

# Asymptotic Mutual Information for the Binary Stochastic Block Model

Yash Deshpande  
Electrical Engineering  
Stanford University

Emmanuel Abbe  
Electrical Engineering and PACM  
Princeton University

Andrea Montanari  
Electrical Engineering and Statistics  
Stanford University

**Abstract**—We develop an information-theoretic view of the stochastic block model, a popular statistical model for the large-scale structure of complex networks. A graph  $G$  from such a model is generated by first assigning vertex labels at random from a finite alphabet, and then connecting vertices with edge probabilities depending on the labels of the endpoints. In the case of the symmetric two-group model, we establish an explicit ‘single-letter’ characterization of the per-vertex mutual information between the vertex labels and the graph, when the graph average degree diverges.

The explicit expression of the mutual information is intimately related to estimation-theoretic quantities, and—in particular—reveals a phase transition at the critical point for community detection. Below the critical point the per-vertex mutual information is asymptotically the same as if edges were independent of the vertex labels. Correspondingly, no algorithm can estimate the partition better than random guessing. Conversely, above the threshold, the per-vertex mutual information is strictly smaller than the independent-edges upper bound. In this regime there exists a procedure that estimates the vertex labels better than random guessing.

## I. INTRODUCTION AND MAIN RESULTS

The stochastic block model is the simplest statistical model for networks with a community (or cluster) structure. As such, it has attracted considerable amount of work across statistics, machine learning, and theoretical computer science [1], [2], [3], [4], [5]. A random graph  $G = (V, E)$  from this model has its vertex set  $V$  partitioned into  $r$  groups, which are assigned  $r$  distinct labels. The probability of edge  $(i, j)$  being present depends on the group labels of vertices  $i$  and  $j$ .

In the context of social network analysis, groups correspond to social communities [1]. For other data-mining applications, they represent latent attributes of the nodes [6]. In all of these cases, we are interested in inferring the vertex labels from a single realization of the graph.

In this paper we develop an information-theoretic viewpoint on the stochastic block model. Namely, we develop an explicit (‘single-letter’) expression for the per-vertex conditional entropy of the vertex labels given the graph. Equivalently, we compute the asymptotic per-vertex mutual information between the graph and the vertex labels. Our results hold for in the large network limit, under suitable assumptions on the model parameters. In particular, we require that the mean vertex degree in the graph diverge (albeit arbitrarily slowly). While the asymptotic mutual information is of independent in-

terest but, as we will see, it is intimately related to estimation-theoretic quantities.

In this paper, we focus on the symmetric binary model. Namely, we assume the vertex set  $V = [n] \equiv \{1, 2, \dots, n\}$  to be partitioned into two sets  $V = V_+ \cup V_-$ , with  $\mathbb{P}(i \in V_+) = \mathbb{P}(i \in V_-) = 1/2$  independently across vertices  $i$ . Conditional on the edge labels, edges are independent with

$$\mathbb{P}((i, j) \in E | V_+, V_-) = \begin{cases} p_n & \text{if } \{i, j\} \subseteq V_+ \text{ or } V_- \\ q_n & \text{otherwise.} \end{cases} \quad (1)$$

Throughout we will denote by  $\mathbf{X} = (X_i)_{i \in V}$  the set of vertex labels  $X_i \in \{+1, -1\}$ , and we will be interested in the conditional entropy  $H(\mathbf{X} | \mathbf{G})$  or—equivalently—the mutual information  $I(\mathbf{X}; \mathbf{G})$  in the limit  $n \rightarrow \infty$ . We will write  $\mathbf{G} \sim \text{SBM}(n; p, q)$  (or  $(\mathbf{X}, \mathbf{G}) \sim \text{SBM}(n; p, q)$ ) to imply that the graph  $G$  is distributed according to the stochastic block model with  $n$  vertices and parameters  $p, q$ .

Since we are interested in the large  $n$  behavior, two preliminary remarks are in order:

- 1) *Normalization.* We obviously have<sup>1</sup>  $0 \leq H(\mathbf{X} | \mathbf{G}) \leq n \log 2$ . It is therefore natural to study the per-vertex entropy  $H(\mathbf{X} | \mathbf{G})/n$ . As we will see, depending on the model parameters, this will take a value between 0 and  $\log 2$ .
- 2) *Scaling.* The reconstruction problem becomes easier when  $p_n$  and  $q_n$  are well separated, and more difficult when they are closer to each other. For instance, in an early contribution, Dyer and Frieze [2] proved that the labels can be reconstructed exactly—modulo an overall flip—if  $p_n = p > q_n = q$  are distinct and independent of  $n$ . In particular, such exact recovery results imply that  $H(\mathbf{X} | \mathbf{G})/n \rightarrow 0$ . In this regime, the ‘signal’ is so strong that the conditional entropy is trivial. Letting  $\bar{p}_n = (p_n + q_n)/2$  be the average edge probability. It turns out that the relevant ‘signal-to-noise ratio’ (SNR) is given by the following parameter:

$$\lambda_n = \frac{n(p_n - q_n)^2}{4\bar{p}_n(1 - \bar{p}_n)}. \quad (2)$$

Indeed, we will see that  $H(\mathbf{X} | \mathbf{G})/n$  of order 1, and has a strictly positive limit when  $\lambda_n$  converges to  $\lambda$  of order

<sup>1</sup>Unless explicitly stated otherwise, logarithms will be in base  $e$ , and entropies will be measured in nats.

one. In this regime, the fraction of incorrectly labeled vertices has a limit in  $(0, 1)$ .

#### A. Main results: Limiting mutual information and MMSE

Our main result provides a single-letter characterization for the per-vertex mutual information. This is given in terms of an *effective Gaussian scalar channel*. Namely, define the Gaussian channel

$$Y_0 = Y_0(\gamma) = \sqrt{\gamma} X_0 + Z_0, \quad (3)$$

where  $X_0 \sim \text{Uniform}(\{+1, -1\})$  independent<sup>2</sup> of  $Z_0 \sim \mathcal{N}(0, 1)$ . We denote by  $\text{mmse}(\gamma)$  and  $I(\gamma)$  the corresponding minimum mean square error and mutual information:

$$I(\gamma) = \mathbb{E} \log \left\{ \frac{dp_{Y|X}(Y_0(\gamma)|X_0)}{dp_Y(Y_0(\gamma))} \right\}, \quad (4)$$

$$\text{mmse}(\gamma) = \mathbb{E} \{ (X_0 - \mathbb{E} \{X_0|Y_0(\gamma)\})^2 \}. \quad (5)$$

In the present case, these quantities can be written explicitly as Gaussian integrals of elementary functions:

$$I(\gamma) = \gamma - \mathbb{E} \log \cosh(\gamma + \sqrt{\gamma} Z_0), \quad (6)$$

$$\text{mmse}(\gamma) = 1 - \mathbb{E} \{ \tanh(\gamma + \sqrt{\gamma} Z_0)^2 \}. \quad (7)$$

The following theorem gives a single-letter characterization of the limiting per-vertex mutual information.

**Theorem I.1.** *For any  $\lambda > 0$ , let  $\gamma_* = \gamma_*(\lambda)$  be the largest non-negative solution of the equation:*

$$\gamma = \lambda(1 - \text{mmse}(\gamma)). \quad (8)$$

We refer to  $\gamma_*(\lambda)$  as to the effective signal-to-noise ratio. Further, define  $\Psi(\gamma, \lambda)$  by:

$$\Psi(\gamma, \lambda) = \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma). \quad (9)$$

Let the graph  $\mathbf{G}$  and vertex labels  $\mathbf{X}$  be distributed according to the stochastic block model with  $n$  vertices and parameters  $p_n, q_n$  (i.e.  $(\mathbf{G}, \mathbf{X}) \sim \text{SBM}(n; p_n, q_n)$ ) and define  $\lambda_n \equiv n(p_n - q_n)^2 / (4\bar{p}_n(1 - \bar{p}_n))$ .

Assume that, as  $n \rightarrow \infty$ , (i)  $\lambda_n \rightarrow \lambda$  and (ii)  $n\bar{p}_n(1 - \bar{p}_n) \rightarrow \infty$ . Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{X}; \mathbf{G}) = \Psi(\gamma_*(\lambda), \lambda) = \inf_{\gamma \in [0, \infty]} \Psi(\gamma, \lambda). \quad (10)$$

This theorem and its proof has implications on the minimum error that can be achieved in estimating the labels  $\mathbf{X}$  from the graph  $\mathbf{G}$ . For reasons that will become clear below, a natural metric is given by the matrix minimum mean square error

$$\text{MMSE}_n(\lambda) \equiv \frac{1}{n(n-1)} \mathbb{E} \left\{ \left\| \mathbf{X} \mathbf{X}^\top - \mathbb{E} \{ \mathbf{X} \mathbf{X}^\top | \mathbf{G} \} \right\|_F^2 \right\}. \quad (11)$$

<sup>2</sup>Throughout the paper, we will generally denote scalar equivalents of vector/matrix quantities with the 0 subscript

**Theorem I.2.** *Under the assumptions of Theorem I.1 (in particular assuming  $\lambda_n \rightarrow \lambda$  as  $n \rightarrow \infty$ ), the following limit holds for the matrix minimum mean square error*

$$\lim_{n \rightarrow \infty} \text{MMSE}_n(\lambda_n) = 1 - \frac{\gamma_*(\lambda)^2}{\lambda^2}. \quad (12)$$

Further, this implies  $\lim_{n \rightarrow \infty} \text{MMSE}_n(\lambda_n) = 1$  for  $\lambda \leq 1$  and  $\lim_{n \rightarrow \infty} \text{MMSE}_n(\lambda_n) < 1$  for  $\lambda > 1$ .

A couple of remarks are in order

**Remark I.3.** Notice that our assumptions require  $n\bar{p}_n(1 - \bar{p}_n) \rightarrow \infty$  at any, arbitrarily slow, rate. In words, this corresponds to the graph average degree diverging at any, arbitrarily slow, rate. Recently (see Section II for a discussion of this literature), there has been considerable interest in the case of bounded average degree, namely

$$p_n = \frac{a}{n}, \quad q_n = \frac{b}{n}, \quad (13)$$

with  $a, b$  bounded. Our proof gives an explicit error bound in terms of problem parameters even when  $n\bar{p}_n(1 - \bar{p}_n)$  is of order one. Hence we are able to characterize the asymptotic mutual information for large-but-bounded average degree up to an offset:

$$\limsup_{n \rightarrow \infty} \left| \frac{1}{n} I(\mathbf{X}; \mathbf{G}) - \Psi(\gamma_*(\lambda), \lambda) \right| \leq \frac{C\lambda^3}{\sqrt{a+b}}, \quad (14)$$

Further, our mild condition on diverging average degree should be contrasted with the phase transition of naive spectral methods. It is well understood that the community structure can be estimated by the principal eigenvector of the centered adjacency matrix  $\mathbf{G} - \mathbb{E}\{\mathbf{G}\} = (\mathbf{G} - \bar{p}_n \mathbf{1}\mathbf{1}^\top)$ . (We denote by  $\mathbf{G}$  the graph as well as its adjacency matrix.) This approach is successful for  $\lambda > 1$  but requires average degree  $n\bar{p}_n \geq (\log n)^c$  for  $c$  a constant [7], [8].

**Remark I.4.** The reader will notice that the limiting expressions for the mutual information and MMSE are related by the identity:

$$\frac{d\Psi(\gamma_*(\lambda), \lambda)}{d\lambda} = \frac{1}{4} \left( 1 - \frac{\gamma_*(\lambda)^2}{\lambda^2} \right). \quad (15)$$

Such an identity relating mutual information and MMSE is well-known for estimation in Gaussian noise [9], and its appearance in the setting of the SBM is *a priori* surprising. However, it is hardly coincidental, as our proof strategy involves reducing the SBM to an analogous ‘‘Gaussian noise’’ setting (see Section III).

The rest of the paper is organized as follows. We discuss related work on the stochastic block model in Section II. In Section III, we discuss some consequences of our main results for other estimation metrics. Finally, in Section IV, we outline our proof strategy. We refer the reader to the full version [10] for proofs of the results.

## II. RELATED WORK

The stochastic block model was first introduced within the social science literature in [1]. Around the same time, it was studied within theoretical computer science [11], [2], [12], under the name of ‘planted partition model’. A large part of the literature has focused on the problem of *exact recovery*, providing algorithms and conditions on the gap between  $p_n$  and  $q_n$  that guarantee exact recovery of the vertex labels (up to a sign flip), with phase transitions recently discovered [13], [14], [15]. On the other hand, recent works have also studied the *detection* problem, i.e., recovering a positively correlated community structure. This was highlighted by [16], who conjectured the following intriguing phase transition phenomenon: the detection problem is solvable if and only if  $(a - b)^2 > 2(a + b)$  where  $p_n = a/n$  and  $q_n = b/n$ . The two parts of the conjecture were settled in [17] and [18], [19] respectively. Results for detection with more than two communities were also obtained in [20], [21], [15], [22].

In a sense, the present paper bridges detection and exact recovery, by characterizing the minimum estimation error when this is non-zero, but –for  $\lambda > 1$ – smaller than for random guessing. Note that [23] introduced an information-theoretic view of the SBM, showing that  $I(\mathbf{X}; \mathbf{G})/n$  admits a limit as  $n \rightarrow \infty$ , when  $p_n = a/n \leq q_n = b/n$ . The key argument establishes a certain sub-additivity property for the conditional entropy  $H(\mathbf{X}; \mathbf{G})$ . While these techniques might extend to a broad family of planted models, the sub-additivity property does not establish the limiting value itself.

In terms of proof techniques, our arguments are closest to [24], [25]. We use the well-known Lindeberg strategy to reduce computation of mutual information in the SBM to mutual information of the Gaussian observation model. We then compute the latter mutual information by developing sharp algorithmic upper bounds, which are then shown to be asymptotically tight via an area theorem. The Lindeberg strategy builds from [24], [26] while the area theorem argument also appeared in [27].

Lesieur, Krzakala and Zdeborová [28] studied estimation of low-rank matrices observed through noisy memoryless channels. They conjectured that the resulting minimal estimation error is universal across a variety of channel models. Our proof establishes universality across two such models: the Gaussian and the binary output channels. We expect that similar techniques can be useful to prove universality for other models as well.

Finally, we expect that the result obtained in this paper are likely to extend to more general models, such as censored or labelled models [23], [29], [30], [31], [21].

## III. CONSEQUENCES FOR ESTIMATION

Theorem I.2 establishes that a phase transition takes place at  $\lambda = 1$  for the matrix minimum mean square error  $\text{MMSE}_n(\lambda_n)$  defined in Eq. (11). Throughout this section, we will omit the subscript  $n$  to denote the  $n \rightarrow \infty$  limit (for instance, we write  $\text{MMSE}(\lambda) \equiv \lim_{n \rightarrow \infty} \text{MMSE}_n(\lambda_n)$ ).

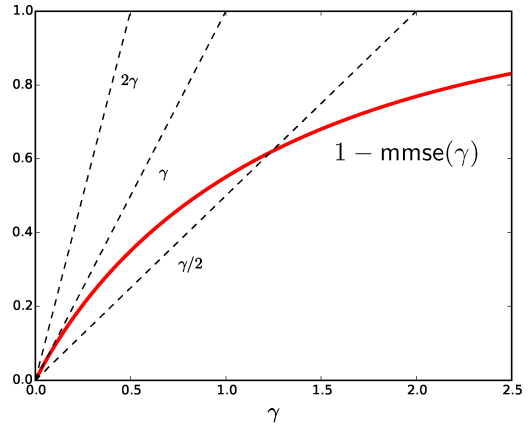


Fig. 1: Illustration of the fixed point equation Eq. (8). The ‘effective signal-to-noise ratio’  $\gamma_*(\lambda)$  is given by the intersection of the curve  $\gamma \mapsto G(\gamma) = 1 - \text{mmse}(\gamma)$ , and the line  $\gamma/\lambda$ .

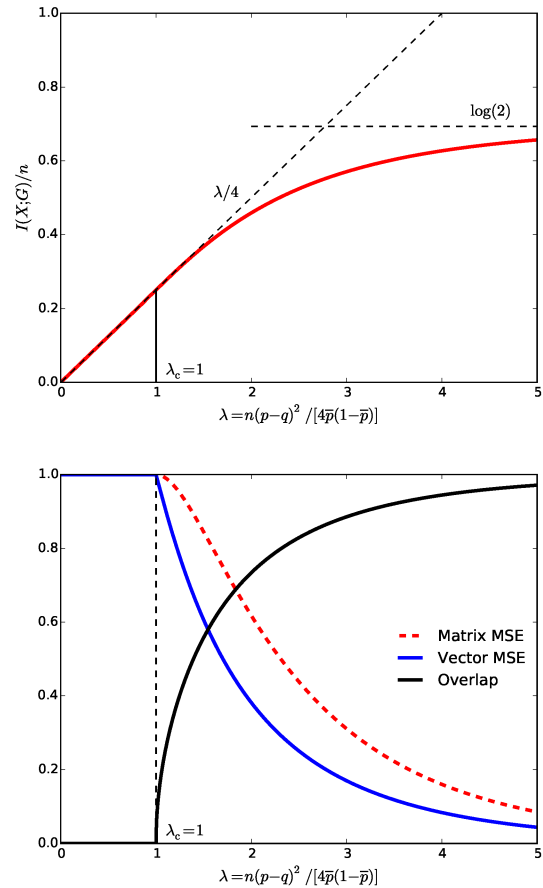


Fig. 2: Left frame: Asymptotic mutual information per vertex of the two-groups stochastic block model, as a function of the signal-to-noise ratio  $\lambda$ . The dashed lines are simple upper bounds (see [10]):  $\lim_{n \rightarrow \infty} I(\mathbf{X}; \mathbf{G})/n \leq \lambda/4$  and  $I(\mathbf{X}; \mathbf{G})/n \leq \log 2$ . Right frame: Asymptotic estimation error under different metrics. Note the phase transition at  $\lambda = 1$  in both frames.

Figure 2 reports the asymptotic prediction for  $\text{MMSE}(\lambda)$  stated in Theorem I.2, and evaluated as discussed above. The error decreases rapidly to 0 for  $\lambda > 1$ .

In this section we discuss two other estimation metrics. In both cases we define these metrics by optimizing a suitable risk over a class of estimators: it is understood that randomized estimators are admitted as well.

- The first metric is the *vector minimum mean square error*:

$$\text{vmmse}_n(\lambda_n) = \frac{1}{n} \inf_{\hat{\mathbf{x}}: \mathcal{G}_n \rightarrow \mathbb{R}^n} \mathbb{E} \left\{ \min_{s \in \{\pm 1\}} \|\mathbf{X} - s \hat{\mathbf{x}}(\mathbf{G})\|_2^2 \right\}. \quad (16)$$

Note the minimization over the sign  $s$ : this is necessary because the vertex labels can be estimated only up to an overall flip. Of course  $\text{vmmse}_n(\lambda_n) \in [0, 1]$ , since it is always possible to achieve vector mean square error equal to one by returning  $\hat{\mathbf{x}}(\mathbf{G}) = 0$ .

- The second metric is the *overlap*:

$$\text{Overlap}_n(\lambda_n) = \frac{1}{n} \sup_{\hat{\mathbf{s}}: \mathcal{G}_n \rightarrow \{\pm 1\}^n} \mathbb{E} \{ |\langle \mathbf{X}, \hat{\mathbf{s}}(\mathbf{G}) \rangle| \}. \quad (17)$$

Again  $\text{Overlap}_n(\lambda_n) \in [0, 1]$  (but now large overlap corresponds to good estimation). Indeed by returning  $\hat{x}_i(\mathbf{G}) \in \{+1, -1\}$  uniformly at random, we obtain  $\mathbb{E} \{ |\langle \mathbf{X}, \hat{\mathbf{s}}(\mathbf{G}) \rangle| \} / n = O(n^{-1/2}) \rightarrow 0$ .

Note that the main difference between overlap and vector minimum mean square error is that in the latter case we consider estimators  $\hat{\mathbf{x}}: \mathcal{G}_n \rightarrow \mathbb{R}^n$  taking arbitrary real values, while in the former we assume estimators  $\hat{\mathbf{s}}: \mathcal{G}_n \rightarrow \{+1, -1\}^n$  taking binary values.

The next lemma clarifies the relationship between matrix, vector minimum mean square error and the overlap.

**Lemma III.1.** *With the above definitions, we have*

$$\text{vmmse}_n(\lambda) \geq 1 - \sqrt{1 - (1 - n^{-1})\text{MMSE}_n(\lambda)}, \quad (18)$$

$$\text{vmmse}_n(\lambda) \leq \text{MMSE}_n(\lambda), \quad (19)$$

$$\text{Overlap}_n(\lambda) \leq \sqrt{1 - \text{MMSE}_n(\lambda)}, \quad (20)$$

$$\text{Overlap}_n(\lambda) \geq 1 - \text{MMSE}_n(\lambda). \quad (21)$$

As an immediate corollary of these lemmas (together with Theorem I.2), we obtain that  $\lambda = 1$  is the critical point for other estimation metrics as well.

**Corollary III.2.** *The vector minimum mean square error and the overlap exhibit a phase transition at  $\lambda = 1$ . Namely, under the assumptions of Theorem I.1 (in particular,  $\lambda_n \rightarrow \lambda$  and  $n\bar{p}_n(1 - \bar{p}_n) \rightarrow \infty$ ), we have*

- If  $\lambda \leq 1$ , then estimation cannot be performed asymptotically better than without any information:

$$\lim_{n \rightarrow \infty} \text{vmmse}_n(\lambda_n) = 1, \quad (22)$$

$$\lim_{n \rightarrow \infty} \text{Overlap}_n(\lambda_n) = 0. \quad (23)$$

- If  $\lambda > 1$ , then estimation can be performed better than without any information, even in the limit  $n \rightarrow \infty$ :

$$\liminf_{n \rightarrow \infty} \text{vmmse}_n(\lambda_n) \geq 1 - \frac{\gamma_*(\lambda)}{\lambda} > 0, \quad (24)$$

$$\limsup_{n \rightarrow \infty} \text{vmmse}_n(\lambda_n) \leq 1 - \frac{\gamma_*(\lambda)^2}{\lambda^2} < 1, \quad (25)$$

$$\liminf_{n \rightarrow \infty} \text{Overlap}_n(\lambda_n) \geq \frac{\gamma_*(\lambda)^2}{\lambda^2}. \quad (26)$$

#### IV. PROOF STRATEGY:

In this section we describe the main elements used in the proof of Theorems I.1 and I.2. As an intermediate step, we introduce the Gaussian observation model:

$$\mathbf{Y} = \sqrt{\frac{\lambda}{n}} \mathbf{X} \mathbf{X}^\top + \mathbf{Z}, \quad (27)$$

where  $\mathbf{Z}$  is a symmetric matrix with independent Gaussian entries  $Z_{ij} \sim \mathcal{N}(0, 1 + \delta_{ij})$  ( $\delta_{ij}$  is the Kronecker delta). Note that this model matches the first two moments of the SBM. More precisely, if  $G_{ij}^{\text{res}} \equiv (G_{ij} - \bar{p}_n) / \sqrt{\bar{p}_n(1 - \bar{p}_n)}$ , then  $\mathbb{E}\{G_{ij}^{\text{res}} | \mathbf{X}\} = \mathbb{E}\{Y_{ij} | \mathbf{X}\}$  and  $\text{Var}(G_{ij}^{\text{res}} | \mathbf{X}) = \text{Var}(Y_{ij} | \mathbf{X}) + O(n^{-1/2})$ .

Our strategy consists of two main steps:

- 1) Show that the mutual information in the SBM and the Gaussian observation model (i.e.  $I(\mathbf{X}; \mathbf{G})$  and  $I(\mathbf{X}; \mathbf{Y})$ ) coincide to leading order.
- 2) Prove an asymptotic characterization of the mutual information  $I(\mathbf{X}; \mathbf{Y})$  via an appropriately defined *approximate message passing* (AMP) algorithm.

The first step is established in the following proposition that also gives explicit error bounds on the difference between the two mutual information terms.

**Proposition IV.1.** *Assume that, as  $n \rightarrow \infty$ , (i)  $\lambda_n \rightarrow \lambda$  and (ii)  $n\bar{p}_n(1 - \bar{p}_n) \rightarrow \infty$ . Then there is a constant  $C$  independent of  $n$  such that*

$$\frac{1}{n} |I(\mathbf{X}; \mathbf{G}) - I(\mathbf{X}; \mathbf{Y})| \leq \frac{C\lambda^{3/2}}{\sqrt{n\bar{p}_n(1 - \bar{p}_n)}} + C|\lambda_n - \lambda|. \quad (28)$$

The proof of this proposition relies on the celebrated Lindeberg method, whereby we replace the non-Gaussian random variables  $G_{ij}$  with Gaussian counterparts  $Y_{ij}$  and control the error incurred. Similar arguments have been used to establish universality of various macroscopic properties across microscopic details of the underlying probabilistic models (see [24], [26] for applications to compressed sensing, information theory and spin glass theory).

The second step of characterizing the Gaussian model of Eq.(27) involves developing algorithmic upper bounds via an appropriately defined Approximate Message Passing algorithm. This algorithm yields estimates of  $\mathbf{X} \mathbf{X}^\top$ , and a corresponding mean squared error  $\text{MSE}_{\text{AMP}}(\lambda)$  incurred by these estimates. This mean squared error is shown to be sharp, via an area theorem argument and the I-MMSE identity. Such

an area theorem argument has been used in coding theory [32], multiuser detection [27] and sparse PCA [25]. Finally we state the characterization of the Gaussian model, since it is of independent interest.

**Theorem IV.2.** *For any  $\lambda > 0$ , let  $\gamma_*(\lambda)$  be the largest non-negative solution of the equation:*

$$\gamma = \lambda (1 - \text{mmse}(\gamma)). \quad (29)$$

Further, define  $\Psi(\gamma, \lambda)$  by:

$$\Psi(\gamma, \lambda) = \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma). \quad (30)$$

Then, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{X}; \mathbf{Y}) = \Psi(\gamma_*(\lambda), \lambda). \quad (31)$$

#### ACKNOWLEDGMENTS

Y.D. and A.M. were partially supported by NSF grants CCF-1319979 and DMS-1106627 and the AFOSR grant FA9550-13-1-0036. E.A. was partially supported by NSF CAREER Award CCF-1552131. Part of this work was done while the authors were visiting Simons Institute for the Theory of Computing, UC Berkeley.

#### REFERENCES

- [1] P. W. Holland, K. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983. [Online]. Available: [http://d.wanfangdata.com.cn/NSTLQK\\_10.1016-0378-8733\(83\)90021-7.aspx](http://d.wanfangdata.com.cn/NSTLQK_10.1016-0378-8733(83)90021-7.aspx)
- [2] M. Dyer and A. Frieze, "The solution of some random NP-hard problems in polynomial expected time," *Journal of Algorithms*, vol. 10, no. 4, pp. 451 – 489, 1989. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0196677489900011>
- [3] T. A. B. Snijders and K. Nowicki, "Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure," *Journal of Classification*, vol. 14, no. 1, pp. 75–100, Jan. 1997. [Online]. Available: <http://dx.doi.org/10.1007/s003579900004>
- [4] A. Condon and R. M. Karp, "Algorithms for graph partitioning on the planted partition model," *Lecture Notes in Computer Science*, vol. 1671, pp. 221–232, 1999.
- [5] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," *J. Mach. Learn. Res.*, vol. 9, pp. 1981–2014, Jun. 2008. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1390681.1442798>
- [6] F. McSherry, "Spectral partitioning of random graphs," in *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, 2001, pp. 529–537.
- [7] M. Capitaine, C. Donati-Martin, and D. Féral, "The largest eigenvalues of finite rank deformation of large wigner matrices: convergence and nonuniversality of the fluctuations," *The Annals of Probability*, vol. 37, no. 1, pp. 1–47, 2009.
- [8] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011.
- [9] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 4, pp. 1261–1282, 2005.
- [10] Y. Deshpande, E. Abbe, and A. Montanari, "Asymptotic mutual information for the two-groups stochastic block model," *arXiv preprint arXiv:1507.08685*, 2015.
- [11] T. Bui, S. Chaudhuri, F. Leighton, and M. Sipser, "Graph bisection algorithms with good average case behavior," *Combinatorica*, vol. 7, no. 2, pp. 171–191, 1987. [Online]. Available: <http://dx.doi.org/10.1007/BF02579448>
- [12] R. Boppana, "Eigenvalues and graph bisection: An average-case analysis," In *28th Annual Symposium on Foundations of Computer Science*, pp. 280–285, 1987.
- [13] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *Available at ArXiv:1405.3267*, May 2014.
- [14] E. Mossel, J. Neeman, and A. Sly, "Consistency thresholds for binary symmetric block models," *Arxiv:arXiv:1407.1591*. To appear in *STOC15*, July 2014.
- [15] E. Abbe and C. Sandon, "Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms," *arXiv:1503.00609*, March 2015.
- [16] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, p. 066106, December 2011.
- [17] E. Mossel, J. Neeman, and A. Sly, "Stochastic block models and reconstruction," *Available online at arXiv:1202.1499 [math.PR]*, 2012.
- [18] —, "A proof of the block model threshold conjecture," *Available online at arXiv:1311.4115 [math.PR]*, January 2014.
- [19] L. Massoulié, "Community detection thresholds and the weak Ramanujan property," in *STOC 2014: 46th Annual Symposium on the Theory of Computing*, New York, United States, Jun. 2014, pp. 1–10. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00969235>
- [20] O. Guédon and R. Vershynin, "Community detection in sparse networks via Grothendieck's inequality," *ArXiv:1411.4686*, Nov. 2014.
- [21] P. Chin, A. Rao, and V. Vu, "Stochastic block model and community detection in the sparse graphs: A spectral algorithm with optimal rate of recovery," *arXiv:1501.05021*, January 2015.
- [22] C. Bordenave, M. Lelarge, and L. Massoulié, "Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs," *Available at arXiv:1501.06087*, 2015.
- [23] E. Abbe and A. Montanari, "Conditional random fields, planted constraint satisfaction and entropy concentration," *Theory of Computing*, vol. 11, 2015.
- [24] S. B. Korada and A. Montanari, "Applications of the lindeberg principle in communications and statistical learning," *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 2440–2450, 2011.
- [25] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse pca," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 2197–2201.
- [26] S. Chatterjee, "A generalization of the lindeberg principle," *The Annals of Probability*, vol. 34, no. 6, pp. 2061–2076, 2006.
- [27] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of cdma (or: How to prove tanaka's formula)," in *Information Theory Workshop, 2006. ITW'06 Punta del Este. IEEE. IEEE*, 2006, pp. 160–164.
- [28] T. Lesieur, F. Krzakala, and L. Zdeborová, "Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel," *arXiv preprint arXiv:1507.03857*, 2015.
- [29] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, "Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, 2014.
- [30] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," *arXiv:1209.2910*, 2012.
- [31] J. Xu, M. Lelarge, and L. Massoulié, "Edge label inference in generalized stochastic block models: from spectral theory to impossibility results," *Proceedings of COLT 2014*, 2014.
- [32] C. Méasson, A. Montanari, T. J. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4793–4821, 2009.