Phase transitions in semidefinite relaxations

Adel Javanmard\textsuperscript{a}, Andrea Montanari\textsuperscript{b,c,1}, and Federico Ricci-Tersenghi\textsuperscript{d,e,f}

\textsuperscript{a}Marshall School of Business, University of Southern California, Los Angeles, CA 90089; \textsuperscript{b}Department of Statistics, Stanford University, Stanford, CA 94305; \textsuperscript{c}Department of Electrical Engineering, Stanford University, Stanford, CA 94305; \textsuperscript{d}Dipartimento di Fisica, Università La Sapienza, 00185 Rome, Italy; \textsuperscript{e}Istituto Nazionale di Fisica Nucleare, Sezione di Roma1, 00185 Rome, Italy; and \textsuperscript{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.

\textsuperscript{1}To whom correspondence should be addressed. Email: montanari@stanford.edu.

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

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.

E2218–E2223 | PNAS | Published online March 21, 2016

PNAS is a registered trademark of the National Academy of Sciences. Copyright © 2016 by the National Academy of Sciences, U.S.A.
As in the previous case, we assume observations to be distorted by Gaussian noise; that is, for each \( i < j \) we observe \( Y_{ij} = (\lambda/n) x_{0i} x_{0j} + W_{ij} \), where \( Z \) denotes complex conjugation\(^1\) and \( W_{ij} \sim \mathcal{CN}(0,1/n) \).

In matrix notations, this model takes the same form as [1], provided we interpret \( x_0^* \) as the conjugate transpose of vector \( x_0 \in \mathbb{C}^n \), with components \( x_{0i} |x_{0i}| = 1 \). 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 (\( Z_2 \) 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 partitioned) plant model (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 \( P(i \in V_+) = P(i \in V_-) = 1/2 \). Conditional on the partition, edges are independent with

\[
P((i,j) \in E|V_+, \overline{V}_-) = \begin{cases} a/n & \text{if } \{i,j\} \subseteq V_+ \text{ or } \{i,j\} \subseteq V_-; \\ b/n & \text{otherwise} \end{cases}
\]

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 \( x_0 \in \{+1,-1\}^n \), letting \( x_{0i} = +1 \) if \( i \in V_+ \) and \( x_{0i} = -1 \) if \( i \in V_- \). An important insight, which we will further develop below (25, 26), is that this problem is analogous to \( Z_2 \) synchronization, with signal strength \( \lambda \equiv (a-b)/\sqrt{2(a+b)} \). The parameters’ correspondence is obtained, at a heuristics level, by noting that if \( A_G \) is the adjacency matrix of \( G \), then \( E[x_{0i} A_G x_{0j}/|nE[A_G]|^{1/2}] = (a-b)/\sqrt{2(a+b)} \). (Here and below, \( [a,b] \equiv \sum a_ib_i \) 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 \( Z_2 \) synchronization problem, comparing them with numerical simulations using semidefinite programming (SDP). An estimator is a map \( \hat{x} : \mathbb{R}^{2n} \to \mathbb{R}^n \), \( \hat{x}(Y) \). We compare various estimators in terms of their per-coordinate mean square error (MSE):

\[
\text{MSE}_{\lambda}(\hat{x}) \equiv \frac{1}{n} \sum_{x_i \in \{+1,-1\}} \min_{c(x_0 \in \{+1,-1\}^n \text{ uniformly random}} \|x_0 - \hat{x}(Y)\|_2^2.
\]

where expectation is with respect to the noise model [1] and \( x_0 \in \{+1,-1\}^n \) 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 \( x_0 \) 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 \( \text{MSE}(\hat{x}) \)---to denote this limit. Note that a trivial estimator that always returns 0 has error \( \text{MSE}_0(\hat{x}) = 0 \).

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:

\[
\hat{x}_{\text{Bayes}}(Y) = \mathbb{E}\{x_0 | \lambda/n, XX^* + W = Y\}.
\]

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

### Maximum Likelihood

The estimator \( \hat{x}_{\text{ML}}(Y) \) is given by the solution of

\[
\hat{x}_{\text{ML}}(Y) = c(\lambda) \arg \max_{x_0 \in \{+1,-1\}^n} \langle x_0, XX^* \rangle.
\]

Here \( c(\lambda) \) is a scaling factor\(^3\) that is chosen according to the asymptotic theory as to minimize the MSE. As for the Bayes optimal curve, we obtain \( \text{MSE}(\hat{x}_{\text{ML}}) = 1 \) for \( \lambda \leq \lambda_{\text{ML}}^\text{Bayes} = 1 \) and \( \text{MSE}(\hat{x}_{\text{ML}}) < 1 \) (and rapidly decaying to 0) for \( \lambda > \lambda_{\text{ML}}^\text{Bayes} \). (We refer to SI Appendix for this result.)

\(^1\)In practical applications, \( \lambda \) 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 \( x_{\text{ML}}(Y) \) to keep MSE\((x_{\text{ML}}) \) ≤ 0.1.

\(^3\)Javanmard et al.
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 $X \in \mathbb{R}^{n \times n}$ (18, 29, 30):

$$\begin{align*}
\text{maximize } & \langle X, Y \rangle, \\
\text{subject to } & X \succeq 0, \quad X_{ii} = 1 \forall i \in [n].
\end{align*}$$

We use $\langle \cdot, \cdot \rangle$ to denote the scalar product between matrices, namely, $(A, B) \equiv \text{Tr}(A^T B)$, and $\Delta \succeq 0$ to indicate that $\Delta$ is positive semidefinite ($\text{PSD}$). If we assume $X = A x^*$, the SDP [6] reduces to the maximum-likelihood problem $\text{ML}$ ([7]). By dropping this condition, we obtain a convex optimization problem that is solvable in polynomial time. Given an optimizer $X_{\text{opt}} = X_{\text{opt}}(Y)$ 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 $X_{\text{opt}} = \sum_{i=1}^{\infty} \lambda_i v_i^T v_i$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and eigenvectors $v_i = v_i(X_{\text{opt}}(Y))$, with $\|v_i\|_2 = 1$. We then return the estimate

$$x_{\text{SDP}}(Y) = \sqrt{n} c_{\text{SDP}}(\xi) v_1(X_{\text{opt}}(Y)),$$

with $c_{\text{SDP}}(\xi)$ a certain scaling factor (SI Appendix).

Our analytical prediction for $\text{MSE}(x_{\text{SDP}})$ 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 $\text{MSE}(x_{\text{PCA}})$, the mean square error of principal component analysis (PCA). This method simply returns the principal eigenvector of $Y$, 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 $\xi > \xi_{\text{Bayes}}^{\text{opt}} = 1$. The same is achieved by a method as simple as PCA (blue-dashed curve). On the other hand, for $\xi < 1$, no method can achieve $\text{MSE}(x) < 1$ strictly [whereas $\text{MSE}(x^*) = 1$ is trivial by $x^* = 0$].

Second, PCA is suboptimal at large signal strength. PCA can be implemented efficiently but does not exploit the information $x_{0i} \in \{+1, -1\}$. As a consequence, its estimation error is significantly suboptimal at large $\xi$ (SI Appendix). Third, the SDP-based estimator is nearly optimal. The tractable estimator $x_{\text{SDP}}(Y)$ achieves the best of both worlds. Its phase transition coincides with the Bayes optimal one $\xi_{\text{Bayes}}^{\text{opt}} = 1$, and $\text{MSE}(x_{\text{SDP}})$ decays exponentially at large $\xi$, staying close to $\text{MSE}(x_{\text{Bayes}})$ and strictly smaller than $\text{MSE}(x_{\text{PCA}})$, for $\lambda > 1$.

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 $x_0 \in \{+1, -1\}^n$. 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 $G_n$, namely, $x: G_n \rightarrow \{+1, -1\}^n, G \mapsto x(G)$.

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

$$\text{Overlap}_n(x) = \frac{1}{n} \mathbb{E}\{|\langle x(G), x_0 \rangle|\},$$

and its asymptotic $n \rightarrow \infty$ 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,

$$x_{\text{ML}}(G) \equiv \arg \max_{x \in \{+1, -1\}^n} \left\{ \sum_{(i,j) \in E} x_i x_j : \langle x, 1 \rangle = 0 \right\},$$

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

$$\begin{align*}
\text{maximize } & \sum_{(i,j) \in E} X_{ij}, \\
\text{subject to } & X \succeq 0, \quad X_{11} = 0, \quad X_{ii} = 1 \forall i \in [n].
\end{align*}$$

Given an optimizer $X_{\text{opt}} = X_{\text{opt}}(G)$, we extract a partition of the vertices $V$ as follows. As for the $Z_2$ synchronization problem, we compute the principal eigenvector $v_1(X_{\text{opt}})$. We then partition $V$ according to the sign of $v_1(X_{\text{opt}})$. Formally,

$$x_{\text{SDP}}(G) = \text{sign}(v_1(X_{\text{opt}}(G))).$$

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 $Z_2$ synchronization model, with Gaussian noise (the GOE theory). This can be shown to capture the large degree limit: $d = (a + b)/2 \rightarrow \infty$, 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 $A$ is said to be PSD if all of its eigenvalues are nonnegative.

E2220 | www.pnas.org/cgi/doi/10.1073/pnas.1523097113
Javanmard et al.
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 OverlapP (x) ≥ Δ > 0 as n → ∞, if λ = (a−b)/√(a+b) < 1. Fig. 2 (and the theory developed in the next section) suggests that SDP has a phase transition threshold. Namely, there exists λSDP ≥ λSDP (d) such that if

\[ \lambda = \frac{a - b}{\sqrt{2(a + b)}} \geq \lambda_{SDP}(d = (a + b)/2), \]

then SDP achieves overlap bounded away from zero: OverlapP,SDP > 0. Fig. 2 also suggests λSDP(d) ≈ λBayes = 1; that is, SDP is nearly optimal.

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

Fig. 3 plots our prediction for the function \( \lambda_{SDP}(d) \), together with deterministically 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 \( \lambda_{SDP}(d) > 1 \) strictly, but the gap \( \lambda_{SDP}(d) - 1 \) 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 with 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 synchronizability under Gaussian noise, treating together the \( Z_2 \) 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 \( (Z_2) \) 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 \( \mathbb{C} \).

**Gibbs Measures and Vector Spin Models.** We start by recalling that a matrix \( \mathbf{X} \in \mathbb{F}^{n \times n} \) is PSD if and only if it can be written as \( \mathbf{X} = \mathbf{a} \mathbf{a}^T \) for some \( \mathbf{a} \in \mathbb{F}^{n \times 1} \). Indeed, without loss of generality, one can take \( m = n \), and any \( m > n \) is equivalent.

Letting \( \sigma_1, \ldots, \sigma_n \in \mathbb{F}^n \) be the rows of \( \sigma \), the SDP \( [6] \) can be rewritten as

\[ \text{maximize} \sum_{ij} Y_{ij}(\sigma_i, \sigma_j), \]

subject to \( \sigma_i \in \mathbb{F}^{m-1} \forall i \in [n] \), with \( \mathbb{F}^{m-1} = (x \in \mathbb{F}^m : \|x\|_2 = 1) \) the unit sphere in \( m \) dimensions. The SDP relaxation corresponds to any case \( m \leq n \) 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 \( \sigma_i \in \{-1, +1\} \) (for the real case) or \( \sigma_i \in U(1) \subset \mathbb{C} \) (for the complex case). Hence, we recover the maximum-likelihood estimator setting \( m = 1 \). It is also known that (under suitable conditions on \( Y \)) 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, for \( m < n \). Denote by OPT\( (Y;m) \) the optimum value when the dimension is \( m \) (in particular, OPT\( (Y;m) \)) is also the value of \( [6] \) for \( m = n \). It was proven in ref. 25 that there exists a constant \( C \) independent of \( m \) and \( n \) such that

\[ (1 - \frac{C}{m}) \text{OPT}(Y;\infty) \leq \text{OPT}(Y;m) \leq \text{OPT}(Y;\infty), \]

with probability converging to one as \( n \to \infty \) (whereby \( Y \) 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:

\[ p_{\beta,m}(\sigma) = \frac{1}{Z} \exp \left\{ 2m\beta \sum_{i<j} \Re(Y_{ij}(\sigma_i, \sigma_j)) \right\} \prod_{i=1}^{n} p_{0}(d\sigma_i), \]

where \( p_{0}(d\sigma_i) \) is the uniform measure over the \( m \)-dimensional sphere \( \mathbb{S}^{m-1} \) 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: \( Z_n \) 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 \( \lambda_{SDP} \).

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 \( \text{MSE}(\sigma_{\text{Bayes}}) \); on the basis of ref. 39, we **Fig. 3.** Phase transition for the SDP estimator: for \( \lambda > \lambda_{SDP}(d) \), the SDP estimator has positive correlation with the ground truth; for \( \lambda \leq \lambda_{SDP}(d) \) the correlation is vanishing (here \( \lambda = (a - b)/\sqrt{2(a + b)} \) and \( d = (a + b)/2 \)). Solid line denotes prediction \( \lambda_{SDP}(d) \) 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.
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 MSE$(x_{\text{ML}})$. Although this prediction is not expected to be exact (because of replica symmetry breaking), it should be nevertheless rather accurate, especially for large $\lambda$. (iv) By setting $m \to \infty$ and $\beta \to \infty$, we obtain the SDP estimation error MSE$(x_{\text{SDP}})$, 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 $\mu$, $q$, and $b \in \mathbb{R}$ that we state next. Let $Z \sim \text{N}(0,1)$ (in the real case) or $Z \sim \mathcal{CN}(0,1)$ (in the complex case). Define $\rho = \rho(Z; \mu, q, r)$ as the only nonnegative solution of the following equation in $(0, \infty)$:

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

Then $\mu$, $q$, and $r$ satisfy

$$\mu = \mathbb{E}\left[\frac{\mu + \sqrt{q} \mathbb{E}[Z]}{\rho + r}\right], \quad q = \mathbb{E}\left[\frac{|\mu + \sqrt{q} \mathbb{E}[Z]|^2}{(\rho + r)^2}\right],$$ \hspace{1cm} \text{(17)}

$$r = \mathbb{E}\left[\frac{1}{\rho} - \frac{\mu + \sqrt{q} \mathbb{E}[Z]}{\rho + r} - \frac{|Z|^2}{\rho + r}\right].$$ \hspace{1cm} \text{(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

$$\text{MSE}(x_{\text{SDP}}) = 1 - \frac{\mu(\lambda)^2}{2q(\lambda)}. \hspace{1cm} \text{(19)}$$

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

$$\text{Overlap}(x_{\text{SDP}}) = 1 - 2\Phi\left(-\frac{\mu(\lambda)}{\sqrt{2q(\lambda)}}\right). \hspace{1cm} \text{(20)}$$

The corresponding curve is plotted in Fig. 2.

More generally, for any dimension $m$ and inverse temperature $\beta$, we obtain equations that are analogous to Eqs. 17 and 18. The parameters $\mu$, $q$, and $b$ characterize the asymptotic structure of the probability measure $p_{\beta}(\sigma)$ defined in Eq. 15, as follows. We assume, for simplicity $\kappa_0 = (\ldots, +1, \ldots)$. Define the following probability measure on unit sphere $S^{m-1}$, parametrized by $\xi \in \mathbb{R}^m - 1, r \in \mathbb{R}:

$$\nu_{\lambda}(\sigma) = \frac{1}{2(\xi, r)} \exp\left\{2b m \mathbb{E}[R(\xi, \sigma)] - \beta mr |\xi|^2\right\} p_{\beta}(\sigma). \hspace{1cm} \text{(21)}$$

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

$$p_{\beta}^{(m, \beta)}(\sigma_{(1)}, \ldots, \sigma_{(k)}) \Rightarrow \int J_{\lambda_{(1)}^R, \ldots, \lambda_{(k)}^R} R(\cdot) \times \ldots R(\cdot) dR. \hspace{1cm} \text{(22)}$$

Here $dR$ denotes the uniform (Haar) measure on the orthogonal group, $\Rightarrow$ denotes convergence in distribution (note that $p_{\beta}^{(m, \beta)}(\sigma_{(1)}, \ldots, \sigma_{(k)})$ is a random variable), and $\xi_1, \ldots, \xi_k \sim N(\mu, \mathbb{Q})$ with $\mathbb{Q} = \text{diag}(q, q_0, \ldots, q_0)$, $q_0 = (1 - q)/(m - 1)$.  

\textbf{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 $\lambda \geq \lambda_{\text{SDP}}(d)$, for some $\lambda_{\text{SDP}}(d) > 1$.

Our prediction for the curve $\lambda_{\text{SDP}}(d)$ will be denoted by $\lambda_{\text{SDP}}^*(d)$ 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 $\lambda_{\text{SDP}}^*(d)$ approaches very quickly the ideal value $\lambda = 1$ for $d \to \infty$. Indeed, our prediction implies $\lambda_{\text{SDP}}^*(d) \to 1 + 1/(8d) + O(d^{-2})$.

Also, $\lambda_{\text{SDP}}^*(d) \to 1$ as $d \to 1$. This is to be expected because the constraints $a \geq b \geq 0$ imply $(a - b)/2 \leq d$, 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 $\lambda_{\text{SDP}}^*(d) / \sqrt{d}$.

More interestingly, $\lambda_{\text{SDP}}^*(d)$ 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)

$$c = \sum_{i=1}^{L} \frac{c_i}{1 + c_i}. \hspace{1cm} \text{(23)}$$

Here $c$ denotes equality in distribution, $L \sim \text{Poisson}(d)$, and $c_1, \ldots, c_L$ 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 $i \geq 0$, let $c_{i+1} = \sum_{k=1}^{\min(i, L)} (1 + c_k)$. It is proved in ref. 42 that the resulting sequence converges in distribution to a solution of Eq. 23: $c^* \overset{d}{=} c$.

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, $c_* = 0$ almost surely for $d \leq 1$, and $c_* > 0$ with positive probability if $d > 1$ (see ref. 42 and SI Appendix).

Next consider the distributional recursion

$$(c_{i+1}, R_{i+1}) \overset{d}{=} \left(\sum_{i=1}^{\min(i, L)} \frac{c_j}{1 + c_j}, \sum_{i=1}^{\min(i, L)} \frac{s_{L,i}^j}{\sqrt{1 + c_j}} \right). \hspace{1cm} \text{(24)}$$

where $s_{1,i} = +1$, $s_{L+1,i} = 1$, $\ldots$, $s_{L,L} + = 1$, and we use initialization $(c_{0, R}) = (+, +1)$. This recursion determines sequentially...
the distribution of \((c^{\text{rel}}, h^{\text{rel}})\) from the distribution of \((c^0, h^0)\). Here \(L, l \sim \text{Poiss}(d + \lambda/2),\) \(L, l \sim \text{Poiss}(d - \lambda/2),\) and \((c^0, h^0), \ldots, (c^L, h^L)\) are i.i.d. copies of \((c^0, h^0)\), independent of \(L, l\). Notice that because \(L + l \sim \text{Poiss}(d)\), we have \(c^{\text{rel}} \equiv c^{\text{rel},C}\). The threshold \(\lambda_c(d)\) is defined as the smallest \(\lambda\) such that the \(h^L\) diverges exponentially:

\[
\lambda_c(d) \equiv \inf \left\{ \lambda \in [0, \sqrt{2}] : \liminf_{i \to \infty} \frac{1}{\lambda} \log \mathbb{E} \left[ |h^L|^2 \right] > 0 \right\}. \tag{25}
\]

This value can be computed numerically, for instance, by sampling the recursion \cite{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 \([\text{in } \mathbb{Z}_2\text{ and } U(1)\text{ synchronization}]\) 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—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 \((a-b)/\sqrt{2(a+b)} = 1\) \cite{25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 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 \(\alpha \leq \{0.1\}\), we draw \(n \times n\) vertices \(i_1, \ldots, i_n\) uniformly at random in \(G\). For each such vertex \(i_k\) we connect by edges all of the neighbors of \(i_k\). In our case, this results in adding \(O(n\alpha^2)\) 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.