SEMIDEFINITE PROGRAMS ON SPARSE RANDOM GRAPHS

ANDREA MONTANARI AND SUBHABRATA SEN

ABSTRACT. Denote by \( A \) the adjacency matrix of an Erdős-Rényi graph with bounded average degree. We consider the problem of maximizing \( \langle A - E\{A\}, X \rangle \) over the set of positive semidefinite matrices \( X \) with diagonal entries \( X_{ii} = 1 \). We prove that for large (bounded) average degree \( \gamma \), the value of this semidefinite program (SDP) is \( \gamma \cdot n \) with high probability.

Our proof is based on two tools from different research areas. First, we develop a new ‘higher-rank’ Grothendieck inequality for symmetric matrices. In the present case, our inequality implies that the value of the above SDP is arbitrarily well approximated by optimizing over rank-\( k \) matrices for \( k \) large but bounded. Second, we use the interpolation method from spin glass theory to approximate this problem by a second one concerning Wigner random matrices instead of sparse graphs. As an application of our results, we prove new bounds on community detection via SDP that are substantially more accurate than the state of the art.

1. Main results

1.1. Semidefinite programs on sparse random graphs. Let \( G = (V, E) \) be a random graph with vertex set \( V = [n] \), and let \( A_G \in \{0, 1\}^{n \times n} \) denote its adjacency matrix. Spectral algorithms have proven extremely successful in analyzing the structure of such graphs under various probabilistic models. Interesting tasks include finding clusters, communities, latent representations, and so on [AKS98, McS01, NJW+02, CO06]. The underlying mathematical justification for these applications can be informally summarized as follows (more precise statements are given below):

If \( G \) is dense enough, then \( \langle A_G - E\{A_G\}, X \rangle \) is ‘much smaller’ than \( E\{A_G\} \).

However, it was repeatedly observed that this principle breaks down for random graphs with bounded average degree [FO05, CO10, KMO10, DPKMZ11, KMM+13], and that spectral methods consequently fail in this case. In order to focus on the simplest non-trivial instance of this phenomenon, assume that \( G \sim G(n, \gamma/n) \) is an Erdős-Rényi random graph with edge probability \( \gamma/n \). Then –letting \( \lambda_{\max}(\cdot) \) denote the largest eigenvalue– we have\(^1\) \( \lambda_{\max}(E(A_G)) = \gamma \). On the other hand, with high probability [KS03, Vu05]

\[
\lambda_{\max}(A_G - E(A_G)) = \begin{cases} 2\sqrt{\gamma}(1 + o(1)) & \text{if } \gamma \gg (\log n)^4, \\ \sqrt{\log n/(\log \log n)(1 + o(1))} & \text{if } \gamma = O(1). \end{cases} \tag{1.1}
\]

(The same behavior holds for the second-largest eigenvalue \( \lambda_2(A_G) \)).

In particular, \( \lambda_{\max}(A_G - E(A_G)) \gg \lambda_{\max}(E(A_G)) \) for bounded average degree \( \gamma \). This phenomenon is not limited to Erdős-Rényi random graphs, and instead leads to failures of spectral methods for

---

\(^1\)For mathematical convenience, we assume \( (A_G)_{ii} \in \{0, 1\} \) with \( P((A_G)_{ii} = 1) = \gamma/n \).
many probabilistic models in the same regime. The origin of this behavior is also well-understood: large eigenvalues correspond to eigenvectors localized close to vertices with high degree in \( G \).

A natural approach to overcome this problem is to use semidefinite programming (SDP), instead of spectral methods. Informally, SDP allows to rule out solutions that are localized on a small subset of vertices. Concretely, we consider the following graph parameters:

\[
M^+(G) \equiv \max \left\{ \langle A_G - \mathbb{E} A_G, X \rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n] \right\},
\]

\[
M^-(G) \equiv -\min \left\{ \langle A_G - \mathbb{E} A_G, X \rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n] \right\}.
\]

(Throughout, for matrices \( A, B \), we let \( \langle A, B \rangle = \text{Tr}(AB^T) \) denote the standard scalar product.) Equivalently, \( M^+(G) \) is the value of the following optimization problem (dropping the subscript \( G \) from the adjacency matrix and assuming \( G \sim G(n, \gamma/n) \))

\[
\text{maximize} \sum_{i,j=1}^{n} (A_{ij} - \gamma/n)\langle u_i, u_j \rangle,
\]

subject to \( u_i \in \mathbb{R}^n, \|u_i\|_2 = 1 \) for \( i \in [n] \).

and \( -M^-(G) \) is the solution of the corresponding minimization problem.

Our intuition is that the constraints \( \|u_i\|_2 = 1 \) rule out solutions concentrated on a few high-degree vertices, and hence the SDP should be insensitive to the ‘discontinuity’ exemplified by Eq. (1.1). Our main result yields a precise formalization of this idea\(^2\).

**Theorem 1.1.** Let \( G \sim G(n, \gamma/n) \) be an Erdős-Rényi random graph with edge probability \( \gamma/n \). Then with high probability we have

\[
\frac{M^+(G)}{\sqrt{n}} = 2\sqrt{\gamma} + o_\gamma(\sqrt{\gamma}), \quad \frac{M^-(G)}{\sqrt{n}} = 2\sqrt{\gamma} + o_\gamma(\sqrt{\gamma}).
\]

**Remark 2.** A possible interpretation of this result is that – unlike the largest, or second-largest eigenvalue - \( M^+(G)/n \) behaves ‘continuously’ for large bounded degrees and remains much smaller than the largest eigenvalue of the expected adjacency matrix\(^3\) \( \lambda_{\max}(\mathbb{E} A_G) \).

**Remark 3.** The quantity \( M^+(G) \) can also be thought as a relaxation of the problem of maximizing \( \sum_{i,j=1}^{n} A_{ij}u_iu_j \) over \( u_i \in \{+1, -1\}, \sum_{i=1}^{n} u_i = 0 \) (the \( -\gamma/n \) term being a Lagrangian version of the latter constraint). The result of our companion paper [DMS15] implies that this has – with high probability – value \( n(P_*/2)\sqrt{\gamma} + o(\sqrt{\gamma}) + o(n) \) (see [DMS15] for a definition of \( P_* \)). We deduce that – with high probability – the SDP relaxation overestimates the optimum by a factor \( 4/P_* + o_\gamma(1) \) (where \( 4/P_* \approx 5.241 \)).

**Remark 1.4.** For the sake of simplicity, we stated Eq. (1.6) in asymptotic form. However, our proof provides quantitative bounds on the error terms. In particular, the \( o_\gamma(\sqrt{\gamma}) \) term is upper bounded by \( C\gamma^{2/5} \log(\gamma) \), for \( C \) a numerical constant.

\(^2\) Throughout the paper, \( O(\cdot), o(\cdot), \) and \( \Theta(\cdot) \) stands for the usual \( n \to \infty \) asymptotic, while \( O_*(\cdot), o_*(\cdot) \) and \( \Theta_*(\cdot) \) are used to describe the \( \gamma \to \infty \) asymptotic regime. We say that a sequence of events \( A_n \) occurs with high probability (w.h.p.) if \( P(A_n) \to 1 \) as \( n \to \infty \). Finally, for random \( \{X_n\} \) and non-random \( f : \mathbb{R}^+ \to \mathbb{R}^+ \), we say that \( X_n = o_*(f(\gamma)) \) w.h.p. as \( n \to \infty \) if there exists non-random \( g(\gamma) = o_*(f(\gamma)) \) such that the sequence \( A_n = \{X_n \leq g(\gamma)\} \) occurs w.h.p. (as \( n \to \infty \)).

\(^3\)Notice in passing that, letting \( \mathcal{A} = \mathbb{E} A_G, \lambda_{\max}(\mathcal{A}) = n^{-1} \max \{\langle \mathcal{A} 11^T, X \rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n] \} \). In other words, it makes sense to compare \( M^+(G)/n \) to \( \lambda_{\max}(\mathcal{A}) \).
1.2. Higher-Rank Grothendieck Inequalities. A useful tool we develop is a Grothendieck-type inequality, which is of independent interest (see [KN12] for background). The inequality is more general than our graph setting and applies to general symmetric matrices.

In analogy with the definitions given in the previous section, we introduce the following quantities, for any symmetric matrix $B \in \mathbb{R}^{n \times n}$

$$Q_k(B) \equiv \max \left\{ \langle B, X \rangle : X \succeq 0, \ rank(X) \leq k, \ X_{ii} = 1 \forall i \in [n] \right\},$$

$$Q(B) \equiv \max \left\{ \langle B, X \rangle : X \succeq 0, \ X_{ii} = 1 \forall i \in [n] \right\}.$$

Note that the graph parameters introduced in the previous section can be expressed in terms of these, e.g. $M^+(G) = Q(A_G - \mathbb{E}A_G)$.

**Theorem 1.5.** For $k \geq 1$, let $g \sim \mathcal{N}(0, I_{k \times k}/k)$ be a vector with i.i.d. centered normal entries with variance $1/k$, and define

$$\alpha_k \equiv (\mathbb{E}\|g\|^2)^2.$$

Then, for any symmetric matrix $B$, we have the inequalities

$$Q(B) \geq Q_k(B) \geq \alpha_k Q(B) - (1 - \alpha_k) Q(-B),$$

$$Q_k(B) \geq \left( 2 - \frac{1}{\alpha_k} \right) Q(B) + \left( \frac{1}{\alpha_k} - 1 \right) Q_k(-B).$$

**Remark 1.6.** The upper bound in Eq. (1.10) is trivial. It follows from Cauchy-Schwartz that $\alpha_k \in (0, 1)$ for all $k$. Also $\|g\|^2_2$ is a chi-squared random variable with $k$ degrees of freedom and hence

$$\alpha_k = \frac{2\Gamma((k + 1)/2)^2}{k\Gamma(k/2)^2} = 1 - \frac{1}{2k} + O(1/k^2).$$

Substituting in Eq. (1.10) we get, for all $k \geq k_0$ with $k_0$ a sufficiently large constant, and assuming $Q(B) > 0,$

$$\left( 1 - \frac{1}{k} \right) Q(B) - \frac{1}{k} |Q(-B)| \leq Q_k(B) \leq Q(B).$$

In particular, if $|Q(-B)|$ is of the same order as $Q(B)$, we conclude that $Q_k(B)$ approximates $Q(B)$ with a relative error of order $O(1/k)$.

The classical Grothendieck inequality concerns non-symmetric bilinear forms [Gro96]. A Grothendieck inequality for symmetric matrices was established in [NRT99, Meg01] (see also [AMMN06] for generalizations) and states that, for a constant $C$,

$$Q_1(B) \geq \frac{1}{C \log n} Q(B).$$

Higher-rank Grothendieck inequalities were developed in the setting of general graphs in [Bri10, BdOFV10]. However, constant-factor approximations were not established for the present problem (the complete graph, in the setting of [Bri10]).

Constant factor approximations exist for $B$ positive semidefinite [BdOFV10]. However we want to apply Theorem 1.5 – among others – to $B = A_G - \mathbb{E}A_G$ with $A_G$ the adjacency matrix of a Erdős-Rényi graph with average degree $\gamma = O(1)$. This matrix is non-positive definite, and in a dramatic way (the smallest eigenvalue being approximately $-(\log n/\log \log n)^{1/2}$). On the other hand, Eq. (1.10) can be weakened by using $Q(-B) \leq -\lambda_{\text{min}}(B)$, yielding Grothendieck inequality of [BdOFV10] for the positive semidefinite matrix $B - \lambda_{\text{min}}(B)$.
In summary, we could not use the vast literature on Grothendieck-type inequality to prove our main result, Theorem 1.1, which motivated us to develop Theorem 1.5.

1.3. Application to community detection. Significant attention has been devoted recently to the community detection problem under the so-called ‘stochastic block model’ in the regime of bounded average degree [DKMZ11, KMM+13, Mas14]. This regime is particularly challenging mathematically because of the heterogeneity of vertex degrees.

To be definite we will formalize this as a hypothesis testing problem, whereby we want to determine—with high probability of success—whether the random graph under consideration has a community structure or not. The estimation version of the problem, i.e. the question of determining—an approximately—a partition into communities, can be addressed by similar techniques and will be considered in a future publication. We are given a single graph \( G = (V, E) \) over \( n \) vertices and we have to decide which of the following holds:

**Hypothesis 0**: \( G \) is an Erdős-Rényi random graph with edge probability \((a+b)/(2n), \ G \sim G(n, (a+b)/(2n))\). We denote the corresponding distribution over graphs by \( \mathbb{P}_0 \).

**Hypothesis 1**: There is a set \( S \subseteq [n] \), that is uniformly random given its size \(|S| = n/2\). Conditional on \( S \), edges are independent with

\[
\mathbb{P}_1((i,j) \in E|S) = \begin{cases} 
\frac{a}{n} & \text{if } \{i,j\} \subseteq S \text{ or } \{i,j\} \subseteq [n]\setminus S, \\
\frac{b}{n} & \text{if } i \in S, j \in [n]\setminus S \text{ or } i \in [n]\setminus S, j \in S. 
\end{cases}
\] (1.15)

The corresponding distribution over graphs is denoted by \( \mathbb{P}_1 \).

A statistical test takes as input a graph \( G \), and returns \( T(G) \in \{0, 1\} \) depending on which hypothesis is estimated to hold. We say that it is successful with high probability if \( \mathbb{P}_0(T(G) = 1) + \mathbb{P}_1(T(G) = 0) \to 0 \) as \( n \to \infty \).

It is convenient to generalize slightly the definition (1.2) by letting, for \( \lambda \in \mathbb{R} \) a regularization parameter,

\[
M^+(G; \lambda) \equiv \max \left\{ \langle A_G - \lambda I I^T, X \rangle : X \succeq 0, X_{ii} = 1 \\forall i \in [n] \right\}. \] (1.16)

Theorem 1.1 indicates that, under Hypothesis 0, and letting \( \gamma = (a+b)/2 \) be the average degree, we have \( M^+(G; \gamma/n)/n = 2\sqrt{\gamma} + o_\gamma(\sqrt{\gamma}) \). This suggests the following test:

\[
T(G; \delta) = \begin{cases} 
1 & \text{if } M^+(G; (a+b)/(2n)) \geq n(1+\delta)\sqrt{2(a+b)}, \\
0 & \text{otherwise.} 
\end{cases}
\] (1.17)

**Theorem 1.7.** Assume, for some \( \varepsilon > 0 \),

\[
\frac{a-b}{\sqrt{2(a+b)}} \geq 2 + \varepsilon. \] (1.18)

Then there exists \( \delta_\varepsilon = \delta_\varepsilon(\varepsilon) > 0 \) and \( \gamma_\varepsilon = \gamma_\varepsilon(\varepsilon) > 0 \) such that the following holds. If \( (a+b)/2 \geq \gamma_\varepsilon \), then the SDP-based test \( T(\cdot; \delta_\varepsilon) \) succeeds with high probability.

**Remark 1.8.** Mossel, Neeman, Sly [MNS12] proved that no test can be successful with high probability if \( (a-b)/\sqrt{2(a+b)} \). The above theorem guarantees that SDP is successful—roughly—within a factor 2 from this threshold.

**Remark 1.9.** The factor 2 on the right hand side of Eq. (1.18) is not the best possible one. Indeed a somewhat more involved proof (see Appendix C and Appendix E) yields the condition \( (a-b)/\sqrt{2(a+b)} \geq \xi_0 + \varepsilon \) for some \( \xi_0 \in (1, 2) \) strictly. In fact we expect the optimal constant \( \xi_\varepsilon \) for which this theorem holds to be \( \xi_\varepsilon = 1 \).
In Appendix C we develop a characterization of the constant $\xi_*$ in terms of the value of an SDP with gaussian data. We believe that this might give a way to prove $\xi_* = 1$.

Remark 1.10. One might wonder why we consider large degree asymptotics $(a + b) \to \infty$ instead of trying to establish a threshold at $(a - b)^2 / 2(a + b) = 1$ for fixed $a, b$. Preliminary non-rigorous calculation [JMRT15] suggest that indeed this is necessary. For fixed $(a + b)$ the SDP threshold does not coincide with the optimal one.

The semidefinite program (1.2) was analyzed mostly in a regime of diverging degrees $a, b = \Theta(\log n)$ [ABH14, HWX14, HWX15]. In that setting SDP recovers exactly the correct partition of vertices and hence one can apply more standard dual witness techniques to carry out the analysis.

The only earlier result that compares to ours was recently proven by Guedon and Vershynin [GV14]. Their work establishes upper bounds on the estimation error of SDP, that apply only under the condition $(a - b)^2 \geq 10^4 (a + b)$. The same paper also provides guarantees for a more general community structure. We defer the analysis of estimation error and more general community structures using our methods to future work.

1.4. Organization of the paper. In Section 2, we state some auxiliary results and prove Theorem 1.1. As mentioned, the proof is based on two tools that come from different research communities. On one hand, we use the Grothendieck-type inequality of Theorem 1.5, whose proof uses a Rietz’s randomized rounding method, and is deferred to Appendix A. On the other, we exploit a connection with statistical physics, and develop an interpolation between sparse graphs and Wigner matrices in Section 3. The interpolation argument requires to prove uniform continuity ‘at zero temperature’ as well as a separate analysis of the Wigner matrix problem, carried out in Appendix B. Finally, Appendices C, D, E contain proofs and complementary results on community detection.

2. Other results and proof of Theorem 1.1

We begin by defining a rank-$k$ version of the parameters $M^\pm(G)$. Namely, for $k$ an integer we let

$$M^+_k(G) \equiv \max \left\{ \langle A_G - \mathbb{E} A_G, X \rangle : X \succeq 0, \; \text{rank}(X) \leq k, \; X_{ii} = 1 \forall i \in [n] \right\},$$

or equivalently (assuming $G \sim G(n, \gamma/n)$), $M^+_k(G)$ is the value of the optimization problem

$$\text{maximize} \sum_{i,j=1}^{n} (A_{ij} - \gamma/n) \langle u_i, u_j \rangle,$$

subject to $u_i \in \mathbb{R}^k$, $\|u_i\|_2 = 1$ for $i \in [n]$.

We define $M^-_k(G)$ through the corresponding minimization problem.

The basic tool we will employ is an interpolation with Wigner matrices, that we will use together with the Grothendieck-type inequality Theorem 1.5.

2.1. Interpolation lemmas. For $W \in \mathbb{R}^{n \times n}$ a symmetric random matrix with $(W_{ij})_{i \leq j}$ independent and $W_{ij} \sim \mathcal{N}(0, 1/n)$ if $i < j$, $W_{ii} \sim \mathcal{N}(0, 2/n)$, we define

$$q_l(k) \equiv \lim \inf_{n \to \infty} \frac{1}{n} \mathbb{E} Q_k(W), \quad q_u(k) \equiv \lim \sup_{n \to \infty} \frac{1}{n} \mathbb{E} Q_k(W).$$

The first lemma connects these quantities to the sparse graph setting. For its proof we refer to Section 3.1.
Lemma 2.1. Let \( G \sim G(n, \gamma/n) \) be an Erdős-Rényi random graph with edge probability \( \gamma/n, \gamma \geq k \). Letting \( q_k\), \( q_u\) be defined as above, there exists an absolute constant \( C_0 \), such that, with high probability

\[
\frac{M_+^k(G)}{n}, \frac{M^-_k(G)}{n} \in \left[ q_k(\sqrt{\gamma} - C_0k^{2/3}\gamma^{1/3}\log(k\gamma)), q_u(\sqrt{\gamma} + C_0k^{2/3}\gamma^{1/3}\log(k\gamma)) \right]. \tag{2.5}
\]

This proof follows an interpolation method, following the approach developed by Guerra and Toninelli [GT04] in spin glass-theory, and recently applied in our companion paper [DMS15] to combinatorics. The basic idea is to continuously interpolate between the problem defined on a sparse Erdős-Rényi random graph, and the one in terms of a Wigner matrix.

Interpolation (or ‘smart path’) methods have a long history in probability theory, dating back to Lindeberg method [Lin22]. More specifically, interpolation methods from statistical physics have proven successful in the study of combinatorial problems on random graphs [FL03, FLT03, BGT13, PT04]. However, the type of interpolation used in these papers is different from the one studied here, and aimed at proving existence of the \( n \to \infty \) limit for certain graph parameters. Here we are interested in the large (bounded) degree asymptotics.

The next lemma controls the constants \( q_k\), \( q_u\) for large \( k \). Its proof can be found in Appendix B.

Lemma 2.2. Let \( q_k\), \( q_u\) be defined as above. Then, there exists an integer \( k_0\), such that, for all \( k \geq k_0\),

\[
2 - \frac{4}{\alpha_k} \leq q_k \leq q_u \leq 2. \tag{2.6}
\]

(Lemma C.1 proves the additional result that \( q_k = q_u\).)

2.2. Proof of Theorem 1.1. We will limit ourselves to proving the claim for \( M^+(G)\), since the analogous result for \( M^-(G)\) follows from a very similar argument.

Fix \( k \) a large integer. Then, applying Eq. (1.11) to \( B = A_G - \lambda 11^T\), we get

\[
M_+^k(G) \leq M^+(G) \leq \frac{\alpha_k}{2\alpha_k - 1} M^+_k(G) + \frac{1 - \alpha_k}{2\alpha_k - 1} M^-_k(G). \tag{2.7}
\]

(The first inequality is trivial.)

By Lemma 2.1, we have, letting \( \psi(k, \gamma) = C_0k^{2/3}\gamma^{1/3}\log(k\gamma)\),

\[
\lim_{n \to \infty} \mathbb{P} \left\{ q_k(\sqrt{\gamma} - \psi(k, \gamma) \leq \frac{1}{n} M^+(G) \leq \frac{1}{2\alpha_k - 1} \left[q_u(\sqrt{\gamma} + \psi(k, \gamma)\right] \right\} = 1. \tag{2.8}
\]

Since this holds for any \( k \) and \( \gamma \), we can take \( k = k(\gamma) = \gamma^{1/10} \to \infty \) with \( \gamma \). Notice that, with this setting \( \psi(k(\gamma), \gamma) = o(\sqrt{\gamma})\). Using \( \alpha_k = 1 - O(1/k)\), cf. Eq. (1.12) the claim follows.

Note indeed that the proof yields a more quantitative version of Theorem 1.1, namely \(|M^\pm(G)/n - 2\sqrt{\gamma}| \leq C\gamma^{2/5}\log(\gamma)\) with high probability.

3. Interpolation and Proof of Lemma 2.1

We consider \( M_+^k(G)\) in the proof. The proof for \( M^-_k(G)\), being similar, is omitted.

Throughout let \( W \in \mathbb{R}^{n \times n} \) be a symmetric random matrix with \((W_{ij})_{i \leq j}\) independent, \( W_{ij} \sim N(0, 1/n)\) if \( i < j \), \( i, j \in [n]\), and \( W_{ii} \sim N(0, 2/n)\) for \( i \in [n]\).

We further consider a slightly different (Poisson) random multi-graph model \( G_{n, \gamma}^{\text{Pois}} \). Under this model, the number of edges is Poisson\((n\gamma/2)\) and each edge has endpoints \((I, J)\) that are independent and uniformly random in \(\{1, 2, \ldots, n\}\). This can be coupled to the Erdős-Rényi random
graph in such a way that they differ in $O(1)$ edges with high probability. It is therefore sufficient to prove our claims under the Poisson model. We denote by $A$ the adjacency matrix of this graph (with every edge counted with its multiplicity).

Let $S^{k-1} = \{u \in \mathbb{R}^k : ||u||_2 = 1\}$ be the unit sphere in $k$ dimensions. We will consider the following Hamiltonians (measurable functions $(S^{k-1})^n \to \mathbb{R}$):

$$H^{\text{spar}}(\sigma) = \frac{1}{\sqrt{n}} \sum_{i,j=1}^{n} \left(A_{ij} - \frac{\gamma}{n}\right) \langle \sigma_i, \sigma_j \rangle,$$

$$H^{\text{den}}(\sigma) = \sum_{i,j=1}^{n} W_{ij} \langle \sigma_i, \sigma_j \rangle,$$

where $\sigma = (\sigma_1, \ldots, \sigma_n)$, $\sigma_i \in S^{k-1}$.

We note that the process $\{H^{\text{den}}(\sigma)\}$ is a Gaussian process with mean 0 and $\text{Cov}(H^{\text{den}}(\sigma), H^{\text{den}}(\sigma')) = \frac{2}{n} \sum_{ij} \langle \sigma_i, \sigma_j \rangle \langle \sigma'_i, \sigma'_j \rangle$.

We denote by $d\nu(\cdot)$ the uniform measure on $(S^{k-1})^n$ (normalized to 1, i.e. $\int d\nu(\sigma) = 1$). For each of $\star \in \{\text{spar}, \text{den}\}$ we define the (expected) free energy density

$$\phi^{\star}(\beta) = \frac{1}{n} \mathbb{E} \log \left\{ \int e^{\beta H^{\star}(\sigma)} d\nu(\sigma) \right\}.$$  \hfill (3.3)

Our proof of Lemma 2.1 is based on two auxiliary results, denoted as Lemma 3.1 and 3.2. The first result bounds the difference between $\phi^{\text{spar}}(\beta)$ and $\phi^{\text{den}}(\beta)$.

**Lemma 3.1.** With the above definitions, there exists a constant $C$ independent of $n$, $\gamma$ and $\beta$ such that, for all $\beta \leq \sqrt{\gamma}/10$,

$$|\phi^{\text{spar}}(\beta) - \phi^{\text{den}}(\beta)| \leq C \frac{\beta^3}{\gamma^{1/2}} + o(1).$$  \hfill (3.4)

**Proof.** We define an interpolating Hamiltonian. Namely, for $t \in [0, 1]$, we define $H_t : (S^{k-1})^n \to \mathbb{R}$, by

$$H_t(\sigma) = \frac{1}{\sqrt{n}} \sum_{i,j=1}^{n} \left(A_{ij}(t) - \frac{\gamma(1-t)}{n}\right) \langle \sigma_i, \sigma_j \rangle + \sqrt{t} \sum_{i,j=1}^{n} W_{ij} \langle \sigma_i, \sigma_j \rangle.$$ \hfill (3.5)

Here $A_{ij}(t)$ is the adjacency matrix of a graph with distribution $G^{\text{Poisson}}_{n, \gamma(1-t)}$. We further define the interpolating free energy

$$\phi(\beta; t) = \frac{1}{n} \mathbb{E} \log \left\{ \int e^{\beta H_t(\sigma)} d\nu(\sigma) \right\},$$ \hfill (3.6)

and notice that $\phi(\beta; 0) = \phi^{\text{spar}}(\beta)$, $\phi(\beta; 1) = \phi^{\text{den}}(\beta)$, whence

$$|\phi^{\text{spar}}(\beta) - \phi^{\text{den}}(\beta)| \leq \int_0^1 \left| \frac{\partial \phi(\beta; t)}{\partial t} \right| dt.$$ \hfill (3.7)

Therefore, it suffices to bound the derivative uniformly in $[0, 1]$. To this end, we introduce the interpolating Gibbs measure

$$\mu_{\beta; t}(\sigma) = \frac{\exp(\beta H_t(\sigma))}{\int \exp(\beta H_t(\tau)) d\nu(\tau)} d\nu(\sigma)$$ \hfill (3.8)
and obtain that
\[ \frac{\partial \phi(\beta; t)}{\partial t} = I + II, \] (3.9)
where
\[ I = \beta \sqrt{n} \frac{1}{n^2} \sum_{ij} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)] + \frac{\beta^2}{n^2} \sum_{ij} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle^2) - \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2], \] (3.10)
\[ II = -\frac{\gamma}{2n^2} \sum_{ij} \mathbb{E} \left[ \log \mu_{\beta,t} \left( \exp \left( \frac{2\beta}{\sqrt{n}} (\langle \sigma_i, \sigma_j \rangle) \right) \right) \right] + o(1). \] (3.11)

The calculation of these derivatives is relatively straightforward, and we present it in Section A.1. Using Taylor expansion with \( b = \frac{2\beta}{\sqrt{n}} \) and noting that \( |\langle \sigma_i, \sigma_j \rangle| \leq 1 \) for all \( i, j \), one has
\[ \exp(b(\langle \sigma_i, \sigma_j \rangle)) = 1 + b(\langle \sigma_i, \sigma_j \rangle) + \frac{b^2}{2} \langle \sigma_i, \sigma_j \rangle^2 + R_{ij}, \] (3.12)
where \( |R_{ij}| \leq b^3 e^b / 6. \)

Notice that, since \( \beta \leq \sqrt{n} / 10 \), we have \( b \leq 1 / 5 \). Defining \( \psi = b \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle) + \frac{b^2}{2} \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2 + \mu_{\beta,t}(R_{ij}) \), we have \( |\psi| \leq b + b^2 / 2 + b^3 e^b / 6 \leq 1 / 2 \). We can then use the Taylor approximation
\[ \left| \log(1 + \psi) - \psi - \frac{1}{2} \psi^2 \right| \leq \frac{|\psi|^3}{6(1 + \psi)^4} \leq C_1 b^3. \] (3.13)

Further substituting \( \psi \), we get
\[ \left| \log \mu_{\beta,t}(\exp(b(\langle \sigma_i, \sigma_j \rangle))) - \left( b \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle) + \frac{b^2}{2} \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2 \right) - \frac{b^2}{2} \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle) \right| \leq C_1' b^3. \] (3.14)

Combining Eqs. (3.9) to (3.11) with Eq. (3.14) completes the proof.

Next, for \( * \in \{ \text{spar, den} \} \), we define the ground state energy density \( e_n^* = \mathbb{E} \left[ \frac{1}{n} \sup_\alpha H^*(\alpha) \right] \). The next result bounds the difference between the free energy density and the ground state energy density.

**Lemma 3.2.** There exist constants \( C, C_0 \) and \( c \) independent of \( n, \gamma \) and \( k \) such that, for every \( \varepsilon > 0 \),
\[ \left| e_n^\text{den} - \frac{1}{\beta} \phi^\text{den}(\beta) \right| \leq 5 \varepsilon \sqrt{k} + C \frac{1}{\beta} \log \frac{k}{c \varepsilon} + o(1), \] (3.15)
\[ \left| e_n^\text{spar} - \frac{1}{\beta} \phi^\text{spar}(\beta) \right| \leq C_0 \varepsilon \sqrt{k} \gamma + C \frac{1}{\beta} \log \frac{k}{c \varepsilon} + o(1). \] (3.16)

**Proof.** For \( * \in \{ \text{spar, den} \} \), define the partition function \( Z^*(\beta) = \int \exp(\beta H^*(\alpha))d\nu(\alpha) \). Let \( H^*(\alpha) \) denote the maximum of the Hamiltonian, and denote by \( \alpha^* \) a maximizer (i.e. \( H^*(\alpha^*) = H^* \)). Each \( \alpha \in (S^{k-1})^n \) will also be thought of as a matrix in \( \mathbb{R}^{n \times k} \). We also define the event
\[ \mathcal{E}^* \equiv \{|H^*(\alpha) - H^*| \leq C \sqrt{n} \alpha \| \alpha - \alpha^* \|_F \ \forall \alpha \in (S^{k-1})^n \}, \] (3.17)
where \( \| \cdot \|_F \) is the Frobenius norm of a matrix. We choose \( C^\text{den} = 5 \) and \( C^\text{spar} = C_0 \sqrt{n} \) where \( C_0 \) is a sufficiently large constant independent of \( n, \gamma \) and \( k \). We claim that –with these choices– \( \mathbb{P}(\mathcal{E}^*) = 1 - o(1 / n) \) for both \( * \in \{ \text{den, spar} \} \). Before proving this claim, let us see how it implies the Lemma.
On the event $\mathcal{E}^*$, we have

$$e^{\beta H^*} \geq Z^*(\beta) \geq e^{\beta H^*} \int \exp(-\beta C^* \sqrt{k n} \|\overline{\sigma} - \overline{\sigma}^*\|_F) \, d\nu(\overline{\sigma}).$$  \hfill (3.18)

For any $\varepsilon > 0$, we have (here $\mathbb{I}(\cdot)$ denotes the indicator function)

$$\int \exp(-\beta C^* \sqrt{k n} \|\overline{\sigma} - \overline{\sigma}^*\|_F) \, d\nu(\overline{\sigma}) \geq \int \exp(-\beta C^* \sqrt{k n} \|\overline{\sigma} - \overline{\sigma}^*\|_F) \mathbb{I}(\max_{i \in [n]} \|\sigma_i - \sigma_i^*\|_2 \leq \varepsilon) \, d\nu(\overline{\sigma})$$

$$\geq \exp(-\beta C^* n \varepsilon \sqrt{k} (V_k(\varepsilon))^n),$$  \hfill (3.19)

where $V_k(\varepsilon)$ is volume of the spherical cap $\{y \in S^{k-1} : \|y - y^*\|_2 \leq \varepsilon\}$. Combining (3.18) and (3.19), we get

$$\left| e_n^* - \frac{1}{\beta} \phi^*(\beta) \right| \leq C^* \varepsilon \sqrt{k} + \frac{1}{\beta} \log \frac{1}{V_k(\varepsilon)} + \mathbb{E}\left[ \left\| \frac{H^*_n}{n} - \frac{1}{n \beta} \log Z^*(\beta) \right\|_2 \mathbb{I}(\mathcal{E}^*)^c \right].$$  \hfill (3.20)

We show that the last term in the RHS of (3.20) is $o(1)$. By Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[ \left\| \frac{H^*_n}{n} - \frac{1}{n \beta} \log Z^*(\beta) \right\|_2 \mathbb{I}(\mathcal{E}^*)^c \right] \leq \sqrt{\mathbb{P}(\mathcal{E}^*)^c} \frac{1}{n^2} \mathbb{E}\left\{ \sup_{\overline{\sigma}} H^*(\overline{\sigma}) - \inf_{\overline{\sigma}} H^*(\overline{\sigma})^2 \right\} \leq k \sqrt{\mathbb{P}(\mathcal{E}^*)^c} \mathbb{E}\left\{ \lambda_{\max}(A^* - \lambda_{\min}(A^*)^2 \right\}$$  \hfill (3.21)

where, for the dense case we have $A^\text{den} = W$, and for the sparse case $A^\text{spar} = (A_G - (\gamma/n)11^T)/\sqrt{\gamma}$, with $A_G$ the adjacency matrix of $G$. In the first inequality we used the elementary bounds $\beta \inf_{\overline{\sigma}} H^*(\overline{\sigma}) \leq \log Z^*(\beta) \leq \beta \sup_{\overline{\sigma}} H^*(\overline{\sigma})$. The second inequality is the obvious spectral relaxation bound.

We next show that the last expression in Eq. (3.21) is $o(1)$ as $n \to \infty$, treating separately the dense and sparse case (as a consequence of our claim $\mathbb{P}(\mathcal{E}^*)^c = 1 - o(1/n)$ which will be proven below). For the dense case, $\mathbb{P}(\lambda_{\max}(W) - \lambda_{\min}(W) \geq 4 + t) \leq c_1 e^{vct^2}$ for some constant $c_1, c_2 > 0$ [AGZ09], whence $\mathbb{E}\{\lambda_{\max}(A^\text{den}) - \lambda_{\min}(A^\text{den})\}^2 \leq 20$, thus yielding the claim.

To handle the sparse case, note that by Efron-Stein inequality [BLM13], $\var{\lambda_{\max}(A^\text{spar}) - \lambda_{\min}(A^\text{spar})} = O(n)$ while $\var{\lambda_{\max}(A^\text{spar}) - \lambda_{\min}(A^\text{spar})} \leq C \sqrt{\log n} / (\log \log n)$ with high probability [KS03]. So the RHS is $o(1)$ under our claim $\mathbb{P}(\mathcal{E}^\text{spar}) \geq 1 - o(1/n)$.

Finally, $V_k(\varepsilon) = \pi^{k/2} / \Gamma(k/2) \mathbb{P}[X < \varepsilon^2 - (\varepsilon^4/4)]$ where $X \sim \text{Beta}(k-1/2, 1/2)$. Hence

$$\mathbb{P}(X < \varepsilon^2 - \varepsilon^4/4) \geq \frac{1}{\text{Beta}(k-1/2, 1/2)} \int_0^{\varepsilon^2 - \varepsilon^4/4} \frac{t^{k-1/2}}{\varepsilon^2 - \varepsilon^4/4} \frac{k-1/2}{2} \, dt \geq \frac{c}{\sqrt{k}} (\varepsilon^2 - \varepsilon^4/4)^{k-1/2}.$$  \hfill (3.22)

Plugging this back into (3.20) gives the required upper bounds.

It remains to prove $\mathbb{P}(\mathcal{E}^*) = 1 - o(1/n)$ for $* \in \{\text{den}, \text{spar}\}$. Using the triangle inequality, we get

$$|H^*(\overline{\sigma}) - H^*_n| \leq \|\overline{\sigma} - \overline{\sigma}^*, A^*\overline{\sigma}\| + \|\overline{\sigma} - \overline{\sigma}^*, A^*\overline{\sigma}^*\| \leq \|\overline{\sigma} - \overline{\sigma}^*\|_F \max\{\|A^*\overline{\sigma}\|_F, \|A^*\overline{\sigma}^*\|_F\}$$  \hfill (3.23)

where we recall that, for the dense case we have $A^\text{den} = W$, and for the sparse case $A^\text{spar} = (A_G - (\gamma/n)11^T)/\sqrt{\gamma}$, with $A_G$ the adjacency matrix of the Erdős-Rényi graph $G \sim G(n, \gamma/n)$.

For any $\overline{\sigma} \in (S^{k-1})^n$ we have

$$\|A^*\overline{\sigma}\|_2^2 \leq k \sup_{s \in \mathbb{R}^n : \|s\|_\infty \leq 1} \|A^*s\|_2^2,$$  \hfill (3.25)
and therefore
\[\mathbb{P}((\mathcal{E}^* c) \leq s_{\infty} < 1) \leq \sup_{s \in \mathbb{R}^n : \|s\|_2 \geq 1} \|A^* s\|_2 \geq C^* \sqrt{n}. \quad (3.26)\]
We conclude the proof by bounding the right-hand side separately in the dense and sparse cases.

The dense case is straightforward. Indeed, for \( \|s\|_\infty \leq 1 \)
\[\|A_{\text{den}} s\|_2^2 \leq \lambda_{\max}(A_{\text{den}}) s \|s\|_2^2 \leq \lambda_{\max}(A_{\text{den}}) n. \quad (3.27)\]
The claim follows since \( \lambda_{\max}(A_{\text{den}}) \leq 2 + t \) with probability larger than \( 1 - c_1 e^{-c_2 n t^2} \) for some \( c_1, c_2 > 0 \). [AGZ09]

For the sparse case, we observe that \( s \mapsto \|A_{\text{spar}} s\|_2^2 \) is a convex function on \( \|s\|_\infty \leq 1 \), and thus attains it maxima at one of the corners of the hypercube \([-1, 1]^n\). In other words, \( \sup_{\|s\|_\infty \leq 1} \|A_{\text{spar}} s\|_2^2 = \max_{s \in \{\pm 1\}^n} \|A_{\text{spar}} s\|_2^2 \). For \( s \in \{\pm 1\}^n \), we get
\[\|A_{\text{spar}} s\|_2^2 \leq \frac{2}{\gamma} \left( \sum_{i=1}^n d_i^2 + n \gamma^2 \right), \quad (3.28)\]
where \( d_i \) is the degree of vertex \( i \) in \( G \). The desired bound follows since \( \sum_{i=1}^n d_i^2 \leq C_0' \gamma^2 n \) with probability at least \( 1 - o(1/n) \) for some constant \( C_0' \) large enough.

3.1. Proof of Lemma 2.1. Using Lemma 3.1 and Lemma 3.2, we can finally prove Lemma 2.1.

We compare the ”ground state energies” for the sparse and the dense models. By triangle inequality and lemmas 3.1 and 3.2, we have
\[|e_n^{\text{spar}} - e_n^{\text{den}}| \leq \left| e_n^{\text{den}} - \frac{1}{\beta} \phi^{\text{den}}(\beta) \right| + \frac{1}{\beta} \phi^{\text{den}}(\beta) - \phi^{\text{spar}}(\beta) + \left| e_n^{\text{spar}} - \frac{1}{\beta} \phi^{\text{spar}}(\beta) \right| \quad (3.29)\]
\[\leq C \left\{ \varepsilon \sqrt{k} \gamma + \frac{k}{\beta} \log \left( k / \varepsilon \right) + \frac{\beta^2}{\gamma^{1/2}} \right\} + o(1) \quad (3.30)\]
\[\leq C \frac{k^{2/3}}{\gamma^{1/6}} \log(\sqrt{k}) + o(1). \quad (3.31)\]
Here, in the last inequality, we substituted \( \varepsilon = k^{1/6} / \gamma^{2/3}, \beta = k^{1/3} \gamma^{1/6} \) (which satisfies the assumptions of Lemma 3.1 for \( \gamma \geq k \)).

Recall that \( \mathbb{E} M_{k}^G (G) / n = e_n^{\text{spar}} \sqrt{\gamma} \) and \( \mathbb{E} k(W) / n = e_n^{\text{den}} \). Hence, to complete the proof, it is sufficient to show that \( M_{k}^G (G) - \mathbb{E} M_{k}^G (G) = o(n) \) with high probability.

In order to prove this, let \( Z(A) = M_{k}^G (G) \) (with \( A = (a_{ij}) \) the adjacency matrix of \( G \)). Let \( Z^{(i,j)}(A) \) be the random variable obtained by replacing \( a_{ij} \) by an independent copy \( a_{ij}' \). It is easy to see that \( |Z(A) - Z^{(i,j)}(A)| \leq |a_{ij} - a_{ij}'| \). An application of Efron-Stein inequality [BLM13] yields that for any \( \varepsilon > 0 \),
\[\mathbb{P} \left[ |M_{k}^G (G, \lambda) - \mathbb{E}[M_{k}^G (G, \lambda)]| \geq n \varepsilon \right] = O(1/n), \quad (3.32)\]
Combining this with (3.31) completes the proof.
APPENDIX A. PROOF OF THEOREM 1.5

As mentioned already, the upper bound in Eq. (1.10) is trivial. The proof of the lower bound follows Rietz’s method [Rie74].

Let \( X \) be a solution of the problem (1.8) and through its Cholesky decomposition write \( X_{ij} = \langle u_i, u_j \rangle \), with \( u_i \in \mathbb{R}^n \), \( \| u_i \|_2 = 1 \). In other words we have, letting \( B = (B_{ij})_{i,j \in [n]} \),
\[
Q(B) = \sum_{i,j=1}^n B_{ij} \langle u_i, u_j \rangle .
\] (A.1)

Let \( J \in \mathbb{R}^{k \times n} \) be a matrix with i.i.d. entries \( J_{ij} \sim N(0, 1/k) \). Define, \( x_i \in \mathbb{R}^k \), for \( i \in [n] \), by letting
\[
x_i = \frac{J_{ui}}{\| J_{ui} \|_2}.
\] (A.2)

We next need a technical lemma.

**Lemma A.1.** Let \( u, v \in \mathbb{R}^n \) with \( \| u \|_2 = \| v \|_2 = 1 \) and \( J \in \mathbb{R}^{k \times n} \) be defined as above. Further, for \( w \in \mathbb{R}^n \), let \( z(w) \equiv (1 - \alpha_k^{-1/2} \| Jw \|_2^{-1}) Jw \). Then
\[
\mathbb{E} \langle \frac{J_{uw}}{\| J_{uw} \|_2}, \frac{J_{vw}}{\| J_{vw} \|_2} \rangle = \alpha_k \langle u, v \rangle + \alpha_k \mathbb{E} \langle z(u), z(v) \rangle .
\] (A.3)

**Proof.** Let \( g_1, g_2 \sim N(0, I_{k \times k}/k) \) be independent vectors (distributed as the first two columns of \( J \). Let \( a = \langle u, v \rangle \) and \( b = \sqrt{1 - a^2} \). Then by rotation invariance
\[
\mathbb{E} \langle Ju, Jv \rangle = \mathbb{E} \langle g_1, ag_1 + g_2 \rangle = a \mathbb{E} \langle \| g_1 \|_2^2 \rangle = \langle u, v \rangle ,
\] (A.4)

and
\[
\mathbb{E} \langle \frac{J_{uw}}{\| J_{uw} \|_2}, \frac{J_{vw}}{\| J_{vw} \|_2} \rangle = \mathbb{E} \langle \frac{g_1}{\| g_1 \|_2}, ag_1 + g_2 \rangle
\]
\[
= a \mathbb{E} \langle \| g_1 \|_2 \rangle = \alpha_k^{1/2} \langle u, v \rangle .
\] (A.5)

By expanding the product we have
\[
\mathbb{E} \langle z(u), z(v) \rangle = \langle u, v \rangle - \alpha_k^{-1/2} \mathbb{E} \langle \frac{J_{uw}}{\| J_{uw} \|_2}, Jv \rangle - \alpha_k^{-1/2} \mathbb{E} \langle Jw, \frac{J_{vw}}{\| J_{vw} \|_2} \rangle + \frac{1}{\alpha_k} \mathbb{E} \langle \frac{J_{uw}}{\| J_{uw} \|_2}, \frac{J_{vw}}{\| J_{vw} \|_2} \rangle
\] (A.7)
\[
= -\langle u, v \rangle + \frac{1}{\alpha_k} \mathbb{E} \langle \frac{J_{uw}}{\| J_{uw} \|_2}, \frac{J_{vw}}{\| J_{vw} \|_2} \rangle
\] (A.8)

which is equivalent to the statement of our lemma. \( \square \)

Now, by definition of the \( x_i \)'s we have
\[
\mathbb{E} \left\{ \sum_{i,j=1}^n B_{ij} \langle x_i, x_j \rangle \right\} = \sum_{i,j=1}^n B_{ij} \mathbb{E} \langle \frac{J_{ui}}{\| J_{ui} \|_2}, \frac{J_{uj}}{\| J_{uj} \|_2} \rangle
\]
\[
= \alpha_k \sum_{i,j=1}^n B_{ij} \langle u_i, u_j \rangle + \alpha_k \sum_{i,j=1}^n B_{ij} \mathbb{E} \langle z(u_i), z(u_j) \rangle
\] (A.9)
\[
= \alpha_k Q(B) + \alpha_k \sum_{i,j=1}^n B_{ij} \mathbb{E} \langle z(u_i), z(u_j) \rangle .
\] (A.10)
Now we interpret \( z(u_i) \) as a vector in a Hilbert space with scalar product \( \mathbb{E}(\cdot, \cdot) \). Further by the rounding lemma A.1, these vectors have norm
\[
\mathbb{E}(\| z(u_i) \|_2^2) = \frac{1}{\alpha_k} - 1. \tag{A.12}
\]
Hence, by definition of \( Q(\cdot) \), we have
\[
- \sum_{i,j=1}^n B_{ij} \mathbb{E}(z(u_i) z(u_j)) \leq \left( \frac{1}{\alpha_k} - 1 \right) Q(-B). \tag{A.13}
\]
Substituting this in Eq. (A.11), we obtain
\[
Q_k(B) \geq \mathbb{E}\left\{ \sum_{i,j=1}^n B_{ij} \langle x_i, x_j \rangle \right\} \geq \alpha_k Q(B) - (1 - \alpha_k)Q(-B), \tag{A.14}
\]
which coincides with the claim (1.10).

In order to prove Eq. (1.11), we apply Eq. (1.10) to \(-B\), thus getting
\[
Q(-B) \leq \frac{1}{\alpha_k} Q_k(-B) + \frac{1 - \alpha_k}{\alpha_k} Q(B). \tag{A.15}
\]
Substituting this in Eq. (1.10), we obtain Eq. (1.11).

A.1. The interpolation derivatives. In this subsection we carry out the calculation of the derivative in Eqs. (3.9), (3.10), (3.11). We note that
\[
\phi(\beta; t) = \int \frac{1}{n} \log \left[ \int \exp(\beta H_t(\sigma)) d\nu(\sigma) \right] p(t, A(t)) d\rho_0(A(t), \mathbb{W}) \tag{A.16}
\]
where \( p(t, A(t)) = \prod_{i \leq j} p(t, a_{ij}(t)), \) and \( p(t, a_{ij}(t)) \) is the PMF of Pois\( (\gamma(1-t)) \) if \( i = j \) and is the PMF of Pois\( (\gamma(1-t)/n) \) if \( i < j \). Further \( \rho_0 = \rho_{\mathbb{W}}^n(\alpha(n+1)/2) \otimes \rho_\mathbb{R}^n(\alpha(n+1)/2) \), where \( \rho_\mathbb{W} \) is the counting measure on \( \mathbb{W} \) and \( \rho_\mathbb{R} \) is the standard Gaussian measure on \( \mathbb{R} \). Then we have,
\[
\frac{\partial \phi(\beta; t)}{\partial t} = M_1 + M_2, \tag{A.17}
\]
\[
M_1 = \frac{\beta}{n} \mathbb{E}\left[ \mu_{\beta,t} \left( \frac{\partial H_t(\sigma)}{\partial t} \right) \right], \tag{A.18}
\]
\[
M_2 = \int \frac{1}{n} \log \left[ \int \exp(\beta H_t(\sigma)) d\nu \right] \frac{\partial p(t, A(t))}{\partial t} d\rho_0. \tag{A.19}
\]
Consider the term \( M_1 \):
\[
M_1 = \beta \sqrt{\gamma} \frac{1}{2n} \sum_{i,j} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)] + \frac{\beta}{2n\sqrt{t}} \mathbb{E}[\mu_{\beta,t}(H_{\text{den}}(\sigma))]. \tag{A.20}
\]
Further
\[
\mathbb{E}[\mu_{\beta,t}(H_{\text{den}}(\sigma))] = \sum_{i,j} \int \frac{\langle \sigma_i, \sigma_j \rangle}{\sqrt{n}} \mathbb{E}\left[ \frac{W_{ij} \exp(\beta H_t(\sigma))}{\int \exp(\beta H_t(\tau)) d\nu(\tau)} \right] d\nu(\sigma). \tag{A.21}
\]
An application of Gaussian integration by parts yields
\[
\mathbb{E}[\mu_{\beta,t}(H_{\text{den}}(\sigma))] = 2\beta \sqrt{\gamma} \frac{1}{n} \sum_{i,j} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle^2)] - \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2$. \tag{A.22}
\]
Plugging this back into (A.20) gives \( I \), as in Eq. (3.10).
Next, let \( h_{ij}(a_{ij}(t)) = \mathbb{E}[n^{-1}\log[\int \exp(\beta H_t(\sigma))d\sigma]|a_{ij}(t)] \). The product form of \( p(t, A(t)) \) implies

\[
M_2 = \sum_{i \leq j} \int h_{ij}(a_{ij}(t)) \frac{\partial p(t, a_{ij}(t))}{\partial t} d\rho_n(a_{ij}(t)). \tag{A.23}
\]

The \( ij \)-th integral in the RHS of (A.23) is simply \(-\frac{\gamma}{2n} g'(\lambda)\), where \( g(\lambda) = \mathbb{E}[f(X)] \) for \( f = h_{ij} \) and \( X \sim \text{Pois}(\lambda) \) at \( \lambda = \frac{n(1-t)}{n} \). Differentiating the PMF of the Poisson distribution, one observes that \( g'(\lambda) = \mathbb{E}[f(X + 1) - f(X)] \). With this observation, we have,

\[
M_2 = -\frac{\gamma}{2n} \sum_{i \leq j} \mathbb{E}[h_{ij}(a_{ij}(t) + 1) - h_{ij}(a_{ij}(t))]. \tag{A.24}
\]

Adding 1 to \( a_{ij}(t) \) corresponds to adding an \((i, j)\) edge in the sparse hamiltonian \( H^{\text{spar}}(\cdot) \). Thus we have,

\[
\mathbb{E}[h_{ij}(a_{ij}(t) + 1) - h_{ij}(a_{ij}(t))] = \mathbb{E}[\log \mu_\beta(t) \exp(\frac{\psi_{ij}^2}{\sqrt{\gamma}}(\sigma_i, \sigma_j))]. \tag{A.25}
\]

where \( \psi_{ij} = 1 \) if \( i = j \) and 2 otherwise. Plugging this back to (A.24) establishes \( \text{II}, \) as in Eq. (3.10).

**Appendix B. The complete graph:** Proof of Lemma 2.2

Throughout this section, \( W \in \mathbb{R}^{n \times n} \) is a symmetric random matrix with \((W_{ij})_{i \leq j}\) independent, \( W_{ij} \sim N(0, 1/n) \) if \( i < j \), \( i, j \in [n] \), and \( W_{ii} \sim N(0, 2/n) \) for \( i \in [n] \).

By Theorem 1.5, applied to \( B = W \), we have \( Q(W) \geq Q_k(W) \geq \alpha_k Q(W) - (1 - \alpha_k) Q(-W) \), and hence using the symmetry of \( W \),

\[
(2\alpha_k - 1) \lim \inf_{n \to \infty} \frac{1}{n} \mathbb{E}Q(W) \leq q(k) \leq q_u(k) \leq \lim \sup_{n \to \infty} \frac{1}{n} \mathbb{E}Q(W). \tag{B.1}
\]

The proof of Lemma 2.2 therefore follows from \( \alpha_k = 1 - O(1/k) \), once we prove the following.

**Lemma B.1.** We have

\[
\lim_{n \to \infty} \frac{1}{n} \mathbb{E}Q(W) = 2. \tag{B.2}
\]

**Proof.** First notice that \( Q(W) \leq n \lambda_{\text{max}}(W) \) (the maximum eigenvalue of \( W \)). Hence

\[
\lim \inf_{n \to \infty} \frac{1}{n} \mathbb{E}Q(W) \leq \lim_{n \to \infty} \mathbb{E} \lambda_{\text{max}}(W) = 2. \tag{B.3}
\]

For the last (classical) equality, see, for instance, [AGZ09].

We are left to prove a lower bound on \( Q(W) \). Fix \( \varepsilon > 0 \), and let \( \varphi_1, \varphi_2, \ldots, \varphi_{n\varepsilon} \) be the eigenvectors of \( W \) corresponding to the top \( n\varepsilon \) eigenvalues. Denote by \( U \in \mathbb{R}^{n \times (n\varepsilon)} \) the matrix whose columns are \( \varphi_1, \varphi_2, \ldots, \varphi_{n\varepsilon} \), and let \( D \in \mathbb{R}^{n \times n} \) be the diagonal matrix with entries

\[
D_{ii} = (UU^T)_{ii}. \tag{B.4}
\]

Note that –by the invariance of \( W \) under rotations– \( P = UU^T \) is a projector onto a uniformly random subspace of \( n\varepsilon \) dimension in \( \mathbb{R}^n \), and \( D_{ii} = \langle \varepsilon_i, Pe_i \rangle = \|Pe_i\|_2^2 \). Inverting the role of \( P \) and \( \varepsilon_i \), we see that \( D_{ii} \) is distributed as the square norm of the first \( n\varepsilon \) components of a uniformly random unit vector of \( n \) dimensions. Hence

\[
D_{ii} = \frac{Z_{n\varepsilon}}{Z_{n\varepsilon} + Z_{n(1-\varepsilon)}}, \tag{B.5}
\]
where $Z_\ell \sim \chi^2(\ell)$, $\ell \in \{n\varepsilon, n(1 - \varepsilon)\}$ denote two independent chi-squared random variable with $\ell$ degrees of freedom. Standard tail bounds on chi-squared random variables, plus union bound over $i \in [n]$, imply

$$\mathbb{P}\left(\max_{i \in [n]} |D_{ii} - \varepsilon| \leq C \sqrt{\frac{\log n}{n}}\right) \geq 1 - \frac{1}{n^{10}}, \quad (B.6)$$

for $C$ a suitable constant.

We then define

$$X = D^{-1/2}UU^T D^{-1/2}.$$  

(B.7)

This is clearly a feasible point of the optimization problem that defines $Q(W)$, cf. Eq. (1.8), i.e $X \succeq 0$ and $X_{ii} = 1$. Therefore, letting $E = \varepsilon^{1/2}D^{-1/2}$

$$Q(W) \geq \langle W, X \rangle \quad (B.8)$$

$$= \frac{1}{\varepsilon}(W, UU^T) + \frac{1}{\varepsilon}(W - EWE, UU^T) \quad (B.9)$$

$$\geq \frac{1}{\varepsilon} \sum_{\ell=1}^{n\varepsilon} \lambda_\ell(W) - \frac{1}{\varepsilon} \|W - EWE\|_2 \|UU^T\|_* \quad (B.10)$$

$$\geq n\lambda_{n\varepsilon}(W) - \frac{1}{\varepsilon} \|W\|_2 (1 + \|E\|_2) \|E - I\|_2 \|UU^T\|_*.$$  

(B.11)

Here $\|Z\|_*$ denotes the nuclear norm of $Z$ (sum of the absolute values of eigenvalues) and in the last inequality we used $\|W - EWE\|_2 \leq \|W - EW\|_2 + \|EW - EWE\|_2 \leq \|W\|_2 \|E - I\|_2 + \|E\|_2 \|W\|_2 \|E - I\|_2$.

Next, since $UU^T$ is a projector on $n\varepsilon$ dimensions, we have $\|UU^T\|_* = n\varepsilon$, whence

$$\frac{1}{n} Q(W) \geq \lambda_{n\varepsilon}(W) - \|W\|_2 (1 + \|E\|_2) \|E - I\|_2.$$  

(B.12)

By Eq. (B.6), we have $\|E - I\|_2 \to 0$ almost surely, and by a classical result [AGZ09], also the following limits hold almost surely

$$\lim_{n \to \infty} \|W\|_2 = 2, \quad (B.13)$$

$$\lim_{n \to \infty} \lambda_{n\varepsilon}(W) = 2 - \delta(\varepsilon), \quad (B.14)$$

where $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \to 0$. Indeed $\delta(\varepsilon)$ can be expressed explicitly in terms of Wigner semicircle law, namely, for $\varepsilon \in (0,1)$ it is the unique positive solution of the following equation.

$$\int_{2-\delta}^{2} \frac{\sqrt{4 - x^2}}{2\pi} \, dx = \varepsilon.$$  

(B.15)

Substituting in Eq. (B.12), we get, almost surely,

$$\lim \inf_{n \to \infty} \frac{1}{n} Q(W) \geq 2 - \delta(\varepsilon).$$  

(B.16)

Note that $\lambda_{\min}(W) \leq Q(W) \leq \lambda_{\max}(W)$ and $\mathbb{E}|\lambda_{\min}(W)|, \mathbb{E}|\lambda_{\max}(W)| < \infty$. Hence by dominated convergence

$$\lim \inf_{n \to \infty} \frac{1}{n} \mathbb{E}Q(W) \geq 2 - \delta(\varepsilon), \quad (B.17)$$

And this implies $\lim_{n \to \infty} \mathbb{E}Q(W)/n = 2$ since $\varepsilon$ can be taken arbitrarily small. \hfill \square
APPENDIX C. Community detection: A Stronger Version of Theorem 1.7

Throughout this section we will assume \( n \) even, to avoid unnecessary technicalities. We will also fix, without loss of generality \( S = \{1, 2, \ldots, n/2\} \). Further \( v \) is the vector with \( v_i = 1/\sqrt{n} \) for \( i \in \{1, \ldots, n/2\} \) and \( v_i = -1/\sqrt{n} \) for \( i \in \{(n/2) + 1, \ldots, n\} \). Finally, throughout this section we let \( \gamma \equiv (a + b)/2 \).

Notice that Theorem 1.7 follows immediately if we can prove that, assuming Eq. (1.5) is the following.

This yields the desired bound (C.1) with \( \delta_*(\varepsilon) = \varepsilon/4 \).

In the rest of this section, we will establish a characterization of the optimal constant \( \xi_* \) that can be substituted to 2 on the right-hand side of Eq. (1.18) in Theorem 1.7. This characterization is not explicit: it yields \( \xi_* \) in terms of the asymptotic value of a sequence of SDP’s with random Gaussian data. However, we believe it provides an insightful equivalence and can open the way to determining \( \xi_* \).

C.1. Interpolation lemmas. Let us begin by defining a problem with Gaussian data. For \( \xi \in \mathbb{R} \), we let \( B(\xi) \in \mathbb{R}^{n \times n} \) be a symmetric matrix given by

\[
B(\xi) \equiv \xi v v^T + W.
\]  

Here \( W \) is a symmetric random matrix with \( (W_{ij})_{i \leq j} \) independent and \( W_{ij} \sim \mathcal{N}(0, 1/n) \) if \( i < j \), \( W_{ii} \sim \mathcal{N}(0, 2/n) \).

We then introduce a generalization and strengthening of the definition in Eq. (2.4).

Lemma C.1. With the above definitions, the following limit exists, for all fixed \( k \in \mathbb{N} \), \( \xi \in \mathbb{R} \)

\[
q(\xi; k) \equiv \lim \limits_{n \to \infty} \frac{1}{n} \mathbb{E} Q_k(B(\xi)).
\]  

Its proof can be found in Appendix D and uses results from [GT02]. Note that, applying this for \( \xi = 0 \), we obtain that indeed \( q_0(k) = q_0(k) = q(0; k) \). Another straightforward consequence of this Lemma and Theorem 1.5 is the following.

Corollary C.2. The following limits exist and coincide

\[
q(\xi) \equiv \lim \limits_{k \to \infty} q(\xi; k) = \lim \limits_{n \to \infty} \frac{1}{n} \mathbb{E} Q(B(\xi)).
\]
Finally we have the following interpolation lemma that generalizes Lemma 2.1 to the stochastic block model.

**Lemma C.3.** Let \( G \) be an random graph distributed according to Hypothesis 1, i.e. according to the two-groups stochastic block model with \( n \) vertices, and parameters \( a, b \). Define \( \xi \equiv \frac{a-b}{\sqrt{2(a+b)}} \) and \( \gamma \equiv (a+b)/2 \).

Letting \( q(\xi; k) \) be defined as above, there exists a constant \( C_0 = C_0(\xi) \) such that, with high probability
\[
\frac{1}{n} M_k^+(G; \gamma/n) \in \left[ q(\xi; k) \sqrt{\gamma} - C_0 k^{2/3} \gamma^{1/3} \log(k\gamma), q(\xi; k) \sqrt{\gamma} + C_0 k^{2/3} \gamma^{1/3} \log(k\gamma) \right].
\] (C.7)

**C.2. A stronger form of Theorem 1.7.** Lemma C.3, together with Theorem 1.5 implies that, for large \( M^+(G; \gamma/n)/n = q(\xi) \sqrt{\gamma} + o(\sqrt{\gamma}) \). The following properties of the function \( q \) are proven in Appendix D.2.

**Lemma C.4.** The function \( \xi \mapsto q(\xi) \) is non-decreasing in \( \xi \in \mathbb{R}_+ \). Further, we have the bounds
\[
0 \leq \xi \leq 1 \implies q(\xi) = 2,
\] (C.8)
\[
\xi \geq 1 \implies \max(2, \xi) \leq q(\xi) \leq \frac{1}{\xi}.
\] (C.9)

We therefore define \( \xi_* \in [1, 2] \) by
\[
\xi_* = \inf \{ \xi \geq 0 : q(\xi) > 0 \}.
\] (C.10)

Indeed, as mentioned in Remark 1.9, it is possible to prove by a perturbative argument that \( \xi_* < 2 \) strictly, cf. Appendix E.

**Theorem C.5.** Let \( G \) be an random graph distributed according to Hypothesis 1, i.e. according to the two-groups stochastic block model with \( n \) vertices, and parameters \( a, b \). Define \( \xi \equiv \frac{a-b}{\sqrt{2(a+b)}} \) and \( \gamma \equiv (a+b)/2 \).

Then, with high probability, for any \( \xi \) fixed,
\[
\frac{1}{n} M(G; \gamma/n) \in \left[ q(\xi) \sqrt{\gamma} - o_\gamma(\sqrt{\gamma}), q(\xi) \sqrt{\gamma} + o_\gamma(\sqrt{\gamma}) \right].
\] (C.11)

In particular, if for some \( \varepsilon > 0 \),
\[
a - b > \frac{2\gamma}{\sqrt{2(a+b)}} \geq \xi_* + \varepsilon,
\] (C.12)
then there exists \( \delta_* = \delta_*(\varepsilon) > 0 \) and \( \gamma_* = \gamma_*(\varepsilon) > 0 \) such that the following holds. If \( (a+b)/2 \geq \gamma_* \), then the SDP-based test \( T(\cdot; \delta_*) \) succeeds with high probability.

**Proof.** The fact that \( T(\cdot; \delta_*) \) succeeds under condition (C.12) follows immediately from Theorem 1.1 and Eq. (C.11). We can therefore focus on proving the latter.

Now, by Theorem 1.5 and Lemma C.3, we have, letting \( \psi(\xi, k; \gamma) = C_0 k^{2/3} \gamma^{1/3} \log(k\gamma) \),
\[
\lim_{n \to \infty} \mathbb{P} \left\{ q(\xi; k) \sqrt{\gamma} - \psi(\xi, k; \gamma) \leq \frac{1}{n} M^+(G; \gamma/n) \leq \frac{1}{2\alpha_k - 1} \left[ q(\xi; k) \sqrt{\gamma} + \psi(\xi, k; \gamma) \right] \right\} = 1.
\] (C.13)

The conclusion follows by taking \( k = k(\gamma) = \gamma^{1/10} \) and using Corollary C.2 alongside Eq. (1.12). \( \square \)
C.3. **Proof of Lemma C.3.** The interpolation argument in this case proceeds in two steps — we first compare the sparse problem to a problem on the complete graph. Next, we compare two distinct problems on the complete graph. Without loss of generality, we assume that \( S = \{1, \cdots, n/2\} \).

Let \( H_{\text{spar}}(\sigma) = \frac{1}{n} \sum_{i,j=1}^{n} (A_{ij} - \frac{\gamma}{n}) \langle \sigma_i, \sigma_j \rangle \), where \( A = (A_{ij}) \) is the adjacency matrix corresponding to the random graph distributed according to Hypothesis 1 with parameters \( a \) and \( b \). Next, let \( H_{\text{den}}(\sigma) = \sum_{i,j=1}^{n} D_{ij} \langle \sigma_i, \sigma_j \rangle \), where \( D = \xi v v^T + U \), where \( U \) is a symmetric Gaussian matrix with mean 0, \((U_{ij})_{i,j \leq j}\) are independent normal random variables, and, for \( i \leq j \)

\[
\text{Var}(U_{ij}) = \begin{cases} \frac{a}{(n\gamma)} & \text{if } \{i,j\} \subseteq S \text{ or } \{i,j\} \subseteq [n]\backslash S, \\ \frac{b}{(n\gamma)} & \text{if } i \in S, j \in [n]\backslash S \text{ or } i \in [n]\backslash S, j \in S. \end{cases} \tag{C.14}
\]

The free energy densities corresponding to these Hamiltonians is defined using (3.3) and denoted by \( \phi_{\text{spar}}(\beta) \) and \( \phi_{\text{den}}(\beta) \) respectively.

The proof of Theorem C.3 will use two lemmas. Our first lemma will bound the difference between \( \phi_{\text{spar}}(\beta) \) and \( \phi_{\text{den}}(\beta) \).

**Lemma C.6.** There exists \( C = C(\xi) \) independent of \( n, \gamma \) and \( k \) such that, if \( \beta \leq \sqrt{\gamma}/C \), then

\[
|\phi_{\text{spar}}(\beta) - \phi_{\text{den}}(\beta)| \leq C \frac{\beta}{\sqrt{\gamma}} + o(1) \tag{C.15}
\]

**Proof.** The proof will proceed along lines similar to the proof of Lemma 2.1. We start with a “Poissonized” version of the model in Hypothesis 1, with parameters \( a \) and \( b \). The graph is formed by adding \( \text{Pois}(n\gamma/2) \) edges to \([n] \), with the end-points being chosen as follows. The first end-point of the edge is selected uniformly at random in \( \{1, \cdots, n\} \). Given the first end-point \( i, j \in [n] \) is chosen with probability \( \frac{a}{n\gamma} \) if \( \{i,j\} \subseteq S \) or \( \{i,j\} \subseteq [n]\backslash S \); otherwise \( j \) is chosen with probability \( \frac{b}{n\gamma} \). It is easy to see that it suffices to establish the result for this “Poissonized” graph. Indeed, this can be coupled to the original graph in such a way that they differ -with high probability- in \( O(1) \) edges.

Let \( A(a,b) = (A_{ij}) \) denote the adjacency matrix of the modified graph. For simplicity of notation, the arguments will be suppressed when the underlying parameters are clear from the context. We observe that \( A(a,b) \) is a symmetric matrix having \( (A_{ij})_{i,j \leq j} \) independent with \( A_{ij} \sim \text{Pois}(a/n) \) if \( i \neq j \) and \( \{i,j\} \subseteq S \) or \( \{i,j\} \subseteq [n]\backslash S \), \( A_{ii} \sim \text{Pois}(a/2n) \) and \( A_{ij} \sim \text{Pois}(b/n) \) otherwise.

Next we define the interpolating Hamiltonian

\[
H_t(\sigma) = \frac{1}{\sqrt{\gamma}} \sum_{i,j=1}^{n} \left( A_{ij}(t) - \frac{(1-t)\gamma}{n} \right) \langle \sigma_i, \sigma_j \rangle + \sum_{i,j=1}^{n} D_{ij}(t) \langle \sigma_i, \sigma_j \rangle, \tag{C.16}
\]

where \( A(t) = A((1-t)a, (1-t)b) \) and \( D(t) = t \xi v v^T + \sqrt{t} U \). Let

\[
\phi(\beta; t) = \frac{1}{n} \mathbb{E} \log \left\{ \int \exp(\beta H_t(\sigma)) d\nu(\sigma) \right\} \tag{C.17}
\]

denote the interpolating free energy. Then we see that \( \phi(\beta; 0) = \phi_{\text{spar}}(\beta) \) while \( \phi(\beta, 1) = \phi_{\text{den}}(\beta) \). Similar to (3.7), it suffices to uniformly bound the derivative of the interpolating free energy. To this end, we introduce the Gibbs measure \( \mu_{\beta,t}(\cdot) \) corresponding to the interpolating Hamiltonian \( H_t(\cdot) \). Further, let \( S_1 \) denote the set of \((i,j)\) pairs such that either \( \{i,j\} \subseteq S \) or \( \{i,j\} \subseteq [n] \backslash S \).
Setting $vv^T = (\delta_{ij})$, by calculations similar to Section A.1, we have, \[ \frac{\partial \phi_t}{\partial t} = I + II, \] where,

\[
I = \frac{\beta \sqrt{\gamma}}{n^2} \sum_{i,j=1}^n \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)] + \frac{\beta}{n^2} \delta_{ij} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)] + \frac{\beta}{2n\sqrt{\gamma}} \mathbb{E}[\mu_{\beta,t}(\langle U \sigma, \sigma \rangle)], \tag{C.18}
\]

\[
II = -\frac{a}{2n^2} \sum_{S_i} \mathbb{E} \left[ \log \mu_{\beta,t} \left( \exp \left( \frac{2\beta}{\sqrt{\gamma}} \langle \sigma_i, \sigma_j \rangle \right) \right) \right] - \frac{b}{2n^2} \sum_{S_i} \mathbb{E} \left[ \log \mu_{\beta,t} \left( \exp \left( \frac{2\beta}{\sqrt{\gamma}} \langle \sigma_i, \sigma_j \rangle \right) \right) \right] + o(1). \tag{C.19}
\]

An application of Gaussian integration by parts as in (A.22) yields,

\[
\mathbb{E}[\mu_{\beta,t}(\langle U \sigma, \sigma \rangle)] = 2\beta \sqrt{\gamma} \left\{ \frac{a}{n\gamma} \sum_{S_i} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2] - \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2] \right\} + 2\beta \sqrt{\gamma} \left\{ \frac{b}{n\gamma} \sum_{S_i} \mathbb{E}[\mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2] - \mu_{\beta,t}(\langle \sigma_i, \sigma_j \rangle)^2] \right\}. \tag{C.20}
\]

The proof is completed by using Taylor expansion and bounding the derivatives as in the proof of Lemma 3.1.

We next notice that the $\beta \to \infty$ regularity proved in Lemma 3.2 holds for the random graph $G$ and the random matrix $D$ introduced here as well. The proof 3.2 does indeed go through with essentially no change. Proceeding as in the proof of Lemma 2.1, cf. Section 3.1, we obtain the following. (Note that $\max_a H^\text{par}(\sigma) = M_k^+(G; \gamma/n)/\sqrt{\gamma}$)

**Corollary C.7.** Let $G$ be a random graph distributed according to Hypothesis 1 with parameters $a$ and $b$, and $D = \xi v^T v + U$ be a random matrix as defined above, with variance of the entries given by Eq. (C.14). Then, the following holds with high probability, for a constant $C = C(\xi)$,

\[
\left| \frac{1}{n \sqrt{\gamma}} M_k^+(G; \gamma/n) - \frac{1}{n} \mathbb{E} Q_k(D) \right| \leq Ck^{2/3} \gamma^{-1/6} \log(k \gamma). \tag{C.21}
\]

Next, define three independent, symmetric, random matrices $Y_0, Y_1, Y_2$ with centered Gaussian entries. We choose variances as follows for $i \leq j$

\[
\text{Var}((Y_0)_{ij}) = \frac{b}{n\gamma}, \tag{C.22}
\]

\[
\text{Var}((Y_1)_{ij}) = \begin{cases} (a - b)/(n\gamma) & \text{if } \{i,j\} \subseteq S \text{ or } \{i,j\} \subseteq [n] \setminus S, \\ 0 & \text{if } i \in S, j \in [n] \setminus S \text{ or } i \in [n] \setminus S, j \in S. \end{cases} \tag{C.23}
\]

Finally $\text{Var}((Y_2)_{ij}) = (a - b)/(2n\gamma)$ for all $i < j$ and $\text{Var}((Y_2)_{ii}) = a/(n\gamma)$ on the diagonal.

With these definition, we can couple the random matrices $B$ and $D$ defined above by letting

\[
B = \xi v^T v + Y_0 + Y_2, \tag{C.24}
\]

\[
D = \xi v^T v + Y_0 + Y_1. \tag{C.25}
\]
We then have
\[
\left| \frac{1}{n} \mathbb{E} Q_k(B) - \frac{1}{n} \mathbb{E} Q_k(D) \right| \leq \mathbb{E} \left[ \| B - D \|_2 \right] 
\leq \mathbb{E} \| Y_1 \|_2 + \mathbb{E} \| Y_2 \|_2 
\leq 2 \sqrt{\frac{a - b}{\gamma}} + 2 \sqrt{\frac{a - b}{2\gamma}} + \frac{2}{n}, \quad (C.28)
\]
where the last bound holds by standard estimates on the eigenvalues of GOE matrices [AGZ09].

Recalling that, by definition, \( a - b = 2 \xi \sqrt{\gamma} \) we conclude that
\[
\left| \frac{1}{n} \mathbb{E} Q_k(B) - \frac{1}{n} \mathbb{E} Q_k(D) \right| \leq 5 \frac{\xi}{\sqrt{\gamma}} . \quad (C.29)
\]
Substituting in Eq. (C.21) we obtain, for a constant \( C = C(\xi) \), with high probability
\[
\left| \frac{1}{n} \mathbb{E} Q_k(B) - \frac{1}{n} \mathbb{E} Q_k(D) \right| \leq C \frac{\gamma}{\sqrt{n}}. \quad (C.30)
\]
Recalling that \( n^{-1} \mathbb{E} Q_k(B) \to q(\xi; k) \) by Lemma C.1, we obtain the claim of Lemma C.3.

**Appendix D. Auxiliary lemmas for community detection**

In this Appendix we prove Lemma C.1 and Lemma C.4. Note that by a change of variables in the optimization problem defining \( Q_k(B(\xi)) \) we can redefine \( B(\xi) \) as per Eq. (C.4) with \( v = 1/\sqrt{n} \).

To see this, it is sufficient to replace the decision variable \( X \) in Eq. (1.7) by \( X' \) defined as follows
\[
X'_{ij} = \begin{cases} 
X_{ij} & \text{if } i, j \in S \text{ or } i, j \in S^c, \\
-X_{ij} & \text{if } i \in S, j \in S^c \text{ or } i \in S^c, j \in S,
\end{cases} \quad (D.1)
\]
and note that this change leave unchanged the constraints.

Hence, in the proof below we will take
\[
B(\xi) = \frac{\xi}{n} \mathbb{1} \mathbb{1}^T + W. \quad (D.2)
\]

**D.1. Proof of Lemma C.1.** We define the following Hamiltonian \( H : (\mathbb{S}^{k-1})^n \to \mathbb{R} \):
\[
H_\xi(\sigma) = \sum_{i,j=1}^n B(\xi)_{ij} \langle \sigma_i, \sigma_j \rangle 
= \frac{\xi}{n} \left( \sum_{i=1}^n \sigma_i \right)^2 + \sum_{i,j=1}^n W_{ij} \langle \sigma_i, \sigma_j \rangle , \quad (D.3)
\]
and the corresponding free energy density
\[
\phi_n(\beta; \xi) = \frac{1}{n} \mathbb{E} \log \left\{ \int e^{\beta H_\xi(\sigma)} d\nu(\sigma) \right\}. \quad (D.5)
\]
Using [GT02, Theorem 1], we conclude that the limit
\[
\phi(\beta; \xi) = \lim_{n \to \infty} \phi_n(\beta; \xi) \quad (D.6)
\]
exists for every \( \beta > 0 \).

Further note that \( \lim_{\beta \to \infty} \phi_n(\beta; \xi)/\beta = \mathbb{E} Q_k(B(\xi))/n \). By inspection of the proof of Lemma 3.2, we conclude that a bound analogous to (3.15) holds for the present Hamiltonian as well (the only
property used is that the $\|B(\xi)\|_2 \leq 5 + \xi$ with probability $1 - e^{-cn}$). Hence, we get, for a universal constant $C$,

$$\left| \frac{1}{n} \mathbb{E} q_k(B(\xi)) - \frac{1}{\beta} \phi_n(\beta; \xi) \right| \leq (5 + \xi) \varepsilon \sqrt{k} + C \frac{k}{\beta} \log \frac{k}{c \varepsilon} + o(1). \quad (D.7)$$

Define

$$q_l(\xi; k) = \lim_{n \to \infty} \inf \left\{ \frac{1}{n} \mathbb{E} q_k(B(\xi)) : \text{for all } n \right\}, \quad q_u(\xi; k) = \lim_{n \to \infty} \sup \left\{ \frac{1}{n} \mathbb{E} q_k(B(\xi)) : \text{for all } n \right\}. \quad (D.8)$$

Taking the $n \to \infty$ in Eq. $(D.7)$ and using Eq. $(D.6)$, we get

$$\left| q_l(\xi; k) - \frac{1}{\beta} \phi(\beta; \xi) \right| \leq (5 + \xi) \varepsilon \sqrt{k} + C \frac{k}{\beta} \log \frac{k}{c \varepsilon}, \quad (D.9)$$

$$\left| q_u(\xi; k) - \frac{1}{\beta} \phi(\beta; \xi) \right| \leq (5 + \xi) \varepsilon \sqrt{k} + C \frac{k}{\beta} \log \frac{k}{c \varepsilon}. \quad (D.10)$$

and hence by triangular inequality

$$\left| q_l(\xi; k) - q_u(\xi; k) \right| \leq 2(5 + \xi) \varepsilon \sqrt{k} + 2C \frac{k}{\beta} \log \frac{k}{c \varepsilon}. \quad (D.11)$$

Taking $\varepsilon = 1/\beta$ and letting $\beta \to \infty$ we get $q_l(\xi; k) = q_u(\xi; k)$, and hence the limit $q(\xi; k)$ exists as claimed.

D.2. Proof of Lemma C.4. Note that $Q(\cdot)$ is non-decreasing in its argument, i.e.

$$B_1 \preceq B_2 \Rightarrow Q(B_1) \leq Q(B_2). \quad (D.12)$$

Indeed, if $X_1$ is an optimizer for the SDP with argument $B_1$, then $Q(B_1) = \langle B_1, X_1 \rangle \leq \langle B_2, X_1 \rangle \leq Q(B_2)$ (the first inequality follows from positive semidefiniteness of $X$).

In particular, this implies $\mathbb{E} Q(B(\xi_1)) \leq \mathbb{E} Q(B(\xi_2))$ whenever $\xi_1 \leq \xi_2$. Since by Corollary C.2, $q(\xi) = \lim_{n \to \infty} n^{-1} \mathbb{E} Q(B(\xi))$, it follows that $\xi \mapsto q(\xi)$ is non-decreasing.

Further, since $B(\xi) \succeq B(0)$ for all $\xi \geq 0$, it follows from Lemma 2.2 that $q(\xi) \geq 2$ for all $\xi \geq 0$. Finally [FP07] yields that

$$\lim_{n \to \infty} \lambda_{\max}(B(\xi)) = \begin{cases} 2 & \text{if } \xi \in [0, 1], \\ \xi + 1/\xi & \text{if } \xi > 1. \end{cases} \quad (D.13)$$

(These limits hold almost surely but also in $L_1$, and hence in expectation.)

Since $Q(B(\xi)) \leq n \lambda_{\max}(B(\xi))$, it follows that $q(\xi) \leq 2$ for $\xi \in [0, 1]$, and $q(\xi) \leq \xi + 1/\xi$ for $\xi > 1$, thus concluding the proof.

Appendix E. A better bound on $\xi^*$

In this appendix we prove that the factor 2 in Theorem 1.7 can be replaced by $\xi_0 \in [1, 2)$ strictly. Indeed we prove the following.

Proposition E.1. Let $\xi^*$ be defined as per Eq. $(C.10)$. Then $\xi^* < 2$ strictly.

Proof. Without loss of generality, we can define $B(\xi)$ Appendix D, i.e.

$$B(\xi) \equiv \frac{\xi}{n} \mathbf{1} \mathbf{1}^T + W. \quad (E.1)$$
Fixing $\rho > 0$ bounded, we let
\[ V(\rho) \equiv \left\{ i \in [n] : \sum_{j=1}^{n} W_{ij} \leq -\rho \sqrt{(n+1)/n} \right\}. \tag{E.2} \]

Note that $\sum_{j=1}^{n} W_{ij} \sim \mathcal{N}(0, (n + 1)/n)$, whence
\[ \mathbb{E}[V(\rho)] = n \Phi(-\rho), \tag{E.3} \]
with $\Phi(x) \equiv \int_{-\infty}^{x} e^{-t^2/2} \, dt / \sqrt{2\pi}$ denoting the Gaussian distribution.

We also have, for $i \neq j$, $i, j \in [n]$, and letting $\rho_n \equiv \rho \sqrt{(n+1)/n}$, and $Z, Z', Z_1, Z_2 \sim \mathcal{N}(0, 1)$ independent,
\[ P(i \in V(\rho), j \in V(\rho)) - P(i \in V(\rho))P(j \in V(\rho)) = \]
\[ = P(Z_1 + n^{-1/2}Z \leq -\rho_n; Z_2 + n^{-1/2}Z \leq -\rho_n) - \Phi(-\rho)^2 \tag{E.4} \]
\[ = \mathbb{E}\{\Phi(-\rho_n + n^{-1/2}Z)^2\} - \mathbb{E}\{\Phi(-\rho_n + n^{-1/2}Z')^2\} \tag{E.5} \]
\[ = \frac{1}{2} \mathbb{E}\left\{[\Phi(-\rho_n + n^{-1/2}Z) - \Phi(-\rho_n + n^{-1/2}Z')]^2\right\}. \tag{E.6} \]

Since $\Phi'(x) \leq 1/\sqrt{2\pi}$, this implies
\[ \left| \mathbb{P}(i \in V(\rho), j \in V(\rho)) - \mathbb{P}(i \in V(\rho))\mathbb{P}(j \in V(\rho)) \right| \leq \frac{1}{4\pi n} \mathbb{E}\{(Z - Z')^2\} = \frac{1}{2\pi n}. \tag{E.7} \]

Hence
\[ \text{Var}(\|V(\rho)\|) \leq n\Phi(-\rho)(1 - \Phi(-\rho)) + n(n - 1)\left\{\mathbb{P}(1 \in V(\rho), 2 \in V(\rho)) - \mathbb{P}(1 \in V(\rho))\mathbb{P}(2 \in V(\rho))\right\} \]
\[ \leq \frac{n}{2}. \tag{E.8} \]

In particular, by Chebyshev inequality,
\[ \mathbb{P}\left(\|V(\rho)\| - n\Phi(-\rho) \geq n\varepsilon\right) \leq \frac{1}{2n\varepsilon}. \tag{E.9} \]

Next let $\ell = \lambda n$ for some $\lambda \in (0, 1/2)$. For any $R \subseteq [n], |R| = \ell$, we have $\langle 1_R, W 1_R \rangle \sim \mathcal{N}(0, 2\ell^2/n)$. Then
\[ \mathbb{P}\left(\max_{R \subseteq [n], |R| = \ell} \langle 1_R, W 1_R \rangle \geq nt \right) \leq \frac{n}{\ell} \mathbb{P}\left(\langle 1_R, W 1_R \rangle \geq nt \right) \]
\[ \leq e^{nH(\lambda)} \Phi\left(-n^3t^2/(2\ell^2)\right) \leq \exp \left\{n\left[\lambda \log(\varepsilon/\lambda) - t^2/4\lambda^2\right]\right\}, \tag{E.10} \]
with $H(\lambda) = -\lambda \log \lambda - (1 - \lambda) \log(1 - \lambda)$ the entropy function. Note that the exponent is negative for $t > F_*(\lambda) = (4\lambda^2H(\lambda))^{1/2}$. By a standard calculation, there exists a constant $C = C(\lambda)$ such that
\[ \mathbb{E}\left(\max_{R \subseteq [n], |R| = \ell} \langle 1_R, W 1_R \rangle \right) \leq n F_*(\lambda) + C. \tag{E.11} \]

Now define $s \in \{+1, -1\}^n$ by
\[ s_i = \begin{cases} +1 & \text{if } i \in V(\rho), \\ -1 & \text{otherwise.} \end{cases} \tag{E.12} \]
We then have
\[ Q(B(\xi)) \geq \langle s, B(\xi)s \rangle \quad (E.16) \]
\[ = \langle 1, B(\xi)1 \rangle - 4 \langle 1_{V(\rho)}, B(\xi)1 \rangle + 4 \langle 1_{V(\rho)}, B(\xi)1_{V(\rho)} \rangle \quad (E.17) \]
\[ = \xi \left( n - 4|V(\rho)| + 4 \frac{|V(\rho)|^2}{n} \right) + \langle 1, W1 \rangle - 4 \langle 1_{V(\rho)}, W1 \rangle + 4 \langle 1_{V(\rho)}, W1_{V(\rho)} \rangle \quad (E.18) \]
\[ \geq \xi \left( n - 4|V(\rho)| + 4 \frac{|V(\rho)|^2}{n} \right) + \langle 1, W1 \rangle + 4 \rho |V(\rho)| - 4 \min_{R \subseteq \{n\}} \langle 1_R, W1_R \rangle. \quad (E.19) \]

Using Eqs. (E.11) and (E.13), and letting \( \lambda \equiv \Phi(-\rho) \in (0, 1/2) \) we get
\[ \frac{1}{n} \mathbb{E}Q(B(\xi)) \geq \xi \left( 1 - 2\lambda \right)^2 + 4 \lambda \rho \Phi^{-1}(1 - \lambda) - 4 \sqrt{4\lambda^2 H(\lambda)} - o(1), \quad (E.20) \]
and hence, for any \( \lambda \in (0, 1/2) \),
\[ q(\xi) \geq q_0(\xi; \lambda) \equiv \xi \left( 1 - 2\lambda \right)^2 + 4 \lambda \rho \Phi^{-1}(1 - \lambda) - 4 \sqrt{4\lambda^2 H(\lambda)}. \quad (E.21) \]

We conclude that
\[ \xi_* \leq \xi_0 \equiv \inf \{ \xi \geq 1 : \exists \lambda \in (0, 1/2) \text{ s.t. } q_0(\xi; \lambda) > 2 \}. \quad (E.22) \]

We now claim that \( \xi_0 < 2 \) strictly. Indeed, expanding \( q_0(2; \lambda) \) for small \( \lambda \) yields
\[ q_0(2; \lambda) = 2 + 4\lambda (2 \log(1/\lambda))^{1/2} (1 + o_\lambda(1)). \quad (E.23) \]
It follows that there exists \( \lambda_0 \) such that \( q_0(2; \lambda_0) > 2 \) strictly. By continuity of the function \( \xi \to q_0(\xi; \lambda_0) \), there exists \( \varepsilon > 0 \) such that \( q_0(2 - \varepsilon; \lambda_0) > 2 \) as well. This proves our claim. \( \square \)
References

[ABH14] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall, Exact recovery in the stochastic block model, arXiv:1405.3267 (2014).

[AGZ09] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, An introduction to random matrices, Cambridge studies in advanced mathematics., Cambridge University Press, 2009.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov, Finding a large hidden clique in a random graph, Random Structures and Algorithms 13 (1998), no. 3-4, 457–466.

[AMMN06] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor, Quadratic forms on graphs, Inventiones mathematicae 163 (2006), no. 3, 499–522.

[BdOFV10] Jop Briët, Fernando Mário de Oliveira Filho, and Frank Vallentin, The positive semidefinite grothendieck problem with rank constraint, Automata, Languages and Programming, Springer, 2010, pp. 31–42.

[BGT13] Mohsen Bayati, David Gamarnik, and Prasad Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs, Annals of Probability 41 (2013), no. 6, 4080–4115.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart, Concentration inequalities: A nonasymptotic theory of independence, Oxford University Press, 2013.

[Bri10] Briët, Jop and de Oliveira Filho, Fernando Mário and Vallentin, Frank, Grothendieck inequalities for semidefinite programs with rank constraint, arXiv:1011.1754 (2010).

[CO06] Amin Coja-Oghlan, A spectral heuristic for bisecting random graphs, Random Structures & Algorithms 29 (2006), no. 3, 351–398.

[CO10] ______, Graph partitioning via adaptive spectral techniques, Combinatorics, Probability and Computing 19 (2010), no. 2, 227–284.

[DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, Physical Review E 84 (2011), no. 6, 066106.

[DMS15] Amir Dembo, Andrea Montanari, and Subhabrata Sen, Extremal cuts of sparse random graphs, arXiv:1503.03923 (2015).

[FLT03] Silvio Franz, Michele Leone, and Fabio L. Toninelli, Replica bounds for diluted non-poissonian spin systems, J. Phys. A 36 (2003), 10967–10985.

[FO05] Uriel Feige and Eran Ofek, Spectral techniques applied to sparse random graphs, Random Structures & Algorithms 27 (2005), no. 2, 251–275.

[FP07] Delphine Féra and Sandrine Péché, The largest eigenvalue of rank one deformation of large wigner matrices, Communications in mathematical physics 272 (2007), no. 1, 185–228.

[Gro06] Alexander Grothendieck, Résumé de la théorie métrique des produits tensoriels topologiques, Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo 2 (1996), no. 4, 401–481.

[GT02] Francesco Guerra and Fabio L. Toninelli, The infinite volume limit in generalized mean field disordered models, Markov Processes Relat. Fields 9 (2002), no. cond-mat/0208579, 195–207.

[GT04] ______, The high temperature region of the Viana-Bray diluted spin glass models, J. Stat. Phys 115 (2004), 531–555.

[GV14] Olivier Guédon and Roman Vershynin, Community detection in sparse networks via grothendieck’s inequality, arXiv:1411.4686 (2014).

[HXX14] Bruce Hajek, Yihong Wu, and Jiaming Xu, Achieving exact cluster recovery threshold via semidefinite programming, arXiv:1412.6156 (2014).

[HXX15] ______, Achieving exact cluster recovery threshold via semidefinite programming: Extensions, arXiv:1502.07738 (2015).

[JMMRT15] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi, Phase transitions in semidefinite relaxations: A statistical physics analysis, In preparation, 2015.

[KMM+13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang, Spectral redemption in clustering sparse networks, Proceedings of the National Academy of Sciences 110 (2013), no. 52, 20935–20940.

[KMO10] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh, Matrix completion from noisy entries, Journal of Machine Learning Research 11 (2010), 2057–2078.

[KN12] Subhash Khot and Assaf Naor, Grothendieck-type inequalities in combinatorial optimization, Communications on Pure and Applied Mathematics 65 (2012), no. 7, 992–1035.
[KS03] Michael Krivelevich and Benny Sudakov, *The largest eigenvalue of sparse random graphs*, Combinatorics, Probability and Computing **12** (2003), no. 01, 61–72.

[Lin22] Jarl Waldemar Lindeberg, *Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift **15** (1922), no. 1, 211–225.

[Mas14] Laurent Massoulié, *Community detection thresholds and the weak Ramanujan property*, Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, 2014, pp. 694–703.

[McS01] Frank McSherry, *Spectral partitioning of random graphs*, Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on, IEEE, 2001, pp. 529–537.

[Meg01] Alexandre Megretski, *Relaxations of quadratic programs in operator theory and system analysis*, Systems, approximation, singular integral operators, and related topics, Springer, 2001, pp. 365–392.

[MNS12] Elchanan Mossel, Joe Neeman, and Allan Sly, *Stochastic block models and reconstruction*, arXiv:1202.1499 (2012).

[NJW*02] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al., *On spectral clustering: Analysis and an algorithm*, Advances in neural information processing systems **2** (2002), 849–856.

[NRT99] Arkadi Nemirovski, Cornelis Roos, and Tamás Terlaky, *On maximization of quadratic form over intersection of ellipsoids with common center*, Mathematical Programming **86** (1999), no. 3, 463–473.

[PT04] Dmitry Panchenko and Michel Talagrand, *Bounds for diluted mean-fields spin glass models*, Probability Theory and Related Fields **130** (2004), no. 3, 319–336.

[Rie74] Ronald E Rietz, *A proof of the Grothendieck inequality*, Israel Journal of Mathematics **19** (1974), no. 3, 271–276.

[Vu05] Van H Vu, *Spectral norm of random matrices*, Proceedings of the thirty-seventh annual ACM symposium on Theory of computing, ACM, 2005, pp. 423–430.

**Department of Electrical Engineering and Department of Statistics, Stanford University**

**Department of Statistics, Stanford University**