The Annals of Statistics
2019, Vol. 47, No. 4, 1805–1834
https://doi.org/10.1214/17-AOS1620

EXACT RECOVERY IN THE ISING BLOCKMODEL

BY QUENTIN BERTHET[1], PHILIPPE RIGOLLET[2] AND PIYUSH SRIVASTAVA[3]

University of Cambridge, Massachusetts Institute of Technology and Tata Institute of Fundamental Research

We consider the problem associated to recovering the block structure of an Ising model given independent observations on the binary hypercube. This new model, called the Ising blockmodel, is a perturbation of the mean field approximation of the Ising model known as the Curie–Weiss model: the sites are partitioned into two blocks of equal size and the interaction between those of the same block is stronger than across blocks, to account for more order within each block. We study probabilistic, statistical and computational aspects of this model in the high-dimensional case when the number of sites may be much larger than the sample size.

1. Introduction. The past decades have witnessed an explosion of the amount of data collected routinely. Along with this expansion comes the promise of a better understanding of the nature of *interactions* between basic entities. Understanding this web of interactions has profound implications in social sciences, structural biology, neuroscience, marketing and finance for example. Graphical models (a.k.a. Markov Random Fields) have proved to be a very useful tool to turn raw data into networks that capture such interactions between random variables. Specifically, given observations of a random vector $\sigma = (\sigma_1, \dots, \sigma_p)^\top \in \mathbb{R}^p$, the goal is to output a graph on p nodes, one for each variable, where edges encode conditional independence between said variables [Lauritzen (1996)]. Graphical models have been successfully employed in a variety of applications such as image analysis [Besag (1986)], natural language processing [Manning and Schütze (1999)] and genetics [Lauritzen and Sheehan (2003); Sebastiani et al. (2005)] for example. While for many of such applications, the goal is to perform subsequent clustering of the variables, graphical models with a community structure have seldom been studied. Instead, stochastic blockmodels that were introduced in Holland, Laskey

Received January 2017; revised July 2017.

[1]Supported by an Isaac Newton Trust Early Career Support Scheme and by The Alan Turing Institute under the EPSRC Grant EP/N510129/1.

[2]Supported by Grants NSF CAREER DMS-1541099, NSF DMS-1541100, NSF DMS-1712596, DARPA W911NF-16-1-0551, ONR N00014-17-1-2147 and a grant from the MIT NEC Corporation.

[3]Part of this work was performed while PS was at the Center for the Mathematics of Information at the California Institute of Technology supported in part by NSF Grant CCF-1319745.

MSC2010 subject classifications. Primary 62H30; secondary 82B20.

Key words and phrases. Ising blockmodel, Curie–Weiss, stochastic blockmodel, planted partition, spectral partitioning.

and Leinhardt (1983) [see Abbe (2017) for an overview of recent developments] have played a central role in community detection. Despite their apparent simplicity, stochastic blockmodels have proved to be quite rich and interesting threshold phenomena have been revealed in recent years [Abbe and Sandon (2015); Banks et al. (2016); Decelle et al. (2011); Massoulié (2014); Mossel, Neeman and Sly (2013, 2015)]. However, such models are fundamentally different from graphical models for two reasons. First, in stochastic blockmodels, one observes a single realization of a graph of interactions between basic entities and the goal is to recover the community structure. Second, it is assumed that interactions between entities are independent.

In several problems concerned with inferring community structure from a graph, a common motivation is that this graph can be constructed by connecting individuals with a similar behavior. A natural question is therefore whether it is possible to use this observed behavior as the data itself. In this work, we raise the question of recovering a community structure from independent copies of σ. Such a question was recently investigated in the context of G-latent models [Bunea, Giraud and Luo (2015); Bunea et al. (2016)] where σ is assumed to have a covariance matrix with a block structure similar to the one employed in stochastic blockmodels, chiefly for a Gaussian σ. It is worth mentioning G-latent models are not graphical models as the community structure is not imposed on conditional independence relations but rather on the covariance directly and the question of learning Gaussian graphical models with a community structure is still open.

We focus on binary random variables $\sigma_1, \dots, \sigma_p \in \{-1, 1\}$, hereafter called *spins*. Graphical models on such variables are known as Ising models and were introduced in the context of statistical physics as a mathematical model for ferromagnetism [Ising (1925)]. Together with Gaussian graphical models, they form the preponderant class of graphical models in the literature, perhaps owing to their shared maximum entropy property. The Ising model has applications reaching far beyond the limits of statistical physics. For example, it has been proposed to model social interactions such as political affinities, where σ_j may represent the vote of U.S. senator j on a random bill in Banerjee, El Ghaoui and d'Aspremont (2008) [see also the data used in Diaconis, Goel and Holmes (2008) for the U.S. House of Representatives]. It has also been used as a model for neural activity in Schneidman et al. (2006) and it is known to be the stationary distribution of the so-called Glauber dynamics, which has been used to model the spread of information in social networks [Montanari and Saberi (2010)]. In the context of such applications, much effort has been devoted to estimating the underlying structure of the graphical model [Bresler (2015); Bresler, Gamarnik and Shah (2014); Bresler, Mossel and Sly (2008); Ravikumar, Wainwright and Lafferty (2010)] primarily under *sparsity* assumptions.

We introduce the *Ising blockmodel*, which forms a canonical graphical model for binary random variables with a simple community structure: the choice $\sigma_i = 1$ or -1 of each individual is influenced by the choice of others, and more markedly by members of its own community with whom it is more likely to agree. It gives us a framework to translate the common notions of *homophily*—the tendency to connect or be aligned with others—and of *assortativity*—a stronger tendency for members of the same group—from network analysis in a statistical setting. This model lends itself to a variety of possible applications, for example, identifying groups of elected representatives based on their voting record, or of customers based on their reviews of various products. Unlike G-latent models, the community structure is not imposed on the covariance matrix but rather directly on interactions. Translating these structural assumptions to the covariance matrix requires a careful analysis of the ground states of this model that we carry out in Section 4.

From a computational point of view, the Ising blockmodel bears similarity to stochastic blockmodels where the community structure may be recovered using semidefinite programing like in many problems with the same flavor [Abbe, Bandeira and Hall (2016); Bunea et al. (2016); Goemans and Williamson (1995); Hajek, Wu and Xu (2016)]. Our contribution falls into this line of work and we describe a semidefinite program (SDP) that enjoys optimal rates of structure recovery.

Our contribution. In Section 2, we define the Ising blockmodel with distribution $\mathbb{P}_{\alpha,\beta}$ as a structured perturbation of the Curie–Weiss model for which $\alpha = \beta$, that is, the interaction parameter—in this context, the inverse temperature β—is constant between all spins. In the Ising blockmodel, spins are instead divided into two balanced communities, and the interaction parameter β within each community is greater than the interaction parameter α across different communities. Our objective is to recover the two communities based on the observation of n independent draws from this distribution.

We obtain bounds on the sample size needed for an estimator based on an SDP relaxation of the maximum likelihood estimator to recover the two communities with high probability. We prove that these bounds are optimal in a minimax sense, up to constants. Our proof is based on geometric arguments, an approach that we find more transparent than more traditional arguments based on dual certificates—see Section 3.

Our bounds depend on an unknown quantity, called *gap* that may scale with the size p of the system but is not explicitly given in the model description. To get a handle on this quantity, in Section 4 we perform an analysis of the ground states for the mean magnetization in each block, a variation of the asymptotic analysis of the Curie–Weiss model. In particular, we show that these mean magnetizations jointly converge to a mixture of bivariate Gaussian distributions with explicit centers and covariance matrices for all temperature parameters (α, β), except critical ones. This asymptotic characterization allows us to quantify the optimal scaling of the sample size—as a function of the system size p—required to recover the community structure exactly. As a byproduct, our results exhibit sharp phase transitions that have a significant impact on the optimal rates for exact recovery. Informally, we find that there are two regions for (α, β): one in which the optimal sample size is of order $\log p$, and another where it is of order $p \log p$, with a sharp phase transition between those. The formal statements of these results appear in Theorems 3.3 and 3.7 and Corollary 4.6.

Finally, note that the size p of the system has to be large enough to observe interesting phenomena. Such high dimensional systems are especially pertinent to the applications described above and our results are valid for large enough p, potentially much larger than the number of observations. In particular, we often consider asymptotic statements as $p \to \infty$. However, in the statistical applications of Section 3 we are interested in understanding the scaling of the number of observations as a function of p. To that end, we keep track of the first-order terms in p and only let higher order terms vanish when convenient.

2. The Ising block model. In this work, we introduce the Ising blockmodel in order to combine the notions of the stochastic blockmodel and that of graphical model by assuming that we observe independent copies of a vector of spins $\sigma = (\sigma_1, \dots, \sigma_p) \in \{-1, 1\}^p$ distributed according to an Ising model with a block structure analogous to the one arising in the stochastic blockmodel.

2.1. *Definition.* Let $p \geq 2$ be an even integer and let $S \subset [p] := \{1, \dots, p\}$ be a subset of size $|S| = m = p/2$. For any partition $(S, \bar{S})$, where $\bar{S} = [p] \setminus S$ denotes the complement of S, write $i \sim j$ if $(i, j) \in S^2 \cup \bar{S}^2$ and $i \nsim j$ if $(i, j) \in [p]^2 \setminus (S^2 \cup \bar{S}^2)$. Fix $\beta, \alpha \in \mathbb{R}$ and let $\sigma \in \{-1, 1\}^p$ have density $f_{S,\alpha,\beta}$ with respect to the counting measure on $\{-1, 1\}^p$ given by

$$f_{S,\alpha,\beta}(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right], \tag{2.1}$$

where

$$Z_{S,\alpha,\beta} := \sum_{\sigma \in \{-1,1\}^p} \exp\left[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right] \tag{2.2}$$

is a normalizing constant traditionally called *partition function*. Let $\mathbb{P}_{S,\alpha,\beta}$ denote the probability distribution over $\{-1, 1\}^p$ that has density $f_{S,\alpha,\beta}$ with respect to the counting measure on $\{-1, 1\}^p$. We call this model the *Ising Blockmodel* (IBM). In the sequel, we write simply $f_{\alpha,\beta}$ and $\mathbb{P}_{\alpha,\beta}$ to emphasize the dependency on α, β and simply $\mathbb{P}_S$ to emphasize the dependency on S. We study in this work the statistical problem of identifying the partition $(S, \bar{S})$ based on n independent draws from $\mathbb{P}_{\alpha,\beta}$, when $\alpha < \beta$.

2.2. *Link with the Curie–Weiss model.* When $\alpha = \beta > 0$, the model (2.1) is the mean field approximation of the (ferromagnetic) Ising model and is called the *Curie–Weiss* model (without external field). It can be readily seen from (2.1) that vectors $\sigma \in \{-1, 1\}^p$ that present a lot of pairs (i, j) with opposite spins, that is, $\sigma_i \sigma_j < 0$, receive less probability than vectors where most of the spins agree. However, there are many more vectors σ of the former kind in the discrete hypercube $\{-1, 1\}^p$. This tension between *energy* and *entropy* is responsible for phase transitions in such systems.

When positive, the parameter $\beta > 0$ is called *inverse temperature* and it controls the strength of interactions and, therefore, the weight given to the energy term. When $\beta \to 0$, the entropy term dominates and $\mathbb{P}_{\beta,\beta}$ tends to the uniform density over $\{-1, 1\}^p$. When $\beta \to \infty$, $\mathbb{P}_{\beta,\beta} \to 0.5\delta_{\mathbf{1}} + 0.5\delta_{-\mathbf{1}}$, where δ_x denotes the Dirac point mass at x and $\mathbf{1} = (1, \dots, 1) \in \{-1, 1\}^p$ denotes the all-ones vector of dimension p. In this case, energy dominates entropy.

The Curie–Weiss model is tractable enough that the above considerations can be made precise. Specifically, the be behavior of the system as the temperature $1/\beta$ varies can be described accurately. To that end, let $\mu^{\mathrm{CW}} = \sigma^\top \mathbf{1}/p$ denote the *magnetization* of σ. When $\mu^{\mathrm{CW}} \simeq 0$, then σ has a balanced number of positive and negative spins (paramagnetic behavior) and when $|\mu^{\mathrm{CW}}| \gg 0$, then σ has a large proportion of spins with a given sign (ferromagnetic behavior). When p is large enough, the Curie–Weiss model is known to obey a phase-transition from ferromagnetic to paramagnetic behavior when the temperature crosses a threshold (see Appendix A for details). This striking result indicates that when the temperature decreases (β increases), the model changes from that of a disordered system (no preferred inclination towards -1 or $+1$) to that of an ordered system (a majority of the spins agree to the same sign). This behavior is interesting in the context of modeling social interactions and indicates that if the strength of interactions is large enough ($\beta > 1$) then a partial consensus may be found. Formally, the Curie–Weiss model may also be defined in the anti-ferromagnetic case $\beta < 0$—we abusively call it "inverse temperature" in this case also—to model the fact that negative interactions are encouraged. For such choices of β, the distribution is concentrated around balanced configurations σ that have magnetization close to 0. Moreover, as $\beta \to -\infty$, $\mathbb{P}_{\beta,\beta}$ converges to the uniform distribution on configurations with zero magnetization (assuming that p is even so that such configurations exist for simplicity). As a result, the anti-ferromagnic case arises when no consensus may be found and the spins are evenly split between positive and negative.

In reality though, a collective behavior may be fragmented into communities and the IBM is meant to reflect this structure. Specifically, since $\beta > \alpha$, the strength β of interactions within the blocks S and $\bar{S}$ is larger than that across blocks S and $\bar{S}$. As will become clear from our analysis, the case where $\alpha < 0$ presents interesting configurations whereby the two blocs S and $\bar{S}$ have polarized behaviors, that is, opposite magnetization in each block.

2.3. *Covariance.* The covariance matrix $\Sigma = \mathbb{E}_{\alpha,\beta}[\sigma\sigma^\top]$ captures the block structure of IBM, and thus plays a major role in the statistical applications of Section 3. Moreover, the coefficients of Σ can be expressed explicitly in terms of two statistics of σ. For any $A \subset [p]$, define $\mathbf{1}_A \in \{0,1\}^p$ to be the indicator vector of A and let $\mu_A = \sigma^\top \mathbf{1}_A/|A|$ denote the *local magnetization*. Akin to the Curie–Weiss model, the density $f_{\alpha,\beta}$ puts uniform weights on configurations that have the same local magnetization μ_S and $\mu_{\bar{S}}$. This symmetry can be exploited to obtain the following estimates.

LEMMA 2.1. *Let $\Sigma = \mathbb{E}_{\alpha,\beta}[\sigma\sigma^\top]$ denote the covariance matrix of a random configuration $\sigma \sim \mathbb{P}_{\alpha,\beta}$. For any $i \neq j \in [p]$, it holds*

$$\Delta := \Sigma_{ij} = \frac{m}{2(m-1)}\mathbb{E}[\mu_S^2 + \mu_{\bar{S}}^2] - \frac{1}{m-1} \qquad \text{if } i \sim j,$$
$$\Omega := \Sigma_{ij} = \mathbb{E}[\mu_S \mu_{\bar{S}}] \qquad \text{if } i \nsim j.$$

PROOF. In this proof, we rely on symmetry of the problem: all the spins σ_i in a given block, S or $\bar{S}$ have the same marginal distribution. Fix $i \neq j$.

If $i \sim j$, for example, if $i, j \in S$, we have by linearity of expectation:

$$\begin{aligned}\Sigma_{ij} = \mathbb{E}[\sigma_i\sigma_j] &= \frac{1}{m(m-1)}\Bigg(\mathbb{E}\sum_{(i,j)\in S^2}\sigma_i\sigma_j - m\Bigg)\\ &= \frac{m}{m-1}\mathbb{E}[\mu_S^2] - \frac{1}{m-1}.\end{aligned}$$

Since μ_S and $\mu_{\bar{S}}$ are identically distributed, we obtain the desired result.

For any $i \nsim j$, we have

$$\Sigma_{ij} = \mathbb{E}[\sigma_i\sigma_j] = \frac{1}{m^2}\mathbb{E}\sum_{(i,j)\in S\times\bar{S}}\sigma_i\sigma_j = \mathbb{E}[\mu_S\mu_{\bar{S}}].$$

□

Unlike many models in the statistical literature, computing Σ exactly is difficult in the IBM. In particular, it is not immediately clear from Lemma 2.1 that $\Delta > \Omega$, while this should be intuitively true since $\beta > \alpha$ and, therefore, the spin interactions are stronger within blocks than across blocks. It turns out that this simple fact can be checked by other means. For example, Lemma 3.6 implies that $\Delta - \Omega$ and $\beta - \alpha$ have the same sign. We derive in Section 4 asymptotic approximations as $m \to \infty$ to prove effective upper and lower bound on the gap $\Delta - \Omega$, by analyzing precisely the behavior of μ_S and $\mu_{\bar{S}}$, and using the result of Lemma 2.1.

3. Exact recovery. In this section, we focus on the following clustering task: given n i.i.d. observations drawn from $\mathbb{P}_{\alpha,\beta}$ with $\alpha < \beta$, recover the partition $(S, \bar{S})$. We study the properties of an efficient clustering algorithm together with the fundamental limitations associated to this task. Guarantees in terms of necessary sample size are expressed in terms of the parameters of the problem and of $\Delta - \Omega$. This last term is analyzed in Section 4.

3.1. *Maximum likelihood estimation.* Fix a sample size $n \geq 1$. Given n independent copies $\sigma^{(1)}, \dots, \sigma^{(n)}$ of $\sigma \sim \mathbb{P}_{\alpha,\beta}$, the log-likelihood is given by

$$\mathcal{L}_n(S) = \sum_{t=1}^{n} \log\big(\mathbb{P}_{\alpha,\beta}\big(\sigma^{(t)}\big)\big) = -n \log Z_{\alpha,\beta} - \sum_{t=1}^{m} \mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}\big(\sigma^{(t)}\big),$$

where $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ denotes the *IBM Hamiltonian* defined on $\{-1, 1\}^p$ by

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\left(\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right), \tag{3.1}$$

and $Z_{\alpha,\beta}$ is the partition function defined in (2.2).

While both $Z_{\alpha,\beta}$ and $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ could depend on the choice of the block S, it turns out that $Z_{\alpha,\beta}$ is constant over choices of S such that $|S| = m = p/2$.

LEMMA 3.1. *The partition function $Z_{\alpha,\beta} = Z_{\alpha,\beta}(S)$ defined in (2.2) is such that $Z_{\alpha,\beta}(S) = Z_{\alpha,\beta}([m])$ for all S of size $|S| = m$. This statement remains true even if $m \neq p/2$.*

PROOF. Fix $S \subset [p]$ such that $|S| = m$ and denote by $\pi : [p] \to [p]$ any bijection that maps $[m]$ to S. By (2.2) and (4.1), it holds

$$\begin{aligned} Z_{\alpha,\beta}(S) &= \sum_{\sigma \in \{-1,1\}^p} \exp\left[\frac{1}{4m}\big(2\alpha\big(\sigma^\top \mathbf{1}_S\big)\big(\sigma^\top \mathbf{1}_{\bar{S}}\big) - \beta\big(\big(\sigma^\top \mathbf{1}_S\big)^2 + \big(\sigma^\top \mathbf{1}_{\bar{S}}\big)^2\big)\big)\right] \\ &= \sum_{\substack{\tau = \pi(\sigma) \\ \sigma \in \{-1,1\}^p}} \exp\left[\frac{1}{4m}\big(2\alpha\big(\tau^\top \mathbf{1}_S\big)\big(\tau^\top \mathbf{1}_{\bar{S}}\big) - \beta\big(\big(\tau^\top \mathbf{1}_S\big)^2 + \big(\tau^\top \mathbf{1}_{\bar{S}}\big)^2\big)\big)\right] \end{aligned}$$

since π is a bijection. Moreover, $\tau^\top \mathbf{1}_S = \pi(\sigma)^\top \mathbf{1}_S = \sigma^\top \mathbf{1}_{[m]}$ and $\tau^\top \mathbf{1}_{\bar{S}} = \sigma^\top \mathbf{1}_{\overline{[m]}}$. Hence $Z_{\alpha,\beta}(S) = Z_{\alpha,\beta}([m])$. $\square$

To emphasize the fact that the partition function does not depend on S, we simply write $Z_{\alpha,\beta} = Z_{\alpha,\beta}(S)$. It implies that the log-likelihood is a simple function of S. Indeed, define the matrix $Q = Q_S \in \mathbb{R}^{p \times p}$ such that $Q_{ij} = \beta/p$ for $i \sim j$ and $Q_{ij} = \alpha/p$ for $i \nsim j$. Observe that (3.1) can be written as

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\frac{1}{2}\sigma^\top Q \sigma = -\frac{1}{2}\mathbf{Tr}\big(\sigma\sigma^\top Q\big).$$

This in turns implies

$$\mathcal{L}_n(S) = -n \log Z_{\alpha,\beta} + \frac{n}{2}\mathbf{Tr}[\hat{\Sigma} Q],$$

where $\hat{\Sigma} := \frac{1}{n}\sum_{t=1}^{n} \sigma^{(t)} \sigma^{(t)\top}$ denotes the empirical covariance matrix. Since $\alpha < \beta$, the likelihood maximization problem $\max_{S \subset [p], |S| = m} \mathcal{L}_n(S)$ is equivalent to

$$\max_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Sigma} V], \qquad \mathcal{P} = \big\{vv^\top : v \in \{-1, 1\}^p, v^\top \mathbf{1}_{[p]} = 0\big\}. \tag{3.2}$$

In particular, estimating the blocks $(S, \bar{S})$ amounts to estimating $v_S v_S^\top \in \mathcal{P}$, where $v_S = \mathbf{1}_S - \mathbf{1}_{\bar{S}} \in \{-1, 1\}^p$. Note that $v_S v_S^\top = v_{\bar{S}} v_{\bar{S}}^\top$. For an adjacency matrix A, the optimization problem $\max_{V \in \mathcal{P}} \mathbf{Tr}[AV]$ is a special case of the *minimum bisection* problem and it is known to be NP-hard in general [Garey, Johnson and Stockmeyer (1976)]. To overcome this limitation, various approximation algorithms were suggested over the years, culminating with a poly-logarithmic approximation algorithm [Feige and Krauthgamer (2002)]. Unfortunately, such approximations are not directly useful in the context of maximum likelihood estimation. Nevertheless, the maximum likelihood estimation problem at hand is not worst case, but rather a random problem. It can be viewed as a variant of the planted partition model (aka stochastic blockmodel) introduced in Dyer and Frieze (1989). Indeed the block structure of Σ unveiled in Lemma 2.1 can be viewed as similar to the adjacency matrix of a weighted graph with a small bisection. Moreover, $\hat{\Sigma}$ can be viewed as the matrix Σ *planted* in some noise. It is therefore not surprising that we can use the same methodology in both cases. In particular, we will use the semidefinite relaxation to the MAXCUT problem of Goemans and Williamson (1995) that was already employed in the planted partition model [Abbe, Bandeira and Hall (2016); Hajek, Wu and Xu (2016)]. Here, unlike the original planted partition problem, the noise is correlated and, therefore, requires a different analysis. In random matrix terminology, the observed matrix in the stochastic block model is of Wigner type, whereas in the IBM, it is of Wishart type.

It can actually be impractical to use directly the matrix $\hat{\Sigma}$ in the above relaxations, and we apply a pre-preprocessing that amounts to a centering procedure, which simplifies our analysis. Given $\sigma \in \{-1, 1\}^p$, define its centered version $\bar{\sigma}$ by

$$\bar{\sigma} = \sigma - \frac{\mathbf{1}_{[p]}^\top \sigma}{p} \mathbf{1}_{[p]} = P\sigma,$$

where $P = I_p - \frac{1}{p}\mathbf{1}_{[p]}\mathbf{1}_{[p]}^\top$ is the projector onto the subspace orthogonal to $\mathbf{1}_{[p]}$. Moreover, let $\Gamma = P\Sigma P$ and $\hat{\Gamma} = P\hat{\Sigma}P$, respectively, denote the covariance and empirical covariance matrices of the vector $\bar{\sigma}$.

Note that for all $V \in \mathcal{P}$, we have that $\mathbf{Tr}[\hat{\Gamma}V] = \mathbf{Tr}[\hat{\Sigma}V]$ since $V\mathbf{1}_{[p]}\mathbf{1}_{[p]}^\top = 0$, so that $PVP = V$. It implies that the likelihood function is unchanged over $\mathcal{P}$ when substituting $\hat{\Sigma}$ by $\hat{\Gamma}$. Moreover, $\mathbb{E}[\hat{\Gamma}] = \Gamma$ and the spectral decomposition of Γ is given by

$$\Gamma = (1 - \Delta)P + p\frac{\Delta - \Omega}{2} u_S u_S^\top, \tag{3.3}$$

where $u_S = v_S/\sqrt{p}$ is a unit vector. Therefore, the matrix Γ has leading eigenvalue $(1 - \Delta) + p(\Delta - \Omega)/2$ with associated unit eigenvector u_S. Moreover, its eigengap is $p(\Delta - \Omega)/2$. It is well known in matrix perturbation theory that the eigengap plays a key role in the stability of the spectral decomposition of Γ when observed with noise.

3.2. *Exact recovery via semidefinite programming.* In this subsection, we consider the following semidefinite programming (SDP) relaxation of the optimization problem (3.2):

$$\max_{V\in\mathcal{E}} \mathbf{Tr}[\hat{\Gamma}V], \qquad \mathcal{E}=\{V\in\mathcal{S}_p:\mathbf{diag}(V)=\mathbf{1}_{[p]}, V\succeq 0\}, \tag{3.4}$$

where $\mathcal{S}_p$ denotes the set of $p\times p$ symmetric real matrices and $V\succeq 0$ denotes that V is positive semidefinite. The set $\mathcal{E}$ is the set of correlation matrices, and it is known as the *elliptope*. We recall the definition of the vector $v_S=\mathbf{1}_S-\mathbf{1}_{\bar{S}}\in\{-1,1\}^p$ and note that $v_Sv_S^\top\in\mathcal{P}\subset\mathcal{E}$. Moreover, we denote by $\hat{V}^{\mathsf{SDP}}$ any solution to the the above program. Our goal is to show that (3.4) has a unique solution given by $\hat{V}^{\mathsf{SDP}}=v_Sv_S^\top$, that is, the SDP relaxation is tight. In contrast to the MLE, this estimator can be computed efficiently by interior-point methods [Boyd and Vandenberghe (2004)].

While the dual certificate approach of Abbe, Bandeira and Hall (2016) could be used in this case [see also Hajek, Wu and Xu (2016)] we employ a slightly different proof technique, more geometric, that we find to be more transparent. This approach is motivated by the idea that the relaxation is tight in the population case, suggesting that it might be the case as well when $\hat{\Gamma}$ is close to Γ.

Recall that for any $X_0\in\mathcal{E}$, the normal cone to $\mathcal{E}$ at X_0 is denoted by $\mathcal{N}_{\mathcal{E}}(X_0)$ and defined by

$$\mathcal{N}_{\mathcal{E}}(X_0)=\{C\in\mathcal{S}_p:\mathbf{Tr}(CX)\le\mathbf{Tr}(CX_0),\forall X\in\mathcal{E}\}.$$

It is the cone of matrices $C\in\mathcal{S}_p$ such that $\max_{X\in\mathcal{E}}\mathbf{Tr}(CX)=\mathbf{Tr}(CX_0)$. Therefore, $v_Sv_S^\top$ is a solution of (3.4), that is, the SDP relaxation is tight, whenever $\hat{\Gamma}\in\mathcal{N}_{\mathcal{E}}(v_Sv_S^\top)$. The normal cone can be described using the following Laplacian operator. For any matrix $C\in\mathcal{S}_p$, define

$$L_S(C):=\mathbf{diag}(Cv_Sv_S^\top)-C,$$

and observe that $L_S(C)v_S=0$. Indeed, since $v_S\in\{-1,1\}^p$, it holds

$$\mathbf{diag}(Cv_Sv_S^\top)v_S=\mathbf{diag}(Cv_S\mathbf{1}_{[p]}^\top)\mathbf{1}_{[p]}=Cv_S.$$

PROPOSITION 3.2. *For any matrix $C\in\mathcal{S}_p$, the following are equivalent*:

1. $C\in\mathcal{N}_{\mathcal{E}_p}(v_Sv_S^\top)$.
2. $L_S(C)=\mathbf{diag}(Cv_Sv_S^\top)-C\succeq 0$.

Moreover, if $L_S(C)\succeq 0$ has only one eigenvalue equal to 0, then $v_Sv_S^\top$ is the unique maximizer of $\mathbf{Tr}(CV)$ over $V\in\mathcal{E}$.

PROOF. It is known [see Laurent and Poljak (1996)] that the normal cone $\mathcal{N}_{\mathcal{E}}(v_Sv_S^\top)$ is given by

$$\mathcal{N}_{\mathcal{E}}(v_Sv_S^\top)=\{C\in\mathcal{S}_p:C=D-M, D \text{ diagonal}, M\succeq 0, v_S^\top Mv_S=0\},$$

where $M \succeq 0$ denotes that M is a symmetric, semidefinite positive matrix. We are going to make use of the following facts. First, for any diagonal matrix D and any $V \in \mathcal{E}$, we have $\mathbf{diag}(DV) = D$. Second, taking $V = v_S v_S^\top$, we have

$$L_S(C) v_S v_S^\top = \mathbf{diag}(C v_S v_S^\top) v_S v_S^\top - C v_S v_S^\top.$$

Taking the diagonal on both sides directly yields that

$$\mathbf{diag}(L_S(C) v_S v_S^\top) = 0. \tag{3.5}$$

2. $\Rightarrow$ 1. Let $C \in \mathcal{S}_p$ be such that $L_S(C) \succeq 0$. By definition, we have $C = \mathbf{diag}(C v_S v_S^\top) - L_S(C)$ and it remains to check that $v_S^\top L_S(C) v_S = 0$, which follows readily from (3.5).

1. $\Rightarrow$ 2. Let $C = D - M \in \mathcal{N}_{\mathcal{E}_p}(v_S v_S^\top)$ where D is diagonal and $M \succeq 0$, $v_S^\top M v_S = 0$, which implies that $M v_S = 0$. It yields $C v_S v_S^\top = D v_S v_S^\top$ and $\mathbf{diag}(C v_S v_S^\top) = \mathbf{diag}(D v_S v_S^\top) = D$ so that the decomposition is necessarily $D = \mathbf{diag}(C v_S v_S^\top)$ and $M = L_S(C) = \mathbf{diag}(C v_S v_S^\top) - C$. In particular, $L_S(C) \succeq 0$.

Thus, if $L_S(C) \succeq 0$ then $v_S v_S^\top$ is a maximizer of $\mathbf{Tr}(CV)$ over $V \in \mathcal{E}$. To prove uniqueness, recall that for any maximizer $V \in \mathcal{E}$, we have $\mathbf{Tr}(CV) = \mathbf{Tr}(C v_S v_S^\top)$. Plugging $C = \mathbf{diag}(C v_S v_S^\top) - L_S(C)$ and using (3.5) yields

$$\begin{aligned}\mathbf{Tr}\big(\mathbf{diag}(C v_S v_S^\top) V\big) - \mathbf{Tr}\big(L_S(C) V\big) &= \mathbf{Tr}\big(\mathbf{diag}(C v_S v_S^\top) v_S v_S^\top\big) \\ &= \mathbf{Tr}\big(\mathbf{diag}(C v_S v_S^\top)\big).\end{aligned}$$

Recall that $\mathbf{Tr}(\mathbf{diag}(C v_S v_S^\top) V) = \mathbf{Tr}(\mathbf{diag}(C v_S v_S^\top))$ so that the above display yields $\mathbf{Tr}(L_S(C) V) = 0$. Since $V \succeq 0$ and the kernel of the semidefinite positive matrix $L_S(C)$ is spanned by v_S, we have that $V = v_S v_S^\top$. $\square$

It follows from Proposition 3.2 that if $L_S(\hat{\Gamma}) \succeq 0$ and it has only one eigenvalue equal to zero, then $v_S v_S^\top$ is the unique solution to (3.4). In particular, in this case, the SDP allows exact recovery of the block structure $(S, \bar{S})$. Observe that the conditions of Proposition 3.2 hold if $\hat{\Gamma}$ is replaced by the population matrix Γ. Indeed, using (3.3), we obtain

$$\begin{aligned}L_S(\Gamma) &= \left(1 - \Delta + p\frac{\Delta - \Omega}{2}\right) I_p - (1 - \Delta) P - p\frac{\Delta - \Omega}{2} u_S u_S^\top \\ &= (1 - \Delta) \frac{\mathbf{1}_{[p]}}{\sqrt{p}} \frac{\mathbf{1}_{[p]}^\top}{\sqrt{p}} - p\frac{\Delta - \Omega}{2} u_S u_S^\top + p\frac{\Delta - \Omega}{2} I_p,\end{aligned}$$

where we used the fact that $I_p - P$ is the projector onto the linear span of $\mathbf{1}_{[p]}$. Therefore, the eigenvalues of $L_S(\Gamma)$ are 0, $1 - \Delta + p(\Delta - \Omega)/2$, both with multiplicity 1 and $p(\Delta - \Omega)/2$ with multiplicity $p - 1$. In particular, for $p \geq 2$, $L_S(\Gamma) \succeq 0$ and it has only one eigenvalue equal to zero.

Extending this result to $L_S(\hat{\Gamma})$ yields the following theorem, as illustrated in Figure 1. Let $C_{\alpha,\beta} > 0$ be a positive constant such that $\Delta - \Omega > C_{\alpha,\beta}/p$, whose existence is guaranteed by Proposition 4.4.

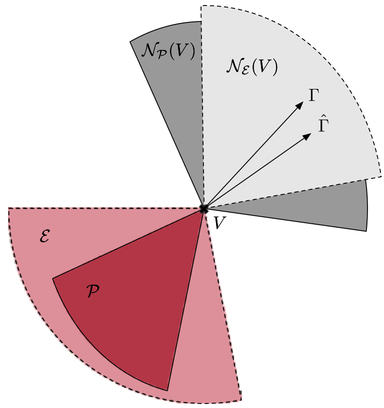

FIG. 1. *The geometric interpretation for the analysis of this convex relaxation. In the population case, the true value of the parameter $V = v_S v_S^\top$ is the unique solution of both the maximum likelihood problem on $\mathcal{P}$ and of the convex relaxation on $\mathcal{E}$, as Γ belongs to both normal cones at V. The relaxation is therefore tight with Γ as input. We show that when the sample size is large enough, the sample matrix $\hat{\Gamma}$ is close enough to Γ and also in both normal cones, making V the solution to both problems.*

THEOREM 3.3. *The SDP relaxation* (3.4) *has a unique maximum at $V = v_S v_S^\top$ with probability $1-\delta$ whenever*

$$n > 64\Big(3 + \frac{2}{C_{\alpha,\beta}}\Big)\frac{\log(4p/\delta)}{\Delta - \Omega}\big(1 + o_p(1)\big).$$

In particular, the SDP relaxation recovers exactly the block structure $(S, \bar{S})$.

PROOF. Recall that $L_S(\hat{\Gamma})v_S = 0$ and for any $C \in \mathcal{S}_p$, denote by $\lambda_2[C]$ its second smallest eigenvalue. Our goal is to show that $\lambda_2[L_S(\hat{\Gamma})] > 0$. To that end, observe that

$$L_S(\hat{\Gamma}) = L_S(\Gamma) + \mathbf{diag}\big((\hat{\Gamma} - \Gamma)v_S v_S^\top\big) + \Gamma - \hat{\Gamma}.$$

Therefore, using Weyl's inequality and the fact $\lambda_2[L_S(\Gamma)] = p(\Delta - \Omega)/2$, we get

$$\lambda_2[L_S(\hat{\Gamma})] \geq p\frac{\Delta - \Omega}{2} - \big\|\mathbf{diag}\big((\hat{\Gamma} - \Gamma)v_S v_S^\top\big)\big\|_{\mathrm{op}} - \|\hat{\Gamma} - \Gamma\|_{\mathrm{op}}, \tag{3.6}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm. Therefore, it is sufficient to upper bound the above operator norms. This is ensured by the following lemma.

LEMMA 3.4. *Fix $\delta > 0$ and define*

$$\mathcal{R}_{n,p}(\delta) = 2p\max\Big(\sqrt{\frac{(1 + 2/C_{\alpha,\beta})(\Delta - \Omega)\log(4p/\delta)}{n}}, \frac{(6 + 4/C_{\alpha,\beta})\log(4p/\delta)}{n}\Big).$$

With probability $1-\delta$, *it holds simultaneously that*

$$\|\hat{\Gamma}-\Gamma\|_{\mathrm{op}} \le \mathcal{R}_{n,p}(\delta)\big(1+o_p(1)\big) \tag{3.7}$$

and

$$\big\|\mathbf{diag}\big((\hat{\Gamma}-\Gamma)v_S v_S^\top\big)\big\|_{\mathrm{op}} \le \mathcal{R}_{n,p}(\delta)\big(1+o_p(1)\big). \tag{3.8}$$

PROOF. To prove (3.7), we use a matrix Bernstein inequality for sum of independent matrices from Tropp (2015). To that end, note that

$$\hat{\Gamma}-\Gamma = \frac{1}{n}\sum_{t=1}^{n} M_t,$$

where $M_1,\dots,M_n$ are i.i.d. random matrices given by $M_t = (\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}-\Gamma)$, $t=1,\dots,n$. We have

$$\|M_t\|_{\mathrm{op}} \le \big\|\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}\big\|_{\mathrm{op}} + \|\Gamma\|_{\mathrm{op}} \le p + \|\Gamma\|_{\mathrm{op}}.$$

Furthermore, we have that

$$\begin{aligned}\mathbb{E}[M_t^2] &= \mathbb{E}\big[\|\bar{\sigma}^{(t)}\|^2\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top} - \bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}\Gamma - \Gamma\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top} + \Gamma^2\big] \\ &= p\mathbb{E}\big[\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}\big] - \Gamma^2 - \Gamma^2 + \Gamma^2 \preceq p\Gamma.\end{aligned}$$

As a consequence, $\sum_{t=1}^n \mathbb{E}[M_t^2] \preceq p\Gamma$. By Theorem 1.6.2 in Tropp (2015), this yields

$$\mathbb{P}\big(\|\hat{\Gamma}-\Gamma\|_{\mathrm{op}} > t\big) \le 2p\exp\bigg(-\frac{nt^2}{2p\|\Gamma\|_{\mathrm{op}} + 2(p+\|\Gamma\|_{\mathrm{op}})t}\bigg). \tag{3.9}$$

We have $\|\hat{\Gamma}-\Gamma\|_{\mathrm{op}} \le t$ with probability $1-\delta$ for any t such that

$$\log(2p/\delta) \le \frac{nt^2}{2p\|\Gamma\|_{\mathrm{op}} + 2(p+\|\Gamma\|_{\mathrm{op}})t}.$$

This holds for all

$$t \ge \max\bigg(\sqrt{\frac{4p\|\Gamma\|_{\mathrm{op}}\log(2p/\delta)}{n}}, \frac{4(p+\|\Gamma\|_{\mathrm{op}})\log(2p/\delta)}{n}\bigg).$$

To conclude the proof of (3.7), observe that

$$\|\Gamma\|_{\mathrm{op}} = p\frac{\Delta-\Omega}{2} + 1 - \Delta \le \bigg(1+\frac{1}{C_{\alpha,\beta}}\bigg)(\Delta-\Omega)p,$$

where $C_{\alpha,\beta} > 0$ is defined immediately before the statement of Theorem 3.3.

We now turn to the proof of (3.8). Recall that $v_S \in \{-1,1\}^p$ so that the ith diagonal element satisfies

$$\big|\mathbf{diag}\big((\hat{\Gamma}-\Gamma)v_S v_S^\top\big)_{ii}\big| = \big|e_i^\top(\hat{\Gamma}-\Gamma)v_S\big|,$$

where e_i denotes the ith vector of the canonical basis of $\mathbb{R}^p$. Hence,

$$\|\mathbf{diag}((\hat{\Gamma}-\Gamma)v_S v_S^\top)\|_{\mathrm{op}} = \max_{i\in[p]}|\mathbf{diag}((\hat{\Gamma}-\Gamma)v_S v_S^\top)_{ii}| = \max_{i\in[p]}|e_i^\top(\hat{\Gamma}-\Gamma)v_S|.$$

We bound the right-hand side of the above inequality by noting that

$$e_i^\top(\hat{\Gamma}-\Gamma)v_S = \frac{m}{n}\sum_{t=1}^{n}\bigl(\bar{\sigma}_i^{(t)}(\mu_S^{(t)}-\mu_{\bar{S}}^{(t)}) - \mathbb{E}[\bar{\sigma}_i^{(t)}(\mu_S^{(t)}-\mu_{\bar{S}}^{(t)})]\bigr),$$

where $\mu_S^{(t)} = \mathbf{1}_S^\top\bar{\sigma}^{(t)}/m \in [-1,1]$ and $\mu_{\bar{S}}^{(t)}$ is defined analogously. The random variables $\bar{\sigma}_i^{(t)}(\mu_S^{(t)}-\mu_{\bar{S}}^{(t)}) - \mathbb{E}[\bar{\sigma}_i^{(t)}(\mu_S^{(t)}-\mu_{\bar{S}}^{(t)})]$ are centered, i.i.d., and are bounded in absolute value by 2 for all $t\in[n]$. Moreover, it follows from Lemma 2.1 that the variance of these random variables is bounded by (for $p\geq 4$)

$$\mathbb{E}\bigl[(\mu_S^{(t)}-\mu_{\bar{S}}^{(t)})^2\bigr] \leq 2(\Delta-\Omega)+\frac{4}{p-2} \leq 2(\Delta-\Omega)+\frac{8}{p} =: \nu^2.$$

By a one-dimensional Bernstein inequality, and a union bound over p terms, we have therefore that

$$\mathbb{P}\biggl(\max_{i\in[p]}|e_i^\top(\hat{\Gamma}-\Gamma)v_S| > \frac{pt}{n}\biggr) \leq 2p\exp\biggl(-\frac{2t^2}{n\nu^2+4t/3}\biggr),$$

which yields

$$\max_{i\in[p]}|e_i^\top(\hat{\Gamma}-\Gamma)v_S| \leq p\max\biggl(\sqrt{\frac{\nu^2\log(2p/\delta)}{n}}, \frac{4\log(2p/\delta)}{3n}\biggr),$$

with probability $1-\delta$. It completes the proof of (3.8). $\square$

To conclude the proof of Theorem 3.3, note that for the prescribed choice of n, we have

$$2\mathcal{R}_{n,p}(\delta)\bigl(1+o_p(1)\bigr) < p\frac{\Delta-\Omega}{2}$$

and it follows from (3.6) that $\lambda_2[L_S(\hat{\Gamma})]>0$. $\square$

REMARK 3.5. We have not attempted to optimize the constant term $64(3+2/C_{\alpha,\beta})$ that appears in Theorem 3.3 and it is arguably suboptimal. One way to see how it can be reduced at least by a factor 2 is by noting that the factor p in the right-hand side of (3.9) is in fact superfluous thus resulting in a extra logarithmic factor in (3.7). This is because, akin to the stochastic blockmodel analysis in Abbe, Bandeira and Hall (2016), the matrix deviation inequality from Tropp (2015) is too coarse for this problem. The extra factor p may be removed using the concentration results of Section 4.3 but at the cost of a much longer argument. Indeed, using

Theorem 4.3, we can establish the concentration of local magnetization around the ground states and conditionally on these magnetizations, the configurations are uniformly distributed. These conditional distributions can be shown to exhibit sub-Gaussian concentration so that $\sigma^\top u$, and thus $\bar{\sigma}^\top u$ are sub-Gaussian with constant variance proxy for any unit vector $u \in \mathbb{R}^p$. This result can yield a bound for $\|\hat{\Gamma} - \Gamma\|_{\mathrm{op}}$ using an ε-net argument that is standard in covariance matrix estimation. With this in mind, we could get an upper bound in (3.7) that is negligible with respect to $\mathcal{R}_{n,p}$ thereby removing a factor 2. Nevertheless, in absence of a tight control of the constant $C_{\alpha,\beta}$, exact constants are hopeless and beyond the scope of this paper.

3.3. *Information theoretic limitations.* In this section, we present lower bounds on the sample size needed to recover the partition $(S, \bar{S})$ and compare them to the upper bounds of Theorem 3.3. In the sequel, we write $\hat{S} \asymp S$ if either $(\hat{S}, \bar{\hat{S}}) = (S, \bar{S})$ or $(\hat{S}, \bar{\hat{S}}) = (\bar{S}, S)$ to indicate that the two partitions are the same. We write $\hat{S} \not\asymp S$ to indicate that the two partitions are different.

For any balanced partition $(S, \bar{S})$, consider a "neighborhood" $\mathcal{T}_S$ of $(S, \bar{S})$ composed of balanced partitions such that for all $(T, \bar{T}) \in \mathcal{T}_S$, we have $\rho(S, T) = 1$ and $\rho(\bar{S}, \bar{T}) = 1$. We first compute the Kullback–Leibler divergence between the distributions $\mathbb{P}_S$ and $\mathbb{P}_T$.

LEMMA 3.6. *For any positive β, $\alpha < \beta$, and $T \in \mathcal{T}_S$, it holds that*

$$\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) = \frac{p-2}{p}(\beta - \alpha)(\Delta - \Omega).$$

PROOF. By definition of the divergence and of the distributions, we have that

$$\begin{aligned}
\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) &= \mathbb{E}_T\left[\log\left(\frac{\mathbb{P}_T}{\mathbb{P}_S}(\sigma)\right)\right] \\
&= \mathbb{E}_T\big[\mathbf{Tr}\big[(Q_T - Q_S)\sigma\sigma^\top\big]\big] \\
&= \mathbf{Tr}\big[(Q_T - Q_S)\Sigma_T\big].
\end{aligned}$$

Note that most of the coefficients of $Q_T - Q_S$ are equal to 0. In fact, noting $\{s\} = S \cap \bar{T}$ and $\{t\} = \bar{S} \cap T$, we have

$$(Q_T - Q_S)_{ij} = \frac{\alpha - \beta}{p} \qquad \text{if } \begin{cases} i \in S \setminus \{s\}, \ j = s, \\ i = s, \ j \in S \setminus \{s\}, \\ i \in \bar{S} \setminus \{t\}, \ j = t, \\ i = t, \ j \in \bar{S} \setminus \{t\}, \end{cases}$$

and

$$(Q_T - Q_S)_{ij} = \frac{\beta - \alpha}{p} \qquad \text{if } \begin{cases} i \in S \setminus \{s\},\ j = t, \\ i = s,\ j \in \bar{S} \setminus \{t\}, \\ i \in \bar{S} \setminus \{t\},\ j = s, \\ i = t,\ j \in S \setminus \{s\}, \end{cases}$$

and 0 otherwise. There are therefore $p - 2$ coefficients of each sign. Furthermore, whenever $(Q_T - Q_S)_{ij} = (\alpha - \beta)/p$, we have $(\Sigma_T)_{ij} = \Omega$, and whenever $(Q_T - Q_S)_{ij} = (\beta - \alpha)/p$, we have $(\Sigma_T)_{ij} = \Delta$. Computing $\mathbf{Tr}[(Q_T - Q_S)\Sigma_T]$ explicitly yields the desired result. □

From this lemma we derive the following lower bound.

THEOREM 3.7. *For $\gamma \in (0, 3/5)$ and $p \geq 6$ and*

$$n \leq \frac{\gamma \log(p/4)}{(\beta - \alpha)(\Delta - \Omega)}.$$

We have

$$\inf_{\hat{S}} \max_{S \in \mathcal{S}} \mathbb{P}_S^{\otimes n}\big((\hat{S}, \bar{\hat{S}}) \not\asymp (S, \bar{S})\big) \geq \frac{p-2}{p}(1 - \gamma - \sqrt{\gamma}) > 0,$$

where the infimum is taken over all estimators of S. Note that the right-hand side of the above inequality goes to 1 as $p \to \infty$ and $\gamma \to 0$.

PROOF. First note that by Lemma 3.6, for any $T \in \mathcal{T}_S$, it holds $|\mathcal{T}_S| = (p/2 - 1)^2$ so that

$$\mathsf{KL}\big(\mathbb{P}_T^{\otimes n}, \mathbb{P}_S^{\otimes n}\big) = n\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) \leq n(\beta - \alpha)(\Delta - \Omega) \leq \gamma \log(p/4) \leq \frac{\gamma}{2} \log |\mathcal{T}_S|.$$

Thus, Theorem 2.5 in Tsybakov (2009) yields

$$\begin{aligned} \inf_{\hat{S}} \max_{S \in \mathcal{P}} \mathbb{P}_S^{\otimes n}(\hat{S} \not\asymp S) &\geq \frac{\sqrt{|\mathcal{T}_S|}}{1 + \sqrt{|\mathcal{T}_S|}}\left(1 - \gamma - \sqrt{\frac{\gamma}{\log(|\mathcal{T}_S|)}}\right) \\ &\geq \frac{p-2}{p}(1 - \gamma - \sqrt{\gamma}) > 0, \end{aligned}$$

for $\gamma \in (0, 3/5)$. □

As we will see in the next section, the gap $\Delta - \Omega$ scales with the size p of the system. Since our lower bound Theorem 3.7 exhibits the same dependence in $\Delta - \Omega$ as the upper bound of Theorem 3.3, we conclude that the SDP relaxation studied in this paper is rate optimal: it achieves exact recovery of the community structure for a sample size that scales optimally with p. In the next section we quantify this scaling in p using an asymptotic analysis of the ground states of the Ising blockmodel.

4. Optimal rates of exact recovery. As seen in Section 3, given $\sigma^{(1)}, \dots, \sigma^{(n)}$ independent copies of $\sigma \sim \mathbb{P}_{\alpha,\beta}$, the sample covariance matrix $\hat{\Sigma}$ defined by

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \sigma^{(t)} \sigma^{(t)\top},$$

is a sufficient statistic for S. From basic concentration results (see Section 3), we have shown that this matrix concentrates around the true covariance matrix $\Sigma = \mathbb{E}_{\alpha,\beta}[\sigma\sigma^\top]$ where $\mathbb{E}_{\alpha,\beta}$ denotes the expectation associated to $\mathbb{P}_{\alpha,\beta}$. Unfortunately, computing Σ directly is quite challenging. As a consequence, there is no simple expression of the quantity of interest $\Delta - \Omega$ as a function of the parameters of the problem p and α, β. Instead, we show that $\mathbb{P}_{\alpha,\beta}$ converges to a mixture of bivariate Gaussians with known centers and covariance matrices. In turn, it give us a handle of quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\sigma)]$ for some test function φ and in particular, it allows us to quantify the gap $\Delta - \Omega$. Obtaining an asymptotic expression of these values is therefore important to derive theoretical guarantees for the estimators considered above. Beyond our statistical task, we show phase transitions that are interesting from a probabilistic point of view.

4.1. *Free energy.* Recall that $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ denotes the *IBM Hamiltonian* (or "energy") defined on $\{-1, 1\}^p$ by

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\left(\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right),$$

so that

$$f_{\alpha,\beta}(\sigma) = \frac{e^{-\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma)}}{Z_{\alpha,\beta}}.$$

As noted above, the density $f_{\alpha,\beta}$ assigns the same probability to configurations that have the same magnetization structure. It follows from elementary computations that

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\frac{m}{4}\left(2\alpha\mu_S\mu_{\bar{S}} + \beta\left(\mu_S^2 + \mu_{\bar{S}}^2\right)\right), \tag{4.1}$$

where we recall that $m = p/2$. Moreover, the number of configurations σ with local magnetizations $\mu = (\mu_S, \mu_{\bar{S}}) \in [-1, 1]^2$ is given by

$$\binom{m}{\frac{\mu_S+1}{2}m}\binom{m}{\frac{\mu_{\bar{S}}+1}{2}m}.$$

This quantity can be approximated using Stirling's formula (see Lemma C.2): For any $\mu \in (-1+\varepsilon, 1-\varepsilon)$, there exists two positive constants $\underline{c}$, $\bar{c}$ such that

$$\frac{\underline{c}}{\sqrt{m}} e^{-mh(\frac{\mu+1}{2})} \le \binom{m}{\frac{\mu+1}{2}m} \le \frac{\bar{c}}{\sqrt{m}} e^{mh(\frac{\mu+1}{2})} \qquad \forall m \ge 1,$$

where $h : [0, 1] \to \mathbb{R}$ is the binary entropy function defined by $h(0) = h(1) = 1$ and for any $s \in (0, 1)$ by

$$h(s) = -s \log(s) - (1 - s) \log(1 - s).$$

Thus, IBM induces a marginal distribution on the local magnetizations that has density

$$\frac{\ell_m(\mu_S, \mu_{\bar{S}})}{m Z_{\alpha,\beta}} \exp\left[-\frac{m}{4} g(\mu_S, \mu_{\bar{S}})\right], \tag{4.2}$$

where $\underline{c}^2 \leq \ell_m(\mu_S, \mu_{\bar{S}}) \leq \bar{c}^2$ and

$$g(\mu_S, \mu_{\bar{S}}) = -2\alpha\mu_S\mu_{\bar{S}} - \beta(\mu_S^2 + \mu_{\bar{S}}^2) - 4h\left(\frac{\mu_S + 1}{2}\right) - 4h\left(\frac{\mu_{\bar{S}} + 1}{2}\right). \tag{4.3}$$

Note that the support of this density is implicitly the set of possible values for pairs of local magnetizations of vectors in $\{-1, 1\}^p$, that is, the set $\mathcal{M}^2$, where

$$\mathcal{M} := \left\{\frac{s^\top \mathbf{1}_{[m]}}{m}, s \in \{-1, 1\}^m\right\} \subset [-1, 1]. \tag{4.4}$$

We call the function g the *free energy* of the Ising blockmodel and its structure of minima is known to control the behavior of the system. Indeed, let g^* denote the minimum value of g over $\mathcal{M}^2$. It follows from (4.2) that any local magnetization $(\mu_S, \mu_{\bar{S}}) \in \mathcal{M}^2$ such that $g(\mu_S, \mu_{\bar{S}}) > g^*$ has a probability exponentially smaller than any magnetization that minimizes g over $\mathcal{M}^2$. Intuitively, this results in a distribution that is concentrated around its modes. Before quantifying this effect, we study the minima, known as *ground states* of the free energy g, when defined over the continuum $[-1, 1]^2$.

4.2. *Ground states.* Recall that when $\alpha = \beta$, the block structure vanishes and the IBM reduces to the well-known Curie–Weiss model. We gather in the Supplementary Material [Berthet, Rigollet and Srivastava (2019)] facts about the Curie–Weiss model that we use in this section.

The following proposition characterizes the ground states of the Ising blockmodel. For any $p \in [1, \infty]$, we denote by $\|\cdot\|_p$ the ℓ_p norm of $\mathbb{R}^2$ and by $\mathcal{B}_p = \{x \in \mathbb{R}^2 : \|x\|_p \leq 1\}$ the unit ball with respect to that norm.

PROPOSITION 4.1. *For any $b \in \mathbb{R}$, let $\pm\tilde{x}(b) \in (-1, 1)$, $\tilde{x}(b) \geq 0$ denote the ground state(s) of the Curie–Weiss model with inverse temperature b. The free energy $g_{\alpha,\beta}$ of the IBM defined in (4.3) has the following minima:*

If $\beta + |\alpha| \leq 2$, then $g_{\alpha,\beta}$ has a unique minimum at $(0, 0)$.

If $\beta + |\alpha| > 2$, then three cases arise:

1. *If $\alpha = 0$, then $g_{\alpha,\beta}$ has four minima at $(\pm\tilde{x}(\beta/2), \pm\tilde{x}(\beta/2))$.*
2. *If $\alpha > 0$, $g_{\alpha,\beta}$ has two minima at $\tilde{s} = (\tilde{x}(\frac{\beta+\alpha}{2}), \tilde{x}(\frac{\beta+\alpha}{2}))$ and $-\tilde{s}$.*

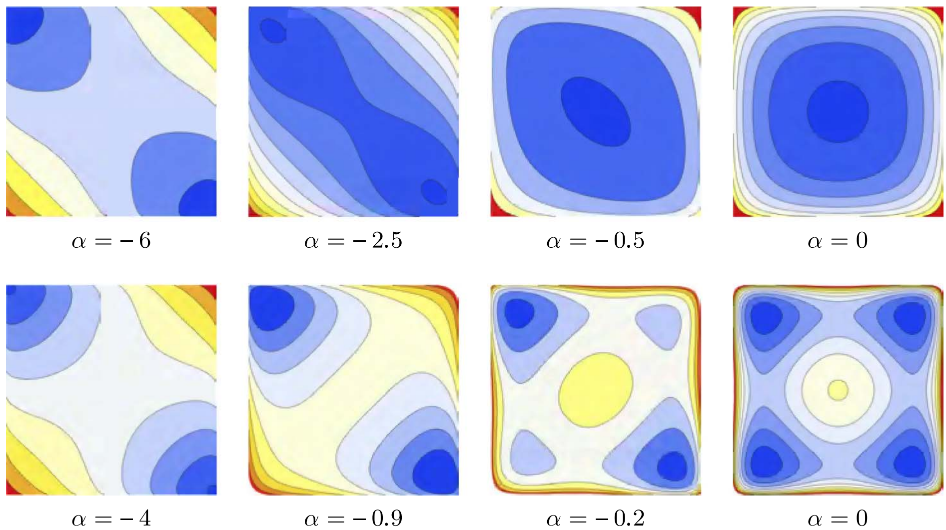

FIG. 2. *Contour plots of the values of the free energy $g_{\alpha,\beta}$ with higher values in red and lower values in blue, corresponding to ground states*. Top row: *Several choices for $\alpha < 0$, and $\beta = 1.5 < 2$.* Bottom row: *Several choices for $\alpha < 0$, and $\beta = 2.5 > 2$. The same plots with $\alpha > 0$ can be obtained by a 90° rotation, by symmetry of the function.*

3. *If $\alpha < 0$, $g_{\alpha,\beta}$ has two minima at $\tilde{s} = (\tilde{x}(\frac{\beta-\alpha}{2}), -\tilde{x}(\frac{\beta-\alpha}{2}))$ and $-\tilde{s}$.*

In particular, for all values of the parameters α and β, all ground states $(\tilde{x}, \tilde{y})$ satisfy $\tilde{x}^2 = \tilde{y}^2 < 1$.

This result is illustrated in Figure 2, composed of contour plots of the free energy $g_{\alpha,\beta}$ on the square $[-1, 1]^2$, for several values of the parameters. The different regions are also represented in Figure 3.

PROOF OF PROPOSITION 4.1. Throughout this proof, for any $b \in \mathbb{R}$, we denote by $g_b^{\mathrm{CW}}(x)$, $x \in [-1, 1]$, the free energy of the Curie–Weiss model with inverse temperature b. We write $g := g_{\alpha,\beta}$ for simplicity to denote the free energy of the IBM.

Note that

$$g(x, y) = g_{\frac{\beta+\alpha}{2}}^{\mathrm{CW}}(x) + g_{\frac{\beta+\alpha}{2}}^{\mathrm{CW}}(y) + \alpha(x - y)^2. \tag{4.5}$$

We split our analysis according to the sign of α. Note first that if $\alpha = 0$, we have

$$g(x, y) = g_{\frac{\beta}{2}}^{\mathrm{CW}}(x) + g_{\frac{\beta}{2}}^{\mathrm{CW}}(y).$$

It yields that:

- If $\beta \le 2$, then $g_{\frac{\beta}{2}}^{\mathrm{CW}}$ has a unique local minimum at $x = 0$ which implies that g has a unique minimum at $(0, 0)$.

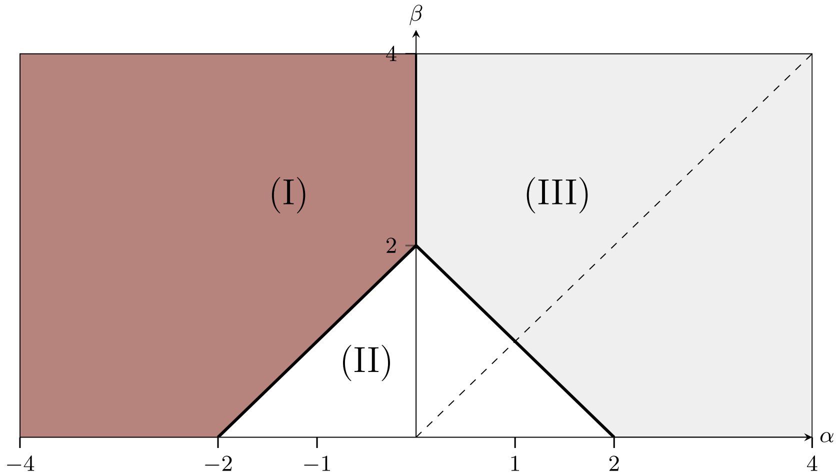

FIG. 3. *Phase diagram of the Ising block model, with three regions for* α *and* $\beta > 0$. *In region* (I), *where* $\alpha < 0$ *and* $\beta + |\alpha| > 2$, *there are two ground states of the form* $(x, -x)$ *and* $(-x, x)$. *In region* (II), *where* $\beta + |\alpha| < 2$, *there is one ground state at* $(0, 0)$. *In region* (III), *where* $\alpha > 0$ *and* $\beta + |\alpha| > 2$, *there are two ground states of the form* (x, x) *and* $(-x, -x)$. *The dotted line has equation* $\alpha = \beta$, *we only consider parameters in the region to its left.*

- If $\beta > 2$, then $g^{\mathrm{CW}}_{\frac{\beta}{2}}$ has exactly two minima at $\tilde{x}(\beta/2)$ and $-\tilde{x}(\beta/2)$, where $\tilde{x}(\beta/2) \in (-1, 1)$. It implies that g has four minima at $(\pm\tilde{x}(\beta/2), \pm\tilde{x}(\beta/2))$.

  Next, if $\alpha > 0$, in view of (4.5) we have

$$g(x, y) \geq g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(y)$$

with equality iff $x = y$. It follows that:

- If $\alpha + \beta \leq 2$, then g has a unique minimum at $(0, 0)$.
- If $\alpha + \beta > 2$, then g has two minima on $\mathcal{A}$ at $(\tilde{x}(\frac{\beta+\alpha}{2}), \tilde{x}(\frac{\beta+\alpha}{2}))$ and at $(-\tilde{x}(\frac{\beta+\alpha}{2}), -\tilde{x}(\frac{\beta+\alpha}{2}))$.

Finally, note that $(x - y)^2 \leq 2x^2 + 2y^2$ with equality iff $x = -y$. Thus, if $\alpha < 0$, in view of (4.5) we have

$$g(x, y) \geq g^{\mathrm{CW}}_{\frac{\beta-\alpha}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta-\alpha}{2}}(y) \tag{4.6}$$

with equality iff $x = -y$. It implies that:

- If $\beta - \alpha \leq 2$, then g has a unique minimum at $(0, 0)$.
- If $\beta - \alpha > 2$, then g has two minima at $(\tilde{x}(\frac{\beta-\alpha}{2}), -\tilde{x}(\frac{\beta-\alpha}{2}))$ and at $(-\tilde{x}(\frac{\beta-\alpha}{2}), \tilde{x}(\frac{\beta-\alpha}{2}))$. $\square$

Using the localization of the ground states from Lemma A.1, we also get the following local and global behaviors of the free energy of the IBM around the ground states.

LEMMA 4.2. *Assume that $\beta + |\alpha| \neq 2$. Denote by $(\tilde{x}, \tilde{y})$ any ground state of Ising blockmodel and recall that $\tilde{x}^2 = \tilde{y}^2$. Then the following holds*:

1. *The Hessian $H_{\alpha,\beta}$ of $g_{\alpha,\beta}$ at $(\tilde{x}, \tilde{y})$ is given by*

$$H_{\alpha,\beta} = -2\begin{pmatrix} \beta & \alpha \\ \alpha & \beta \end{pmatrix} + \frac{4}{1-\tilde{x}^2} I_2.$$

In particular $H_{\alpha,\beta}$ has eigenvalues $2(\alpha-\beta)+4/(1-\tilde{x}^2)$ and $-2(\alpha+\beta)+4/(1-\tilde{x}^2)$ associated with eigenvectors $(1,-1)$ and $(1,1)$, respectively.

2. *There exists positive constants $\delta = \delta(\beta + |\alpha|)$, $\kappa^2 = \kappa^2(\beta + |\alpha|)$ such that the following holds. For any $(x, y) \in (-1, 1)^2$, we have*

$$g(x, y) \geq g(\tilde{x}, \tilde{y}) + \frac{\kappa^2}{2}\big(\|(x, y) - (\tilde{x}, \tilde{y})\|_\infty \wedge \delta\big)^2. \tag{4.7}$$

Moreover:

If $\beta + |\alpha| > 2$, we can take $\delta = e^{-2(\beta+|\alpha|)\frac{\beta+|\alpha|-2}{4(\beta+|\alpha|)}}$ and $\kappa^2 = 1 - \frac{2}{\beta+|\alpha|}$.

If $\beta + |\alpha| < 2$, we can take $\delta = \sqrt{(2-(\beta+|\alpha|))/6}$ and $\kappa^2 = 2 - (\beta + |\alpha|)$.

PROOF. Elementary calculus yields directly that

$$H_{\alpha,\beta} = \begin{pmatrix} -2\beta + \dfrac{4}{1-\tilde{x}^2} & -2\alpha \\ -2\alpha & -2\beta + \dfrac{4}{1-\tilde{y}^2} \end{pmatrix}.$$

Moreover, it follows from Proposition 4.1 that all ground states satisfy $\tilde{x}^2 = \tilde{y}^2$. This completes the proof of the first point.

We now turn to the proof of the second point and split the analysis into five cases: (i) $\alpha \geq 0$ and $\beta + \alpha < 2$, (ii) $\alpha > 0$ and $\beta + \alpha > 2$, (iii) $\alpha < 0$ and $\beta - \alpha < 2$, (iv) $\alpha < 0$ and $\beta - \alpha > 2$, (v) $\alpha = 0$ and $\beta + \alpha > 2$.

Case (i): $\alpha \geq 0$ *and* $\beta + \alpha < 2$. Recall that in this case, g has a unique minimum at $(0, 0)$. Therefore, in view of (4.5) and Lemma A.1, we have

$$\begin{aligned} g(x, y) - g(0, 0) &= g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + \alpha(x-y)^2 \\ &\geq \frac{1}{2}\big(2 - (\beta + |\alpha|)\big)\big[\big(|x - 0| \wedge \varepsilon'\big)^2 + \big(|y - 0| \wedge \varepsilon'\big)^2\big] \\ &\geq \frac{1}{2}\big(2 - (\beta + |\alpha|)\big)\big(\|(x, y) - (0, 0)\|_\infty \wedge \varepsilon'\big)^2, \end{aligned}$$

where $\varepsilon' = \sqrt{(2 - (\beta + |\alpha|))/6}$, which concludes this case.

Case (ii): $\alpha > 0$ *and* $\beta + \alpha > 2$. Recall that in this case, g has two minima denoted generically by $(\tilde{x}, \tilde{y})$ where $\tilde{x} = \tilde{y}$. Therefore, in view of (4.5) and Lemma A.1, we have

$$\begin{aligned} g(x, y) - g(\tilde{x}, \tilde{y}) &= g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{y}) + \alpha(x-y)^2 \\ &\geq \frac{1}{2}\left(1 - \frac{2}{\beta + |\alpha|}\right)\left[(|x - 0| \wedge \varepsilon)^2 + (|y - 0| \wedge \varepsilon)^2\right] \\ &\geq \frac{1}{2}\left(1 - \frac{2}{\beta + |\alpha|}\right)\left(\|(x, y) - (0, 0)\|_\infty \wedge \varepsilon\right)^2, \end{aligned}$$

where $\varepsilon = e^{-2(\beta+|\alpha|)\frac{\beta+|\alpha|-2}{4(\beta+|\alpha|)}}$ which concludes this case.

Case (iii): $\alpha < 0$ *and* $\beta - \alpha < 2$. Recall that in this case, g has a unique minimum at $(0, 0)$. Moreover, in view of (4.6) and Lemma A.1, it holds

$$\begin{aligned} g(x, y) - g(0, 0) &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) \\ &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) \\ &\geq \frac{1}{2}\left(2 - (\beta + |\alpha|)\right)\left(\|(x, y) - (0, 0)\|_\infty \wedge \varepsilon'\right)^2, \end{aligned}$$

where in the second inequality, we used the fact that

$$g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) = g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) = -4h(1/2),$$

and we concluded as in Case (i).

Case (iv): $\alpha < 0$ *and* $\beta - \alpha > 2$. Recall that in this case, g has two minima denoted generically by $(\tilde{x}, \tilde{y})$ where $\tilde{x} = -\tilde{y}$. Therefore, in view of (4.5) and (4.6), we have

$$g(x, y) - g(\tilde{x}, \tilde{y}) \geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(-\tilde{x}) - 4\alpha\tilde{x}^2.$$

Next, observe that from the definition (A.1) of the free energy in the Curie–Weiss model, we have

$$-g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(\tilde{x}) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(-\tilde{x}) - 4\alpha\tilde{x}^2 = -g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(-\tilde{x}).$$

The above two displays yield

$$\begin{aligned} g(x, y) - g(\tilde{x}, \tilde{y}) &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(-\tilde{x}) \\ &\geq \frac{1}{2}\left(1 - \frac{2}{\beta + |\alpha|}\right)\left(\|(x, y) - (0, 0)\|_\infty \wedge \varepsilon\right)^2, \end{aligned}$$

where we concluded as in Case (ii).

Case (v): $\alpha = 0$ *and* $\beta + \alpha > 2$. This case can be handled by combining the arguments made in cases (ii) and (iv), but it is easier to proceed directly. Recall that in this case, there are four distinct minima of g, of the form $(\pm\tilde{x}, \pm\tilde{x})$, where $\tilde{x}$ is the unique positive minimum of $g^{\mathrm{CW}}_{\beta/2}$. For any such minimum $(\tilde{x}, \tilde{y})$, we have, in view of (4.5) followed by an application of Lemma A.1:

$$\begin{aligned} g(x, y) - g(\tilde{x}, \tilde{y}) &\geq g^{\mathrm{CW}}_{\frac{\beta}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta}{2}}(-\tilde{x}) \\ &\geq \frac{1}{2}\left(1 - \frac{2}{\beta}\right)\left[\left(|x - \tilde{x}| \wedge \varepsilon\right)^2 + \left(|y - \tilde{y}| \wedge \varepsilon\right)^2\right] \\ &\geq \frac{1}{2}\left(1 - \frac{2}{\beta}\right)\left(\left\|(x, y) - (\tilde{x}, \tilde{y})\right\|_\infty \wedge \varepsilon\right)^2, \end{aligned}$$

where $\varepsilon = e^{-2\beta}\frac{\beta-2}{4\beta}$, which concludes this case. $\square$

4.3. *Concentration.* As mentioned above, quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\sigma)]$ cannot in general be computed explicitly in the IBM. Fortunately, it will be sufficient for us to compute quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\mu)]$, where we recall that $\mu = (\mu_S, \mu_{\bar{S}})$ denotes the pair of local magnetizations of a random configuration $\sigma \in \{-1, 1\}^p$ drawn according to $\mathbb{P}_{\alpha,\beta}$. While exact computation is still a hard problem, these quantities can be be well approximated using the fact that $\mathbb{P}_{\alpha,\beta}$ is highly concentrated around its ground states for large enough p.

To leverage concentration, we need to consider the "large m" (or equivalently "large p") asymptotic framework. As a result, it will be convenient to write for two sequences a_m, b_m that $a_m \simeq_m b_m$ if $a_m = (1 + o_m(1))b_m$.

Our main result hinges on the following theorem that compares the distribution of $\mu = (\mu_S, \mu_{\bar{S}}) \in [-1, 1]$ to a certain mixture of Gaussians that are centered at the ground states.

THEOREM 4.3. *Consider the IBM with parameters α, β such that $\beta + |\alpha| \neq 2$. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}^{\geq 0}$ be a nonnegative test function such that $\varphi([-1, 1]^2) \subseteq [0, 1]$, and for which there exists a positive constant $\gamma < 3/2$ such that for any ground state $\tilde{s}$,*

$$\mathbb{E}\left[\varphi\left(\tilde{s} + \frac{2}{\sqrt{m}} H^{-1/2} Z\right)\right] \geq C m^{-\gamma}, \tag{4.8}$$

where $Z \sim \mathcal{N}_2(0, I_2)$ and $H = H_{\alpha,\beta}$ denotes the Hessian of the free energy $g_{\alpha,\beta}$ at $\tilde{s}$. Assume further that there exist $D, \epsilon > 0$ and a positive integer d such that $\varphi(x) \leq D + D\|x\|_2^{2d}$ for all $x \in \mathbb{R}^2$, and such that in the ϵ size ℓ_∞-neighborhood η of each ground state $\tilde{s}$, φ satisfies one of the following two regularity conditions:

1. *There exist positive constants C_1, C_2 such that $0 < C_1 < \varphi(x)$ and φ is C_2-Lipschitz in η in the sense that $|\varphi(x) - \varphi(y)| \leq C_2\|x - y\|_\infty$ for all $x, y \in \eta$, or*
2. *There exists a positive constant C_1 such that*

$$\left|\varphi(x) - \varphi(y)\right| \leq C_1 \max\left(\|x - \tilde{s}\|_\infty, \|y - \tilde{s}\|_\infty\right) \cdot \|x - y\|_1$$

for all $x, y \in \eta$.

Then

$$\mathbb{E}_{\alpha,\beta}\big[\varphi(\mu)\big] \simeq_m \frac{1}{|G|}\sum_{\tilde{s}\in G}\mathbb{E}\Big[\varphi\Big(\tilde{s}+\frac{2}{\sqrt{m}}H^{-1/2}Z\Big)\Big],$$

where $G \subset \{(\pm\tilde{x}, \pm\tilde{x})\}$ *denotes the set of ground states of the IBM.*

After this paper had gone to print, we came to learn of some related work on inhomogenous Ising models in the statistical physics literature, where the ground states of the model are identified in a manner similar to our lemma [Fedele and Unguendoli (2012)]. However, for obtaining the tight sample complexity bounds for our recovery algorithm, we crucially need the sharper estimates derived in the lemma and in Theorem 4.3.

Applying this result to the covariance statistic yields the following.

PROPOSITION 4.4. *Let* Δ *and* Ω *be defined as in Lemma 2.1 and recall that* G *denotes the set of ground states of the IBM with parameters* α, β *such that* $\beta + |\alpha| \neq 2$. *Then*

$$\Delta - \Omega = \frac{1}{m}\left(\frac{(\beta-\alpha)(1-\tilde{x}^2)^2}{2-(\beta-\alpha)(1-\tilde{x}^2)}\right) + \frac{1+o_m(1)}{2|G|}\sum_{(\tilde{x},\tilde{y})\in G}(\tilde{x}-\tilde{y})^2 + o_m\left(\frac{1}{m}\right),$$

where $\tilde{x}$ *is the* x *coordinate of an arbitrary* $(\tilde{x}, \tilde{y}) \in G$. *In particular*:

- *If* $\beta + |\alpha| < 2$, *then* $\Delta - \Omega \simeq_m \frac{1}{m}(\frac{\beta-\alpha}{2-(\beta-\alpha)})$, *which is positive when* $\alpha < \beta$.
- *If* $\beta + |\alpha| > 2$, *then three cases arise*:
  1. *if* $\alpha = 0$, *then* $\Delta - \Omega \simeq_m \tilde{x}^2$,
  2. *if* $\alpha > 0$, *then* $\Delta - \Omega \simeq_m \frac{1}{m}(\frac{(\beta-\alpha)(1-\tilde{x}^2)^2}{2-(\beta-\alpha)(1-\tilde{x}^2)})$, *which is positive when* $\alpha < \beta$,
  3. *if* $\alpha < 0$, *then* $\Delta - \Omega \simeq_m 2\tilde{x}^2$.

The proofs of Theorem 4.3 and Proposition 4.4 are deferred to the Supplementary Material. It follows from Proposition 4.4 that if $\beta + |\alpha| \neq 2$ then the covariance matrix Σ takes two values that are separated by a term of order at least $1/m$ and even sometimes of order 1. This result makes it possible to express the theoretical guarantees on the sample size of Sections 3.2 and 3.3 only in terms of p, β, α.

REMARK 4.5. Note that Theorem 4.3, and hence Proposition 4.4, do not cover the case $\beta + |\alpha| = 2$, which includes the boundary along which the phase transitions in the model and the sample complexity occur. The technical reason for this is that the Hessian of the free energy at the ground state becomes singular on this boundary. Thus, the free energy is no more approximately quadratic in the vicinity of the ground states and the Gaussian behavior exhibited in Theorem 4.3 is lost.

A similar subtlety arises at the critical boundary in many related contexts. An example is the recent identification, following a long line of work, of phase transitions in the hard core lattice gas and anti-ferromagnetic Ising model with a complexity theoretic phase transition [Li, Lu and Yin (2012); Sinclair, Srivastava and Thurley (2014); Sly and Sun (2014); Weitz (2006)], where the behavior of the problem on the critical boundary is still not fully resolved. We leave the exact determination of the behavior of the IBM at the phase transition boundary to future work.

4.4. *Rates of exact recovery.* Combining the results of Proposition 4.4 that quantifies the gap $\Delta - \Omega$ in terms of the dimension p and of Theorem 3.3 readily yields the following corollary.

COROLLARY 4.6. *Let β and α be parameters for the IBM such that $\beta + |\alpha| \neq 2$ and $\alpha < \beta$. There exist positive constants C_1 and C_2 that depend on α and β such that the following holds. The SDP relaxation* (3.4) *recovers the block structure $(S, \bar{S})$ exactly with probability $1 - \delta$ whenever*:

1. $n \geq C_1 p \log(p/\delta)$ *if* $\beta + |\alpha| < 2$ *or* $\alpha > 0$,
2. $n \geq C_2 \log(p/\delta)$ *otherwise.*

In particular, if $\beta - \alpha > 2$, $\alpha \leq 0$ a number of observations that is logarithmic in the dimension p is sufficient to recover the blocks exactly.

These results suggest that there is a sharp phase transition in sample complexity for this problem, depending on the value of the parameters α and β. We address this question further in Section 5. As noted above, our upper and lower bounds on the sample complexity (in Theorems 3.3 and 3.7, resp.) match up to a numerical constant. The sample complexity stated in Corollary 4.6 has optimal dependence on the dimension p. Note that past work on exact recovery in the stochastic blockmodel [Abbe, Bandeira and Hall (2016); Hajek, Wu and Xu (2016)] has shown that some SDP relaxation was also optimal with respect to constants. We do not pursue this question in the present paper.

5. Conclusion and open problems. This paper introduces the Ising blockmodel (IBM) for large binary random vectors with an underlying cluster structure. In this model, we studied the sample complexity of recovering exactly the clusters. Unsurprisingly, this paper bears similarities with the stochastic blockmodel, but also differences. For example, in the stochastic blockmodel one is given only one observation of the graph. In the IBM, given one realization $\sigma^{(1)} \in \{-1, 1\}^p$, the maximum likelihood estimator is the trivial clustering that assigns $i \in [p]$ to a cluster according to the sign of $\sigma_i^{(1)}$, up to a trivial reassignment to keep the partition balanced.

Below is a summary of our main findings:

1. The model exhibits three phases depending on the values taken by two parameters.
2. In one phase, where the two clusters tend to have opposite behavior, the sample complexity is logarithmic in the dimension; in the other two, it is near linear. These sample complexities are proved to be optimal in an information theoretic sense.
3. Akin to the stochastic blockmodel, the optimal sample complexity is achieved using the natural semidefinite relaxation to the MAXCUT problem.

Many questions regarding this model remain open. The first and most natural is the determination of exact constants. Theorems 3.3 and 3.7 suggest that there exists a universal constant $C^\star$ such that the optimal sample complexity is

$$\frac{C^\star \log(p)}{(\beta-\alpha)(\Delta-\Omega)}\bigl(1+o_p(1)\bigr).$$

Throughout this paper, we have only loosely kept track of the correct dependency of the constants in the upper and lower bounds as function of the constants (α, β). We have shown that the optimal sample complexity is a product of $\log(p)/(\Delta-\Omega)$ and of a constant term that only becomes arbitrarily large when α is arbitrarily close to β, with a divergence of order $(\beta-\alpha)^{-1}$, which is consistent with our lower bound. In the spirit of exact thresholds for the stochastic blockmodel [Abbe, Bandeira and Hall (2016); Massoulié (2014); Mossel, Neeman and Sly (2015)], we find that proving existence of the constant $C^\star$ and computing it worthy of investigation but is beyond the scope of the present paper.

Another possible development is the extension of this model to settings with multiple blocks, possibly of unbalanced sizes. This has been studied in the case of the stochastic blockmodel for graphs in the sparse case [Abbe and Sandon (2015); Banks et al. (2016)] and in the dense case [Gao et al. (2015, 2016); Rohe, Chatterjee and Yu (2011)], as well as in the case of clustering for Gaussian variables [Bunea, Giraud and Luo (2015); Bunea et al. (2016)]. For the Ising blockmodel, the main challenge is that the population covariance matrix cannot be directly computed from the parameters of the problem, and an analysis of the ground states of the free energy is required. Developing a general approach to this task, rather than having to do an ad hoc analysis for each case would be an important step in this direction. In another direction, a possible extension would be to study the impact of a an external magnetic field μ on the Ising blockmodel. Problems related to the influence of a sparse external magnetic field in a Curie–Weiss model have recently been investigated [Mukherjee, Mukherjee and Yuan (2016)], and similar phenomena of phase transitions for sample complexities have been exhibited.

As pointed out in the Introduction, the Ising model is in general related to a corresponding Glauber dynamics, a natural Markov chain with the desired Gibbs distribution as a stationary distribution. Exact recovery from dependent observations arising from the Glauber dynamics is a natural extension. Tools to study this question have been developed in Bresler, Gamarnik and Shah (2014).

We have only analyzed in this work the performance of the semidefinite positive relaxation of the maximum likelihood problem, but other methods can be considered for total or partial recovery. In related problems, belief propagation is used to recover communities [see, e.g., Abbe and Sandon (2016a, 2016b); Lesieur, Krzakala and Zdeborová (2017); Moitra, Perry and Wein (2016); Mossel, Neeman and Sly (2016), and work cited above]. In particular, Lesieur, Krzakala and Zdeborová (2017) covers Hopfield models, which are a generalization of our model.

It is possible that studying these types of algorithms is necessary in order to obtain sharper rates.

Finally, in view of the simple spectral decomposition (3.3) of Γ, one may wonder about the behavior of a simple method that consists in computing the leading eigenvector of $\hat{\Gamma}$ and clustering according to the sign of its entries. Such a method is the basis of the approach in denser graph models in McSherry (2001) or Alon, Krivelevich and Sudakov (1998). The results of such an approach are easily implementable as follows.

Let $\hat{u}$ denote a leading unit eigenvectors of $\hat{\Gamma}$ and consider the following estimate for the partition $(S, \bar{S})$:

$$\hat{S} \asymp \{i \in [p] \mid \hat{u}_i > 0\}. \tag{5.1}$$

It follows from the Perron–Frobenius theorem that $\hat{S} \asymp S$ whenever $\mathrm{sign}(\hat{\Gamma}) = \mathrm{sign}(\Gamma)$. This allows for perfect recovery of S, but only holds with high probability when n is of order $\log(p)/(\Delta - \Omega)^2$, which is suboptimal. It is however possible to obtain partial recovery guarantees for the spectral recovery. In order to state our result, for any two partitions $(S, \bar{S})$, $(T, \bar{T})$ define

$$|S \diamond T| = \min(|S \triangle T|, |S \triangle \bar{T}|),$$

where $\triangle$ denotes the symmetric difference.

PROPOSITION 5.1. *Fix $\delta \in (0, 1)$ and let $\hat{S} \subset [p]$ be defined in (5.1). Then there exits a constant $\gamma_{\alpha,\beta} > 0$ such that with probability $1 - \delta$,*

$$\frac{1}{p}|S \diamond \hat{S}| \le \gamma_{\alpha,\beta} \frac{\log(4p/\delta)}{n(\Delta - \Omega)}.$$

PROOF. Let $\hat{u}$ denote the leading unit eigenvector of $\hat{\Gamma}$ and let $\hat{v} = \sqrt{p}\hat{u}$. Recall that $v_S = \mathbf{1}_S - \mathbf{1}_{\bar{S}}$ and observe that

$$\begin{aligned} |S \diamond \hat{S}| &= \min\left(\sum_{i=1}^{p} \mathbb{1}(\hat{v}_i \cdot (v_S)_i \le 0), \sum_{i=1}^{p} \mathbb{1}(\hat{v}_i \cdot (v_S)_i \ge 0)\right) \\ &\le \min(\|\hat{v} - v_S\|^2, \|\hat{v} + v_S\|^2) = p \min(\|\hat{u} - u_S\|^2, \|\hat{u} + u_S\|^2), \end{aligned}$$

where in the inequality we used the fact that $v_S \in \{-1, 1\}^p$ so that

$$\mathbb{1}(\hat{v}_i \cdot (v_S)_i \le 0) \le |\hat{v}_i - (v_S)_i| \mathbb{1}(\hat{v}_i \cdot (v_S)_i \le 0) \le |\hat{v}_i - (v_S)_i|^2.$$

Using a variant of the Davis–Kahan lemma [see, e.g, Wang, Berthet and Samworth (2016)], we get

$$\frac{1}{p}|S \diamond \hat{S}| \le \frac{\|\hat{\Gamma} - \Gamma\|_{\mathrm{op}}^2}{(\lambda_1(\Gamma) - \lambda_2(\Gamma))^2},$$

and the result follows readily from (3.7) and the fact that the eigengap of Γ is given by $p(\Delta - \Omega)/2$. □

In terms of exact recovery, this result is quite weak as it only gives guarantees for a sample complexity of the order of $p \log(p/\delta)/(\Delta - \Omega)$, which is suboptimal by a factor of p. Moreover, for the bound of Proposition 5.1 to be nontrivial, one already needs the sample size to be of the same order as the one required for exact recovery by semidefinite programming. Nevertheless, Proposition 5.1 raises the question of the optimal rates of estimation of S with respect to the metric $|S \diamond \hat{S}|/p$. While partial recovery is beyond the scope of this paper, it would be interesting to establish the optimal rate.

Acknowledgements. The authors thank Andrea Montanari for pointing out a connection to the Hopfield model, and the anonymous referees for their very helpful input.

SUPPLEMENTARY MATERIAL

Supplement to "Exact recovery in the Ising blockmodel" (DOI: 10.1214/17-AOS1620SUPP; .pdf). The Supplementary Material contains additional facts about the Curie–Weiss model in Appendix A and proofs of technical results in Appendix B.

REFERENCES

ABBE, E. (2017). Community detection and stochastic block models: Recent developments. Preprint. Available at arXiv:1703.10146.

ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. MR3447993

ABBE, E. and SANDON, C. (2015). Detection in the stochastic block model with multiple clusters: Proof of the achievability conjectures, acyclic BP, and the information-computation gap. Preprint. Available at arXiv:1512.09080.

ABBE, E. and SANDON, C. (2016a). Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 1334–1342. Curran Associates, Inc., New York.

ABBE, E. and SANDON, C. (2016b). Crossing the KS threshold in the stochastic block model with information theory. In *Proceedings of the* 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 840–844. IEEE, New York.

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. In *Proceedings of the* 1998 *ACM–SIAM Symposium on Discrete Algorithms* 594–598. SIAM, Philadelphia, PA.

BANERJEE, O., EL GHAOUI, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516.

BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. Preprint. Available at arXiv:1601.02658.

BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2019). Supplement to "Exact recovery in the Ising blockmodel." DOI:10.1214/17-AOS1620SUPP.

BESAG, J. (1986). On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575

BRESLER, G. (2015). Efficiently learning Ising models on arbitrary graphs [extended abstract]. In *Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 771–782. ACM, New York.

BRESLER, G., GAMARNIK, D. and SHAH, D. (2014). Learning graphical models from the glauber dynamics. Preprint. Available at arXiv:1410.7659.

BRESLER, G., MOSSEL, E. and SLY, A. (2008). Reconstruction of Markov random fields from samples: Some observations and algorithms. In *Approximation, Randomization and Combinatorial Optimization. Lecture Notes in Computer Science* **5171** 343–356. Springer, Berlin.

BUNEA, F., GIRAUD, C. and LUO, X. (2015). Minimax optimal variable clustering in G-models via Cord. Preprint. Available at arXiv:1508.01939.

BUNEA, F., GIRAUD, C., ROYER, M. and VERZELEN, N. (2016). PECOK: A convex optimization approach to variable clustering. Preprint. Available at arXiv:1606.05100.

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* (3) **84** 066106.

DIACONIS, P., GOEL, S. and HOLMES, S. (2008). Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat.* **2** 777–807.

DYER, M. E. and FRIEZE, A. M. (1989). The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489.

FEDELE, M. and UNGUENDOLI, F. (2012). Rigorous results on the bipartite mean-field model. *J. Phys. A* **45** 385001. MR2970551

FEIGE, U. and KRAUTHGAMER, R. (2002). A polylogarithmic approximation of the minimum bisection. *SIAM J. Comput.* **31** 1090–1118 (electronic).

GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2015). Achieving optimal misclassification proportion in stochastic block model. Preprint. Available at arXiv:1505.03772.

GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2016). Community detection in degree-corrected block models. Preprint. Available at arXiv:1607.06993.

GAREY, M. R., JOHNSON, D. S. and STOCKMEYER, L. (1976). Some simplified NP-complete graph problems. *Theoret. Comput. Sci.* **1** 237–267.

GOEMANS, M. X. and WILLIAMSON, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.* **42** 1115–1145.

HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inform. Theory* **62** 2788–2797.

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137.

ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31** 253–258.

LAURENT, M. and POLJAK, S. (1996). On the facial structure of the set of correlation matrices. *SIAM J. Matrix Anal. Appl.* **17** 530–547.

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. MR1419991

LAURITZEN, S. L. and SHEEHAN, N. A. (2003). Graphical models for genetic analyses. *Statist. Sci.* **18** 489–514. MR2059327

LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.* **2017** 073403. MR3683819

LI, L., LU, P. and YIN, Y. (2012). Correlation decay up to uniqueness in spin systems. In *Proceedings of the Twenty-Fourth Annual ACM–SIAM Symposium on Discrete Algorithms* 67–84. SIAM, Philadelphia, PA. MR3185380

MANNING, C. D. and SCHÜTZE, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *Proceedings of the* 46*th Annual ACM Symposium on Theory of Computing*. ACM, New York.

MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA.

MOITRA, A., PERRY, W. and WEIN, A. S. (2016). How robust are reconstruction thresholds for community detection? In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* 828–841.

MONTANARI, A. and SABERI, A. (2010). The spread of innovations in social networks. *Proc. Natl. Acad. Sci. USA* **107** 20196–20201.

MOSSEL, E., NEEMAN, J. and SLY, A. (2013). A proof of the block model threshold conjecture. Preprint. Available at arXiv:1311.4115.

MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461.

MOSSEL, E., NEEMAN, J. and SLY, A. (2016). Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.* **26** 2211–2256. MR3543895

MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2016). Global testing against sparse alternatives under Ising models. Preprint. Available at arXiv:1611.08293.

RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319.

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

SCHNEIDMAN, E., BERRY, M. J., SEGEV, R. and BIALEK, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440** 1007–1012.

SEBASTIANI, P., RAMONI, M. F., NOLAN, V., BALDWIN, C. T. and STEINBERG, M. H. (2005). Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat. Genet.* **37** 435–440.

SINCLAIR, A., SRIVASTAVA, P. and THURLEY, M. (2014). Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.* **155** 666–686.

SLY, A. and SUN, N. (2014). Counting in two-spin models on d-regular graphs. *Ann. Probab.* **42** 2383–2416. MR3265170

TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438

WEITZ, D. (2006). Counting independent sets up to the tree threshold. In *Proceedings of the* 2006 *ACM Symposium on the Theory of Computing* 140–149. ACM, New York.

Q. BERTHET
DEPARTMENT OF PURE MATHEMATICS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
WILBEFORCE ROAD
CAMBRIDGE, CB3 0WB
UNITED KINGDOM
E-MAIL: q.berthet@statslab.cam.ac.uk

P. RIGOLLET
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 MASSACHUSETTS AVENUE
CAMBRIDGE, MASSACHUSETTS 02139-4307
USA
E-MAIL: rigollet@math.mit.edu

P. SRIVASTAVA
SCHOOL OF TECHNOLOGY AND COMPUTER SCIENCE
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
1 DR. HOMI BHABHA ROAD, NAVYNAGAR, COLABA
MUMBAI, MH 400 005
INDIA
E-MAIL: piyush.srivastava@tifr.res.in