The Geometry of Community Detection via the MMSE Matrix

Galen Reeves*†, Vaishakhi Mayya*, and Alexander Volfovsky†
*Department of Electrical and Computer Engineering, Duke University
†Department of Statistical Science, Duke University

***Abstract*—The information-theoretic limits of community detection have been studied extensively for network models with high levels of symmetry or homogeneity. The contribution of this paper is to study a broader class of network models that allow for variability in the sizes and behaviors of the different communities, and thus better reflect the behaviors observed in real-world networks. Our results show that the ability to detect communities can be described succinctly in terms of a matrix of effective signal-to-noise ratios that provides a geometrical representation of the relationships between the different communities. This characterization follows from a matrix version of the I-MMSE relationship and generalizes the concept of an effective scalar signal-to-noise ratio introduced in previous work. We provide explicit formulas for the asymptotic per-node mutual information and upper bounds on the minimum mean-squared error. The theoretical results are supported by numerical simulations.**

***Index Terms*—community detection, I-MMSE, matrix factorization, recovery thresholds, stochastic block model.**

I. Introduction

Modern data problems often ask questions about how individuals (or computers or countries) interact or relate to each other within a network. A frequently studied problem in this context is that of community detection: how does one partition a network into clusters (or communities or groups) of nodes? A natural partition of a network is into communities that exhibit similar connection patterns, both within and between communities. A generative model for random networks called the stochastic block model (SBM) exhibits such behavior and hence much of the theoretical analysis of community detection has focused on it [1]. Under the SBM each individual belongs to one of k communities, and the probability of an edge between two individuals is exclusively a function of their community memberships.

The problem of community detection can be modeled in terms of a joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$ where $\boldsymbol{G}$ is a simple graph on n vertices and $\boldsymbol{X} = (X_1, \dots, X_n)$ is a collection of labels associated with the vertices. In the SBM this joint distribution is governed by two parameters: a probability vector p of each node being assigned to one of k labels, and a $k \times k$ matrix of probabilities Q where Q_{ab} is the probability of an edge between nodes in communities a and b. The community detection task is recovering the labels $\boldsymbol{X}$ given the graph $\boldsymbol{G}$ and potentially side information.

This was supported in part by funding from the Laboratory for Analytic Sciences (LAS) and by the NSF under Grant No. 1750362. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

Inspired by the work of Decelle et al. [2], a recent line of work has studied the information-theoretic limits of recovery when the distribution of $(\boldsymbol{X}, \boldsymbol{G})$ is known. Most of this work has focused on either the two-community SBM [3]–[9] or the so-called k-community symmetric SBM [7], [10]–[12]. In all of these cases, performance is summarized in terms of a single numerical value, which is often referred to as the effective signal-to-noise ratio of the problem. General SBMs have been considered by Abbe and Sandon [10] who characterize conditions for weak recovery and also by Lesieuir et al. [7] who analyze the performance of an approximate message passing algorithm.

A different line of research within the statistics community has focused on settings where the parameters of the distribution, such as the distribution of communities and the conditional probabilities of edges, are unknown quantities that must also be inferred, along with the community memberships [13], [14]. While the models considered in this literature are highly flexible, the conditions needed for consistent recovery of communities corresponds to a very high SNR regime relative to the information theoretic analysis.

A. Our Contributions

The contribution of this paper is to characterize the information-theoretic limits for a large class of degree-balanced SBMs. In comparison to the symmetric SBM, these models allow for variability in the sizes and behaviors of the different communities, and thus reflect behaviors observed in real-world networks. While previous work is limited to a scalar measure of performance for the overall community detection problem, we introduce a multivariate measure of performance, the minimum mean-squared error (MMSE) matrix, which describes detection limits for individual communities. For example, this matrix allows us to characterize settings where some of the communities can be detected while other cannot.

Our analysis of the community detection problem leverages a matrix version of the I-MMSE [15] relation, which both simplifies and generalizes techniques used in previous work. In particular, the upper bound on the mutual information in Theorem 2 is a consequence of a novel non-asymptotic inequality that holds under *any* distribution on the community labels. Many of our techniques can be applied more generally to other high-dimensional inference problems, including matrix and tensor factorization.

B. Overview of Approach

This paper introduces a multivariate measure of performance, which we refer to as the MMSE matrix:

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{G}}[\mathsf{Cov}(X_i \mid \boldsymbol{G})]. \tag{1}$$

In this expression, $\mathsf{Cov}(X_i \mid \boldsymbol{G})$ is the covariance matrix of the i-th node's label after is has been embedded in to an ℓ-dimensional Euclidean space (where ℓ is either k or $k-1$). We show that the MMSE matrix provides important geometrical information about the uncertainty in the community memberships. While the trace of the MMSE matrix corresponds to standard measures of performance such as the average overlap, the information provided by individual entries in the MMSE matrix can be used to answer more nuanced questions about which of the community relationships can (or cannot) be recovered.

One of the key ideas in this paper is to focus on community detection in the setting where there is additional covariate information about the labels. Specifically, we assume that one has side-information from the signal-plus-noise model:

$$\boldsymbol{Y} = \boldsymbol{X} S^{1/2} + \boldsymbol{N}, \tag{2}$$

where S is an $\ell \times \ell$ positive semidefinite matrix, known as the matrix SNR, and $\boldsymbol{N}$ is an $n \times \ell$ matrix with i.i.d. standard Gaussian entries.

The introduction of the signal-plus-noise model plays an important role both for our analysis and for our interpretation of the results. For example, it allows us to leverage the matrix I-MMSE relation [15] to characterize the MMSE matrix in terms of the gradient of the mutual information:

$$\nabla_S I(\boldsymbol{X}; \boldsymbol{G}, \boldsymbol{Y}) = \frac{n}{2}\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}). \tag{3}$$

Remarkably, this relationship holds generally for any joint distribution on the pair $(\boldsymbol{X}, \boldsymbol{G})$. Notice that the matrix MMSE in (1) is obtained by evaluating this expression at $S = 0$.

The signal-plus-noise model also provides a natural way to address non-identifiability issues that arise when the distribution over the labels is invariant to permutations. The key idea is that in the large-n limit, an arbitrarily small amount of side-information is sufficient to break the symmetry in the model. Hence, focusing on the double limit

$$\lim_{S \to 0} \lim_{n \to \infty} \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}),$$

provides a meaningful and interpretable measure of average performance that bypasses the need to optimize over an equivalence class of permutations.

Section IV provides explicit formulas for the per-vertex mutual information and MMSE matrix in the large-n limit. These formulas are stated for a degree-balanced stochastic block model and are given in terms of single-letter formulas. Numerical simulations are provided in Section V. The proofs are omitted due to space constraints and are available in the online version of the paper

C. Notation

We use $\mathbb{S}^d$, $\mathbb{S}^d_+$ to denote the space $d \times d$ symmetric matrices and symmetric positive semi-definite matrices, respectively. Given a symmetric positive semi-definite matrix S, we use $S^{1/2}$ to denote the unique positive semi-definite square root. Given matrix $A, B \in \mathbb{S}^d$, the relation $A \preceq B$ means that $B - A \in \mathbb{S}^d_+$.

II. MMSE Matrix

Without loss of generality, the community labels can be embedded into finite dimensional Euclidean space. Two useful representations are considered in the following.

A. Standard Basis Representation

A natural embedding associates the labels with the standard basis vectors $\{e_1, \dots, e_k\}$ in $\mathbb{R}^k$, i.e., the columns of the identity matrix. Under this representation, the expected value of a label vector X_i is a point on the probability simplex. The conditional covariance is defined by

$$\mathsf{Cov}(X_i \mid \boldsymbol{G}) \triangleq \mathbb{E}_{\boldsymbol{X}|\boldsymbol{G}}\Big[(X_i - \mathbb{E}[X_i|\boldsymbol{G}])^T (X_i - \mathbb{E}[X_i|\boldsymbol{G}])\Big],$$

and the MMSE matrix is defined according to (1). By the data processing inequality for MMSE, this matrix satisfies

$$0 \preceq \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \preceq \mathsf{MMSE}(\boldsymbol{X}) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathsf{Cov}(X_i).$$

As a consequence, the difference between the MMSE matrix and covariance provides a measure of the difference between the prior and posterior marginals of the labels.

Proposition 1. *Under the standard basis representation, the $k \times k$ MMSE matrix satisfies*

$$\begin{aligned}&\operatorname{tr}(\mathsf{MMSE}(\boldsymbol{X}) - \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}))\\ &\qquad = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{G}}\Big[\big\|P_{X_i|\boldsymbol{G}}(\cdot \mid G) - P_{X_i}(\cdot)\big\|_2^2\Big].\end{aligned}$$

Furthermore, the individual entries of the MMSE matrix also provide information about different recovery tasks. For example, consider the problem of determining whether a label belongs to a subset $A \subset [k]$. If we define $\mathbf{1}_A = \sum_{\ell \in A} e_\ell$, then $\mathbf{1}_A^T X_i$ is binary random variable indicating whether the i-th label belongs to A. Summing the entries in the MMSE matrix indexed by the set A provides a measures of the average error probability:

$$\mathbf{1}_A^T \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \mathbf{1}_A = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_G\big[\mathsf{Var}(\mathbf{1}_A^T X_i \mid \boldsymbol{G})\big].$$

B. Whitened Representation

Next, we focus on the setting where the labels are identically distributed with probability vector $p = (p_1, \dots, p_k)$. The whitened representation is defined to be of a set of k points $\{\mu_1, \dots, \mu_k\}$ in $\mathbb{R}^{k-1}$ with the property that

$$\sum_{\ell} p_\ell \mu_\ell = 0, \qquad \sum_{\ell} p_\ell \mu_\ell \mu_\ell^T = I_{k-1}.$$

Under the whitened representation, each label vector has zero mean and identity covariance and thus the MMSE matrix satisfies $0 \preceq \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \preceq I_{k-1}$.

Remark 1 (Unique Specification of Whitened Representation)**.** The whitened representation can be defined explicitly as a function of p as follows. Let $\tilde{p} = (\sqrt{p_1}, \dots, \sqrt{p_k})^T$ and apply the Gram-Schmidt process to the vectors $\{\tilde{p}, e_1, \dots, e_{k-1}\}$ to obtain an orthonormal basis for $\mathbb{R}^k$ of the form $[\tilde{p}, B]$ where B is $k \times (k-1)$. Then, the support of the whitened representation is related to the standard basis vectors according to

$$\mu_\ell = B^T P^{-1/2} e_\ell \quad \iff \quad e_\ell = p + P^{1/2} B \mu_\ell, \qquad (4)$$

where $P = \operatorname{diag}(p)$. This construction is unique and has the useful property that μ_ℓ lies in the span of $\{e_1, \dots, e_\ell\}$.

Proposition 2. *If the labels are identically distributed then the $(k-1) \times (k-1)$ MMSE matrix of the whitened representation satisfies*

$$\operatorname{tr}(I - \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G})) = \frac{1}{n} \sum_{i=1}^{n} \chi^2(P_{X_i, \boldsymbol{G}} \,\|\, P_{X_i} P_{\boldsymbol{G}}),$$

where $\chi^2(P \,\|\, Q) = \int (\mathrm{d}P/\mathrm{d}Q)^2 \, \mathrm{d}Q$ denotes the chi-squared divergence.

For the purposes of analysis, the two representations described above are equivalent in the sense that there is a one-to-one mapping between the $k \times k$ MMSE matrix defined under the standard basis representation and the $(k-1) \times (k-1)$ MMSE matrix defined under the whitened representation. For notational convenience we work in the whitened representation.

III. Signal-Plus-Noise Problem

Our analysis uses properties of the signal-plus-noise model given in (2). Throughout this section we will assume the labels are drawn i.i.d. according to a probability vector $p = (p_1, \dots, p_k)$ with strictly positive entries and are supported on the whitened representation described in Section II-B. For each $S \in \mathbb{S}_+^{k-1}$, the task of recovering $\boldsymbol{X}$ from $\boldsymbol{Y}$ decouples into n independent copies of the problem

$$Y = S^{1/2} X + N,$$

where X is supported on $\{\mu_1, \dots, \mu_k\}$ with probability vector p and $N \sim \mathcal{N}(0, I)$ is independent Gaussian noise.

Following [15] we define the the mutual information function $I_X : \mathbb{S}_+^{k-1} \to [0, \infty)$ and matrix-valued MMSE function $M_X : \mathbb{S}_+^{k-1} \to \mathbb{S}_+^{k-1}$ according to

$$I_X(S) = I(X; Y) \qquad (5)$$

$$M_X(S) = \mathbb{E}[\mathsf{Cov}(X \mid Y)]. \qquad (6)$$

The gradient and Hessian of $I_X(S)$ are given by [15, Lemma 4]

$$\nabla_S I_X(S) = \frac{1}{2} M_X(S) \qquad (7)$$

$$\nabla_S^2 I_X(S) = -\frac{1}{2} \mathbb{E}[\mathsf{Cov}(X \mid Y) \otimes \mathsf{Cov}(X \mid Y)], \qquad (8)$$

where $\otimes$ denotes the Kronecker product. We note that these functions can be approximated using numerical integration methods or Monte-Carlo sampling.

IV. Main Results

A. Degree-Balanced SBM

An SBM is frequently parameterized in terms of the tuple (n, p, Q) where $p = (p_1, \dots, p_k)$ is a distribution over k communities and $Q \in [0,1]^{k \times k}$ is a symmetric matrix such that Q_{ab} is the probability of an edge between nodes in communities a and b. The average degree of an SBM corresponds to the expected number of edges for a node chosen uniformly at random and is denoted by d. An SBM is said to be *degree-balanced* if the expected degree of a node does not depend on its community assignments. This condition is equivalent to saying that Qp is proportional to the all ones vector.

For the purposes of this paper, it is useful to consider a different parameterization of the degree-balanced SBM in terms of the tuple (n, d, p, R) where d is the average degree and $R \in \mathbb{S}^{k-1}$. Using this parameterization, the entries of Q are given by

$$Q_{ab} = \frac{d}{n} + \frac{\sqrt{d(1 - d/n)}}{n} \mu_a^T R \mu_b, \qquad (9)$$

where $\{\mu_1, \dots, \mu_k\}$ are defined as a function of p using the procedure described in Remark 1. The tuple (n, d, p, R) is valid only if the entries of Q are between zero and one.

The matrix R quantifies the relative strength of relationships between different communities. The eigenvalue decomposition is given by

$$R = U \operatorname{diag}(\lambda) U^T,$$

where $\lambda = (\lambda_1, \dots, \lambda_{k-1})$ are real numbers. To simplify the analysis, we will assume throughout that all the eigenvalues are nonzero so that R is invertible.

We remark that the definition of signal-to-noise ratio given by Abbe and Sandon [10, Section 2.1] corresponds to $\max_i \lambda_i^2$. Furthermore, for the special case of $k = 2$ communities, the representation of X_i is one-dimensional and our formulation is equivalent to the one given by Lelarge and Miolane [5].

B. Formulas for Mutual Information and MMSE

Our analysis focuses on a sequence of degree-balanced SBMs where the parameters (p, R) are fixed as the size of the network n scales to infinity. Additionally, we make two assumptions.

Assumption 1 (Diverging Average Degree)**.** The average degree of the network d increases with n such that both d and $(n - d)$ tend to infinity.

Assumption 2 (Definite Matrix)**.** The matrix R is either positive definite or negative definite.

Our first result is stated in terms of the potential function $\mathcal{F} : \mathbb{S}_+^{k-1} \to \mathbb{R}_+$ defined by

$$\mathcal{F}(\Delta) = I_X(\Delta) + \frac{1}{4} \operatorname{tr}\Big(\big(R - R^{-1} \Delta\big)^2 \Big). \qquad (10)$$

where $I_X(\cdot)$ is defined by (5). Notice that the first term in the potential function is defined by the distribution the labels p whereas the second term is defined by the matrix R. By the matrix I-MMSE relation [15], it can be verified that every stationary point of $\mathcal{F}(\Delta)$ satisfies the fixed-point equation

$$M_X(\Delta) = I - R^{-1}\Delta R^{-1}. \tag{11}$$

where $M_X(\cdot)$ is defined by (6). Noting that $M_X(0) = I$, we see that $\Delta = 0$ is always a stationary point. Furthermore, every solution of (11) belongs to the set $\{\Delta : 0 \preceq \Delta \preceq R^2\}$.

Theorem 1. *Under Assumptions 1 and 2,*

$$\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{G}) = \min_{\Delta \in \mathbb{S}_+^{k-1}} \mathcal{F}(\Delta),$$

where $\mathcal{F}(\Delta)$ is given in (10).

The next result provides an upper bound on the mutual information in the setting where side information is generated according to the signal-plus-noise model (2) parameterized by a positive semi-definite matrix S. To characterize this setting, we define the modified potential function:

$$\mathcal{F}(\Delta, S) = I_X(S + \Delta) + \frac{1}{4}\operatorname{tr}\Big(\big(R - R^{-1}\Delta\big)^2\Big). \tag{12}$$

Notice that the main difference from (11) is that the side information changes the prior information about the labels.

Theorem 2. *Suppose that $\boldsymbol{Y}$ is generated according to the signal-plus-noise model* (2) *with matrix $S \in \mathbb{S}_+^{k-1}$. Under Assumption 1,*

$$\limsup_{n\to\infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{G}, \boldsymbol{Y}) \le \min_{\Delta \in \mathbb{S}_+^{k-1}} \mathcal{F}(\Delta, S).$$

where $\mathcal{F}(\Delta, S)$ is given in (12).

Remark 2. Similar to previous work [3]–[8], our proofs of Theorems 1 and 2 use a channel universality argument to relate the community detection problem to a low-rank estimation problem. Assumption 2 is needed for the proof of Theorem 1, which leverages [5, Theorem 12]. To prove Theorem 2 we develop a novel variation of the Guerra interpolation method that exploits the matrix I-MMSE relationship [15] to provide a general and non-asymptotic upper bound.

Next, we recall that that by the data processing inequality, the MMSE matrix satisfies

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \succeq \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}),$$

for all $S \in \mathbb{S}_+^{k-1}$. For any fixed problem size n, the difference between these matrices converges to zero as $S \to 0$. However, in the large-n limit it is possible that the limiting behavior is discontinuous with respect to S. This can occur, for example, when the SBM is invariant to permutations of the labels and hence $\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) = \mathsf{MMSE}(\boldsymbol{X})$. The presence of side-information with an arbitrarily small positive definite matrix S is sufficient to break the permutation invariance, and thus the small-S limit provides a meaningful measure of recovery performance that overcomes the non-identifiability issues.

The following result follows from the matrix I-MMSE relation and Theorems 1 and 2.

Theorem 3. *Consider Assumptions 1 and 2. For every $S \succ 0$,*

$$\limsup_{n\to\infty} \lambda_{\max}(\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}) - M_X(\Delta^*)) \le 0$$

where Δ^* denotes any minimizer of $\mathcal{F}(\Delta)$. In other words,

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}) \preceq M_X(\Delta^*) + o_n(1),$$

where $o_n(1)$ denotes a sequence of symmetric matrices that converges to zero as $n \to \infty$.

Following heuristic arguments, we postulate that upper bounds in Theorem 2 is asymptotically tight and that the MMSE matrix satisfies

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}) = M_X(S + \Delta^*) + o_n(1)$$

for almost all S, where Δ^* is the unique minimizer of $\mathcal{F}(\cdot, S)$. These conjectures are supported by the numerical experiments in Section V.

C. Implications for Weak Recovery

In the context of community detection, weak recovery refers to the ability to produce an estimate of the community labels that is positively correlated with the ground truth. Within the literature, the measure of correlation usually includes an additional step that maximizes over all permutations of the labels; see e.g., [10, Section 2].

Using the results in this paper, we see that a natural alternative is to focus on the small-S behavior of the MMSE matrix. In particular, we say that weak recovery is possible if

$$\inf_{S \succ 0} \liminf_{n\to\infty} \|\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}, \boldsymbol{Y}) - \mathsf{MMSE}(\boldsymbol{X})\| > 0. \tag{13}$$

In view of this definition, we see that Theorem 3 provides a sufficient condition for weak recovery.

Theorem 4. *Consider Assumptions 1 and 2. If $\mathcal{F}(\cdot)$ has a minimizer Δ^* with $M_X(\Delta^*) \prec M_X(0)$ then weak recovery in the sense of* (13) *is possible.*

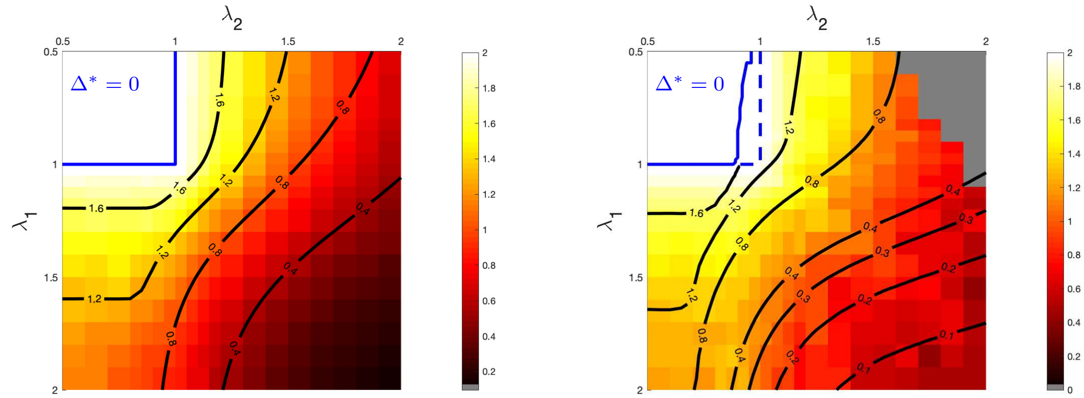
Evaluating the Hessian of the potential function at zero provides a simple test to determine whether $\Delta = 0$ is a local minimum. Using (8), it can be shown that

$$\nabla^2 \mathcal{F}(\Delta)\Big|_{\Delta=0} \propto R^{-1} \otimes R^{-1} - I_{(k-1)^2}.$$

Therefore, if $\max_i \lambda_i^2 > 1$ then $\Delta = 0$ is not a local minimizer.

V. Numerical Experiments

This section compares the asymptotic bounds given in Section IV with the MSE obtained using belief propagation. The case of the three-community degree balanced SBM (n, d, p, R) is illustrated in Figure 1. The black contour lines correspond to the trace of $M_X(\Delta^*)$ where Δ^* is the global minimizer of the potential function defined in (10). The heat map values correspond to the empirical MSE of the belief propagation algorithm described in [2] applied to a network of size $n = 10^5$ with average degree $d = 10$. Each pixel is the average of ten independent trials and the MSE is measured with respect to the whitened basis representation.



(a) $p = (1/3, 1/3, 1/3)$ (b) $p = (0.6, 0.3, 0.1)$

Fig. 1: Comparison of upper bound on $\mathrm{tr}(\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}))$ given in Theorem 3 (black contour lines) and the empirical MSE of belief propagation (heat map) on a network of size $n = 10^5$ with average degree $d = 10$. In both cases, $R = \mathrm{diag}(\lambda_1, \lambda_2)$. The upper bound on the weak recovery threshold given in Theorem 4 (solid blue line) corresponds to the boundary where $\Delta^* = 0$. The weak recovery threshold for acyclic BP [10] (dashed blue line) corresponds to $\max(\lambda_1, \lambda_2) = 1$.

In the case of uniform community assignments (Figure 1a), our upper bound on the weak detection threshold is equal to the weak recovery limit for acyclic BP [10]. Furthermore, we see that there is a close correspondence between the asymptotic formula and the empirical results. Note that the special case $\lambda_1 = \lambda_2$ corresponds to the three-community symmetric SBM.

In the case of non-uniform community assignments (Figure 1b) there exists a region of the parameter space where weak recovery is possible with $\max(\lambda_1, \lambda_2) < 1$. The existence of such a region has been shown previously in the special case of the two-community asymmetric SBM [4]. We also see that the asymptotic formulas match the empirical behavior qualitatively, although the empirical MSE is worse than is suggested by the formulas. The grey region in Figure 1b corresponds to settings where (n, d, p, R) does not define a valid SBM.

Numerical Approximation of Formulas: We use Monte Carlo sampling to approximately evaluate the functions I_X and M_X, and we use the concave-convex procedure [16] to explore the local minima of the potential function. Starting is an initialization point Δ^0, a sequence of iterates is obtained according to

$$\Delta^{t+1} = (1-\epsilon)\big(R^2 - RM_X(\Delta^t)R\big) + \epsilon\Delta^t,$$

where $\epsilon \in [0, 1)$ is a dampening parameter.

Acknowledgment

The authors thank Lenka Zdeborová for providing initial direction on this problem and Jiaming Xu for a generalization of the channel universality theorem [8, Theorem I.1].

References

[1] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[2] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physitcal Review Eq*, vol. 84, no. 6, Dec. 2011.

[3] Y. Deshpande, E. Abbe, and A. Montanari, "Asymptotic mutual information for the balanced binary stochastic block model," *Information and Inference*, vol. 6, no. 2, pp. 125–170, Jun. 2017.

[4] F. Caltagirone, M. Lelarge, and L. Miolane, "Recovering asymmetric communities in the stochastic block model," *IEEE Transactions on Network Science and Engineering*, vol. 5, no. 3, pp. 237–246, 2018.

[5] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probability Theory and Related Fields*, 2018.

[6] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Adv. Neural Inf. Process. Syst.*, vol. 29, Barcelona, Spain, 2016, pp. 424–432.

[7] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications," *Journal of Statistical Mechanics: Theory and Experiment*, Jul. 2017.

[8] F. Krzakala, J. Xu, and L. Zdeborová, "Mutual information in rank-one matrix estimation," in *Proc. IEEE Inform. Theory Workshop*, 2016.

[9] Y. Deshpande, A. Montanari, E. Mossel, and S. Sen, "Contextual stochastic block models," in *NeurIPS*, 2018.

[10] E. Abbe and C. Sandon, "Proof of the achievability conjectures for the general stochastic block model," *Communications on Pure and Applied Mathematics*, vol. 71, no. 7, pp. 1334–1406, 2018.

[11] J. Banks, C. Moore, J. Neeman, and P. Netrapalli, "Information-theoretic thresholds for community detection in sparse networks," in *Conference On Learning Theory*, 2016.

[12] E. Abbe, "Community detection and stochastic block models: Recent developments," *Journal of Machine Learning Research*, vol. 18, no. 177, pp. 1–86, 2018.

[13] K. Rohe, S. Chatterjee, B. Yu *et al.*, "Spectral clustering and the high-dimensional stochastic blockmodel," *The Annals of Statistics*, vol. 39, no. 4, pp. 1878–1915, 2011.

[14] S. Suwan, D. S. Lee, R. Tang, D. L. Sussman, M. Tang, and C. E. Priebe, "Empirical Bayes estimation for the stochastic blockmodel," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 761–782, 2016.

[15] G. Reeves, H. D. Pfister, and A. Dytso, "Mutual information as a function of matrix SNR for linear gaussian channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Vail, CO, Jun. 2018.

[16] A. L. Yuille and A. Rangarajan, "The concave-convex procedure (CCCP)," in *Adv. Neural Inf. Process. Syst.*, 2002, pp. 1033–1040.