Phase transitions in semidefinite relaxations

Adel Javanmard[a], Andrea Montanari[b,c,1], and Federico Ricci-Tersenghi[d,e,f]

[a]Marshall School of Business, University of Southern California, Los Angeles, CA 90089; [b]Department of Statistics, Stanford University, Stanford, CA 94305; [c]Department of Electrical Engineering, Stanford University, Stanford, CA 94305; [d]Dipartimento di Fisica, Università La Sapienza, 00185 Rome, Italy; [e]Istituto Nazionale di Fisica Nucleare, Sezione di Roma1, 00185 Rome, Italy; and [f]Rome unit, Nanotec, Consiglio Nazionale delle Ricerche, 00185 Rome, Italy

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved February 17, 2016 (received for review November 21, 2015)

Statistical inference problems arising within signal processing, data mining, and machine learning naturally give rise to hard combinatorial optimization problems. These problems become intractable when the dimensionality of the data is large, as is often the case for modern datasets. A popular idea is to construct convex relaxations of these combinatorial problems, which can be solved efficiently for large-scale datasets. Semidefinite programming (SDP) relaxations are among the most powerful methods in this family and are surprisingly well suited for a broad range of problems where data take the form of matrices or graphs. It has been observed several times that when the statistical noise is small enough, SDP relaxations correctly detect the underlying combinatorial structures. In this paper we develop asymptotic predictions for several detection thresholds, as well as for the estimation error above these thresholds. We study some classical SDP relaxations for statistical problems motivated by graph synchronization and community detection in networks. We map these optimization problems to statistical mechanics models with vector spins and use nonrigorous techniques from statistical mechanics to characterize the corresponding phase transitions. Our results clarify the effectiveness of SDP relaxations in solving high-dimensional statistical problems.

semidefinite programming | phase transitions | synchronization | community detection

Many information processing tasks can be formulated as optimization problems. This idea has been central to data analysis and statistics at least since Gauss and Legendre's invention of the least-squares method in the early 19th century (1).

Modern datasets pose new challenges to this centuries-old framework. On one hand, high-dimensional applications require the simultaneous estimation of millions of parameters. Examples span genomics (2), imaging (3), web services (4), and so on. On the other hand, the unknown object to be estimated has often a combinatorial structure: In clustering we aim at estimating a partition of the data points (5). Network analysis tasks usually require identification of a discrete subset of nodes in a graph (6, 7). Parsimonious data explanations are sought by imposing combinatorial sparsity constraints (8).

There is an obvious tension between the above requirements. Although efficient algorithms are needed to estimate a large number of parameters, the maximum likelihood (ML) method often requires the solution of NP-hard (nondeterministic polynomial-time hard) combinatorial problems. A flourishing line of work addresses this conundrum by designing effective convex relaxations of these combinatorial problems (9–11).

Unfortunately, the statistical properties of such convex relaxations are well understood only in a few cases [compressed sensing being the most important success story (12–14)]. In this paper we use tools from statistical mechanics to develop a precise picture of the behavior of a class of semidefinite programming relaxations. Relaxations of this type appear to be surprisingly effective in a variety of problems ranging from clustering to graph synchronization. For the sake of concreteness we will focus on three specific problems.

$\mathbb{Z}_2$ Synchronization

In the general synchronization problem, we aim at estimating $x_{0,1},x_{0,2},\dots,x_{0,n}$, which are unknown elements of a known group $\mathfrak{G}$. This is done using data that consist of noisy observations of relative positions $Y_{ij}=x_{0,i}^{-1}x_{0,j}+\text{noise}$. A large number of practical problems can be modeled in this framework. For instance, the case $\mathfrak{G}=SO(3)$ (the orthogonal group in three dimensions) is relevant for camera registration and molecule structure reconstruction in electron microscopy (15).

$\mathbb{Z}_2$ synchronization is arguably the simplest problem in this class and corresponds to $\mathfrak{G}=\mathbb{Z}_2$ (the group of integers modulo 2). Without loss of generality, we will identify this with the group $(\{+1,-1\},\cdot)$ (elements of the group are $+1$, -1, and the group operation is ordinary multiplication). We assume observations to be distorted by Gaussian noise; namely, for each $i<j$ we observe $Y_{ij}=(\lambda/n)\,x_{0,i}x_{0,j}+W_{ij}$, where $W_{ij}\sim\mathsf{N}(0,1/n)$ are independent standard normal random variables. This fits the general definition because $x_{0,i}^{-1}=x_{0,i}$ for $x_{0,i}\in\{+1,-1\}$.

In matrix notation, we observe a symmetric matrix $\boldsymbol{Y}=\boldsymbol{Y}^*\in\mathbb{R}^{n\times n}$ given by

$$\boldsymbol{Y}=\frac{\lambda}{n}\,\boldsymbol{x}_0\,\boldsymbol{x}_0^*+\boldsymbol{W}. \qquad [1]$$

(Note that entries on the diagonal carry no information.) Here $\boldsymbol{x}_0\in\{+1,-1\}^n$ and $\boldsymbol{x}_0^*$ denote the transpose of $\boldsymbol{x}_0$, and $\boldsymbol{W}=(W_{ij})_{i,j\le n}$ is a random matrix from the Gaussian orthogonal ensemble (GOE), i.e., a symmetric matrix with independent entries (up to symmetry) $(W_{ij})_{1\le i<j\le n}\sim_{i.i.d.}\mathsf{N}(0,1/n)$ and $(W_{ii})_{1\le i\le n}\sim_{i.i.d.}\mathsf{N}(0,2/n)$.

A solution of the $\mathbb{Z}_2$ synchronization problem can be interpreted as a bipartition of the set $\{1,\dots,n\}$. Hence, this has been used as a model for partitioning signed networks (16, 17).

$U(1)$ Synchronization

This is again an instance of the synchronization problem. However, we take $\mathfrak{G}=U(1)$. This is the group of complex number of modulus one, with the operation of complex multiplication $\mathfrak{G}=(\{x\in\mathbb{C}:|x|=1\},\cdot)$.

Significance

Modern data analysis requires solving hard optimization problems with a large number of parameters and a large number of constraints. A successful approach is to replace these hard problems by surrogate problems that are convex and hence tractable. Semidefinite programming relaxations offer a powerful method to construct such relaxations. In many instances it was observed that a semidefinite relaxation becomes very accurate when the noise level in the data decreases below a certain threshold. We develop a new method to compute these noise thresholds (or phase transitions) using ideas from statistical physics.

Author contributions: A.J., A.M., and F.R.-T. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 4238.

[1]To whom correspondence should be addressed. Email: montanari@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523097113/-/DCSupplemental.

As in the previous case, we assume observations to be distorted by Gaussian noise; that is, for each $i<j$ we observe $Y_{ij}=(\lambda/n)\, x_{0,i}\,\bar{x}_{0,j}+W_{ij}$, where $\bar{z}$ denotes complex conjugation[†] and $W_{ij}\sim \mathrm{CN}(0,1/n)$.

In matrix notations, this model takes the same form as [**1**], provided we interpret $\boldsymbol{x}_0^*$ as the conjugate transpose of vector $\boldsymbol{x}_0\in\mathbb{C}^n$, with components $x_{0,i}$, $|x_{0,i}|=1$. We will follow this convention throughout.

$U(1)$ synchronization has been used as a model for clock synchronization over networks (18, 19). It is also closely related to the phase-retrieval problem in signal processing (20–22). An important qualitative difference with respect to the previous example ($\mathbb{Z}_2$ synchronization) lies in the fact that $U(1)$ is a continuous group. We regard this as a prototype of synchronization problems over compact Lie groups [e.g., $SO(3)$].

Hidden Partition

The hidden (or planted) partition (also known as community detection) model is a statistical model for the problem of finding clusters in large network datasets (see refs. 7, 23, 24 and references therein for earlier work). The data consist of graph $G=(V,E)$ over vertex set $V=[n]\equiv\{1,2,\dots,n\}$ generated as follows. We partition $V=V_+\cup V_-$ by setting $i\in V_+$ or $i\in V_-$ independently across vertices with $\mathbb{P}(i\in V_+)=\mathbb{P}(i\in V_-)=1/2$. Conditional on the partition, edges are independent with

$$\mathbb{P}\{(i,j)\in E|V_+,V_-\}=\begin{cases}a/n & \text{if } \{i,j\}\subseteq V_+ \text{ or } \{i,j\}\subseteq V_-.\\ b/n & \text{otherwise.}\end{cases} \quad [2]$$

Here $a>b>0$ are model parameters that will be kept of order one as $n\to\infty$. This corresponds to a random graph with bounded average degree $d=(a+b)/2$ and a cluster (also known as block or community) structure corresponding to the partition $V_+\cup V_-$. Given a realization of such a graph, we are interested in estimating the underlying partition.

We can encode the partition V_+, V_- by a vector $\boldsymbol{x}_0\in\{+1,-1\}^n$, letting $x_{0,i}=+1$ if $i\in V_+$ and $x_{0,i}=-1$ if $i\in V_-$. An important insight, which we will further develop below (25, 26), is that this problem is analogous to $\mathbb{Z}_2$ synchronization, with signal strength $\lambda=(a-b)/\sqrt{2(a+b)}$. The parameters' correspondence is obtained, at a heuristics level, by noting that if $\boldsymbol{A}_G$ is the adjacency matrix of G, then $\mathbb{E}\langle\boldsymbol{x}_0,\boldsymbol{A}_G\boldsymbol{x}_0\rangle/(n\mathbb{E}\|\boldsymbol{A}_G\|_F^2)^{1/2}\approx(a-b)/\sqrt{2(a+b)}$. (Here and below, $\langle\boldsymbol{a},\boldsymbol{b}\rangle=\sum_i a_i b_i$ denotes the standard scalar product between vectors.)

A generalization of this problem to the case of more than two blocks has been studied since the 1980s as a model for social network structure (27), under the name of "stochastic block model." For the sake of simplicity, we will focus here on the two-blocks case.

Illustrations

As a first preview of our results, Fig. 1 reports our analytical predictions for the estimation error in the $\mathbb{Z}_2$ synchronization problem, comparing them with numerical simulations using semidefinite programming (SDP). An estimator is a map $\hat{\boldsymbol{x}}:\mathbb{R}^{n\times n}\to\mathbb{R}^n$, $\boldsymbol{Y}\mapsto\hat{\boldsymbol{x}}(\boldsymbol{Y})$. We compare various estimators in terms of their per-coordinate mean square error (MSE):

$$\mathrm{MSE}_n(\hat{\boldsymbol{x}})\equiv\frac{1}{n}\mathbb{E}\left\{\min_{s\in\{+1,-1\}}\|\hat{\boldsymbol{x}}(\boldsymbol{Y})-s\,\boldsymbol{x}_0\|_2^2\right\}, \quad [3]$$

where expectation is with respect to the noise model [**1**] and $\boldsymbol{x}_0\in\{+1,-1\}^n$ uniformly random. Note the minimization with respect to the sign $s\in\{+1,-1\}$ inside the expectation: because of the symmetry of [**1**], the vector $\boldsymbol{x}_0$ can only be estimated up to a global sign. We will be interested in the high-dimensional limit $n\to\infty$ and will omit the subscript n—thus writing $\mathrm{MSE}(\hat{\boldsymbol{x}})$—to denote this limit. Note that a trivial estimator that always returns 0 has error $\mathrm{MSE}_n(\boldsymbol{0})=1$.

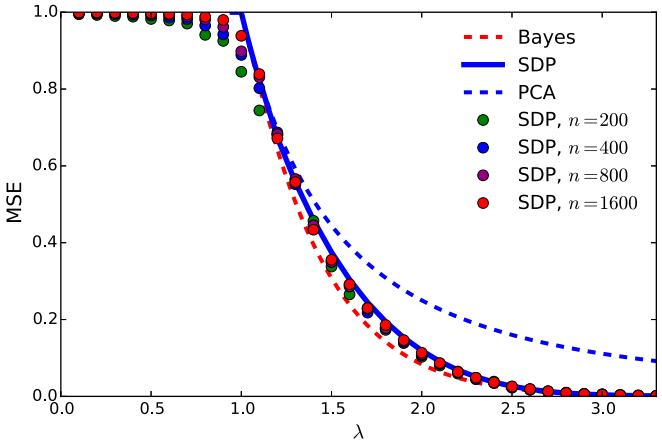


Fig. 1. Estimating $\boldsymbol{x}_0\in\{+1,-1\}^n$ under the noisy $\mathbb{Z}_2$ synchronization model of Eq. **1**. Curves correspond to (asymptotic) analytical predictions, and dots correspond to numerical simulations (averaged over 100 realizations).

Classical statistical theory suggests two natural reference estimators: the Bayes optimal and the maximum likelihood estimators. We will discuss these methods first, to set the stage for SDP relaxations.

Bayes Optimal Estimator. The Bayes optimal estimator (also known as minimum MSE) provides a lower bound on the performance of any other approach. It takes the conditional expectation of the unknown signal given the observations:

$$\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}}(\boldsymbol{Y})=\mathbb{E}\{\boldsymbol{x}|(\lambda/n)\boldsymbol{x}\boldsymbol{x}^*+\boldsymbol{W}=\boldsymbol{Y}\}. \quad [4]$$

Explicit formulas are given in *SI Appendix*. We note that $\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}}(\boldsymbol{Y})$ assumes knowledge of the prior distribution. The red dashed curve in Fig. 1 presents our analytical prediction for the asymptotic MSE for $\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}}(\,\cdot\,)$. Notice that $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}})=1$ for all $\lambda\le 1$ and $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}})<1$ strictly for all $\lambda>1$, with $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{Bayes}})\to 0$ quickly as $\lambda\to\infty$. The point $\lambda_c^{\mathrm{Bayes}}=1$ corresponds to a phase transition for optimal estimation, and no method can have nontrivial MSE for $\lambda\le\lambda_c^{\mathrm{Bayes}}$.

Maximum Likelihood. The estimator $\hat{\boldsymbol{x}}^{\,\mathrm{ML}}(Y)$ is given by the solution of

$$\hat{\boldsymbol{x}}^{\,\mathrm{ML}}(\boldsymbol{Y})=c(\lambda)\arg\max_{\boldsymbol{x}\in\{+1,-1\}^n}\langle\boldsymbol{x},\boldsymbol{Y}\boldsymbol{x}\rangle. \quad [5]$$

Here $c(\lambda)$ is a scaling factor[‡] that is chosen according to the asymptotic theory as to minimize the MSE. As for the Bayes optimal curve, we obtain $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{ML}})=1$ for $\lambda\le\lambda_c^{\mathrm{ML}}=1$ and $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{ML}})<1$ (and rapidly decaying to 0) for $\lambda>\lambda_c^{\mathrm{ML}}$. (We refer to *SI Appendix* for this result.)

[†]Here and below, $\mathrm{CN}(\mu,\sigma^2)$, with $\mu=\mu_1+i\,\mu_2$ and $\sigma^2\in\mathbb{R}_{\ge 0}$, denotes the complex normal distribution. Namely, $X\sim\mathrm{CN}(\mu,\sigma^2)$ if $X=X_1+i\,X_2$, with $X_1\sim\mathrm{N}(\mu_1,\sigma^2/2)$ and $X_2\sim\mathrm{N}(\mu_2,\sigma^2/2)$ independent Gaussian random variables.

[‡]In practical applications, λ might not be known. We are not concerned by this at the moment because maximum likelihood is used as a idealized benchmark here. Note that strictly speaking, this is a scaled maximum likelihood estimator. We prefer to scale $\hat{\boldsymbol{x}}^{\,\mathrm{ML}}(\boldsymbol{Y})$ to keep $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\,\mathrm{ML}})\in[0,1]$.

Semidefinite Programming. Neither the Bayes nor the maximum likelihood approaches can be implemented efficiently. In particular, solving the combinatorial optimization problem in Eq. **5** is a prototypical NP-complete problem. Even worse, approximating the optimum value within a sublogarithmic factor is computationally hard (28) (from a worst case perspective). SDP relaxations allow us to obtain tractable approximations. Specifically, and following a standard lifting idea, we replace the problem [**5**] by the following semidefinite program over the symmetric matrix $\boldsymbol{X}\in\mathbb{R}^{n\times n}$ (18. 29, 30):

$$\begin{aligned}&\text{maximize } \langle \boldsymbol{X},\boldsymbol{Y}\rangle,\\ &\text{subject to } \boldsymbol{X}\succcurlyeq 0,\quad X_{ii}=1\forall i\in[n].\end{aligned} \qquad [6]$$

We use $\langle\cdot,\cdot\rangle$ to denote the scalar product between matrices, namely, $\langle \boldsymbol{A},\boldsymbol{B}\rangle\equiv \mathsf{Tr}(\boldsymbol{A}^*\boldsymbol{B})$, and $\boldsymbol{A}\succcurlyeq 0$ to indicate that $\boldsymbol{A}$ is positive semidefinite[§] (PSD). If we assume $\boldsymbol{X}=\boldsymbol{x}\boldsymbol{x}^*$, the SDP [**6**] reduces to the maximum-likelihood problem [**5**]. By dropping this condition, we obtain a convex optimization problem that is solvable in polynomial time. Given an optimizer $\boldsymbol{X}_{\rm opt}=\boldsymbol{X}_{\rm opt}(\boldsymbol{Y})$ of this convex problem, we need to produce a vector estimate. We follow a different strategy from standard rounding methods in computer science, which is motivated by our analysis below. We compute the eigenvalue decomposition $\boldsymbol{X}_{\rm opt}=\sum_{i=1}^{n}\xi_i\,\boldsymbol{v}_i\boldsymbol{v}_i^*$, with eigenvalues $\xi_1\ge\xi_2\ge\cdots\ge\xi_n\ge 0$ and eigenvectors $\boldsymbol{v}_i=\boldsymbol{v}_i(\boldsymbol{X}_{\rm opt}(\boldsymbol{Y}))$, with $\|\boldsymbol{v}_i\|_2=1$. We then return the estimate

$$\hat{\boldsymbol{x}}^{\rm SDP}(\boldsymbol{Y})=\sqrt{n}\;c^{\rm SDP}(\lambda)\;\boldsymbol{v}_1\big(\boldsymbol{X}_{\rm opt}(\boldsymbol{Y})\big), \qquad [7]$$

with $c^{\rm SDP}(\lambda)$ a certain scaling factor (*SI Appendix*).

Our analytical prediction for $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\rm SDP})$ is plotted as blue solid line in Fig. 1. Dots report the results of numerical simulations with this relaxation for increasing problem dimensions. The asymptotic theory appears to capture these data very well already for $n=200$. For further comparison, alongside the above estimators, we report the asymptotic prediction for $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\rm PCA})$, the mean square error of principal component analysis (PCA). This method simply returns the principal eigenvector of $\boldsymbol{Y}$, suitably rescaled (*SI Appendix*).

Fig. 1 reveals several interesting features.

First, it is apparent that optimal estimation undergoes a phase transition. Bayes optimal estimation achieves nontrivial accuracy as soon as $\lambda>\lambda_c^{\rm Bayes}=1$. The same is achieved by a method as simple as PCA (blue-dashed curve). On the other hand, for $\lambda<1$, no method can achieve $\mathrm{MSE}(\hat{\boldsymbol{x}})<1$ strictly [whereas $\mathrm{MSE}(\hat{\boldsymbol{x}})=1$ is trivial by $\hat{\boldsymbol{x}}=0$].

Second, PCA is suboptimal at large signal strength. PCA can be implemented efficiently but does not exploit the information $x_{0,i}\in\{+1,-1\}$. As a consequence, its estimation error is significantly suboptimal at large λ (*SI Appendix*).

Third, the SDP-based estimator is nearly optimal. The tractable estimator $\hat{\boldsymbol{x}}^{\rm SDP}(\boldsymbol{Y})$ achieves the best of both worlds. Its phase transition coincides with the Bayes optimal one $\lambda_c^{\rm Bayes}=1$, and $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\rm SDP})$ decays exponentially at large λ, staying close to $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\rm Bayes})$ and strictly smaller than $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\rm PCA})$, for $\lambda\ge 1$.

We believe that the above features are generic: as shown in *SI Appendix*, $U(1)$ synchronization confirms this expectation.

Fig. 2 illustrates our results for the community detection problem under the hidden partition model of Eq. **2**. Recall that we encode the ground truth by a vector $\boldsymbol{x}_0\in\{+1,-1\}^n$. In the present context, an estimator is required to return a partition of the vertices of the graph. Formally, it is a function on the space of graphs with n vertices $\mathcal{G}_n$, namely, $\hat{\boldsymbol{x}}:\mathcal{G}_n\to\{+1,-1\}^n$, $G\mapsto\hat{\boldsymbol{x}}(G)$.

[§]Recall that a symmetric matrix $\boldsymbol{A}$ is said to be PSD if all of its eigenvalues are nonnegative.

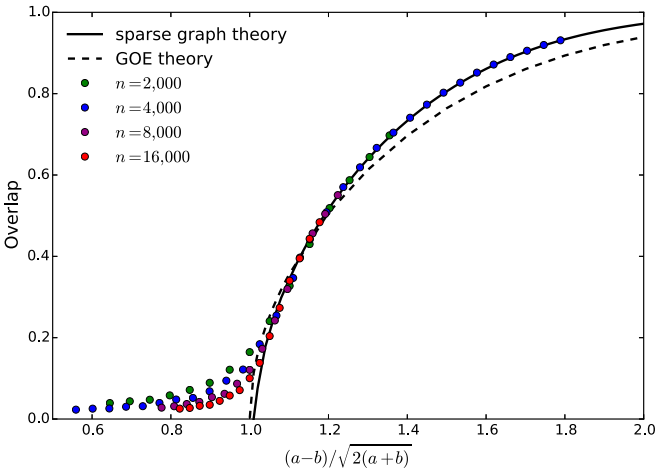


Fig. 2. Community detection under the hidden partition model of Eq. **2**, for average degree $(a+b)/2=5$. Dots indicate performance of the SDP reconstruction method (averaged over 500 realizations). Dashed curve indicates asymptotic analytical prediction for the Gaussian model (which captures the large-degree behavior). Solid curve indicates analytical prediction for the sparse graph case (within the vectorial ansatz; *SI Appendix*).

We will measure the performances of such an estimator through the overlap,

$$\mathrm{Overlap}_n(\hat{\boldsymbol{x}})=\frac{1}{n}\mathbb{E}\{|\langle\hat{\boldsymbol{x}}(G),\boldsymbol{x}_0\rangle|\}, \qquad [8]$$

and its asymptotic $n\to\infty$ limit (for which we omit the subscript). To motivate the SDP relaxation we note that the maximum likelihood estimator partitions V in two sets of equal size to minimize the number of edges across the partition (the minimum bisection problem). Formally,

$$\hat{\boldsymbol{x}}^{\rm ML}(G)\equiv\arg\max_{\boldsymbol{x}\in\{+1,-1\}^n}\left\{\sum_{(i,j)\in E}x_ix_j:\langle\boldsymbol{x},1\rangle=0\right\}, \qquad [9]$$

where $\mathbf{1}=(1,1,\dots,1)$ is the all-ones vector. Once more, this problem is hard to approximate (31), which motivates the following SDP relaxation:

$$\begin{aligned}&\text{maximize } \sum_{(i,j)\in E}X_{ij},\\ &\text{subject to } \boldsymbol{X}\succcurlyeq 0,\quad \boldsymbol{X}1=0,\quad X_{ii}=1\;\forall i\in[n].\end{aligned} \qquad [10]$$

Given an optimizer $\boldsymbol{X}_{\rm opt}=\boldsymbol{X}_{\rm opt}(G)$, we extract a partition of the vertices V as follows. As for the $\mathbb{Z}_2$ synchronization problem, we compute the principal eigenvector $\boldsymbol{v}_1(\boldsymbol{X}_{\rm opt})$. We then partition V according to the sign of $\boldsymbol{v}_1(\boldsymbol{X}_{\rm opt})$. Formally,

$$\hat{\boldsymbol{x}}^{\rm SDP}(G)=\mathrm{sign}\big(\boldsymbol{v}_1\big(\boldsymbol{X}_{\rm opt}(G)\big)\big). \qquad [11]$$

Let us emphasize a few features of Fig. 2:

First, both the GOE theory and the cavity method are accurate. The dashed curve of Fig. 2 reports the analytical prediction within the $\mathbb{Z}_2$ synchronization model, with Gaussian noise (the GOE theory). This can be shown to capture the large degree limit: $d=(a+b)/2\to\infty$, with $\lambda=(a-b)/\sqrt{2(a+b)}$ fixed, and is an excellent approximation already for $d=5$. The continuous curve is our prediction for $d=5$, obtained by applying the cavity method from statistical mechanics to the community detection problem (see next section and *SI Appendix*). This approach describes very accurately the empirical data and the small discrepancy from the GOE theory.

Second, SDP is superior to PCA. A sequence of recent papers (ref. 7 and references therein) demonstrate that classical spectral methods—such as PCA—fail to detect the hidden partition in graphs with bounded average degree. In contrast, Fig. 2 shows that a standard SDP relaxation does not break down in the sparse regime. See refs. 25, 32 for rigorous evidence toward the same conclusion.

Third, SDP is nearly optimal. As proven in ref. 33, no estimator can achieve $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}) \geq \delta > 0$ as $n \to \infty$, if $\lambda = (a-b)/\sqrt{2(a+b)} < 1$. Fig. 2 (and the theory developed in the next section) suggests that SDP has a phase transition threshold. Namely, there exists $\lambda_c^{\mathrm{SDP}} = \lambda_c^{\mathrm{SDP}}(d)$ such that if

$$\lambda = \frac{a-b}{\sqrt{2(a+b)}} \geq \lambda_c^{\mathrm{SDP}}(d=(a+b)/2), \qquad [12]$$

then SDP achieves overlap bounded away from zero: $\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{SDP}}) > 0$. Fig. 2 also suggests $\lambda_c^{\mathrm{SDP}}(5) \approx \lambda_c^{\mathrm{Bayes}} = 1$; that is, SDP is nearly optimal.

Below we will derive an accurate approximation for the critical point $\lambda_c^{\mathrm{SDP}}(d)$. The factor $\lambda_c^{\mathrm{SDP}}(d)$ measures the suboptimality of SDP for graphs of average degree d.

Fig. 3 plots our prediction for the function $\lambda_c^{\mathrm{SDP}}(d)$, together with empirically determined values for this threshold, obtained through Monte Carlo experiments for $d \in \{2,5,10\}$ (red circles). These were obtained by running the SDP estimator on randomly generated graphs with size up to $n = 64{,}000$ (total CPU time was about 10 y). In particular, we obtain $\lambda_c^{\mathrm{SDP}}(d) > 1$ strictly, but the gap $\lambda_c^{\mathrm{SDP}}(d) - 1$ is very small (at most of the order of 2%) for all d. This confirms in a precise quantitative way the conclusion that SDP is nearly optimal for the hidden partition problem.

Simulations results are in broad agreement with our predictions but present small discrepancies (below 0.5%). These discrepancies might be due to the extrapolation from finite-n simulations to $n \to \infty$ or to the inaccuracy of our analytical approximation.

Analytical Results. Our analysis is based on a connection with statistical mechanics. The models arising from this connection are spin models in the so-called "large-N" limit, a topic of intense study across statistical mechanics and quantum field theory (34). Here we exploit this connection to apply nonrigorous but sophisticated tools from the theory of mean field spin glasses (35, 36). The paper (25) provides partial rigorous evidence toward the predictions developed here.

We will first focus on the simpler problem of synchronization under Gaussian noise, treating together the $\mathbb{Z}_2$ and $U(1)$ cases. We will then discuss the new features arising within the sparse hidden partition problem. Most technical derivations are presented in *SI Appendix*. To treat the real ($\mathbb{Z}_2$) and complex [$U(1)$] cases jointly, we will use $\mathbb{F}$ to denote any of the fields of reals or complex numbers, i.e., either $\mathbb{R}$ or $\mathbb{C}$.

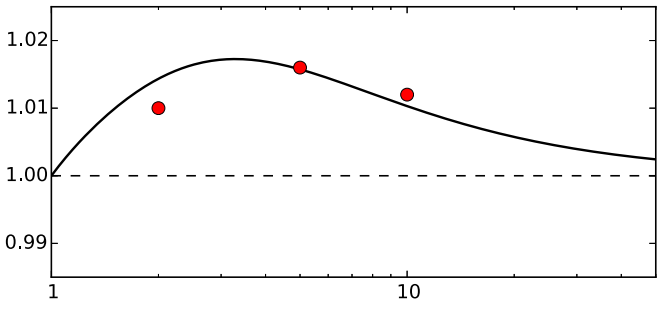


Fig. 3. Phase transition for the SDP estimator: for $\lambda > \lambda_c^{\mathrm{SDP}}(d)$, the SDP estimator has positive correlation with the ground truth; for $\lambda \leq \lambda_c^{\mathrm{SDP}}(d)$ the correlation is vanishing [here $\lambda = (a-b)/\sqrt{2(a+b)}$ and $d = (a+b)/2$]. Solid line indicates prediction $\lambda_c^{\mathrm{SDP}}(d)$ from the cavity method (vectorial ansatz; *SI Appendix*) (compare Eq. **25**). Dashed line indicates ideal phase transition $\lambda = 1$. Red circles indicate numerical estimates of the phase transition location for $d = 2$, 5, and 10.

Gibbs Measures and Vector Spin Models. We start by recalling that a matrix $\boldsymbol{X} \in \mathbb{F}^{n\times n}$ is PSD if and only if it can be written as $\boldsymbol{X} = \boldsymbol{\sigma}\boldsymbol{\sigma}^*$ for some $\boldsymbol{\sigma} \in \mathbb{F}^{n\times m}$. Indeed, without loss of generality, one can take $m = n$, and any $m \geq n$ is equivalent.

Letting $\boldsymbol{\sigma}_1, \dots \boldsymbol{\sigma}_n \in \mathbb{F}^m$ be the rows of $\boldsymbol{\sigma}$, the SDP [**6**] can be rewritten as

$$\begin{aligned} &\text{maximize } \sum_{(i,j)} Y_{ij}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle, \\ &\text{subject to } \boldsymbol{\sigma}_i \in S^{m-1} \ \forall i \in [n], \end{aligned} \qquad [13]$$

with $S^{m-1} = \{\boldsymbol{z} \in \mathbb{F}^m : \|\boldsymbol{z}\|_2 = 1\}$ the unit sphere in m dimensions. The SDP relaxation corresponds to any case $m \geq n$ or, following the physics parlance, $m = \infty$. Note, however, that cases with bounded (small) m are of independent interest. In particular, for $m = 1$ we have $\boldsymbol{\sigma}_i \in \{-1, +1\}$ (for the real case) or $\boldsymbol{\sigma}_i \in U(1) \subset \mathbb{C}$ (for the complex case). Hence, we recover the maximum-likelihood estimator setting $m = 1$. It is also known that (under suitable conditions on $\boldsymbol{Y}$) for $m > \sqrt{2n}$, the problem [**13**] has no local optima except the global ones (37).

A crucial question is how the solution of [**13**] depends on the spin dimensionality m, for $m \ll n$. Denote by $\mathrm{OPT}(\boldsymbol{Y}; m)$ the optimum value when the dimension is m (in particular, $\mathrm{OPT}(\boldsymbol{Y}; m)$ is also the value of [**6**] for $m \geq n$). It was proven in ref. 25 that there exists a constant C independent of m and n such that

$$\left(1 - \frac{C}{m}\right)\mathrm{OPT}(\boldsymbol{Y}; \infty) \leq \mathrm{OPT}(\boldsymbol{Y}; m) \leq \mathrm{OPT}(\boldsymbol{Y}; \infty), \qquad [14]$$

with probability converging to one as $n \to \infty$ (whereby $\boldsymbol{Y}$ is chosen with any of the distributions studied in the present paper). The upper bound in Eq. **14** follows immediately from the definition. The lower bound is a generalization of the celebrated Grothendieck inequality from functional analysis (38).

The above inequalities imply that we can obtain information about the SDP [**6**] in the $n \to \infty$ limit, by taking $m \to \infty$ after $n \to \infty$. This is the asymptotic regime usually studied in physics under the term "large-N limit."

Finally, we can associate to the problem [**13**] a finite-temperature Gibbs measure as follows:

$$p_{\beta,m}(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{Z}\exp\left\{2m\beta \sum_{i<j} \Re\left(Y_{ij}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle\right)\right\} \prod_{i=1}^{n} p_0(\mathrm{d}\boldsymbol{\sigma}_i), \qquad [15]$$

where $p_0(\mathrm{d}\boldsymbol{\sigma}_i)$ is the uniform measure over the m-dimensional sphere S^{m-1} and $\Re(z)$ denotes the real part of z. This allows us to treat in a unified framework all of the estimators introduced above. The optimization problem [**13**] is recovered by taking the limit $\beta \to \infty$ (with maximum likelihood for $m = 1$ and SDP for $m \to \infty$). The Bayes optimal estimator is recovered by setting $m = 1$ and $\beta = \lambda/2$ (in the real case) or $\beta = \lambda$ (in the complex case).

Cavity Method: $\mathbb{Z}_2$ and $U(1)$ Synchronization. The cavity method from spin-glass theory can be used to analyze the asymptotic structure of the Gibbs measure [**15**] as $n \to \infty$. Below we will state the predictions of our approach for the SDP estimator $\hat{\boldsymbol{x}}^{\mathrm{SDP}}$.

Here we list the main steps of our analysis for the expert reader, deferring a complete derivation to the *SI Appendix*: (*i*) We use the cavity method to derive the replica symmetric predictions for the model (15) in the limit $n \to \infty$. (*ii*) By setting $m = 1$, $\beta = \lambda/2$ (in the real case), or $\beta = \lambda$ (in the complex case) we obtain the Bayes optimal error $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{Bayes}})$: on the basis of ref. 39, we

expect the replica symmetric assumption to hold and these predictions to be exact. (See also ref. 40 for related work.) (*iii*) By setting $m=1$ and $\beta\to\infty$ we obtain a prediction for the error of maximum likelihood estimation $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{ML}})$. Although this prediction is not expected to be exact (because of replica symmetry breaking), it should be nevertheless rather accurate, especially for large λ. (*iv*) By setting $m\to\infty$ and $\beta\to\infty$, we obtain the SDP estimation error $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{SDP}})$, which is our main object of interest. Notice that the inversion of limits $m\to\infty$ and $n\to\infty$ is justified (at the level of objective value) by Grothendieck inequality. Further, because the $m=\infty$ case is equivalent to a convex program, we expect the replica symmetric prediction to be exact in this case.

The properties of the SDP estimator are given in terms of the solution of a set of three nonlinear equations for the three scalar parameters μ, q, and $b\in\mathbb{R}$ that we state next. Let $Z\sim\mathsf{N}(0,1)$ (in the real case) or $Z\sim\mathsf{CN}(0,1)$ (in the complex case). Define $\rho=\rho(Z;\mu,q,r)$ as the only nonnegative solution of the following equation in $(0,\infty)$:

$$1=\frac{|\mu+\sqrt{q}Z|^2}{(\rho+r)^2}+\frac{1-q}{\rho^2}. \qquad [16]$$

Then μ, q, and r satisfy

$$\mu=\lambda\mathbb{E}\left\{\frac{\mu+\sqrt{q}\,\Re(Z)}{\rho+r}\right\},\quad q=\mathbb{E}\left\{\frac{|\mu+\sqrt{q}\,Z|^2}{(\rho+r)^2}\right\}, \qquad [17]$$

$$r=\mathbb{E}\left\{\frac{1}{\rho}-\frac{\mu}{\sqrt{q}}\frac{\Re(Z)}{\rho+r}-\frac{|Z|^2}{\rho+r}\right\}. \qquad [18]$$

These equations can be solved by iteration, after approximating the expectations on the right-hand side numerically. The properties of the SDP estimator can be derived from this solution. Concretely, we have

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{SDP}})=1-\frac{\mu(\lambda)^2}{\lambda^2 q(\lambda)}. \qquad [19]$$

The corresponding curve is reported in Fig. 1 for the real case $\mathfrak{G}=\mathbb{Z}_2$. We can also obtain the asymptotic overlap from the solution of these equations. The cavity prediction is

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{SDP}})=1-2\Phi\left(-\frac{\mu(\lambda)}{\sqrt{q(\lambda)}}\right). \qquad [20]$$

The corresponding curve is plotted in Fig. 2.

More generally, for any dimension m and inverse temperature β, we obtain equations that are analogous to Eqs. **17** and **18**. The parameters μ,q, and b characterize the asymptotic structure of the probability measure $p_{\beta,m}(\mathrm{d}\boldsymbol{\sigma})$ defined in Eq. **15**, as follows. We assume, for simplicity $\boldsymbol{x}_0=(+1,\dots,+1)$. Define the following probability measure on unit sphere S^{m-1}, parametrized by $\boldsymbol{\xi}\in\mathbb{R}^m$, $r\in\mathbb{R}$:

$$\nu_{\boldsymbol{\xi},r}(\mathrm{d}\boldsymbol{\sigma})=\frac{1}{z(\boldsymbol{\xi},r)}\exp\left\{2\beta\, m\Re\langle\boldsymbol{\xi},\boldsymbol{\sigma}\rangle-\beta\, mr|\sigma_1|^2\right\}p_0(\mathrm{d}\boldsymbol{\sigma}). \qquad [21]$$

For ν a probability measure on S^{m-1} and R an orthogonal (or unitary) matrix, let ν^R be the measure obtained by[¶] rotating ν. Finally, let $p^{(m,\beta)}_{i(1),\dots,i(k)}$ denote the joint distribution of $\boldsymbol{\sigma}_{i(1)},\cdots,\boldsymbol{\sigma}_{i(k)}$ under $p_{m,\beta}$. Then, for any fixed k, and any sequence of k-uples $(i(1),\dots,i(k))_n\in[n]$, we have

$$p^{(m,\beta)}_{i(1),\dots,i(k)}\Rightarrow\int\nu^R_{\boldsymbol{\xi}_1,r}(\,\cdot\,)\times\cdots\times\nu^R_{\boldsymbol{\xi}_k,r}(\,\cdot\,)\mathrm{d}R. \qquad [22]$$

Here $\mathrm{d}R$ denotes the uniform (Haar) measure on the orthogonal group, $\Rightarrow$ denotes convergence in distribution (note that $p^{(m,\beta)}_{i(1),\dots,i(k)}$ is a random variable), and $\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_k\sim_{iid}\mathsf{N}(\mu\boldsymbol{e}_1,\boldsymbol{Q})$ with $\boldsymbol{Q}=\mathrm{diag}(q,q_0\dots,q_0)$, $q_0=(1-q)/(m-1)$.

Cavity Method: Community Detection in Sparse Graphs. We next consider the hidden partition model, defined by Eq. **2**. As above, we denote by $d=(a+b)/2$ the asymptotic average degree of the graph G and by $\lambda=(a-b)/\sqrt{2(a+b)}$ the signal-to-noise ratio. As illustrated by Fig. 2 (and further simulations presented in *SI Appendix*), $\mathbb{Z}_2$ synchronization appears to be a very accurate approximation for the hidden partition model already at moderate d.

The main change with respect to the dense case is that the phase transition at $\lambda=1$, is slightly shifted, as per Eq. **12**. Namely, SDP can detect the hidden partition with high probability if and only if $\lambda\ge\lambda_c^{\mathrm{SDP}}(d)$, for some $\lambda_c^{\mathrm{SDP}}(d)>1$.

Our prediction for the curve $\lambda_c^{\mathrm{SDP}}(d)$ will be denoted by $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$ and is plotted in Fig. 3. It is obtained by finding an approximate solution of the RS cavity equations, within a scheme that we name "vectorial ansatz" (see *SI Appendix* for details). We see that $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$ approaches very quickly the ideal value $\lambda=1$ for $d\to\infty$. Indeed, our prediction implies $\tilde{\lambda}_c^{\mathrm{SDP}}(d)=1+1/(8d)+O(d^{-2})$. Also, $\tilde{\lambda}_c^{\mathrm{SDP}}(d)\to1$ as $d\to1$. This is to be expected because the constraints $a\ge b\ge0$ imply $(a-b)/2\le d$, with $b=0$ at $(a-b)/2=d$. Hence, the problem becomes trivial at $(a-b)/2=d$: it is sufficient to identify the connected components in G, whence $\lambda_c^{\mathrm{SDP}}(d)\le\sqrt{d}$.

More interestingly, $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$ admits a characterization in terms of a distributional recursion, which can be evaluated numerically and is plotted as a continuous line in Fig. 3. Surprisingly, the SDP detection threshold appears to be suboptimal at most by 2%. To state this characterization, consider first the recursive distributional equation (RDE)

$$\mathsf{c}\overset{\mathrm{d}}{=}\sum_{i=1}^{L}\frac{\mathsf{c}_i}{1+\mathsf{c}_i}. \qquad [23]$$

Here $\overset{\mathrm{d}}{=}$ denotes equality in distribution, $L\sim\mathrm{Poisson}(d)$, and $\mathsf{c}_1,\dots,\mathsf{c}_L$ are independent and identically distributed (i.i.d.) copies of c. This has to be read as an equation for the law of the random variable c (see, e.g., ref. 41 for further background on RDEs). We are interested in a specific solution of this equation, constructed as follows. Set $c^0=\infty$ almost surely, and for $\ell\ge0$, let $\mathsf{c}^{\ell+1}\overset{\mathrm{d}}{=}\sum_{i=1}^{L}\mathsf{c}_i^\ell/(1+\mathsf{c}_i^\ell)$. It is proved in ref. 42 that the resulting sequence converges in distribution to a solution of Eq. **23**: $\mathsf{c}^\ell\overset{\mathrm{d}}{\Rightarrow}\mathsf{c}_*$.

The quantity c_* has a useful interpretation. Consider a (rooted) Poisson Galton–Watson tree with branching number d, and imagine each edge to be a conductor with unit conductance. Then c_* is the total conductance between the root and the boundary of the tree at infinity. In particular, $\mathsf{c}_*=0$ almost surely for $d\le1$, and $\mathsf{c}_*>0$ with positive probability if $d>1$ (see ref. 42 and *SI Appendix*).

Next consider the distributional recursion

$$(\mathsf{c}^{\ell+1};h^{\ell+1})\overset{\mathrm{d}}{=}\left(\sum_{i=1}^{L_++L_-}\frac{\mathsf{c}_i^\ell}{1+\mathsf{c}_i^\ell};\sum_{i=1}^{L_++L_-}\frac{s_ih_i^\ell}{\sqrt{1+\mathsf{c}_i^\ell}}\right), \qquad [24]$$

where $s_1,\dots,s_{L_+}=+1$, $s_{L_++1},\dots,s_{L_++L_-}=-1$, and we use initialization $(\mathsf{c}^0,h^0)=(+\infty,1)$. This recursion determines sequentially

[¶]Formally, $\nu^R(\sigma\in A)\equiv\nu(R^{-1}\sigma\in A)$.

the distribution of $(c^{\ell+1}, h^{\ell+1})$ from the distribution of (c^{ℓ}, h^{ℓ}). Here $L_+ \sim \text{Poisson}((d+\lambda)/2)$, $L_- \sim \text{Poisson}((d-\lambda)/2)$, and $(c_1^{\ell}, h_1^{\ell}), \dots, (c_L^{\ell}, h_L^{\ell})$ are i.i.d. copies of (c^{ℓ}, h^{ℓ}), independent of L_+, L_-. Notice that because $L_+ + L_- \sim \text{Poisson}(d)$, we have $c^{\ell} \Rightarrow c_*$. The threshold $\tilde{\lambda}_c^{\text{SDP}}(d)$ is defined as the smallest λ such that the h^t diverges exponentially:

$$\tilde{\lambda}_c^{\text{SDP}}(d) \equiv \inf\left\{\lambda \in \left[0, \sqrt{d}\right] : \lim\inf_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left(|h^t|^2\right) > 0\right\}. \quad [25]$$

This value can be computed numerically, for instance, by sampling the recursion [24]. The results of such an evaluation are plotted as a continuous line in Fig. 3.

Final Algorithmic Considerations

We have shown that ideas from statistical mechanics can be used to precisely locate phase transitions in SDP relaxations for high-dimensional statistical problems. In the problems investigated here, we find that SDP relaxations have optimal thresholds [in $\mathbb{Z}_2$ and $U(1)$ synchronization] or nearly optimal thresholds (in community detection under the hidden partition model). Here near-optimality is to be interpreted in a precise quantitative sense: SDP's threshold is suboptimal—at most—by a 2% factor. As such, SDPs provide a very useful tool for designing computationally efficient algorithms that are also statistically efficient.

Let us emphasize that other polynomial–time algorithms can be used for the specific problems studied here. In the synchronization problem, naive PCA achieves the optimal threshold $\lambda = 1$. In the community detection problem, several authors recently developed ingenious spectral algorithms that achieve the information theoretically optimal threshold $(a-b)/\sqrt{2(a+b)} = 1$ (see, e.g., refs. 7, 23, 24, 43, 44).

However, SDP relaxations have the important feature of being robust to model misspecifications (see also refs. 30, 45 for independent investigations of robustness issues). To illustrate this point, we perturbed the hidden partition model as follows. For a perturbation level $\alpha \in [0,1]$, we draw $n\alpha$ vertices $i_1, \dots, i_{n\alpha}$ uniformly at random in G. For each such vertex i_ℓ we connect by edges all of the neighbors of i_ℓ. In our case, this results in adding $O(nd^2\alpha)$ edges.

In *SI Appendix*, we compare the behavior of SDP and the Bethe Hessian algorithm of ref. 44 for this perturbed model: although SDP appears to be rather insensitive to the perturbation, the performance of Bethe Hessian are severely degraded by it. We expect a similar fragility in other spectral algorithms.

ACKNOWLEDGMENTS. A.M. was partially supported by National Science Foundation Grants CCF-1319979 and DMS-1106627 and Air Force Office of Scientific Research Grant FA9550-13-1-0036. A.J. was partially supported by the Center for Science of Information (CSoI) fellowship.

1. Gauss CF (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss* (Friedrich Perthes und I. H. Besser, Hamburg, Germany).
2. Ben-Dor A, Shamir R, Yakhini Z (1999) Clustering gene expression patterns. *J Comput Biol* 6(3-4):281–297.
3. Plaza A, et al. (2009) Recent advances in techniques for hyperspectral image processing. *Remote Sens Environ* 113(1):S110–S122.
4. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.
5. Von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17(4):395–416.
6. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826.
7. Krzakala F, et al. (2013) Spectral redemption in clustering sparse networks. *Proc Natl Acad Sci USA* 110(52):20935–20940.
8. Wasserman L (2000) Bayesian model selection and model averaging. *J Math Psychol* 44(1):92–107.
9. Tibshirani R (1996) Regression shrinkage and selection with the Lasso. *J R Stat Soc, B* 58(1):267–288.
10. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM J Sci Comput* 20(1):33–61.
11. Candès EJ, Tao T (2010) The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans Information Theory* 56(5):2053–2080.
12. Donoho DL, Tanner J (2005) Neighborliness of randomly projected simplices in high dimensions. *Proc Natl Acad Sci USA* 102(27):9452–9457.
13. Candés EJ, Tao T (2007) The Dantzig selector: Statistical estimation when p is much larger than n. *Ann Stat* 35(6):2313–2351.
14. Donoho DL, Maleki A, Montanari A (2009) Message-passing algorithms for compressed sensing. *Proc Natl Acad Sci USA* 106(45):18914–18919.
15. Singer A, Shkolnisky Y (2011) Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM J Imaging Sci* 4(2):543–572.
16. Cucuringu M (2015) Synchronization over F_2 and community detection in signed multiplex networks with constraints. *J Complex Networks* cnu050.
17. Abbe E, Bandeira AS, Bracher A, Singer A (2014) Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Trans Network Sci Eng* 1(1):10–22.
18. Singer A (2011) Angular synchronization by eigenvectors and semidefinite programming. *Appl Comput Harmon Anal* 30(1):20–36.
19. Bandeira AS, Boumal N, Singer A (2014) Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. arXiv:1411.3272.
20. Candes EJ, Eldar YC, Strohmer T, Voroninski V (2015) Phase retrieval via matrix completion. *SIAM Rev* 57(2):225–251.
21. Waldspurger I, d'Aspremont A, Mallat S (2015) Phase recovery, maxcut and complex semidefinite programming. *Math Program* 149(1-2):47–81.
22. Alexeev B, Bandeira AS, Fickus M, Mixon DG (2014) Phase retrieval with polarization. *SIAM J Imaging Sci* 7(1):35–66.
23. Massoulié L (2014) Community detection thresholds and the weak Ramanujan property. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York), pp 694–703.
24. Mossel E, Neeman J, Sly A (2013) A proof of the block model threshold conjecture. arXiv:1311.4115.
25. Montanari A, Sen S (2016) Semidefinite programs on sparse random graphs and their application to community detection. *Proceedings of the 48th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York).
26. Bandeira AS (2015) Random Laplacian matrices and convex relaxations. arXiv: 1504.03987.
27. Holland PW, Laskey K, Leinhardt S (1983) Stochastic blockmodels: First steps. *Soc Networks* 5(2):109–137.
28. Arora S, Berger E, Hazan E, Kindler G, Safra M (2005) On non-approximability for quadratic programs. *46th Annual IEEE Symposium on Foundations of Computer Science, 2005. FOCS 2005* (Inst of Electr and Electron Eng, Washington, DC), pp 206–215.
29. Nesterov Y (1998) Semidefinite relaxation and nonconvex quadratic optimization. *Optim Methods Softw* 9(1-3):141–160.
30. Feige U, Kilian J (2001) Heuristics for semirandom graph problems. *J Comput Syst Sci* 63(4):639–671.
31. Khot S (2006) Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J Comput* 36(4):1025–1071.
32. Guédon O, Vershynin R (2014) Community detection in sparse networks via grothendieck's inequality. arXiv:1411.4686.
33. Mossel E, Neeman J, Sly A (2012) Stochastic block models and reconstruction. arXiv: 1202.1499.
34. Brézin E, Wadia SR (1993) *The Large N Expansion in Quantum Field Theory and Statistical Physics: From Spin Systems to 2-Dimensional Gravity* (World Scientific, Singapore).
35. Mézard M, Montanari A (2009) *Information, Physics, and Computation* (Oxford University Press, Oxford, United Kingdom).
36. Mézard M, Parisi G, Virasoro MA (1987) *Spin Glass Theory and Beyond* (World Scientific, Singapore).
37. Burer S, Monteiro RDC (2003) A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math Program* 95(2):329–357.
38. Khot S, Naor A (2012) Grothendieck-type inequalities in combinatorial optimization. *Commun Pure Appl Math* 65(7):992–1035.
39. Deshpande Y, Abbe E, Montanari A (2015) Asymptotic mutual information for the two-groups stochastic block model. arXiv:1507.08685.
40. Lesieur T, Krzakala F, Zdeborová L (2015) . MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. arXiv:1507.03857.
41. Aldous DJ, Bandyopadhyay A (2005) A survey of max-type recursive distributional equations. *Ann Appl Probab* 15(2):1047–1110.
42. Lyons R, Pemantle R, Peres Y (1997) Unsolved problems concerning random walks on trees. *Classical and Modern Branching Processes* (Springer, New York), pp 223–237.
43. Decelle A, Krzakala F, Moore C, Zdeborová L (2011) Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys Rev E Stat Nonlin Soft Matter Phys* 84(6 Pt 2):066106.
44. Saade A, Krzakala F, Zdeborová L (2014) Spectral clustering of graphs with the Bethe Hessian. *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, La Jolla, CA), pp 406–414.
45. Moitra A, Perry W, Wein AS (2015) How robust are reconstruction thresholds for community detection? *Proceedings of the 48th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York).

Supplementary Information for: Phase Transitions in Semidefinite Relaxations

Adel Javanmard[*], Andrea Montanari[†] and Federico Ricci-Tersenghi[‡]

February 12, 2016

Contents

[*]USC Marshall School of Business, University of Southern California
[†]Department of Electrical Engineering and Department of Statistics, Stanford University
[‡]Dipartimento di Fisica, Universitá di Roma, La Sapienza

Notice

Most of the derivations in this documents are based on non-rigorous method from statistical physics. All the results that are rigorously proved will be stated as lemmas, propositions, and so on.

1 Notations

1.1 General notations

We will often treat $\mathbb{Z}_2$ and $U(1)$ synchronization simultaneously. Throughout $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ depending on whether we are treating the real case ($\mathbb{Z}_2$ synchronization) or the complex case ($U(1)$ synchronization).

We let S^{m-1} denotes the radius one sphere in $\mathbb{R}^m$ or $\mathbb{C}^m$ depending on the context. Namely $S^{m-1} = \{\boldsymbol{z} \in \mathbb{F}^m : \|\boldsymbol{z}\|_2 = 1\}$. In particular $S^0 = \{+1, -1\}$ in the real case, and $S^0 = \{z \in \mathbb{C} : |z| = 1\}$ in the complex case.

Some of our formulae depends upon the domain that we are considering (real or complex). In order to write them in a compact form, we introduce the notation $s_{\mathfrak{G}} = 1$ for $\mathfrak{G} = \mathbb{Z}_2$, and $s_{\mathfrak{G}} = 2$ for $\mathfrak{G} = U(1)$.

We write $X \sim \text{Poisson}(a)$ to indicate that X is a Poisson random variable with mean a. A Gaussian random vector $\boldsymbol{z}$ with mean $\boldsymbol{a} = \mathbb{E}(\boldsymbol{z})$ and covariance $\boldsymbol{C} = \mathbb{E}((\boldsymbol{z} - \boldsymbol{a})(\boldsymbol{z} - \boldsymbol{a})^*)$ is denoted by $\boldsymbol{z} \sim \mathsf{N}(\boldsymbol{a}, \boldsymbol{C})$. Note that in the complex case, this means that $\boldsymbol{C}$ is Hermitian and $\mathbb{E}((\boldsymbol{z}-\boldsymbol{a})(\boldsymbol{z}-\boldsymbol{a})^{\mathsf{T}}) = 0$. Occasionally, we will write $\boldsymbol{z} \sim \mathsf{CN}(\boldsymbol{a}, \mathcal{C})$ for complex Gaussians, whenever it is useful to emphasize that $\boldsymbol{z}$ is complex.

The standard Gaussian density is denoted by $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$, and the Gaussian distribution by $\Phi(x) = \int_{-\infty}^{x} \phi(t)\,\mathrm{d}t$.

Given two un-normalized measures p and q on the same space, we write $p(s) \cong q(s)$ if they are equal up to an overall normalization constant. We use $\doteq$ to denote equality up to subexponential factors, i.e. $f(n) \doteq g(n)$ if $\lim_{n\to\infty} n^{-1} \log[f(n)/g(n)] = 0$.

1.2 Estimation metrics

We recall the definition of some estimation metrics used in the main text. For the sake of uniformity, we consider estimators $\hat{\boldsymbol{x}} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$.

It is convenient to define a *scaled* MSE, with scaling factor $c \in \mathbb{F}$:

$$\mathrm{MSE}_n(\hat{\boldsymbol{x}}; c) \equiv \frac{1}{n}\mathbb{E}\big\{ \min_{s\in S^0} \big\|\boldsymbol{x_0} - s\, c\, \hat{\boldsymbol{x}}(\boldsymbol{Y})\big\|_2^2\big\}\,. \tag{1}$$

We also define the overlap as follows in the real case

$$\mathrm{Overlap}_n(\hat{x}) \equiv \frac{1}{n}\mathbb{E}\big\{|\langle \mathrm{sign}(\hat{\boldsymbol{x}}(\boldsymbol{Y})), \boldsymbol{x_0}\rangle|\big\}\,. \tag{2}$$

In the complex case, we replace $\mathrm{sign}(z)$ by $z/|z|$ (defined to be 0 at $z = 0$):

$$\mathrm{Overlap}_n(\hat{x}) \equiv \frac{1}{n}\mathbb{E}\left\{\left|\sum_{i=1}^{n} \frac{\hat{x}_i(\boldsymbol{Y})}{|\hat{x}_i(\boldsymbol{Y})|}\, x_{0,i}\right|\right\}\,. \tag{3}$$

This formula applies to the real case as well. (Note that, in the main text, we defined the overlap only for estimators taking values in $\{+1, -1\}^n$, in the real case. Throughout these notes, we generalize that definition for the sake of uniformity.)

We omit the subscript n to refer to the $n \to \infty$ limit of these quantities.

2 Preliminary facts

2.1 Some estimation identities

Lemma 2.1. *Let $p_0(\,\cdot\,)$ be a probability measure on the real line $\mathbb{R}$, symmetric around 0 (i.e. $p_0((a, b)) = p_0((-b, -a))$ for any interval (a, b)). For $\gamma \geq 0$, define $f : \mathbb{R} \to \mathbb{R}$ as*

$$f(y;\gamma) \equiv \frac{\int \sigma\, e^{\sqrt{\gamma} y\sigma - \frac{1}{2}\gamma\sigma^2}\, p_0(\mathrm{d}\sigma)}{\int e^{\sqrt{\gamma} y\sigma - \frac{1}{2}\gamma\sigma^2}\, p_0(\mathrm{d}\sigma)}\,. \tag{4}$$

Then we have the identity

$$\mathbb{E}\{|\sigma|\, f(\sqrt{\gamma}|\sigma| + Z; \gamma)\} = \mathbb{E}\big\{f(\sqrt{\gamma}|\sigma| + Z; \gamma)^2\big\}\,. \tag{5}$$

where the expectation is with respect to the independent random variables $Z \sim \mathsf{N}(0, 1)$, and $\sigma \sim p_0(\,\cdot\,)$.

Analogously, let $p_0(\,\cdot\,)$ be a probability measure on $\mathbb{C}$, symmetric under rotations (i.e. $p_0(e^{i\theta}R) = p_0(R)$, $p_0(R) = p_0(R^*)$ for any Borel set $R \subseteq \mathbb{C}$ and any $\theta \in (0, 2\pi]$). For $\gamma \geq 0$, define $f : \mathbb{C} \to \mathbb{R}$ as

$$f(y;\gamma) \equiv \frac{\int \sigma\, e^{2\sqrt{\gamma}\Re(y^*\sigma) - \gamma|\sigma|^2}\, p_0(\mathrm{d}\sigma)}{\int e^{2\sqrt{\gamma}\Re(y^*\sigma) - \gamma|\sigma|^2}\, p_0(\mathrm{d}\sigma)}\,. \tag{6}$$

Then we have the identity (with $Z \sim \mathsf{CN}(0, 1)$ a complex normal)

$$\mathbb{E}\{|\sigma|\, f(\sqrt{\gamma}|\sigma| + Z; \gamma)\} = \mathbb{E}\big\{|f(\sqrt{\gamma}|\sigma| + Z; \gamma)|^2\big\}\,. \tag{7}$$

Proof. Consider, to be definite, the real case, and define the observation model

$$Y = \sqrt{\gamma}\,\sigma + Z\,, \tag{8}$$

where $\sigma \sim p_0(\,\cdot\,)$ independent of the noise $Z \sim \mathsf{N}(0,1)$. Then a straightforward calculation shows that

$$f(y;\gamma) = \mathbb{E}\{\sigma | Y = y\}\,. \tag{9}$$

Then, by the tower property of conditional expectation $\mathbb{E}\{\sigma\, f(Y;\gamma)\} = \mathbb{E}\{f(Y;\gamma)^2\}$ or, equivalently

$$\mathbb{E}\{\sigma\, f(\sqrt{\gamma}\,\sigma + Z;\gamma)\} = \mathbb{E}\{f(\sqrt{\gamma}\,\sigma + Z;\gamma)^2\}\,. \tag{10}$$

The identity (5) follows by exploiting the symmetry of p_0, which implies $f(-y;\gamma) = -f(y;\gamma)$.

The proof follows a similar argument in the complex case. □

We apply the above lemma to specific cases that will be of interest to us. Below, $\mathrm{I}_k(z)$ denotes the modified Bessel function of the second kind. Explicitly, for k integer, we have the integral representation

$$\mathrm{I}_k(z) = \frac{1}{2\pi}\int_0^{2\pi} e^{z\cos\theta}\cos(k\theta)\,\mathrm{d}\theta\,. \tag{11}$$

Corollary 2.2. *For any $\gamma \geq 0$, we have the identities*

$$\mathbb{E}\big\{\tanh(\gamma + \sqrt{\gamma}\,Z)\big\} = \mathbb{E}\big\{\tanh(\gamma + \sqrt{\gamma}\,Z)^2\big\}\,, \tag{12}$$

$$\mathbb{E}\left\{\frac{\gamma + \sqrt{\gamma}\,Z}{|\gamma + \sqrt{\gamma}\,Z|}\,\frac{\mathrm{I}_1(2|\gamma + \sqrt{\gamma}\,Z|)}{\mathrm{I}_0(2|\gamma + \sqrt{\gamma}\,Z|)}\right\} = \mathbb{E}\left\{\frac{\mathrm{I}_1(2|\gamma + \sqrt{\gamma}\,Z|)^2}{\mathrm{I}_0(2|\gamma + \sqrt{\gamma}\,Z|)^2}\right\}\,, \tag{13}$$

where the expectation is with respect to $Z \sim \mathsf{N}(0,1)$ (first line) or $Z \sim \mathsf{CN}(0,1)$ (second line).

Proof. These follows from Lemma 2.1. For the first line we apply the real case (5) with $p_0 = (1/2)\delta_{+1} + (1/2)\delta_{-1}$, whence

$$f(y;\gamma) = \tanh(\sqrt{\gamma}\,y)\,. \tag{14}$$

For the second line we apply the complex case (6) with p_0 the uniform measure over the unit circle. Consider the change of variables $y = |y|e^{j\phi}$ and $\sigma = e^{j(\phi+\theta)}$. Computing the curve integral, we have

$$\begin{aligned} f(y;\gamma) &= \frac{\int_0^{2\pi} e^{j(\phi+\theta)} e^{2\sqrt{\gamma}|y|\cos(\theta)}\mathrm{d}\theta}{\int_0^{2\pi} e^{2\sqrt{\gamma}y\cos(\theta)}\mathrm{d}\theta} \\ &= \frac{e^{j\phi}\int_0^{2\pi} e^{2\sqrt{\gamma}|y|\cos(\theta)}\cos(\theta)\mathrm{d}\theta}{\int_0^{2\pi} e^{2\sqrt{\gamma}|y|\cos(\theta)}\mathrm{d}\theta} = \frac{y}{|y|}\frac{\mathrm{I}_1(2\sqrt{\gamma}|y|)}{\mathrm{I}_0(2\sqrt{\gamma}|y|)}\,, \end{aligned} \tag{15}$$

where in the second equality we used the fact that $\int_0^{2\pi} e^{2\sqrt{\gamma}|y|\cos(\theta)}\sin(\theta)\mathrm{d}\theta = 0$. □

3 Analytical results for $\mathbb{Z}_2$ and and $U(1)$ synchronization

As explained in the main text, our approach is based on defineing a probability measure in a space that suitably 'relaxes' the posterior distribution of the unknown signal. In the present case, we are interested in the following probability measure over $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)$, where $\boldsymbol{\sigma}_i \in S^{m-1}$:

$$p_{\beta,m}(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{Z_{n,\boldsymbol{Y}}(\beta, m)} \exp\left\{2m\beta \sum_{i<j} \Re(Y_{ij}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle)\right\} \prod_{i=1}^{n} p_0(\mathrm{d}\boldsymbol{\sigma}_i)\,. \tag{16}$$

Here $p_0(\mathrm{d}\boldsymbol{\sigma}_i)$ is the uniform measure over $\boldsymbol{\sigma}_i \in S^{m-1}$.

We define a general m, β estimator as follows.

1. In order to break the $\mathcal{O}(m)$ symmetry, we add a term $\beta \sum_{i=1}^{n} \langle \boldsymbol{h}, \boldsymbol{\sigma}\rangle$ in the exponent of Eq. (16), for $\boldsymbol{h}$ an arbitrary small vector. It is understood throughout that $\|\boldsymbol{h}\|_2 \to 0$ after $n \to \infty$.

   As is customary in statistical physics, we will not explicitly carry out calculations with the perturbation $\boldsymbol{h} \neq 0$, but only using this device to select the relevant solution at $\boldsymbol{h} = 0$.

2. We compute the expectation

$$\boldsymbol{z}(\boldsymbol{Y}; \beta, m) = \int \boldsymbol{\sigma}\; p_{\beta,m}(\mathrm{d}\boldsymbol{\sigma})\,. \tag{17}$$

   Note that for $\beta \to \infty$ this amounts to maximizing the exponent term in equation (16).

3. Compute the empirical covariance

$$\widehat{\boldsymbol{Q}} \equiv \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^*\,. \tag{18}$$

   Let $\widehat{\boldsymbol{u}}$ be its principal eigenvector.

4. Return $\hat{\boldsymbol{x}}^{(\beta,m)}$

$$\hat{x}_i^{(\beta,m)} = c\,\langle \widehat{\boldsymbol{u}}, \boldsymbol{z}_i\rangle\,, \tag{19}$$

   where $c = c(\lambda)$ is the optimal scaling predicted by the asymptotic theory.

3.1 Derivation of the Gibbs measure

The Gibbs measure (16) encodes several estimators of interests. Here we briefly describe this connections.

Bayes-optimal estimators. As mentioned in the main text, this is obtained by setting $m = 1$ and $\beta = \lambda/2$ (in the real case) or $\beta = \lambda$ (in the complex case). To see this, recall our observation model (for $i < j$)

$$Y_{ij} = \frac{\lambda}{n} x_{0,i} x_{0,j}^* + W_{ij}\,, \tag{20}$$

with $x_{0,i} \in S^0$ and $W_{ij} \sim \mathsf{N}(0, 1/n)$. Hence, by an application of Bayes formula, the conditional density of $\boldsymbol{x_0} = \boldsymbol{\sigma} \in \mathbb{F}^n$ given $\boldsymbol{Y}$ is becomes

$$p(\mathrm{d}\boldsymbol{\sigma}|\boldsymbol{Y}) = \frac{1}{Z'_{\boldsymbol{Y}}} \exp\Big\{ -\frac{n s_{\mathfrak{G}}}{2} \sum_{i<j} \Big| Y_{ij} - \frac{\lambda}{n} \sigma_i \sigma_j^* \Big|^2 \Big\} \prod_{i=1}^n p_0(\mathrm{d}\sigma_i)\,, \tag{21}$$

where we recall that $s_{\mathfrak{G}} = 1$ for $\mathfrak{G} = \mathbb{Z}_2$ (real case), and $s_{\mathfrak{G}} = 2$ for $\mathfrak{G} = U(1)$ (complex case). Further, $p_0(\mathrm{d}\sigma)$ denotes the uniform measure over $S^0 \in \mathbb{F}$ (in particular, this is the uniform measure over $\{+1, -1\}$ for $\mathfrak{G} = \mathbb{Z}_2$). Expanding the square and re-absorbing terms independent of $\boldsymbol{\sigma}$ in the normalization constant, we get

$$p(\mathrm{d}\boldsymbol{\sigma}|\boldsymbol{Y}) = \frac{1}{Z_{\boldsymbol{Y}}} \exp\Big\{ \lambda s_{\mathfrak{G}} \sum_{i<j} \Re(Y_{ij} \sigma_i^* \sigma_j) \Big\} \prod_{i=1}^n p_0(\mathrm{d}\sigma_i)\,. \tag{22}$$

As claimed, this coincides with Eq. (16) if we set $\beta = \lambda/2$ (in the real case) or $\beta = \lambda$ (in the complex case).

Maximum-likelihood and SDP estimators. By letting $\beta \to \infty$ in Eq. (16), we obtain that $p_{\beta,m}$ concentrates on the maximizers of the problem

$$\begin{aligned} \text{maximize} \quad & \sum_{i<j} Y_{ij} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle\,, \\ \text{subject to} \quad & \boldsymbol{\sigma}_i \in S^{m-1} \ \ \forall i \in [n]\,, \end{aligned} \tag{23}$$

In the case $m \geq n$ we recover the SDP relaxation. In the case $m = 1$, this is equivalent to the maximum likelihood problem

$$\begin{aligned} \text{maximize} \quad & \sum_{i<j} \big| Y_{ij} - \sigma_i \sigma_j^* \big|^2\,, \\ \text{subject to} \quad & \sigma_i \in S^0 \ \ \forall i \in [n]\,. \end{aligned} \tag{24}$$

3.2 Cavity derivation for $\mathbb{Z}_2$ and $U(1)$ synchronization

In this section we use the cavity method to derive the asymptotic properties of the measure (16).

3.2.1 General m and β

In the replica-symmetric cavity method, we consider adding a single variable $\boldsymbol{\sigma}_0$ to a problem with n variables $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n$. We compute the marginal distribution of $\boldsymbol{\sigma}_0$ in the system with $n+1$ variables, to be denoted by $\nu_0^{n+1}(\mathrm{d}\boldsymbol{\sigma}_0)$. This is expressed in terms of the marginals of the other variables in the system with n variables $\nu_1^n(\mathrm{d}\boldsymbol{\sigma}_0), \dots \nu_n^n(\mathrm{d}\boldsymbol{\sigma}_1)$. We will finally impose the consistency condition that ν_0^{n+1} is distributed as any of $\nu_1^n, \dots \nu_n^n$ in the $n \to \infty$ limit.

Assuming that $\boldsymbol{\sigma}_1, \dots \boldsymbol{\sigma}_n$ are, for this purpose, approximately independent, we get

$$\nu_0^{n+1}(\mathrm{d}\boldsymbol{\sigma}_0) \ \cong \ p_0(\mathrm{d}\boldsymbol{\sigma}_0) \prod_{k=1}^n \int \exp\Big\{ 2\beta m \Re(Y_{0k} \langle \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_k \rangle) \Big\} \, \nu_k^n(\mathrm{d}\boldsymbol{\sigma}_k)\,. \tag{25}$$

We will hereafter drop the superscripts n, $n+1$ assuming they will be clear from the range of the subscripts.

Next we consider a fixed $k \in [n]$ and estimate the integral by expanding the exponential term. This expansion proceeds slightly different in the real and the complex cases. We give details for the first one, leaving the second to the reader. Write

$$\begin{aligned}
&\int \exp\Big\{2\beta m Y_{0k}\langle \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_k\rangle)\Big\}\, \nu_k(\mathrm{d}\boldsymbol{\sigma}_k) = \\
&= 1 + 2\beta m Y_{0k}\langle \boldsymbol{\sigma}_0, \mathrm{E}_k(\boldsymbol{\sigma}_k)\rangle + \frac{1}{2}\, 4\beta^2 m^2 Y_{0k}^2 \langle \boldsymbol{\sigma}_0, \mathrm{E}_k(\boldsymbol{\sigma}_k\boldsymbol{\sigma}_k^*)\boldsymbol{\sigma}_0\rangle + O(n^{-3/2}) \\
&= \exp\Big\{2\beta m Y_{0k}\langle \boldsymbol{\sigma}_0, \mathrm{E}_k(\boldsymbol{\sigma}_k)\rangle + \frac{1}{2}\, 4\beta^2 m^2 Y_{0k}^2 \langle \boldsymbol{\sigma}_0, \big(\mathrm{E}_k(\boldsymbol{\sigma}_k\boldsymbol{\sigma}_k^*) - \mathrm{E}_k(\boldsymbol{\sigma}_k)\mathrm{E}_k(\boldsymbol{\sigma}_k^*)\big)\boldsymbol{\sigma}_0\rangle + O(n^{-3/2})\Big\},
\end{aligned}$$

where $\mathrm{E}_k(\,\cdot\,) \equiv \int(\,\cdot\,)\,\nu_k(\mathrm{d}\boldsymbol{\sigma}_k)$ denotes expectation with respect to ν_k. Here, we used the fact that $Y_{ij} = O(1/\sqrt{n})$, as per equation (20).

Substituting in Eq. (25), and neglecting $O(n^{-1/2})$ terms, we get (both in the real and complex case)

$$\nu_i(\mathrm{d}\boldsymbol{\sigma}_i) = \frac{1}{z_i}\, \exp\Big\{2\beta\, m\, \Re\langle \boldsymbol{\xi}_i, \boldsymbol{\sigma}_i\rangle + \frac{2\beta m}{s_{\mathfrak{G}}}\, \langle \boldsymbol{\sigma}_i, \boldsymbol{C}_i\boldsymbol{\sigma}_i\rangle\Big\}\, p_0(\mathrm{d}\boldsymbol{\sigma}_i)\ , \tag{26}$$

where $\boldsymbol{\xi}_i \in \mathbb{F}^m$, $\boldsymbol{C}_i \in \mathbb{F}^{m\times m}$, with $\boldsymbol{C}_i^* = \boldsymbol{C}_i$, z_i is a normalization constant, and $s_{\mathfrak{G}} = 1$ (real case) or $s_{\mathfrak{G}} = 2$ (complex case). This expression holds for $i = 0$ and, by the consistency condition, for all $i \in \{1, \dots, n\}$.

We further have the following equations for $\boldsymbol{\xi}_0$, $\boldsymbol{C}_0$:

$$\boldsymbol{\xi}_0 = \sum_{k=1}^{n} Y_{0k}^*\mathrm{E}_k(\boldsymbol{\sigma}_k)\,, \tag{27}$$

$$\boldsymbol{C}_0 = \beta m \sum_{k=1}^{n} |Y_{0k}|^2\big(\mathrm{E}_k(\boldsymbol{\sigma}_k\boldsymbol{\sigma}_k^*) - \mathrm{E}_k(\boldsymbol{\sigma}_k)\mathrm{E}_k(\boldsymbol{\sigma}_k^*)\big)\,. \tag{28}$$

Notice that the expectations $\mathrm{E}_k(\,\cdot\,)$ on the right-hand side are in fact functions of $\boldsymbol{\xi}_k$, $\boldsymbol{C}_k$ through Eq. (26).

We next pass to studying the distribution of $\{\boldsymbol{\xi}_k\}$ and $\{\boldsymbol{C}_k\}$. For large n, the pairs $\{(\boldsymbol{\xi}_k, \boldsymbol{C}_k)\}$ appearing on the right-hand side of Eqs. (27), (28) can be treated as independent. By the law of large numbers and central limit theorem, we obtain that

$$\boldsymbol{\xi}_i \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q})\,, \qquad \boldsymbol{C}_i = \boldsymbol{C}\,, \tag{29}$$

for some deterministic quantities $\boldsymbol{\mu}$, $\boldsymbol{Q}$, $\boldsymbol{C}$. Note that the law of $\boldsymbol{\xi}_i$ can be equivalently described by

$$\boldsymbol{\xi}_i = \boldsymbol{\mu} + \boldsymbol{Q}^{1/2}\boldsymbol{g}\,, \tag{30}$$

where $\boldsymbol{g} = (g_1, g_2, \dots, g_m) \sim \mathsf{N}(0, \mathrm{I}_m)$.

Using these and the consistency condition in Eqs. (27), (28), we obtain the following equations for the unknowns $\boldsymbol{\mu}, \boldsymbol{Q}, \boldsymbol{C}$:

$$\boldsymbol{\mu} = \lambda\mathbb{E}\big\{\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma})\big\}\,, \tag{31}$$

$$\boldsymbol{Q} = \mathbb{E}\big\{\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma})\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma})^*\big\}\,, \tag{32}$$

$$\boldsymbol{C} = \beta m\, \mathbb{E}\big\{\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma}\boldsymbol{\sigma}^*) - \mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma})\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}(\boldsymbol{\sigma})^*\big\}\,. \tag{33}$$

Here $\mathbb{E}$ denotes expectation with respect to $\boldsymbol{\xi} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q})$. Further $\mathrm{E}_{\boldsymbol{\xi},\boldsymbol{C}}$ denotes expectation with respect to the following probability measure on S^{m-1}:

$$\nu_{\boldsymbol{\xi},\boldsymbol{C}}(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{z(\boldsymbol{\xi},\boldsymbol{C})}\, \exp\Big\{2\beta\, m \Re\langle \boldsymbol{\xi}, \boldsymbol{\sigma}\rangle + \frac{2\beta\, m}{s_{\mathfrak{G}}}\langle \boldsymbol{\sigma}, \boldsymbol{C}\boldsymbol{\sigma}\rangle\Big\}\, p_0(\mathrm{d}\boldsymbol{\sigma})\,. \tag{34}$$

The prediction of the replica-symmetric cavity methods have been summarized in the main text. We generalize the discussion here. Assume, for simplicity $\boldsymbol{x_0} = (+1, \dots, +1)$. For ν a probability measure on S^{m-1} and R an orthogonal (or unitary) matrix, let ν^R be the measure obtained by 'rotating' ν, i.e. $\nu^R(\boldsymbol{\sigma} \in A) \equiv \nu(R^{-1}\boldsymbol{\sigma} \in A)$ for any measurable set A. Finally, let $p^{(m,\beta)}_{i(1),\dots,i(k)}$ denote the joint distribution of $\boldsymbol{\sigma}_{i(1)}, \cdots, \boldsymbol{\sigma}_{i(k)}$ under $p_{m,\beta}$. Then, for any fixed k, and any sequence of k-tuples $(i(1), \dots, i(k))_n \in [n]$, we have

$$p^{(m,\beta)}_{i(1),\dots,i(k)} \Rightarrow \int \nu^R_{\boldsymbol{\xi}_1,\boldsymbol{C}}(\,\cdot\,) \times \cdots \times \nu^R_{\boldsymbol{\xi}_k,\boldsymbol{C}}(\,\cdot\,)\, \mathrm{d}R \tag{35}$$

Here $\mathrm{d}R$ denotes the uniform (Haar) measure on the orthogonal group, '$\Rightarrow$' denotes convergence in distribution, and $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_k \sim_{iid} \mathsf{N}(\mu, \boldsymbol{Q})$ as above. Note that the original measure (16) is unaltered under multiplication by a phase. Specifically, if we add an arbitrary phase ϕ_ℓ to coordinate ℓ of the spins, the measure remains unchanged. In order to break this invariance, we consider marginals $\nu_{\boldsymbol{\xi},\boldsymbol{C}}$ that corresponds to $\boldsymbol{\mu}$ being real-valued. Henceforth, without loss of generality we stipulate that $\boldsymbol{\mu}$ is a real-valued vector.

While in general this prediction is only a good approximation (because of replica symmetry breaking) we expect to be asymptotically exact for the Bayes-optimal, ML and SDP estimator. In the next sections we will discuss special estimators.

3.2.2 Bayes-optimal: $m = 1$ and $\beta \in \{\lambda/2, \lambda\}$

For $m = 1$, $\boldsymbol{\sigma} = \sigma$ is a scalar satisfying $\sigma\sigma^* = |\sigma|^2 = 1$. Hence, the term proportional to $\boldsymbol{C}$ in Eq. (34) is a constant and can be dropped. Also $\boldsymbol{\xi} = \xi$, $\boldsymbol{\mu} = \mu$ and $\boldsymbol{Q} = q$ are scalar in this case.

The expression (34) thus reduces to

$$\nu_{\xi}(\mathrm{d}\sigma) = \frac{1}{z(\xi)}\, e^{2\beta\, \Re(\xi^*\sigma)}\, p_0(\mathrm{d}\sigma)\,, \tag{36}$$

with $p_0(\mathrm{d}\sigma)$ the uniform measure on $\{+1, -1\}$ (in the real case) or on $\{z \in \mathbb{C} :\ |z| = 1\}$ (in the complex case). Substituting in Eqs. (31), (32), we get

$$\mu = \lambda \mathbb{E}\big\{\mathrm{E}_\xi(\sigma)\big\}\,, \tag{37}$$

$$q = \mathbb{E}\big\{|\mathrm{E}_\xi(\sigma)|^2\big\}\,, \tag{38}$$

with E_ξ denoting expectation with respect to ν_ξ, and $\mathbb{E}$ expectation with respect to $\xi \sim \mathsf{N}(\mu, q)$.

We will write these equations below in terms of classical functions both in the real and in the complex cases. Before doing that, we derive expressions for the estimation error in the $n \to \infty$ limit. The estimator $\hat{\boldsymbol{x}}(\boldsymbol{Y})$ is given in this case by $\hat{\boldsymbol{x}}(\boldsymbol{Y}) = \hat{\boldsymbol{x}}^{\beta,m=1}(\boldsymbol{Y})$, cf. Eq. (17). Therefore, the

scaled MSE, cf. Eq. (1), reads

$$\mathrm{MSE}(\hat{\boldsymbol{x}}; c) = \lim_{n\to\infty} \left\{ 1 - \frac{2c}{n}\sum_{i=1}^{n} \mathbb{E}\big\{\Re(x_{0,i}\hat{x}_i(\boldsymbol{Y}))\big\} + \frac{c^2}{n}\sum_{i=1}^{n} \mathbb{E}\big\{|\hat{x}_i(\boldsymbol{Y})|^2\big\} \right\} \tag{39}$$

$$= 1 - 2c\mathbb{E}\big\{\Re \mathrm{E}_\xi(\sigma)\big\} + c^2\mathbb{E}\big\{|\mathrm{E}_\xi(\sigma)|^2\big\} \tag{40}$$

$$= 1 - \frac{2c\mu}{\lambda} + c^2 q\,. \tag{41}$$

Note that the optimal scaling is $c = \mu/(\lambda q)$, leading to minimal error –for the ideally scaled estimator–

$$\mathrm{MSE}(\hat{\boldsymbol{x}}) = 1 - \frac{\mu^2}{\lambda^2 q}\,. \tag{42}$$

For the overlap we have, from Eq. (3),

$$\mathrm{Overlap}(\hat{x}) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left\{ \left| \sum_{i=1}^{n} \frac{\hat{x}_i(\boldsymbol{Y})}{|\hat{x}_i(\boldsymbol{Y})|}\, x_{0,i} \right| \right\} \tag{43}$$

$$= \mathbb{E}\Big\{ \frac{\mathrm{E}_\xi(\sigma)}{|\mathrm{E}_\xi(\sigma)|} \Big\}\,. \tag{44}$$

Real case. In this case $\mathrm{E}_\xi(\sigma) = \tanh(2\beta\xi)$ and therefore Eqs. (37), (38) yield

$$\mu = \lambda\mathbb{E}\big\{ \tanh\big(2\beta\mu + 2\beta\sqrt{q}Z\big)\big\}\,, \tag{45}$$

$$q = \mathbb{E}\big\{ \tanh\big(2\beta\mu + 2\beta\sqrt{q}Z\big)^2\big\}\,, \tag{46}$$

where expectation is with respect to $Z \sim \mathsf{N}(0,1)$. As discussed in Section 3.1, the Bayes optimal estimator is recovered by setting $\beta = \lambda/2$ above. Using the identity (12) in Corollary 2.2, we obtain the solution

$$\mu = \frac{\kappa}{\lambda}\,, \qquad q = \frac{\kappa}{\lambda^2}\,. \tag{47}$$

where κ satisfies the fixed point equation

$$\kappa = \lambda^2\mathbb{E}\big\{ \tanh\big(\kappa + \sqrt{\kappa}Z\big)\big\} \tag{48}$$

We denote by $\kappa_* = \kappa_*(\lambda)$ the largest non-negative solution of this equation. Using Eqs. (41) and (44) we obtain the following predictions for the asymptotic estimation error

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{Bayes}}) = 1 - \frac{\kappa_*(\lambda)}{\lambda^2}\,, \tag{49}$$

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{Bayes}}) = 1 - 2\Phi(-\sqrt{\kappa_*(\lambda)})\,. \tag{50}$$

(Note that in this case, the optimal choice of a scaling is $c = 1$.)

Complex case. In this case

$$\mathrm{E}_\xi(\sigma) = \frac{\xi}{|\xi|}\,\frac{\mathrm{I}_1(2\beta|\xi|)}{\mathrm{I}_0(2\beta|\xi|)}\,, \tag{51}$$

where, as mentioned above, $\mathrm{I}_k(z)$ denotes the modified Bessel function of the second kind.

The general fixed point equations (37) and (38) yield

$$\mu = \lambda \mathbb{E}\Big\{ \frac{\mu + \sqrt{q}\Re(Z)}{|\mu + \sqrt{q}Z|} \frac{\mathrm{I}_1(2\beta|\mu + \sqrt{q}Z|)}{\mathrm{I}_0(2\beta|\mu + \sqrt{q}Z|)} \Big\}, \tag{52}$$

$$q = \lambda \mathbb{E}\Big\{ \frac{\mathrm{I}_1(2\beta|\mu + \sqrt{q}Z|)^2}{\mathrm{I}_0(2\beta|\mu + \sqrt{q}Z|)^2} \Big\}. \tag{53}$$

As discussed in Section 3.1, the Bayes optimal estimator is recovered by setting $\beta = \lambda$ in these equations. In this case we can use the identity (13) in Corollary 2.2, to obtain the solution

$$\mu = \frac{\kappa}{\lambda}, \qquad q = \frac{\kappa}{\lambda^2}. \tag{54}$$

where κ satisfies the fixed point equation

$$\kappa = \lambda^2 \mathbb{E}\Big\{ \frac{\kappa + \sqrt{\kappa}\Re(Z)}{|\kappa + \sqrt{\kappa}Z|} \frac{\mathrm{I}_1(2|\kappa + \sqrt{\kappa}Z|)}{\mathrm{I}_0(2|\kappa + \sqrt{\kappa}Z|)} \Big\}, \tag{55}$$

where the expectation is taken with respect to $Z \sim \mathsf{CN}(0,1)$. We denote by $\kappa_* = \kappa_*(\lambda)$ the largest non-negative solution of these equations.

Using again Eqs. (41) and (44) , we obtain

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{Bayes}}) = 1 - \frac{\kappa_*(\lambda)}{\lambda^2}, \tag{56}$$

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{Bayes}}) = \mathbb{E}\Big\{ \frac{\kappa_*(\lambda) + \sqrt{\kappa_*(\lambda)}\,\Re(Z)}{|\kappa_*(\lambda) + \sqrt{\kappa_*(\lambda)}Z|} \Big\}. \tag{57}$$

3.2.3 Maximum likelihood: $m = 1$ and $\beta \to \infty$

As discussed in Section 3.1, the maximum likelihood estimator is recovered by setting $m = 1$ and $\beta \to \infty$. Notice that *in this case our results are only approximate because of replica symmetry breaking.*

We can take the limit $\beta \to \infty$ in Eqs. (36), (37), (38). In this limit, the measure $\nu_\xi(\,\cdot\,)$ concentrates on the single point $\sigma^* = \xi/|\xi| \in S^0$. We thus obtain

$$\mu = \lambda \mathbb{E}\Big\{ \frac{\Re(\xi)}{|\xi|} \Big\}, \tag{58}$$

$$q = 1. \tag{59}$$

We next specialize our discussion to the real and complex cases.

Real case. Specializing Eq. (58) to the real case, we get the equation

$$\mu = \lambda\big(1 - 2\Phi(-\mu)\big). \tag{60}$$

Taylor expanding near $\mu = 0$, this yields $\mu = 2\phi(0)\lambda\,\mu + O(\mu^2)$ which yields the critical point (within the replica symmetric approximation)

$$\lambda_c^{\mathrm{ML,RS}} = \sqrt{\frac{\pi}{2}} \approx 1.253314. \tag{61}$$

We denote by $\mu_* = \mu_*(\lambda)$ the largest non-negative solution of Eq. (60). The asymptotic estimation metrics (for optimally scaled estimator) at level λ are given by

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = 1 - \frac{\mu_*(\lambda)^2}{\lambda^2}\,, \tag{62}$$

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = \frac{\mu_*(\lambda)}{\lambda}\,. \tag{63}$$

It follows immediately from Eq. (60) that, as $\lambda \to \infty$, $\mu_*(\lambda) = \lambda[1 - 2\Phi(-\lambda) + O(\Phi(-\lambda)^2)]$, whence

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = 4\Phi(-\lambda) + O(\Phi(-\lambda)^2) = \sqrt{\frac{8}{\pi\lambda^2}}\, e^{-\lambda^2/2}\left(1 + O(\lambda^{-1})\right), \tag{64}$$

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = 1 - 2\Phi(-\lambda) + O(\Phi(-\lambda)^2) = 1 - \sqrt{\frac{2}{\pi\lambda^2}}\, e^{-\lambda^2/2}\left(1 + O(\lambda^{-1})\right). \tag{65}$$

Complex case. Specializing Eq. (58), we get

$$\mu = \lambda \mathbb{E}\Big\{\frac{\mu + \Re(Z)}{|\mu + Z|}\Big\}\,, \tag{66}$$

where the expectation is with respect to $Z \sim \mathsf{CN}(0,1)$. Taylor-expanding around $\mu = 0$, we get $\mu = (\lambda\mu/2)\mathbb{E}\{|Z|^{-1}\} + O(\mu^2)$. Using $\mathbb{E}\{|Z|^{-1}\} = \sqrt{\pi}$, we obtain the replica-symmetric estimate for the critical point

$$\lambda_c^{\mathrm{ML,RS}} = \frac{2}{\sqrt{\pi}} \approx 1.128379\,. \tag{67}$$

Denoting by $\mu_* = \mu_*(\lambda)$ the largest non-negative solution of Eq. (66), the estimation metrics are obtained again via Eqs. (62) and (63).

For large λ, it is easy to get $\mu_*(\lambda)/\lambda = 1 - (4\lambda^2)^{-1} + O(\lambda^{-3})$ whence

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = \frac{1}{2\lambda^2} + O(\lambda^{-3})\,, \tag{68}$$

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = 1 - \frac{1}{4\lambda^2} + O(\lambda^{-3})\,. \tag{69}$$

3.2.4 General m and $\beta \to \infty$

In the limit $\beta \to \infty$, the measure $\nu_{\boldsymbol{\xi},\boldsymbol{C}}(\,\cdot\,)$ of Eq. (34) concentrates on the single point $\boldsymbol{\sigma}_*(\boldsymbol{\xi}, \boldsymbol{C})$ that maximizes the exponent. A simple calculation yields

$$\boldsymbol{\sigma}_*(\boldsymbol{\xi}, \boldsymbol{C}) = \big(\rho \mathrm{I} - (2/s_{\mathfrak{G}})\boldsymbol{C}\big)^{-1}\boldsymbol{\xi}\,, \tag{70}$$

where ρ is a Lagrange multiplier determined by the normalization condition $\|\boldsymbol{\sigma}_*\|_2 = 1$, or

$$\langle \boldsymbol{\xi}, \big(\rho \mathrm{I} - (2/s_{\mathfrak{G}})\boldsymbol{C}\big)^{-2}\boldsymbol{\xi}\rangle = 1\,. \tag{71}$$

Further $\nu_{\boldsymbol{\xi},\boldsymbol{C}}(\,\cdot\,)$ has variance of order $1/\beta$ around $\boldsymbol{\sigma}_*$.

In order to solve Eqs. (31) to (33) we next assume that the $\mathcal{O}(m)$ symmetry is –at most– broken vectorially to $\mathcal{O}(m-1)$. Without loss of generality, we can assume that it is broken along the direction $\boldsymbol{e}_1 = (1, 0, \dots, 0)$. Further, since $\nu_{\boldsymbol{\xi},\boldsymbol{C}}(\,\cdot\,)$ is a measure on the unit sphere $\{\boldsymbol{\sigma} : \|\boldsymbol{\sigma}\|_2 = 1\}$, the matrix $\boldsymbol{C}$ is only defined up to a shift $\boldsymbol{C} - c_0\mathrm{I}$. This leads to the following ansatz for the order parameters.

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \qquad \boldsymbol{C} = \begin{pmatrix} -s_{\mathfrak{G}} r/2 & & & & \\ & 0 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 0 \end{pmatrix}, \qquad \boldsymbol{Q} = \begin{pmatrix} q_1 & & & & \\ & q_0 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & q_0 \end{pmatrix}, \tag{72}$$

where it is understood that out-of-diagonal entries vanish. Note that the above structure on $(\boldsymbol{\mu}, \boldsymbol{C}, \boldsymbol{Q})$ is the only one that remains invariant under rotations in $\mathbb{F}^m$ that map $\boldsymbol{e}_1$ to itself. We then can represent $\boldsymbol{\xi}$ as follows. For $Z_1, \dots, Z_m \sim_{i.i.d.} \mathsf{N}(0,1)$:

$$\boldsymbol{\xi} = \big(\mu + \sqrt{q_1}\, Z_1, \sqrt{q_0}\, Z_2, \dots, \sqrt{q_0}\, Z_m\big)^{\mathsf{T}}, \tag{73}$$

and $\boldsymbol{\sigma}_*(\boldsymbol{\xi}, \boldsymbol{C})$ reads

$$\boldsymbol{\sigma}_* = \Big(\frac{\mu + \sqrt{q_1}\, Z_1}{\rho + r}, \frac{\sqrt{q_0}\, Z_2}{\rho}, \dots, \frac{\sqrt{q_0}\, Z_m}{\rho}\Big)^{\mathsf{T}}. \tag{74}$$

Taking the limit $\beta \to \infty$ of Eqs. (31) to (33) we obtain the following four equations for the four parameters μ, r, q_0, q_1:

$$\mu = \lambda\, \mathbb{E}\left[\frac{\mu + \sqrt{q_1}\, \Re(Z_1)}{\rho + r}\right], \tag{75}$$

$$q_1 = \mathbb{E}\left[\frac{|\mu + \sqrt{q_1}\, Z_1|^2}{(\rho + r)^2}\right], \tag{76}$$

$$q_0 = q_0 \mathbb{E}\left[\frac{|Z_2|^2}{\rho^2}\right], \tag{77}$$

$$r = \mathbb{E}\left[\frac{|Z_2|^2}{\rho} - \frac{\mu\, \Re(Z_1)}{\sqrt{q_1}(\rho + r)} - \frac{|Z_1|^2}{\rho + r}\right]. \tag{78}$$

Further, the normalization condition $\mathsf{Tr}(\boldsymbol{Q}) = \mathbb{E}(\|\boldsymbol{\sigma}_*(\boldsymbol{\xi}, \boldsymbol{C})\|_2^2) = 1$ yields

$$q_1 + (m-1) q_0 = 1\,. \tag{79}$$

In the above expressions, expectation is with respect to the Gaussian vector $\boldsymbol{Z} = (Z_1, \dots, Z_m) \sim \mathsf{N}(0, \mathrm{I}_m)$, and $\rho = \rho(Z_1, \dots, Z_m)$ is defined as the solution of the equation

$$1 = \frac{|\mu + \sqrt{q_1}\, Z_1|^2}{(\rho + r)^2} + \frac{q_0}{\rho^2} \sum_{i=2}^{m} |Z_i|^2\,. \tag{80}$$

The simplest derivation of these equations is obtained by differentiating the ground state energy, for which we defer to Section 3.3.

We can then compute the performance of the estimator $\hat{\boldsymbol{x}}^{(\beta,m)}$ defined at the beginning of this section. Note that $\widehat{\boldsymbol{Q}} \to \boldsymbol{Q}$ as $n \to \infty$, and therefore its principal vector is $\widehat{\boldsymbol{u}} \to \boldsymbol{e}_1$ (within the above ansatz), and therefore, for a test function f, we have

$$\frac{1}{n}\sum_{i=1}^{n} f(x_{0,i}, \hat{x}_i^{(\infty,m)}) = \mathbb{E}\Big\{ f\Big(X_0, c\,\frac{\mu X_0 + \sqrt{q_1} Z_1}{\rho + r}\Big)\Big\}\,, \tag{81}$$

where $X_0 \sim \mathsf{Unif}(S^0)$ independent of Z_1.

Applying (81) and after a simple calculation we obtain

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{(\infty,m)}) = 1 - \frac{\mu_*(\lambda)^2}{\lambda^2 q_{1,*}(\lambda)}\,, \tag{82}$$

where $\mu_*(\lambda)$, $q_{1,*}(\lambda)$ denote the solutions of the above equations. Also, invoking (81) the asymptotic overlap is given by

$$\begin{aligned}\mathrm{Overlap}(\hat{\boldsymbol{x}}^{(\infty,m)}) &= \mathbb{E}\Big\{ X_0 \frac{\mu_* X_0 + \sqrt{q_{1,*}} Z_1}{|\mu_* X_0 + \sqrt{q_{1,*}} Z_1|}\Big\} \\ &= 1 - 2\Phi\Big(-\frac{\mu_*(\lambda)}{\sqrt{q_{1,*}(\lambda)}}\Big)\,. \end{aligned} \tag{83}$$

Spin-glass phase. The spin-glass phase is described by the completely symmetric solution with $\mu = 0$, $b = 0$ and $q_0 = q_1 = 1/m$. From Eq. (80) we get

$$\rho^2 = \frac{1}{m}\,\|\boldsymbol{Z}\|_2^2\,. \tag{84}$$

Critical signal-to-noise ratio. We next compute the critical value of λ. We begin by expanding Eq. (80). Define

$$F(s) \equiv \frac{q_1}{(s+r)^2}\,|Z_1|^2 + \frac{q_0}{s^2}\sum_{i=2}^{m} |Z_i|^2\,, \tag{85}$$

and let ρ_0 be the solution of the equation $1 = F(\rho_0)$. Notice that ρ_0 is unaltered under sign change $Z_1 \to -Z_1$. Further, comparing with the equation for ρ, see Eq. (80), we obtain the following perturbative estimate

$$\rho = \rho_0 - \frac{1}{F'(\rho_0)}\,\frac{2\mu\sqrt{q_1}\,\Re(Z_1)}{(\rho_0 + r)^2} + O(\mu^2)\,. \tag{86}$$

By the results for the spin glass phase, we have $q_0, q_1 \to (1/m)$ and $\rho, (\rho + r) \to \|\boldsymbol{Z}\|_2/\sqrt{m}$ as $\mu \to 0$, whence

$$\rho = \rho_0 + \frac{\Re(Z_1)}{\|\boldsymbol{Z}\|_2}\,\mu + o(\mu)\,. \tag{87}$$

Now consider Eq. (75). Retaining only $O(\mu)$ terms we get

$$\mu = \lambda\mu\,\mathbb{E}\left[\frac{1}{\rho+r}\right] + \lambda\sqrt{q_1}\,\mathbb{E}\left[\frac{\Re(Z_1)}{\rho+r}\right] \tag{88}$$

$$= \lambda\mu\,\mathbb{E}\left[\frac{\sqrt{m}}{\|\boldsymbol{Z}\|_2}\right] + \lambda\sqrt{q_1}\,\mathbb{E}\left[\frac{\Re(Z_1)}{\rho_0+r}\right] - \lambda\frac{1}{\sqrt{m}}\,\mathbb{E}\left[\frac{\Re(Z_1)}{(\rho_0+r)^2}\frac{\Re(Z_1)}{\|\boldsymbol{Z}\|_2}\right]\,\mu + o(\mu)\,. \tag{89}$$

where in the last step we used Eq. (87) and $q_1 = 1/m + o(1)$ as $\mu \to 0$. Now recalling that ρ_0 is even in Z_1, the second term vanishes and we obtain

$$\mu = \lambda\sqrt{m}\mathbb{E}\left\{\frac{1}{\|\boldsymbol{Z}\|_2} - \frac{\Re(Z_1)^2}{\|\boldsymbol{Z}\|_2^3}\right\}\,\mu + o(\mu)\,. \tag{90}$$

We therefore get the critical point $\lambda_{\rm c}(m)$ by setting to 1 the coefficient of μ above. In the real case, we get

$$\lambda_{\rm c}^{\rm RS}(m)^{-1} = \sqrt{m}\Big(1-\frac{1}{m}\Big)\,\mathbb{E}\big\{1/\|\boldsymbol{Z}\|_2\big\} \tag{91}$$

$$= \sqrt{\frac{2}{m}}\frac{\Gamma((m+1)/2)}{\Gamma(m/2)}\,. \tag{92}$$

In the complex case

$$\lambda_{\rm c}^{\rm RS}(m)^{-1} = \sqrt{m}\Big(1-\frac{1}{2m}\Big)\,\mathbb{E}\big\{1/\|\boldsymbol{Z}\|_2\big\} \tag{93}$$

$$= \sqrt{\frac{1}{m}}\frac{\Gamma(m+(1/2))}{\Gamma(m)}\,. \tag{94}$$

Summarizing the (replica symmetric) critical point is

$$\lambda_{\rm c}^{\rm RS}(m) = \begin{cases} \sqrt{m/2}\,\Gamma(m/2)\Gamma((m+1)/2)^{-1} & \text{for the real case,} \\ \sqrt{m}\,\Gamma(m)\Gamma(m+(1/2))^{-1} & \text{for the complex case.} \end{cases} \tag{95}$$

In particular, for $m = 1$ we recover $\lambda_{\rm c}^{\rm RS}(1) = \sqrt{\pi/2}$ for the real case, and $\lambda_{\rm c}^{\rm RS}(1) = 2/\sqrt{\pi}$ for the complex case. These are the values derived in Section 3.2.3. For large m, we get

$$\lambda_{\rm c}^{\rm RS}(m) = 1 + \frac{1}{4s_{\mathfrak{G}}m} + O(m^{-2})\,, \tag{96}$$

with $s_{\mathfrak{G}} = 1$ (real case), or $s_{\mathfrak{G}} = 2$ (complex case).

Let us emphasize once more: we do not expect the replica symmetric calculation above to be exact, but only an excellent approximation. In other words, *for any bounded m, we expect* $\lambda_{\rm c}(m) \approx \lambda_{\rm c}^{RS}(m)$ *but* $\lambda_{\rm c}(m) \neq \lambda_{\rm c}^{RS}(m)$. However, as $m \to \infty$ the problem becomes convex, and hence we expect $\lim_{m\to\infty}|\lambda_{\rm c}(m) - \lambda_{\rm c}^{\rm RS}(m)| = 0$. Hence

$$\lambda_c^{\rm SDP} = \lim_{m\to\infty}\lambda_{\rm c}(m) = 1\,. \tag{97}$$

3.2.5 SDP: $m \to \infty$ and $\beta \to \infty$

In the limit $m \to \infty$, Eqs. (75) to (78) simplify somewhat. We set $q_1 = q$ and eliminate q_0 using Eq. (79). Applying the law of large numbers, the equation for ρ reads

$$1 = \frac{|\mu + \sqrt{q}\, Z_1|^2}{(\rho + r)^2} + \frac{1 - q}{\rho^2}\,. \tag{98}$$

As a consequence, ρ becomes independent of Z_2. Hence, Eqs. (75) to (78) reduce to

$$\mu = \lambda\, \mathbb{E}\left[\frac{\mu + \sqrt{q}\, \Re(Z_1)}{\rho + r}\right], \tag{99}$$

$$q = \mathbb{E}\left[\frac{|\mu + \sqrt{q}\, Z_1|^2}{(\rho + r)^2}\right], \tag{100}$$

$$r = \mathbb{E}\left[\frac{1}{\rho} - \frac{\mu\, \Re(Z_1)}{\sqrt{q}(\rho + r)} - \frac{|Z_1|^2}{\rho + r}\right], \tag{101}$$

$$1 = \mathbb{E}\left\{\frac{1}{\rho^2}\right\}. \tag{102}$$

Denoting by $\mu_*(\lambda)$ and $q_*(\lambda)$ the solutions to the above equations, we have

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{(\infty,\infty)}) = 1 - \frac{\mu_*(\lambda)^2}{\lambda^2 q_*(\lambda)}\,, \tag{103}$$

Further,

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{(\infty,\infty)}) = 1 - 2\Phi\Big(-\frac{\mu_*(\lambda)}{\sqrt{q_*(\lambda)}}\Big)\,. \tag{104}$$

We solution of the above equations displays a phase transition at the critical point $\lambda_c^{\mathrm{SDP}} = 1$, which we next characterize.

Spin glass phase and critical point. The spin-glass phase corresponds to a symmetric solution $\mu = q = r = 0$.

In order to investigate the critical behavior, we expand the equations (99) to (101) for $\lambda = 1 + \varepsilon$, $\varepsilon \ll 1$. To leading order in ε, we get the following solution

$$\mu = 2\,\varepsilon^{3/2} + O(\varepsilon^2)\,, \tag{105}$$

$$q = s_{\mathfrak{G}}\,\varepsilon^2 + O(\varepsilon^{5/2})\,, \tag{106}$$

$$r = \varepsilon + O(\varepsilon^{3/2})\,, \tag{107}$$

$$\rho = 1 + \varepsilon^2\,(|Z_1|^2 - 1) + o(\varepsilon^{3/2})\,. \tag{108}$$

Hence

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{SDP}}) = 1 - \frac{\mu^2}{q} = 1 - \frac{4}{s_{\mathfrak{G}}}\,\varepsilon + O(\varepsilon^{3/2})\,. \tag{109}$$

To check the above perturbative solution, note that expanding the denominator of Eq. (100) and using $\rho = 1 + O(\varepsilon^2)$, we get

$$q = \mathbb{E}|\lambda\mu + \sqrt{q}Z_1|^2(1 - 2r + O(\varepsilon^2))\,, \tag{110}$$

$$\Leftrightarrow \quad q = \lambda^2\mu^2 + q(1-2r) + O(\varepsilon^4) \tag{111}$$

$$\Leftrightarrow \quad \mu^2 = 2rq + O(\varepsilon^4)\,. \tag{112}$$

Multiplying Eq. (98) by ρ^2 and expanding the right-hand side, we get

$$\rho^2 = 1 - q + q|Z_1|^2 \left(\frac{\rho}{r+\rho}\right)^2 + O(\mu\sqrt{q}) \tag{113}$$

$$\Leftrightarrow \quad \rho = 1 + \frac{q}{2}(|Z_1|^2 - 1) + O(\varepsilon^{5/2})\,. \tag{114}$$

Finally, expanding Eq. (101), we get

$$r = 1 - \frac{\mu}{\sqrt{q}}\mathbb{E}\Big\{\frac{Z_1}{1 + b + (q/2)(|Z_1|^2 - 1)}\Big\} - \mathbb{E}\Big\{\frac{|Z_1|^2}{1 + b + (q/2)(|Z_1|^2 - 1)}\Big\} + O(\varepsilon^{5/2})\,, \tag{115}$$

whence

$$(\mathbb{E}|Z_1|^4 - 1)\, q = 2\, r^2 + O(\varepsilon^{5/2})\,. \tag{116}$$

Finally, expanding Eq. (99) we get $r = \varepsilon + O(\varepsilon^{3/2})$.

3.3 Free energy and energy

It is easier to derive the free energy using the replica method. This also give an independent verification of the cavity calculations in the previous section.

3.3.1 Replica calculation

In this section, apply the replica method to compute the free energy of model (16). Our aim is to compute asymptotics for the partition function

$$Z = \int \exp\Big\{2m\beta \sum_{i<j} \Re(Y_{ij}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle)\Big\} \prod_{i=1}^{n} p_0(\mathrm{d}\boldsymbol{\sigma}_i)\,, \tag{117}$$

where we recall that $p_0(\mathrm{d}\boldsymbol{\sigma}_i)$ is the uniform measure over $\boldsymbol{\sigma}_i \in S^{m-1}$. The k-th moment is given by

$$\mathbb{E}\{Z^k\} = \int \prod_{i<j} \mathbb{E}\exp\Big\{2\beta m \sum_{a=1}^{k} \Re(Y_{ij}\langle \boldsymbol{\sigma}_i^a, \boldsymbol{\sigma}_j^a\rangle)\Big\}\, \overline{p}_0(\mathrm{d}\boldsymbol{\sigma})\,, \tag{118}$$

where we introduced replicas $\boldsymbol{\sigma}_i^1, \dots, \boldsymbol{\sigma}_i^k \in S^{m-1}$, along with the notation $\overline{p}_0(\mathrm{d}\boldsymbol{\sigma}) \equiv \prod_{i=1}^n \prod_{a=1}^k p_0(\mathrm{d}\boldsymbol{\sigma}_i^a)$. Taking the expectation over $Y_{ij} = (\lambda/n) + W_{ij}$, we get

$$\mathbb{E}\{Z^k\} = \int \prod_{i<j} \exp\Big\{\frac{2\beta m\lambda}{n} \sum_{a=1}^{k} \Re\langle \boldsymbol{\sigma}_i^a, \boldsymbol{\sigma}_j^a\rangle + \frac{2\beta^2 m^2}{s_{\mathfrak{G}} n}\Big|\sum_{a=1}^{k}\langle \boldsymbol{\sigma}_i^a, \boldsymbol{\sigma}_j^a\rangle\Big|^2\Big\}\, \overline{p}_0(\mathrm{d}\boldsymbol{\sigma}) \tag{119}$$

$$\doteq \int \exp\Big\{\frac{\beta m\lambda}{n} \sum_{a=1}^{k} \Big\|\sum_{i=1}^{n} \boldsymbol{\sigma}_i^a\Big\|_2^2 + \frac{\beta^2 m^2}{s_{\mathfrak{G}} n} \sum_{a,b=1}^{k} \Big\|\sum_{i=1}^{n} \boldsymbol{\sigma}_i^a (\boldsymbol{\sigma}_i^b)^*\Big\|_F^2\Big\}\, \overline{p}_0(\mathrm{d}\boldsymbol{\sigma})\,. \tag{120}$$

We next use the identity

$$\exp\Big\{\frac{1}{2}\zeta^2\|\boldsymbol{v}\|^2\Big\} \doteq \int \exp\Big\{-\frac{\|\boldsymbol{w}\|^2}{2\zeta^2} + \Re\langle \boldsymbol{w},\boldsymbol{v}\rangle\Big\}\mathrm{d}\boldsymbol{w}\,, \tag{121}$$

where the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ take their entries in $\mathbb{F}$. We apply this identity and introduce Gaussian integrals over the variables $\boldsymbol{\mu}_a \in \mathbb{F}^m$, $\boldsymbol{Q}_{ab} \in \mathbb{F}^{m\times m}$ (with $\boldsymbol{Q}_{ba} = \boldsymbol{Q}_{ab}^*$)

$$\begin{aligned}\mathbb{E}\{Z^k\} \doteq \int \exp\Big\{ &-\frac{\beta m n}{\lambda}\sum_{a=1}^{k}\|\boldsymbol{\mu}_a\|_2^2 + 2\beta m\sum_{a=1}^{k}\sum_{i=1}^{n}\Re\langle\boldsymbol{\mu}_a,\boldsymbol{\sigma}_i^a\rangle \\ &-\frac{\beta^2 m^2 n}{s_{\mathfrak{G}}}\sum_{a,b=1}^{k}\mathsf{Tr}(\boldsymbol{Q}_{ab}\boldsymbol{Q}_{ab}^*) + \frac{2\beta^2 m^2}{s_{\mathfrak{G}}}\sum_{a,b=1}^{k}\sum_{i=1}^{n}\Re\langle\boldsymbol{\sigma}_i^a,\boldsymbol{Q}_{ab}\boldsymbol{\sigma}_i^b\rangle\Big\}\,\overline{p}_0(\mathrm{d}\boldsymbol{\sigma})\,\mathrm{d}\boldsymbol{Q}\mathrm{d}\boldsymbol{\mu}\,.\end{aligned} \tag{122}$$

The final formula for the free energy density is obtained by integrating with respect to $\boldsymbol{\sigma}$ (now the integrand is in product form) and taking the saddle point in $\boldsymbol{Q}$, $\boldsymbol{\mu}$, and is reported in the next section, see Eq. (124) below.

3.3.2 Non-zero temperature ($\beta < \infty$)

The final result of the calculations in the previous section is obtaining the moments

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\{Z^k\} = \mathrm{ext}_{\boldsymbol{Q},\boldsymbol{\mu}}S_k(\boldsymbol{Q},\boldsymbol{\mu})\,, \tag{123}$$

$$S_k(\boldsymbol{Q},\boldsymbol{\mu}) = -\frac{\beta m}{\lambda}\sum_{a=1}^{k}\|\boldsymbol{\mu}_a\|_2^2 - \frac{\beta^2 m^2}{s_{\mathfrak{G}}}\sum_{a,b=1}^{k}\mathsf{Tr}(\boldsymbol{Q}_{ab}\boldsymbol{Q}_{ab}^*) + \log W_k(\boldsymbol{Q},\boldsymbol{\mu})\,. \tag{124}$$

Here, for each $a,b \in \{1,2,\dots,k\}$, $\boldsymbol{\mu}_a \in \mathbb{R}^m$, $\boldsymbol{Q}_{a,b} \in \mathbb{F}^{m\times m}$ with $\boldsymbol{Q}_{b,a} = \boldsymbol{Q}_{a,b}^*$. In particular $\boldsymbol{Q}_{a,a}$ are Hermitian (or symmetric) matrices. The notation $\mathrm{ext}_{\boldsymbol{Q},\boldsymbol{\mu}}$ indicates that we need to take a stationary point over $\boldsymbol{Q}_{ab}, \boldsymbol{\mu}_a$. As usual in the replica method, this will be a local minimum over some of the parameters, and local maximum over the others. Finally, the one-site replicated partition function is

$$W_k(\boldsymbol{Q},\boldsymbol{\mu}) \equiv \int \exp\Big\{2\beta\, m\sum_{a=1}^{k}\Re\langle\boldsymbol{\mu}_a,\boldsymbol{\sigma}_a\rangle + \frac{2\beta^2 m^2}{s_{\mathfrak{G}}}\sum_{a,b=1}^{k}\Re\langle\boldsymbol{\sigma}_a,\boldsymbol{Q}_{a,b}\boldsymbol{\sigma}_b\rangle\Big\}\prod_{a=1}^{k}p_0(\mathrm{d}\boldsymbol{\sigma}_a)\,, \tag{125}$$

where we used the following identity in its derivation

$$\Big(\int f(\boldsymbol{\sigma})p_0(\mathrm{d}\boldsymbol{\sigma})\Big)^n \equiv \int f(\boldsymbol{\sigma}_1)f(\boldsymbol{\sigma}_2)\dots f(\boldsymbol{\sigma}_n)\,p_0(\mathrm{d}\boldsymbol{\sigma}_1)\dots p_0(\mathrm{d}\boldsymbol{\sigma}_n)\,. \tag{126}$$

Recall that $s_{\mathfrak{G}} = 1$ for $\mathfrak{G} = \mathbb{Z}_2$ (real case), and $s_{\mathfrak{G}} = 2$ for $\mathfrak{G} = U(1)$ (complex case).

Replica-symmetric free energy. The replica-symmetric (RS) ansatz is

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}\,, \tag{127}$$

$$\boldsymbol{Q}_{a,b} = \begin{cases}\boldsymbol{Q} + \frac{1}{\beta m}\boldsymbol{C} & \text{if } a = b,\\ \boldsymbol{Q} & \text{if } a \neq b.\end{cases} \tag{128}$$

It follows from the above that $\boldsymbol{C}$, $\boldsymbol{Q} \in \mathbb{F}^{m\times m}$ must be Hermitian (symmetric) matrices. We next compute the RS free energy functional

$$\phi(\boldsymbol{Q},\boldsymbol{C},\boldsymbol{\mu}) = \lim_{k\to 0}\frac{1}{k}S_k(\boldsymbol{Q},\boldsymbol{\mu})\,. \tag{129}$$

By the RS ansatz we have

$$\lim_{k\to 0}\frac{\beta m}{k\lambda}\sum_{a=1}^{k}\|\boldsymbol{\mu}_a\|_2^2 = \frac{\beta m}{\lambda}\|\boldsymbol{\mu}\|_2^2 \tag{130}$$

$$\begin{aligned}\lim_{k\to 0}\frac{\beta^2 m^2}{k s_{\mathfrak{G}}}\sum_{a,b=1}^{k}\mathsf{Tr}(\boldsymbol{Q}_{ab}\boldsymbol{Q}_{ab}^*) &= \lim_{k\to 0}\frac{\beta^2 m^2}{s_{\mathfrak{G}}}\Big\{\mathsf{Tr}\Big(\Big(\boldsymbol{Q}+\frac{1}{\beta m}\boldsymbol{C}\Big)^2\Big) + (k-1)\mathsf{Tr}(\boldsymbol{Q}^2)\Big\}\\ &= \frac{\beta^2 m^2}{s_{\mathfrak{G}}}\Big\{\mathsf{Tr}\Big(\Big(\boldsymbol{Q}+\frac{1}{\beta m}\boldsymbol{C}\Big)^2\Big) - \mathsf{Tr}(\boldsymbol{Q}^2)\Big\}\end{aligned} \tag{131}$$

For computing the third term, we use the following identity. For a fixed arbitrary vector $\boldsymbol{v}$,

$$\mathbb{E}\Big\{\exp(\Re\langle \boldsymbol{v},\boldsymbol{\xi}\rangle)\Big\} = \exp\Big\{\Re\langle \boldsymbol{v},\boldsymbol{\mu}\rangle + \frac{1}{2s_{\mathfrak{G}}}\Re\langle \boldsymbol{v},\boldsymbol{Q}\boldsymbol{v}\rangle\Big\}\,, \tag{132}$$

where $\boldsymbol{\xi}\sim \mathsf{N}(\boldsymbol{\mu},\boldsymbol{Q})$ in the real case, and $\boldsymbol{\xi}\sim \mathsf{CN}(\boldsymbol{\mu},\boldsymbol{Q})$ in the complex case. Further, the expectation $\mathbb{E}$ is with respect to $\boldsymbol{\xi}$.

Applying this identity, we write

$$\begin{aligned}&W_k(\boldsymbol{Q},\boldsymbol{\mu})\\ &= \int \exp\Big\{2\beta m\,\Re\langle\boldsymbol{\mu},\sum_{a=1}^{k}\boldsymbol{\sigma}_a\rangle + \frac{2\beta m}{s_{\mathfrak{G}}}\sum_{a=1}^{k}\Re\langle\boldsymbol{\sigma}_a,\boldsymbol{C}\boldsymbol{\sigma}_a\rangle + \frac{2\beta^2 m^2}{s_{\mathfrak{G}}}\Re\Big\langle\sum_{a=1}^{k}\boldsymbol{\sigma}_a,\boldsymbol{Q}\sum_{a=1}^{k}\boldsymbol{\sigma}_a\Big\rangle\Big\}\prod_{a=1}^{k}p_0(\mathrm{d}\boldsymbol{\sigma}_a)\\ &= \int \mathbb{E}\exp\Big\{\frac{2\beta m}{s_{\mathfrak{G}}}\sum_{a=1}^{k}\Re\langle\boldsymbol{\sigma}_a,\boldsymbol{C}\boldsymbol{\sigma}_a\rangle + 2\beta m\sum_{a=1}^{k}\Re\langle\boldsymbol{\sigma}_a,\boldsymbol{\xi}\rangle\Big\}\prod_{a=1}^{k}p_0(\mathrm{d}\boldsymbol{\sigma}_a)\\ &= \mathbb{E}\Big(\int\exp\Big\{\frac{2\beta m}{s_{\mathfrak{G}}}\Re\langle\boldsymbol{\sigma},\boldsymbol{C}\boldsymbol{\sigma}\rangle + 2\beta m\Re\langle\boldsymbol{\sigma},\boldsymbol{\xi}\rangle\Big\}p_0(\mathrm{d}\boldsymbol{\sigma})\Big)^k\,.\end{aligned} \tag{133}$$

Hence,

$$\lim_{k\to 0}\frac{1}{k}\log W_k(\boldsymbol{Q},\boldsymbol{\mu}) = \mathbb{E}\log\Big(\int\exp\Big\{2\beta m\Re\langle\boldsymbol{\xi},\boldsymbol{\sigma}\rangle + \frac{2\beta m}{s_{\mathfrak{G}}}\langle\boldsymbol{\sigma},\boldsymbol{C}\boldsymbol{\sigma}\rangle\Big\}p_0(\mathrm{d}\boldsymbol{\sigma})\Big)\,. \tag{134}$$

Combining Eqs. (130), (131) and (134) we arrive at

$$\begin{aligned}\phi(\boldsymbol{Q},\boldsymbol{C},\boldsymbol{\mu}) = &-\frac{\beta m}{\lambda}\|\boldsymbol{\mu}\|_2^2 - \frac{1}{s_{\mathfrak{G}}}\beta^2 m^2\Big\{\mathsf{Tr}\Big(\Big(\boldsymbol{Q}+\frac{1}{\beta m}\boldsymbol{C}\Big)^2\Big) - \mathsf{Tr}(\boldsymbol{Q}^2)\Big\}\\ &+\mathbb{E}\log\Big(\int\exp\Big\{2\beta m\Re\langle\boldsymbol{\xi},\boldsymbol{\sigma}\rangle + \frac{2\beta m}{s_{\mathfrak{G}}}\langle\boldsymbol{\sigma},\boldsymbol{C}\boldsymbol{\sigma}\rangle\Big\}p_0(\mathrm{d}\boldsymbol{\sigma})\Big)\,,\\ &\boldsymbol{\xi}\sim\mathsf{N}(\boldsymbol{\mu},\boldsymbol{Q})\,.\end{aligned} \tag{135}$$

In the complex case, the last line should be interpreted as $\boldsymbol{\xi}\sim\mathsf{CN}(\boldsymbol{\mu},\boldsymbol{Q})$. Differentiating this expression against $\boldsymbol{\mu},\boldsymbol{C},\boldsymbol{Q}$ we recover Eqs. (31) to (33) as saddle point conditions.

3.3.3 Zero temperature ($\beta \to \infty$)

As $\beta \to \infty$, the free energy behaves as

$$\phi(\boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{\mu}) = 2\beta m \, u(\boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{\mu}) + o(\beta)\,, \tag{136}$$

where $u(\boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{\mu})$ is the replica-symmetric ground state energy

$$\begin{aligned} u(\boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{\mu}) = -\frac{1}{2\lambda}\|\boldsymbol{\mu}\|_2^2 - \frac{1}{s_{\mathfrak{G}}}\mathsf{Tr}(\boldsymbol{C}\boldsymbol{Q}) + \mathbb{E} \max_{\boldsymbol{\sigma} \in S^{m-1}} \big[\Re\langle \boldsymbol{\xi}, \boldsymbol{\sigma}\rangle + \frac{1}{s_{\mathfrak{G}}}\langle \boldsymbol{\sigma}, \boldsymbol{C}\boldsymbol{\sigma}\rangle\big]\,, \\ \boldsymbol{\xi} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q})\,. \end{aligned} \tag{137}$$

Let us stress that expectation is with respect to $\boldsymbol{\xi}$. Denote by $\boldsymbol{\sigma}_{\mathrm{M}} = \boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})$ the solution of the above maximization problem. It is immediate to see that this is given by

$$\boldsymbol{\sigma}_{\mathrm{M}} = \big(\rho \mathrm{I} - (2/s_{\mathfrak{G}})\boldsymbol{C}\big)^{-1}\boldsymbol{\xi}\,, \tag{138}$$

where $\rho = \rho(\boldsymbol{\xi}, \boldsymbol{C}) \in \mathbb{R}$ is a Lagrange multiplier determined by solving the equation

$$\langle \boldsymbol{\xi}, \big(\rho \mathrm{I} - (2/s_{\mathfrak{G}})\boldsymbol{C}\big)^{-2}\boldsymbol{\xi}\rangle = 1\,. \tag{139}$$

The equations for $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ are immediate by taking the $\beta \to \infty$ limit on Eqs. (31), (32). In zero temperature, measure $\nu_{\boldsymbol{\xi}, \boldsymbol{C}}$ concentrates around $\boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})$.

$$\boldsymbol{\mu} = \mathbb{E}\{\Re \boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})\}\,, \tag{140}$$

$$\boldsymbol{Q} = \mathbb{E}\{\boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})\boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})^*\}\,. \tag{141}$$

Equivalently, we obtain the above equations by differentiating $u(\boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{C}$, as follows. We write $\boldsymbol{\sigma}_{\mathrm{M}} = \boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})$ to lighten the notation. Since ρ is a Lagrange multiplier, we have

$$\nabla_{\boldsymbol{\sigma}_{\mathrm{M}}}\Big(\Re\langle \boldsymbol{\xi}, \boldsymbol{\sigma}_{\mathrm{M}}\rangle + (1/s_{\mathfrak{G}})\langle \boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{C}, \boldsymbol{\sigma}_{\mathrm{M}}\rangle\Big) = \rho \boldsymbol{\sigma}_{\mathrm{M}}\,.$$

Hence,

$$\begin{aligned} \nabla_{\boldsymbol{C}} u &= -\frac{1}{s_{\mathfrak{G}}}\boldsymbol{Q} + \mathbb{E}\Big\{\frac{1}{s_{\mathfrak{G}}}\boldsymbol{\sigma}_{\mathrm{M}}\boldsymbol{\sigma}_{\mathrm{M}}^* + \rho\sum_{\ell=1}^{m}\boldsymbol{\sigma}_{\mathrm{M},\ell}\nabla_{\boldsymbol{C}}\boldsymbol{\sigma}_{\mathrm{M},\ell}\Big\} \\ &= -\frac{1}{s_{\mathfrak{G}}}\boldsymbol{Q} + \mathbb{E}\Big\{\frac{1}{s_{\mathfrak{G}}}\boldsymbol{\sigma}_{\mathrm{M}}\boldsymbol{\sigma}_{\mathrm{M}}^*\Big\}\,, \end{aligned} \tag{142}$$

where the second equation follows from the constraint $\|\boldsymbol{\sigma}_{\mathrm{M}}\|_2 = 1$.

We next substitute $\boldsymbol{\xi} = \boldsymbol{\mu} + \boldsymbol{Q}^{1/2}\boldsymbol{Z}$. By a similar calculation, we have

$$\nabla_{\boldsymbol{\mu}} u = -\frac{1}{\lambda}\boldsymbol{\mu} + \mathbb{E}\{\Re \boldsymbol{\sigma}_{\mathrm{M}}\}\,. \tag{143}$$

$$\nabla_{\boldsymbol{Q}^{1/2}} u = -[\boldsymbol{C}\boldsymbol{Q}^{1/2}]_s + \frac{s_{\mathfrak{G}}}{2}\,\mathbb{E}\big\{[\boldsymbol{Z}\,\boldsymbol{\sigma}_{\mathrm{M}}(\boldsymbol{\xi}, \boldsymbol{C})^*]_s\big\}\,. \tag{144}$$

Recall that expectation $\mathbb{E}$ is with respect to $\boldsymbol{\xi} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q})$ or, equivalently, $\boldsymbol{Z} \sim \mathsf{N}(0, \mathrm{I}_m)$. Further $[\boldsymbol{A}]_s$ denotes the symmetric (Hermitian) part of the matrix $\boldsymbol{A}$, i.e. $[\boldsymbol{A}]_s = (\boldsymbol{A} + \boldsymbol{A}^*)/2$.

Using the ansatz (72), we recover Eqs. (75) to (78). Specifically, Eq (75) follows readily from Eq. (140), restricting to the $(1,1)$ entry and plugging in for $\boldsymbol{\sigma}_{\mathrm{M}}$ from Eq. (74). Also, Eqs. (76) and (77) follow from Eq. (141), restricting to $(1,1)$ and $(2,2)$ entries, respectively. Derivation of Eq. (78) requires more care. Note that since $\nu_{\xi,\boldsymbol{C}}$, given by (34), is a measure on the unit sphere, the matrix $\boldsymbol{C}$ is only defined up to a diagonal shift. Let $s_{\mathfrak{G}}\eta/2$ denote the slack shift parameter. The ansatz (72) for $\boldsymbol{C}$ then becomes

$$\boldsymbol{C} = \begin{pmatrix} -s_{\mathfrak{G}}(r+\eta)/2 & & & & \\ & -s_{\mathfrak{G}}\eta/2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & -s_{\mathfrak{G}}\eta/2 \end{pmatrix}.$$

We set $\nabla_{\boldsymbol{Q}^{1/2}} u = 0$. Applying Eq. (144), this results in the following two equations for $\boldsymbol{C}_{11}$ and $\boldsymbol{C}_{22}$:

$$\frac{-s_{\mathfrak{G}}}{2}(r+\eta)\sqrt{q_1} = \frac{s_{\mathfrak{G}}}{2}\mathbb{E}\Big\{\frac{\mu\Re(Z_1)}{\rho+r} + \frac{\sqrt{q_1}|Z_1|^2}{\rho+r}\Big\}, \tag{145}$$

$$-\frac{s_{\mathfrak{G}}}{2}\eta\sqrt{q_0} = \frac{s_{\mathfrak{G}}}{2}\mathbb{E}\Big\{\frac{\sqrt{q_0}|Z_2|^2}{\rho}\Big\}. \tag{146}$$

Solving for η from Eq. (146) and substituting for that in Eq. (145), we obtain Eq. (78).

3.4 On the maximum likelihood phase transition

In Section 3.2.3 we computed the replica symmetric approximation for the phase tranition of the maximum likelihood estimator. We obtained $\lambda_c^{\mathrm{ML,RS}}(\mathbb{Z}_2) = \sqrt{\pi/2}$ (real case), $\lambda_c^{\mathrm{ML,RS}}(U(1)) = 2/\sqrt{\pi}$ (complex case). This is somewhat surprising because it suggests that the maximum likelihood estimator has a worse threshold than the Bayes optimal estimator.

It turns out that this is an artifact of the replica symmetric approximation and instead

$$\lambda_c^{\mathrm{ML}}(\mathbb{Z}_2) = \lambda_c^{\mathrm{ML}}(U(1)) = 1\,. \tag{147}$$

We next outline the heuristic argument that support this claim. (For the sake of simplicity, we will consider the $\mathbb{Z}_2$ case.)

For a given noise realization $\boldsymbol{W}$, the maximum likelihood estimator is

$$\hat{\boldsymbol{x}}^{\mathrm{ML}}(\boldsymbol{Y}) = \arg\max_{\boldsymbol{x}\{+1,-1\}^n} \left\{\frac{\lambda}{2n}\langle \boldsymbol{x}, \boldsymbol{x_0}\rangle^2 + \frac{1}{2}\langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x}\rangle\right\}. \tag{148}$$

We then define the correlation

$$M(\boldsymbol{W}) \equiv \frac{1}{n}\big|\langle \boldsymbol{x_0}, \hat{\boldsymbol{x}}^{\mathrm{ML}}(\boldsymbol{Y})\rangle\big|\,, \tag{149}$$

and recall that $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\mathrm{ML}}) = \mathbb{E}\{M(\boldsymbol{W})\}$. We want to show that $M(\boldsymbol{W})$ is (with high probability) bounded away from 0 for $\lambda > 1$. Setting, without loss of generality, $\boldsymbol{x_0} = \mathbf{1}$, we have

$$M(\boldsymbol{W}) = \arg\max_{m\in[0,1]} \Big\{F_{\boldsymbol{W},n}(m) + \frac{\lambda}{2}\,m^2\Big\}, \tag{150}$$

$$F_{\boldsymbol{W},n}(m) \equiv \max_{\boldsymbol{x}\in\{+1,-1\}^n} \Big\{\frac{1}{2n}\langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x}\rangle :\ \frac{1}{n}\sum_{i=1}^n x_i \in \big[m, m+1/n\big)\Big\}. \tag{151}$$

We expect $\lim_{n\to\infty} F_{\boldsymbol{W},n}(m) \equiv F(m)$ to exist and to be non-random. This implies that the asymptotic overlap is given by

$$\text{Overlap}(\hat{\boldsymbol{x}}^{\text{ML}}) = \arg\max_{m\in[0,1]} \left\{ F(m) + \frac{\lambda}{2}\, m^2 \right\} \tag{152}$$

By symmetry we have $F(m) = F(-m)$. Assuming $m \mapsto F(m)$ to be differentiable, this implies $F'(0) = 0$. Hence $m = 0$ is a local maximum for $\lambda < -F''(0)$ and a local minimum for $\lambda > -F''(0)$. Since at $\lambda = 0$ we obviously have $\text{Overlap}(\hat{\boldsymbol{x}}^{\text{ML}}) = 0$, $F''(0) < 0$. Further, if $m = 0$ is a local minimum, we necessarily have $\text{Overlap}(\hat{\boldsymbol{x}}^{\text{ML}}) > 0$. Hence $\lambda_c^{\text{ML}} \le -F''(0)$.

On the other hand, we know that we cannot estimate $\boldsymbol{x_0}$ with non-vanishing overlap for $\lambda < 1$. This is a consequence –for instance– of [DAM15, Theorem 4.3] or can, in alternative, be proved directly using the technique of [MRZ14]. This implies that $\lambda_c^{\text{ML}} \ge 1$. Summarizing, we have

$$1 \le \lambda_c^{\text{ML}} \le -F''(0)\,. \tag{153}$$

We next claim that earlier work on the Sherrington-Kirkpatrick model implies $F''(0) = -1$, thus yielding $\lambda_c^{\text{ML}} = 1$. Indeed, alternative expressions can be obtained by studying the modified problem

$$\widehat{F}_{\boldsymbol{W},n}(h) \equiv \frac{1}{n} \max_{\boldsymbol{x}\in\{+1,-1\}^n} \left\{ \frac{1}{2}\langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x}\rangle + h\langle \boldsymbol{1}, \boldsymbol{x}\rangle \right\}, \tag{154}$$

where h is an added magnetic field. Then, we have $\lim_{n\to\infty} \widehat{F}_{\boldsymbol{W},n}(h) = \widehat{F}(h)$, the Legendre transform of F, and we get the alternative upper bound

$$\lambda_c^{\text{ML}} \le -F''(0) = \frac{1}{\widehat{F}''(0)}\,. \tag{155}$$

Note that $\widehat{F}(h)$ is the zero-temperature free energy density of the Sherrington-Kirkpatrick model in a magnetic field h [MPV87], whose $n \to \infty$ limit exists by [GT02]. Using well-known thermodynamic identities, we get

$$\lambda_c^{\text{ML}} \le -F''(0) = = \lim_{\beta\to\infty} \lim_{n\to\infty} \frac{1}{\chi(\beta, h = 0+)} \tag{156}$$

$$= \lim_{\beta\to\infty} \lim_{n\to\infty} \frac{1}{\beta \mathbb{E}_{\beta,h=0+}\{1 - Q\}}\,, \tag{157}$$

where $\chi(\beta, h)$ is the magnetic susceptibility of the Sherrington-Kirkpatrick model at inverse temperature β, and magnetic field h, and Q is the random overlap.

To the best of our knowledge, the above connection between response to a magneric field, and couplings with non-zero mean was first described by Gérard Toulouse[1] in [Tou80].

The relation $\beta\mathbb{E}_{\beta,h=0+}\{1 - Q\} = 1$ was derived in [Som83] from Sompolisky's formulation of mean field theory. This result is also confirmed by recent high-precision numerical approximations of the ∞-RSB solution of the Sherrington-Kirkpatrick model [OS08].

[1] In [Tou80], this argument was put forward within the context of the so-called Parisi-Toulouse (PaT) scaling hypothesis. Let us emphasize that here we are not assuming PaT to hold (and indeed, it has been convincingly shown that PaT is not correct, albeit an excellent approximation, see e.g. [CRT03]).

4 Analysis of PCA estimator for synchronization problem

Here, we study the PCA estimator for the synchronization problem. Recall the observation model

$$\boldsymbol{Y} = \frac{\lambda}{n}\boldsymbol{x}_0\boldsymbol{x}_0^* + \boldsymbol{W}\,, \tag{158}$$

with $\boldsymbol{x}_0 \in \mathbb{F}^n$ and $\|\boldsymbol{x}_0\| = n$, where $\|\cdot\|$ refers to the norm on $\mathbb{F}$.

Let $\boldsymbol{v}_1(Y)$ denote the leading eigenvector of $\boldsymbol{Y}$. The PCA estimator $\hat{\boldsymbol{x}}^{\mathrm{PCA}}$ is defined as

$$\hat{\boldsymbol{x}}^{\mathrm{PCA}}(\boldsymbol{Y}) \equiv \sqrt{n}\, c^{\mathrm{PCA}}(\lambda)\boldsymbol{v}_1(\boldsymbol{Y})\,, \tag{159}$$

with c^{PCA} a certain scaling factor discussed below.

In order to characterize the error of $\hat{\boldsymbol{x}}^{\mathrm{PCA}}$, we use a simplified version of the main theorem in [CDMF09].

Lemma 4.1 ([CDMF09])**.** *Let $\boldsymbol{Y} = \lambda\boldsymbol{v}_0\boldsymbol{v}_0^* + \boldsymbol{W}$ be a rank-one deformation of the Gaussian symmetric matrix $\boldsymbol{W}$, with $W_{ij} \sim \mathsf{N}(0, 1/n)$ independent for $i < j$, and $\|\boldsymbol{v}_0\| = 1$. Then, we have, almost surely*

$$\lim_{n\to\infty} |\langle \boldsymbol{v}_1(\boldsymbol{Y}), \boldsymbol{v}_0\rangle| = \begin{cases} 0 & \text{if } \lambda \le 1\,, \\ \sqrt{1-\lambda^{-2}} & \text{if } \lambda > 1\,. \end{cases} \tag{160}$$

Further, letting $\lambda_1(\boldsymbol{Y})$ be the top eigenvalue of $\boldsymbol{Y}$, the following holds true almost surely

$$\lim_{n\to\infty} \lambda_1(\boldsymbol{Y}) = \begin{cases} 2 & \text{if } \lambda \le 1\,, \\ \lambda + 1/\lambda & \text{if } \lambda > 1\,. \end{cases} \tag{161}$$

Applying this lemma, we compute $\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{PCA}}; c)$ as follows

$$\begin{aligned} \mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{PCA}}; c) &= \lim_{n\to\infty} \Big\{1 - \frac{2c}{n}\mathbb{E}\{\Re\langle \boldsymbol{x}_0, \hat{\boldsymbol{x}}^{\mathrm{PCA}}\rangle\} + \frac{c^2}{n}\|\hat{\boldsymbol{x}}^{\mathrm{PCA}}\|^2\Big\} \\ &= 1 - 2c\sqrt{1-\lambda^{-2}} + c^2\,, \end{aligned} \tag{162}$$

which is optimized for $c = c^{\mathrm{PCA}}(\lambda) \equiv \sqrt{\max(1-\lambda^{-2}, 0)}$. Note that this choice can be written in terms of $\lambda_1(\boldsymbol{Y})$ as well and so knowledge of λ is not required. We then obtain

$$\mathrm{MSE}(\hat{\boldsymbol{x}}^{\mathrm{PCA}}(\boldsymbol{Y}); c^{\mathrm{PCA}}(\lambda)) = \min(1, \lambda^{-2})\,. \tag{163}$$

5 Analytical results for community detection

In this Section we use the cavity method to analyze the semidefinite programming approach to community detection. We refer, for instance, to [MM09] for general background on the cavity method for sparse graphs. Also, see [BSS87, SW87] for early statistical mechanics work on the related graph bisection problem.

While we focus on the cases of two communities, we believe that the methods developed here are applicable to a significantly broader class of graph models. Semidefinite programming relaxations for community detection problems where studied in [ABH16, HWX15a, Ban15]. for the case of two

communities, and in [HWX15b, ABKK15] for multiple communities. These works focus however on the regime in which the average node degree is of order $\Theta(\log n)$ and exact reconstruction of the vertex labels is possible. We are interested in bounded average dgree, a case that was studied in [MS16, GV14].

Recall (from the main text) that we are interested in the hidden partition model. Namely, consider a random graph $G_n = (V_n, E_n)$ over vertex set $V_n = [n]$, generated according to the following distribution. We let $\boldsymbol{x_0} \in \{+1, -1\}^n$ be uniformly random: this vector contains the vertex labels (equivalently, it encodes a partition of the vertex set $V = V_+ \cup V_-$, in the obvious way). Conditional on $\boldsymbol{x_0}$, edges are independent with distribution

$$\mathbb{P}\big\{(i,j) \in E_n \big| \boldsymbol{x_0}\big\} = \begin{cases} a/n & \text{if } x_{0,i}x_{0,j} = +1, \\ b/n & \text{if } x_{0,i}x_{0,j} = -1. \end{cases} \tag{164}$$

As explained in the main text, we tackle this problem via the semidefinite relaxation

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E_n} X_{ij}\,, \\ \text{subject to} \quad & \boldsymbol{X} \succeq 0\,, \\ & \boldsymbol{X}\boldsymbol{1} = \boldsymbol{0}\,, \quad X_{ii} = 1 \ \ \forall i \in [n]\,. \end{aligned} \tag{165}$$

For our analysis, we use the non-convex formulation

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E_n} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle\,, \\ \text{subject to} \quad & \sum_{i=1}^n \boldsymbol{\sigma}_i = \boldsymbol{0}\,, \\ & \boldsymbol{\sigma}_i \in S^{m-1} \subseteq \mathbb{R}^m \ \ \forall i \in [n]\,. \end{aligned} \tag{166}$$

This is equivalent to the above SDP provided $m \geq n$. Note that, throughout this section, the spin variables $\boldsymbol{\sigma}_i$ are *real* vectors.

We introduce the following Boltzmann-Gibbs distribution

$$p_{\beta,m}(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{Z_G(\beta,m)} \exp\Big\{2m\beta \sum_{(i,j)\in E} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle\Big\}\, \overline{p}_0(\mathrm{d}\boldsymbol{\sigma})\,. \tag{167}$$

Here $\overline{p}_0(\mathrm{d}\boldsymbol{\sigma}_i)$ is the uniform measure over $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)$ with $\boldsymbol{\sigma}_i \in S^{m-1}$ and $\sum_{i=1}^n \boldsymbol{\sigma}_i = \boldsymbol{0}$. In order to extract information about the SDP (165), we take the limits $m \to \infty$, $\beta \to \infty$ *after* $n \to \infty$.

As $n \to \infty$, the graph G_n converges locally to a rooted multi-type Galton-Watson tree with vertices of type + (corresponding to $x_{0,i} = +1$) or − (corresponding to $x_{0,i} = -1$). Each vertex has Poisson$(a/2)$ offsprings of the same type, and Poisson$(b/2)$ offsprings of the other type (see, e.g. [DM10] for background on local weak convergence in statistical mechanics).

We write the sum-product fixed point equations to compute the marginals at different nodes.

$$\nu_{i\to j}(\mathrm{d}\boldsymbol{\sigma}_i) \cong \prod_{\ell\in\partial i\setminus\{j\}} \int e^{2\beta m\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_\ell\rangle}\, \nu_{\ell\to i}(\mathrm{d}\boldsymbol{\sigma}_\ell)\,, \tag{168}$$

where $\nu_{i\to j}$ are the messages associated to the directed edges of the graph. The marginal $\nu_0(\mathrm{d}\boldsymbol{\sigma}_0)$, for an arbitrary node 0, is given by

$$\nu_0(\mathrm{d}\boldsymbol{\sigma}_0) \cong \prod_{i\in\partial 0} \int e^{2\beta m\langle \boldsymbol{\sigma}_0,\boldsymbol{\sigma}_i\rangle}\, \nu_{i\to 0}(\mathrm{d}\boldsymbol{\sigma}_i)\,. \tag{169}$$

We rewrite the above equations from another perspective. We designate node 0 as the root of the tree and denote its neighbors by $\{1, 2, \dots, k\}$. Let T_i be the subtree rooted at node i and induced by its descendants. We call $\nu_i(\mathrm{d}\boldsymbol{\sigma}_i)$ the marginal for $\boldsymbol{\sigma}_i$ w.r.t the graphical model in the subtree T_i. Replica symmetric cavity equations relate the marginal $\nu_0(\mathrm{d}\boldsymbol{\sigma}_0)$ to the marginals at the descendant subtrees, i.e., $\nu_i(\mathrm{d}\boldsymbol{\sigma}_i)$. Note that, in the above notation, $\nu_i(\mathrm{d}\boldsymbol{\sigma}_i) \equiv \nu_{i\to 0}(\mathrm{d}\boldsymbol{\sigma}_i)$ and therefore we obtain

$$\nu_0(\mathrm{d}\boldsymbol{\sigma}_0) \cong \prod_{i=1}^{k} \widehat{\nu}_i(\boldsymbol{\sigma}_0)\,. \tag{170}$$

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \int e^{2\beta m\langle \boldsymbol{\sigma}_0,\boldsymbol{\sigma}_i\rangle}\, \nu_i(\mathrm{d}\boldsymbol{\sigma}_i)\,. \tag{171}$$

(The measures $\nu_i(\mathrm{d}\boldsymbol{\sigma}_i)$ are probability measures over S^{m-1} and the right-hand side should be interpreted as a density with respect to the uniform measure on S^{m-1}.)

We will use the notation of Eq. (170) but both interpretations are useful.

Notice that the (non-local) constraint $\sum_{i=1}^n \boldsymbol{\sigma}_i = \mathbf{0}$ does not enter these equations. However, it is enforced by selecting a solution of the cavity equations such that $\mathbb{E}\{\int \boldsymbol{\sigma}_0 \nu_0(\mathrm{d}\boldsymbol{\sigma}_0)\} = \mathbf{0}$, where $\mathbb{E}$ is expectation with respect to the underlying graph which is a Galton-Watson tree with Poisson offspring distribution.

Remark 5.1. *We will carry out our calculations within a simple 'vectorial' ansatz, whereby $\nu_i(\mathrm{d}\boldsymbol{\sigma}_i)$ depends only on a two-dimensional projection of $\boldsymbol{\sigma}_i$. The ansatz is defined in its most general form in Eq. (227) below. It turns out that, for computing the critical point $\lambda_c^{SDP}(d)$, the quadratic coefficient r_i can be set to 0, which is what we will do in Sections 5.1 through 5.3.*

While this ansatz is not exact, it turns out to yield very accurate results. Also, it can be systematically improved upon, a direction that we leave for future work.

5.1 Symmetric phase

For small $(a - b)$, we expect the solution to the cavity equation to be symmetric (in distribution) under rotations in $\mathcal{O}(m)$. By this we mean that, for any rotation $R \in \mathcal{O}(m)$, $\nu_i^R(\,\cdot\,)$ is distributed as $\nu_i(\,\cdot\,)$. ($\nu_i^R(\,\cdot\,)$ is defined as the measure induced by action $\boldsymbol{\sigma} \mapsto R\boldsymbol{\sigma}$ on S^{m-1}, cf. Section 3).

In the symmetric phase, assuming the 'vectorial' ansatz, cf. Remark 5.1, we look for an approximate solution of the form

$$\nu_i(\mathrm{d}\boldsymbol{\sigma}_i) \cong \exp\Big\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\sigma}_i\rangle + O_m(1)\Big\}\, p_0(\mathrm{d}\boldsymbol{\sigma}_i) \tag{172}$$

$$\cong \exp\Big\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\sigma}_i\rangle + O_m(1)\Big\}\, \delta\Big(\|\boldsymbol{\sigma}_i\|_2^2 - 1\Big)\, \mathrm{d}\boldsymbol{\sigma}_i\,, \tag{173}$$

where $\boldsymbol{z}_i \sim \mathsf{N}(0, \mathrm{I}_m)$, and $O_m(1)$ represents a term of order one as $m \to \infty$.

Using the Fourier representation of the δ function (with associated parameter ρ), and performing the Gaussian integral over $\boldsymbol{\sigma}$, we get

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \int \exp\Big\{2\beta m\langle \boldsymbol{\sigma}_0,\boldsymbol{\sigma}\rangle + \beta m\rho + 2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i,\boldsymbol{\sigma}\rangle - \beta m\rho\|\boldsymbol{\sigma}\|_2^2 + O_m(1)\Big\}\,\mathrm{d}\rho\,\mathrm{d}\boldsymbol{\sigma} \tag{174}$$

$$\cong \int \rho^{-m/2}\,\exp\Big\{\frac{\beta}{\rho m}\Big\|\sqrt{m\mathsf{c}_i}\boldsymbol{z}_i + m\boldsymbol{\sigma}_0\Big\|_2^2 + \beta m\rho + O_m(1)\Big\}\,\mathrm{d}\rho \tag{175}$$

$$\cong \int \exp\Big\{\beta m\, S_i(\rho) + \frac{2\beta}{\rho}\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle + \frac{\beta\mathsf{c}_i}{\rho}(\|\boldsymbol{z}_i\|_2^2 - m) + O_m(1)\Big\}\mathrm{d}\rho\,, \tag{176}$$

where

$$S_i(\rho) \equiv \rho + \frac{1}{\rho}\big(\mathsf{c}_i+1\big) - \frac{1}{2\beta}\log\rho\,. \tag{177}$$

Here the indegral over ρ runs along the imaginary axis in the complex plane, from $-i\infty$ and $+i\infty$.

Note, for $\boldsymbol{\sigma}_0$ uniformly random on the unit sphere, the term $(2\beta m/\rho)\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle$ is of order $\sqrt{m}$, i.e. of lower order with respect to the term including $S(\,\cdot\,)$. Also, the term $(\beta m\mathsf{c}_i/\rho)(\|\boldsymbol{z}_i\|_2^2-1)$ is of order $\sqrt{m}$ and does not depend on $\boldsymbol{\sigma}_0$. Hence, up to an additional $O_m(1)$ term, we can reabsorb this in the normalization constant. We therefore get

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \int \exp\Big\{\beta m\, S_i(\rho) + \frac{2\beta}{\rho}\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle + O_m(1)\Big\}\mathrm{d}\rho\,. \tag{178}$$

We next perform integration over ρ by the saddle point method. Since $\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle = O(m^{-1/2})$ and $S_i(\rho) = O(1)$, the saddle point is given by the stationary equation $S_i'(\rho) = 0$. The saddle point $\rho_{i,*}$ lies on the real axis and is a minimum along the real axis but a maximum with respect to the imaginary direction, i.e.,

$$\rho_{i,*}(\beta) = \arg\min_{\rho\in\mathbb{R}_+} S_i(\rho)\,. \tag{179}$$

By Cauchy's theorem, we can deform the contour of integral to pass the saddle point along the imaginary direction. This in fact corresponds to the path that descents most steeply from the saddle point. The integral is dominated by $\rho = \rho_{i,*} + O(m^{-1/2})$ and hence,

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \exp\Big\{\frac{2\beta}{\rho_{i,*}}\sqrt{m\mathsf{c}_i}\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle + O_m(1)\Big\}\,. \tag{180}$$

While this expression for $\widehat{\nu}_i(\boldsymbol{\sigma}_0)$ is accurate when $\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle$ is small, it breaks down for large $\langle \boldsymbol{z}_i,\boldsymbol{\sigma}_0\rangle$. In Section 5.5 we will discuss the regimes of validity of this approximation. Namely, we expect it to be accurate for d large and for d close to one.

Substituting in Eq. (170), we get the recursion

$$\mathsf{c}_0 = \sum_{i=1}^{k}\frac{\mathsf{c}_i}{\rho_{i,*}^2}\,, \tag{181}$$

$$\rho_{i,*}^2 = \big(\mathsf{c}_i+1\big) + \frac{1}{2\beta}\,\rho_{i,*}\,. \tag{182}$$

In particular, as $\beta \to \infty$, we get the simple equation

$$\mathsf{c}_0 = \sum_{i=1}^{k} \frac{\mathsf{c}_i}{1+\mathsf{c}_i}\,. \tag{183}$$

Note that $\mathsf{c}_0, \mathsf{c}_1, \dots, \mathsf{c}_k$ are random variables, because of the randomness in the underlying limiting tree, which is Galton-Watson tree with Poisson offspring distribution. We get

$$\mathsf{c} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{L} \frac{\mathsf{c}_i}{1+\mathsf{c}_i}\,, \tag{184}$$

where $L \sim \mathrm{Poisson}(d)$, $d = (a+b)/2$, and $\mathsf{c}_1, \dots, \mathsf{c}_L$ are i.i.d. copies of c. We are interested in solutions supported on $\mathbb{R}_{\geq 0}$, i.e. such that $\mathbb{P}(\mathsf{c} \geq 0) = 1$.

This recursion is connected to potential theory on Galton-Watson trees, see e.g. [LPP97, LP13]. The next proposition establishes a few basic facts about its solutions. For the readers' convenience, we provide a proof of the simplest statements, referring to [LPP97] for the other facts. Given two random variables X, Y, we write $X \preceq Y$ if Y dominates stochastically X, i.e. if there exists a coupling of X, Y such that $\mathbb{P}(X \leq Y) = 1$.

Proposition 5.2 ([LPP97])**.** *There exists random variables c_0, c_* supported on $\mathbb{R}_{\geq 0}$, that solve Eq. (184) and such that:*

1. *$\mathsf{c}_0 = 0$ almost surely.*
2. *$\mathsf{c}_* \preceq L$, $L \sim \mathrm{Poisson}(d)$.*
3. *For $d \leq 1$, $\mathsf{c}_* = \mathsf{c}_0 = 0$ identically. Hence the distributional equation (184) has a unique solution.*
4. *For $d > 1$, $\mathsf{c}_* > 0$ with positive probability and further equation (184) admits no other solution than c_0, c_*.*

Proof. Let $\mathfrak{P}$ denote the space of probability measures over $[0, \infty]$, and $\mathsf{T}_d : \mathfrak{P} \to \mathfrak{P}$ the map defined by the right-hand side of Eq. (184). Namely $\mathsf{T}_d(\mu)$ is the probability distribution of the right-hand side of Eq. (184) when $\mathsf{c}_i \sim_{i.i.d.} \mu$. Notice that this is well defined on the extended real line because the summands are non-negative. It is immediate to see that this map is monotone, i.e.

$$\mu_1 \preceq \mu_2 \;\Rightarrow\; \mathsf{T}_d(\mu_1) \preceq \mathsf{T}_d(\mu_2)\,. \tag{185}$$

We define $\mathsf{c}_0 = 0$ identically and $\mathsf{c}_* \sim \mu^+$, where $\mu^+ = \lim_{\ell\to\infty} \mathsf{T}_d^{\ell}(\delta_{+\infty})$ where "lim" refers to the weak limit, which exists by monotonicity. Points 1, 2 follows from an immediate monotonicity argument (see e.g. [AB05]).

For point 3, note that by Jensen inequality

$$\mathbb{E}\mathsf{c} \leq \frac{d\mathbb{E}\mathsf{c}}{1+\mathbb{E}\mathsf{c}}\,, \tag{186}$$

which implies $\mathbb{E}\mathsf{c} = 0$, and hence $\mathsf{c} = 0$ identically.

Point 4 is just Theorem 4.1 in [LPP97]. □

The next proposition establishes an appealing interpretation of the random variable c_*. Again, this puts together results of [Lyo90] and [LPP97]. We give here a proof of this connection for the readers' convenience. We refer to [LP13] for further background on discrete potential theory (electrical networks) and trees.

Proposition 5.3 ([Lyo90, LPP97])**.** *Let $\mathcal{T}$ be a rooted Galton-Watson tree with offspring distribution* Poisson(d)*, and consider the associated (infinite) electric network, whereby each edge of $\mathcal{T}$ is replaced by a conductor with unit resistance. Let $\mathcal{C}(0 \leftrightarrow \ell)$ be the conductance between the root and vertices at level ℓ (when nodes at level ℓ are connected together). Let $\mathcal{C}(0 \leftrightarrow \infty) \equiv \lim_{\ell\to\infty} \mathcal{C}(0 \leftrightarrow \ell)$. Then:*

$$\mathsf{c}_* \stackrel{\mathrm{d}}{=} \mathcal{C}(0 \leftrightarrow \infty)\,. \tag{187}$$

In particular, if $d > 1$, then $\mathsf{c}_* > 0$ with positive probability.

Proof. Let 0 denote the root of $\mathcal{T}$, and $\{1, 2, \dots, k\}$ be its children. Let the conductance between the root and level ℓ on $\mathcal{T}$ be $\gamma_0(\ell)$,. Also, let $\mathcal{T}_1, \dots \mathcal{T}_k$ denote the subtrees rooted at $1, \dots, k$. Let the conductance between the root and level ℓ on $\mathcal{T}_i$ (i.e. between vertex i and vertices at distance ℓ from i on $\mathcal{T}_i$) be denoted by $\gamma_i(\ell)$. Let $\widehat{\gamma}_i(\ell)$, $i \in \{1, \dots, k\}$ be the conductance of the tree obtained from $\mathcal{T}$ by removing edges $\{(0, j) : j \in [k] \setminus \{k\}\}$. Equivalently $\widehat{\gamma}_i(\ell)$ is the conductance of the tree obtained from $\mathcal{T}_i$ by connecting the root of $\mathcal{T}_i$ (vertex i) to 0 and moving the root to 0. Since the resistance of a series is the sum of resistances of each component, we have

$$\frac{1}{\widehat{\gamma}_i(\ell+1)} = 1 + \frac{1}{\gamma_i(\ell)}\,. \tag{188}$$

This is of course equivalent to $\widehat{\gamma}_i(\ell+1) = \gamma_i(\ell)/(1+\gamma_i(\ell))$.

Now since the conductance of several resistances in parallel is equal to the sum of the conductances of the components, we get

$$\gamma_0(\ell+1) = \sum_{i=1}^{k} \widehat{\gamma}_i(\ell+1) = \sum_{i=1}^{k} \frac{\gamma_i(\ell)}{(1+\gamma_i(\ell))}\,, \tag{189}$$

with boundary condition $\gamma_0(0) = \infty$. Notice that this coincides with the recursion for c^ℓ, included the boundary condition, thus proving our claim.

It follows from Theorem [Lyo90, Theorem 4.3] and [Lyo90, Proposition 6.4] that $\mathsf{c}_* > 0$ with positive probability whenever $d > 1$. □

We will hereafter consider the case $d > 1$ and focus on the maximal solution c_*.

We first study the behavior of c_* for large d limit. When $d \to \infty$, by the law of large numbers the right-hand of Eq. (184) concentrates around a deterministic value, and hence so does c. In order to characterize this value for large d, we write $\mathsf{c} = \overline{\mathsf{c}} + \Delta\mathsf{c}$ where $\mathbb{E}(\Delta\mathsf{c}) = 0$ and $\overline{\mathsf{c}} = \mathbb{E}(\mathsf{c})$ is deterministic. Further, we expect $\Delta\mathsf{c} = \Theta(\sqrt{d})$ and $\overline{\mathsf{c}} = \Theta(d)$. Expanding Eq. (184), we get

$$\overline{\mathsf{c}} + \Delta\mathsf{c} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{L} \Big\{ \frac{\overline{\mathsf{c}}}{1+\overline{\mathsf{c}}} + \frac{\Delta\mathsf{c}_i}{(1+\overline{\mathsf{c}})^2} - \frac{\Delta\mathsf{c}_i^2}{(1+\overline{\mathsf{c}})^3} + \frac{\Delta\mathsf{c}_i^3}{(1+\overline{\mathsf{c}})^4} + O(d^{-3}) \Big\} \tag{190}$$

$$= \frac{L\,\overline{\mathsf{c}}}{1+\overline{\mathsf{c}}} + \frac{d\,(\mathbb{E}(\Delta\mathsf{c}^2))^{1/2}}{(1+\overline{\mathsf{c}})^2}\, Z_1 - \frac{d\,\mathbb{E}(\Delta\mathsf{c}^2)}{(1+\overline{\mathsf{c}})^3} + O(d^{-3/2})\,, \tag{191}$$

where $Z_1 \sim \mathsf{N}(0,1)$ is independent of $L \sim \text{Poisson}(d)$. Note that we used central limit theorem and law of large numbers in obtaining (191).

Taking expectation we get

$$\overline{\mathsf{c}} = \frac{d\,\overline{\mathsf{c}}}{1+\overline{\mathsf{c}}} - \frac{d\,\mathbb{E}(\Delta\mathsf{c}^2)}{(1+\overline{\mathsf{c}})^3} + O(d^{-3/2})\,, \tag{192}$$

whence

$$\overline{\mathsf{c}} = d - 1 - \frac{\mathbb{E}(\Delta\mathsf{c}^2)}{d^2} + O(d^{-3/2})\,. \tag{193}$$

On the other hand, taking the variance, we obtain

$$\mathbb{E}(\Delta\mathsf{c}^2) = d\Big(\frac{\overline{\mathsf{c}}}{1+\overline{\mathsf{c}}}\Big)^2 + \frac{d^2\,\mathbb{E}(\Delta\mathsf{c}^2)}{(1+\overline{\mathsf{c}})^4} + O(d^{-3})\,. \tag{194}$$

Therefore,

$$\mathbb{E}(\Delta\mathsf{c}^2) = d - 2 + O(d^{-1})\,. \tag{195}$$

Using equation (195) in equation (193), we get

$$\overline{\mathsf{c}} = d - 1 - \frac{1}{d} + O(d^{-3/2})\,. \tag{196}$$

Summarizing, we found that

$$\mathsf{c}_* = d - 1 - \frac{1}{d} + \sqrt{d-2}\;Z + O(d^{-3/2})\,, \tag{197}$$

for $Z \sim \mathsf{N}(0,1)$.

5.2 Linear stability of the symmetric phase and critical point

We next study the stability of the symmetric solution (172). We break the $\mathcal{O}(m)$ symmetry by letting

$$\nu_i(\mathrm{d}\boldsymbol{\sigma}_i) \cong \exp\Big\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\tau}_i\rangle + 2\beta m h_i\, s_i + O_m(1)\Big\}\,\delta\Big(\|\boldsymbol{\tau}_i\|_2^2 + s_i^2 - 1\Big)\,\mathrm{d}s_i \mathrm{d}^{m-1}\boldsymbol{\tau}_i\,, \tag{198}$$

where $\boldsymbol{\sigma}_i = (s_i, \boldsymbol{\tau}_i)$ and $\boldsymbol{z}_i \sim \mathsf{N}(0, \mathrm{I}_{m-1})$ is $(m-1)$-dimensional. Note that each coordinate of $\boldsymbol{z}_i$ is of order 1, and is multiplied by a factor $\sqrt{m}$ in the above expression. We will consider $h_i \lll 1$, and expand all expressions to linear order in h_i.

Proceeding as in the symmetric phase, we get

$$\begin{aligned}
\widehat{\nu}_i(\boldsymbol{\sigma}_0) &\cong -j \int \exp\Big\{2\beta m s_0 s + \beta m\rho - \beta m\rho s^2 + 2\beta m h_i s \\
&\qquad + 2\beta m\langle\boldsymbol{\tau}_0,\boldsymbol{\tau}\rangle + 2\beta\sqrt{m\mathsf{c}_i}\langle \boldsymbol{z}_i,\boldsymbol{\tau}\rangle - \beta m\rho\|\boldsymbol{\tau}\|_2^2 + O_m(1)\Big\}\,\mathrm{d}\rho\,\mathrm{d}^{m-1}\boldsymbol{\tau}\,\mathrm{d}s \\
&\cong -j\int \rho^{-m/2}\,\exp\Big\{\frac{\beta}{\rho m}\Big\|\sqrt{m\mathsf{c}_i}\boldsymbol{z}_i + m\boldsymbol{\tau}_0\Big\|_2^2 + \frac{\beta m}{\rho}\Big(h_i + s_0\Big)^2 + \beta m\rho + O_m(1)\Big\}\,\mathrm{d}\rho \\
&\cong -j\int \exp\Big\{\beta m\, S_i(\rho) + \frac{2\beta}{\rho}\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i,\boldsymbol{\tau}_0\rangle + \frac{2\beta m}{\rho}\,h_i s_0 + \frac{\beta\mathsf{c}_i}{\rho}(\|\boldsymbol{z}_i\|_2^2 - m) + \frac{\beta m}{\rho}h_i^2 + O_m(1)\Big\}\,\mathrm{d}\rho\,,
\end{aligned} \tag{199}$$

where $S_i(\rho)$ is defined as in the symmetric phase, namely

$$S_i(\rho) = \rho + \frac{1}{\rho}\big(\mathsf{c}_i + 1\big) - \frac{1}{2\beta}\log\rho\,. \tag{200}$$

We perform the integral by saddle point, around $\rho_{i,*}(\beta) = \arg\min_{\rho\in\mathbb{R}_+} S_i(\rho)$. Proceeding as in the symmetric case, and neglecting terms quadratic in h_i, we have

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \exp\Big\{\frac{2\beta m}{\rho_{i,*}}\,h_i s_0 + \frac{2\beta}{\rho_{i,*}}\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\tau}_0\rangle + O_m(1) + O(h^2)\Big\}\,. \tag{201}$$

Substituting in Eq. (170), we obtain the equations

$$\mathsf{c}_0 = \sum_{i=1}^{k}\frac{\mathsf{c}_i}{\rho_{i,*}^2}\,, \qquad h_0 = \sum_{i=1}^{k}\frac{h_i}{\rho_{i,*}}\,, \tag{202}$$

which simplify at zero temperature to

$$\mathsf{c}_0 = \sum_{i=1}^{k}\frac{\mathsf{c}_i}{1+\mathsf{c}_i}\,, \qquad h_0 = \sum_{i=1}^{k}\frac{h_i}{\sqrt{1+\mathsf{c}_i}}\,. \tag{203}$$

Recall that the graph G_n converges locally to a two-types Galton-Watson tree, whereby each vertex has Poisson$(a/2)$ vertices of the same type, and Poisson$(b/2)$ vertices of the opposite type. We look for solutions that break the symmetry $+1 \leftrightarrow -1$. If (c_i, h_i) is the pair of random variables introduced above, for vertex i, we therefore assume $(\mathsf{c}_{i(+)}, h_{i(+)}) \stackrel{\mathrm{d}}{=} (\mathsf{c}_{i(-)}, -h_{i(-)})$ for $i(+) \in V_+$, $i(-) \in V_-$. This leads to the following distributional recursion for the sequence of random vectors $\{(\mathsf{c}^t, h^t)\}_{t\ge 0}$:

$$\big(\mathsf{c}^{t+1}; h^{t+1}\big) \stackrel{\mathrm{d}}{=} \Big(\sum_{i=1}^{L_+ + L_-}\frac{\mathsf{c}_i^t}{1+\mathsf{c}_i^t}; \sum_{i=1}^{L_+ + L_-}\frac{s_i h_i^t}{\sqrt{1+\mathsf{c}_i^t}}\Big)\,, \tag{204}$$

where $L_+ \sim$ Poisson$(a/2)$, $L_- \sim$ Poisson$(b/2)$, $s_1, \dots, s_{L_+} = +1$, $s_{L_+ +1}, \dots, s_{L_+ + L_-} = -1$, and $\{(\mathsf{c}_i^t, h_i^t)\}$ are i.i.d. copies of (c^t, h^t)

Let us pause for two important remarks:

1. The recursion (204) is invariant under the rescaling $h^t \to a\,h^t$ for $a \in \mathbb{R}$. Hence, only properties that are invariant under thus rescaling are meaningful.

2. It admits the fixed point $(\mathsf{c}^t, h^t) \stackrel{\mathrm{d}}{=} (\mathsf{c}_*, \sqrt{\mathsf{c}_*}\,Z)$ where c_* is the distributional fixed point of the symmetric phase, constructed in the previous section, and $Z \sim \mathsf{N}(0,1)$ is independent of c_*. This is a fixed point[2] that corresponds to the symmetric phase, and does not break the $+1 \leftrightarrow -1$ symmetry.

Therefore, in order to investigate stability, we initialize the above recursion in a way that breaks the symmetry, $(\mathsf{c}^0, h^0) = (\infty, 1)$. Note that by monotonicity property (185), starting with $\mathsf{c}^0 = \infty$,

[2]Indeed, by the scaling invariance, $(\mathsf{c}^t, h^t) \stackrel{\mathrm{d}}{=} (\mathsf{c}_*, a\sqrt{\mathsf{c}_*}\,Z)$ is a fixed point for any fixed scale factor $a \in \mathbb{R}$.

we have $\mathsf{c}^t \stackrel{\mathrm{d}}{\Rightarrow} \mathsf{c}_*$. We ask whether this perturbation grows, by computing the exponential growth rate

$$G_\alpha(d,\lambda) \equiv \lim \inf_{t\to\infty} \frac{1}{t\alpha} \log \mathbb{E}(|h^t|^\alpha)\,, \tag{205}$$

where $d = (a+b)/2$, and $\lambda = (a-b)/\sqrt{2(a+b)}$ parametrize the model. We define the critical point as the smallest λ such that the growth rate is strictly positive:

$$\tilde{\lambda}_c^{\mathrm{SDP}}(d) \equiv \inf\big\{\lambda \in [0,\sqrt{d}] :\ G_2(d,\lambda) > 0\big\}\,. \tag{206}$$

Notice that in the definition we used the second moment, i.e. set $\alpha = 2$. However, the result appear to be insensitive to the choice of α. In the next section we will discuss the numerical solution of the above distributional equations and our analytical prediction for $\lambda_c^{\mathrm{SDP}}(d)$.

In the rest of this section we analyze the behavior of $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$ for large d. Along the way, we analyze the behavior of perturbation h^t, which in turn clarifies why the critical point $\tilde{\lambda}_c^{\mathrm{SDP}}$ is defined based on the *exponential growth rate* of the perturbation.

For the sake of simplicity we assume the initialization $(\mathsf{c}^0, h^0) \stackrel{\mathrm{d}}{=} (\mathsf{c}_*, 1)$. Since with initialization $\mathsf{c}^0 = \infty$ we have $\mathsf{c}^t \stackrel{\mathrm{d}}{\Rightarrow} \mathsf{c}_*$, this should be equivalent to our initialization $(\mathsf{c}^0, h^0) = (\infty, 1)$. We let $\mathsf{c}^t = \overline{\mathsf{c}} + \Delta\mathsf{c}^t$, $\overline{\mathsf{c}} = \mathbb{E}(\mathsf{c}_*) = \mathbb{E}(\mathsf{c}^t)$ as in the previous section. Note that although c^t is distributed as per c_*, joint distribution of (c^t, h^t) varies over time and so we make the iteration number explicit in $\Delta\mathsf{c}^t$.

We start by taking expectation of Eq. (204).

$$\mathbb{E}(h^{t+1}) = \frac{a-b}{2}\,\mathbb{E}\Big\{\frac{h^t}{\sqrt{1+\mathsf{c}^t}}\Big\} \tag{207}$$

$$= \frac{a-b}{2}\,\mathbb{E}\Big\{h^t\Big[\frac{1}{(1+\overline{\mathsf{c}})^{1/2}} - \frac{\Delta\mathsf{c}^t}{2(1+\overline{\mathsf{c}})^{3/2}} + \frac{3(\Delta\mathsf{c}^t)^2}{8(1+\overline{\mathsf{c}})^{5/2}} + O(d^{-2})\Big]\Big\}\,. \tag{208}$$

By taking the covariance of h^t and c^t, we obtain

$$\mathbb{E}(\Delta\mathsf{c}^t h^t) = \mathrm{Cov}(\mathsf{c}^t; h^t) = \frac{a-b}{2}\,\mathbb{E}\Big\{\frac{h^{t-1}\mathsf{c}^{t-1}}{(1+\mathsf{c}^{t-1})^{3/2}}\Big\} \tag{209}$$

$$= \frac{a-b}{2}\frac{\overline{\mathsf{c}}}{(1+\overline{\mathsf{c}})^{3/2}}\mathbb{E}(h^{t-1})\big(1+O(d^{-1})\big)\,. \tag{210}$$

Further we have

$$\mathbb{E}\{h^t(\Delta\mathsf{c}^t)^2\} = \mathbb{E}\{h^t\}\mathbb{E}\{(\Delta\mathsf{c}^t)^2\}\big(1+O(d^{-1/2})\big) \tag{211}$$

$$= d\,\mathbb{E}\{h^t\}\big(1+O(d^{-1/2})\big)\,. \tag{212}$$

Substituting in Eq. (208), using $a-b = 2\lambda\sqrt{d}$, and setting $\overline{h}^t = \mathbb{E}\{h^t\}$, we get

$$\overline{h}^{t+1} = \lambda\sqrt{d}\Big\{\frac{1}{(1+\overline{\mathsf{c}})^{1/2}}\,\overline{h}^t - \frac{\overline{\mathsf{c}}\lambda\sqrt{d}}{2(1+\overline{\mathsf{c}})^3}\,\overline{h}^{t-1} + \frac{3d}{8(1+\overline{\mathsf{c}})^{5/2}}\,\overline{h}^t + O(d^{-2}\overline{h}^t)\Big\} \tag{213}$$

$$= \lambda\Big\{\Big(1+\frac{3}{8d}\Big)\overline{h}^t - \frac{1}{2d}\,\overline{h}^{t-1} + O(d^{-3/2}\overline{h}^t)\Big\}\,, \tag{214}$$

Hence, to this order

$$\tilde{\lambda}_c^{\rm SDP}(d) = \rho_{\rm sp}(M_d)^{-1} + O(d^{-3/2})\,, \tag{215}$$

$$M_d \equiv \begin{bmatrix} 1+3/(8d) & -1/2d \\ 1 & 0 \end{bmatrix}. \tag{216}$$

where $\rho_{\rm sp}(\,\cdot\,)$ denotes the spectral radius of a matrix. A simple calculation yields

$$\tilde{\lambda}_c^{\rm SDP}(d) = 1 + \frac{1}{8d} + O(d^{-3/2})\,. \tag{217}$$

5.3 Numerical solution of the distributional recursions

We solved numerically the distributional recursions (184), (204) through a sampling algorithm that is known as 'population dynamics' within spin glass theory [MP01]. The algorithm updates a sample that, at iteration t, is meant to be an approximately iid samples with the same law as the one defined by the distributional equation, at iteration t. For concreteness, we define the algorithm here in the case of the iteration corresponding to Eq. (184):

$$\mathsf{c}^{t+1} \stackrel{\rm d}{=} \sum_{i=1}^{L} \frac{\mathsf{c}_i^t}{1+\mathsf{c}_i^t}\,. \tag{218}$$

The distribution of c^t will be approximated by a sample $\underline{\mathsf{c}}^t \equiv (\mathsf{c}_1^t, \mathsf{c}_2^t, \dots, \mathsf{c}_N^t)$ (we represent this by a vector but ordering is irrelevant).

Algorithm 1 Population dynamics algorithm

Input: Sample size N; Number of iterations T
Output: Samples $\underline{\mathsf{c}}^1, \underline{\mathsf{c}}^2, \dots, \underline{\mathsf{c}}^T$

1: $\underline{\mathsf{c}}^1 \leftarrow (L_1, L_2, \dots, L_N)$, with $L_i \sim_{i.i.d.}$ Poisson(d)
2: **for** $t = 1, 2, \dots, T = 1$ **do**
3: $\quad \underline{\mathsf{c}}^{t+1} \leftarrow (\,)$
4: $\quad$ **for** $i = 1, 2, \dots, N$ **do**
5: $\quad\quad$ Generate $L \sim$ Poisson(d)
6: $\quad\quad$ Generate $i(1), \dots, i(L) \sim_{i.i.d.}$ $\mathsf{Unif}(\{1, 2, \dots, N\})$
7: $\quad\quad$ Compute

$$\mathsf{c}^{\rm new} = \sum_{a=1}^{L} \frac{\mathsf{c}_{i(a)}^t}{1+\mathsf{c}_{i(a)}^t}\,. \tag{219}$$

8: $\quad\quad$ Set $\underline{\mathsf{c}}^{t+1} \leftarrow [\underline{\mathsf{c}}^{t+1} | \mathsf{c}^{\rm new}]$ (append entry $\mathsf{c}^{\rm new}$ to vector $\underline{\mathsf{c}}^{t+1}$)
9: **return** $\underline{\mathsf{c}}^1, \dots, \underline{\mathsf{c}}^T$

The notation $[a|b]$ in the step 8 of the algorithm denotes appending element b to vector a. Note that with initial point $c_0 = \infty$, we have $c_1 = L \sim$ Poisson(d). In the population dynamic algorithm we start from c^1.

As an illustration, Figure 1 presents the results of some small-scale calculations using this algorithm.

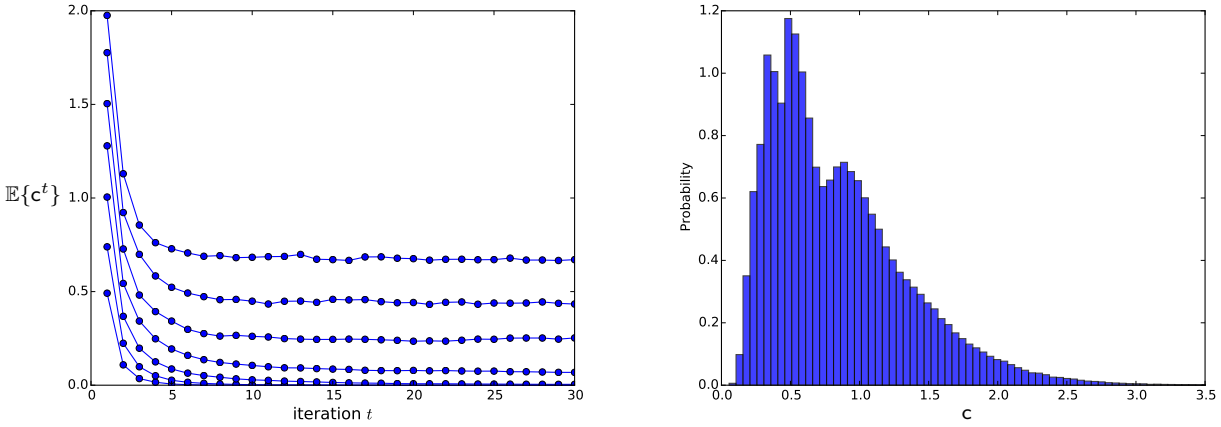


Figure 1: Solution of the recursive distributional equation (184) using the population dynamics algorithm. Left frame: evolution of the mean $\mathbb{E}\{\mathsf{c}^t\}$ versus the number of iterations t, as estimated by the algorithm. Various curves refer (from bottom to top) to $d = 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2$. Here sample size is $N = 5 \cdot 10^3$. Right frame: histogram of the samples at convergence, for $d = 2$. Here $N = 5 \cdot 10^3$.

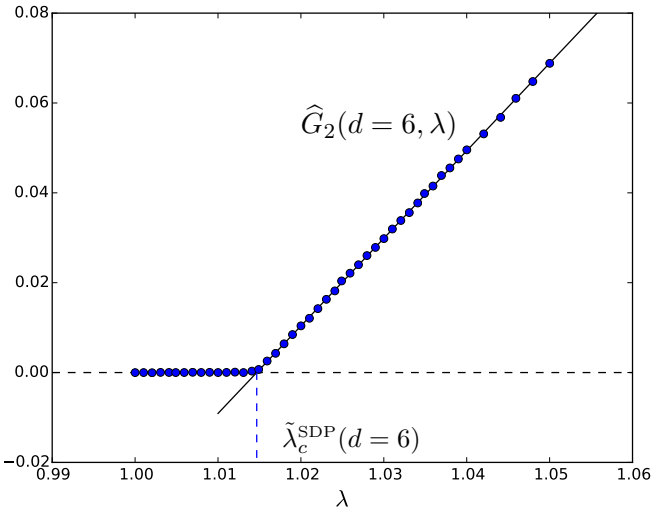


Figure 2: Local stability of the $\mathcal{O}(m)$ symetric phase. We evaluate $G_2(d, \lambda)$ defined by Eq. (205) using the population dynamics algorithm. Data here reder to average degree $d = 6$. The phase transition point $\tilde{\lambda}_c^{\rm SDP}(d)$ (within the vectorial ansatz) is determined by a local linear fit to the estimated $G_2(d, \lambda)$, when it is significantly larger than 0.

We used the obvious modification of this algorithm to implement the recursion (204), whereby a population is now formed of pairs $(\mathsf{c}_1^t, h_1^t), \dots (\mathsf{c}_N^t, h_N^t)$. An important difference is that the overall

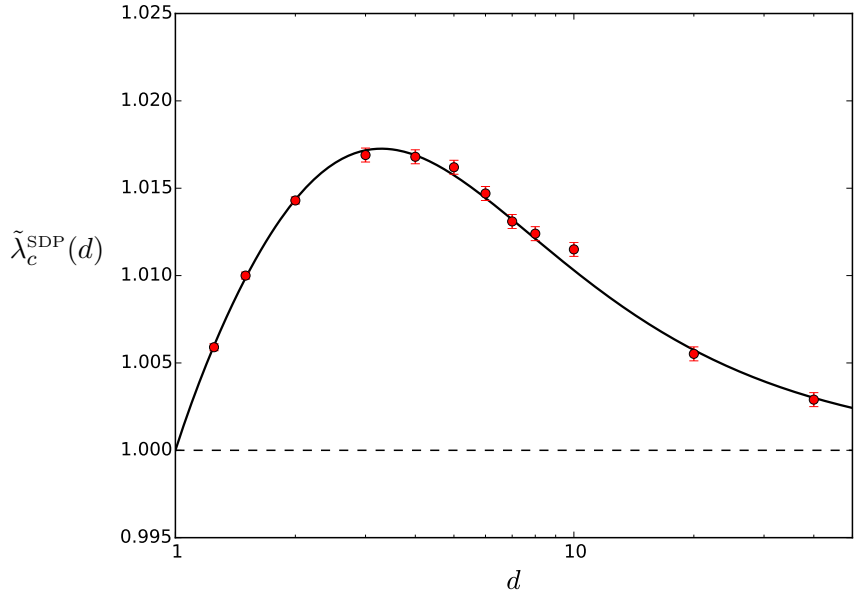


Figure 3: Theoretical prediction of the phase transition location for detecting hidden partition in sparse graphs using semidefinite programming. here the phase transition location is reported in the rescaled variable $\lambda = (a-b)/\sqrt{2(a+b)}$ as a function of the average degree $d = (a+b)/2$. Points with error bars correspond to the definition (206), evaluated numerically by approximating the recursion (204) with the population dynamics algorithm. The continuous curve is a rational fit to the data, constrained at $d = 1$, and for large d.

scaling of the h_i^t is immaterial. We hence normalize them at each iteration as follows

$$(h_i^t)^{\rm new} = \frac{h_i^t}{\sqrt{M_t}}\,, \qquad M_t \equiv \frac{1}{N}\sum_{i=1}^{N}(h_i^t)^2\,. \tag{220}$$

The normalization constant M_t also allow us to estimate $G_2(d,\lambda)$, namely

$$\widehat{G}_2(d,\lambda) \equiv \frac{1}{2(t_{\rm max}-t_{\rm min}+1)}\sum_{t=t_{\rm min}}^{t_{\rm max}}\log M_t\,. \tag{221}$$

Figure 2 presents the typical results of this calculation, using $N = 10^7$, $t_{\rm min} = 100$, $t_{\rm max} = 400$, at average degree $d = 6$.

The behavior in Figure 2 is generic. For small λ, the estimate $\widehat{G}_2(d,\lambda)$ is statistically indistinguishable from 0. Above a critical point, that we identify with $\tilde{\lambda}_c^{\rm SDP}(d)$, $\widehat{G}_2(d,\lambda)$ is strictly positive, and essentially linear in λ, close to $\tilde{\lambda}_c^{\rm SDP}(d)$. In order to estimate $\tilde{\lambda}_c^{\rm SDP}(d)$ we use a local linear fit, with parameters g_0, g_1

$$\widehat{G}_2(d,\lambda) = g_0 + g_1\lambda\,, \tag{222}$$

d	$\tilde{\lambda}_c^{\rm SDP}(d)$
1.25	1.0059 ± 0.0002
1.5	1.0100 ± 0.0002
2	1.0143 ± 0.0002
3	1.0169 ± 0.0004
4	1.0168 ± 0.0004
5	1.0162 ± 0.0004
6	1.0147 ± 0.0004
7	1.0131 ± 0.0004
8	1.0124 ± 0.0004
10	1.0115 ± 0.0004
20	1.0055 ± 0.0004
40	1.0029 ± 0.0004

Table 1: Numerical determination of the critical point $\tilde{\lambda}_c^{\rm SDP}(d)$ (within the vectorial ansatz) for a few values of d.

and set $\tilde{\lambda}_c^{\rm SDP}(d) = -\hat{g}_0/\hat{g}_1$. In the fit we include only the values of λ such that $\widehat{G}_2(d,\lambda)$ is significantly different from 0. Note that the linear fit here is in a local neighborhood of $\lambda = 1$. By equation (214), it is easy to see that for large fixed d, $G_2(d,\lambda)$ is logarithmic in λ.

The resulting values of $\tilde{\lambda}_c^{\rm SDP}(d)$ are plotted in Figure 3, together with statistical errors. We report the same values in Table 1 (note that the value for $d = 10$ appears to be an outlier).

In order to interpolate these results, we fitted a rational function, with the correct asymptotic behavior at $d = 1$ (namely, $\tilde{\lambda}_c^{\rm SDP}(d = 1) = 1$) and at $d \to \infty$ (namely, $\tilde{\lambda}_c^{\rm SDP}(d) = 1+1/(8d)+o(d^{-1})$). This fit is reported as a continuous line in Figure 2, and is given by

$$\lambda_c^{\rm FIT}(d) = 1 + \frac{a_1(d-1) + a_2(d-1)^2}{1 + b_1(d-1) + (8a_1 + 19.5a_2)(d-1)^2 + 8a_2(d-1)^3}\,, \tag{223}$$

with parameters

$$a_1 = 0.0307569\,, \tag{224}$$
$$a_2 = 0.030035\,, \tag{225}$$
$$b_1 = 2.16454\,. \tag{226}$$

This is also the curve reported in the main text.

5.4 The recovery phase (broken $\mathcal{O}(m)$ symmetry)

In this section we describe an approximate solution of the cavity equations within the region $\lambda > \lambda_c^{\rm SDP}(d)$. In this regime the SDP estimator has positive correlation with the ground truth (in the $n \to \infty$ limit). We will work within a generalization of the vectorial ansatz introduced in Section 5.2. Namely, we look for a solution which breaks the $\mathcal{O}(m)$ symmetry to $\mathcal{O}(m-1)$ along the first direction, as follows. Letting $\boldsymbol{\sigma}_i = (s_i, \boldsymbol{\tau}_i)$, $s_i \in \mathbb{R}$, $\boldsymbol{\tau}_i \in \mathbb{R}^{m-1}$, we adopt the ansatz

$$\nu_i(\mathrm{d}\boldsymbol{\sigma}_i) \cong \exp\Big\{2\beta\sqrt{m\mathsf{c}_i}\,\langle \boldsymbol{z}_i, \boldsymbol{\tau}_i\rangle + 2\beta m h_i\, s_i - \beta m r_i s_i^2 + O_m(1)\Big\}\,\delta\Big(s_i^2 + \|\boldsymbol{\tau}_i\|_2^2 - 1\Big)\,\mathrm{d}^m\boldsymbol{\sigma}_i\,, \tag{227}$$

where we will assume $\boldsymbol{z}_i \sim \mathsf{N}(0, \mathrm{I}_{m-1})$. In the following, we let $\mathsf{P}_1 = \boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}$ be the projector along the first direction and $\mathsf{P}_1^{\perp} = \mathrm{I} - \mathsf{P}_1$ denote the orthogonal projector.

Recalling the cavity equations (170), and using the Fourier representation of the delta function, we get

$$\widehat{\nu}_i(\boldsymbol{\sigma}_0) \cong \int \exp\Big\{\beta m\rho - \beta m\big((\rho + r_i)s_i^2 + \rho\|\boldsymbol{\tau}_i\|_2^2\big) \tag{228}$$
$$+ 2\beta m\big(s_0 s_i + \langle\boldsymbol{\tau}_0, \boldsymbol{\tau}_i\rangle\big) + 2\beta m h_i s_i + 2\beta\sqrt{m\mathsf{c}_i}\langle\boldsymbol{z}_i, \boldsymbol{\tau}_i\rangle\Big\}\,\mathrm{d}\rho\,\mathrm{d}^{m-1}\boldsymbol{\tau}_i\mathrm{d}s_i$$
$$\cong \int \rho^{-m/2}\,\exp\Big\{\beta m\rho + \frac{(2\beta m s_0 + 2\beta m h_i)^2}{4\beta m(\rho + r_i)} + \frac{\|2\beta m\boldsymbol{\tau}_0 + 2\beta\sqrt{m\mathsf{c}_i}\boldsymbol{z}_i\|_2^2}{4\beta m\rho}\Big\}\,\mathrm{d}\rho \tag{229}$$
$$\cong \int \exp\Big\{\beta m S_i(\rho; s_0) + \frac{2\beta\sqrt{m\mathsf{c}_i}}{\rho}\langle\boldsymbol{z}_i, \boldsymbol{\tau}_0\rangle\Big\}\,\mathrm{d}\rho\,, \tag{230}$$

where

$$S_i(\rho; s_0) = S_{i,0}(\rho) + S_{i,1}(\rho)\,s_0 + \frac{1}{2}S_{i,2}(\rho)\,s_0^2\,, \tag{231}$$
$$S_{i,0}(\rho) \equiv \rho - \frac{1}{2\beta}\log\rho + \frac{h_i^2}{\rho + r_i} + \frac{1 + \mathsf{c}_i}{\rho}\,, \tag{232}$$
$$S_{i,1}(\rho) = \frac{2h_i}{\rho + r_i}\,, \tag{233}$$
$$S_{i,2}(\rho) = \frac{2}{\rho + r_i} - \frac{2}{\rho}\,. \tag{234}$$

where we approximated $\|\boldsymbol{z}_i\|_2^2/m \approx 1$, and used the identity $\|\boldsymbol{\tau}_0\|_2^2 = 1 - s_0^2$. In the $m \to \infty$ limit, we approximate the integral over ρ by its saddle point. In order to obtain a set of equations for the parameters of the ansatz (227), we will expand the exponent to second order in s_0. The saddle point location is given by

$$\rho_{i,*}(s_0) \equiv \arg\min_{\rho\in\mathbb{R}_{>0}} S_i(\rho; s_0) = \rho_i + \rho_{i,1}s_0 + O(s_0^2)\,. \tag{235}$$

Here ρ_i solves the equation $S'_{i,0}(\rho_i) = 0$. Henceforth we shall focus on the $\beta \to \infty$ limit, in which the equation $S'_{i,0}(\rho_i) = 0$ reduces to

$$1 = \frac{h_i^2}{(\rho_i + r_i)^2} + \frac{1 + \mathsf{c}_i}{\rho_i^2}\,. \tag{236}$$

The first order correction is given by $\rho_{i,1} = -S'_{i,1}(\rho_i)/S''_{i,0}(\rho_i)$. Substituting in Eq. (231), we get the saddle point value of $S_i(\rho_i; s_0)$:

$$\min_{\rho\in\mathbb{R}_{>0}} S_i(\rho; s_0) = S_{i,0}(\rho_i) + S_{i,1}(\rho_i)\,s_0 + \Delta_2 s_0^2 + O(s_0^3)\,, \tag{237}$$
$$\Delta_2 = \frac{1}{2}\left[S_{i,2}(\rho_i) - \frac{S'_{i,1}(\rho_i)^2}{S''_{i,0}(\rho_i)}\right] \tag{238}$$
$$= -\frac{1}{\rho_i} + \frac{1}{\rho_i + r_i} - \left(\frac{h_i^2}{(\rho_i + r_i)^3} + \frac{1 + \mathsf{c}_i}{\rho_i^3}\right)^{-1}\frac{h_i^2}{(\rho_i + r_i)^4} \tag{239}$$
$$= -\frac{1}{\rho_i} + \frac{1}{\rho_i + r_i} - \left(1 + \frac{(1 + \mathsf{c}_i)r_i}{\rho_i^3}\right)^{-1}\frac{h_i^2}{(\rho_i + r_i)^3}\,, \tag{240}$$

where in the last expression we used Eq. (236). Substituting in Eq. (170), we get a recursion for the triple h_i, c_i, r_i:

$$\mathsf{c}_0 = \sum_{i=1}^{k} \frac{\mathsf{c}_i}{\rho_i^2}\,, \tag{241}$$

$$h_0 = \sum_{i=1}^{k} \frac{h_i}{\rho_i + r_i}\,, \tag{242}$$

$$r_0 = \sum_{i=1}^{k} \left\{ \frac{1}{\rho_i} - \frac{1}{\rho_i + r_i} + \left(1 + \frac{(1+\mathsf{c}_i) r_i}{\rho_i^3}\right)^{-1} \frac{h_i^2}{(\rho_i + r_i)^3} \right\} \tag{243}$$

If G_n is distributed according to the two-groups stochastic block model, the distributions of this triples on different type vertices are related by symmetry $(\mathsf{c}_{i(+)}, h_{i(+)}, r_{i(+)}) \stackrel{\mathrm{d}}{=} (\mathsf{c}_{i(-)}, -h_{i(-)}, r_{i(-)})$ for $i(+) \in V_+$, $i(-) \in V_-$. This leads to the following distributional recursion for the sequence of random vectors $\{(\mathsf{c}^t, h^t, r^t)\}_{t\ge 0}$:

$$\mathsf{c}^{t+1} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{L_+ + L_-} \frac{\mathsf{c}_i^t}{\rho_i^2}\,, \tag{244}$$

$$h^{t+1} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{L_+ + L_-} \frac{s_i h_i^t}{\rho_i + r_i^t}\,, \tag{245}$$

$$r^{t+1} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{L_+ + L_-} \left\{ \frac{1}{\rho_i} - \frac{1}{\rho_i + r_i^t} + \left(1 + \frac{(1+\mathsf{c}_i^t) r_i^t}{\rho_i^3}\right)^{-1} \frac{(h_i^t)^2}{(\rho_i + r_i^t)^3} \right\}, \tag{246}$$

where $L_+ \sim \mathrm{Poisson}(a/2)$, $L_- \sim \mathrm{Poisson}(b/2)$, $s_1, \dots, s_{L_+} = +1$, $s_{L_+ + 1}, \dots, s_{L_+ + L_-} = -1$, and $\{(\mathsf{c}_i^t, h_i^t, r_i^t)\}$ are i.i.d. copies of (c^t, h^t, r^t). Finally, ρ_i is a function of $(\mathsf{c}_i^t, h_i^t, r_i^t)$ implicitly defined as the solution of

$$1 = \frac{(h_i^t)^2}{(\rho_i + r_i^t)^2} + \frac{1 + \mathsf{c}_i^t}{\rho_i^2}\,. \tag{247}$$

Finally, the asymptotic overlap achieved by SDP is given, in terms of the distributional fixed point h^* of this recursion, by

$$\mathrm{Overlap}(\hat{\boldsymbol{x}}^{\mathrm{SDP}}) = \mathbb{E}\big\{\mathrm{sign}(h^*)\big\}\,. \tag{248}$$

As a check of the above derivation, we next verify that we recover the correct $d \to \infty$ limit, captured by the $\mathbb{Z}_2$ synchronization model, cf. Section 3.2.5. We will focus on the fixed point of recursions (244) to (246), and hence omit iteration index t. For large degree, we expect central limit theorem to imply the following asymptotic behaviors:

$$\mathsf{c}_i = d\,\mu_{\mathsf{c}} + O(\sqrt{d})\,, \tag{249}$$

$$h_i \stackrel{\mathrm{d}}{=} \sqrt{d}\big(\mu_h + \sigma_h\, Z\big) + o(\sqrt{d})\,, \tag{250}$$

$$r_i = \sqrt{d}\,\mu_r + O(1)\,, \tag{251}$$

where μ_{c}, μ_h, σ_h, μ_r are deterministic parameters to be determined. Equation (247) thus implies $\rho_i = \sqrt{d}\rho$, with ρ solution of

$$1 = \frac{(\mu_h + \sigma_h Z)^2}{(\rho + \mu_r)^2} + \frac{\mu_{\mathsf{c}}}{\rho^2}\,. \tag{252}$$

Equations (244) to (246) then yield the following. From Eq. (244) we get

$$\mu_{\mathsf{c}} = \mu_{\mathsf{c}} \mathbb{E}\left\{\frac{1}{\rho^2}\right\}, \tag{253}$$

i.e., assuming $\mu_{\mathsf{c}} \neq 0$, $\mathbb{E}\{\rho^{-2}\} = 1$. From Eq. (245) we get

$$\mu_h = \lambda\, \mathbb{E}\left\{\left(\frac{\mu_h + \sigma_h Z}{\rho + \mu_r}\right)\right\}, \tag{254}$$

$$\sigma_h^2 = \mathbb{E}\left\{\left(\frac{\mu_h + \sigma_h Z}{\rho + \mu_r}\right)^2\right\}. \tag{255}$$

Finally, from Eq. (246) we get

$$\mu_r = \mathbb{E}\left\{\frac{1}{\rho} - \frac{1}{\rho + \mu_r} + \left(1 + \frac{\mu_{\mathsf{c}}\mu_r}{\rho^3}\right)^{-1} \frac{(\mu_h + \sigma_h Z)^2}{(\rho + \mu_r)^3}\right\}. \tag{256}$$

We claim that these equations are equivalent to the ones for the $\mathbb{Z}_2$ synchronization problem, derived using the replica method, namely Eqs. (99) to (102). Indeed Eq. (102) coincides with Eq. (253), for $\mu_{\mathsf{c}} \neq 0$. Setting $\mu_h = \mu$, $\sigma_h^2 = q$, and $\mu_r = r$, we obtain that Eqs. (99), (100) coincide with Eqs. (254), (255). Further, taking expectation of Eq. (252), and using $\mathbb{E}\{\rho^{-2}\} = 1$, we get $\mu_{\mathsf{c}} = 1 - q$. Hence, we obtain that Eq. (252) coincides with Eq. (98).

Substituting the values of various parameters in Eq. (256), we obtain

$$r = \mathbb{E}\left\{\frac{1}{\rho} - \frac{1}{\rho + r} + \left(1 + \frac{r(1-q)}{\rho^3}\right)^{-1} \frac{(\mu + \sqrt{q}Z)^2}{(\rho + r)^3}\right\}. \tag{257}$$

We claim that this is equivalent to Eq. (101). To see this, notice that differentiating Eq. (252) with respect to Z we get

$$\frac{\partial \rho}{\partial Z} = \left[\frac{(\mu + \sqrt{q}Z)^2}{(\rho + r)^3} + \frac{(1-q)}{\rho^3}\right]^{-1} \cdot \frac{(\mu + \sqrt{q}Z)\sqrt{q}}{(\rho + r)^2} \tag{258}$$

$$= \left(1 + \frac{r(1-q)}{\rho^3}\right)^{-1} \frac{(\mu + \sqrt{q}Z)\sqrt{q}}{\rho + r}\,, \tag{259}$$

where the second equality follows again from Eq. (252). Using this identity, we can rewrite Eq. (257) as

$$r = \mathbb{E}\left\{\frac{1}{\rho} - \frac{1}{\rho + r} + \frac{1}{\sqrt{q}} \frac{(\mu + \sqrt{q}Z)}{(\rho + r)^2} \frac{\partial \rho}{\partial Z}\right\}, \tag{260}$$

or

$$r = \mathbb{E}\left\{\frac{1}{\rho} - \frac{1}{\sqrt{q}} \frac{\partial}{\partial Z} \frac{\mu + \sqrt{q}Z}{\rho + r}\right\}, \tag{261}$$

Using Gaussian integration by parts in the second term we finally obtain Eq. (101). This concludes our verification for the case of $d \to \infty$.

5.5 Limitations of the vectorial ansatz

Our analytical estimate of the SDP phase transition location, $\tilde{\lambda}_c^{\rm SDP}(d)$ was carried out within the vectorial ansatz in Eqs. (172), (198). Let us stress once again that this ansatz is only approximate.

The origin of this approximation can be gleaned from the calculation in Section 5.1. As we have seen Eq. (180) is only accurate when $\langle \boldsymbol{\sigma}_0, \boldsymbol{z}_i\rangle$ is small. However, according to the same ansatz, $\boldsymbol{\sigma}_0$ will be aligned to $\boldsymbol{z}_0$, which –in turn– can be aligned with $\boldsymbol{z}_i$.

We expect this approximation to be accurate in the following regimes:

- For large average degree d. Indeed, in this case, $\boldsymbol{z}_0$ is weakly correlated with $\boldsymbol{z}_i$.
- For d close to 1. In this case c_0 is small and hence, under ν_0, $\boldsymbol{\sigma}_0$ is approximately uniformly distributed, and hence has a small scalar product $\langle \boldsymbol{z}_i, \boldsymbol{\sigma}_0\rangle$.

Let us also notice that the vectorial ansatz can be systematically improved upon by considering quadratic terms tepending in two-dimensional projections, and so on. We leave this direction for future work.

6 Numerical experiments for community detection

In this section we provide details about our numerical simulations with the SDP estimator for the community detection problem. For the reader's convenience we begin by recalling some definitions.

We denote by $G_n = (V_n, E_n)$ the random graph over vertex set $V_n = [n]$, generated according the hidden partition model, and by $\boldsymbol{x_0} \in \{+1, -1\}^n$ the vertex labels. Conditional on $\boldsymbol{x_0}$, edges are independent with distribution

$$\mathbb{P}\big\{(i,j) \in E_n \big| \boldsymbol{x_0}\big\} = \begin{cases} a/n & \text{if } x_{0,i}x_{0,j} = +1, \\ b/n & \text{if } x_{0,i}x_{0,j} = -1. \end{cases} \tag{262}$$

We denote by $d = (a+b)/2$ the average degree, and by $\lambda = (a-b)/\sqrt{2(a+b)}$ the 'signal strength.'

Throughout this section, ∂i indicates the set of neighbors of vertex i, i.e. $\partial i \equiv \{j \in [n] : (i,j) \in E\}$.

We next recall the SDP relaxation for estimating community memberships:

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} X_{ij}\,, \\ \text{subject to} \quad & \boldsymbol{X} \succeq 0\,, \\ & \boldsymbol{X}\mathbf{1} = \mathbf{0}\,, \quad X_{ii} = 1 \ \ \forall i \in [n]\,. \end{aligned} \tag{263}$$

Denote by $\boldsymbol{X}_{\rm opt} = \boldsymbol{X}_{\rm opt}(G)$ an optimizer of the above problem. The estimated membership vector is then obtained by 'rounding' the principal eigenvector of $\boldsymbol{X}_{\rm opt}$ as follows. Letting $\boldsymbol{v}_1 = \boldsymbol{v}_1(\boldsymbol{X}_{\rm opt})$ be the principle eigenvector of $\boldsymbol{X}_{\rm opt}$, the SDP estimate is given by

$$\hat{\boldsymbol{x}}^{\rm SDP}(G) = \mathrm{sign}\big(\boldsymbol{v}_1(\boldsymbol{X}_{\rm opt}(G))\big)\,. \tag{264}$$

We measure the performance of such an estimator via the overlap:

$$\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP}) = \frac{1}{n}\mathbb{E}\big\{\big|\langle \hat{\boldsymbol{x}}^{\rm SDP}(G), \boldsymbol{x_0}\rangle\big|\big\}\,, \tag{265}$$

where $\boldsymbol{x_0} \in \{-1,+1\}^n$ encodes the ground truth memberships with $x_{0,i} = +1$ if $i \in V_+$ and $x_{0,i} = -1$ if $i \in V_-$. Note that $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\mathrm{SDP}}) \in [0,1]$ and a random guessing estimator yields overlap of order $O(1/\sqrt{n})$.

The majority of our calculations were run on a cluster with 160 cores (Intel Xeon), taking roughly a month (hence total CPU time was roughly 10 years).

6.1 Optimization algorithms

We write SDP optimization (263) in terms of the vector spin model. Let $\boldsymbol{\sigma}_i \in \mathbb{R}^m$ be the vector spin assigned to node i, for $i \in [n]$, and define $\boldsymbol{\sigma} \equiv (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)$. We then rewrite the SDP (263) as

$$\begin{aligned} \underset{\boldsymbol{\sigma}}{\text{maximize}} \quad & F(\boldsymbol{\sigma}) \equiv \sum_{(i,j)\in E} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathcal{M}(n,m) \,, \end{aligned} \tag{266}$$

where the manifold $\mathcal{M}(n,m)$ is defined as below:

$$\mathcal{M}(n,m) = \Big\{ \boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_n) \in (\mathbb{R}^m)^n : \ \|\boldsymbol{\sigma}_i\|_2 = 1 \,, \ \sum_{i=1}^n \boldsymbol{\sigma}_i = 0 \Big\} \,. \tag{267}$$

We will omit the dimensions when they are clear from the context.

As discussed in the main text, the two optimization problems (263) and (266) have a value that differ by a relative error of $O(1/m)$, uniformly in the size n. In particular, the asymptotic value of the SDP is the same, if we let $m \to \infty$ *after* $n \to \infty$.

In fact the following empirical findings (further discussed below) point at a much stronger connection:

1. With high probability, the optimizer appears to be essentially independent of m already for moderate values of m (in practice, already for $m = 40 \sim 100$, when $n \lesssim 10^4$).

2. Again, for moderate values of m, optimization methods do not appear to be stuck in local minima. Roughly speaking, while the problem is non-convex from a worst case perspective, typical instances are nearly convex.

Motivated by these findings, we solve optimization problem (266) in lieu of SDP problem (263), using the two algorithms described below: (*i*) Projected gradient ascent; (*ii*) Block-coordinate ascent.

The rank-constrained formulation also allows to accelerate the rounding step to compute $\hat{\boldsymbol{x}}^{\mathrm{SDP}}(G)$, which can be obtained in time $O(nm^2+m^3)$, instead of the naive $O(n^3)$. Namely, given an optimizer $\boldsymbol{\sigma}^{\mathrm{opt}}$, we compute the $m \times m$ empirical covariance matrix

$$\widehat{\boldsymbol{\Sigma}} \equiv \frac{1}{n} \sum_{i=1}^n \boldsymbol{\sigma}_i^{\mathrm{opt}} (\boldsymbol{\sigma}_i^{\mathrm{opt}})^{\mathsf{T}} \,. \tag{268}$$

Denoting by $\boldsymbol{\varphi}$ the principal eigenvector of $\widehat{\boldsymbol{\Sigma}}$, we obtain the estimator $\hat{\boldsymbol{x}}^{\mathrm{SDP}}(G) \in \{1,-1\}^n$ via

$$\hat{x}_i^{\mathrm{SDP}} = \mathrm{sign}\big(\langle \boldsymbol{\varphi}, \boldsymbol{\sigma}_i^{\mathrm{opt}} \rangle\big) \,. \tag{269}$$

This approach allows us to carry out high-precision simulations for large instances, namely up to $n = 64,000$. By comparison, standard SDP solvers are based on interior-point methods and cannot scale beyond n of the order of a few hundreds.

6.1.1 Projected gradient ascent

The tangent space at $\boldsymbol{\sigma} \in \mathcal{M}$ is given by

$$\mathsf{T}_{\boldsymbol{\sigma}}\mathcal{M} \simeq \Big\{ \boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n) : \langle \boldsymbol{\sigma}_i, \boldsymbol{z}_i \rangle = 0 \,,\ \sum_{i=1}^n \boldsymbol{z}_i = 0 \Big\} \,. \tag{270}$$

Define the orthogonal projectors in $\mathbb{R}^m$:

$$\mathsf{P}_i \equiv \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^{\mathsf{T}} \,, \quad \mathsf{P}_i^\perp \equiv \mathrm{I} - \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^{\mathsf{T}} \,. \tag{271}$$

By identification (270), the manifold gradient of F reads

$$\begin{aligned} \nabla F(\boldsymbol{\sigma}) &= \big(\nabla F(\boldsymbol{\sigma})_1, \dots, \nabla F(\boldsymbol{\sigma})_n \big) \,, \\ \nabla F(\boldsymbol{\sigma})_i &= \mathsf{P}_i^\perp (\boldsymbol{v}_i - \overline{\boldsymbol{v}}_i) \,, \end{aligned} \tag{272}$$

where

$$\begin{aligned} \boldsymbol{v}_i(\boldsymbol{\sigma}) &\equiv \sum_{j \in \partial i} \boldsymbol{\sigma}_j \,, \\ \overline{\boldsymbol{v}} &\equiv \Big(\sum_{i=1}^n \mathsf{P}_i^\perp \Big)^{-1} \Big(\sum_{i=1}^n \mathsf{P}_i^\perp \boldsymbol{v}_i \Big) \,. \end{aligned} \tag{273}$$

We next define the convex envelope of $\mathcal{M}$:

$$\mathrm{conv}(\mathcal{M}) = \Big\{ \boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_n) \in (\mathbb{R}^m)^n : \ \|\boldsymbol{\sigma}_i\| \le 1 \,,\ \sum_{i=1}^n \boldsymbol{\sigma}_i = 0 \Big\} \,, \tag{274}$$

and the corresponding orthogonal projector

$$\mathsf{P}_{\mathcal{M}}(\boldsymbol{y}) \equiv \arg \min_{\boldsymbol{z} \in \mathrm{conv}(\mathcal{M})} \|\boldsymbol{y} - \boldsymbol{z}\|_2 \,. \tag{275}$$

The projected gradient method alternates between a step in the direction of $\nabla F(\boldsymbol{\sigma})$ and a projection onto $\mathrm{conv}(\mathcal{M})$. Pseudocode is given as Algorithm 2.

The projected gradient method requires a subroutine for computing the projection $\mathsf{P}_{\mathcal{M}}$ onto $\mathrm{conv}(\mathcal{M})$. In order to compute this projection, we write the Lagrangian corresponding to problem (275):

$$\mathcal{L} \equiv \frac{1}{2} \sum_{i=1}^n \|\boldsymbol{y}_i - \boldsymbol{z}_i\|_2^2 + \boldsymbol{w}^{\mathsf{T}} \sum_{i=1}^n \boldsymbol{z}_i + \sum_{i=1}^n \frac{\mu_i}{2} (\|\boldsymbol{z}_i\|_2^2 - 1) \,, \tag{278}$$

with $\mu \ge 0$ for $1 \le i \le n$. Setting $\nabla_{\boldsymbol{z}_i} \mathcal{L} = 0$, we obtain

$$\boldsymbol{z}_i = \frac{\boldsymbol{y}_i - \boldsymbol{w}}{\mu_i + 1} \,. \tag{279}$$

Further, the constraint $\sum_{i=1}^n \boldsymbol{z}_i = 0$ implies

$$\boldsymbol{w} = \Big(\sum_{i=1}^n \frac{1}{\mu_i + 1} \Big)^{-1} \Big(\sum_{i=1}^n \frac{\boldsymbol{y}_i}{\mu_i + 1} \Big) \,. \tag{280}$$

Algorithm 2 Projected gradient ascent for relaxed min bisection

Input: n,m, edge set $E \subseteq [n] \times [n]$, $\texttt{tol}_2$
Output: $\boldsymbol{\sigma}^{\rm opt} \equiv (\boldsymbol{\sigma}_1^{\rm opt}, \dots, \boldsymbol{\sigma}_n^{\rm opt})$
1: Initialize $\boldsymbol{\sigma}^0 \in \mathcal{M}$ at random
(e.g. by letting $\boldsymbol{\sigma}_i \in \{\pm \boldsymbol{e}_1, \pm \boldsymbol{e}_2, \dots, \pm \boldsymbol{e}_m\}$ uniformly random conditional on $\sum_{i=1}^n \boldsymbol{\sigma}_i^0 = 0$).
2: **for** $t = 0, 1, 2, \dots$ **do**
3: Gradient step

$$\tilde{\boldsymbol{\sigma}}^{t+1} = \boldsymbol{\sigma}^t + \varepsilon_t \nabla F(\boldsymbol{\sigma}^t)\,, \tag{276}$$

with $\varepsilon_t = 1/\sqrt{t}$ the step size.
4: Projection step:

$$\boldsymbol{\sigma}^{t+1} = \mathsf{P}_{\mathcal{M}}(\tilde{\boldsymbol{\sigma}}^{t+1})\,. \tag{277}$$

5: **if** $\|\nabla F(\boldsymbol{\sigma}^{t+1})\|_2/\sqrt{n} \le \texttt{tol}_2$ **then**
6: set $T = t$
7: break
8: **return** $\boldsymbol{\sigma}^{\rm opt} = \boldsymbol{\sigma}^t$

Due to constraint $\|\boldsymbol{z}_i\| \le 1$, we have $\mu_i \ge \|\boldsymbol{y}_i - \boldsymbol{w}\| - 1$. Also, by the KKT conditions, if the inequality is strict we have $\mu_i = 0$. Therefore,

$$\mu_i = \max(\|\boldsymbol{y}_i - \boldsymbol{w}\|_2 - 1, 0)\,. \tag{281}$$

Substituting for μ_i from Eq. (281) into (280) we arrive at

$$\boldsymbol{w} = \Big(\sum_{i=1}^n \frac{1}{\max(\|\boldsymbol{y}_i - \boldsymbol{w}\|_2, 1)}\Big)^{-1} \Big(\sum_{i=1}^n \frac{\boldsymbol{y}_i}{\max(|\boldsymbol{y}_i - \boldsymbol{w}\|, 1)}\Big)\,. \tag{282}$$

We compute the Lagrange multiplier $\boldsymbol{w}$ in an iterative manner as described in Algorithm 3.

6.1.2 Block coordinate ascent

We present here a second algorithm to solve problem (266), that uses block-coordinate descent. This provides independent check of numerical results. Further, this second method appears to be faster than projected gradient ascent.

We start by considering an unconstrained version of the optimization problem (with $\boldsymbol{A}_G$ the adjacency matrix of the graph G):

$$\text{maximize } \langle \boldsymbol{\sigma}, \big(\boldsymbol{A}_G - \eta \boldsymbol{1}\boldsymbol{1}^{\mathsf{T}}\big)\boldsymbol{\sigma}\rangle\,. \tag{285}$$

Equivalently, this objective function can be written as $F(\underline{\boldsymbol{\sigma}}) - \eta\|\boldsymbol{M}(\boldsymbol{\sigma})\|_2^2/2$, where $\boldsymbol{M}(\boldsymbol{\sigma}) \equiv \sum_{i=1}^n \boldsymbol{\sigma}_i$. As $\eta \to \infty$, this is of course equivalent to problem (266).

For G distributed according to the hidden partition model (262), with random vertex labels $\boldsymbol{x_0}$, we have $\mathbb{E}\{(\boldsymbol{A}_G)_{ij}\} = d/n$. This suggests that $\eta \ge d/n$ should be sufficient to obtain a balanced

Algorithm 3 Algorithm for computing the projection $\mathsf{P}_{\mathcal{M}}(\boldsymbol{y})$

Input: $\boldsymbol{y}$, $\mathtt{tol}_1$
Output: $\mathsf{P}_{\mathcal{M}}(\boldsymbol{y})$
1: $\boldsymbol{w}^0 \leftarrow \mathbf{0}$
2: **for** $t = 0, 1, 2, \dots$ **do**
3: update

$$\boldsymbol{w}^{t+1} = \Big(\sum_{i=1}^{n} \frac{1}{\max(\|\boldsymbol{y}_i - \boldsymbol{w}^t\|_2, 1)} \Big)^{-1} \Big(\sum_{i=1}^{n} \frac{\boldsymbol{y}_i}{\max(\|\boldsymbol{y}_i - \boldsymbol{w}^t\|_2, 1)} \Big) \,. \tag{283}$$

4: **if** $\|\boldsymbol{w}^{t+1} - \boldsymbol{w}^t\|_2 \le \mathtt{tol}_1$ **then**
5: set $T = t$
6: break
7: **return**

$$\mathsf{P}_{\mathcal{M}}(\boldsymbol{y})_i = \frac{\boldsymbol{y}_i - \boldsymbol{w}^T}{\max(\|\boldsymbol{y}_i - \boldsymbol{w}^T\|, 1)} \,, \qquad \text{for} \quad i = \{1, 2, \dots, n\} \,. \tag{284}$$

partition. In practice we will take $\eta \approx 1$ and check that this yields well balanced partitions (i.e. $\boldsymbol{M} \approx 0$), cf. Section 6.3.1 below.

We maximize the objective by iteratively maximizing over each of the vectors $\boldsymbol{\sigma}_i$. The latter optimization has a close form expression. More precisely, at each step of this dynamics, we sort the variables in a random order and we update them sequentially, by maximizing the objective function. This is easily done by aligning $\boldsymbol{\sigma}_i$ along the 'local field'

$$\boldsymbol{h}_i \equiv -\eta \, \boldsymbol{M}(\underline{\boldsymbol{\sigma}}) + \sum_{j:(i,j)\in E} \boldsymbol{\sigma}_j \,. \tag{286}$$

We check the convergence by measuring the largest variation in a spin variable during the last iteration. Namely we define

$$\Delta_{\max}(t) = \max_{i\in[n]} \|\boldsymbol{\sigma}_i^{t+1} - \boldsymbol{\sigma}_i^t\|_2 \,, \tag{287}$$

and we use as convergence criterion $\Delta_{\max} < \mathtt{tol}_3$. The corresponding pseudocode is presented as Algorithm 4.

The resulting algorithm is very simple and depends on two parameters (η and $\mathtt{tol}_3$) that will be discussed in the Section 6.3, together with dependence on the number m of spin components.

6.2 Numerical experiments with the projected gradient ascent

In this section we report our results with the projected gradient algorithm. As mentioned above, we found that the block coordinate ascent method was somewhat faster, and therefore we used the latter for large-statistics simulations, and high-precision determinations of the critical point $\lambda_c^{\rm SDP}(d)$. We defer to the next section for further discussion.

Algorithm 4 Block coordinate ascent for relaxed min bisection

Input: n, m, edge set $E \subseteq [n] \times [n]$, η, $\texttt{tol}_3$
Output: $\underline{\boldsymbol{\sigma}}^{\rm opt} \equiv (\boldsymbol{\sigma}_1^{\rm opt}, \dots, \boldsymbol{\sigma}_n^{\rm opt})$

Initialize $\underline{\boldsymbol{\sigma}}^0 \in \mathcal{M}$ at random
(e.g. by letting $\boldsymbol{\sigma}_i \in \{\pm \boldsymbol{e}_1, \pm \boldsymbol{e}_2, \dots, \pm \boldsymbol{e}_m\}$ at random so that $\sum_i \boldsymbol{\sigma}_i^0 = 0$).
for $t = 1, 2, \dots$ **do**
  Choose a random permutation $\pi : [n] \to [n]$
  for $i = 1, 2, \dots, n$ **do**
    Update spins, aligning to the local field

$$\boldsymbol{\sigma}_{\pi(i)}^{t+1} = \frac{-\eta\, \boldsymbol{M}(\boldsymbol{\sigma}^t) + \sum_{j:(\pi(i),j)\in E} \boldsymbol{\sigma}_j^t}{\| - \eta\, \boldsymbol{M}(\boldsymbol{\sigma}^t) + \sum_{j:(\pi(i),j)\in E} \boldsymbol{\sigma}_j^t \|_2}\,, \tag{288}$$

  if $\Delta_{\max}(t) \le \texttt{tol}_3$ **then**
    set $T = t$
    break
return $\boldsymbol{\sigma}^{\rm opt} = \boldsymbol{\sigma}^T$

We use the projected gradient ascent discussed in Section 6.1.1, with $\texttt{tol}_1 = \texttt{tol}_2 = 10^{-6}$. For each value of $d \in \{5, 10, 15, 20, 25, 30\}$ and $n \in \{2000,\ 4000,\ 8000,\ 16000\}$, we generate 500 realizations of graph G from the stochastic block model defined in Eq. (262). In these experiments, we observed that the estimated membership vector does not change for $m \ge 40$, cf. Section 6.2.1. The results reported here correspond to $m = 40$.

Figure 4 reports the estimated overlap $\text{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP})$ (across realizations) achieved by the SDP estimator, for different values of d and n. The solid curve corresponds to the cavity prediction, cf. equation (104), for large d. As we see the empirical results are in good agreement with the analytical curve even for small average degrees $d = 5, 10$.

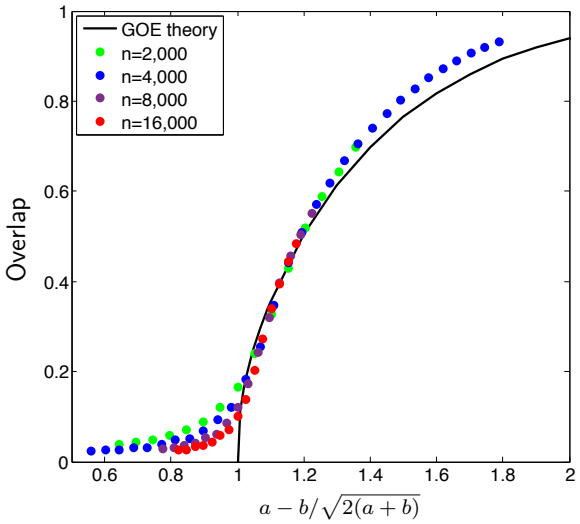
In Figure 2 in the main text, we also plot the prediction from the cavity method in the sparse graph model, cf. Section 5.4, Eq. (248), with the cavity equations solved by population dynamics. This approach appears to capture wery well the small discrepancy between the large-degree (Gaussian) theory, and the small degree behavior.

In particular, *the phase transition location seems indistinguishable, on this scale, from* $(a - b)/\sqrt{2(a+b)} = 1$.

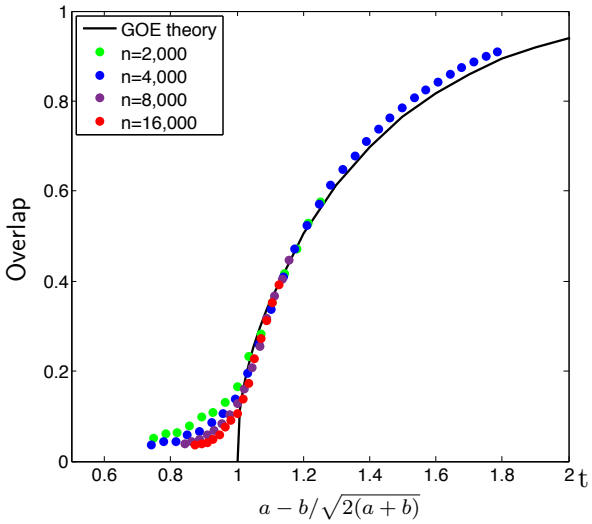
6.2.1 Dependence on m

As we explained before, optimization problem (266) and SDP (263) are equivalent provided $m \ge n$. In principle, one can solve (266) applying Algorithm 2 with $m = n$. However, this choice leads to a computationally expensive procedure. On the other hand, we expect the solution to be essentially independent of m already for moderate values of m. Several theoretical arguments point to this (in particular, the Grothendieck-type inequality of [MS16]). We provide numerical evidence in this section (supporting in particular the choice $m = 40$).

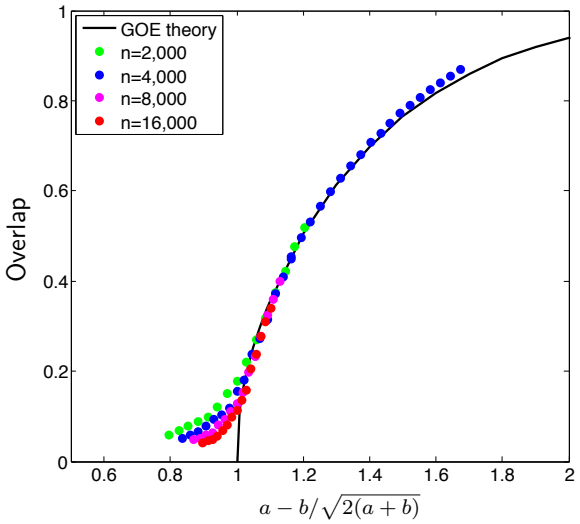
In the first experiment, we set the average degree $d = (a + b)/2 = 5$, $n = 4000$ and vary $\lambda = (a - b)/\sqrt{2(a+b)} \in \{0.9, 1, 1.1, 1.2\}$. For each λ, we solve for a and b and generate 100



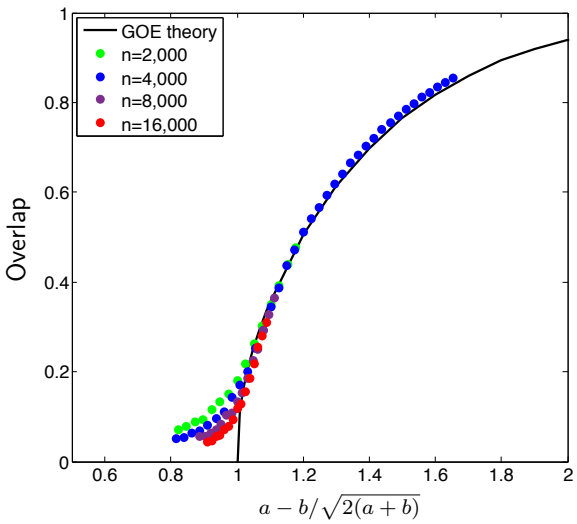
(a) $d = 5$



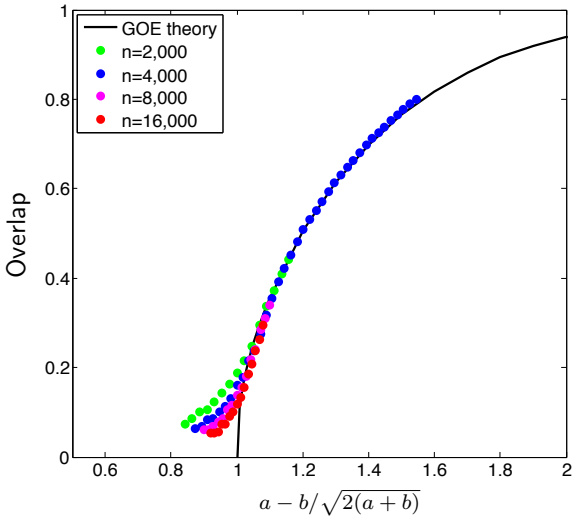
(b) $d = 10$



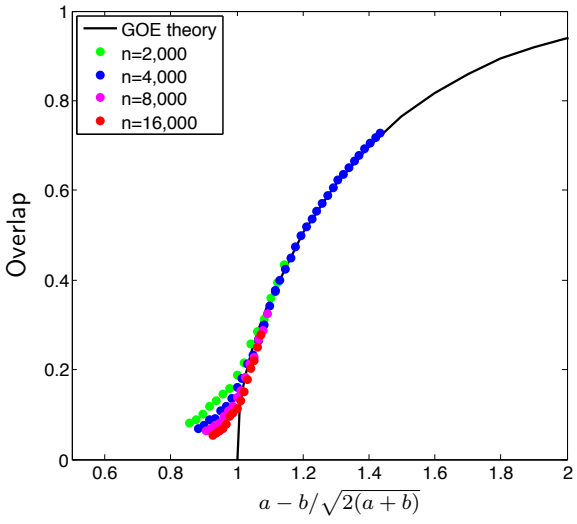
(c) $d = 15$



(d) $d = 20$



(e) $d = 25$



(f) $d = 30$

Figure 4: Community detection under the hidden partition model of Eq. (262), for several average degree $d = (a+b)/2$. Dots corresponds to the performance of the SDP reconstruction method (averaged over 500 realizations). The continuous curve is the asymptotic analytical prediction for the Gaussian model (which captures the large-degree behavior).

realizations of the graph as per model (262) with parameters a, b. For several values of m, we solve optimization (266) and report the average overlap and its standard deviation. The results are summarized in Table 2. As we see, for $m \geq 40$ the changes in average overlaps are comparable with the corresponding standard deviations. An interesting observation is that error bars for smaller m are larger, indicating more variations of overlaps across different realizations.

Table 3 demonstrates the results for an analogous experiment with $d = 10$.

We observe a similar trend for other several values of a, b, n. Based on these observations, we use $m = 40$ in our numerical experiments throughout this section.

	$\lambda = 0.9$	$\lambda = 1$	$\lambda = 1.1$	$\lambda = 1.2$
$m = 5$	0.1231 ± 0.0037	0.0898 ± 0.0059	0.2637 ± 0.0074	0.4627 ± 0.0017
$m = 10$	0.1544 ± 0.0008	0.0659 ± 0.0045	0.3303 ± 0.0018	0.4612 ± 0.0006
$m = 20$	0.1663 ± 0.0002	0.1363 ± 0.0008	0.3358 ± 0.0005	0.4598 ± 0.0002
$m = 40$	0.1668 ± 0.0001	0.1398 ± 0.0002	0.3358 ± 0.0002	0.4599 ± 0.0001
$m = 80$	0.1669 ± 0.0001	0.1397 ± 0.0002	0.3358 ± 0.0002	0.4596 ± 0.0001
$m = 160$	0.1669 ± 0.0001	0.1397 ± 0.0002	0.3358 ± 0.0002	0.4596 ± 0.0001

Table 2: We fix $d = (a + b)/2 = 5$ and vary $\lambda = (a - b)/\sqrt{2(a + b)} \in \{0.9, 1, 1.1., 1.2\}$. The reported values are the average overlaps (over 100 realizations) with one standard deviations, for several values of m. As we see for $m \geq 40$, the changes in the average overlaps are comparable to the corresponding standard deviations.

	$\lambda = 0.9$	$\lambda = 1$	$\lambda = 1.1$	$\lambda = 1.2$
$m = 5$	0.0875 ± 0.0035	0.1385 ± 0.0072	0.3760 ± 0.0021	0.5045 ± 0.0008
$m = 10$	0.1322 ± 0.0007	0.1943 ± 0.0014	0.3873 ± 0.0006	0.5050 ± 0.0003
$m = 20$	0.1346 ± 0.0002	0.2100 ± 0.0003	0.3890 ± 0.0001	0.5070 ± 0.0001
$m = 40$	0.1353 ± 0.0001	0.2089 ± 0.0001	0.3885 ± 0.0001	0.5088 ± 0.0001
$m = 80$	0.1354 ± 0.0000	0.2090 ± 0.0000	0.3885 ± 0.0000	0.5089 ± 0.0000
$m = 160$	0.1354 ± 0.0000	0.2090 ± 0.0000	0.3885 ± 0.0000	0.5089 ± 0.0000

Table 3: We fix $d = (a + b)/2 = 10$ and vary $\lambda = (a - b)/\sqrt{2(a + b)} \in \{0.9, 1, 1.1., 1.2\}$. The reported values are the average overlaps (over 100 realizations) with one standard deviations, for several values of m. As we see for $m \geq 40$, the changes in the average overlaps are comparable to the corresponding standard deviations.

6.2.2 Robustness and comparison with spectral methods

Spectral methods are among the most popular nonparametric approaches to clustering. These methods classify nodes according to a the eigenvectors of a matrix associated with the graph, for instance its adjacency matrix or Laplacian. While standard spectral clustering works well when the graph is sufficiently dense or is regular, it is significantly suboptimal for sparse graphs. The reason is that the leading eigenvector of the adjacency matrix is localized around the high degree

nodes.[3]

Recently, [KMM+13] proposed a class of very interesting spectral methods based on the non-backtracking walk on the directed edges of the graph G. The spectrum of non-backtracking matrix is more robust to high-degree nodes because a walk starting at a node cannot return to it immediately. Later, [SKZ14] proposed another spectral method, based on the Bethe Hessian operator, that is computationally more efficient than the non-backtracking operator. Further, the (determinant of the) Bethe Hessian is closely related to the spectrum of the non-backtracking operator and exhibits the same convenient properties for the aim of clustering. Rigorous analysis of spectral methods under the model (262) was carried out in [Mas14, MNS13, BLM15]. The main result of these papers is that spectral methods allow to estimate the hidden partition significantly better than random guessing immediately above the ideal threshold $\lambda = (a-b)/\sqrt{2(a+b)} = 1$.

As we saw in the previous section, the threshold of the SDP-based method is extremely close to the ideal one. Here, we compare the Bethe Hessian algorithm with SDP approach in terms of robustness to model miss-specification. We perturb the hidden partition model (262) as follows. For a perturbation level $\alpha \in [0,1]$, we draw $n\alpha$ vertices $i_1, i_2, \dots, i_{n\alpha}$ uniformly at random and for each vertex i_ℓ, we add to graph G connecting all of the neighbors of i_ℓ. This results in adding $O(nd^2\alpha)$ edges to the underlying model (262). This perturbation mimics an important feature of real networks that is absent from the stochastic block model (262), the so-called triangle closure property [EK10] (two friends of a person are often friends).

For perturbation levels $\alpha \in \{0, 0.025, 0.05\}$, we compare the performance of SDP and Bethe Hessian algorithms in terms of Overlap, defined by (265). Figure 5 summarizes the results for $n = 16,000$ and average degree $d = (a+b)/2 = 10$. The reported overlaps are averaged over 100 realizations of the model.

In absence of any perturbation (curves $\alpha = 0$), the two algorithms have nearly equivalent performances. However, already for $\alpha = 0.025$, SDP is substantially superior. While SDP appears to be rather insensitive to the perturbation, the performance of the Bethe Hessian algorithm is severely degraded by it. This is because the added triangles perturb the spectrum of the non-backtracking operator (and similarly of the Bethe Hessian operator) significantly, resulting in poor classification of the nodes.

6.3 Numerical experiments with block coordinate ascent

In this section we present our simulations with the block coordinate ascent algorithm, cf. Algorithm 4. We first discuss the choice of the algorithm parameters η and $\texttt{tol}_3$. cf. Section 6.3.2. In Section 6.3.2 we analyze the dependence on the number of dimensions m and the behavior of the convergence time. We conclude by determining the phase transition point in Section 6.3.3, and comparing this location with our analytical predictions.

6.3.1 Selection of the algorithm parameters

Algorithm 4 requires specifying the parameters η (that penalizes $\boldsymbol{M}(\boldsymbol{\sigma}) \neq 0$) and $\texttt{tol}_3$ (for the convergence criterion). In order to investigate the dependance on these parameters, we set $m = 100$ which, as we will see, is large enough to approximate the behavior at $m = n$.

[3]Note that for sparse stochastic block models as in (262), node degrees do not concentrate and we observe highly heterogeneous degrees.

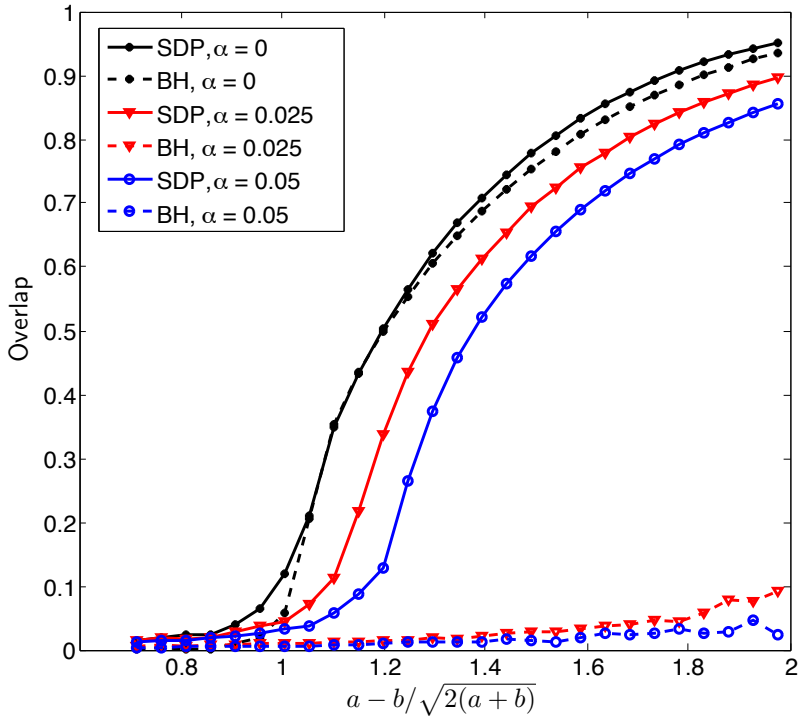


Figure 5: Comparison of SDP and Bethe Hessian algorithm on perturbed hidden partition model. Here, $n = 16,000$, $d = (a+b)/2 = 10$ and we report average overlaps over 100 realizations. Different colors (markers) represent different perturbation levels α. Solid curves are for SDP algorithm and dashed curves are for Bethe-Hessian algorithm. As we see SDP algorithm is more robust than Bethe-Hessian algorithm to the perturbation level α.

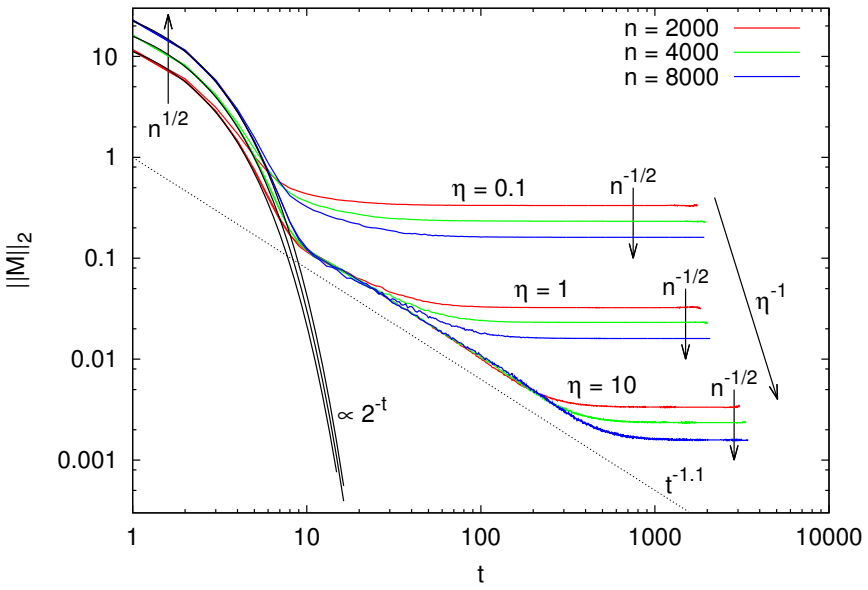


Figure 6: Decrease of the norm of the total magnetization $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2$ during block-coordinate ascent. The starting value $\|\boldsymbol{M}(\boldsymbol{\sigma}^0)\|_2$ is $O(\sqrt{n})$ since the initial configuration is randomly chosen. The asymptotic value is always very small for the sizes studied and decreases as $1/\eta$ and $1/\sqrt{n}$.

In Figure 6 we plot the evolution of the norm of the 'global magnetization,' $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2$, as a function of the number of iterations t. Notice that each iteration corresponds to n updates, one update of each vector $\boldsymbol{\sigma}_i$, $i \in [n]$. We used $d = 5$ and $\lambda = 1.1$, and we averaged over a number of samples ranging from 100 (for $n = 8000$) to 400 (for $n = 2000$).

We observe three regimes:

(*i*) Initially the magnetization decays exponentially, $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2 \approx \|\boldsymbol{M}(\boldsymbol{\sigma}^0)\|_2\, 2^{-t}$. Further, it increases slowly with n. Indeed from central limit theorem, we have $\|\boldsymbol{M}(\boldsymbol{\sigma}^0)\|_2 = \Theta(\sqrt{n})$. The same behavior $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2 = \Theta(\sqrt{n})$ is found empirically at small t.

(*ii*) In an intermediate interval of times, we have a power law decay $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2 \propto t^{-a}$, with exponent $a \approx 1.1$. This intermediate regime is present only for η large enough.

(*iii*) For large t, $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2$ reaches a plateau whose value scales like $\|\boldsymbol{M}(\boldsymbol{\sigma}^t)\|_2 = \Theta(1/\sqrt{n})$ with the system size and is proportional to $1/\eta$.

Already for $\eta = 1$, the value of the plateau is very small, namely

$$\left\| \sum_{i=1}^{n} \boldsymbol{\sigma}_i^t \right\|_2 \lesssim 0.1\,. \tag{289}$$

Further, this value is decreasing with n. Given that $\|\boldsymbol{\sigma}_i\|_2 = 1$, we interpret the above as evidence that the constraint $\boldsymbol{X}\mathbf{1} = 0$ is satisfied with good approximation. We will therefore use $\eta = 1$ in our simulations.

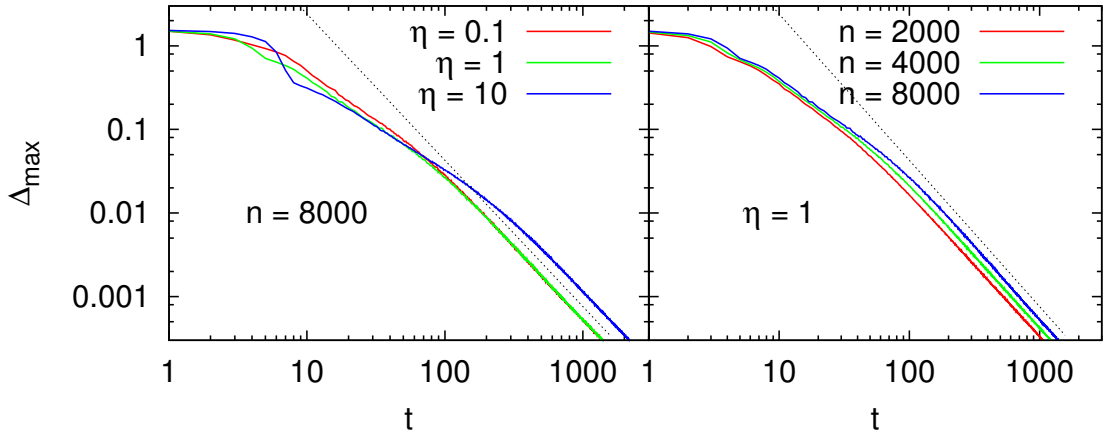


Figure 7: Evolution of $\mathbb{E}\Delta_{\max}(t) = \mathbb{E}\max_{i\in[n]} \|\boldsymbol{\sigma}_i^{t+1} - \boldsymbol{\sigma}_i^t\|_2$ during block-coordinate ascent. Left panel: dependence on the choice of the Lagrange parameter η. Right panel: dependence on the number of vertices.

As an additional remark, notice that there is no special reason to enforce the constraint $\boldsymbol{X}\mathbf{1} = 0$ strictly. Indeed the SDP (165) can be replaced by

$$\begin{aligned} \text{maximize} \quad & \langle \big(\boldsymbol{A}_G - \eta \mathbf{1}\mathbf{1}^{\mathsf{T}}\big), \boldsymbol{X}\rangle\,, && (290)\\ \text{subject to} \quad & \boldsymbol{X} \succeq 0\,, && (291)\\ & X_{ii} = 1 \ \ \forall i \in [n]\,, && (292) \end{aligned}$$

with an arbitrary value of η. Of course this is useful provided η is large enough to rule out the solution $\boldsymbol{X} = \mathbf{1}\mathbf{1}^{\mathsf{T}}$. As mentioned above, $\eta \geq d/n$ should be already large enough [MS16].

In Figure 7 we show how $\Delta_{\max}(t)$ decreases with time in Algorithm 4, again with $d = 5$ and $\lambda = 1.1$.

In the left panel we fix $n = 8000$ and study the dependence on the Lagrange parameter η. We observe two regimes. While for $\eta \lesssim 1$, the convergence rate is roughly independent of η, for $\eta \gtrsim 1$, it becomes somewhat slower with η. This supports the choice $\eta = 1$.

In the right we fix $\eta = 1$ and study the dependence of the convergence time on the graph size n. The number of iterations appears to increase slowly with n (see also Figure 9). Both datasets are consistent with a power law convergence

$$\Delta_{\max}(t) \approx C(n)\, t^{-b}\,, \tag{293}$$

with $b \approx 1.75$ (dotted line), and $C(n)$ polynomially increasing with n (see below for a discussion of the overall scaling of computational complexity with n).

In order to select the tolerance parameter for convergence, $\texttt{tol}_3$, we study the evolution of estimation error. Define the overlap achieved after t iteration as follows. First estimate the vertex labels by computing the top-left singular vector of $\boldsymbol{\sigma}^t$, namely

$$\hat{\boldsymbol{x}}^t(G) = \mathrm{sign}\big(\boldsymbol{v}_1(\boldsymbol{\sigma}^t(\boldsymbol{\sigma}^t)^{\mathsf{T}})\big)\,. \tag{294}$$

Then define, as before

$$\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^t) = \frac{1}{n}\mathbb{E}\big\{\big|\langle \hat{\boldsymbol{x}}^t(G), \boldsymbol{x_0}\rangle\big|\big\}\,. \tag{295}$$

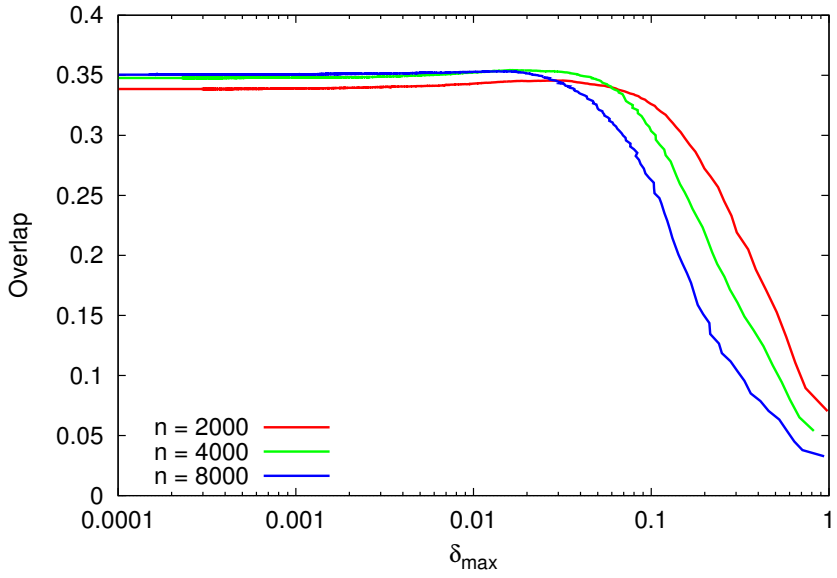


Figure 8: The mean overlap does not depend on the value of $\mathbb{E}\Delta_{\max}$ at which the greedy algorithm is stopped, as soon as $\delta_{\max} \lesssim 10^{-3}$.

Of course, the accuracy of the SDP estimator is given by

$$\text{Overlap}_n(\hat{\boldsymbol{x}}^{\text{SDP}}) = \lim_{t\to\infty} \text{Overlap}_n(\hat{\boldsymbol{x}}^t)\,. \tag{296}$$

In Figure 8, we plot $\text{Overlap}_n(\hat{\boldsymbol{x}}^t)$ as a function of $\mathbb{E}\Delta_{\max}(t)$ for several values of n, $d = 5$ and $\lambda = 1.1$. These data suggest that $\texttt{tol}_3 = 10^{-3}$ is small enough to approximate the $t \to \infty$ behavior. We will fix such a value hereafter.

6.3.2 Selection of m and scaling of convergence times

The last important choice is the value of the dimension (rank) parameter m. We know from [MS16] that the optimal value of the rank constrained problem (266) is within a relative error of order $O(1/m)$ of the value of the SDP (263). Also, a result by Burer and Monteiro [BM03] implies that, for $m \geq 2\sqrt{n}$, the objective function (266) has no local maxima that are not also global maxima (barring accidental degeneracies).

We empirically found that m of the order of 10 or larger is sufficient to obtain accurate results. Through most of our simulations, we fixed however $m = 100$, and we want to provide evidence that this is a safe choice

For each realization of the problem we compute the convergence time t_{conv} as the first time such that the condition $\Delta_{\max}(t) \leq \texttt{tol}_3 = 10^{-3}$ is met. In Figure 9 we plot histograms of $\log(t_{\text{conv}})$ for $n \in \{2000, 4000, 8000, 16000, 24000\}$ and $m \in \{20, 40, 100\}$. Here $d = 10$ and $\lambda = 1$, but t_{conv} does not seems to depend strongly on λ, d in the range we are interested in.

We observe that, for m large enough (in particular $m = 100$, see also data in Figure 10), the histogram of $\log(t_{\text{conv}})$ concentrates around its median. We interpret this as evidence of convergence towards a well defined global minimum, whose properties concentrate for n large. On the other

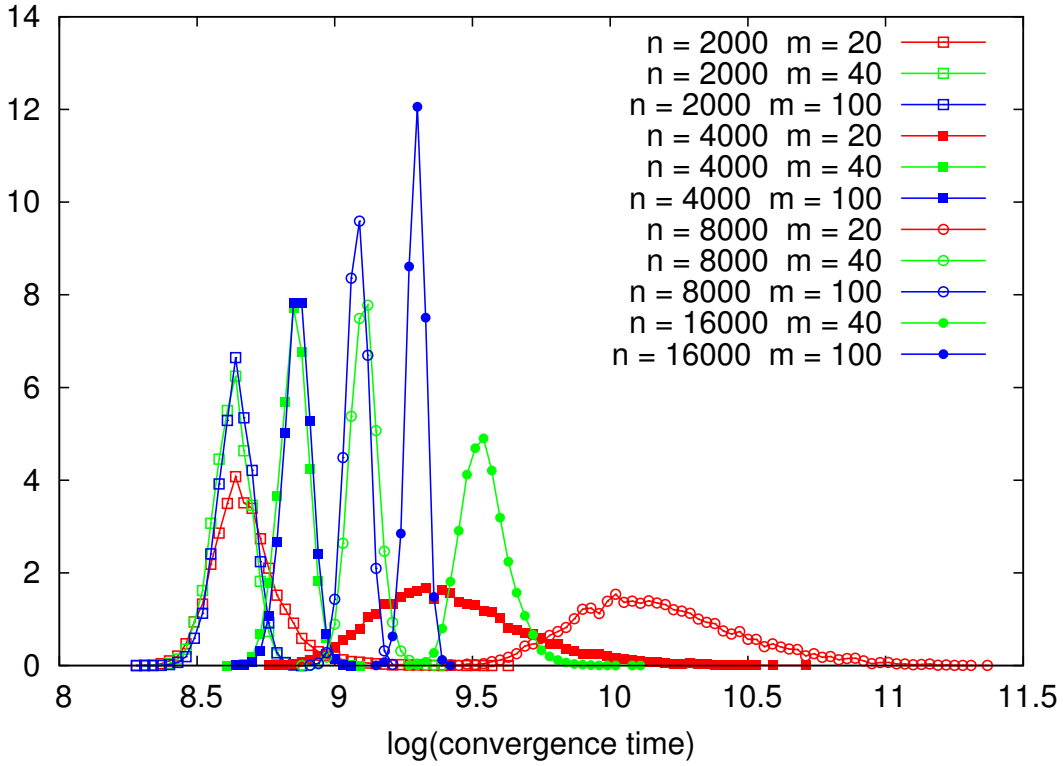


Figure 9: Histograms of the logarithm of the convergence time for the greedy algorithm run with $d = 10$ and $\lambda = 1$. Red, green and blue histograms are for $m = 20, 40$ and 100 respectively. System size increases from left to right histograms.

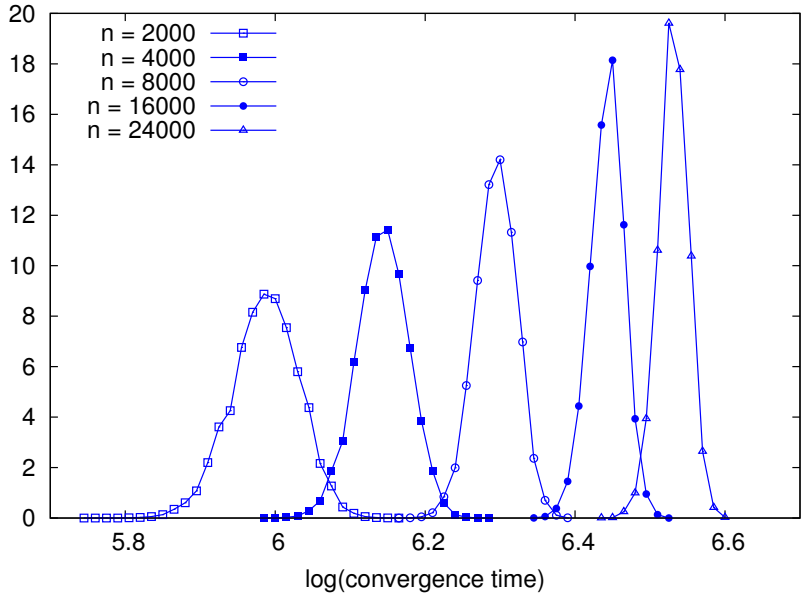


Figure 10: Histograms of the logarithm of the convergence times for $m = 100$, $d = 10$ and $\lambda = 1$.

hand, for m small, e.g. $m = 20$, the histogram broadens as n increases. This is a typical signature of convergence towards local minima, whose properties fluctuate from one graph realization to the other.

Intermediate values of m display a mixed behavior, with the histogram of convergence times concentrating for small n and broadening for larger n. This crossover behavior is consistent with the analytical results of [BA06]. Extrapolating this crossover suggests that $m = 100$ is sufficient for obtaining very accurate results for the range $n \lesssim 10^5$ of interest to us (and most likely, well above).

Focusing on $m = 100$ (data in Figure 10), we computed the mean and variance of $\log(t_{\rm conv})$, for each value of n. These appear to be well fitted by the following expressions

$$\mathbb{E}\log(t_{\rm conv}) \approx 4.3 + 0.22\,\log n\,, \tag{297}$$

$$\mathrm{Var}\Big(\log(t_{\rm conv})\Big) \approx 0.063 \cdot n^{-1/2}\,. \tag{298}$$

In other words the typical time complexity of our block coordinate ascent algorithm is –empirically– $O(m\,n^{1.22})$ (recall that each iteration comprises n updates).

6.3.3 Determination of the phase transition location

As already shown in Section 6.2, the overlap $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP})$ undergoes a phase transition at a critical point $\lambda_c^{\rm SDP}(d)$ close to 1. Namely $\lim_{n\to\infty}\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP}) = 0$ for $\lambda \le \lambda_c^{\rm SDP}(d)$, while $\lim_{n\to\infty}\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP}) > 0$ strictly for $\lambda > \lambda_c^{\rm SDP}(d)$. In order to determine more precisely the phase transition location, we use the Binder's cumulant method, which is standard in statistical physics [Bin81, LB14]. We summarize the main ideas of this method for the readers that might not be familiar with this type of analysis.

For a given graph realization G, we define $Q(G)$ to be the overlap achieved by the SDP estimator on that realization, i.e.

$$Q(G) \equiv \frac{1}{n}\langle \hat{\boldsymbol{x}}^{\rm SDP}(G), \boldsymbol{x_0}\rangle\,, \tag{299}$$

Notice that $Q(G)$ is a random variable taking values in $[-1, 1]$. Also, by the symmetry of the model, its distribution is symmetric around 0. Further $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP}) = \mathbb{E}\{|Q(G)|\}$.

We define the Binder cumulant by

$$\mathsf{Bind}(n,\lambda,d) \equiv \frac{\mathbb{E}\big\{Q(G)^4\big\}}{\mathbb{E}\big\{Q(G)^2\big\}^2}\,. \tag{300}$$

For $\lambda > \lambda_c^{\rm SDP}(d)$, we expect $|Q(G)|$ to concentrate around its expectation $\mathrm{Overlap}_n(\hat{\boldsymbol{x}}^{\rm SDP})$, which converges to a non-zero limit. Hence $\lim_{n\to\infty}\mathsf{Bind}(n,\lambda,d) = 1$. On the other hand, for $\lambda > \lambda_c^{\rm SDP}(d)$, $Q(G)$ concentrates around 0, and we expect it to obey a central limit theorem asymptotics, namely $Q(G) \approx \mathsf{N}(0, \sigma_Q^2(n))$, with $\sigma_Q^2(n) \approx \sigma_{Q,*}^2/n$. This implies $\lim_{n\to\infty}\mathsf{Bind}(n,\lambda,d) = 3$. Summarizing

$$\lim_{n\to\infty}\mathsf{Bind}(n,\lambda,d) = \begin{cases} 3 & \text{if } \lambda < \lambda_c^{\rm SDP}(d), \\ 1 & \text{if } \lambda > \lambda_c^{\rm SDP}(d). \end{cases} \tag{301}$$

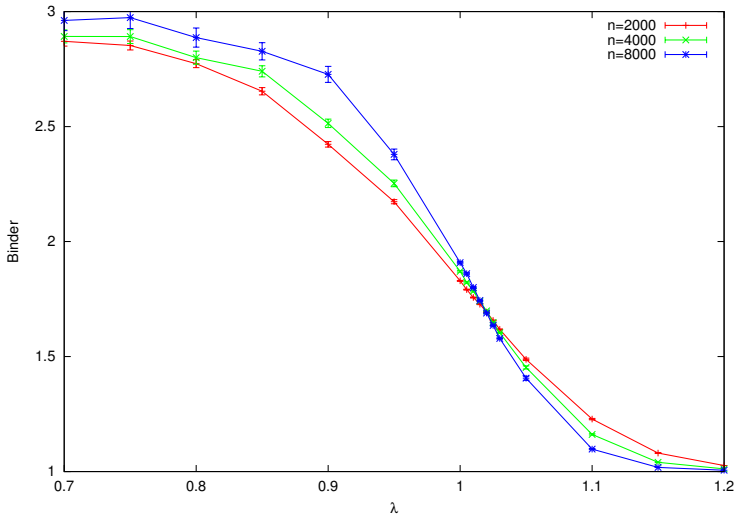


Figure 11: Empirical estimates of the Binder cumulant for $d = 5$.

We carried out extensive simulations with the block coordinate ascent, in order to evaluate the Binder cumulant, and will present our data in the next plots. In order to approximate the expectation over the random graph G, we computed empirical averages over N_{sample} random graph samples, with N_{sample} chosen so that $N_{\text{sample}} \times n = 6.4\ 10^8$. (The rationale for using less samples for larger graph sizes is that we expect statistical uncertainties to decrease with n.)

Figure 11 reports a first evaluation of $\mathsf{Bind}(n, \lambda, d)$ for $d = 5$ and a grid of values of λ. The results are consistent with the prediction of Eq. (301). The approach to the $n \to \infty$ limit is expected to be described by a finite-size scaling ansatz [Car12, LB14]

$$\mathsf{Bind}(n, \lambda, d) \approx \mathcal{F}\big(n^{1/\nu}(\lambda - \lambda_c^{\text{SDP}}(d))\big)\,, \tag{302}$$

for a certain scaling function $\mathcal{F}$, and exponent ν. Formally, the above approximation is meant to be asymptotically exact in the sense that, for any z fixed, letting $\lambda(z, n) = \lambda_c^{\text{SDP}}(p) + n^{-1/\nu} z$, we have $\lim_{n\to\infty} \mathsf{Bind}(n, \lambda(z, n), d) = \mathcal{F}(z)$. We refer to [BBC+01, DM08] for recent examples of rigorous finite-size scaling results in random graph problems.

In particular, finite size scaling suggests to estimate λ_c^{SDP} by the value of λ at which the curves $\lambda \mapsto \mathsf{Bind}(n, \lambda, d)$, corresponding to different values of n, intersect. In Figure 12 we report our data for $d = 2$, 5, 10, focusing on a small window around the crossing point. Continuous lines are linear fit to the data, and vertical lines correspond to the analytical estimates of Section 5.3.

We observe that, for large n, the crossing point is roughly independent of the the value of n, in agreement with the finite-size scaling ansatz. As a nominal estimate for the critical point, we use the crossing point $\lambda_\#(d)$ of the two Binder cumulant curves corresponding to the two largest values of n, see Fig. 12. These are $n = 32,000$ and $64,000$ for $d = 2$, and $n = 16,000$ and $32,000$ for $d = 5$, 10. We obtain

$$\lambda_\#(d = 2) = 1.010\,, \tag{303}$$
$$\lambda_\#(d = 5) = 1.016\,, \tag{304}$$
$$\lambda_\#(d = 10) = 1.012\,. \tag{305}$$

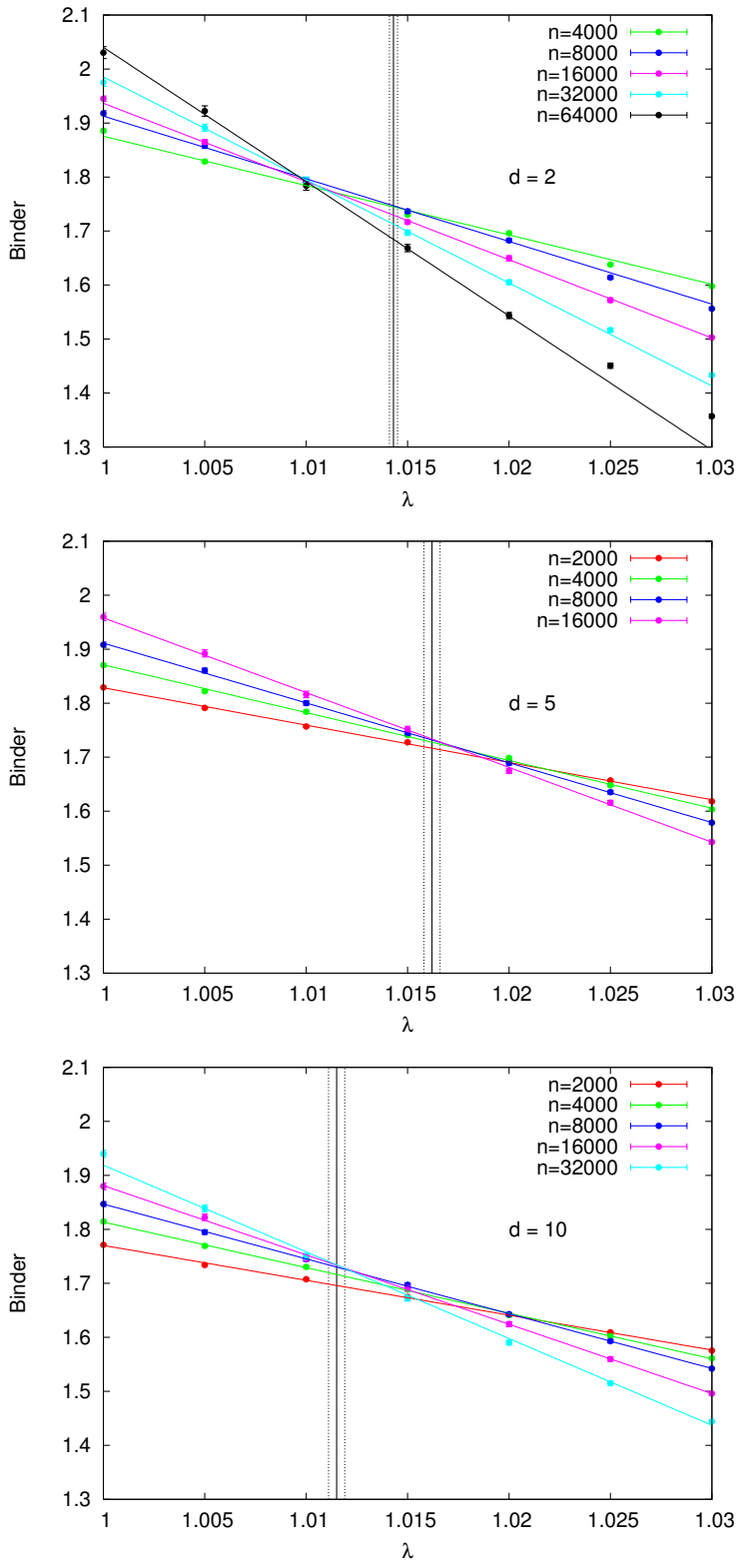


Figure 12: Crossings of the Binder parameters mark the critical value of λ. Vertical lines are the analytical estimates for the critical point $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$ (with dashed lines indicating the uncertainty in this estimate, due to numerical solution of the recursive distributional equation).

These values are broadly consistent with our analytical prediction for $\tilde{\lambda}_c^{\mathrm{SDP}}(d)$. There appear to be some discrepancy, especially for $d = 2$. This might be due to the graph-size being still too small for extrapolating to $n \to \infty$, or to the inaccuracy of our calculation based on the vectorial ansatz.

6.4 Improving numerical results by restricting to the 2-core

In order to accelerate our numerical experiments presented in Section 6.2 and 6.3, we preprocessed the graph G by reducing it to its 2-core. Recall that the k-core of a graph G is the largest subgraph of G, with minimum degree at least k. It can be constructed in linear time by recursively removing vertices with degree at most $(k-1)$.

In numerical experiments we first generated G_0 according to the model (262), then reduced G_0 to its 2-core G, and finally solved the SDP (263) on G. If G_0 has size n, and $d > 1$, the size of G is still of order n albeit somewhat smaller [PSW96].

The pruned graph $G \setminus G_0$ is formed with high probability by a collection of trees with size of order 1. It is not hard to see that the SDP estimator can achieve strictly positive overlap on G_0 (as $n \to \infty$) if and only if it does on G. Hence, this reduction does not change the phase transition location. We confirmed numerically this argument as well.

References

[1] Y. Deshpande, E. Abbe, and A. Montanari. Asymptotic mutual information for the two-groups stochastic block model. arXiv:1507.08685, 2015. 21

[2] A. Montanari, D. Reichman, and O. Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank one perturbations of gaussian tensors. arXiv:1411.6149, 2014. 21

[3] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, 1987. 21

[4] F. Guerra and F.L. Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230(1):71–79, 2002. 21

[5] G. Toulouse. On the mean field theory of mixed spin glass-ferromagnetic phases. *Journal de Physique Lettres*, 41(18):447–449, 1980. 21

[6] A. Crisanti, T. Rizzo, and T. Temesvari. On the parisi-toulouse hypothesis for the spin glass phase in mean-field theory. *The European Physical Journal B-Condensed Matter and Complex Systems*, 33(2):203–207, 2003. 21

[7] H.-J. Sommers. Properties of sompolinsky's mean field theory of spin glasses. *Journal of Physics A: Mathematical and General*, 16(2):447, 1983. 21

[8] R. Oppermann and M. J. Schmidt. Universality class of replica symmetry breaking, scaling behavior, and the low-temperature fixed-point order function of the sherrington-kirkpatrick model. *Physical Review E*, 78(6):061124, 2008. 21

[9] 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*, pages 1–47, 2009. 22

[10] M. Mézard and A. Montanari. *Information, Physics and Computation*. Oxford, 2009. 22

[11] J.R. Banavar, D. Sherrington, and N. Sourlas. Graph bipartitioning and statistical mechanics. *Journal of Physics A: Mathematical and General*, 20(1):L1, 1987. 22

[12] D. Sherrington and K.Y.M. Wong. Graph bipartitioning and the bethe spin glass. *Journal of Physics A: Mathematical and General*, 20(12):L785, 1987. 22

[13] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *Information Theory, IEEE Transactions on*, 62(1):471–487, 2016. 22

[14] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. In *Information Theory (ISIT), 2015 IEEE International Symposium on*, pages 1442–1446. IEEE, 2015. 22

[15] A. S. Bandeira. Random laplacian matrices and convex relaxations. arXiv:1504.03987, 2015. 22

[16] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. arXiv:1502.07738, 2015. 23

[17] N. Agarwal, A. S. Bandeira, K. Koiliaris, and A. Kolla. Multisection in the stochastic block model using semidefinite programming. arXiv:1507.02323, 2015. 23

[18] A. Montanari and S. Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*. ACM, 2016. 23, 43, 49, 50

[19] O. Guédon and R. Vershynin. Community detection in sparse networks via grothendieck's inequality. arXiv:1411.4686, 2014. 23

[20] A. Dembo and A. Montanari. Gibbs measures and phase transitions on sparse random graphs. *Brazilian Journal of Probability and Statistics*, 24(2):137–211, 2010. 23

[21] R. Lyons, R. Pemantle, and Y. Peres. Unsolved problems concerning random walks on trees. In *Classical and modern branching processes*, pages 223–237. Springer, 1997. 26, 27

[22] R. Lyons and Y. Peres. Probability on trees and networks. Citeseer, 2013. 26, 27

[23] D. J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Annals of Applied Probability*, pages 1047–1110, 2005. 26

[24] R. Lyons. Random walks and percolation on trees. *The Annals of Probability*, pages 931–958, 1990. 27

[25] M. Mézard and G. Parisi. The Bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001. 31

[26] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013. 46

[27] A. Saade, F. Krzakala, and L. Zdeborová. Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems*, pages 406–414, 2014. 46

[28] L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014. 46

[29] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. arXiv:1311.4115, 2013. 46

[30] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *Foundations of Computer Science (FOCS), 2015 IEEE 55th Annual Symposium on*, 2015. 46

[31] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010. 46

[32] S. Burer and R. D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003. 50

[33] A. Braun and T. Aspelmeier. The m-component spin glass on a bethe lattice. *Physical Review B*, 74(14):144205, 2006. 52

[34] K. Binder. Finite size scaling analysis of ising model block distribution functions. *Zeitschrift für Physik B Condensed Matter*, 43(2):119–140, 1981. 52

[35] D. P. Landau and K. Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, 2014. 52, 53

[36] J. Cardy. *Finite-size scaling*. Elsevier, 2012. 53

[37] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim, and D. B. Wilson. The scaling window of the 2-sat transition. *Random Structures & Algorithms*, 18(3):201–256, 2001. 53

[38] A. Dembo and A. Montanari. Finite size scaling for the core of large random hypergraphs. *The Annals of Applied Probability*, 18(5):1993–2040, 2008. 53

[39] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giantk-core in a random graph. *Journal of Combinatorial Theory, Series B*, 67(1):111–151, 1996. 55