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 $x_{0,1}, x_{0,2}, \ldots, x_{0,n}$, 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

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

(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 $\{1, \ldots, n\}$. 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 $V = [n] \equiv \{1, 2, \ldots, n\}$ 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

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

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):

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

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 $\mathrm{MSE}_n(\boldsymbol{0}) = 1$.

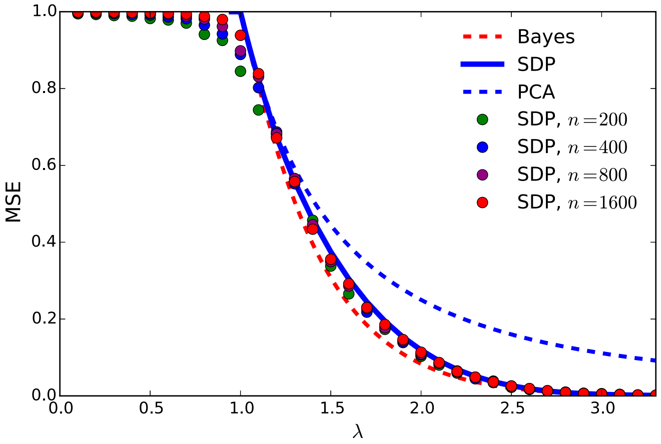

Fig. 1. Estimating $\boldsymbol{x}_0 \in \{+1, -1\}^n$ under the noisy $\mathbb{Z}_2$ synchronization model of Eq. **1**. Curves correspond to (asymptotic) analytical predictions, and dots correspond to numerical simulations (averaged over 100 realizations).

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:

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

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

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

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):

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

We use $\langle\cdot,\cdot\rangle$ to denote the scalar product between matrices, namely, $\langle \boldsymbol{A},\boldsymbol{B}\rangle\equiv \mathrm{Tr}(\boldsymbol{A}^{*}\boldsymbol{B})$, 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 $\xi_1\geq\xi_2\geq\cdots\geq\xi_n\geq 0$ 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

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

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 $\lambda\geq 1$.

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, $\hat{\boldsymbol{x}}:\mathcal{G}_n\to\{+1,-1\}^n$, $G\mapsto\hat{\boldsymbol{x}}(G)$. We will measure the performances of such an estimator through the overlap,

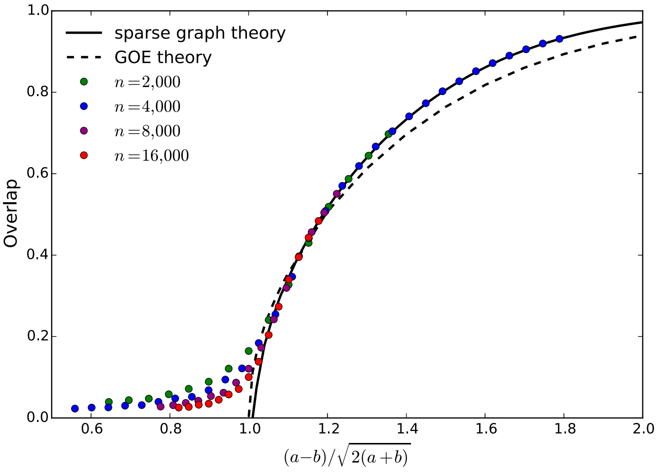

Fig. 2. Community detection under the hidden partition model of Eq. **2**, for average degree $(a+b)/2=5$. Dots indicate performance of the SDP reconstruction method (averaged over 500 realizations). Dashed curve indicates asymptotic analytical prediction for the Gaussian model (which captures the large-degree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

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

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,

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

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

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

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,

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

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.

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

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 $\text{Overlap}_n(\hat{\boldsymbol{x}}) \geq \delta > 0$ 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 $\lambda_c^{\text{SDP}} = \lambda_c^{\text{SDP}}(d)$ such that if

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

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

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

Fig. 3 plots our prediction for the function $\lambda_c^{\text{SDP}}(d)$, 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 $\lambda_c^{\text{SDP}}(d) > 1$ strictly, but the gap $\lambda_c^{\text{SDP}}(d) - 1$ 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 $\mathbb{C}$.

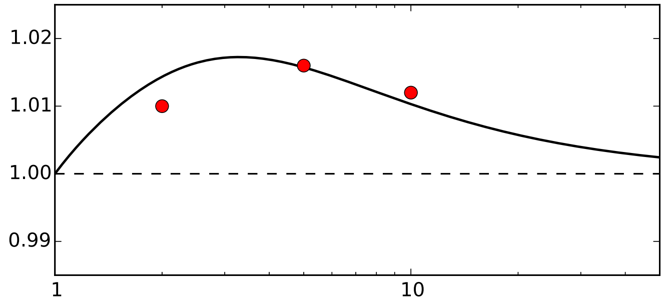

Fig. 3. Phase transition for the SDP estimator: for $\lambda > \lambda_c^{\text{SDP}}(d)$, the SDP estimator has positive correlation with the ground truth; for $\lambda \leq \lambda_c^{\text{SDP}}(d)$ the correlation is vanishing [here $\lambda = (a-b)/\sqrt{2(a+b)}$ and $d = (a+b)/2$]. Solid line indicates prediction $\tilde{\lambda}_c^{\text{SDP}}(d)$ 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 $\boldsymbol{\sigma}_1, \ldots \boldsymbol{\sigma}_n \in \mathbb{F}^m$ be the rows of $\boldsymbol{\sigma}$, the SDP [**6**] can be rewritten as

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

with $S^{m-1} = \{z \in \mathbb{F}^m : \|z\|_2 = 1\}$ 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 $\text{OPT}(\boldsymbol{Y}; m)$ the optimum value when the dimension is m (in particular, $\text{OPT}(\boldsymbol{Y}; m)$ 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

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

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:

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

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 $\hat{\boldsymbol{x}}^{\text{SDP}}$.

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

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

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 $i_1, \ldots, i_{n\alpha}$ 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).