}) d\epsilon' &= H(X_{\partial B_n} \mid Y) - H(X_{\partial B_n} \mid Y, X^\epsilon) && (101)\\
&\le H(X_{\partial B_n}) && (102)\\
&\le \sum_{u\in \partial B_n} H(X_u) && (103)\\
&= \log 2 |\partial B_n|. && (104)
\end{aligned}$$

If we divide by $|B_n|$, we get that for all $\epsilon < 1$

$$\int_\epsilon^1 \frac{1}{|B_n|} \sum_{u \in B_n} I\left(X_u; X_{\partial B_n} \mid Y_{B_n}, X^{\epsilon'}_{B_n}\right) \mathrm{d}\epsilon' \leq \log 2 \cdot \frac{|\partial B_n|}{|B_n|}. \tag{105}$$

Since G is amenable, the RHS is vanishing as $n \to \infty$. Note that the integrand in the LHS is bounded by $\log 2$, hence by bounded convergence theorem, we get that for all $\epsilon \in [0, 1)$

$$\lim_{n \to \infty} \frac{1}{|B_n|} \sum_{u \in B_n} I\left(X_u; X_{\partial B_n} \mid Y_{B_n}, X^{\epsilon}_{B_n}\right) = 0. \tag{106}$$

To conclude, notice that there exists $k \in \mathbb{N}$ such that

$$I(X_o; X_{\partial B_{k \cdot n}(o)} | Y, X^\epsilon) \leq \frac{1}{|B_n(o)|} \sum_{u \in B_n(o)} I(X_u; X_{\partial B_n(o)} \mid Y, X^\epsilon). \tag{107}$$

■