Community detection in sparse networks via Grothendieck's inequality

Olivier Guédon, Roman Vershynin

COMMUNITY DETECTION IN SPARSE NETWORKS VIA GROTHENDIECK'S INEQUALITY

OLIVIER GUÉDON AND ROMAN VERSHYNIN

ABSTRACT. We present a simple and flexible method to prove consistency of semidefinite optimization problems on random graphs. The method is based on Grothendieck's inequality. Unlike the previous uses of this inequality that lead to constant relative accuracy, we achieve any given relative accuracy by leveraging randomness. We illustrate the method with the problem of community detection in sparse networks, those with bounded average degrees. We demonstrate that even in this regime, various natural semidefinite programs can be used to recover the community structure up to an arbitrarily small fraction of misclassified vertices. The method is general; it can be applied to a variety of stochastic models of networks and semidefinite programs.

1. INTRODUCTION

1.1. **Semidefinite problems on random graphs.** In this paper we present a simple and general method to prove consistency of various semidefinite optimization problems on random graphs.

Suppose we observe one instance of an $n \times n$ symmetric random matrix A with unknown expectation $\bar{A} := \mathbb{E} A$. We would like to estimate the solution of the discrete optimization problem

$$\text{maximize } x^{\mathsf{T}} \bar{A} x \quad \text{subject to} \quad x \in \{-1, 1\}^n. \tag{1.1}$$

A motivating example of A is the adjacency matrix of a random graph; the Boolean vector x can represent a partition of vertices of the graph into two classes. Such Boolean problems can be encountered in the context of *community detection in networks* which we will discuss shortly. For now, let us keep working with the general class of problems (1.1).

Since $\bar{A}$ is unknown, one might hope to estimate the solution $\bar{x}$ of (1.1) by solving the random instance of this problem, that is

$$\text{maximize } x^{\mathsf{T}} A x \quad \text{subject to} \quad x \in \{-1, 1\}^n. \tag{1.2}$$

The integer quadratic problem (1.2) is NP-hard for general (non-random) matrices A. Semidefinite relaxations of many problems of this type have been proposed; see [34, 49, 58, 6] and the references therein. Such relaxations are known to have *constant* relative accuracy. For example, a semidefinite

1

relaxation in [6] computes, for any given positive semidefinite matrix A, a vector $x_0 \in \{-1,1\}^n$ such that $x_0^\mathsf{T} A x_0 \geq 0.56 \max_{x \in \{-1,1\}^n} x^\mathsf{T} A x$.

In this paper we demonstrate how semidefinite relaxations of (1.2) can recover a solution of (1.1) with *any given relative accuracy*. Like several previously known methods, our approach is based on Grothendieck's inequality. We refer the reader to the surveys [62, 41] for many reformulations and applications of this inequality in mathematics, computer science, optimization and other fields. In contrast to the previous methods, we are going to apply Grothendieck's inequality for the (random) error $A - \bar{A}$ rather that the original matrix A, and this will be responsible for the arbitrary accuracy.

We will describe the general method in Section 2. It is simple and flexible, and it can be used for showing consistency of a variety of semidefinite programs, which may or may not be related to Boolean problems like (1.1). But before describing the method, we would like to pause and give some concrete examples of results it yields for *community detection*, a fundamental problem in network analysis.

For simplicity, we will first focus on the classical stochastic block model, which is a random network whose nodes are split into two equal-sized clusters. In Section 1.3 we will extend our discussion for much broader models of networks almost without extra effort.

1.2. Community detection: the classical stochastic block model. It
is now customary to model networks as inhomogeneous random graphs [13], which generalize the classical Erdös-Rényi model $G(n,p)$. A benchmark example is the stochastic block model [40]. In this section we focus on the basic model with two communities of equal sizes; in Section 1.3 we will consider a much more general situation.

We define a random graph on vertices $\{1, \ldots, n\}$ as follows. Partition the set of vertices into two communities $\mathcal{C}_1$ and $\mathcal{C}_2$ of size $n/2$ each. For each pair of distinct vertices, we draw an edge independently with probability p if both vertices belong to the same community, and q (with $q \leq p$) if they belong to different communities. For convenience we include the loops, so each vertex has an edge connecting it to itself with probability 1. This defines a distribution on random graphs which is denoted $G(n,p,q)$ and called the (classical) *stochastic block model*. When $p = q$, we recover the classical Erdös-Rényi model of random graphs $G(n,p)$.

The *community detection problem* asks to recover the communities $\mathcal{C}_1$ and $\mathcal{C}_2$ by observing one instance of a random graph drawn from $G(n,p,q)$. As we will discuss in detail in Section 1.4, an array of algorithms is known to succeed for this problem for *relatively dense graphs*, those whose expected average degree (which is of order pn) is $\Omega(\log n)$, while much less is known for *totally sparse graphs* – those with bounded average degrees, i.e. with $pn = O(1)$. Our paper focuses on this challenging, sparse regime.

Recovery of the communities $\mathcal{C}_1$ and $\mathcal{C}_2$ is equivalent to estimating the community membership vector, which we can define as

$$\bar{x} \in \{-1, 1\}^n, \quad \bar{x}_i = \begin{cases} 1, & i \in \mathcal{C}_1 \\ -1, & i \in \mathcal{C}_2. \end{cases} \tag{1.3}$$

We will estimate $\bar{x}$ using the following semidefinite optimization problem:

$$\begin{aligned} & \text{maximize } \langle A, Z \rangle - \lambda \langle E_n, Z \rangle \\ & \text{subject to } Z \succeq 0, \ \text{diag}(Z) \preceq \mathbf{I}_n. \end{aligned} \tag{1.4}$$

Here the inner product of matrices is defined in the usual way, that is $\langle A, B \rangle = \text{tr}(AB) = \sum_{i,j} A_{ij} B_{ij}$, $\mathbf{I}_n$ denotes the identity matrix, the matrix E_n has all entries equal 1, and $A \succeq B$ means that $A - B$ is positive semidefinite. Observe that $E_n = \mathbf{1}_n \mathbf{1}_n^\mathsf{T}$ where $\mathbf{1}_n \in \mathbb{R}^n$ is the vector whose all coordinates equal 1. The constraint $\text{diag}(Z) \preceq \mathbf{I}_n$ in (1.4) simply means that all diagonal entries of Z are bounded by 1.

For the value of λ we choose the average degree of the graph (with loops removed), which is

$$\lambda = \frac{2}{n(n-1)} \sum_{i<j} a_{ij} \tag{1.5}$$

where $a_{ij} \in \{0, 1\}$ denote the entries of the adjacency matrix A.

Theorem 1.1 (Community detection in classical stochastic block model). *Let $\varepsilon \in (0, 1)$ and $n \geq 10^4 \varepsilon^{-2}$. Let A be the adjacency matrix of the random graph drawn from the stochastic block model $G(n, p, q)$ with $\max\{p(1 - p), q(1 - q)\} \geq \frac{20}{n}$. Assume that $p = \frac{a}{n} > q = \frac{b}{n}$, and*

$$(a - b)^2 \geq 10^4 \, \varepsilon^{-2} (a + b). \tag{1.6}$$

Let $\widehat{Z}$ be a solution of the semidefinite program (1.4). Then, with probability at least $1 - e^3 5^{-n}$, we have

$$\|\widehat{Z} - \bar{x}\bar{x}^\mathsf{T}\|_2^2 \leq \varepsilon n^2 = \varepsilon \|\bar{x}\bar{x}^\mathsf{T}\|_2^2. \tag{1.7}$$

Here and in the rest of this paper, $\| \cdot \|_2$ denotes the Frobenius norm of matrices and the Euclidean norm of vectors.

Once we have estimated the rank-one matrix $\bar{x}\bar{x}^\mathsf{T}$ using Theorem 1.1, we can also estimate the community membership vector $\bar{x}$ itself in a standard way, namely by computing the leading eigenvector.

Corollary 1.2 (Community detection with $o(n)$ misclassified vertices). *In the setting of Theorem 1.1, let $\widehat{x}$ denote an eigenvector of $\widehat{Z}$ corresponding to the largest eigenvalue, and with $\|\widehat{x}\|_2 = \sqrt{n}$. Then*

$$\min_{\alpha = \pm 1} \|\alpha \widehat{x} - \bar{x}\|_2^2 \leq \varepsilon n = \varepsilon \|\bar{x}\|_2^2.$$

In particular, the signs of the coefficients of $\widehat{x}$ correctly estimate the partition of the vertices into the two communities, up to at most εn misclassified vertices.

As we will discuss in Section 1.4.2 in more detail, there are previously known algorithms for recovery of two communities under conditions similar to (1.6). These include a spectral clustering algorithm based on truncating the high degree vertices (whose analysis can be derived from [32, 31]), combinatorial algorithms of [51, 55] based on path counting, and an algorithm [53] based on belief propagation, which minimizes the fraction of misclassified vertices.

Prior to this work, it was not known whether semidefinite programming could be successful for the analysis of sparse networks. Theorem 1.1 provides a positive answer to this question. Moreover, the method of this paper is flexible enough to analyze many semidefinite programs, and it can be applied for much more general models of sparse networks than any previous results. To illustrate this point, we will now choose a different semidefinite program and show that it succeeds for a large class of stochastic models of networks.

1.3. Community detection: general stochastic block models. Let us describe a model of networks where one can have multiple communities of arbitrary sizes, arbitrarily many outliers, and unequal edge probabilities.

To define such *general stochastic block model*, we assume that the set of vertices $\{1, \ldots, n\}$ is partitioned into communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$ of arbitrary sizes. We do not restrict the sizes of the communities, so in particular this model can automatically handle outliers, the vertices that form communities of size 1. For each pair of distinct vertices (i, j), we draw an edge between i and j independently and with certain fixed probability p_{ij}. For convenience we include the loops like in the classical stochastic block model, so $p_{ii} = 1$. To promote more edges within than across the communities, we assume that there exist numbers $p > q$ (thresholds) such that

$$
\begin{aligned}
p_{ij} &\geq p \quad \text{if } i \text{ and } j \text{ belong to the same community;} \\
p_{ij} &\leq q \quad \text{if } i \text{ and } j \text{ belong to different communities.}
\end{aligned}
\tag{1.8}
$$

The community structure of such a network is captured by the cluster matrix matrix $\bar{Z} \in \{0, 1\}^{n \times n}$ defined as

$$
\bar{Z}_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ belong to the same community;} \\ 0 & \text{if } i \text{ and } j \text{ belong to different communities.} \end{cases}
\tag{1.9}
$$

We will estimate $\bar{Z}$ using the following semidefinite optimization program:

$$
\begin{aligned}
&\text{maximize } \langle A, Z \rangle \\
&\text{subject to } Z \succeq 0, \ Z \geq 0, \ \text{diag}(Z) \preceq \mathbf{I}_n, \ \sum_{i,j=1}^n Z_{ij} = \lambda.
\end{aligned}
\tag{1.10}
$$

Here as usual $Z \succeq 0$ means that Z is positive semidefinite, and $Z \geq 0$ means that all entries of Z are non-negative. We choose the value of λ to be the number of elements in the cluster matrix, that is

$$
\lambda = \sum_{i,j=1}^n \bar{Z}_{ij} = \sum_{k=1}^K |\mathcal{C}_k|^2.
\tag{1.11}
$$

If all communities have the same size s, then $\lambda = Ks^2 = ns$.

Theorem 1.3 (Community detection in general stochastic block model).
*Let $\varepsilon \in (0,1)$. Let A be the adjacency matrix of the random graph drawn
from the general stochastic block model described above. Denote by $\bar{p}$ the
expected variance of the edges, that is $\bar{p} = \frac{2}{n(n-1)} \sum_{i<j} p_{ij}(1-p_{ij})$. Assume
that $p = \frac{a}{n} > q = \frac{b}{n}$, $\bar{p} = \frac{g}{n}$, $g \geq 9$ and*

$$(a-b)^2 \geq 484\, \varepsilon^{-2} g. \tag{1.12}$$

Let $\widehat{Z}$ be a solution of the semidefinite program (1.10). *Then, with probability
at least $1 - e^3 5^{-n}$, we have*

$$\|\widehat{Z} - \bar{Z}\|_2^2 \leq \|\widehat{Z} - \bar{Z}\|_1 \leq \varepsilon n^2. \tag{1.13}$$

Here as usual $\| \cdot \|_2$ denotes the Frobenius norm of matrices, and $\| \cdot \|_1$
denotes the ℓ_1 norm of the matrices considered as vectors, that is $\|(a_{ij})\|_1 = \sum_{i,j} |a_{ij}|$.

Remark 1.4 (If the sizes of communities are not known). Our choice of the
parameter λ in (1.11) assumes that we know the sizes of the communities.
What if they are not known? From the proof of Theorem 1.3 it will be clear
what happens when $\lambda > 0$ is chosen *arbitrarily*. Assume that we choose
λ so that $\lambda \leq \lambda_0 := \sum_k |\mathcal{C}_k|^2$. Then instead of estimating the full cluster
graph (described in Remark 1.5), the solution $\widehat{Z}$ will only estimate a certain
subgraph of the cluster graph, which may miss at most $\lambda_0 - \lambda$ edges. On the
other hand, if we choose λ so that $\lambda \geq \lambda_0$, then the solution $\widehat{Z}$ will estimate
a certain *supergraph* of the cluster graph, which may have at most $\lambda - \lambda_0$
extra edges. In either case, such solution could be meaningful in practice.

Remark 1.5 (Cluster graph). It may be convenient to view the cluster ma-
trix $\bar{Z}$ as the adjacency matrix of the *cluster graph*, in which all vertices
within each community are connected and there are no connections across
the communities. This way, the semidefinite program (1.10) takes a sparse
graph as an input, and it returns an estimate of the cluster graph as an
output. The effect of the program is thus to "densify" the network inside
the communities and "sparsify" it across the communities.

Remark 1.6 (Other semidefinite programs). There is nothing special about
the semidefinite programs (1.4) and (1.10). For example, one can tighten
the constraints and instead of $\mathrm{diag}(Z) \preceq \mathbf{I}_n$ require that $\mathrm{diag}(Z) = \mathbf{I}_n$ in
both programs. Similarly, instead of placing in (1.10) the constraint on
the sum of all entries of Z, one can place constraints on the sums of each
row. In a similar fashion, one should be able to analyze other semidefi-
nite relaxations, both new and those proposed in the previous literature on
community detection, see [9].

For one more illustration for the method of this paper, we refer the reader
to Section 8 of [36]. There we consider a minor modification of the semi-
definite program (1.4), and we show that it succeeds in presence of multiple

communities of equal sizes (the so-called *balanced planted partition model*). The sufficient condition for that is $(a-b)^2 \geq 50^2 \varepsilon^{-2}(a+b(K-1))$ where K is the number of communities, s the size of the communities and $p = a/s$, $q = b/s$.

1.4. **Related work.** Community detection in stochastic block models is a fundamental problem that has been extensively studied in theoretical computer science and statistics. A plethora or algorithmic approaches have been proposed, in particular those based on combinatorial techniques [18, 30], spectral clustering [14, 5, 4, 50, 59, 64, 22, 57, 43, 63, 38], likelihood maximization [65, 10, 8], variational methods [3, 20, 11], Markov chain Monte Carlo [65, 60], belief propagation [29], and convex optimization including semidefinite programming [37, 7, 61, 2, 23, 24, 25, 19, 9, 18].

1.4.1. *Relatively dense networks: average degrees are* $\Omega(\log n)$. Most known rigorous results on community detection are proved for relatively dense networks whose expected degrees go to infinity with n. If the degrees grow no slower than $\log n$, it may be possible to recover the community structure *perfectly*, without any misclassified vertices. A variety of community detection methods are known to succeed in this regime, including those based on spectral clustering, likelihood maximization and convex optimization mentioned above; see e.g. [50, 18] and the references therein.

The semidefinite programs (1.4) and (1.10) are similar to those proposed in the recent literature, most notably in [23, 25, 19, 9, 18]. The semidefinite relaxations discussed in [25, 19] can perfectly recover the community structure if $(a-b)^2 \geq C(a \log n + b)$ for a sufficiently large constant C; see [9] for a review of these results.

1.4.2. *Totally sparse networks: bounded average degrees.* The problem becomes more difficult for sparser networks, whose expected average degrees grow to infinity arbitrarily slowly or even remain bounded in n. Although studying such networks is well motivated from the practical perspective [45, 66], little has been known on the theoretical level.

If the degrees grow slower than $\log n$, it is impossible to correctly classify all vertices, since with high probability a positive fraction of the vertices will be isolated. Still, the fraction of isolated vertices tends to zero with n, so we can hope to correctly classify a *majority* of the vertices in this regime.

A powerful spectral method developed by J. Kahn and E. Szemeredi for random regular graphs [32] can be adapted for Erdös-Rényi random graphs [5, 31] and, more generally, for the stochastic block model $G(n, \frac{a}{n}, \frac{b}{n})$. If one *truncates the graph* by removing all vertices with too large degrees (say, larger than $10(a+b)$), then the argument of [32, 31] can be adapted to conclude that the truncated adjacency matrix concentrates near its expectation in the spectral norm. The communities can then be approximately recovered using the *spectral clustering*, which is based on the signs of the coefficients of

the second eigenvector. Working out the details, one finds that a sufficient condition for this method to succeed is similar to (1.6), that is

$$(a - b)^2 \geq C_\varepsilon (a + b) \tag{1.14}$$

where C_ε depends only on the desired accuracy ε or recovery. However, for real networks it is usually undesirable and impractical to remove high degree vertices, since these vertices are usually the most important ones.

A. Coja-Oghlan [27] proposed a different, complicated adaptive spectral algorithm that can approximately recover communities under the condition $(a - b)^2 \geq C_\varepsilon (a + b) \log(a + b)$. Recently, L. Massoulié [51] and E. Mossel, J. Neeman and A. Sly [55] came up with combinatorial algorithms based on path counting, which can approximately recover communities under the condition (1.14). These results are stated in the asymptotic regime for $n \to \infty$ and without explicit dependence of C_ε on the desired accuracy ε. Furthermore, E. Mossel, J. Neeman and A. Sly developed an algorithm based on belief propagation [53], which minimizes the fraction of misclassified vertices.

Condition (1.14) has the optimal form. Indeed, it was shown in [56] that the lower bound (1.14) is required for any algorithm to be able to recover communities with at most εn misclassified vertices, where $C_\varepsilon \to \infty$ as $\varepsilon \to 0$. A conjecture of A. Decelle, F. Krzakala, C. Moore and L. Zdeborova proved recently by E. Mossel, J. Neeman and A. Sly [54, 55] and Massouile [51] states that one can find a partition *correlated* with the true community partition (i.e. with the fraction of misclassified vertices bounded away from 50% as $n \to \infty$) if $(a - b)^2 \geq C(a + b)$ with some constant $C > 2$. Moreover, this result achieves information-theoretic limit: no algorithm can succeed if $C \leq 2$.

It remains an open question whether semidefinite programing can achieve similar information-theoretic limits. Theorem 1.1 does not achieve them; addressing this problem will require to tighten the absolute constant and the dependence on ε in (1.6).

1.4.3. *The new results in historical perspective.* Prior to this work, it has not been known whether community detection in totally sparse networks is possible using semidefinite programming. We show that this is possible under the condition (1.14) that is optimal up to an absolute constant. A variety of simple and general semidefinite programs succeed for this purpose, in particular (1.4) and (1.10).

Furthermore, the method of the present paper generalizes smoothly for a broad classes of sparse networks. Stochastic block models with multiple communities and outliers have been studied in the statistical literature before; the semidefinite relaxations proposed in [23, 25, 19, 9] were designed for multiple communities and outliers. However, previous theoretical results for multiple communities were only available for dense regime where the degrees grow as $\Omega(\log n)$, in which case perfect community detection is possible.

1.4.4. *Follow up work.* After this paper had been submitted, several new results appeared on community detection in stochastic block models. We will mention here only results that apply for totally sparse networks. The initial discovery of [51, 55] mentioned in Section 1.4.2 was followed by the work [16]. Semidefinite programs on random graphs were further analyzed in [52] using higher-rank Grothendieck inequalities and insights from mathematical physics. Stochastic block models with labeled edges were addressed in [44] using truncated spectral clustering (with high degree vertices removed, based on [31]) and semidefinite programming (whose analysis is based on the method of the present paper). A two-stage algorithm based on truncated spectral clustering and swapping vertices (like e.g. in [56]) was analyzed in [26]; the swapping stage leads to the sufficient condition (1.14) with with an optimal dependence on the accuracy, $C_\varepsilon \sim \log(1/\varepsilon)$. A different combinatorial method was proposed and analyzed in [1]; regularized spectral clustering was shown to succeed in [46, 47]; and a computationally feasible likelihood-based algorithm that minimizes the risk for misclassification proportion was found in [33]. Some of the mentioned work can be used for networks with multiple communities, see [26, 1, 46, 47, 33].

1.5. **Plan of the paper.** We discuss the method in general terms in Section 2. We explain how Grothendieck's inequality can be used to show tightness of various semidefinite programs on random graphs. Section 3 is devoted to Grothendieck's inequality and its implications for semidefinite programming. In Section 4 we prove a simple concentration inequality for random matrices in the cut norm. In Section 5 we specialize to the community detection problem for the classical stochastic block model, and we prove Theorem 1.1 and Corollary 1.2 there. In Section 6 we consider the general classical stochastic block model, and we prove Theorem 1.3 there.

Acknowledgement. This work was carried out while the first author was a Gerhing Visiting Professor at the University of Michigan. He thanks this institution for hospitality. The second author is grateful to Alexander Barvinok for drawing his attention to Y. Nesterov's work [58] on combinatorial optimization and to Grothendieck's inequality in this context. We also thank Elchanan Mossel for useful discussions, and the anonymous referees whose suggestions helped to improve the presentation.

2. SEMIDEFINITE OPTIMIZATION ON RANDOM GRAPHS: THE METHOD IN A NUTSHELL

In this section we explain the general method of this paper, which can be applied to a variety of optimization problems. To be specific, let us return to the problem we described in Section 1.1, which is to estimate the solution $\bar{x}$ of the optimization problem (1.1) from a single observation of the random matrix A. We suggested there to approximate $\bar{x}$ by the solution of

the (random) program (1.2), which we can rewrite as follows:

$$\text{maximize } \langle A, xx^\mathsf{T} \rangle \quad \text{subject to} \quad x \in \{-1, 1\}^n. \tag{2.1}$$

Note that if we maximized $\langle A, xx^\mathsf{T} \rangle$ over the Euclidean ball $B(0, \sqrt{n})$, then the problem would be simple – the solution x would be the eigenvector corresponding to the eigenvalue of A of largest magnitude. This simpler problem underlies the most basic algorithm for community detection called *spectral clustering*, where the communities are recovered based on the signs of an eigenvector of the adjacency matrix (going back to [39, 14, 50], see [64]). The optimization problem (2.1) is harder and more subtle; the replacement of the Euclidean ball by the cube introduces a strong restriction on the coordinates of x. This restruction rules out *localized* solutions x where most of the mass of x is concentrated on a small fraction of coordinates. Since eigenvectors of sparse matrices tend to be localized (see [15]), basic spectral clustering is often unsuccessful for sparse networks.

Let us choose a convex subset $\mathcal{M}_{\text{opt}}$ of the set of positive semidefinite matrices whose all entries are bounded by 1 in absolute value. (For now, it can be any subset.) Note that xx^T appearing in (2.1) are examples of such matrices. We consider the following semidefinite relaxation of (2.1):

$$\text{maximize } \langle A, Z \rangle \quad \text{subject to} \quad Z \in \mathcal{M}_{\text{opt}}. \tag{2.2}$$

We might hope that the solution $\widehat{Z}$ of this program would enable us to estimate the solution $\bar{x}$ of (1.1).

To realize this hope, one needs to check a few things, which may or may not be true depending on the application. First, one needs to design the feasible set $\mathcal{M}_{\text{opt}}$ in such a way that *the semidefinite relaxation of the expected problem* (1.1) *is tight*. This means that the solution $\bar{Z}$ of the program

$$\text{maximize } \langle \bar{A}, Z \rangle \quad \text{subject to} \quad Z \in \mathcal{M}_{\text{opt}} \tag{2.3}$$

satisfies

$$\bar{Z} = \bar{x}\bar{x}^\mathsf{T}. \tag{2.4}$$

This condition can be arranged for in various applications. In particular, this is the case in the setting of Theorem 1.1; we show this in Lemma 5.1.

Second, one needs a *uniform deviation inequality*, which would guarantee with high probability that

$$\max_{x,y \in \{-1,1\}^n} |\langle A - \bar{A}, xy^\mathsf{T} \rangle| \leq \varepsilon. \tag{2.5}$$

This can often be proved by applying standard deviation inequalities for a fixed pair (x, y), followed by a union bound over all such pairs. We prove such a deviation inequality in Section 4.

Now we make the crucial step, which is an application of *Grothendieck's inequality*. A reformulation of this remarkable inequality, which we explain

in Section 3, states that (2.5) automatically implies that

$$\max_{Z \in \mathcal{M}_{\mathrm{opt}}} |\langle A - \bar{A}, Z \rangle| \leq C\varepsilon. \tag{2.6}$$

This will allow us to conclude that the solution $\widehat{Z}$ of (2.2) approximates the solution $\bar{Z}$ of (2.3). To see this, let us compare the value of the expected objective function $\langle \bar{A}, Z \rangle$ at these two vectors. We have

$$
\begin{aligned}
\langle \bar{A}, \widehat{Z} \rangle &\geq \langle A, \widehat{Z} \rangle - C\varepsilon \quad \text{(replacing } \bar{A} \text{ by } A \text{ using (2.6))} \\
&\geq \langle A, \bar{Z} \rangle - C\varepsilon \quad \text{(since } \widehat{Z} \text{ is the maximizer in (2.2))} \\
&\geq \langle \bar{A}, \bar{Z} \rangle - 2C\varepsilon \quad \text{(replacing } A \text{ by } \bar{A} \text{ back using (2.6)).} \tag{2.7}
\end{aligned}
$$

This means that $\widehat{Z}$ almost maximizes the objective function $\langle \bar{A}, Z \rangle$ in (2.3).

The final piece of information we require is that the expected objective function $\langle \bar{A}, Z \rangle$ *distinguishes points near its maximizer* $\bar{Z}$. This would allow one to automatically conclude from (2.7) that the almost maximizer $\widehat{Z}$ is close to the true maximizer, i.e. that

$$\|\widehat{Z} - \bar{Z}\| \leq \text{something small} \tag{2.8}$$

where $\| \cdot \|$ can be the Frobenius or operator norm. Intuitively, the requirement that the objective function distinguishes points amounts to a non-trivial curvature of the feasible set $\mathcal{M}_{\mathrm{opt}}$ at the maximizer $\bar{Z}$. In many situations, this property is easy to verify. In the setting of Theorems 1.1 and 1.3, we check it in Lemma 5.2 and Lemmas 6.2–6.3 respectively.

Finally, we can recall from (2.4) that $\bar{Z} = \bar{x}\bar{x}^{\mathsf{T}}$. Together with (2.8), this yields that $\widehat{Z}$ is approximately a rank-one matrix, and its leading eigenvector $\widehat{x}$ satisfies

$$\|\widehat{x} - \bar{x}\|_2 \leq \text{something small}.$$

Thus we estimated the solution $\bar{x}$ of the problem (1.1) as desired.

Remark 2.1 (General semidefinite programs). For this method to work, it is not crucial that the semidefinite program be a relaxation of any vector optimization problem. Indeed, one can analyze semidefinite programs of the type (2.2) without any vector optimization problem (2.1) in the background. In such cases, the requirement (2.4) of tightness of relaxation can be dropped. The solution $\bar{Z}$ may itself be informative. An example of such situation is Theorem 1.3 where the community membership matrix $\bar{Z}$ is important by itself. However, $\bar{Z}$ can not be represented as $\bar{x}\bar{x}^{\mathsf{T}}$ for any $\bar{x}$, since $\bar{Z}$ is not a rank one matrix.

3. Grothendieck's inequality and semidefinite programming

Grothendieck's inequality is a remarkable result proved originally in the functional analytic context [35] and reformulated in [48] in the form we are going to describe below. This inequality had found applications in several areas [62, 41]. It has already been used to analyze semidefinite relaxations

of hard combinatorial optimization problems [58, 6], although previous relaxations lead to constant (rather than arbitrary) accuracy.

Theorem 3.1 (Grothendieck's inequality). *Consider an $n \times n$ matrix of real numbers $B = (b_{ij})$. Assume that*

$$\left| \sum_{i,j} b_{ij} s_i t_j \right| \leq 1$$

for all numbers $s_i, t_i \in \{-1, 1\}$. Then

$$\left| \sum_{i,j} b_{ij} \langle X_i, Y_j \rangle \right| \leq K_{\mathrm{G}}$$

for all vectors $X_i, Y_i \in B_2^n$.

Here $B_2^n = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ is the unit ball for the Euclidean norm, and K_{G} is an absolute constant referred to as *Grothendieck's constant*. The best value of K_{G} is still unknown, and the best known bound [17] is

$$K_{\mathrm{G}} < \frac{\pi}{2 \ln(1 + \sqrt{2})} \leq 1.783. \tag{3.1}$$

3.1. **Grothendieck's inequality in matrix form.** To restate Grothendieck's inequality in a matrix form, let us assume for simplicity that $m = n$ and observe that $\sum_{i,j} b_{ij} s_i t_j = \langle B, st^{\mathsf{T}} \rangle$ where s and t are the vectors in $\mathbb{R}^n$ with coordinates s_i and t_j respectively. Similarly, $\sum_{i,j} b_{ij} \langle X_i, Y_j \rangle = \langle B, XY^{\mathsf{T}} \rangle$ where X and Y are the $n \times n$ matrices with rows X_i^{T} and Y_j^{T} respectively. This motivates us to consider the following two sets of matrices:

$$\mathcal{M}_1 := \left\{ st^{\mathsf{T}} : s, t \in \{-1, 1\}^n \right\}, \quad \mathcal{M}_{\mathrm{G}} := \left\{ XY^{\mathsf{T}} : \text{all rows } X_i, Y_j \in B_2^n \right\}.$$

Clearly, $\mathcal{M}_1 \subset \mathcal{M}_{\mathrm{G}}$. Grothendieck's inequality can be stated as follows:

$$\forall B \in \mathbb{R}^{n \times n}, \quad \max_{Z \in \mathcal{M}_{\mathrm{G}}} |\langle B, Z \rangle| \leq K_{\mathrm{G}} \max_{Z \in \mathcal{M}_1} |\langle B, Z \rangle|. \tag{3.2}$$

We can view this inequality as a relation between two matrix norms. The right side of (3.2) defines the $\ell_\infty \to \ell_1$ norm of $B = (b_{ij})$, which is

$$\|B\|_{\infty \to 1} = \max_{\|s\|_\infty \leq 1} \|Bs\|_1 = \max_{s,t \in \{-1,1\}^n} \langle B, st^{\mathsf{T}} \rangle = \max_{s,t \in \{-1,1\}^n} \sum_{i,j=1}^n b_{ij} s_i t_j$$

$$= \max_{Z \in \mathcal{M}_1} |\langle B, Z \rangle|. \tag{3.3}$$

We note in passing that this norm is equivalent to the so-called *cut norm*, whose importance in algorithmic problems is well understood in theoretical computer science community, see e.g. [6, 41].

Let us restrict our attention to the part of Grothendieck's set $\mathcal{M}_{\mathrm{G}}$ consisting of positive semidefinite matrices. To do so, we consider the following set of $n \times n$ matrices:

$$\mathcal{M}_{\mathrm{G}}^+ := \{Z : Z \succeq 0, \ \mathrm{diag}(Z) \preceq \mathbf{I}_n\} \subset \mathcal{M}_{\mathrm{G}} \subset [-1, 1]^{n \times n}. \tag{3.4}$$

To check the first inclusion in (3.4), let $Z \in \mathcal{M}_{\mathrm{G}}^+$. Since $Z \succeq 0$, there exists a matrix X such that $Z = X^2$. The rows X_i^{T} of X satisfy $\|X_i\|_2^2 = \langle X_i, X_i \rangle = (X^{\mathsf{T}} X)_{ii} = Z_{ii} \leq 1$, where the last inequality follows from the assumption $\mathrm{diag}(Z) \preceq \mathbf{I}_n$. Choosing $Y = X$ in the definition of $\mathcal{M}_{\mathrm{G}}$, we conclude that $Z \in \mathcal{M}_{\mathrm{G}}$. To check the second inclusion in (3.4), note that for every matrix $XY^{\mathsf{T}} \in \mathcal{M}_{\mathrm{G}}$, we have $(XY^{\mathsf{T}})_{ij} = \langle X_i, Y_j \rangle \leq \|X_i\|_2 \, \|Y_j\|_2 \leq 1$.

Combining (3.2) with (3.4) and the identity (3.3), we obtain the following form of Grothendieck inequality for positive semidefinite matrices.

Fact 3.2 (Grothendieck's inequality, PSD). *Every matrix $B \in \mathbb{R}^{n \times n}$ satisfies*

$$\max_{Z \in \mathcal{M}_{\mathrm{G}}^+} |\langle B, Z \rangle| \leq K_{\mathrm{G}} \|B\|_{\infty \to 1}.$$

3.2. **Semidefinite programming.** To keep the discussion sufficiently general, let us consider the following class of optimization programs:

$$\text{maximize } \langle B, Z \rangle \quad \text{subject to} \quad Z \in \mathcal{M}_{\mathrm{opt}}. \tag{3.5}$$

Here $\mathcal{M}_{\mathrm{opt}}$ can be any subset of the Grothendieck's set $\mathcal{M}_{\mathrm{G}}^+$ defined in (3.4). A good example is where B is the adjacency matrix of a random graph, possibly dilated by a constant matrix. For example, the semidefinite program (1.4) is of the form (3.5) with $\mathcal{M}_{\mathrm{opt}} = \mathcal{M}_{\mathrm{G}}^+$ and $B = A - \lambda E_n$.

Imagine that there is a similar but simpler problem where B is replaced by a certain *reference matrix* R, that is

$$\text{maximize } \langle R, Z \rangle \quad \text{subject to} \quad Z \in \mathcal{M}_{\mathrm{opt}}. \tag{3.6}$$

A good example is where B is a random matrix and $R = \mathbb{E} B$; this will be the case in the proof of Theorem 1.3. Let $\widehat{Z}$ and Z_R be the solutions of the original problem (3.5) and the reference problem (3.6) respectively, thus

$$\widehat{Z} := \arg \max_{Z \in \mathcal{M}_{\mathrm{opt}}} \langle B, Z \rangle, \quad Z_R := \arg \max_{Z \in \mathcal{M}_{\mathrm{opt}}} \langle R, Z \rangle.$$

The next lemma shows that $\widehat{Z}$ provides an almost optimal solution to the reference problem if the original and reference matrices B and R are close.

Lemma 3.3 ($\widehat{Z}$ almost maximizes the reference objective function). *We have*

$$\langle R, Z_R \rangle - 2K_{\mathrm{G}} \|B - R\|_{\infty \to 1} \leq \langle R, \widehat{Z} \rangle \leq \langle R, Z_R \rangle. \tag{3.7}$$

Proof. The upper bound is trivial by definition of Z_R. The lower bound is based on Fact 3.2, which implies that for every $Z \in \mathcal{M}_{\mathrm{opt}}$, one has

$$|\langle B - R, Z \rangle| \leq K_{\mathrm{G}} \|B - R\|_{\infty \to 1} =: \varepsilon. \tag{3.8}$$

Now, to prove the lower bound in (3.7), we will first replace R by B using (3.8), then replace $\widehat{Z}$ by Z_R using the fact that $\widehat{Z}$ is a maximizer for $\langle B, Z \rangle$, and finally replace back B by R using (3.8) again. This way we obtain

$$\langle R, \widehat{Z} \rangle \geq \langle B, \widehat{Z} \rangle - \varepsilon \geq \langle B, Z_R \rangle - \varepsilon \geq \langle R, Z_R \rangle - 2\varepsilon.$$

This completes the proof of Lemma 3.3. □

4. Deviation in the cut norm

To be able to effectively use Lemma 3.3, we will now show how to bound the cut norm of random matrices.

Lemma 4.1 (Deviation in $\ell_\infty \to \ell_1$ norm). *Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a symmetric matrix whose diagonal entries equal 1, whose entries above the diagonal are independent random variables satisfying $0 \leq a_{ij} \leq 1$. Assume that*

$$\bar{p} := \frac{2}{n(n-1)} \sum_{i<j} \mathrm{Var}(a_{ij}) \geq \frac{9}{n}. \tag{4.1}$$

Then, with probability at least $1 - e^3 5^{-n}$, we have

$$\|A - \mathbb{E}\,A\|_{\infty \to 1} \leq 3\,\bar{p}^{1/2} n^{3/2}.$$

We will shortly deduce Lemma 4.1 from Bernstein's inequality followed by a union bound over $x, y \in \{-1, 1\}^n$; arguments of this type are standard in the analysis of random graphs (see e.g. [12, Section 2.3]). But before we do this, let us pause to explain the conclusion of Lemma 4.1.

Remark 4.2 (Regularization effect of $\ell_\infty \to \ell_1$ norm). Let us test Lemma 4.1 on the simple example where A is the adjacency matrix of a sparse Erdös-Renyi random graph $G(n, p)$ with $p = a/n$, $a \geq 1$. Here we have $\bar{p} = p(1-p) \leq p = a/n$. Lemma 4.1 states that $\|A - \mathbb{E}\,A\|_{\infty \to 1} \leq 3a^{1/2}n$. This can be compared with $\|\mathbb{E}\,A\|_{\infty \to 1} = (1 + p(n-1))n \geq an$. So we obtain

$$\|A - \mathbb{E}\,A\|_{\infty \to 1} \leq 3a^{-1/2} \|\mathbb{E}\,A\|_{\infty \to 1}.$$

This deviation inequality is good when a exceeds a sufficiently large absolute constant. Since that $a = pn$ is the expected average degree of the graph, it follows that we can handle graphs with *bounded expected degrees.*

This is a good place to note the importance of the $\ell_\infty \to \ell_1$ norm. Indeed, for the *spectral norm* a similar concentration inequality would fail. As is well known and easy to check, for $a = O(1)$ one would have $\|A - \mathbb{E}\,A\| \gg \|\mathbb{E}\,A\|$ due to contributions from high degree vertices. In fact, those are the only obstructions to concentration. Indeed, according to a result of U. Feige and E. Ofek [31], the removal of high-degree vertices forces a non-trivial concentration inequality to hold in the spectral norm. In contrast to this, the $\ell_\infty \to \ell_1$ norm does not feel the vertices with high degrees. It has an automatic *regularization effect*, which averages the contributions of all vertices, and in particular the few high degree vertices.

The proof of Lemma 4.1 will be based on Bernstein's inequality, which we quote here (see, for example, Theorem 1.2.6 in [21]).

Theorem 4.3 (Bernstein's inequality). *Let $Y_1, \ldots, Y_N$ be independent random variables such that $\mathbb{E}\,Y_k = 0$ and $|Y_k| \leq M$. Denote $\sigma^2 = \frac{1}{N} \sum_{k=1}^{N} \mathrm{Var}(Y_k)$.*

Then for any $t \geq 0$, one has

$$\mathbb{P}\left\{\frac{1}{N}\sum_{k=1}^{N}Y_k > t\right\} \leq \exp\left(-\frac{Nt^2/2}{\sigma^2 + Mt/3}\right).$$

Proof of Lemma 4.1. Recalling the definition (3.3) of the $\ell_\infty \to \ell_1$ norm, we see that we need to bound

$$\|A - \mathbb{E}\,A\|_{\infty\to 1} = \max_{x,y\in\{-1,1\}^n}\sum_{i,j=1}^{n}(a_{ij} - \mathbb{E}\,a_{ij})x_iy_j. \tag{4.2}$$

Let us fix $x, y \in \{-1, 1\}^n$. Using the symmetry of $A - \mathbb{E}\,A$, the fact that diagonal entries of $A - \mathbb{E}\,A$ vanish and collecting the identical terms, we can express the sum in (4.2) as a sum of independent random variables

$$\sum_{i<j}X_{ij}, \quad \text{where} \quad X_{ij} = 2(a_{ij} - \mathbb{E}\,a_{ij})x_iy_j.$$

To control the sum $\sum_{i<j}X_{ij}$ we can use Bernstein's inequality, Theorem 4.3. There are $N = \frac{n(n-1)}{2}$ terms in this sum. Since $|x_i| = |y_i| = 1$ for all i, the average variance σ^2 of all terms X_{ij} is at most 2^2 times the average variance of all a_{ij}, which is $\bar{p}$. In other words, $\sigma^2 \leq 4\bar{p}$. Furthermore, $|X_{ij}| \leq 2|a_{ij} - \mathbb{E}\,a_{ij}| \leq 2$ since $0 \leq a_{ij} \leq 1$ by assumption. Hence $M \leq 2$. It follows that

$$\mathbb{P}\left\{\frac{1}{N}\sum_{i<j}X_{ij} > t\right\} \leq \exp\left(-\frac{Nt^2/2}{4\bar{p} + 2t/3}\right). \tag{4.3}$$

Let us substitute $t = 6\,(\bar{p}/n)^{1/2}$ here. Rearranging the terms and using that $N = \frac{n(n-1)}{2}$ and $\bar{p} > 9/n$ (so that $t < 2\bar{p}$), we conclude that the probability in (4.3) is bounded by $\exp(-3(n-1))$.

Summarizing, we have proved that for every $x, y \in \{-1, 1\}^n$

$$\mathbb{P}\left\{\frac{2}{n(n-1)}\sum_{i,j=1}^{n}(a_{ij} - \mathbb{E}\,a_{ij})x_iy_j > 6\left(\frac{\bar{p}}{n}\right)^{1/2}\right\} \leq e^{-3(n-1)}.$$

Taking a union bound over all 2^{2n} pairs (x, y), we conclude that

$$\mathbb{P}\left\{\max_{x,y\in\{-1,1\}^n}\frac{2}{n(n-1)}\sum_{i,j=1}^{n}(a_{ij} - \mathbb{E}\,a_{ij})x_iy_j > 6\left(\frac{\bar{p}}{n}\right)^{1/2}\right\} \leq 2^{2n}\cdot e^{-3(n-1)}$$

$$\leq e^3\cdot 5^{-n}.$$

Rearranging the terms and using the definition (4.2) of the $\ell_\infty \to \ell_1$ norm, we conclude the proof of Lemma 4.1. $\qquad\qquad\square$

Remark 4.4 (The sum of entries). Note that by definition, the quantity $\left|\sum_{i,j=1}^{n}(a_{ij} - \mathbb{E}\,a_{ij})\right|$ is bounded by $\|A - \mathbb{E}\,A\|_{\infty\to 1}$, and thus it can be controlled by Lemma 4.1. Alternatively, a bound on this quantity follows

directly from the last line of the proof of Lemma 4.1. For a future reference, we express it in the following way:

$$\frac{2}{n(n-1)}\Big|\sum_{i<j}(a_{ij} - \mathbb{E}\,a_{ij})\Big| \le 3\,\bar{p}^{1/2}n^{-1/2}.$$

5. Stochastic block model: proof of Theorem 1.1

So far our discussion has been general, and the results could be applied to a variety of semidefinite programs on random graphs. In this section, we specialize to the community detection problem considered in Theorem 1.1. Thus we are going to analyze the optimization problem (1.4), where A is the adjacency matrix of a random graph distributed according to the classical stochastic block model $G(n, p, q)$.

As we already noticed, this is a particular case of the class of problems (3.5) that we analyzed in Section 3.2. In our case,

$$B := A - \lambda E_n$$

with λ defined in (1.5), and the feasible set is

$$\mathcal{M}_{\mathrm{opt}} := \mathcal{M}_{\mathrm{G}}^+ = \{Z : \ Z \succeq 0, \ \mathrm{diag}(Z) \preceq \mathbf{I}_n\}\,.$$

5.1. The maximizer of the reference objective function. In order to successfully apply Lemma 3.3, we will now choose a reference matrix R so that it is close to (but also conveniently simpler than) the expectation of B. To do so, we can assume without loss of generality that $\mathcal{C}_1 = \{1, \ldots, n/2\}$ and $\mathcal{C}_2 = \{n/2 + 1, \ldots, n\}$. Then we define R as a block matrix

$$R = \frac{p-q}{2}\begin{bmatrix} E_{n/2} & -E_{n/2} \\ -E_{n/2} & E_{n/2} \end{bmatrix} \tag{5.1}$$

where as usual $E_{n/2}$ denotes the $n/2 \times n/2$ matrix whose all entries equal 1.

Let us compute the expected value $\mathbb{E}\,B = \mathbb{E}\,A - (\mathbb{E}\,\lambda)E_n$ and compare it to R. To do so, note that the expected value of A has the form

$$\mathbb{E}\,A = \begin{bmatrix} pE_{n/2} & qE_{n/2} \\ qE_{n/2} & pE_{n/2} \end{bmatrix} + (1-p)I_n.$$

(The contribution of the identity matrix I_n is required here since the diagonal entries of A and thus of $\mathbb{E}\,A$ equal 1 due to the self-loops.) Furthermore, the definition of λ in (1.5) easily implies that

$$\mathbb{E}\,\lambda = \frac{1}{n(n-1)}\sum_{i\ne j}\mathbb{E}\,a_{ij} = \frac{p+q}{2}\frac{n^2}{n(n-1)} - \frac{p}{n-1} = \frac{p+q}{2} - \frac{p-q}{n-1}. \tag{5.2}$$

Thus

$$\mathbb{E}\,B = \mathbb{E}\,A - (\mathbb{E}\,\lambda)E_n = R + (1-p)\mathbf{I}_n - \frac{p-q}{n-1}E_n. \tag{5.3}$$

In the near future we will think of R as the leading term and other two terms as being negligible, so (5.3) intuitively states that $R \approx \mathbb{E}\,B$. We save this fact for later.

Using the simple form of R, we can easily determine the form of the solution Z_R of the reference problem (3.6).

Lemma 5.1 (The maximizer of the reference objective function). *We have*

$$Z_R := \arg\max_{Z \in \mathcal{M}_{\mathrm{opt}}} \langle R, Z \rangle = \begin{bmatrix} E_{n/2} & -E_{n/2} \\ -E_{n/2} & E_{n/2} \end{bmatrix}.$$

Proof. Let us first evaluate the maximizer of $\langle R, Z \rangle$ on the larger set $[-1, 1]^{n \times n}$, which contains the feasible set $\mathcal{M}_{\mathrm{opt}}$ according to (3.4). Taking into account the form of R in (5.1), one can quickly check that the maximizer of $\langle R, Z \rangle$ on $[-1, 1]^{n \times n}$ is Z_R. Since Z_R belongs to the smaller set $\mathcal{M}_{\mathrm{opt}}$, it must be the maximizer on that set as well. $\square$

5.2. **Bounding the error.** We are going to conclude from Lemma 4.1 and Lemma 3.3 that the maximizer of the actual objective function,

$$\widehat{Z} = \arg\max_{Z \in \mathcal{M}_{\mathrm{opt}}} \langle B, Z \rangle,$$

must be close to Z_R, the maximizer of the reference objective function.

Lemma 5.2 (Maximizers of random and reference functions are close). *Assume that $\bar{p}$ satisfies (4.1). Then, with probability at least $1 - e^3 5^{-n}$, we have*

$$\|\widehat{Z} - Z_R\|_2^2 \leq \frac{116\,\bar{p}^{1/2}n^{3/2}}{p - q}.$$

Proof. We expand

$$\|\widehat{Z} - Z_R\|_2^2 = \|\widehat{Z}\|_2^2 + \|Z_R\|_2^2 - 2\langle \widehat{Z}, Z_R \rangle \tag{5.4}$$

and control the three terms separately.

Note that $\|\widehat{Z}\|_2^2 \leq n^2$ since $\widehat{Z} \in \mathcal{M}_{\mathrm{opt}} \subset [-1, 1]^{n \times n}$ according to (3.4). Next, we have $\|Z_R\|_2^2 = n^2$ by Lemma 5.1. Thus

$$\|\widehat{Z}\|_2^2 \leq \|Z_R\|_2^2. \tag{5.5}$$

Finally, we use Lemma 3.3 to control the cross term in (5.4). To do this, notice that (5.1) and Lemma 5.1 imply that $R = \frac{p-q}{2} \cdot Z_R$. Then, by homogeneity, the conclusion of Lemma 3.3 implies that

$$\langle Z_R, \widehat{Z} \rangle \geq \langle Z_R, Z_R \rangle - \frac{4K_{\mathrm{G}}}{p - q}\|R - B\|_{\infty \to 1}. \tag{5.6}$$

To bound the norm of $R - B$, let us express this matrix as

$$B - R = (B - \mathbb{E}\,B) + (\mathbb{E}\,B - R) = (A - \mathbb{E}\,A) - (\lambda - \mathbb{E}\,\lambda)E_n + (\mathbb{E}\,B - R) \tag{5.7}$$

and bound each of the three terms separately. According to Lemma 4.1 and Remark 4.4, we obtain that with probability larger than $1 - e^3 5^{-n}$,

$$\|A - \mathbb{E}\,A\|_{\infty \to 1} \leq 3\bar{p}^{1/2}n^{3/2} \quad \text{and} \quad |\lambda - \mathbb{E}\,\lambda| \leq 3\bar{p}^{1/2}n^{-1/2}.$$

Moreover, according to (5.3),

$$\mathbb{E}\, B - R = (1-p)\mathbf{I}_n - \frac{p-q}{n-1}E_n.$$

Substituting these bounds into (5.7) and using triangle inequality along with the facts that $\|E_n\|_{\infty\to 1} = n^2$, $\|\mathbf{I}_n\|_{\infty\to 1} = n$, we obtain

$$\|B - R\|_{\infty\to 1} \le 6\bar{p}^{1/2}n^{3/2} + (1-p)n + \frac{(p-q)n^2}{n-1}.$$

Since $\bar{p} \ge 9/n$, one can check that each of the last two terms is bounded by $\bar{p}^{1/2}n^{3/2}$. Thus we obtain $\|B - R\|_{\infty\to 1} \le 8\bar{p}^{1/2}n^{3/2}$. Substituting into (5.6), we conclude that

$$\langle Z_R, \widehat{Z}\rangle \ge \langle Z_R, Z_R\rangle - 8\bar{p}^{1/2}n^{3/2} \cdot \frac{4K_{\mathrm{G}}}{p-q}.$$

Recalling from (3.1) that Grothendieck's constant K_{G} is bounded by 1.783, we can replace $8 \cdot 4K_{\mathrm{G}}$ by 58 in this bound. Substituting it and (5.5) and into (5.4), we conclude that

$$\|\widehat{Z} - Z_R\|_2^2 \le 2\|Z_R\|_2^2 - 2\langle \widehat{Z}, Z_R\rangle \le \frac{116\,\bar{p}^{1/2}n^{3/2}}{p-q}.$$

The proof of Lemma 5.2 is complete. $\qquad\square$

Proof of Theorem 1.1. The conclusion of the theorem will quickly follow from Lemma 5.2. Let us check the lemma's assumption (4.1) on $\bar{p}$. A quick computation yields

$$\bar{p} = \frac{2}{n(n-1)}\sum_{i<j}\mathrm{Var}(a_{ij}) = \frac{p(1-p)(n-2)}{2(n-1)} + \frac{q(1-q)n}{2(n-1)}. \qquad (5.8)$$

Since $p(1-p) \le 1/4$, we get

$$\bar{p} \ge \frac{1}{2}\max\{p(1-p), q(1-q)\} - \frac{1}{8(n-1)} > \frac{9}{n}$$

where the last inequality follows from an assumption of Theorem 1.1. Thus the assumption (4.1) holds, and we can apply Lemma 5.2. It states that

$$\|\widehat{Z} - Z_R\|_2^2 \le \frac{116\,\bar{p}^{1/2}n^{3/2}}{p-q} \qquad (5.9)$$

with probability at least $1 - e^3 5^{-n}$. From (5.8), it is not difficult to see that $\bar{p} \le \frac{p+q}{2}$. Substituting this into (5.9) and expressing $p = a/n$ and $q = b/n$, we conclude that

$$\|\widehat{Z} - Z_R\|_2^2 \le \frac{116\sqrt{(a+b)/2}}{a-b} \cdot n^2.$$

Rearranging the terms, we can see that this expression is bounded by εn^2 if

$$(a-b)^2 \ge 7 \cdot 10^3 \varepsilon^{-2}(a+b).$$

But this inequality follows from the assumption (1.6).

It remains to recall that according to Lemma 5.1, we have $Z_R = \bar{x}\bar{x}^\mathsf{T}$ where $\bar{x} = [\mathbf{1}_{n/2} \; -\mathbf{1}_{n/2}] \in \mathbb{R}^n$ is the community membership vector defined in (1.3). Theorem 1.1 is proved.   □

Proof of Corollary 1.2. The result follows from Davis-Kahan Theorem [28] about the stability of the eigenvectors under matrix perturbations. The largest eigenvalue of $\bar{x}\bar{x}^\mathsf{T}$ is n while all the others are 0, so the spectral gap equals n. Expressing $\widehat{Z} = (\widehat{Z} - \bar{x}\bar{x}^\mathsf{T}) + \bar{x}\bar{x}^\mathsf{T}$ and using that $\|\widehat{Z} - \bar{x}\bar{x}^\mathsf{T}\|_2 \leq \sqrt{\varepsilon}n$, we obtain from Davis-Kahan's theorem (see for example Corollary 3 in [67]) that

$$\|\widehat{v} - \bar{v}\|_2 = 2|\sin(\theta/2)| \leq C\sqrt{\varepsilon}.$$

Here $\hat{v}$ and $\bar{v}$ denote the unit-norm eigenvectors associated to the largest eigenvalues of $\widehat{Z}$ and $\bar{x}\bar{x}^\mathsf{T}$ respectively, and $\theta \in [0, \pi/2]$ is the angle between these two vectors. By definition, $\widehat{x} = \sqrt{n}\widehat{v}$ and $\bar{x} = \sqrt{n}\bar{v}$. This concludes the proof.   □

6. General stochastic block model: proof of Theorem 1.3

In this section we focus on the community detection problem for the general stochastic block-model considered in Theorem 1.3. The semidefinite program (1.10) is a particular case of the class of problems (3.5) that we analyzed in Section 3.2. In our case, we set $B := A$, choose the reference matrix to be

$$R := \bar{A} = \mathbb{E}\,A,$$

and consider the feasible set

$$\mathcal{M}_{\text{opt}} := \Big\{ Z \succeq 0, \; Z \geq 0, \; \text{diag}(Z) \preceq \mathbf{I}_n, \; \sum_{i,j=1}^n Z_{ij} = \lambda \Big\}.$$

Then $\mathcal{M}_{\text{opt}}$ is a subset of the Grothendieck's set $\mathcal{M}_G^+$ defined in (3.4). Using (3.4), we see that

$$\mathcal{M}_{\text{opt}} \subset \Big\{ Z : \; 0 \leq Z_{ij} \leq 1 \text{ for all } i, j; \; \sum_{i,j=1}^n Z_{ij} = \lambda \Big\}. \tag{6.1}$$

6.1. **The maximizer of the expected objective function.** Unlike before, the reference matrix $R = \bar{A} = \mathbb{E}\,A = (p_{ij})_{i,j=1}^n$ is not necessarily a block matrix like in (5.1) since the edge probabilities p_{ij} may be different for all $i < j$. However, we will observe that the solution Z_R of the reference problem (3.6) is a block matrix, and it is in fact the community membership matrix $\bar{Z}$ defined in (1.9).

Lemma 6.1 (The maximizer of the expected objective function)**.** *We have*

$$Z_R := \arg\max_{Z \in \mathcal{M}_{\text{opt}}} \langle \bar{A}, Z \rangle = \bar{Z}. \tag{6.2}$$

Proof. Let us first compute the maximizer on the larger set $\mathcal{M}'_{\text{opt}}$, which contains the feasible set $\mathcal{M}_{\text{opt}}$ according to (6.1). The maximum of the linear form $\langle \bar{A}, Z \rangle$ on the convex set $\mathcal{M}'_{\text{opt}}$ is attained at an extreme point. These extreme points are $0/1$ matrices with λ ones. Thus the maximizer of $\langle \bar{A}, Z \rangle$ has the ones at the locations of the λ largest entries of $\bar{A}$.

From the definition of the general stochastic block model we can recall that $\bar{A} = (p_{ij})$ has two types of entries. The entries larger than p form the community blocks $\mathcal{C}_k \times \mathcal{C}_k$, $k = 1, \ldots, K$. The number of such large entries is the same as the number of ones in the community membership matrix $\bar{Z}$, which in turn equals λ by the choice we made in Theorem 1.3. All other entries of $\bar{A}$ are smaller than q. Thus the λ largest entries of $\bar{A}$ form the community blocks $\mathcal{C}_k \times \mathcal{C}_k$, $k = 1, \ldots, K$.

Summarizing, we have shown that the maximizer of $\langle \bar{A}, Z \rangle$ on the set $\mathcal{M}'_{\text{opt}}$ is a $0/1$ matrix with ones forming the community blocks $\mathcal{C}_k \times \mathcal{C}_k$, $k = 1, \ldots, K$. Thus the maximizer is the community membership matrix $\bar{Z}$ from (1.9). Since $\bar{Z}$ belongs to the smaller set $\mathcal{M}_{\text{opt}}$, it must be the maximizer on that set as well. $\qquad\square$

6.2. **Bounding the error.** We are going to conclude from Lemma 4.1 and Lemma 3.3 that the maximizer of the actual objective function,

$$\widehat{Z} = \arg \max_{Z \in \mathcal{M}_{\text{opt}}} \langle A, Z \rangle,$$

must be close to $\bar{Z}$, the maximizer of the reference objective function. We will first show that *the reference objective function $\langle \bar{A}, Z \rangle$ distinguishes points near its maximizer $\bar{Z}$*.

Lemma 6.2 (Expected objective function distinguishes points)**.** *Every $Z \in \mathcal{M}_{\text{opt}}$ satisfies*

$$\langle \bar{A}, \bar{Z} - Z \rangle \geq \frac{p - q}{2} \|\bar{Z} - Z\|_1. \tag{6.3}$$

Proof. We will prove that the conclusion holds for every Z in the larger set $\mathcal{M}'_{\text{opt}}$, which contains the feasible set $\mathcal{M}_{\text{opt}}$ according to (6.1). Expanding the inner product, we can represent it as

$$\langle \bar{A}, \bar{Z} - Z \rangle = \sum_{i,j=1}^{n} p_{ij}(\bar{Z} - Z)_{ij} = \sum_{(i,j) \in \text{In}} p_{ij}(\bar{Z} - Z)_{ij} - \sum_{(i,j) \in \text{Out}} p_{ij}(Z - \bar{Z})_{ij}$$

where In and Out denote the set of edges that run within and across the communities, respectively. Formally, $\text{In} = \cup_{k=1}^{K}(\mathcal{C}_k \times \mathcal{C}_k)$ and $\text{Out} = \{1, \ldots, n\}^2 \setminus \text{In}$.

For the edges $(i, j) \in \text{In}$, we have $p_{ij} \geq p$ and $(\bar{Z} - Z)_{ij} \geq 0$ since $\bar{Z}_{ij} = 1$ and $Z_{ij} \leq 1$. Similarly, for the edges $(i, j) \in \text{Out}$, we have $p_{ij} \leq q$ and $(Z - \bar{Z})_{ij} \geq 0$ since $\bar{Z}_{ij} = 0$ and $Z_{ij} \geq 0$. It follows that

$$\langle \bar{A}, \bar{Z} - Z \rangle \geq p S_{\text{In}} - q S_{\text{Out}} \tag{6.4}$$

where

$$S_{\text{In}} = \sum_{(i,j) \in \text{In}} (\bar{Z} - Z)_{ij} \quad \text{and} \quad S_{\text{Out}} = \sum_{(i,j) \in \text{Out}} (Z - \bar{Z})_{ij}.$$

Since both $\bar{Z}$ and Z belong to $\mathcal{M}_{\text{opt}}$, the sum of all entries of both these matrices is the same $(n^2/2)$, so we have

$$S_{\text{In}} - S_{\text{Out}} = \sum_{i,j=1}^{n} \bar{Z}_{ij} - \sum_{i,j=1}^{n} Z_{ij} = 0. \tag{6.5}$$

On the other hand, as we already noticed, the terms in the sums that make S_{In} and S_{Out} are all non-negative. Therefore

$$S_{\text{In}} + S_{\text{Out}} = \sum_{i,j=1}^{n} |(\bar{Z} - Z)_{ij}| = \|\bar{Z} - Z\|_1. \tag{6.6}$$

Substituting (6.5) and (6.6) into (6.4), we obtain the conclusion (6.3). $\square$

Now we are ready to conclude that $\widehat{Z} \approx \bar{Z}$.

Lemma 6.3 (Maximizers of random and expected functions are close)**.** *Assume that $\bar{p}$ satisfies* (4.1)*. With probability at least $1 - e^3 5^{-n}$, we have*

$$\|\widehat{Z} - \bar{Z}\|_1 \leq \frac{12\, K_{\text{G}}\, p_0^{1/2} n^{3/2}}{p - q}.$$

Proof. Using first Lemma 6.2, Lemma 3.3 (with $R = \bar{A}$ and $Z_R = \bar{Z}$ as before) and then Lemma 4.1, we obtain

$$\|\widehat{Z} - \bar{Z}\|_1 \leq \frac{2}{p-q} \langle \bar{A}, \bar{Z} - \widehat{Z} \rangle \leq \frac{4K_{\text{G}}}{p-q} \|A - \bar{A}\|_{\infty \to 1} \leq \frac{12K_{\text{G}}}{p-q} \bar{p}^{1/2} n^{3/2}$$

with probability at least $1 - e^3 5^{-n}$. Lemma 6.3 is proved. $\square$

Proof of Theorem 1.3. The conclusion follows from Lemma 6.3. Indeed, substituting $p = a/n$, $q = b/n$ and $\bar{p} = g/n$ and rearranging the terms, we obtain

$$\|\widehat{Z} - \bar{Z}\|_1 \leq \frac{12K_{\text{G}} g^{1/2}}{a - b} \cdot n^2 \leq \frac{22 g^{1/2}}{a - b} \cdot n^2$$

since we know form (3.1) that Grothendieck's constant K_{G} is bounded by 1.783. Rearranging the terms, we can see that this expression is bounded by εn^2 if $(a - b)^2 \geq 484\, \varepsilon^{-2} g$, which is our assumption (1.12). This proves the required bound for the $\| \cdot \|_1$ norm.

Since for any sequence $\sum |b_{i,j}|^2 \leq \max |b_{i,j}| \sum |b_{i,j}|$, we get

$$\|\widehat{Z} - \bar{Z}\|_2^2 \leq \|\widehat{Z} - \bar{Z}\|_\infty \cdot \|\widehat{Z} - \bar{Z}\|_1.$$

As we noted in (6.1), all entries of $\widehat{Z}$ and $\bar{Z}$ belong to $[0, 1]$ hence $\|\widehat{Z} - \bar{Z}\|_\infty \leq 1$. The bound for the Frobenius norm follows and Theorem 1.3 is proved. $\square$

References

[1] E. Abbe, C. Sandon, *Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms*, Arxiv: 1503.00609.

[2] N. Ailon, Y. Chen, H. Xu, *Breaking the small cluster barrier of graph clustering*, Journal of Machine Learning Research, to appear, 2014.

[3] E. Airoldi, D. Blei, S. Fienberg, E. Xing, *Mixed membership stochastic blockmodels*, J. Machine Learning Research 9 (2008), 1981–2014.

[4] N. Alon, *Spectral techniques in graph algorithms*, LATIN'98: theoretical informatics (Campinas, 1998), 206–215, Lecture Notes in Comput. Sci., 1380, Springer, Berlin, 1998.

[5] N. Alon, N. Kahale, *A spectral technique for coloring random 3-colorable graphs*, SIAM J. Comput. 26 (1997), 1733–1748.

[6] N. Alon, A. Naor, *Approximating the cut-norm via Grothendieck's inequality*, SIAM J. Comput. 35 (2006), 787–803.

[7] B. Ames, S. Vavasis, *Nuclear norm minimization for the planted clique and biclique problems*, Mathematical Programming 129 (2011), 69–89.

[8] A. Amini, A. Chen, P. Bickel, E. Levina, *Pseudo-likelihood methods for community detection in large sparse networks*, Annals of Statistics 41 (2013), 2097–2122.

[9] A. Amini, E. Levina, *On semidefinite relaxations of the block model*, 2014. Arxiv: 1406.5647.

[10] P. Bickel, A. Chen, *A nonparametric view of network models and Newman-Girvan and other modularities*, Proc. Natl. Acad. Sci. USA 106 (2009), 21068–21073.

[11] P. Bickel, D. Choi, X. Chang, H. Zhang, *Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels*, Annals of Statistics 41 (2013), 1922–1943.

[12] B. Bollobas, *Random graphs*, Second edition. Cambridge Studies in Advanced Mathematics, 73. Cambridge University Press, Cambridge, 2001. xviii+498 pp.

[13] B. Bollobas, S. Janson, O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures and Algorithms 31 (2007), 3–122.

[14] R. Boppana, *Eigenvalues and graph bisection: an average-case analysis*, Proc. 28th Annual Symposium on Foundations of Computer Science 1987, pp. 280–285.

[15] C. Bordenave, A. Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices*, Probab. Theory Related Fields 157 (2013), 885–953.

[16] C. Bordenave, M. Lelarge, L. Massoulié, *Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs*, ArXiv: 1501.06087.

[17] M. Braverman, K. Makarychev, Y. Makarychev, A. Naor, *The Grothendieck constant is strictly smaller than Krivine's bound.* 2011 IEEE 52nd Annual Symposium on Foundations of Computer ScienceFOCS 2011, 453–462, IEEE Computer Soc., Los Alamitos, CA, 2011.

[18] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser, *Graph bisection algorithms with good average case behavior*, Combinatorica, 7(1987), 171–191.

[19] T. Cai, X. Li, *Robust and computationally feasible community detection in the presence of arbitrary outlier vertices*, 2014. ArXiv: 1404.6000.

[20] A. Celisse, J.-J. Daudin, L. Pierre, *Consistency of maximum-likelihood and variational estimators in the stochastic block model*, Electronic Journal of Statistics 6 (2012),1847–1899.

[21] D. Chafaï, O. Guédon, G. Lecué, A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry.* Panoramas et Synthèses [Panoramas and Syntheses], 37. Société Mathématique de France, Paris, 2012. 181 pp.

[22] K. Chaudhuri, F. Chung, A. Tsiatas, *Spectral clustering of graphs with general degrees in the extended planted partition model*, JMLR Workshop and Conference Proceedings, 23:35.1 – 35.23, 2012.

[23] Y. Chen, S. Sanghavi, H. Xu, *Clustering Sparse Graphs*, NIPS, 2012. Arxiv: 1210.3335

[24] Y. Chen, A. Jalali, S. Sanghavi, H. Xu, *Clustering partially observed graphs via convex optimization*, Journal of Machine Learning Research, to appear (2014). Arxiv: 1104.4803.

[25] Y. Chen, J. Xu, *Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices*, 2014. ArXiv: 1402.1267.

[26] P. Chin, A. Rao, V. Vu, *Stochastic block model and community detection in the sparse graphs: a spectral algorithm with optimal rate of recovery*, 2015. ArXiv: 1501.05021.

[27] A. Coja-Oghlan, *Graph partitioning via adaptive spectral techniques*, Combinatorics, Probability and Computing 19 (2010), 227–284.

[28] C. Davis, W. M. Kahan. *The rotation of eigenvectors by a perturbation. III.* SIAM J. Numer. Anal. 7 1970 1–46.

[29] A. Decelle, F. Krzakala, C. Moore, L. Zdeborova, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*, Physical Review E, 84:066106, 2012.

[30] M. E. Dyer and A. M. Frieze, *The solution of some random NP-hard problems in polynomial expected time*, Journal of Algorithms, 10 (1989), 451–489.

[31] U. Feige, E. Ofek, *Spectral techniques applied to sparse random graphs*, Random Structures and Algorithms 27 (2005), 251–275.

[32] J. Friedman, J. Kahn, E. Szemeredi, *On the second eigenvalue in random regular graphs*, *Proc. 21 Annual ACM Symp. of Theory of Computing*, 1989, pp. 587–598.

[33] C. Gao, Z. Ma, A. Zhang, H. Zhou, *Achieving optimal misclassification proportion in stochastic block model*, ArXiv: 1505.03772.

[34] M. Goemans, D. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach. 42 (1995), 1115–1145.

[35] A. Grothendieck, *Résumé de la théorie métrique des produits tensoriels topologiques*, Bol. Soc. Mat. São Paulo 8 (1953), 1–79.

[36] O. Guédon, R. Vershynin, *Community detection in sparse networks via Grothendieck's inequality* ArXiv:1411.4686.

[37] A. Jalali, Y. Chen, S. Sanghavi, H. Xu, *Clustering partially observed graphs via convex optimization*, ICML, 2011. ArXiv: 1104.4803.

[38] A. Joseph, B. Yu, *Impact of regularization on spectral clustering*, 2014. Arxiv: 1312.1733.

[39] L. Hagen, A. Kahng, *New spectral methods for ratio cut partitioning and clustering*, IEEE Trans. Computer-aided design of integrated circuits and systems, 11 (1992), 1074–1085.

[40] P. W. Holland, K. B. Laskey, S. Leinhardt, *Stochastic blockmodels: first steps*, Social Networks, 5 (1983), 109–137.

[41] S. Khot, A. Naor, *Grothendieck-type inequalities in combinatorial optimization*, Comm. Pure Appl. Math. 65 (2012), 992–1035.

[42] M. Ledoux, M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*, Reprint of the 1991 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2011. xii+480 pp.

[43] J. Lei, A. Rinaldo, *Consistency of spectral clustering in sparse stochastic block models*, 2013. Arxiv: 1312.2050.

[44] M. Lelarge, L. Massoulié, J. Xu, *Reconstruction in the labeled stochastic block model*, Arxiv: 1502.03365. Preliminary version in Proceedings of the Information Theory Workshop, 2013.

[45] J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney, *Statistical properties of community structure in large social and information networks*, In Proceedings of the 17th international conference on World Wide Web, pages 695–704. ACM, 2008.

[46] E. Levina, C. Le, R. Vershynin, *Sparse random graphs: regularization and concentration of the Laplacian*, Arxiv: 1502.03049.

[47] C. Le, R. Vershynin, *Concentration and regularization of random graphs*, Arxiv: 1506.00669.

[48] J. Lindenstrauss, A. Pelczyński, *Absolutely summing operators in L_p-spaces and their applications*, Studia Math. 29 (1968), 275–326.

[49] L. Lovász, A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization*, SIAM J. Optim. 1 (1991), 166–190.

[50] F. McSherry, *Spectral partitioning of random graphs*, Proc. 42nd Annual Symposium on Foundations of Computer Science, 2001, pp. 529–537.

[51] L. Massoulié, *Community detection thresholds and the weak Ramanujan property*, 2013. ArXiv:1311.3085.

[52] A. Montanari, S. Sen, *Semidefinite programs on sparse random graphs*, ArXiv:1504.05910.

[53] E. Mossel, J. Neeman, A. Sly, *Belief propagation, robust reconstruction, and optimal recovery of block models*, 2013. ArXiv:1309.1380.

[54] E. Mossel, J. Neeman, A. Sly, *Stochastic Block Models and Reconstruction*, Probability Theory and Related Fields, to appear, 2014.

[55] E. Mossel, J. Neeman, A. Sly, *A proof of the block model threshold conjecture*, 2014. ArXiv:1311.4115.

[56] E. Mossel, J. Neeman, A. Sly, *Consistency Thresholds for the Planted Bisection Model*, 2014. ArXiv: 1407.1591.

[57] R. Nadakuditi, M. Newman, *Graph spectra and the detectability of community structure in networks* Phys. Rev. Lett. 108 (2012), 188701.

[58] Y. Nesterov, *Semidefinite relaxation and nonconvex quadratic optimization*, Optim. Methods Softw. 9 (1998), 141–160.

[59] M. Newman, *Modularity and community structure in networks*, Proc. Natl. Acad. Sci. USA 103 (2006), 8577–8582.

[60] K. Nowicki, T. Snijders, *Estimation and prediction for stochastic blockstructures*, Journal of the American Statistical Association 96 (2001), 1077–1087.

[61] S. Oymak, B. Hassibi, *Finding dense clusters via low rank + sparse decomposition*, CoRR, 2011. arXiv:1104.5186

[62] G. Pisier. *Grothendieck's theorem, past and present.* Bull. Amer. Math. Soc. (N.S.) 49 (2012), no. 2, 237–323.

[63] T. Qin, K. Rohe, *Regularized spectral clustering under the degree-corrected stochastic blockmodel*, Arxiv: 1309.4111

[64] K. Rohe, S. Chatterjee, B. Yu, *Spectral clustering and the high-dimensional stochastic block model*, Annals of Statistics, 39 (2011), 1878–1915.

[65] T. Snijders, K. Nowicki, *Estimation and prediction for stochastic block-structures for graphs with latent block structure*, Journal of Classification14 (1997), 75–100.

[66] S. Strogatz, *Exploring complex networks*, Nature 410 (2001), 268–276.

[67] V. Vu, *Singular vectors under random perturbation.* Random Structures Algorithms 39 (2011), 526–538.

Université Paris-Est, Laboratoire d'Analyse et Mathématiques Appliquées (UMR 8050), UPEMLV, F-77454, Marne-la-Vallée Cedex 2, France

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.