Spectral analysis in bipartite biregular graphs and community detection

Gerandy Brito

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Ioana Dumitriu, Chair

Christopher Hoffman, Chair

Soumik Pal

Program Authorized to Offer Degree:
Department of Mathematics

University of Washington

Abstract

Spectral analysis in bipartite biregular graphs and community detection

Gerandy Brito

Co-Chairs of the Supervisory Committee:
Professor Ioana Dumitriu
Department of Mathematics

Professor Christopher Hoffman
Department of Mathematics

This thesis concerns to spectral gap of random regular graphs and consists of two main contributions. First, we prove that almost all bipartite biregular graphs are almost Ramanujan by providing a tight upper bound for the non trivial eigenvalues of its adjacency operator, proving Alon's Conjecture for this family of graphs. Secondly, we use a spectral algorithm to recover hidden communities in a random network model we call regular stochastic block model. We rely on a technique introduced recently by Massoullié, which we develop here for random regular graphs.

ACKNOWLEDGMENTS

I have so many people to thank that seems hard to cover the list in a single page. The act should be more personal than this and I hope I can thank each of you in person, sooner than later. I am deeply grateful to Ioana and Chris for been not only math advisors but also friends and colleagues. Special thanks to Soumik Pal who has been like a third advisor to me. I owe a debt to many other people in the Department of Mathematics at UW: Sthepen Rhode, Tatiana Toro, Sara Billey and Judith Arms have play an important role in my formation. I appreciated Yuval Peres' hospitality at Microsoft Research. I am thankful that my time in Seattle I shared with great mathematicians and friends: Shirshendu Ganguly and Jose Samper, can't thank them enough. Thanks to all my friends in Santa Clara, in La Habana, and in Seattle. *A mis tres Lucias: Blanca, Nancy y Geraidy.* Most of all, I thank Monica for having been with me all these years.

Contents

A Bilo e Isel in memoriam

Chapter 1

Introduction

Graphs have become one of the most important objects in human knowledge, as their presence in a variety of fields confirms. They model relations among individuals in the Social Sciences, represent networks of communications, data structure and flow of computation in Computer Science, and have found applications in many areas of Mathematics, as diverse as Group Theory, Geometry, Topology, Combinatorics, Probability and Number Theory.

Because of there ubiquity, scientists with different backgrounds and interests have devoted efforts to understand graphs from different points of view, leading to deep knowledge of structural properties of families of graphs. The accumulation of data in the modern world and algorithms need to be more powerful to be useful, we started to look for alternatives that rely, in many cases, on randomized techniques. To put it in simple words: if we need to study a certain property in a particular graph, we consider a large set of graphs, which contains our initial one and can be described by some general conditions, and look for what can we say about a *typical* graph in this set. If *most* graphs share certain properties, a typical one will have it too. This notion of *universality* is the language in which many of the most fundamental question in graph theory are formulated.

Expander graphs deserve to be mentioned, for their wide range of applications. These are highly connected graphs, with many properties that have an impact in many day to day applications and they are also the main object in many beautiful results in mathematics. We refer to the survey of Hoory, Linial and Wigderson ([3]) for a comprehensive study of this family of graphs. This thesis studies expander graphs from both a theoretical point of view, by proving results regarding the spectral gap of certain family of graphs, and using their spectral behavior to attack a problem of community detection.

It is worthy to introduce briefly the spectrum of a graph. Given $G = (V, \mathcal{E})$, a graph with vertex set V and edge set $E \subset V \times V$, the adjacency operator $A = A(G)$, commonly known as the adjacency matrix, is the linear application on $\mathbb{R}^{|V|}$ defined as

$$A_{uv} = \begin{cases} 1, & \text{if } uv \in \mathcal{E}; \\ 0, & else. \end{cases}$$

Its spectrum, $\sigma(A) = \{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|V|}\}$, will be referred as the set of eigenvalues of G.

There is a close connection between graph eigenvalues and the structure of G, ([1, 2]). This connection have been used extensively leading to beautiful results and applications in many areas of the human knowledge.

To motivate our discussion, we will focus on a particular family of graphs. A regular $d-$graph with n vertices is a graph for which all vertices have degree d. The family of all $d-$regular graphs on n vertices is denoted by $\mathcal{G}(n, d)$. The Perron-Frobenius Theorem gives us that $\lambda_1 = d$ for any element in $\mathcal{G}(n, d)$. The spectral gap of G is defined as $d - \lambda_2$. There are a number of reasons to study this quantity. First, the spectral gap tells us if the graph G is an *expander*. Intuitively, expander graphs are graphs with good connectivity properties and have numeruous applications in a very diverse list of fields of mathematics, (see [3]). Secondly, and very much related to expanders, the magnitude of λ_2 is the central concept used to define *Ramanujan graphs*, another important family of expanders, which are regular graphs with surprising applications in number theory ([4]). Given the importance of these properties of graphs, there are fundamental questions to ask: Are expanders or Ramanujan graphs *common*? How far is a typical graph from being Ramanujan? What can we say about expanders, outside regular graphs? Motivated by these questions, our purpose with this thesis is to contribute to the understanding of one special family of graphs, random bipartite biregular graphs, by studying their spectrum.

1.1 Alon's conjecture beyond regular graphs.

We give some precise definitions of the properties mentioned in the previous section to extend our motivations.

Definition 1.1. A $d-$regular graph is said to be Ramanujan if

$$\lambda_2 \leq 2\sqrt{d-1}.$$

The constant on the right hand side cannot be improved ([5, 6]). Thus, Ramanujan graphs are the optimal expanders in the sense that they achieve the largest spectral gap. An easy example of a Ramanujan graph is the complete graph.

The following question turns out to be difficult:

Exhibit a family of $d-$regular graphs $\{G_i\}$ with $V(G_i) \to \infty$ that are Ramanujan.

The only known constructions work for values of d equal to an odd prime plus one or a power of a prime plus one, see [7–9]. However, it was conjectured by Noga Alon ([5]) that most regular graphs are *almost* Ramanujan, in the sense that $\lambda_2 \leq 2\sqrt{d-1} + o(1)$. The conjecture was not precise regarding what it means 'most' regular graphs. Friedman ([10]) was able to prove the following

Theorem 1.2. *([10, Theorem 1.1]) For any fixed $\epsilon > 0$, and G Unif($\mathcal{G}(n,d)$) we have*

$$\lambda_2 \leq 2\sqrt{d-1} + \epsilon$$

asymptotically almost sure, i.e. with probability tending to one as $n \to \infty$.

Recently, Bordenave ([11]) was able to prove a slightly stronger result with $\epsilon = \epsilon(n) \to 0$. One possible way to extend Alon's conjecture to a wider class of graphs is to consider the universal cover of a graph. Let us briefly define this notion. Let G be a graph. A cover of G is a pair (H, π) where H is a graph and π is a graph homomorphism $\pi : H \to G$ such that, for each vertex $v \in V(H)$, the restriction of ρ to the neighborhoods of v and $\pi(v)$ is an isomorphism. The universal cover of G is the unique cover (T_G, π) such that T_G is a tree. A concrete example is the universal cover of a $d-$regular graph, which is the $d-$regular tree, $\mathbb{T}_d$. We can define the adjacency operator of a locally finite, infinite graph T, $A_T : \ell^2(V(T)) \to \ell^2(V(T))$, by the relation:

$$(A_T f)(u) = \sum_{vu \in \mathcal{E}(T)} f(v).$$

Then A_T is a self-adjoint operator and we can talk of its spectral radius $\rho(T)$. In the case of $\mathbb{T}_d$, it was proved in [12] that $\rho(\mathbb{T}_d) = 2\sqrt{d-1}$. Thus, Ramanujan graphs are optimal in the sense that their non trivial spectrum is as good as the spectrum of its universal cover. In [13], Alon's conjecture is extended following the ideas above. To do so, one most first consider random covers of a given graph, not necessarily regular, and study their spectrum. These random covers are called *lifts* in the literature, see [14] for a nice exposition on this subject. Friedman conjectured that, for any $\epsilon > 0$, the non trivial eigenvalues λ of a random lift satisfied $\lambda \leq \rho(T_G) + \epsilon$.

We will introduce lifts of graph formally in Chapter 4.

1.2 Main results of this thesis.

1.2.1 Spectral gap in bipartite biregular graphs.

A natural family to extend the results for regular graphs will be to consider bipartite biregular graphs. We will denote by $\mathcal{G}(n, m, d_1, d_2)$ the family of graphs such that their vertex set can be partitioned into two disjoint sets V_1 and V_2, with $|V_1| = n$ and $|V_2| = m$, such that each $v \in V_i$ has degree d_i, $i = 1, 2$ and its edge set is a subset of $V_1 \times V_2$. For any $G \in \mathcal{G}(n, m, d_1, d_2)$, we can write its adjacency matrix

$$A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$$

where X is an $n \times m$ matrix of 0s and 1s defined by the edges between V_1 and V_2. It is an standard linear algebra result that the non-zero eigenvalues of A come in pairs $(-\lambda, \lambda)$ with λ^2 an eigenvalue of XX^*, and that A has at least, $(n - m)$ eigenvalues equal to zero, assuming $n \geq m$. Furthermore, the Perron-Frobenius eigenvalue $\lambda_1 = \sqrt{d_1 d_2}$. When the degrees are constant, Mohar [15] proved that, for any $\epsilon > 0$, a constant fraction of the eigenvalues of any element of $\mathcal{G}(n, m, d_1, d_2)$ are bounded below by $\sqrt{d_1 - 1} + \sqrt{d_2 - 1} - \epsilon$. The constant $\sqrt{d_1 - 1} + \sqrt{d_2 - 1}$ is the right most point of the support of the limiting expected spectral distribution of a bipartite biregular graph chosen uniformly from $\mathcal{G}(n, m, d_1, d_2)$ ([16]), and it is expected to be the right order for the second largest eigenvalue. We confirm this by proving that, with high probability, the second eigenvalue of a bipartite biregular graph chosen uniformly from the set $\mathcal{G}(n, m, d_1, d_2)$, satisfies:

$$\lambda_2 \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon_n$$

where $\epsilon_n \to 0$ as $n \to \infty$. This is done in Chapter 3, where a precise statement of the result can be found.

1.2.2 Community detection in sparse regular networks.

An important source of information about the structure of a graph comes from the eigenvectors associated to certain eigenvalues. A simple example is the eigenvector associated to the leading eigenvalue of a bipartite biregular graph, whos restriction to V_i results in a vector with all entries equal. A slightly more sophisticated example comes from lifts of a graph. In this case, the spectrum of the base graph is a subset of the spectrum of the lifts, and the new corresponding eigenvectors have a block like structure (see Chapter 4 for a more detailed explanation). This leads to a popular application of spectral methods: community detection. The situation is as follows: we are given data of various types and relations among these data. We encode this in a graph, who vertex set is the collection of our data points and the edge set represents *interactions* between these data points. Typically, the nature of interactions depends on the type of data. The task is to cluster the data such that we *recover* the communities in the best possible way, and do so efficiently from the computational point of view. Most of the mathematical models studied assume that for any pair of vertices in our graph, the corresponding edge occurs with probability depending only on the type of its ends, independently of other edges. The most popular of such models is the so called Stochastic Block Model (SBM), which has been extensively studied in the last couple of decades. For a recent survey on the subject see [17] and the reference therein. Beyond the SBM, we encounter in the literature many models that aim to cover a variety of situations that arise in real world data. We introduce a model in which the number of interactions is the same for member in the same community, resulting in a regular graph. For our models, we show how spectral knowledge of certain operators associated with the graph helps us to cluster efficiently the vertex set. We defer to Chapter 4 for the rest of the discussion.

This thesis is based on joint work, some unpublished. The results in Chapter 3 were obtain together with Ioana Dumitriu and Kameron Harris. The content of Chapter 4 is from [18] and [19], which was written in collaboration with Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman and Linh V. Tran.

Chapter 2

Background on random regular graphs

This chapter contains the necessary background on random regular graphs to prove our main results. It is structured as follows: first, we introduce the models of random graphs we are going to use. Then we define the notion of *tangle–free* graphs, a key property for our computations. Lastly, we find estimates on the probability of the event that a random regular graph contains a particular subgraph. Our proof relies on the resutls by Mckay [24].

2.1 Models of random regular graphs.

We will work mainly with two probabilistic models.

The uniform model. Given integers n and d consider the family of all simple graphs on n vertices such that each vertex has degree d. Observe that it is necessary for nd to be even. The uniform model picks one such graphs with equal probability.

The configuration model. This model, introduced by Bender and Canfield [20] and made famous by Bollobas [21], is a well known model to study random regular graphs. Assuming that dn is even, the configuration model outputs a d-regular multigraph with n vertices. This is done by considering an array $\{\xi_{ij}, 1 \leq i \leq d,\ \ 1 \leq j \leq n\}$ and choosing a perfect matching for it, uniformly among all possible matchings. A graph on n vertices is obtained by collapsing all ξ_{ij} for $1 \leq i \leq d$ into a single vertex, and putting an edge between two vertices j and ℓ for each pair (ξ_{ij}, ξ_{kl}) present in the matching. We refer to the family ξ_{ij} as *half edges.*

It is not hard to see that under the condition that the configuration model outputs a simple graph, the corresponding distribution of the graph is uniform in the set of all simple $d-$regular graphs. Furthermore, it is well known that, for any fixed d, as n grows to infinity, the probability that a graph obtained by the configuration model is simple is bounded away from zero. More precisely, denoting by G the resulting graph, one has (see [21]),

$$\mathbb{P}(G \text{ is simple}) = (1 - o(1))e^{\frac{1-d^2}{4}}.$$

Thus, these models are asymptotically *contiguous*. This is, if a sequence of events holds asymptotically almost surely for the uniform measure on simple $d-$regular graphs, it holds asymptotically almost surely for the measure induced on multigraphs by the configuration model as well.

One extremely useful property of the configuration model is the fact that one can construct the graph by exposing the vertices one at a time, each time matching one by one the d half edges of the correspondent vertex, to a uniformly chosen half edge among the set of unmatched half edges. This process will be used crucially in many of our estimates. We give the precise definition.

Definition 2.1. Consider the following procedure to generate a random d-regular graph on n vertices:

- Fix an order of the vertices: $v_1 < v_2 < ... < v_n$ and let $\Xi = \{\xi_{ij}\}, 1 \leq i \leq d$ and $1 \leq j \leq n$, be the set of half edges, where, for any $1 \leq j \leq n$, ξ_{ij} are the d half edges incident to vertex v_j. Consider the usual lexicographic order on Ξ.
- Construct a perfect matching of Ξ as follows: the first pair is $(\xi_{11}, \hat{\xi})$ where $\hat{\xi}$ is chosen uniformly from $\Xi \setminus \{\xi_{11}\}$. Having constructed k pairs, let ξ_{ij} be the smallest half edge not matched yet, chose $\tilde{\xi}$ uniformly from the set of remaining unmatched half edges different from ξ_{ij}, and add the edge $(\xi_{ij}, \tilde{\xi})$.
- Output a multigraph G, with vertex set $\{v_j\}$ and an edge set induced by the matching constructed in the previous step.

This construction outputs a graph with the same law as the one given by the configuration model. Conveniently, with this construction we discover all neighbors of vertex v_1 first, then we move to v_2 and expose its neighbors (it could be the case that some edges are connecting v_1 and v_2 and those were exposed before!) and so on. We will refer to this procedure as the *exploration process*. All the above definitions can be easily adapted to sample bipartite regular graphs as well, and throughout this thesis we will use both sets of definitions.

2.2 Tangle–free graphs.

When using the trace method, it will be necessary to count the number of closed walks, or cycles, starting at a typical vertex. If we account for all possible cycles, the bounds are not good enough. Luckily, most neighborhoods of a certain radius are simple, in the sense that they are trees or almost a tree. This behavior is known as *tangle–free* and has been used extensively in the literature, see [22] and [23]. To introduce this notion, we need some definitions.

For a vertex v in $\mathcal{G}(n, d_1, d_2)$, and for $t \in \mathbb{N}$ let the ball of size t centered at v be denoted by

$$B_t(v) = \{u \in \mathcal{G}(n, d_1, d_2) : d(u, v) \leq t\} ,$$

where $d(u, v)$ is the graph distance between vertices u and v. We define the boundary of $B_t(v)$ by

$$\partial B_t(v) = \{u \in \mathcal{G}(n, d_1, d_2) : d(u, v) = t\} .$$

Definition 2.2. A graph G is said to be ℓ*–tangle–free* if for any vertex v in G the ball $B_\ell(v)$ contains at most one cycle.

The next lemma establishes two important properties of random regular graphs: with high probability, they are ℓ*-tangle–free* for $\ell \leq c\log(n)$ for some universal constant c, and the number of cycles of length less than ℓ is small.

Lemma 2.3. *Let $G \sim \mathcal{G}(n, d)$ and $\ell = c\log n$ such that $\delta := c\log(d) < \frac{1}{4}$, and let $0 < \epsilon < 1 - 4\delta$ be a small constant. Then*

(a) G is ℓ–tangle–free with probability $1 - O(n^{-\epsilon})$.

(b) Denote by $X^{(\ell)} = \#\{v \in V(G) : B_\ell(v)$ contains a cycle$\}$. Assuming that G is ℓ–tangle–free,

$$\mathbb{P}(X^{(\ell)} > n^\delta) < O(n^{-\delta}) .$$

Proof. The first part of the lemma already appears as [23, Lemma 2.1]. To prove the second part we use the following standard variant of the exploration process mentioned in Definition 2.1. (This variant is also used in the proof of [23, Lemma 2.1]) . Choose a vertex v of G and fix some ordering among all other vertices. Consider the process that (in accordance to the ordering) exposes the neighbors of v, then reveals the neighbors of the "exposed" vertices, etc., until we have explored $B_{l+1}(v)$. Note that we always expose all neighbors of $\partial B_s(v)$ before any neighbor of a vertex in $\partial B_{s+1}(v)$.

Consider the events $T_r(v) = \{B_r(v) \text{ is a tree}\}$ for $0 \le r \le l$. Since the events $T_r(v)$ are nested and $\mathbb{P}(T_0(v)) = 1$, we conclude that

$$\mathbb{P}(T_l(v)) = \prod_{r=0}^{l-1} \mathbb{P}(T_{r+1}(v)|T_r(v)) \ . \tag{2.1}$$

As we construct $T_{r+1}(v)$, at each step, half-edge choices for the next match that do not create cycles are all those belonging to vertices not yet considered. There are fewer than $(d)^{r+2}$ vertices that have been considered so far, for a total of less than $(d)^{r+3}$ possible bad matches. Hence, for $T_{r+1}(v)$ to hold, we have for each of the $(d)(d-1)^{r-1} < (d)^r$, vertices at the r^{th} level, at least $nd-(d)^{r+3}$ choices for the half edges, out of the maximum possible nd. (In fact we have fewer than nd possible choices remaining; however, since $r \le l = l = c \log n$, for c small, we still have $nd(1 - o(1))$ possibilities, at every step.) This means that

$$\mathbb{P}(T_{r+1}(v)|T_r(v)) \ge \left(\frac{n(d) - (d)^{r+3}}{n(d)} \right)^{(d)^r} .$$

By (2.1),

$$\mathbb{P}(T_l(v)) \ge \prod_{r=0}^{r=l-1} \left(\frac{n(d) - (d)^{r+3}}{n(d)} \right)^{(d)^r} .$$

Taking logarithms, we obtain

$$\begin{aligned} \log(\mathbb{P}(T_l(v))) &\ge \sum_{r=0}^{r=l-1} (d)^r \log(1 - \frac{(d)^{r+2}}{n}) = -(1 + o(1)) \sum_{r=0}^{r=l-1} \frac{(d)^{2r+2}}{n} \\ &= -(1 + o(1)) \frac{(d)^{2l+1}}{n} = -(1 + o(1))(d) n^{2\delta - 1} \end{aligned}$$

This implies that $\mathbb{P}(T_l(v)) \ge 1 - O(n^{2\delta-1})$ for n large enough. Hence

$$\mathbb{E}(X^{(l)}) = \sum \mathbb{E}(T_l(v)^c) \le O(n^{-2\delta}).$$

The result follows using Markov's Inequality. □

2.3 Probabilistic estimates in random regular graphs.

The previous section will allow us to crucially reduce the number of cycles when applying the trace method. We will need to estimate the probability of certain events to happen in random regular graphs. Our estimates follow from the seminal reults by McKay and coauthors ([24], [**?**]).

We set the next lemma for random bipartite biregular graphs. We use the following notation: Let H be a subgraph of G with vertex set $\{v_1, v_2, \dots, v_k\}$. Let d_i be the degree of v_i in G, so $d_i = d_1$ if v_i is in the set V_1 and d_2 if not. Also, denote by h_i the degree of v_i in H. For natural numbers x and t we use

$$(x)_{[t]} = x(x-1)(x-2)(\dots)(x-t+1)$$

to denote the falling factorial.

Lemma 2.4. *Let $H \subset K_{n,m}$ such that $\mathcal{E}(H) = o(n)$ and $G \sim \mathcal{G}(n, m, d_1, d_2)$. The following holds*

$$\mathbb{P}(H \subset G) \leq \frac{\prod (d_i)_{[h_i]}}{(nd_1 - 4d^2)_{[\mathcal{E}(H)]}}$$

$$\mathbb{P}(H \subset G) \geq \frac{\prod (d_i)_{[h_i]}}{(nd_1 - 1)_{[E(H)]}} \left(\frac{nd_1 - E(H) - 5d^2}{nd_1 - cE(H) - 5d^2} \right)^{\mathcal{E}(H)}$$

for some explicit constant $c < 1$, where $d = \max(d_1, d_2)$.

Proof. It follows directly from [24, Theorem 3.5]. □

We will use this Lemma crucially to show that the appearance of edges in random bipartite biregular graphs are weakly correlated, as long as the number of edges is not to big.

Lemma 2.5. *Let $G \sim G(n, m, d_1, d_2)$ and let $H \subset K_{n,m}$ such that the size of $\mathcal{E}(H)$ is $|H| = o(n)$. Let e be an edge not in H and such that H and $H \cup \{e\}$ have the same number of connected components.*

(i) If e and H share an endpoint of degree d_2 in G,

$$\mathbb{P}(e \in G | H \in G) \leq \frac{d_2 - 1}{n} + O\left(\frac{|H|}{n^2}\right).$$

(ii) If e shares exactly one endpoint with H and this vertex has degree one in H and degree d_2 in G, then

$$\mathbb{P}(e \in G | H \in G) = \frac{d_2 - 1}{n} + O\left(\frac{|H|}{n^2}\right).$$

Proof. For part (i), notice that the graph $H \cup \{e\}$ has increased by one the $H-$degree of at least one vertex. Using Lemma 2.4 we get

$$\mathbb{P}(H \cup \{e\} \in G) \leq \frac{d_1(d_2-1)}{nd_1 - 4d^2 - |H|}\,\mathbb{P}(H \in G) = \frac{d_2-1}{n}\left(1 + O\left(\frac{|H|}{n}\right)\right)\mathbb{P}(H \in G).$$

Now, for part (ii), is sufficient to note, by Lemma 2.4 that under our assumptions the inequality above is an equality. □

Remark 2.6. The results of Lemma 2.5 hold true if we consider vertices with degree d_1 in G. Noticing that $nd_1 = md_2$ we can check that the corresponding bound is $(d_1-1)/m$.

The final result of this section is a simple estimation involving the moments of a Bernoulli random variable.

Lemma 2.7. *Let $X \sim Ber(q)$ with $q \leq p + r$, where $0 \leq q, p \leq 1$. For any integer $m > 1$,*

$$\mathbb{E}\left((X-p)^m\right) \leq p + r.$$

Proof. Assume $q < p$. Then:

$$\mathbb{E}(|X-p|^m) \leq (1-p)^m p + p^m \leq p\ ;$$

the latter inequality follows easily by noting that it is satisfied for $m = 2$, and that $(1-p)^m p + p^m$ is a decreasing function of m for all $0 \leq p \leq 1$.

If $q > p$, write $q = p + r'$ with $0 < r' < r$. We get:

$$\mathbb{E}(|X-p|^m) \leq (1-p)^m(p+r') + p^m \leq (1-p)^m p + p^m + r' \leq p + r\ ,$$

due to similar considerations. □

Chapter 3

Spectral gap in random bipartite biregular graphs

This chapter is devoted to the prove of one of the main contribution of this thesis: an optimal bound on the spectral gap for random bipartite biregular graphs. As mentioned in Chapter 1, a similar question has been answered recently by Friedman ([10]) and Bordenave ([11]) for random regular graphs. For random bipartite biregular graphs, Puder ([25, Corollary 1.6]) obtained the upper bound $|\lambda| \leq 2\sqrt{d-1} + 0.84$ for λ been any non trivial eigenvalue of G, a random bipartite $d-$regular graph. The two main constrains of this result are the fact that all degrees are equal and that the author gets a bound of the right order $(2\sqrt{d-1})$ plus an absolute constant, 0.84. We are able to improve an generalize this result to all random bipartite biregular graph and by having a $o(1)$ quantity added to the expected order.

Theorem 3.1. *(Spectral gap for random bipartite biregular graphs) Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a random bipartite biregular graph. For any nontrivial eigenvalue η of G it holds*

$$\eta \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \tilde{\epsilon}_n$$

asymptotically almost surely, with $\tilde{\epsilon}_n \to 0$.

To proof Theorem 3.1 we proof an equivalent result regarding the spectrum of the non backtracking operator of G. We set this result now and explain later why it is equivalent to the previous.

Theorem 3.2. *Let B be the non bactracking operator associated to $G \sim \mathcal{G}(n, m, d_1, d_2)$. If λ_2 is the second largest eigenvalue of B, it holds*

$$\lambda_2 \leq ((d_1 - 1)(d_2 - 1))^{1/4} + \epsilon_n$$

asymptotically almost surely, with $\epsilon_n \to 0$.

3.1 The non backtracking operator.

The non backtracking operator allows us to count non backtracking walks, in the same way the adjacency matrix counts walks. Formally, this operator, denoted by B, is a linear endomorphism of $\mathbb{R}^{|\vec{E}|}$, where $\vec{E}$ is the set of oriented edges of G and $|\vec{E}| = 2|E|$. For oriented edges $e = (u, v)$ and $f = (s, t)$ (here u is the starting vertex of e and v its end, similarly for f) B_{ef} is defined:

$$B_{ef} = \begin{cases} 1, & \text{if } v = s \text{ and } u \neq t; \\ 0, & \text{else.} \end{cases}$$

Non backtracking walks are one of the central objects of study in graph theory. They turn out to be crucial in Broder and Shamir trace method ([26]), which was used extensively by Friedman ([27], [10], [13]) and also by Bordenave ([11]) in his proof of Alon's conjecure, see 1.1. One of the most stunning connections of the operator B is with the *Ihara Zeta function* of a finite graph, which is the (theoretic) analogue of the Riemann Zeta function in number theory ([28], [29]).

We are interested in studying the connection of the spectrum of B with the spectrum of the adjacency matrix.

3.1.1 Ihara-Bass formula for bipartite biregular graphs.

In this section, G will always denote an element of the set of bipartite biregular graphs $\mathcal{G}(n, m, d_1, d_2)$, defined in 1.2.1, A will denote its adjacency matrix and B the non backtracking operator associated to G.

We may order the elements of $\vec{E}$ as $\{e_1, e_2, \cdots, e_{2|E|}\}$, so that the first $|E|$ have end point in the set V_2. In this way, we can write

$$B = \begin{pmatrix} 0 & M \\ N & 0 \end{pmatrix}.$$

for matrices $M, N \in M_{|E|}(\mathbb{R})$ with entries equal to 0 or 1.

Denote the spectrum of an operator L by $\sigma(L)$. Set:

$$\sigma(B) = \{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2|E|}\}$$

Denote by $\mathbf{1}_\alpha$ the vector with first $|E|$ coordinates equal to 1 and the last $|E|$ equal to $\alpha = \sqrt{d_1 - 1}/\sqrt{d_2 - 1}$. We can check that

$$B\mathbf{1}_\alpha = B^*\mathbf{1}_\alpha = \lambda\mathbf{1}_\alpha$$

for $\lambda = \sqrt{(d_1 - 1)(d_2 - 1)}$. By the Perron-Frobenius Theorem, we conclude that $\lambda_1 = \lambda$ and the associated eigenspace has dimension one. Also, one can check that if λ is an eigenvalue of B with eigenvector $v = (v_1, v_2)^T$, $v_i \in \mathbb{R}^{|E|}$ then $-\lambda$ is also an eigenvalue with eigenvector $v' = (-v_1, v_2)^T$. Thus, $\sigma(B) = -\sigma(B)$ and $\lambda_{2|E|} = -\lambda_1$.

Next, we present the Ihara-Bass formula, which relates $\sigma(A)$ and $\sigma(B)$, proved by Bass ([30]) and Kutani and Sunada ([29]), see also [31, Theorem 3.3].

Theorem 3.3. *Let $G = (V, E)$ be any finite graph and B be its non-backtracking operator. Then*

$$\det(B - \lambda I) = (\lambda^2 - 1)^{|E|-|V|}\det(D - \lambda A + \lambda^2 I),$$

where D is the diagonal matrix with $D_{vv} = \deg(v) - 1$ and A is the adjacency matrix of G.

From the theorem above we get the following relation between $\sigma(A)$ and $\sigma(B)$

$$\sigma(B) = \{\pm 1\} \bigcup \{\lambda : \ D - \lambda A + \lambda^2 I \text{ is not invertible}\}.$$

We use the special structure of G to get a more precise description of $\sigma(B)$. Let $X \in M_{n\times m}(\mathbb{R})$ be the matrix with entries $X_{ij} = 1$ if and only if there is an edge between vertices i and j. The matrices A and D are equal to:

$$A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}, \quad D = \begin{pmatrix} (d_1 - 1)I_n & 0 \\ 0 & (d_2 - 1)I_m \end{pmatrix},$$

where I_k is the $k \times k$ identity matrix.

Let $\lambda \in \sigma(B)\backslash\{-1, 0, 1\}$. There exists a nonzero vector v such that:

$$(D - \lambda A + \lambda^2 I)v = 0$$

Writing $v = (v_1, v_2)$ with $v_1 \in \mathbb{R}^n$, $v_2 \in \mathbb{R}^m$, we obtain:

$$Xv_2 = \frac{d_1 - 1 + \lambda^2}{\lambda}v_1, \quad X^*v_1 = \frac{d_2 - 1 + \lambda^2}{\lambda}v_2$$

which imply that

$$\eta^2 = \frac{(d_1 - 1 + \lambda^2)(d_2 - 1 + \lambda^2)}{\lambda^2}$$

is a non zero eigenvalue of both XX^* and X^*X. The above relation gives us the following claim:

Any $\lambda \in \sigma(B)\backslash\{-1,0,1\}$ satisfies:

$$\lambda^4 - (\eta^2 - d_1 - d_2 + 2)\lambda^2 + (d_1 - 1)(d_2 - 1) = 0 \tag{3.1}$$

where η^2 is an nonzero eigenvalue of X^*X or, equivalently, $-\eta$ and η are eigenvalues of A.

The equivalence of Theorem 3.1 and 3.2 follows directly from (3.1).

3.2 Matrix decomposition

We start working towards the proof of Theorem 3.2. The first step is the following simple lemma we borrow from [11].

Lemma 3.4. *Let T, R be matrices such that $\mathrm{Im}(T) \subset \mathrm{Ker}(R)$, $\mathrm{Im}(T^*) \subset \mathrm{Ker}(R)$. Then all eigenvalues λ of $T + R$ that are not eigenvalues of T satisfy:*

$$|\lambda| \leq \max_{x \in \mathrm{Ker}(T)} \frac{||(T+R)x||}{||x||}$$

Applying the lemma above with $T = \lambda_1^\ell S_\alpha$ and $R = B^\ell - T$. We get:

$$\lambda_2 \leq \sup_{x \in \mathrm{Ker}(T),\ \|x\|_2 = 1} \left(\|B^\ell x\| \right)^{1/\ell} \tag{3.2}$$

where $S_\alpha = \left(\mathbf{1}_\alpha \mathbf{1}_\alpha^* - \mathbf{1}_{-\alpha}\mathbf{1}_{-\alpha}^*\right)$ and $\mathbf{1}_{-\alpha}$ is defined analogous to $\mathbf{1}_\alpha$, see subsection 3.1.1. It will be important to have a more precise description of the set $\mathrm{Ker}(T)$. It is not hard to check that

$$\mathrm{Ker}(T) = \{x :\ \langle x, \mathbf{1}_\alpha \rangle = \langle x, \mathbf{1}_{-\alpha} \rangle = 0\} = \{(v, w) \in \mathcal{R}^{2|E|}\ \langle v, \mathbf{1} \rangle = \langle w, \mathbf{1} \rangle = 0\}.$$

Above, the vectors v, w and $\mathbf{1}$ are $|E|-$dimensional.

We will assume G, sampled uniformly from $\mathcal{G}(n, m, d_1, d_2)$, is ℓ-tangle-free, which hold with high probability. Let Γ_{ef}^ℓ be the set of all non-backtracking paths in the complete bipartite graph $K_{n,m}$ of length $\ell + 1$, starting at oriented edge e and ending at f. For a path $\gamma \in \Gamma_{ef}^\ell$, we write $\gamma = (e_1, e_2, \ldots, e_{\ell+1})$ where $e_i \in \vec{E}$ for all i, $e_1 = e$ and $e_{\ell+1} = f$.

Similarly, define $F_{ef}^{\ell} \subset \Gamma_{ef}^{\ell}$ be the set of all non-backtracking, tangle-free paths of length $\ell+1$, starting at oriented edge e and ending at f. Then, in G

$$(B^{\ell})_{ef} = \sum_{\gamma \in \Gamma_{ef}^{\ell}} \prod_{t=1}^{\ell} B_{e_t e_{t+1}} = \sum_{\gamma \in F_{ef}^{\ell}} \prod_{t=1}^{\ell} B_{e_t e_{t+1}},$$

where we note the last equality requires G to be ℓ-tangle-free. Denote by $\bar{B}$ the matrix with entries equal to

$$(\bar{B}^{\ell})_{ef} = \sum_{\gamma \in F_{ef}^{\ell}} \prod_{t=1}^{\ell} (B-S)_{e_t e_{t+1}},$$

where

$$S = \begin{pmatrix} 0 & \frac{d_2-1}{n}\mathbf{1}\mathbf{1}^* \\ \frac{d_1-1}{m}\mathbf{1}\mathbf{1}^* & 0 \end{pmatrix}.$$

Note that $\bar{B}$ is an *almost* centered version of B, and $\mathrm{Ker}(S) = \mathrm{Ker}(T) = \mathrm{span}(\mathbf{1}_{\alpha}, \mathbf{1}_{-\alpha})$.

Using the telescoping sum formula in [11]

$$\prod_{s=1}^{\ell} x_s = \prod_{s=1}^{\ell} y_s + \sum_{j=1}^{\ell} \prod_{s=1}^{j-1} y_s (x_j - y_j) \prod_{t=j+1}^{\ell} x_t$$

with $x_s = B_{e_s e_{s+1}}$ and $y_s = \bar{B}_{e_s e_{s+1}}$, we obtain the following relation:

$$(B^{\ell})_{ef} = (\bar{B}^{\ell})_{ef} + \sum_{\gamma \in F_{ef}^{\ell}} \sum_{j=1}^{\ell} \prod_{i=1}^{j-1} \bar{B}_{e_i e_{i+1}} S_{e_j e_{j+1}} \prod_{t=j+1}^{\ell} B_{e_t e_{t+1}}. \tag{3.3}$$

This decomposition naturally breaks the elements in F_{ef}^{ℓ} into two subpaths, also non-backtracking and tangle-free, of length j and $\ell - j$, respectively. To recover the matrices B and $\bar{B}$ by rearranging (3.3), we need to also count those tangle-free subpaths that arise from tangled paths. While breaking a tangle-free path will necessarily give us two new tangle-free subpaths, the converse is not always true. This extra term generates a remainder that we define now.

Let $T_{ef}^{\ell,j} \subset \Gamma_{ef}^{\ell}$ be the set of non-backtracking paths in $K_{n,m}$ of length $\ell+1$, starting at e and ending at f, such that overall the path is tangled but the first j and last $\ell - j$ edges form tangle-free subpaths of G. Set the remainder

$$R_{ef}^{\ell,j} = \sum_{\gamma \in T_{ef}^{\ell}} \sum_{j=1}^{\ell} \prod_{i=1}^{j-1} \bar{B}_{e_i e_{i+1}} S_{e_j e_{j+1}} \prod_{i=j+1}^{\ell} B_{e_i e_{i+1}}. \tag{3.4}$$

Adding and subtracting $\sum_{j=1}^{\ell} R_{ef}^{\ell,j}$ to (3.3) and rearranging the sums, we obtain

$$B^{\ell} = \bar{B}^{\ell} + \sum_{j=1}^{\ell} \bar{B}^{j} S B^{\ell-j} - \sum_{k=1}^{\ell} R^{\ell,j}. \tag{3.5}$$

Multiplying (3.5) on the right by $x \in \mathrm{Ker}(T)$ and using that $B^{\ell-j}x$ is also within $\mathrm{Ker}(T)$, since it is just the space spanned by the leading eigenvectors, we find that the middle term is identically zero. Thus,

$$\left\| B^{\ell} x \right\| \leq \left\| \bar{B}^{\ell} x \right\| + \left\| \sum_{k=1}^{\ell} R^{\ell,j} x \right\|. \tag{3.6}$$

Combining with (3.2) we get

$$\lambda_2 \leq \left(\left\| \bar{B}^{\ell} \right\| + \left\| \sum_{k=1}^{\ell} R^{\ell,j} \right\| \right)^{1/\ell}. \tag{3.7}$$

Hence, we need to upper bound the norm of the matrices $\bar{B}^{\ell}$ and $R^{\ell,j}$, $1 \leq j \leq \ell$. We use the trace method. For any k we have

$$\mathbb{E}\left(\|\bar{B}^{\ell}\|^{2k} \right) \leq \mathbb{E}\left(Tr\left((\bar{B}^{\ell})^{2k} \right) \right) = \mathbb{E}\left(\sum_{\gamma} \prod_{i=1}^{\ell} \bar{B}^{\ell}_{e_i e_{i+1}} \right) \tag{3.8}$$

where the sum is taking over the set of all cycles $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{2k}\}$ of length $2k\ell$ forming by concatenation of $2k$ elements $\gamma_i \in \Gamma^{\ell}$, with the convention $e_1 = e_{\ell+1}$. The simplest approach is to bound the number of such cycles and to compute the expectation of he corresponding random variables. Again, the tangle–free property comes to the rescue allowing us to find a proper bound on the number of cycles. We deal with these two tasks in the next two subsections.

3.2.1 Path counting.

Our goal now is to find a reasonable bound for the size of $C^{r}_{v,e}$, defined as the set of cycles of length $2k\ell$ obtained as the concatenation of $2k$ non-backtracking, tangle-free walks in G of length ℓ which visit exactly v different vertices, r of them in V_2, and e different edges. Note, these are edges in $E(G)$ and not directed edges in $\vec{E}(G)$. We denote such a cycle as $\gamma = \{\gamma_1, \gamma_2, \cdots, \gamma_{2k}\}$, where each γ_j is a length ℓ walk. We are assuming that we traverse γ in the following way: we start at the initial vertex of γ_1, move along this path until we meet γ_2 and continue along this path and so on. We relabel the vertices from $\{1, 2, \cdots, v\}$ as they appear in γ. Denote by $\mathcal{T}_{\gamma}$ the spanning tree of those edges leading to new vertices as induced by γ. Notice that the enumeration of the vertices tells us how we traverse the cycle and thus define $\mathcal{T}_\gamma$ uniquely.

We encode each walk γ_j by dividing it into sequences of subpaths of three types, which in our convention always occur as type 1 followed by type 2 followed by type 3, although some may be empty subpaths. Given our current position on the cycle, i.e. the label of the current vertex, and the subtree of $\mathcal{T}_\gamma$ already discovered (over the whole cycle γ not just the current walk γ_j), we define each type as follows:

Type 1: These are paths with the property that all of their edges are edges of $\mathcal{T}_\gamma$ and have been traversed already in the cycle. These paths can be encoded by their end vertex; because this is a path contained in a tree, thus there is a unique path connecting its initial and final vertex. We use 0 if no old edges occur before the type 2 path, i.e. the path is empty.

Type 2: These are paths with all of their edges in $\mathcal{T}_\gamma$ but which are traversed for the first time in the cycle. We can encode these paths by their length, since they are traversing new edges, and we know in what order the vertices are discovered. We use 0 if the path is empty.

Type 3: These paths are simply a single edge, not belonging to $\mathcal{T}_\gamma$, that connects the end of a path of type 1 or 2 to a vertex that has been already discovered. Given our position on the cycle, we can encode an edge by its final vertex. Again, we use 0 if the path is empty.

Now, we decompose γ_j into an ordered sequence of triples to encode its subpaths:

$$(p_1, q_1, r_1)(p_2, q_2, r_2)\cdots(p_t, q_t, r_t),$$

where p_i characterizes subpaths of type 1, q_i characterizes subpaths of type 2, and r_i characterizes subpaths of type 3. These subpaths occur in the order given by the triples. We perform this decomposition using the minimal possible number of triples.

Now p_i and r_i are both numbers in $\{0, 1, ..., v\}$, since our cycle has v vertices. On the other hand, $q_i \in \{0, 1, ..., l\}$ since it represents the length of a subpath of a non-backtracking walk of length ℓ. Hence, there are $(v + 1)^2(l + 1)$ possible triples. Next, we want to bound how many of these triples occur in γ_j. We will use the following lemma.

Lemma 3.5. *Let $(p_1, q_1, r_1)(p_2, q_2, r_2)\cdots(p_t, q_t, r_t)$ the encoding of a non backtracking walk γ_j, as described above. Then $r_i = 0$ can only occur in the last triple $i = t$.*

Proof. We can check this case by case. Assume that for some $i < t$ we have $(p_i, q_i, 0)$, and consider the concatenation with $(p_{i+1}, q_{i+1}, r_{i+1})$. First, notice that both p_{i+1} and q_{i+1} cannot be zero since then we will have $(p_i, q_i, 0)(0, 0, v^*)$ which can be written as (p_i, q_i, v^*). If $q_i \neq 0$, then we must have $p_{i+1} \neq 0$. Otherwise, we split a path of new edges (type 2), and the decomposition is not minimal. This implies that we visit new edges and move to edges already visited, hence we need to go through a type 3 edge, implying that $r_i \neq 0$. Finally, if $p_i \neq 0$ and $q_i = 0$, then we must have $p_{i+1} = 0$; otherwise, we split a path of old edges (type 1). We also require $q_{i+1} \neq 0$, but $(p_i, 0, 0)(0, q_{i+1}, r_{i+1})$ is the same as (p_i, q_{i+1}, r_{i+1}), which contradicts the minimality condition. This covers all possibilities and finishes the proof. □

Using the lemma, any representation of a non-backtracking walk γ_j has at most one triple with $r_i = 0$. All other triples indicates the traversing of a type 3 edge. We now give a very rough upper bound for how many of such encodings there can be. To do so, we will use the tangle-free property and slightly modify the encoding of the paths with cycles. Consider the two cases:

Case 1: **Path γ_j contains no cycle.** This implies that we traverse each edge within γ_j once. Thus, we can have at most $\chi = e - v + 1$ many triples with $r_i \neq 0$. This gives a total of

$$\big((v+1)^2(\ell+1)\big)^{\chi+1}$$

many way to encode one of these paths.

Case 2: **Path γ_j contains a cycle.** Since we are dealing with non-backtracking walks, we enter the cycle once, loop around some number of times, and never come back. We change the encoding of such paths as follows: Let γ_j^a, γ_j^b, and γ_j^c be the segments of the path before, during, and after the cycle. We mark the start of the cycle with $|$ and its end with $\|$. The new encoding of the path is:

$$(p_1^a, q_1^a, r_1^a) \cdots (p_{t^a}^a, q_{t^a}^a, r_{t^a}^a) \,|\, (p_1^b, q_1^b, r_1^b) \cdots (p_{t^b}^b, q_{t^b}^b, r_{t^b}^b) \,\|\, (p_1^c, q_1^c, r_1^c) \cdots (p_{t^c}^c, q_{t^c}^c, r_{t^c}^c),$$

where we encode the segments separately. Observe that each a subpath is connected and self-avoiding. The above encoding tells us all we need to traverse γ_j, including how many times to loop around the cycle: since the total length is ℓ, we can back out the number of circuits around the cycle from the lengths of γ_j^a, γ_j^b, and γ_j^c. Following the analysis made for Case 1, the subpaths γ_j^a, γ_j^b, γ_j^c are encoded by at most $\chi + 1$ triples, but we also have at most ℓ choices each for our marks $|$ and $\|$. We are left with

$$\ell^2 \left((v+1)^2(\ell+1)\right)^{\chi+1}$$

many ways to encode these class of paths.

To conclude, we first choose v vertices, r in the set V_2, and order them, which can occur in $(m)_r(n)_{v-r} \leq m^r n^{v-r}$ different ways, and we are counting concatenations of $2k$ such paths. In total, we conclude that

$$|C^r_{v,e}| \leq m^r n^{v-r}(2\ell)^{4k}\left((v+1)^2(\ell+1)\right)^{2k(\chi+1)}.$$

3.2.2 Expectation bounds.

We turns our attention to the expectation in the right hand side of (3.8). We will use the estimates from Chapter 2.

Let $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_{2k})$ be a cycle obtained by the concatenation of $2k$ non-backtracking walks of length ℓ. Recall that we traverse γ in the following way: we start at the initial vertex of γ_1, move along this path until we meet γ_2 and continue along this path and so on. With this convention, denote by E_γ the set of directed edges traversed by γ. A subpath of γ is just an ordered path of edges traversed as described above. Define

$$X_\gamma = \prod \left(\bar{B}_{ef}\right)^{m_{ef}} \tag{3.9}$$

where $e, f \in E_\gamma$ are such that the oriented path ef is a sub path of γ when traversed as described above, and m_{ef} is the number of times we traversed ef. The main result of this section is the following theorem.

Theorem 3.6. *Let $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_{2k})$ be a cycle obtained by the concatenation of $2k$ non-backtracking walks o length ℓ. If γ visits $K = o(n)$ different edges, it holds*

$$\mathbb{E}\left(X_\gamma\right) \leq \frac{C_\gamma}{m^r n^{K-r}}\left(\frac{K}{n}\right)^{\omega}(1+o(1)) \tag{3.10}$$

where $C_\gamma = (d_1-1)^r(d_2-1)^{K-r}$ and r depends on γ and can be computed explicitly, and $\omega = \omega(\gamma) = \left\lfloor \frac{\sum \mathcal{X}_{\{m_{ef}+m_{fe}=1\}}}{3d} \right\rfloor$ is a fraction of the number of path of length two that we traverse exactly once, regarding the orientation.

The proof of Theorem 3.6 goes as follow: we order the set of undirected edges visited by γ and define a sequence of nested sigma algebras $\{\mathcal{F}_t\}$, $1 \leq t \leq K$ each containing the information of the first t edges in this order. We use the tower of expectation to bound the right hand side of (3.10). At each step, a new edge is removed from the filtration and we are able to improve our current bound via Lemma 2.5. The ordering of the edges is done in a way that allows us to use part (ii) of the Lemma a maximal number of times.

We start by describing the ordering of the edges. Let $E = \{e_i\}_{i=1}^K$ be a set of undirected edges. An element π in the symmetric group S_K can be identified with an ordering in E by taking the first edge to be $e_{\pi(1)}$, the second to be $e_{\pi(2)}$ and so on.

For a subset $F \subset E$ of edges, define

$$N(F) = \{e \in E \ \text{ s.t: } e \text{ shares a vertex with some } f \in F\}$$

of *neighbors* of F. Notice that the orientation is not relevant in this definition.

We said that, in an ordering π the edge $e_{\pi(j)}$ is **good** if the following condition holds:

(i) There is exactly one value of $i \leq j-1$ such that $e_{\pi(i)} \in N(e_\pi(j))$.

Informally, this means that the j^{th} edge is **good** if it has exactly one neighbor among the previous edges. We denote by $\omega(\pi)$ the number of **good** edges in π. We have the following lemma.

Lemma 3.7. *Let γ be a closed walk in a graph with maximal degree d and let E be the set of undirected edges traversed by γ. There exist an ordering π of the edges such that*

$$\omega(\pi) \geq \left\lfloor \frac{|E|}{3d} \right\rfloor .$$

Proof. We construct an ordering with the desired property using the following procedure. At time 0, start with the sets $E'_0 = \emptyset$, $E''_0 = \emptyset$ and $E'''_0 = E$. At time t we:

- Choose $e_i, e_j \in E'''_{t-1}$ such that e_ie_j or e_je_i is a subpath of γ. Set $\pi(2t-1) = i$ and $\pi(2t) = j$.
- Set $E'_t = E'_{t-1} \cup \{e_i, e_j\}$, $E''_t = E''_{t-1} \cup N(\{e_i, e_j\}) \backslash E'_t$ and $E'''_t = E \backslash (E'_t \cup E''_t)$.

Note that, at all times, E'_t, E''_t and E'''_t form a disjoint partition of E corresponding to, respectively, the set of edges already ordered, the set of the neighbors of the edges already ordered and the complement of those two sets. Also, it is not hard to check that e_{2t} is **good** for all t. This process will end when one of the following exclusive events happen:

- $E_k''' = \emptyset$. We had defined $\pi(t)$ for $1 \le t \le 2k$. We now let $\pi(t) \in E_k''$, for $2k+1 \le t \le |E|$ in any way we want. By construction, we have at least k **good** edges. Because each step we remove from E_t''' at most $3d$ edges we conclude that

$$k \ge \left\lfloor \frac{|E|}{3d} \right\rfloor$$

as desired.

- After k steps, no two edges in E_k''' form a subpath of γ. By construction, E_k' and E_k''' are disconnected. Because γ is a closed walk, for each $e \in E_k'''$ there exist $f \in E_k''$ such that ef or fe is a subpath of γ. We set then $e_{\pi(2k+1)} = f$ and $e_{\pi(2k+2)} = e$. With this choice, e is **good**. We update E_{k+1}', E_{k+1}'' and E_{k+1}''' as before and repeat until $E_T''' = \emptyset$. Clearly, this lead to at least

$$T \ge \left\lfloor \frac{|E|}{3d} \right\rfloor$$

good edges, which concludes the proof.

□

We are ready to prove Theorem 3.6.

Proof of Theorem 3.6. Let $\pi = \{e_1, e_2, \dots, e_K\}$ be an ordering of the K edges visited by γ. Furthermore let A_t be the set of oriented edges

$$A_t = \{\vec{e}_1, (\vec{e}_1)^{-1}, \dots, \vec{e}_t, (\vec{e}_t)^{-1}\}$$

containing the first t edges with both possible orientations. Define the sequence of sigma algebras $\{\mathcal{F}_t\}_{1\le t\le K}$ as

$$\mathcal{F}_t = \sigma(A_t)$$

We have that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \mathcal{F}_{K-1}$. Recall the definition of X_γ from (3.9). Consider the random variables $\{X_t\}_{2\le t\le K}$ defined as

$$X_t = \prod \left(\bar{B}_{ef}\right)^{m_{ef}}$$

where $e, f \in A_t$, ef is a subpath of γ, and exactly one of them is on the set $\{\vec{e}_t, (\vec{e}_t)^{-1}\}$. Informally, X_t is the product of all factors in X_γ which involves e_t and smaller (according to π) edges. It is not hard to see that

$$X_\gamma = \prod_{t=2}^{K} X_t,$$

$$Y_j = \prod_{t=2}^{j} X_t \text{ is } \mathcal{F}_j\text{−measurable.}$$

With these definitions, we have

$$\mathbb{E}(X_\gamma) = \mathbb{E}(\mathbb{E}(X_\gamma|\mathcal{F}_{K-1})) = \mathbb{E}(\mathbb{E}(Y_{K-1}X_K|\mathcal{F}_{K-1})) = \mathbb{E}(Y_{K-1}\mathbb{E}(X_K|\mathcal{F}_{K-1})). \quad (3.11)$$

We focus on the term $\mathbb{E}(X_K|\mathcal{F}_{K-1})$). Notice that, for oriented edges e, e^{-1}, f, g and h we have

$$(B_{ef}|f) = (B_{eg}|g) = (B_{e^{-1}h}|h).$$

Informally speaking, these equalities say that, under the event $\{f, g, h$ are edges of the grpah$\}$, the random variables B_{ef}, B_{eg} and $B_{e^{-1}h}$ are identical (we are assuming that the orientation is such that these entries are not identically zero). Let $\{Z_t\}_{2\leq t\leq K}$ be independent random variables with distribution:

$$Z_t =_d B_{e_t f}|f$$

where $f \in A_{t-1}$ and $e_t f$ is a subpath of γ. With this notation, we have

$$\mathbb{E}(X_K|\mathcal{F}_{K-1}) = \mathbb{E}\left(\left(Z_K - \frac{d_1-1}{m}\right)^t \left(Z_K - \frac{d_2-1}{n}\right)^s\right) \quad (3.12)$$

where t is the number of times we traverse $\vec{e}_K$ and s is the number of times we traverse $(\vec{e}_K)^{-1}$. We have several cases, depending on the values of t and s.

Case 1: $t \geq 2$, then

$$\left(Z_K - \frac{d_1-1}{m}\right)^t \left(Z_K - \frac{d_2-1}{n}\right)^s \leq \left(Z_K - \frac{d_1-1}{m}\right)^2$$

and, by Lemmas 2.7 and 2.5,

$$\mathbb{E}\left(\left(Z_K - \frac{d_1-1}{m}\right)^2\right) \leq \frac{d_1-1}{m} + O\left(\frac{K}{m^2}\right).$$

Case 2: $s \geq 2$. Is analogous.

Case 3: $t = s = 1$. Expanding the right hand side of (3.12) and using Lemma 2.5 we get:

$$\begin{aligned}\mathbb{E}\left(Z_K - \frac{d_1-1}{m}\right)\left(Z_K - \frac{d_2-1}{n}\right) &= \mathbb{E}\left(Z_K\left(1 - \frac{d_1-1}{m} - \frac{d_2-1}{n}\right) + O\left(\frac{1}{nm}\right)\right) \\ &\leq \frac{d_2-1}{n} + O\left(\frac{K}{n^2}\right).\end{aligned}$$

Case 4: $t = 0$ and $s > 1$ or $s = 0$ and $t > 1$. Apply Lemma 2.7 dirctly to get the desired bound.

Case 5: $t = 0,\ s = 1$ or $s = 0,\ t = 1$. This case is the only time we may see a bound of order n^{-2}. Say that $s = 1$. We only traverse the edge e_K once, hence, there is only one adjacent edge to e_K in $\mathcal{F}_{K-1}$.

then we have, by Lemma 2.5 part (ii)

$$\mathbb{E}(X_K|\mathcal{F}_{K-1})) = \mathbb{E}\left(Z_K - \frac{d_2-1}{n}\right) = O\left(\frac{K}{n^2}\right).$$

Back to (3.11), we have:

$$\mathbb{E}(X_\gamma) = \mathbb{E}(Y_{K-1}\mathbb{E}(X_K|\mathcal{F}_{K-1})) \leq \left(\frac{d_{e_K}-1}{n} + O\left(\frac{K}{n^2}\right)\right)\mathbb{E}(Y_{K-1})$$

where d_{e_K} equals d_1 of d_2 depending on the way we applied the conditional expectation. Apply the same argument to $\mathbb{E}(Y_{K-1})$, conditioning now on $\mathcal{F}_{K-2}$. After $K-1$ iterations we get a bound of the form:

$$\mathbb{E}(X_\gamma) \leq \left(\frac{C_\gamma}{m^r n^{K-r}}\right)\left(1 + O\left(\frac{K}{n}\right)\right)^K = \frac{C_\gamma}{m^r n^{K-r}}(1 + o(1)).$$

Lastly, to get a better upper bound, we reorder the edges so that we have a maximal number of **good** edges for which case 5 above applies. By Lemma 3.7, this quantity is $\omega(\gamma)$. This concludes the proof.

3.2.3 Bounds on the norm of $\bar{B}^\ell$ and $R^{\ell,j}$.

Theorem 3.8. *Let $\ell \leq c\log(n)$ where c is a universal constant as in Lemma 2.3. It holds*

$$\|\bar{B}^\ell\| \leq \log(n)^{15}\left((d_1-1)(d_2-1)\right)^{\ell/4}.$$

asymptotically almost surely.

Proof. Recall (3.8):

$$\mathbb{E}\left(\|\bar{B}^{\ell}\|^{2k}\right) \leq \mathbb{E}\left(Tr\left((\bar{B}^{\ell})^{2k}\right)\right) = \mathbb{E}\left(\sum_{\gamma}\prod_{i=1}^{\ell}\bar{B}^{\ell}_{e_i e_{i+1}}\right)$$

where the sum is taking over the set of all cycles $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{2k}\}$ of length $2k\ell$ forming by concatenation of $2k$ elements of Γ^{ℓ}, with the convention $e_1 = e_{\ell+1}$.

To compute the expectation of the right hand side we split the sum into subsets attending to the number of vertices, $v(\gamma)$, in the cycle. Recall that two oriented edges, e, f, form a subpath of γ if we traverse one right after the other. For convenience, we will call such subpath a *2-path*. With this notation, let

- $\mathcal{C}_1 := \{\gamma :$ all *2-path* in γ are traversed at least twice, regarding the orientation.$\}$
- $\mathcal{C}_2 := \{\gamma :$ at least one *2-path* in γ is traversed exactly once and $v(\gamma) \leq kl + 1.\}$
- $\mathcal{C}_3 := \{\gamma$ at least one *2-path* in γ is traversed exactly once and $v(\gamma) > kl + 1.\}$

The reason for this division is, intuitively, the following. By Lemma 2.5, when we have *2-paths* traversed exactly once the expectation of the corresponding cycle is smaller. Hence, we will see that the order of the expectation of the right hand side in (3.8) will come from the cycles in $\mathcal{C}_1$.

For $j = 1, 2, 3$, let

$$I_j = \mathbb{E}\left(\sum_{\gamma\in\mathcal{C}_j}\prod_{i=1}^{\ell}\bar{B}^{\ell}_{e_i e_{i+1}}\right)$$

and then,

$$\mathbb{E}\left(\|\bar{B}^{\ell}\|^{2k}\right) \leq I_1 + I_2 + I_3. \tag{3.13}$$

We will bound each term on the right hand side above. Notice that, for $j = 1, 2, 3$:

$$I_j \leq \sum_{\gamma\in\mathcal{C}_j} |\mathcal{C}^r_{v,e}| \frac{C^*_{\gamma}}{m^r n^{e-r}} \left(\frac{e}{n}\right)^{\omega_*} (1 + o(1)) \tag{3.14}$$

where $|\mathcal{C}^r_{v,e}|$ denotes the size of the set $\mathcal{C}^r_{v,e}$, $C^*_{\gamma} = \max_{\gamma\in\mathcal{C}^r_{v,e}}\{C_{\gamma}\}$, $\omega_* = \min_{\gamma\in\mathcal{C}^r_{v,e}}\{\omega(\gamma)\}$ and we used Theorem 3.6.

To deal with I_1, notice that each cycle traverses $2k\ell$ *2-paths*. Hence, for each $\gamma \in I_1$, we have at most $k\ell$ different *2-paths*. Furthermore, since each edge can be in multiple *2-paths*, we have that the total number of different *2-paths* is greater or equal than the total number of edges traversed by γ. Denoting this quantity by $e(\gamma)$, we then have $e(\gamma) \leq k\ell$. Since γ is connected, we have $v(\gamma) \leq k\ell + 1$. Lastly, observe that $\omega(\gamma) = 0$ for any $\gamma \in \mathcal{C}_1$. Using the counting arguments of section 3.2.1, we can simplify the right hand side of (3.14) to get

$$I_1 \leq \sum_{v=\ell+1}^{k\ell+1} \sum_{e=v-1}^{k\ell} n^{v-e}(2\ell)^{4k} \left((v+1)^2(\ell+1)\right)^{2k(\chi+1)} C^*_\gamma.$$

In the sum above, there is still a dependence on r that we choose to drop, for clarity. We will see soon why we are able to do so. Indeed, the leading term on the right hand side corresponds to $v - e = 1$ and $e = k\ell$. Because any γ is connected, for this values of v and e the graph induced by γ is a tree, which implies that $r = \lfloor \frac{k\ell+1}{2} \rfloor$. We conclude that

$$I_1 \leq n \left(2\ell(k\ell+2)^2(\ell+1)\right)^{4k} ((d_1-1)(d_2-1))^{\lfloor \frac{k\ell+1}{2} \rfloor} (1 + o(1)). \tag{3.15}$$

We turn our attention to I_2. Because there is at least a *2-path* traversed exactly once, we have $e(\gamma) \geq v(\gamma)$ for $\gamma \in \mathcal{C}_2$. With the same notation as above, we have:

$$I_2 \leq \sum_{v=\ell+1}^{k\ell+1} \sum_{e=v}^{2k\ell} n^{v-e}(2\ell)^{4k} \left((v+1)^2(\ell+1)\right)^{2k(\chi+1)} C^*_\gamma$$

where we dropped the term $\left(\frac{e}{n}\right)^{\omega_*} \leq 1$. Now the leading term is obtained when $e = v = k\ell + 1$, which give us

$$I_2 \leq \left(2\ell(k\ell+2)^2(\ell+1)\right)^{4k} ((d_1-1)(d_2-1))^{\lfloor \frac{k\ell+1}{2} \rfloor} (1 + o(1)). \tag{3.16}$$

We focus now on I_3. Notice that cycles in $\mathcal{C}_3$ will visit many vertices. We first show that, in this case, $\omega(\gamma)$ is also large. Let $v(\gamma) = k\ell + t$, $p(\gamma)$ the number of different *2-paths* traversed by γ and $\tilde{p}(\gamma)$ the number of *2-paths* traversed exactly once. We have $p(\gamma) \geq e(\gamma) \geq v(\gamma) = k\ell + t$. Furthermore, since γ has length $2k\ell$ we deduce that

$$2(p(\gamma) - \tilde{p}(\gamma)) + \tilde{p}(\gamma) \leq 2k\ell$$

which implies that $\tilde{p}(\gamma) \geq 2t$ and thus $\omega(\gamma) \geq \frac{2t}{3d}$. From (3.14) we have

$$I_3 \leq \sum_{v=k\ell+1}^{2k\ell} \sum_{e=v}^{2k\ell} n^{v-e}(2\ell)^{4k} \left((v+1)^2(\ell+1)\right)^{2k(\chi+1)} C^*_\gamma \left(\frac{e}{n}\right)^{\frac{2(v-k\ell)}{3d}}.$$

Since $C^*_\gamma \leq (d-1)^e$ we have, for $v = e$:

$$C^*_\gamma \left(\frac{e}{n}\right)^{\frac{2(v-k\ell)}{3d}} \leq (d-1)^v \left(\frac{v}{n}\right)^{\frac{2(v-k\ell)}{3d}} = (d-1)^{k\ell}\left(d\left(\frac{v}{n}\right)^{\frac{2}{3d}}\right)^{(v-k\ell)} \leq (d-1)^{k\ell}$$

since $v \leq 2k\ell = o(n)$, d is constant and $v - k\ell \geq 1$. We get

$$I_3 \leq k\ell \left(2\ell(v+1)^2(\ell+1)\right)^{4k} (d-1)^{k\ell}(1+o(1)) \tag{3.17}$$

Plugging (3.15), (3.16) and (3.17) into (3.13) we obtain

$$\mathbb{E}\left(\|\bar{B}^\ell\|^{2k}\right) \leq n\left(2\ell(k\ell+2)^2(\ell+1)\right)^{4k}((d_1-1)(d_2-1))^{\lfloor\frac{k\ell+1}{2}\rfloor}(1+o(1)).$$

To finish the proof, let

$$k = \left\lfloor \frac{\log(n)}{\log(\log(n))} \right\rfloor.$$

With this choice, and using that $\ell \leq 1/4\log(n)$, we have

$$\left(2\ell(k\ell+2)^2(\ell+1)\right)^{4k} \leq O(n^{28})$$

and, by Markov's inequality:

$$\begin{aligned}
\mathbb{P}(\|\bar{B}^\ell\| &> \log(n)^{15}\left((d_1-1)(d_2-1)\right)^{\ell/4}) \leq \frac{\mathbb{E}\left(\|\bar{B}^\ell\|^{2k}\right)}{\log(n)^{30k}\left((d_1-1)(d_2-1)\right)^{k\ell/2}} \\
&\leq n^{-29}\left(2\ell(k\ell+2)^2(\ell+1)\right)^{4k}\left((d_1-1)(d_2-1)\right)(1+o(1)) \\
&= o(1).
\end{aligned}$$

since $\log(n)^{30k} = n^{30}$. □

Theorem 3.9. *Let $1 \leq j \leq \ell \leq c\log(n)$ where c is a universal constant satisfying the condition in Lemma 2.3. It holds*

$$\|R^{\ell,j}\| \leq \log(n)^{16}$$

asymptotically almost surely.

Proof. The proof is analogous to the proof of Theorem 3.8. We have, for any integer k

$$\mathbb{E}\left(\|R^{\ell,j}\|^{2k}\right) \leq \mathbb{E}\left(Tr\left((R^{\ell,j})^{2k}\right)\right) = \mathbb{E}\left(\sum_\gamma \prod_{i=1}^{\ell}(R^{\ell,j})_{e_ie_{i+1}}\right). \tag{3.18}$$

Now, the sum is over cycles $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{2k}\}$ of length $2k\ell$ forming by concatenation of $2k$ elements of $T^{\ell,j}$, with the convention $e_1 = e_{\ell+1}$, see section 3.2 for the definition of $T^{\ell,j}$. Denote by $D^r_{v,e}$ the set of these cycles that visit exactly v vertices, r of which are in V_2, and $e + 2k$ different edges. Here, the extra $2k$ edges are the j^{th} edge in each ℓ path γ_i, which connects the first j edges to the last $\ell - j$. We bound the size of these sets using the same reasoning as in section 3.2.1. For $\gamma \in D^r_{v,e}$, each γ_i is divided into two tangle-free non backtracking walks of length j and $\ell - j$. By encoding each of these paths as in section 3.2.1 we conclude that there are at most

$$4\ell^4((v+1)^4(\ell+1)^2)^{\chi+1}$$

such γ_i. Hence, concatenating $2k$ many of these gives us that

$$|D^r_{v,e}| \leq n^{v-r} m^r (4\ell^4)^{2k}((v+1)^4(\ell+1)^2)^{2k(\chi+1)}. \tag{3.19}$$

Notice that, on the right hand side of (3.18) we have, for each γ, terms of the form

$$(\bar{B}_{ef})^{m_{ef}}(B_{ef})^{m'_{ef}}$$

since now *2-paths* are weighted by entries of both $\bar{B}$ and B, and m'_{ef} is the number of times we traverse the *2-path* ef regarding the orientation. If $m'_{ef} > 0$, we have

$$(\bar{B}_{ef})^{m_{ef}}(B_{ef})^{m'_{ef}} \leq (1 - B_{ef})^{m_{ef}} B_{ef} \leq B_{ef}$$

and the corresponding conditional expectation can be upper bounded by $(d-1)/n$, by Lemma 2.5. If $m'_{ef} = 0$ we proceed as in the proof of Theorem 3.6. After dropping the term K/n, we get

$$\mathbb{E}\left(\sum_\gamma \prod_{i=1}^{\ell} (R^{\ell,j})_{e_i e_{i+1}}\right) \leq \sum_{v,e,r} |D^r_{v,e}| \frac{(d-1)^e}{m^r n^{\ell-r}} \left(\frac{(d-1)}{n}\right)^{2k} (1+o(1))$$

where we included an upper bound on the factor arising from the $2k$ entries of the matrix S, see section 3.2. Using the bound for $|D^r_{v,e}|$ in (3.19), we obtain

$$\mathbb{E}\left(\sum_\gamma \prod_{i=1}^{\ell} (R^{\ell,j})_{e_i e_{i+1}}\right) \leq \sum_{v,e,r} n^{v-e}(4\ell^4)^{2k}((v+1)^4(\ell+1)^2)^{2k(\chi+1)}(d-1)^e \left(\frac{(d-1)}{n}\right)^{2k} (1+o(1)).$$

Finally, we notice that $1 \leq v \leq 2k\ell$ and $v - 1 \leq e \leq 2k\ell$. Furthermore, if $v - e = 1$, which implies that r can take a unique value, we get a term linear in n. Also, for fixed v and e the are less than n different values of r. Since we have at most $4k^2\ell^2$ pairs we conclude that the left hand side above can be bounded by:

$$
\begin{aligned}
4nk^2\ell^2(4\ell^4)^{2k}((2k\ell+1)^4(\ell+1)^2)^{2k}(d-1)^{2k\ell}\left(\frac{(d-1)}{n}\right)^{2k}(1+o(1)) &= \\
4nk^2\ell^2(4\ell^4)^{2k}((2k\ell+1)^4(\ell+1)^2)^{2k}\left(\frac{(d-1)^{\ell+1}}{n}\right)^{2k}(1+o(1)) &\leq \\
4nk^2\ell^2(4\ell^4)^{2k}((2k\ell+1)^4(\ell+1)^2)^{2k}(1+o(1)).&
\end{aligned}
$$

Finally, let

$$k = \left\lfloor \frac{\log(n)}{\log(\log(n))} \right\rfloor .$$

Now

$$4k^2\ell^2(4\ell^4)^{2k}((2k\ell+1)^4(\ell+1)^2)^{2k} \leq n^{29}$$

and, by Markov's inequality

$$
\begin{aligned}
\mathbb{P}(\|R^{\ell,j}\| > \log(n)^{16} \quad &\leq \quad \frac{\mathbb{E}\left(\|R^{\ell,j}\|^{2k}\right)}{\log(n)^{32k}} \\
&\leq \quad n^{-31}4k^2\ell^2(4\ell^4)^{2k}((v+1)^4(\ell+1)^2)^{2k}(1+o(1)) = o(1)
\end{aligned}
$$

since $\log(n)^{32k} = n^{32}$ $\square$

Proof of Theorem 3.2.

Combining Theorem 3.8 and 3.9 into (3.7):

$$
\begin{aligned}
|\lambda_2| \quad &\leq \quad \left(\log(n)^{15}\left((d_1-1)(d_2-1)\right)^{\ell/4} + \ell\log(n)^{16}\right)^{1/\ell} \\
&= \quad ((d_1-1)(d_2-1))^{1/4} + \epsilon_n.
\end{aligned}
$$

as desired.

Chapter 4

The regular stochastic block model

4.1 Definition of the model and main results

The stochastic block model (SBM) is a classical cluster-exhibiting random graph model that has been extensively studied, both empirically and rigorously, across numerous fields. In its simplest form, the SBM is a model of random graphs on $2n$ nodes with two equal-sized clusters $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| = |\mathcal{B}| = n$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$. Edges between various pairs of vertices appear independently with probability $p = p_n$ if the two vertices belong to the same cluster and with probability $q = q_n$ otherwise. Thus, for any vertex, the expected number of same-class neighbors is $a := a_n := p(n-1) \sim pn$, and the expected number of across-class neighbors is $b := b_n := qn$.

Given a realization of the graph, the broad goal is to determine whether it is possible (with high probability) to find the partition $\mathcal{A}, \mathcal{B}$; and if the answer is yes, whether it is possible to do so using an efficient algorithm. Otherwise, the best one can hope for is the existence of an algorithm that will output a partition which is highly (or at least positively) correlated with the underlying cluster. To this end, consider the space $\mathcal{M}$ of all algorithms which take as input a finite graph on $2n$ vertices and output a partition of the vertex set into two sets. Informally, we say that an algorithm in $\mathcal{M}$ allows for **weak recovery** if, with probability going to 1 as n goes to infinity, it outputs a partition (A', B') such that $|\mathcal{A}\Delta A'| + |\mathcal{B}\Delta B'| = o(n)$ (here Δ denotes the symmetric difference). We say that an algorithm allows for **strong recovery** if, with probability going to 1 as n goes to infinity, it outputs the partition $(\mathcal{A}, \mathcal{B})$. Finally, an algorithm in $\mathcal{M}$ will be called **efficient** if its run time is polynomial in n.

The problem of community detection described above is closely related to the min-bisection problem, where one looks for a partition of the vertex set of a given graph into two subsets of equal size such that the number of edges across the subsets is minimal. In general, this problem is known to be NP-hard [32]; however, if the min-bisection is smaller than most of the other bisections, the problem is known to be simpler. This fact was noticed a few decades ago, with the advent of the study of min-bisection in the context of the SBM. In particular, Dyer and Frieze [33] produced one of the earliest results when they showed that if $p > q$ are fixed as $n \to \infty$ then the min-bisection is the one that separates the two classes, and it can be found in expected $O(n^3)$ time. Their results were improved by Jerrum and Sorkin [34] and Condon and Karp [35]. Each of these papers were able to find faster algorithms that worked for sparser graphs. The latter work was able to solve the min-bisection problem when the average degrees were of order $n^{1/2+\epsilon}$.

Until a few years ago most of the literature on both the min-bisection problem and community detection in the SBM had focused on the case of increasing expected degrees (i.e. $a, b \to \infty$ as $n \to \infty$), with the best results at that time showing that if the smallest average degree is roughly $\log n$, then weak recovery is possible (e.g., McSherry [36] showed that spectral clustering arguments can work to detect the clusters in this setting). Recently, the sparse case, i.e. when $a, b = O(1)$ has been the focus of a lot of interest. This regime is interesting both from a theoretical and an applied point of view since a lot of real world networks turn out to be sparse; for more on this see [37]. Coja-Oghlan demonstrated a spectral algorithm that finds a bisection which is positively correlated with the true cluster when the average degree is a large constant [38]. Using ideas from statistical physics, Decelle, Krzakala, Moore and Zdeborová gave a precise prediction for the problem of recovering a partition positively correlated with the true partition in the sparse SBM [39]. The prediction was rigorously confirmed in a series of papers by Mossel, Neeman and Sly [40] [41], and Massoulié [42], where it was shown that this level of recovery is possible iff $(a-b)^2 > (a+b)$. More recently, [43] found necessary and sufficient conditions for a and b under which strong recovery is possible. Before them, Abbe, Bandeira and Hall [44] also characterized strong recovery assuming the edge probabilities to be constant factors of $\frac{\log(n)}{n}$.

In [40] Mossel, Neeman and Sly proposed two regular versions of the SBM in a sparse regime, and they conjectured thresholds for the recovery of a correlated partition for each of the models. They also suggested that spectral methods should help to differentiate between the regular SBM and a random regular graph. We introduce here a slightly different version of a regular SBM where in addition to the graph being regular, the number of neighbors that a vertex has within its own community is also a constant. Formally, we have the following definition.

Definition 4.1. For integers n, d_1 and d_2 denote by $\mathcal{G}(n, d_1, d_2)$, the random regular graph with vertex set $[2n]$, obtained as follows: Choose an equipartition (parts have equal sizes) $(\mathcal{A}, \mathcal{B})$ of the vertex set, uniformly from among the set of such equipartitions. Choose two independent copies of uniform simple d_1-regular graphs with vertex set $\mathcal{A}$, respectively $\mathcal{B}$. Finally, connect the vertices from $\mathcal{A}$ with those from $\mathcal{B}$ by a random d_2-bipartite-regular graph chosen uniformly. We refer to this family of measures on graphs as the regular stochastic block model (RSBM).

Our goal is to investigate the similarities and differences between the RSBM and the classical SBM. We assume that $\min\{d_1, d_2\} \geq 3$. This assumption implies that, with high probability, the resulting graph is connected. This differs from the SBM with bounded average degree, which has a positive density of isolated vertices, which make strong recovery impossible. The constant degree of all the vertices in the RSBM makes the local neighborhoods easier to analyze; however, as this model lacks the edge-independence present in the SBM, some computations become significantly more difficult.

Throughout the rest of this chapter we say a sequence of events happen asymptotically almost surely (a.a.s.) if the probabilities of the events go to 1 along the sequence. The underlying measure will be always clear from context.

Our first result, the next proposition, pertains to the rigidity of RSBM; it says that the RSBM is asymptotically distinguishable from a uniformly chosen random regular graph with the same average degree. Below, $||\cdot,\cdot||_{TV}$ denotes the total variation distance between measures.

Proposition 4.2. *Let μ_n be the measure induced by $\mathcal{G}(n, d_1, d_2)$ on the set Reg$(2n, d_1 + d_2)$ of all $(d_1 + d_2)$-regular graphs on $2n$ vertices and let μ'_n be the uniform measure on the same set Reg$(2n, d_1 + d_2)$. Then for any positive integers $d_1, d_2 \geq 3$,*

$$\lim_{n\to\infty} ||\mu'_n, \mu_n||_{TV} = 1.$$

This result sharply contrasts the RSBM and the SBM (which is indistinguishable from an Erdős-Rényi random graph with the same size and average degrees satisfying $(a-b)^2 \leq (a+b)$ [40]).

In order to determine whether it is possible to recover the partition in the RSBM, one must first answer a basic question about the random graph $\mathcal{G}(n, d_1, d_2)$: is the 'true partition' $(\mathcal{A}, \mathcal{B})$ identifiable. I.e., is $(\mathcal{A}, \mathcal{B})$ the only way to partition the graph such that the subgraphs on the parts are d_1-regular (which then implies that the subgraph across is d_2-bipartite)? The following result shows that the answer is yes if d_1 and d_2 are sufficiently large.

Theorem 4.3. *There exists a constant $d' > 0$ such that, for $d_1 > d_2 > d'$, $\mathcal{G}(n, d_1, d_2)$ has a unique partition a.a.s.*

The particular value of d' that we get is far from optimal; we conjecture that the conclusion of this theorem should be true for $d' = 2$. The proof of Theorem 4.3 is quite technical and is given in section 4.3.

To our knowledge, this is the first uniqueness of partition result for block models with constant degrees. Such a result is not true, however, in the classical setting where the edges are independent, since with constant probability one has isolated vertices.

If the original partition is unique in most cases then one can, in principle, find the original partition by exhaustive search, and hence achieve strong recovery. This is again in sharp contrast with the SBM, where strong recovery is achievable only in the case of growing degrees.

The next natural direction is to look for an efficient algorithm for strong recovery. While we do not answer this question in general, we do exhibit one regime where such an algorithm exists. The nature of the algorithm is essentially spectral. We believe that the regime defined below is the largest one can cover using properties of the graph's spectrum.

Theorem 4.4. *Assume $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$. Then there is an efficient algorithm that allows strong recovery.*

Remark 4.5. In the case that d_1 is even our graph is a special case of the so called "random lifts" and we can use their spectral properties (see Section 4.2.1). We will give the proof of Theorem 4.4 when d_1 is even in the main body of the text and postpone the proof for any value of d_1 to the appendix.

The proof of the above theorem is broken into two parts. The first part uses a spectral argument to prove weak recovery. Formally we have the following lemma:

Lemma 4.6. *Assume $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$. Then there is an efficient algorithm that allows weak recovery.*

Lemma 4.6 gives us weak recovery. Strong recovery is then achieved by recursively applying the majority algorithm where one simultaneously updates the label of each vertex by the majority label among the neighbors. That this can be done is again an example of the rigidity in this model, and highlights one of the main differences between RSBM and the classic SBM. It shows that for the former, existence of an efficient algorithm for weak recovery implies the existence of an algorithm for strong recovery. This contrasts with the separate thresholds in the SBM [43].

We present the majority algorithm in the section below.

4.1.0.1 Majority algorithm.

Recall that $\mathcal{A}$ and $\mathcal{B}$ are the true communities. Let (A, B) be any partition (not necessarily an equipartition) of the vertex set. For each $i \in [2n]$, let $\sigma_i = +1$ if $i \in A$ and $\sigma_i = -1$ if $i \in B$.

> **Initialize** $A_0 = A, B_0 = B$.
>
> **For** $i \in [2n]$ (majority rule)
> $\hat{\sigma}_i = \text{sign}(\sum_{v_j \sim v_i} \sigma_j)$
>
> **Return** $A_1 = \{v_i : \hat{\sigma}_i = +1\}$, $B_1 = \{v_i : \hat{\sigma}_i = -1\}$

Similar applications of the majority algorithm appear in [44] and [43]. There, the authors find criteria for both weak recovery and strong recovery in the SBM. It is not hard to see that weak recovery and strong recovery are not equivalent in the sparse SBM, since the presence of isolated vertices prevents strong recovery.

Throughout the rest of the article we will refer to the majority algorithm as **Majority**. The following theorem, along with Lemma 4.6, completes the proof of Theorem 4.4.

Theorem 4.7. *Assume $d_1 > d_2 + 4$. Then there exists an $\varepsilon = \varepsilon(d_1) > 0$ such that the following is true a.a.s.: given a graph $\mathcal{G}(n, d_1, d_2)$ and any partition (A, B) of its vertex set such that $|A \cap \mathcal{A}| > (1 - \varepsilon)n$ and $|B \cap \mathcal{B}| > (1 - \varepsilon)n$, **Majority** recovers the true partition $(\mathcal{A}, \mathcal{B})$ if started with (A, B), after $O(\log(n))$ iterations. The constant in the $O(\cdot)$ depends on ε, d_1.*

The way we iterate the **Majority** algorithm will be clear from the proof of Theorem 4.7.

4.1.1 Organization

The rest of this chapter is organized as follows: in section 4.3, we prove Proposition 4.2 and Theorem 4.3. We present an informal sketch of the proof of Theorem 4.3 in section 4.2. Section 4.4 is concerned with proving Theorem 4.4 when d_1 is even. Section 4.5 contains the proofs of Theorem 4.7 as well as Theorem 4.4 with no restriction on the parity of d_1. The proof of Lemma 4.6 is deferred to section **??**. Finally, we introduce some useful notions on random lifts and multigraphs in section 4.2.1, where we explain how to obtain Theorem 4.4 when d_1 is even.

4.2 Sketch of the proof of Theorem 4.3.

Recall from Definition 4.1, in the graph $G := \mathcal{G}(n, d_1, d_2)$ on $[2n]$, $(\mathcal{A}, \mathcal{B})$ form the true partition.

Let us introduce the following notation: for any $V \subset [2n]$ let G_V denote the subgraph induced by G on V. For disjoint subsets V_1, V_2, let $G_{(V_1,V_2)}$ denote the subgraph on $V_1 \cup V_2$ induced by the edges in G with one endpoint in V_1 and the other in V_2. For any $v \in [2n]$ and $V \subset [2n]$ let $deg_V(v)$ denote the number of edges incident on v whose other endpoint is in V.

Thus Theorem 4.3 says that, a.a.s., there does not exist any $V \subset [2n]$ with $V \neq \mathcal{A}, \mathcal{B}$ and $|V| = n$ such that the following two conditions hold simultaneously:

- Both G_V and $G_{[2n]\setminus V}$ are d_1-regular graphs.
- $G_{(V,[2n]\setminus V)}$ is a d_2-regular bipartite graph.

However we show that it is even unlikely that G_V is d_1-regular for any $V \neq \mathcal{A}, \mathcal{B}$ with $|V| = n$. To this end we fix such a V and let $V_1 := V \cap \mathcal{A}$, $V_2 := V \cap \mathcal{B}$, and assume $|V_2| = \alpha n$ with $\alpha \leq \frac{1}{2}$. Note that, given G, V and $\mathcal{A}$, the degree sequence $\{deg_{V_1}(v)\}_{v \in V_1}$ is determined; if G_V were d_1-regular graph then for each $v \in V$,

$$deg_{V_1}(v) + deg_{V_2}(v) = d_1,$$

and hence the degree sequence $\{deg_{V_2}(v)\}_{v \in V_1}$ is also determined, i.e. the number of edges going from each vertex in V_1 to V_2 is fixed.

It can be shown using the configuration model (see Section **??** for the definition) that the joint distribution of $\{deg_{V_2}(v)\}_{v \in V_1}$ behaves like i.i.d. $Bin(d_2, \alpha)$'s. The proof now follows by using the above to estimate the probability of a certain degree sequence from this distribution, and by a union bound over all possible choices of V. We remark that the formal proof involves some case analysis depending on the size of $|V_2|$ and relies on the expansion properties of regular graphs when $|V_2|$ is small.

4.2.1 Sketch of the proof of Theorem 4.4 when d_1 is even.

To prove Theorem 4.4 when d_1 is even, we make use of the recent work on the spectra of random lifts of graphs in [45, 46] and the references therein. For a wonderful exposition of lifts of graphs see [47]. We now introduce the notion of lift of a multigraph.

4.2.1.1 Random lifts and multigraphs

By a multigraph we simply mean a graph that allows for multiple edges and loops. Next we define the notion of lift. Informally, an n-lift of a multigraph $X = (V, E)$ is a multigraph $X_n = (V_n, E_n)$, such that for each vertex in V there are n vertices in V_n and locally both graphs look the "same". Formally, let $V_n := V \times \{1, 2, \dots n\}$. To define the edge set in the lift consider the set $S_n^E := \{\pi_e\}_{e \in E}$ where $\pi_e \in S_n$ (the set of permutations of $[n]$). We have:

$$E_n := \{((x, i), (y, \pi_e(i))) : e = (x, y) \in E,\ \ 1 \leq i \leq n\},$$

for $\pi \in S_n^E$. Thus every edge in E "lifts" to a matching in E_n. For every $v \in V$, let $v \times \{1, 2, \dots n\}$ be called the *fiber* of v.

A random lift is the lift constructed from $\pi \in S_n^E$ where $\{\pi_e\}_{e \in E}$ are chosen uniformly and independently from S_n. Let A and A_n be the adjacency matrices of the multigraphs X and X_n, respectively. One can check that all the eigenvalues of A are also eigenvalues of A_n and the corresponding eigenvectors can be "lifted" as well to an eigenvector (which is constant on fibers) of the lifted graph. Let the remaining eigenvalues of A_n be,

$$|\mu_1| \geq |\mu_2| \geq \dots \geq |\mu_r|, \tag{4.1}$$

where $r = n|V| - |V|$. With the above definitions we now state one of the main results in [45].

Theorem 4.8. *Let $d \geq 3$ be an integer and let X be a finite, d-regular multigraph. If X_n is a random n-lift of X then, for any $\varepsilon > 0$,*

$$\lim_{n \to \infty} \mathbb{P}(|\mu_1| \geq 2\sqrt{d-1} + \varepsilon) = 0\ .$$

Recall the definition of strong and weak recovery from Section 4.1. We also need the following definition.

Definition 4.9. Let $e := e_{2n}$ be the vector of all ones of length $2n$. Also let $\sigma = \sigma_{2n}$ be the vector of signs which denotes the partition $\mathcal{A}, \mathcal{B}$ i.e.

$$\sigma(x) = \begin{cases} +1 & x \in \mathcal{A}, \\ -1 & otherwise. \end{cases}$$

The proof of Theorem 4.4 follows by first realizing the graph
$mathcalG(n, d_1, d_2)$ as a random lift and then using the above theorem to show spectral

separation of A_n; moreover, it can be shown that, with high probability, σ in Definition 4.9 is an eigenvector associated to the second eigenvalue of the lift. The proof of Theorem 4.4 is now reduced to finding a good approximation to the unitary eigenvector corresponding to the second eigenvalue. Note that this allows the strong recovery of the partition $(\mathcal{A}, \mathcal{B})$.

4.3 Proof of Proposition 4.2 and Theorem 4.3.

Let K_n be the support of μ_n, i.e., K_n is the set of all graphs which are d_1-regular on $\mathcal{A}$ and $\mathcal{B}$ and d_2-regular and bipartite across, for some equipartition $(\mathcal{A}, \mathcal{B})$ of $[2n]$. Let $|\mathcal{G}(n,d)|$ be the number of d-regular graphs on n labelled vertices and let $|\mathcal{BG}(n,d)|$ be the number of d-regular bipartite graphs on $2n$ vertices. To show that $\mu_n'(K_n) \to 0$ we will use the following enumeration results that can be deduced from [48] and [49]. The idea is to count the number of points in the support of the measures μ_n and μ_n'. We have from [49, Corollary 5.3] :

$$|\mathcal{G}(n,d)| = C\frac{(nd)!}{(nd/2)!2^{nd/2}(d!)^n} \ , \tag{4.2}$$

asymptotically in n, where $C = C(n,d)$ remains bounded as n grows. Similarly, from [48, Theorem 2]:

$$|\mathcal{BG}(n,d)| = C_1\frac{(dn)!}{(d!)^{2n}} \ , \tag{4.3}$$

asymptotically in n, for $C_1 = C_1(n,d)$ a bounded function. We have:

$$\mu_n'(K_n) = \frac{|K_n|}{|\mathcal{G}(2n, d_1+d_2)|}$$

To compute $|K_n|$, recall Definition 4.1, first choose $\mathcal{A}$ and then use (4.2) and (4.3). We get:

$$\mu_n'(K_n) = C_2\binom{2n}{n}\left(\frac{(nd_1)!}{(nd_1/2)!2^{nd_1/2}(d_1)!^n}\right)^2\frac{(nd_2)!}{(d_2)!^{2n}} \\ \times\frac{(n(d_1+d_2))!2^{n(d_1+d_2)}(d_1+d_2)!^{2n}}{(2n(d_1+d_2))!}$$

for $C_2 = C_2(n, d_1, d_2)$ bounded as n grows. Using Stirling's Formula we get:

$$\mu_n'(K_n) = C_3\left(\frac{4\binom{d_1+d_2}{d_1}^2 d_1^{d_1} d_2^{d_2}}{2^{d_1+d_2}(d_1+d_2)^{d_1+d_2}}\right)^n$$

$$= C_3 \left(\frac{2\binom{d_1+d_2}{d_1}}{2^{d_1+d_2}} \right)^n \left(\frac{2\binom{d_1+d_2}{d_1} d_1^{d_1} d_2^{d_2}}{(d_1+d_2)^{d_1+d_2}} \right)^n$$

Where C_3 equals C_2 times a universal constant. Both fractions on the right hand side above are less than 1. This proves Proposition 4.2. □

4.3.0.1 Proof of Theorem 4.3

Recall that $d_1 > d_2$ and that $(\mathcal{A},\mathcal{B})$ are the true clusters. The idea, as discussed in Section **??**, will be to show that, conditioned on the choices of $\mathcal{A}$ and $\mathcal{B}$, if we choose another subset of n vertices, the probability of having a d_1-regular graph on these n vertices is small. The estimate on the above probability is crucial since it will then allow us to take a union bound over all possible subsets of size n to conclude that, a.a.s., there is a unique pair of clusters.

First we need some definitions.

Definition 4.10. Given a graph $G = (V, E)$,

i. For a vertex v and a set of vertices S denote by $deg_S(v)$ the number of neighbors of v in S.

ii. For any subsets $V_1 \subset V_2 \subset V$ define the boundary $\partial_{V_2} V_1$ to be the number of edges in E whose one end point lies in V_1 and the other in $V_2 \setminus V_1$. When $V_2 = V$ we use the simpler notation ∂V_1.

Consider non-empty subsets $A \subset \mathcal{A}$, $B \subset \mathcal{B}$ such that $|A \cup B| = n$. Without loss of generality assume $|A| \geq |B|$ and let α be such that

$$\alpha n = |B|. \tag{4.4}$$

We will prove Theorem 4.3 by showing that given the d_1-regular graph with vertex set $\mathcal{A}$, for any choice of A and B the probability that $A \cup B$ is a d_1-regular graph goes to zero as n goes to infinity. We use the simple observation that since $\mathcal{A}$ is d_1-regular, to have $A \cup B$ d_1-regular, for any vertex $v \in A$, the number of neighbors of v in B must be equal to the number of neighbors of v in $\mathcal{A} \setminus A$. The technical core of the proof involves showing that the probability of this event is small.

We start by proving a lemma. Recall that, in order to have a d_1-regular graph with vertex set $A \cup B$ with $A \subset \mathcal{A}$ and $B \subset \mathcal{B}$ it is necessary that $deg_B(v) = deg_{\mathcal{A}\setminus A}(v)$ for all $v \in A$. For notational brevity let

$$g_v := deg_{\mathcal{A}\setminus A}(v) \tag{4.5}$$

for all $v \in A$.

Lemma 4.11. *Given $A \subset \mathcal{A}, B \subset \mathcal{B}$ and a sequence of non-negative numbers $g = (g_1, g_2, \ldots, g_{|A|})$ let*

$$p(g_1, g_2, \ldots, g_{|A|}) := \mathbb{P}(deg_B(v) = g_v \text{ for all } v \in A).$$

Then, for any such g,

$$\max_{g'} p(g'_1, g'_2, \ldots, g'_{|A|}) = p(g^*_1, g^*_2, \ldots, g^*_{|A|}),$$

where $g^*_i \in \{\ell, \ell+1\}$ for some non negative number $\ell = \ell(g)$. The maximum in the above is taken over all sequences $g' = (g'_1, g'_2, \ldots, g'_{|A|})$ such that $\sum_{i=1}^{|A|} g'_i = \sum_{i=1}^{|A|} g_i$.

The above lemma says that, given the total number of edges going from A to B, the probability of a possible degree sequence is maximized when all the degrees are essentially the same. Clearly $l = \left\lfloor \frac{\sum_{i=1}^{|A|} g_i}{|A|} \right\rfloor$; the number of $(l+1)$ degrees occurring in $g^* = (g^*_1, g^*_2, \ldots, g^*_{|A|})$ is determined by $\sum_i g^*_i = \sum_i g_i$.

Proof. To compute $p(g_1, g_2, \ldots, g_{|A|})$ we use the exploration process, defined in 2.1, for the d_2-regular bipartite graph $(\mathcal{A}, \mathcal{B})$ where the vertices of $\mathcal{A}$ are exposed one by one. We order the vertices so that the vertices of A are exposed first. Let $\mathcal{F}_i$ be the filtration generated by the process up to the i^{th} vertex. Using the exchangeability of the variables $deg_B(v_i)$, given a sequence $\{g_i\}$, w.l.o.g. we can assume $g_1 = \min g_i$ and $g_2 = \max g_i$.

Assume now $g_2 - g_1 > 1$. We will show that $p(g_1, g_2, \ldots, g_{|A|}) < p(g_1+1, g_2-1, \ldots, g_{|A|})$, which implies the lemma. We start with the following simple observation:

$$\begin{aligned}
\mathbb{P}(deg_B(v_i) &= g_i,\ i \geq 3 \mid \mathcal{F}_2, deg_B(v_1) = g_1,\ deg_B(v_2) = g_2) = \\
\mathbb{P}(deg_B(v_i) &= g_i,\ i \geq 3 \mid \mathcal{F}_2, deg_B(v_1) = g_1 + 1, deg_B(v_2) = g_2 - 1).
\end{aligned}$$

This is because under the above two conditionings, the number of remaining unmatched half edges in $A, \mathcal{A}, B, \mathcal{B}$ is the same. Hence it suffices to show that

$$\mathbb{P}(deg_B(v_1) = g_1, deg_B(v_2) = g_2) < \mathbb{P}(deg_B(v_1) = g_1 + 1, deg_B(v_2) = g_2 - 1). \qquad (4.6)$$

Next we note that

$$\mathbb{P}(deg_B(v_1) = g_1, deg_B(v_2) = g_2) = \binom{d_2}{g_1}\binom{d_2}{g_2} \frac{(\alpha n d_2)_{[g_1+g_2]}((1-\alpha)nd_2)_{[2d_2-g_1-g_2]}}{(nd_2)_{[2d_2]}},$$

where $(x)_m$ is the falling factorial $(x)_{[m]} = x(x-1)\dots(x-m+1)$. To see the above, we first choose those half edges of v_1 and v_2 that will connect to half edges in B. Then we choose the $2d_2$ half edges in $\mathcal{B}$ that will match with the corresponding half edges of v_1 and v_2 such that exactly $g_1 + g_2$ are incident on vertices in B.

Substituting now into (4.6) we have:

$$\begin{aligned} p(g_1, g_2, ...g_{|A|}) < p(g_1+1, g_2-1, ...g_{|A|}) &\iff \binom{d_2}{g_1}\binom{d_2}{g_2} < \binom{d_2}{g_1+1}\binom{d_2}{g_2-1} \\ &\iff (g_1+1)(d_2-g_2+1) < g_2(d_2-g_1) \\ &\iff g_1 - g_2 + 1 < d_2(g_2 - g_1 - 1)\ , \end{aligned}$$

which follows immediately from $g_2 > g_1 + 1$. $\square$

Recall that we are interested in the probability that $A \cup B$ is d_1-regular for a fixed choice of A and B. As already discussed,

$$\mathbb{P}(A \cup B \text{ is } d_1\text{-regular}) \leq \mathbb{P}(deg_{\mathcal{A}\setminus A}(v) = deg_B(v), \forall v \in A). \tag{4.7}$$

Our next goal is to bound the probability of such an event. To this end we recall the notion of stochastic dominance.

Let ν_1 and ν_2 be two probability measures on $\mathbb{Z}$, and let $X \sim \nu_1$, $Y \sim \nu_2$. We use $X \preceq Y$ to denote that ν_2 stochastically dominates ν_1.

Recall now Definitions 4.1 and 4.10, as well as (4.5).

Lemma 4.12. *Let $M = \min\{\partial_{\mathcal{A}} A, n/2\}$, and let $Y = (Y_1, Y_2, \dots, Y_M)$ where $Y_i \sim Bin(d_2, 2\alpha)$ are i.i.d.. Then*

$$\mathbb{P}(deg_B(v) = g_v\ ,\ \forall v \in A \mid \mathcal{A}) \ \leq \ \prod_{i=1}^{M} \mathbb{P}(Y_i \geq 1).$$

For notational brevity, we have denoted by $\mathbb{P}(\cdot \mid \mathcal{A})$ the random graph measure $\mathcal{G}(n, d_1, d_2)$ conditioned on the subgraph induced by $\mathcal{A}$.

Proof. First recall that by Lemma 4.11 the quantity on the left hand side is maximized when for all v, $g_v \in \{\ell, \ell+1\}$. Hence we assume that this is the case. Now to prove the lemma we consider the exploration process defined above. The definition requires us to fix an order on the vertices of $\mathcal{A}$; we do this in the following way. Consider the two cases:

i. $\ell = 0$: First come all the vertices $v_i \in A$ with $g_i = 1$, followed by the remaining vertices in A. Then come all the vertices in $\mathcal{A} \setminus A$.

ii. $\ell > 0$: First come all the vertices $v_i \in A$ with $g_i = \ell$, followed by the remaining vertices in A. Then come all the vertices in $\mathcal{A} \setminus A$.

Recall that $\mathcal{F}_i$ is the filtration up to vertex i. Note that, for $1 \le i \le \min(\partial_{\mathcal{A}} A, n/2)$,

$$deg_B(v_i)\big|\mathcal{F}_{i-1} \ \preceq \ Bin\left(d_2, \frac{\alpha n d_2 - (i-1)}{n d_2 - i d_2}\right).$$

This follows from the simple observation that for any of the cases mentioned above for the i^{th} vertex, there are at most $(\alpha n d_2 - (i-1))$ half edges in B that haven't yet been matched. Now note that since by hypothesis $i \le \frac{n}{2}$,

$$\begin{aligned}\frac{\alpha n d_2 - (i-1)}{n d_2 - i d_2} &\le \frac{\alpha n d_2}{n d_2/2} \\ &= 2\alpha.\end{aligned}$$

Thus we are done. $\square$

As already used in the proof of the above lemma,

$$\mathbb{P}(A \cup B \text{ is } d_1\text{-regular} \mid \mathcal{A}) \le p(\ell, \ell, \dots, \ell, \ell+1, \dots, \ell+1)$$

for some $\ell = \ell(\mathcal{A}, A)$. In case i. we see that by Lemma 4.12

$$\begin{aligned} p(0,0,\dots,0,1,\dots,1) = p(1,1,\dots,1,0,\dots,0) &\le \prod_{i=1}^{\min\{n/2, \partial_{\mathcal{A}} A\}} \mathbb{P}(Y_i \ge 1) \\ &\le \prod_{i=1}^{\min\{n/2, \partial_{\mathcal{A}} A\}} (2 d_2 \alpha) \end{aligned} \tag{4.8}$$

The first equality follows by exchangeability. The first inequality follows from Lemma 4.12. The second is a simple consequence of the fact that for a nonnegative variable the probability of being bigger than 1 is at most its expectation.

In case *ii* by similar arguments

$$\begin{aligned} p(\ell, \ell, ...\ell, \ell+1, ..., \ell+1) &\le \prod_{i=1}^{n/2} \mathbb{P}(Y_i \ge 1) \\ &\le \prod_{i=1}^{n/2} (2 d_2 \alpha). \end{aligned} \tag{4.9}$$

Note that in (4.9) the term $\partial_{\mathcal{A}} A$ does not appear. This is because in this case by hypothesis

$$|\partial_{\mathcal{A}} A| \geq \ell |A| \geq \frac{n}{2}.$$

To proceed with the proof of Theorem 4.3 we quote two standard results on the expansion of random d-regular graphs. Let γ be the spectral gap for the operator of the random walk in the uniform random regular graph $G \in \mathcal{G}(n, d)$, i.e.:

$$\gamma = 1 - \frac{\lambda_2}{d} \tag{4.10}$$

where λ_2 is the second largest eigenvalue of the adjacency matrix of G.

Theorem 4.13. *[22, Theorem 1.1] With probability going to* 1 *as* $n \to \infty$,

$$\gamma \geq 1 - \frac{2}{\sqrt{d}}.$$

The next result was proven independently in [50] and [51]. We will use it as it appears in [52, Theorem 13.14].

Theorem 4.14. *Let G be a d-regular graph in n vertices. For any $S \subset V(G)$, with* $|S| \leq \frac{n}{2}$,

$$\frac{\gamma}{2} \leq \frac{|\partial S|}{d|S|}.$$

Putting everything together we get the following: For $d_1 \geq 16$, a.a.s., for all $S \subset \mathcal{A}$ with $|S| \leq \frac{n}{2}$

$$|\partial_{\mathcal{A}} S| \geq \frac{d_1}{4} |S|.$$

In particular since $|A| \geq n/2$ it follows that, a.a.s.,

$$|\partial_{\mathcal{A}} A| = |\partial_{\mathcal{A}}(\mathcal{A} \setminus A)| \ \geq \ \frac{d_1}{4} |\mathcal{A} \setminus A|. \tag{4.11}$$

In case i. ($\ell = 0$) plugging (4.11) in (4.8) we get

$$\begin{aligned} \mathbb{P}(A \cup B \text{ is } d_1\text{-regular}|\mathcal{A}) \leq \prod_{i=1}^{\min(n/2, |\partial_{\mathcal{A}} A|)} \mathbb{P}(Y_i \geq 1) \ &\leq \ \prod_{i=1}^{\frac{|\partial_{\mathcal{A}} A|}{2}} \mathbb{P}(Y_i \geq 1) \\ &\leq \prod_{i=1}^{\frac{d_1}{8}\alpha n} \mathbb{P}(Y_i \geq 1) \end{aligned} \tag{4.12}$$

assuming that the d_1-regular graph on $\mathcal{A}$ satisfies (4.11). The second inequality follows from the simple observation that since $\ell = 0$, we have $|\partial A| \leq n$.

Recall that we want an upper bound on the right hand side of 4.7. Combining Lemma 4.12, (4.9) and (4.12) we get

$$\mathbb{P}(A \cup B \text{ is } d_1\text{-regular} \mid \mathcal{A}) \;\leq\; \mathbb{P}(Y \geq 1)^{\frac{d_1}{8}\alpha n} + \mathbb{P}(Y \geq 1)^{n/2}. \tag{4.13}$$

The two terms on the right hand side correspond to the two cases $\ell = 0$ and $\ell \geq 1$.

Next we show that the bounds in (4.13) are good enough to be able to use union bound over all possible choices of A and B. There are $\binom{n}{\alpha n}^2$ ways to choose A and B. Denote by R_α the event that $A \cup B$ is d_1-regular for *at least* one choice of A and B. Thus by union bound,

$$\mathbb{P}(R_\alpha) \leq \binom{n}{\alpha n}^2 \left[\mathbb{P}(Y \geq 1)^{\frac{d_1}{8}\alpha n} + \mathbb{P}(Y \geq 1)^{n/2}\right]. \tag{4.14}$$

We now estimate the right hand side using Stirling's formula. Let

$$H(x) = -x \log x - (1-x)\log(1-x)$$

be the binary entropy function. Then the two terms in the right hand side of (4.14) are at most

$$\frac{2^{n[2H(\alpha)+\frac{d_1}{8}\alpha \log(\mathbb{P}(Y \geq 1))]}}{\sqrt{\alpha n}} \quad \text{and} \quad \frac{2^{n[2H(\alpha)+\frac{\log(\mathbb{P}(Y \geq 1))}{2}]}}{\sqrt{\alpha n}},$$

up to universal constants involved in Stirling's approximation. Our goal would be to upper bound the two exponents,

$$2H(\alpha) + \frac{d_1}{8}\alpha \log(\mathbb{P}(Y \geq 1)) \quad \text{and} \quad 2H(\alpha) + \frac{\log(\mathbb{P}(Y \geq 1))}{2}. \tag{4.15}$$

Recall that α was defined in (4.4). Consider the three following cases:

CASE 1: $\alpha \leq \frac{1}{d_2^2}$.

In this case we will use the bound $\mathbb{P}(Y \geq 1) \leq 2d_2\alpha$ by Lemma 4.12. Plugging this in (4.15) we get the following upper bounds

$$2H(\alpha) + \frac{d_1}{8}\alpha \log(2d_2\alpha) \quad \text{and} \quad 2H(\alpha) + \frac{\log(2d_2\alpha)}{2}.$$

Now,

$$\begin{aligned}2H(\alpha)+\frac{d_1}{8}\alpha\log(2d_2\alpha) &= -2\alpha\log(\alpha)+\frac{d_1}{8}\alpha\log(2d_2\alpha)-2(1-\alpha)\log(1-\alpha)\\ &\le \alpha\log(\alpha)(\frac{d_1}{32}-2)-2(1-\alpha)\log(1-\alpha)\\ &\le \alpha\log(\alpha)(\frac{d_1}{32}-4).\end{aligned}$$

To see the above inequalities first note that since $\alpha \le \frac{1}{d_2^2}$, $\log(2d_2\alpha) \le \frac{\log(\alpha)}{4}$ as soon as $d_2 \ge 4$, and also $|(1-\alpha)\log(1-\alpha)| \le 4\alpha$. Similarly for large enough d_2 we have

$$\begin{aligned}2H(\alpha)+\frac{\log(2d_2\alpha)}{2} &= -2\alpha\log(\alpha)+\frac{\log(\alpha)}{8}-2(1-\alpha)\log(1-\alpha)\\ &\le \frac{\log(\alpha)}{16}\ .\end{aligned}$$

Thus for large enough $d_2 \le d_1$

$$\mathbb{P}(R_\alpha) \le \frac{2^{3\alpha\log(\alpha)n}}{\sqrt{\alpha n}}.$$

Hence

$$\begin{aligned}\mathbb{P}\left(\bigcup_{\alpha\in I_1} R_\alpha\right) &\le \sum_{\alpha\in I_1}\frac{2^{3\alpha\log(\alpha)n}}{\sqrt{\alpha n}}\\ &\le n2^{-3\frac{1}{n}\log(n)n}\\ &\le \frac{1}{n},\end{aligned}\tag{4.16}$$

where $\alpha \in I_1 = (0, \frac{1}{d_2^2})$. The last term is derived using the following: The function $\alpha\log\alpha$ is decreasing from 0 to 1/2 and the least possible value of $\alpha = \frac{1}{n}$. Plugging this value of α we get the above.

CASE 2: $\frac{1}{d_2^2} \le \alpha \le \frac{C}{d_2}$.

Now clearly in this range of α, by stochastic domination $\mathbb{P}(Bin(d_2,\alpha)\ge 1)$ is maximized when $\alpha = \frac{C}{d_2}$. We now use the Poisson approximation of $Bin(d_2, \frac{2C}{d_2})$ to bound the probability $\mathbb{P}(Y \ge 1)$ by a universal constant c which is a function of C for all α in this range. Using this, we rewrite (4.14) to get

$$\begin{aligned}2H(\alpha)+\frac{d_1}{8}\alpha\log(c) &\le -2\alpha\log(\alpha)+\frac{d_1}{8}\alpha\log(c)-2(1-\alpha)\log(1-\alpha)\\ &\le -4\alpha\log(\alpha)+\frac{d_1}{8}\alpha\log(c)\\ &\le -5\alpha\end{aligned}$$

for large enough d_1. Similarly for large enough d_2 we have

$$2H(\alpha) + \frac{\log(c)}{2} \leq \frac{\log(c)}{4}.$$

Plugging in we get

$$\mathbb{P}\left(\bigcup_{\alpha \in I_2} R_\alpha\right) \leq \sum_{\alpha \in I_2} \frac{2^{-5\alpha n}}{\sqrt{\alpha n}} \leq n 2^{-\frac{5}{d_2^2} n}, \tag{4.17}$$

where $I_2 = [\frac{1}{d_2^2}, \frac{C}{d_2}]$. Thus the proof for the case when $\alpha \leq \frac{C}{d_2}$ is complete.

CASE 3: $\frac{C}{d_2} \leq \alpha \leq \frac{1}{2}$.

We first need a preliminary lemma. For $d_2 \in \mathbb{N}$ and $\alpha \in (0,1)$ let $Z_{d_2,p} \sim Bin(d_2, p)$.

Lemma 4.15. *There exists a constant C_1 such that for all large enough d_2*

$$\sup_{p \in (\frac{C_1}{d_2}, \frac{2}{3})} \sup_{1 \leq i \leq d_2} \mathbb{P}(Z_{d_2,\alpha} = i) \leq \frac{1}{400}.$$

Proof. It is a standard fact that for any d_2, α

$$\sup_{1 \leq i \leq d_2} \mathbb{P}(Z_{d_2,\alpha} = i) = \mathbb{P}(Z_{d_2,\alpha} = \lfloor (d_2+1)\alpha \rfloor).$$

Let $k = \lfloor (d_2+1)\alpha \rfloor$. We now estimate

$$\mathbb{P}(Z_{d_2,\alpha} = k) = \binom{d_2}{k} \alpha^k (1-\alpha)^{d_2-k}.$$

Since $k > C_1$ by hypothesis using Stirling's formula we have

$$\begin{aligned} \mathbb{P}(Z_{d_2,\alpha} = k) &= O\left(\frac{1}{\sqrt{k}} 2^{H(\alpha)d_2} 2^{-H(\alpha)d_2}\right) \\ &= O\left(\frac{1}{\sqrt{C_1}}\right) \leq \frac{1}{400} \end{aligned}$$

for large enough C_1. $\square$

We now need another lemma. Consider the exploration process for sampling the bipartite regular graph given by $\mathcal{A}, \mathcal{B}$ (sketched in Definition 2.1), where vertices of $\mathcal{A}$ are exposed one by one to find out the neighbors in $\mathcal{B}$. We do this first for each half edge incident to the vertices in A, followed by the half edges corresponding to the rest of the vertices in

$\mathcal{A}$. Let us parametrize time by the number of half edges. Consider the Bernoulli variable

$$B_t = \mathbf{1}(\text{the } t^{th} \text{ half edge is matched to a half edge in } B). \tag{4.18}$$

Now note that the first d_2 half edges correspond to $deg_B(v_1)$, the second d_2 half edges correspond to $deg_B(v_2)$, and so on. We now make a simple observation that the Bernoulli probabilities do not change much from time t to $t + d_2$. This then shows that $deg_B(v_i)$ are essentially Binomial variables with probability depending on the filtration at time (id_2). Formally, we have the following lemma: let $\mathcal{F}_i$ be the filtration generated up to time (id_2) (when all the half edges up to vertex i have been matched).

Lemma 4.16. *For any $i \leq \frac{n}{4}$ there exists a p_i which is $\mathcal{F}_{i-1}$-measurable such that*

$$||deg_B(v_i)|\mathcal{F}_{i-1}, Bin(d_2, p_i)||_{TV} = O\left(\frac{1}{n}\right),$$

where $||\cdot,\cdot||_{TV}$ denotes the total variation norm and the constant in the $O(\cdot)$ notation depends only on d_2.

Proof. To show this first note that the random variables B_t in (4.18) are Bernoulli variables with probability

$$\hat{p}_t = \frac{\alpha n d_2 - \sum_{j \leq t-1} B_j}{n d_2 - t} \ .$$

Then clearly for all $t \leq \frac{nd_2}{4}$, $|\hat{p}_t - \hat{p}_{t-1}| \leq \frac{4}{n}$. The proof thus follows since

$$deg_B(v_i) = \sum_{(i-1)d_2 < j \leq id_2} B_j \ .$$

□

Recall ℓ from Lemma 4.11. Now suppose $A \cup B$ is d_1-regular. Then by definition

$$\ell|A| \leq \sum_{i=1}^{|A|} deg_B v_i \leq d_2|B| = \alpha n d_2$$
$$\implies \ell \leq \frac{\alpha}{1-\alpha} d_2 \leq 2\alpha d_2.$$

Using the above we get that for all $j \leq \frac{n}{4}$:

$$\frac{\alpha n d_2 - j(\ell+1)}{n d_2 - j d_2} \geq \frac{\alpha n d_2 - \frac{n}{4}(3\alpha d_2)}{n d_2} \geq \frac{\alpha}{4}. \tag{4.19}$$

Above we used the fact that $\ell+1 \le 2\alpha d_2+1 \le 3\alpha d_2$ since $\alpha d_2 > C > 1$ by hypothesis. Also clearly for $j \le n/4$, since $\alpha \le 1/2$,

$$\frac{\alpha n d_2 - j\ell}{n d_2 - j d_2} \le 2/3. \tag{4.20}$$

Assume that all the $deg_B(v_i) \in \{\ell, \ell+1\}$. We have the following corollary.

Corollary 4.17. *For all $1 \le i \le n/4$, if $deg_B(v_j) \in \{\ell, \ell+1\}$, for some $\ell \le 2d_2\alpha$ for all $j \le i$ then there exists p_i which is $\mathcal{F}_{i-1}$ measurable such that*

$$||deg_B(v_i), Bin(d_2, p_i)||_{TV} = O\left(\frac{1}{n}\right)$$

where $\frac{\alpha}{4} \le p_i \le 2/3$.

Proof. The proof is immediate from (4.19), (4.20) and Lemma 4.16. □

We now complete the proof of Theorem 4.3 in the case $\alpha \in I_3 = [\frac{C}{d_2}, \frac{1}{2}]$. Using the same notation we used before we have:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{\alpha \in I_3} R_\alpha \mid \mathcal{A}\right) &\le \sum_{\alpha \in I_3} \sum_{A,B} \mathbb{P}(deg_B(v_i) = g_i) \\ &\le \sum_{\alpha \in I_3} \binom{n}{\alpha n}^2 \frac{1}{400^{n/4}} \\ &= \sum_{\alpha \in I_3} \frac{1}{\alpha n} 2^{2H(\alpha)n} \frac{1}{400^{n/4}} \\ &\le n \frac{2^{2n}}{400^{n/4}}. \end{aligned} \tag{4.21}$$

The first inequality is by the union bound. To see the second inequality observe first that by Lemma 4.11 it suffices to assume that $g_i^\prime s \in \{\ell, \ell+1\}$. Thus the second inequality follows by Corollary 4.17 and Lemma 4.15 as soon as

$$\frac{\alpha}{4} \ge \frac{C_1}{d_2}$$

which we ensure by choosing $C \ge 4C_1$.

Thus combining (4.16), (4.17) and (4.21) we have shown that

$$\mathbb{P}(\cup R_\alpha) \le \tau^n$$

for some $\tau = \tau(d_2) < 1$. Hence we are done. □

4.4 Theorem 4.4 and connection to the min-bisection problem

Throughout this section we always assume d_1 is even. We first remark that, under the hypothesis of Theorem 4.4, one can make a quick and simple connection to the min-bisection problem. It turns out that, in the case of the RSBM, the two problems are equivalent. More precisely, in the proof of Theorem 4.4 below, we show that the second eigenvalue of $\mathcal{G}(n, d_1, d_2)$ equals $d_1 - d_2$ with high probability, which implies that $\gamma = \frac{2d_2}{d_1+d_2}$ where γ is the spectral gap defined in (4.10). Hence, it follows by Theorem 4.14, that the size of the min bisection of $\mathcal{G}(n, d_1, d_2)$ is at least nd_2. Since the true partition $(\mathcal{A}, \mathcal{B})$ matches this lower bound, it solves the min-bisection problem.

We now proceed towards proving Theorem 4.4 for d_1 even. Recall the notion of random lifts from Section 4.2.1.1. We will now connect $\mathcal{G}(n, d_1, d_2)$ (RSBM) with random lifts of a certain small graph. Consider the following multigraph on two vertices: u and v, with d_2 edges between u and v and $d_1/2$ self loops at both the vertices (recall that d_1 is even).

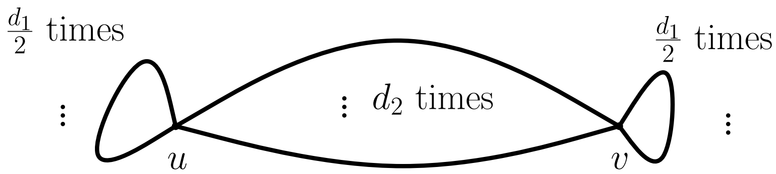

FIGURE 4.1: Multigraph lifting to $\mathcal{G}(n, d_1, d_2)$.

To randomly $n-$lift the above graph according to Section 4.2.1 we choose uniformly $d_1 + d_2$ many permutations:

$$\pi_1, \pi_2, \ldots \pi_{d_1}, \pi'_1, \pi'_2, \ldots \pi'_{d_2} \tag{4.22}$$

from S_n.

Let the lift be $\mathfrak{G}(n, d_1, d_2)$ on the vertex set $\{u, v\} \times \{1, 2, \ldots n\}$. We naturally identify it with $[2n] = \{1, 2, 3, 4, \ldots 2n\}$ with the first n numbers corresponding to $u \times \{1, 2, \ldots n\}$ and the rest corresponding to $v \times \{1, 2, \ldots n\}$.

Note that $\mathfrak{G}_1$, the subgraph induced by $\mathfrak{G}(n, d_1, d_2)$ on $[n]$ has edge set $(i, \pi_j(i))$ for $i \in [n]$ and $j \in [d_1/2]$. Similarly $\mathfrak{G}_2$, on $[2n] \setminus [n]$ has edges $(n + i, n + \pi_j(i))$ for $i \in [n]$ and $j \in [d_1] \setminus [d_1/2]$. The edges between $[n]$ and $[2n] \setminus [n]$ are the edges $(i, n + \pi'_j(i))$ for $i \in [n]$ and $j \in [d_2]$. Recall $\mathcal{G}(n, d_1, d_2)$ from Definition 4.1. A standard model to generate regular graphs is the well known configuration model, as also used in this article (see Section **??**) Now notice that $\mathfrak{G}(n, d_1, d_2)$ is essentially the same as $\mathcal{G}(n, d_1, d_2)$ except the graphs are now generated using permutations in (4.22). This is the Permutation model introduced in Chapter 2. We now use a well known result which says that the two models are contiguous, i.e. any event occurring a.a.s. in one of the models occurs a.a.s. in the other one as well (see [53]). [1]

We now prove Theorem 4.4. Let the graph in Figure 4.1 be called $\mathfrak{C}$. The adjacency matrix of $\mathfrak{C}$ is $A_* := \begin{bmatrix} d_1 & d_2 \\ d_2 & d_1 \end{bmatrix}$ with eigenvalues $d_1 + d_2$ and $d_1 - d_2$ and corresponding eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, respectively. Let $A_{*,n}$ be the adjacency matrix of $\mathfrak{G}(n, d_1, d_2)$ which as discussed above is a random $n-$lift of
$mathfrakC$. From the discussion in Section 4.2.1 we have the following:

- $d_1 + d_2$ and $d_1 - d_2$ are eigenvalues of $A_{*,n}$, with eigenvectors e and σ respectively (see Definition 4.9).
- By Theorem 4.8, for any $\varepsilon > 0$, a.a.s., all the other eigenvalues λ of $A_{*,n}$ satisfy $|\lambda| \leq 2\sqrt{d_1 + d_2 - 1} + \varepsilon$.

Let A_n be the adjacency matrix of $\mathcal{G}(n, d_1, d_2)$. That the first fact above holds for A_n as well is easy to check. Moreover, using the contiguity of the two models, A_n also has the second property a.a.s.. Note that finding the partition $(\mathcal{A}, \mathcal{B})$, in Definition 4.1 is equivalent to finding σ, (the eigenvector corresponding to the eigenvalue $d_1 - d_2$). Now under the hypothesis of Theorem 4.4, by the above discussion we see that $d_1 - d_2$ is the second eigenvalue which is also separated from the first and rest of the eigenvalues. Thus, we can efficiently compute a unitary eigenvector, w, associated to this eigenvalue. To assign the communities, put $v \in \mathcal{A}$ if and only if $w_v > 0$. Strong recovery is then achieved. This proves Theorem 4.4.

4.5 Complete reconstruction from partial reconstruction: proof of Theorem 4.7 and Theorem 4.4

In this section we prove Theorem 4.7. The idea is to show that, because of the rigid nature of the graph, if we initialize the partition with a large number of vertices labeled

[1] [53, Theorem 1.3] actually shows contiguity of regular graphs under configuration model and the permutation model. Note that $\mathcal{G}(n, d_1, d_2)$ and $\mathfrak{G}(n, d_1, d_2)$ are constructed from three independent regular graphs constructed using the configuration model and the permutation model. Since contiguity is preserved under taking product of measures, $\mathcal{G}(n, d_1, d_2)$ and $\mathfrak{G}(n, d_1, d_2)$ are contiguous.

correctly, one can bootstrap to deduce the true labels of even more vertices in the next step. We do this by looking at the majority of a vertex' neighbors. Recall **Majority** from Section 4.1.0.1. We prove that with high probability the graph $\mathcal{G}(n, d_1, d_2)$ is such that if the input (A, B) has a large overlap with the true partition $(\mathcal{A}, \mathcal{B})$, then one round of the algorithm reduces the number of wrongly labeled vertices by a constant factor. Thus it follows then that, with high probability, after $O(\log(n))$ iterations, no further corrections can be made and the algorithm outputs the true communities.

Lemma 4.18. *Assume* $d_1 > d_2 + 4$ *and let* $1/2 < \lambda < 1$*. Then there exists an* $\epsilon = \epsilon(d_1) > 0$ *such that, with probability* $1 - O(n^{1/2-\lambda})$*, the graph has the property that if* (A, B) *(the input) satisfies* $\min\{|A \cap \mathcal{A}|, |B \cap \mathcal{B}|\} > (1 - \epsilon)n$ *and if* $|\mathcal{A} \bigcap B| =: k$ *and* $|\mathcal{B} \bigcap A| =: k'$*, then*

$$|\mathcal{A} \bigcap B_1| \leq \lambda k \text{ and } |\mathcal{B} \bigcap A_1| \leq \lambda k' \ .$$

where (A_1, B_1) *is the output after one round of* ***Majority****.*

The constant in $O(\cdot)$ depends on d_1, λ, ϵ.

Proof. Let $v \in \mathcal{A} \bigcap B_1$ (that is, v has the wrong label after one iteration of **Majority**). We claim that v has more than two neighbors in $\mathcal{A} \bigcap B$, otherwise v will have at least $d_1 - 2$ neighbors in $\mathcal{A} \bigcap A$ and hence its label will be the sign of:

$$\sum_{i \sim v} \sigma_i^1 \geq d_1 - 2 - (d_2 + 2) > 0 \ ,$$

which contradicts the assumption that $v \in \mathcal{A} \bigcap B_1$. Thus the occurrence of the event $|\mathcal{A} \bigcap B_1| \geq \lambda k$ implies the occurrence of the event

$$\begin{gathered} E_k := \{\exists \text{ a subset } S \subset \mathcal{A}, \ |S| = \lambda k : \text{ any } v \in \\ S \text{ has at least three neighbors in } \mathcal{A} \bigcap B\} \ . \end{gathered}$$

Hence an upper bound on the probability of the event E_k will be an upper bound on the failure probability for **Majority** to reduce the size of the set of incorrectly labeled vertices in $\mathcal{A}$ by a fraction $1 - \lambda$.

We compute now an upper bound on the probability of E_k. By the exploration process (see Definition 2.1) it follows that for vertices in the set S, the degree sequence $\{deg_{(\mathcal{A} \bigcap B)}(v)\}_{v \in S}$ is stochastically bounded by a vector of i.i.d. binomial random variables $\{Z_v\}_{v \in S}$, i.e.,

$$\{deg_{(\mathcal{A} \bigcap B)}(v)\}_{v \in S} \preceq \{Z_v\}_{v \in S} \ , \quad \text{where } Z_v \sim Bin(d_1, \frac{k}{n - \lambda k}) \ .$$

By stochastic domination of vectors we mean the existence of a coupling of the two distributions such that the one vector is pointwise at most the other vector. As $\mathbb{P}(Z_v \geq 3) \leq \left(\frac{d_1 k}{n-\lambda k}\right)^3$, by union bound and counting the number of choices for all the possible sets $\mathcal{A}\bigcap B$ of size k and S of size λk, we obtain the following:

$$\mathbb{P}(E_k) \leq \binom{n}{k}\binom{n}{\lambda k}\Big(\frac{d_1 k}{n-\lambda k}\Big)^{3\lambda k} .$$

Adding over all possible k, we obtain

$$\mathbb{P}\left(\left|\mathcal{A}\bigcap B_1\right| \geq \lambda k \mid k \leq \epsilon n\right) \leq \sum_{k=1}^{\epsilon n}\binom{n}{k}\binom{n}{\lambda k}\Big(\frac{d_1 k}{n-\lambda k}\Big)^{3\lambda k} \tag{4.23}$$

$$\leq \sum_{k=1}^{\epsilon n}\Big(\frac{d_1^{3\lambda} e^{1+\lambda}}{\lambda^\lambda (1-\lambda)^{3\lambda}}\Big)^k \Big(\frac{k}{n}\Big)^{(2\lambda-1)k} \tag{4.24}$$

The last inequality follows by using the bound $\binom{n}{m} \leq \left(\frac{ne}{m}\right)^m$, as well as the fact that $n - \lambda n \leq n - \lambda k$. Denote now by $c = c(d_1) := \frac{d_1^{3\lambda} e^{1+\lambda}}{\lambda^\lambda (1-\lambda)^{3\lambda}}$.

We show now that the sum in (4.23) is $O(n^{1/2-\lambda})$. We split this sum into two parts, P_1 and P_2, the first representing the sum of all the terms corresponding to indices up to $\lfloor\sqrt{n}\rfloor$, and the second part representing the rest. For P_1, we obtain that

$$\begin{aligned} P_1 &= \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} c^k \Big(\frac{k}{n}\Big)^{(2\lambda-1)k} \leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} c^k n^{-(\lambda-1/2)k} \\ &\leq \sum_{k=1}^{\infty}\Big(\frac{c}{n^{\lambda-1/2}}\Big)^k \\ &\leq \frac{2c}{n^{\lambda-1/2}} . \end{aligned}$$

The last inequality is true for large n. To bound P_2, we note that $k/n \leq \epsilon$ and we write:

$$P_2 = \sum_{k=\lceil\sqrt{n}\rceil}^{\epsilon n} c^k \Big(\frac{k}{n}\Big)^{(2\lambda-1)k} \leq \sum_{k=\lceil\sqrt{n}\rceil}^{\infty} (c\epsilon^{2\lambda-1})^k \leq \frac{1}{1-c\epsilon^{2\lambda-1}}(c\epsilon^{2\lambda-1})^{\lceil\sqrt{n}\rceil}.$$

The last inequality above follows by choosing ϵ so that $c\epsilon^{2\lambda-1} < 1$. Hence the probability of event E_k is $O\left(n^{1/2-\lambda}\right)$. As the problem is symmetric in $\mathcal{A}$ and $\mathcal{B}$, it follows that a similar bound can be found for the event that $|\mathcal{B}\bigcap A_1| > \lambda k'$. Thus by union bound, the probability of both events is also $O\left(n^{1/2-\lambda}\right)$, and the proof of the lemma is complete. $\square$

4.5.1 Proof of Theorem 4.7

Let $\epsilon = \epsilon(d_1)$ as in Lemma 4.18. Initialize **Majority** as $(A_0, B_0) = (A, B)$ where A, B satisfy the conditions of Lemma 4.18. Denote by (A_i, B_i) the partition after the i^{th} iteration of **Majority** where A_i corresponds to the vertices labeled $+1$, i.e., (A_i, B_i) is the output of the algorithm when we initialize it with (A_{i-1}, B_{i-1}). Consider the random variables $X_i = \max\{|\mathcal{A} \bigcap B_i|; |\mathcal{B} \bigcap A_i|\}$. Note that $\{X_i = 0\}$ iff $\mathcal{A} = A_i$ (and thus $\mathcal{B} = B_i$). Also by the hypothesis $X_0 \leq \varepsilon n$, so Lemma 4.18 implies that

$$\mathbb{P}(X_i \leq \lambda^i k \ , \forall\ 1 \leq i) \geq 1 - O(n^{1/2-\lambda}).$$

Let now $t = \left\lceil \frac{\log(\epsilon n)^{-1}}{\log \lambda} \right\rceil$. Since the X_is are integer-valued random variables, we have

$$\mathbb{P}(X_t = 0) \geq 1 - O(n^{1/2-\lambda}) \ ,$$

which proves the theorem. □

Proof of Theorem 4.4. The proof is a straightforward corollary of Lemma 4.6 and Theorem 4.7. □

4.6 Proof of Lemma 4.6

Recall that we have used spectral properties of random lifts to prove Theorem 4.4 when d_1 is even, see section 4.4. The general proof relies on studying the matrix of self-avoiding walks (formally defined below) of the graph $\mathcal{G}(n, d_1, d_2)$. This is the same matrix used in [42] to prove the blockmodel threshold conjecture. This section adapts the techniques in that paper to the regular setting to prove Lemma 4.6.

In the case of the RSBM, the lack of edge independence increases the complexity of many of the calculations. On the other hand, the rigid nature of the model forces certain other calculations to be much easier for e.g. the size of small neighborhoods.

The key ingredient in the proof of Lemma 4.6 will be Proposition 4.20. Its proof hinges on two technical lemmas we present below. We give here the proof of Proposition 4.20, subject to these two lemmas, whose proofs we defer to Sections 4.6.1 and 4.6.2.

Let us recall the definition of a self-avoiding walk on a graph G. Given two vertices i and j and a length $l > 0$, a self-avoiding walk from i to j of length l is a graph path $(i = v_0, v_1, ..., v_l = j)$ such that $|\{v_0, v_1, ..., v_{l-1}\}| = l$.

We denote by $S^{(l)}$ the matrix whose entry $S^{(l)}_{ij}$ equals the number of self-avoiding walks of length l between i and j, for all $1 \leq i, j \leq 2n$.

Definition 4.19. We say that the sequence of unitary vectors $\{v_n\}_{n\geq 1}$ is asymptotically aligned with the sequence of unitary vectors $\{w_n\}_{n\geq 1}$ if:

$$\lim_{n\to\infty} | < v_n, w_n > | \quad = \quad 1. \tag{4.25}$$

Definition 4.19 means that, asymptotically, v_n and w_n are the same up to a factor of -1. Throughout the rest of the article let

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_{2n} \tag{4.26}$$

be the eigenvalues of $S^{(l)}$,

Proposition 4.20. *Assume* $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$*. Let* $l = c\log(n)$ *where* c *is a constant such that* $c\log(d_1 + d_2) < \frac{1}{4}$*. For any fixed* $\varepsilon > 0$ *the following three events happen with high probability as n grows:*

(a) $\lambda_1 = (d_1 + d_2)(d_1 + d_2 - 1)^{l-1} + o(1)$ *and any unitary eigenvector associated to* λ_1 *is asymptotically aligned with* e *(the vector of all ones).*

(b) There exists a constant $A > 0$ *such that* $\lambda_2 = A\alpha^l(1 + o(1))$*, where*

$$\alpha = \frac{d_1 - d_2 + \sqrt{(d_1 - d_2)^2 - 4(d_1 + d_2 - 1)}}{2} \; ;$$

any unitary eigenvector associated to λ_2 *is asymptotically aligned with* σ *(the vector of labels).*

(c) $|\lambda_k| \leq |\lambda_3| \leq n^{\epsilon}(d_1 + d_2)^{l/2}(1 + o(1))$*, for all* $3 \leq k \leq 2n$*.*

Remark 4.21. Note that as $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$,

$$d_1 - d_2 + \sqrt{(d_1 - d_2)^2 - 4(d_1 + d_2 - 1)} \geq d_1 - d_2 + 1 > 2\sqrt{d_1 + d_2 - 1} + 1 > 2\sqrt{d_1 + d_2} \, ,$$

the latter inequality being true as $d_1 + d_2 \geq 6$. This, in turn, means that

$$\alpha \quad = \quad \frac{d_1 - d_2 + \sqrt{(d_1 - d_2)^2 - 4(d_1 + d_2 - 1)}}{2} > \sqrt{d_1 + d_2} \, , \tag{4.27}$$

and as $l = O(\log n)$, by picking $0 < \epsilon < 1 - 4c\log(d_1 + d_2)$, (4.27) is enough to show that $\lim_{n\to\infty} |\lambda_3|/|\lambda_2| = 0$, so the first two eigenvalues of $S^{(l)}$ are separated from the bulk. Also note that $\alpha < d_1 - d_2 < d_1 + d_2 - 1$, so λ_1 and λ_2 are also separated from each other.

From part (b) of Proposition 4.20 one can see how to construct a labeling that recovers at least $(1-\epsilon)n$ vertices correctly, for any $\epsilon > 0$ and $n = n(\epsilon)$ large enough using an eigenvector associated to the second eigenvalue.

The next two lemmas contains estimates for the leading two eigenvalues and the corresponding eigenvectors which imply spectral separation. The proofs we defer to the next sections.

Lemma 4.22. *Let $S^{(l)}$ the matrix of self-avoiding walks of length $l = c\log(n)$. Recall e and σ from Definition 4.9 . Assume G is l–tangle–free. With high probability, the following two events happen:*

(i) $S^{(l)}e = (d_1+d_2)(d_1+d_2-1)^{l-1}e + \tilde{e}$ for a vector $\tilde{e}$ such that $||\tilde{e}||_2 = o(n)$,

(ii) there exists a constant $A = A(d_1,d_2) \in \mathbb{R}$ such that $S^{(l)}\sigma = A\alpha^l(1+o(1))\sigma + \tilde{\sigma}$ for a vector $\tilde{\sigma}$ such that $||\tilde{\sigma}||_2 = o(n)$, and for $\alpha = \frac{d_1-d_2+\sqrt{(d_1-d_2)^2-4(d_1+d_2-1)}}{2}$.

Lemma 4.23. *Let $1 \le m \le l$, and all the notations as above. For any unitary vector x such that $x'e = x'\sigma = 0$ the following holds with high probability:*

$$||S^{(m)}x||_2 \le (l+1)n^{\epsilon}(d_1+d_2)^{m/2}(1+o(1)) \ ;$$

here $\delta = c\log(d_1+d_2) < 1/4$ and ϵ is as in Proposition 4.20.

Using these two lemmas, we can prove the main result of this section.

Proof of Proposition 4.20. Recall (4.26). Let $\{w_i\}$ be an orthonormal basis of eigenvectors, with w_i associated with λ_i for all $1 \le i \le 2n$. From Lemma 4.22 and Lemma 4.23 we have, with high probability,

$$\sup_{|x|=1} |x'S^{(l)}x| = (d_1+d_2)(d_1+d_2-1)^{l-1} + o(1) \ ,$$

which implies that, with high probability, $\lambda_1 = (d_1+d_2)(d_1+d_2-1)^{l-1} + o(1)$.

As $\sigma \perp e$, it follows that

$$\sup_{|x|=1, x\perp e} |x'S^{(l)}x| \ge |\sigma' S^{(l)}\sigma| = A\alpha^l(1+o(1)) \ ;$$

on the other hand, Lemma 4.23 and the Courant-Fischer theorem (see [54]) guarantee that

$$\sup_{|x|=1, x\perp e,\sigma} |x'S^{(l)}x| = o(|\sigma' S^{(l)}\sigma|) \ .$$

This yields that, with high probability, $\lambda_2 = A\alpha^l(1 + +o(1))$. Finally, Lemma 4.23 also yields that $|\lambda_k| \leq (l+1)n^{\epsilon}(d_1+d_2)^l(1+o(1))$, for all $k \geq 3$.

We address now the issue of eigenvector alignment. Recalling the definition of alignment (4.25), let $\hat{e} = (\sqrt{2n})^{-1}e$. We can write

$$\hat{e} = \sum_{i=1}^{2n} c_i w_i \ ,$$

with $\sum_{i=1}^{2n} c_i^2 = 1$ (as $\hat{e}$ has unit norm). Our goal is to prove that $c_1 \to 1$. Note that $\tilde{w} = \sum_{i=2}^{2n} c_i w_i$ is perpendicular to w_1. Therefore we can write, as per Lemma 4.22

$$\begin{aligned}
(d_1+d_2)(d_1+d_2-1)^{l-1} + o(1) &= ||S^{(l)}\hat{e}||_2 \leq ||S^{(l)}c_1w_1||_2 + ||S^{(l)}\tilde{w}||_2 \\
&\leq c_1(d_1+d_2)(d_1+d_2-1)^{l-1} + o(1) + |\lambda_2| \ ,
\end{aligned}$$

where the last inequality is due to the Courant-Fischer theorem. As $\lambda_2/\lambda_1 \to 0$ with high probability as $n \to \infty$, it follows that, again with high probability, $c_1 \to 1$ and e and w_1 are asymptotically aligned.

Similarly, we show that $\hat{\sigma} = (\sqrt{2n})^{-1}\sigma$ and w_2 are asymptotically aligned; as σ and e are orthogonal and we just proved that e and w_1 are asymptotically aligned, if we write $\hat{\sigma} = \sum_{i=1}^{2n} a_i w_i$, it follows that $\lim\limits_{n\to\infty} a_1 = 0$. Let $w^* = \sum\limits_{i=3}^{2n} a_i w_i$.

Note that

$$\begin{aligned}
a_1\lambda_1 &= \lambda_1\langle w_1, \hat{\sigma}\rangle = \langle S^{(l)}w_1, \hat{\sigma}\rangle \\
&= \langle w_1, S^{(l)}\hat{\sigma}\rangle = A\alpha^l(1+o(1))\langle w_1, \hat{\sigma}\rangle + o(1) \\
&= a_1 A\alpha^l(1+o(1)) + o(1) \ .
\end{aligned}$$

The above implies that $a_1(\lambda_1 - A\alpha^l(1+o(1))) \to 0$, and since $\lambda_1 \gg A\alpha^l$, it follows that we have the much stronger statement $a_1\lambda_1 \to 0$.

Now we use Lemma 4.22 to write:

$$\begin{aligned}
A\alpha^l(1+o(1)) &= ||S^{(l)}\hat{\sigma}||_2 = ||S^{(l)}a_1w_1||_2 + ||S^{(l)}a_2w_2||_2 + ||S^{(l)}w^*||_2 \\
&\leq a_1\lambda_1 + o(1) + a_2A\alpha^l(1+o(1)) + |\lambda_3| \ ,
\end{aligned}$$

since $|\lambda_3| \ll A\alpha^l$ by Lemma 4.23, and we just showed that $a_1\lambda_1 \to 0$ as $n \to \infty$, it follows that $a_2 \to 1$ as $n \to \infty$, and with high probability the vectors σ and w_2 are aligned. This completes the proof of Proposition 4.20. □

As mentioned, the rest of this section is dedicated to proving Lemmas 4.22 and 4.23. In Section 4.6.1 we give a clear description of the neighborhood structure of $G(n, d_1, d_2)$ which leads to the proof of Lemma 4.22, and in Section 4.6.2 we use this neighborhood structure to obtain spectral bounds for $S^{(l)}$ via the moment method, and prove Lemma 4.23.

4.6.1 Proof of Lemma 4.22

We start by analyzing the local neighborhoods in $\mathcal{G}(n, d_1, d_2)$. We will havily use the fact that the graph is tangle–free.

Proof of Lemma 4.22. Let $\mathcal{T} = \{v \in V(G) : \ B_l(v) \text{ is a tree}\}$. Observe that if $S^{(l)}_{uv} > 0$ then $v \in B_l(u)$. Furthermore, if $u \in \mathcal{T}$ then

$$S^{(l)}_{uv} = \begin{cases} 1 & \text{if } v \in \partial B_l(u); \\ 0 & \text{else.} \end{cases} \tag{4.28}$$

If $v \in \mathcal{T}$,

$$(S^{(l)}e)_v = |\partial B_l(v)| = (d_1 + d_2)(d_1 + d_2 - 1)^{l-1} \ . \tag{4.29}$$

Write

$$S^{(l)}e = (d_1 + d_2)(d_1 + d_2 - 1)^{l-1}e + \tilde{e} \ ,$$

where $\tilde{e}$ is an error vector and note that, from (4.29), $\tilde{e}_v = 0$ if $i \in \mathcal{T}$. Note that for all u and v we have $S^l_{uv} \leq 2$, otherwise we have more than one cycle in $B_l(u)$, which contradicts the assumption that G is l–tangle–free. Using that $|B_l(u)| \leq (d_1 + d_2)^l$ we have for $v \notin \mathcal{T}$:

$$\tilde{e}_v \leq 2(d_1 + d_2)^l$$

Lemma 2.3 (b) implies that: $|\mathcal{T}^c| \leq n^\delta$ with high probability. Finally, by our choice of δ in Proposition 4.20 we conclude:

$$||\tilde{e}||_2 = o(n)$$

This proves part (i).

The calculation for part (ii) is slightly more complex. For a fixed vertex v and every $0 \leq k \leq l$ let

$$x_k(v) := |\{w : d(v,w) = k, \sigma_w = \sigma_v\}| \ , \qquad y_k(v) := |\{w : d(v,w) = k, \sigma_w = -\sigma_v\}|$$

and let

$$z_k(v) := x_k(v) - y_k(v).$$

Thus, $x_k(v)$ counts the number of vertices in the boundary of $B_k(v)$ with the same label as v and similarly, $y_k(v)$ counts the vertices in the boundary of $B_k(v)$ with label $-\sigma(v)$. The importance of these quantities is reflected in the following observation: if $v, v' \in \mathcal{T}$ then $x_k(v) = x_k(v')$ and $y_k(v) = y_k(v')$ for all $0 \le k \le l$, so $z_k(v) = z_k(v')$. Also, for *any* vertex v,

$$(S^{(l)}\sigma)_v = \sum_w S^{(l)}_{vw}\sigma_w = (x_l(v) - y_l(v))\sigma_v = z_l(v)\sigma_v \ .$$

Since with high probability all but a negligible number of vertices are in $\mathcal{T}$ and hence have the same $z_l(v)$, this relation suggests that σ is *almost* an eigenvector. We make this understanding rigorous in the claim below.

Claim 1. With the notation introduced before the following holds with high probability:

a) $S^{(l)}\sigma = z_l\sigma + \tilde{\sigma}$ where $z_l = z_l(v)$ for some (any) $v \in \mathcal{T}$, and $||\tilde{\sigma}||_2 = o(n)$.

b) Assume that the equation $x^2 - (d_1 - d_2)x + (d_1 + d_2 - 1) = 0$ has two distinct real roots (which is equivalent to the condition $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$) and denote the biggest root by α (trivially, $\alpha > 0$). Then there is a real constant $A > 0$ such that: $z_l = A\alpha^l(1 + o(1))$ as $n \to \infty$.

Proof. To prove part a) of the claim, let $\tilde{\sigma} = S^{(l)}\sigma - z_l\sigma$. We have $\tilde{\sigma}_v = 0$ if $v \in \mathcal{T}$; else

$$\tilde{\sigma}_v \le |(S^{(l)}\sigma)_v| + z_l(v) \le 2|B_l(v)| < (d_1 + d_2)^l \ .$$

By Lemma 2.3, $|\mathcal{T}^c| \le n^\delta$, with high probability, where $\delta < \frac{1}{4}$. Note that $n^\delta = (d_1 + d_2)^l$, and since $(d_1 + d_2)^l = o(n^{1/4})$, we can conclude that, with high probability,

$$||\tilde{\sigma}||_2 \le n^\delta(d_1 + d_2)^l = o(n) \ .$$

To prove part b), we actually compute z_l. We do this by finding a recurrence for x_k and y_k, which leads to a recurrence for z_k, which we can solve.

Consider a $(d_1 + d_2)$-regular rooted tree and the following labeling process on it: the root is labeled as $+1$. Among its neighbors, choose d_1 vertices uniformly and label them $+1$, and label the others -1. Continue the labeling process in such a way that for each vertex w in the tree, exactly d_1 neighbors have the same label as v. Denote by x_k (respectively, y_k) the number of vertices labeled $+1$ (respectively, -1) at distance k from the root. We have:

$$x_1 = d_1, \ y_1 = d_2, \ x_2 = d_1^2 + d_2^2 - d_1 - d_2, \ y_2 = 2d_1d_2$$

Fix $k \geq 3$; we have that

$$x_k = d_1 x_{k-1} + d_2 y_{k-1} - (d_1 + d_2 - 1) x_{k-2} \ .$$

To see this consider edges going 'out' of the $(k-1)$th level whose other endpoint is a $+1$. Clearly number of such edges is

$$d_1 x_{k-1} + d_2 y_{k-1}.$$

Now to compute the number of $+1$'s at the kth level one needs to subtract the number of edges going from $k-1$ to a $+1$ vertex in level $k-2$ since all the vertices at level k have exactly one edge connecting to level $k-1$. Now number of edges between $k-2$ and $k-1$ where the vertex at level $k-2$ is a $+1$ is $x_{k-2}(d_1+d_2-1)$ since each vertex at level $k-2$ have exactly d_1+d_2-1 edges going down.

By symmetry, using the same counting argument as above, we can obtain that

$$y_k = d_1 y_{k-1} + d_2 x_{k-1} - (d_1 + d_2 - 1) y_{k-2} \ .$$

Subtracting the two recurrences we obtain the recurrence for z_k:

$$z_k = (d_1 - d_2) z_{k-1} - (d_1 + d_2 - 1) z_{k-2} \ .$$

Hence, if α, β are the roots of $x^2 - (d_1 - d_2)x + (d_1 + d_2 - 1) = 0$, which we assume to be real and distinct, there are constants, A, B such that:

$$z_k = A\alpha^k + B\beta^k \ ;$$

A and B can be computed using z_1 and z_2, which are positive, and since z_k eventually will go to ∞, the fact that $A > 0$ follows. □

This finishes the proof of Lemma 4.22. □

4.6.2 Proof of Lemma 4.23

As was observed in [42], the spectrum of $S^{(l)}$ can be studied by relating it to the spectra of $S^{(r)}$ for $0 \leq r < l$. In fact, Theorem 2.2 of [42] is valid here as well; we will not present the proof, as it is applies verbatim, but we will introduce the notation and explain the quantities involved.

Consider the matrix:

$$\bar{A} := \frac{d_1}{n}\Big(\frac{1}{2}(ee' + \sigma\sigma') - I\Big) + \frac{d_2}{2n}(ee' - \sigma\sigma') \tag{4.30}$$

Let $\Delta^{(l)}$ be the matrix whose entries are given by

$$\Delta_{ij}^{(l)} := \sum \prod_{t=1}^{l} (A - \bar{A})_{i_{t-1} i_t}$$

where the sum is taken over all self-avoiding walks from i to j of length l. Finally, consider the matrix $\Gamma^{(l,m)}$, for $1 \le m \le l$, whose entries are given by

$$\Gamma_{ij}^{(l,m)} = \sum \prod_{t=1}^{l-m} (A - \bar{A})_{i_{t-1} i_t} \bar{A}_{i_{l-m} i_{l-m+1}} \prod_{t=l-m+2}^{l} A_{i_{t-1} i_t} \tag{4.31}$$

Here we sum over paths of length l obtained by concatenation of two self-avoiding walks of lengths $l - m$ and $m - 1$ respectively, the first starting at i and the second ending at j, with the additional constrain that they have non-empty intersection.

Theorem 2.2 in [42] gives the following equation

$$S^{(l)} = \Delta^{(l)} + \sum_{m=1}^{l} (\Delta^{(l-m)} \bar{A} S^{(m-1)}) - \sum_{m=1}^{l} \Gamma^{(l,m)} \; ; \tag{4.32}$$

In the decomposition above it turns out that the first and the third terms have small spectral norm and hence understanding the spectrum of $S^{(l)}$ becomes equivalent to understanding the spectrum of the middle term.

Throughout this section, unless otherwise noted, expectations are taken with respect to the randomness in the graph, *given a set of labels* σ. Later, the dependence on σ is removed with the help of Lemma 4.26.

To upper bound the moments of the trace of powers of $\Delta^{(l)}$ and $\Gamma^{(l,m)}$ we will need the estimates from Chapter 2.

Recall Definition 4.1. For any set $E \subset [2n] \times [2n]$ let $\mathcal{X}_E$ be the indicator of the event

$$\mathcal{X}_E := \mathbf{1}_{\{E \subset E(G)\}} \; ,$$

that is, E is a subset of the set of edges of the random graph G. Similarly we denote by $\mathcal{X}_{E^c}$ the indicator of the event

$$\mathcal{X}_{E^c} := \mathbf{1}_{\{E \cap E(G) = \emptyset\}} \; ,$$

when no edge in E is an edge of G. When E has one element we will use e instead of E.

As n Chapter 3, a particular important point related to using Lemma **??**, will be to examine the possible number of disjoint edges in an ordering of a given set of edges.

Given a set of ordered edges $\vec{E} = \{e_i\}$, we say edge e_j is *disconnected* if the sets $\{e_j\}$ and $\{e_1, e_2, ..., e_{j-1}\}$ are disconnected. We denote by $\delta(\vec{E})$ the number of *disconnected* edges of $\vec{E}$. Clearly this number depends on the order of the elements of $\vec{E}$.

Lemma 4.24. *Let G be a graph with maximal degree equal to d and E a subset of edges of G. Then there is an order of the elements of E such that:*

$$\delta(\vec{E}) \geq \left\lfloor \frac{|E|}{2d} \right\rfloor$$

Proof. We denote by $[|E|] = \{1, 2, \ldots, |E|\}$.

We claim that the following algorithm finds a bijection $\pi : [|E|] \to E$ with the required property: Choose an edge e of E and consider the subset $E(e)$ of all the edges of E that are adjacent to e. Since E is a subset of the edges of a graph with maximal degree d we have $|E(e)| < 2d$. We will add e and the edges in $E(e)$ at the end of the ordering; namely, we let $\pi(i) \in E(e)$ for $|E \setminus E(e)| + 1 \leq i \leq |E|$ and $\pi(|E \setminus E(e)|) = e$. We have used at most $2d$ edges. We now exclude all those edges from our set E and continue constructing the bijection by recursion, until no more edges exist. Note that if we add the edges in the order given by π the construction ensures that e is disconnected. Since each time we exclude at most $2d$ edges the results follows. $\square$

The following theorem is the analogous of Theorem 3.6.

Theorem 4.25. *Let E be a set of edges of $\mathcal{G}(n, d_1, d_2)$ with $|E| = K = O(\log(n))$. Let m_i, $i = 1, \ldots, |E|$ be positive integers. The following holds:*

$$\left| \mathbb{E}\left(\prod_{i=1}^{i=K} (A_{e_i} - \bar{A}_{e_i})^{m_i} | \sigma \right) \right| \leq \left(\prod_{i=1}^{i=K} \frac{d_{e_i}}{n} \right) \left(1 + O\left(\frac{\log(n)^2}{n} \right) \right) \left(\frac{K}{n} \right)^{\omega},$$

where $d_{e_i} = d_1$ if e_i has both endpoints in the same clusters and $d_{e_i} = d_2$ if not, and

$$\omega = 1 + \left\lfloor \frac{\sum_i \delta_{\{m_i=1\}}}{2d} \right\rfloor$$

if $\sum_i \delta_{\{m_i=1\}} > 0$ and $\omega = 0$ else.

Proof. For $1 \leq s \leq K$ write:

$$X_s = \prod_{i=1}^{i=s} (A_{e_i} - \bar{A}_{e_i})^{m_i}$$

Also denote by $\mathcal{G}_s$ the σ-algebra generated by $\{A_{e_1}, A_{e_2}, ..., A_{e_s}\}$. As a first step, we will show that

$$|\mathbb{E}(X_K|\sigma)| \leq \prod_{i=1}^{i=K} \left(\frac{d_{e_i}}{n} + O\left(\frac{\log(n)}{n^2}\right)\right) \left(\frac{K}{n}\right)^{\delta} .$$

Thus,

$$|\mathbb{E}(X_K|\sigma)| = |\mathbb{E}(\mathbb{E}(X_K|\sigma)|\mathcal{G}_{K-1})| = |\mathbb{E}((X_{K-1}|\sigma)\mathbb{E}((A_{e_K} - \bar{A}_{e_K})^{m_K}|\sigma)|\mathcal{G}_{K-1})| \quad (4.33)$$

The last equality follows by observing that X_{K-1} is $\mathcal{G}_{K-1}$-measurable. If $m_K > 1$, noting that $\mathcal{X}_{e_K} = A_{e_K}$, Lemma **??** implies:

$$\mathbb{E}(A_{e_K}|\sigma, \mathcal{G}_{K-1}) \leq \frac{d_{e_K}}{n} + O\left(\frac{\log(n)}{n^2}\right) .$$

If we now apply Lemma **??** with $q = \mathbb{E}(A_{e_K}|\sigma, \mathcal{G}_{K-1})$, $p = \frac{d_{e_K}}{n}$ and $r = O\left(\frac{\log(n)}{n^2}\right)$, we obtain

$$|\mathbb{E}((A_{e_K} - \bar{A}_{e_K})^{m_K}|\sigma)|\mathcal{G}_{K-1})| \leq \frac{d_{e_K}}{n} + O\left(\frac{\log(n)}{n^2}\right)$$

Substitute in (4.33) to obtain

$$|\mathbb{E}(X_K|\sigma)| \leq |\mathbb{E}(X_{K-1}|\sigma)| \left(\frac{d_{e_K}}{n} + O\left(\frac{\log(n)}{n^2}\right)\right) .$$

On the other hand, if $m_K = 1$ and e_K is *disconnected* then using the second part of Lemma **??** one can see that

$$|\mathbb{E}(A_{e_K} - \bar{A}_{e_K})|\sigma| \leq \frac{K}{n^2} .$$

To complete the proof, we reorder, if necessary, the edges of E such that we have the maximum possible number of *disconnected* edges with the property that the corresponding exponent $m_i = 1$. By Lemma 4.24 this is equal to ω, as defined in the proposition.

We conclude that,

$$|\mathbb{E}(X_K|\sigma)| \leq \prod_{i=1}^{i=K} \left(\frac{d_{e_i}}{n} + O\left(\frac{\log(n)}{n^2}\right)\right)\left(\frac{K}{n}\right)^{\omega} = \left(\prod_{i=1}^{i=K} \frac{d_{e_i}}{n}\right)\left(1 + O\left(\frac{\log(n)^2}{n}\right)\right)\left(\frac{K}{n}\right)^{\omega} .$$

The proof is complete. □

The next lemma considers the expectation under the measure generated by the labels.

Lemma 4.26. *Let* (T, o) *a subtree of* $\mathcal{G}(n, d_1, d_2)$ *with at most* $O(\log(n))$ *many edges. Then:*

$$\mathbb{E}_\sigma\left(\prod_{e\in T}\frac{d_e}{n}\right) \leq \left(\frac{d_1+d_2}{2n}\right)^{|T|}\left(1+O\left(\frac{\log(n)^2}{n}\right)\right)$$

where $\mathbb{E}_\sigma$ *indicates we are taking the expectation over the measure generated by the labels.*

Proof. Let w be a leaf of T. Let $\mathcal{F}_w$ be the σ-algebra generated by the labels of all vertices in T but σ_w. We have:

$$\mathbb{E}_\sigma\Big(\prod_{e\in T}\frac{d_e}{n}\Big) = \mathbb{E}_\sigma\Big(\mathbb{E}\Big(\prod_{e\in T}\frac{d_e}{n} \mid \mathcal{F}_w\Big)\Big) = \mathbb{E}_\sigma\Big(\prod_{e\in T, e\neq \bar{e}}\frac{d_e}{n}\mathbb{E}\Big(\frac{\bar{e}}{n} \mid \mathcal{F}_w\Big)\Big)$$

Now we check that $Pp_\sigma(d_{\bar{e}} = d_i) \leq 1/2 + O\Big(\frac{\log(n)}{n}\Big)$. Given any event on $\mathcal{F}_w$, this is, any labeling of the vertices of T expect for w, denote by s^+ the number of positive labels and by s^- the number of negative labels. Recalling that $s = s^+ + s^- = O(\log(n))$, we have:

$$\mathbb{P}_\sigma(d_{\bar{e}} = d_1) = \frac{\binom{2n-s-1}{n-s^+-1}}{\binom{2n-s}{n-s^+}} = \frac{n-s^+}{2n-s} \leq \frac{1}{2} + O\Big(\frac{\log(n)}{n}\Big)$$

An analogous bound holds for $\mathbb{P}(d_{\bar{e}} = d_2)$. We conclude that

$$\mathbb{E}\Big(\frac{\bar{e}}{n} \mid \mathcal{F}_w\Big) \leq \frac{d_1+d_2}{2n} + O\Big(\frac{\log(n)}{n^2}\Big)$$

Repeating this argument we get:

$$\mathbb{E}_\sigma\left(\prod_{e\in T}\frac{d_e}{n}\right) \leq \Big(\frac{d_1+d_2}{2n} + O\Big(\frac{\log(n)}{n^2}\Big)\Big)^{|T|} \leq \Big(\frac{d_1+d_2}{2n}\Big)^{|T|}\left(1+O\left(\frac{\log(n)^2}{n}\right)\right)$$

for n large. This completes the proof. □

We now have enough tools to examine the spectral radius of $\Delta^{(l)}$, which we denote by $\rho(\Delta^{(l)})$. For any integer k we have $\rho(\Delta^{(l)})^{2k} \leq \mathrm{Tr}((\Delta^{(l)})^{2k})$ and hence the same inequality holds if one takes expectation. From the definition of $\Delta^{(l)}$ we have:

$$\mathbb{E}(Tr((\Delta^{(l)})^{2k})) = \sum_{c\in\mathcal{C}}\mathbb{E}(\prod(A_{e_i} - \bar{A}_{e_i})^{m_i}) \tag{4.34}$$

where $\mathcal{C}$ is the collection of cycles in $\mathcal{G}(n, d_1, d_2)$ of length $2kl$ obtained from the concatenation of $2k$ self-avoiding walks of length l. Also, m_i is the number of times the edge e_i is traversed in one of such cycle. To bound the expectation from above we need to bound the number of such cycles; this was done in [42], but we include the argument here for the sake of completeness.

The idea is to bound the number of cycles in $\mathcal{C}$ with v vertices and e edges. Given one of these cycles, number the vertices by the order they appear in the cycle, starting at 1, and denote by $\mathcal{T}$ the tree of those edges of the cycle which lead to new vertices.

It is crucial to note that the listing of a cycle is in order and thus it tells us how it was traversed, so the above enumeration and $\mathcal{T}$ are well defined.

Recall that each cycle is the concatenation of $2k$ self-avoiding walks of length l. We will break each path into three types of sub-paths and then we encode these sub-paths. To do so, we start traversing the cycle, and we check each time we found one of the sub-paths described above. Given our position on the cycle and the tree of the previously discovered vertices we represent each type as follows:

Type1 These are paths with the property that all their edges are edges of $\mathcal{T}$ and have been traversed already in the cycle. They can be encoded by its end vertex. This is because our sub-path is part of a self-avoiding walk, and it is a path contained in a tree. Given its initial and its final vertex there will be exactly one such path. We use 0 if the path is empty.

Type2 These are the paths with the property that all their edges are edges of $\mathcal{T}$ but they are traversed for the first time in the cycle. We can encode these paths by its length, since they are traversing new edges and we know in what order the vertices are discovered. We use 0 if the path is empty.

Type3 This is just an edge that connects the end of a path of type 1 or 2 to a vertex that has been already discovered. Given our position on the cycle, it is clear we can encode an edge by its final vertex. Again, we use 0 if the path is empty.

Now we decompose each self-avoiding walk into sequences characterizing its sub-paths:

$$(p_1, q_1, r_1)(p_2, q_2, r_2)(...)(p_t, q_t, r_t)$$

Here, p_i characterizes sub-paths of type 1, q_i characterizes subpaths of type 2 and r_i characterizes sub-paths of type 3.

Note that p_i and r_i are both numbers in $\{0, 1, ..., v\}$, since our cycle has v vertices. On the other hand, $q_i \in \{0, 1, ..., l\}$ since it represents the length of a sub-path of a self-avoiding walk of length l. Hence, there are $(v+1)^2(l+1)$ different triples.

We must now see in how many ways we can concatenate sub-paths encoded by the triples to form a cycle. First, note that $r_i = 0$ only if (p_i, q_i, r_i) is at the end of a self-avoiding walk. Hence, all other triples indicate the traversal of an edge not in $\mathcal{T}$. There are $e-v+1$ such edges and each of it can be traversed at most $2k$ times in the cycle. Hence there are at most $((v+1)^2(l+1))^{2k(e-v+1)}$ triples with $r_i > 0$ and there are at most $((v+1)^2(l+1))^{2k}$ triples with $r_i = 0$.

We conclude there are at most

$$C_{v,e} := ((v+1)^2(l+1))^{2k(1+e-v+1)} \tag{4.35}$$

cycles with v vertices and e edges.

Recall that we want to bound the right hand side of (4.34). Denote by $v(c)$ the number of vertices visited by the cycle c. Let us split $\mathcal{C}$ into three subsets $\mathcal{C}_j,\ j = 1,2,3$ as follows:

- $\mathcal{C}_1 := \{c \in \mathcal{C} :$ all edges in c are traversed at least twice.$\}$
- $\mathcal{C}_2 := \{c \in \mathcal{C} :$ at least one edge in c is traversed exactly once and $v(c) \leq kl+1.\}$
- $\mathcal{C}_3 := \{c \in \mathcal{C} :$ at least one edge in c is traversed exactly once and $v(c) > kl+1.\}$

Clearly $\mathcal{C} = \bigcup C_j$.

For $j = 1,2,3$, let

$$I_j = \sum_{c \in \mathcal{C}_j} \left| \mathbb{E}\left(\prod (A_{e_i} - \bar{A}_{e_i})^{m_i} \right) \right|$$

From (4.34) we then can write:

$$\mathbb{E}(Tr(\Delta^{(l)})^{2k}) \leq I_1 + I_2 + I_3. \tag{4.36}$$

We will bound each I_j separately. For I_1 we have by Proposition 4.25:

$$I_1 \leq \sum_{c \in \mathcal{C}_1} \left(\prod_{i=1}^{e(c)} \frac{d_{e_i}}{n} \right) \left(1 + O\left(\frac{\log(n)^2}{n} \right) \right) ,$$

where $e(c)$ denote the number of different edges traversed by the cycle c. Note that since all edges in cycles of $\mathcal{C}_1$ are traversed at least twice we have $\omega = 0$. The same condition implies that each of these cycles have at most kl different edges, since its total length is

$2kl$, and at most $kl+1$ vertices, since each c is connected. Use (4.35) to get:

$$I_1 \le \sum_{v=l+1}^{kl+1} \sum_{e=v-1}^{kl} (2n)^v [(v+1)^2(l+1)]^{2k(1+e-v+1)} \left(\prod_{i=1}^{e} \frac{d_{e_i}}{n} \right) \left(1 + O\left(\frac{\log(n)^2}{n}\right)\right).$$

Note that the right hand side depends on the label of the graph. We will average under the randomness induced by the label in the following way: for each cycle c, recall that $\mathcal{T}$ is the tree of spanned vertices, this is the tree of those edges which discover new vertices when traversed. For any edge $e \in c$ not in $\mathcal{T}$ use the bound $d_e \le (d_1 \vee d_2)$. Now take expectation with respect to the labels over $\mathcal{T}$. From Lemma 4.26 we conclude:

$$\begin{aligned} I_1 &\le \sum_{v=l+1}^{kl+1} \sum_{e=v-1}^{kl} 2(2n)^v [(v+1)^2(l+1)]^{2k(1+e-v+1)} \left(\frac{d_1+d_2}{2n}\right)^{v-1} \left(\frac{(d_1 \vee d_2)}{n}\right)^{e-v+1} \\ &\le 4kl(1+o(1))[(kl+2)^2(l+1)]^{2k} n (d_1+d_2)^{kl}. \end{aligned} \tag{4.37}$$

To explain the second inequality, note that since $kl = O(\log n)$ the only terms that are asymptotically significant are the ones for which $e - v + 1 = 0$. We now bound from above by kl times the highest term.

Using the same kind of reasoning and Proposition 4.25 we obtain the following bound for I_2:

$$\begin{aligned} I_2 &\le \sum_{v=l+1}^{kl+1} \sum_{e=v}^{kl+1} 2(2n)^v [(v+1)^2(l+1)]^{2k(1+e-v+1)} \left(\frac{d_1+d_2}{2n}\right)^{v-1} \left(\frac{(d_1 \vee d_2)}{n}\right)^{e-v+1} \\ &\le 4kl(1+o(1))[(kl+2)^2(l+1)]^{4k} (d_1 \vee d_2)(d_1+d_2)^{kl}. \end{aligned} \tag{4.38}$$

Note that now $e \ge v$ since each c is a closed path and at least one edge is traversed exactly once, and we have used the trivial bound $\left(\frac{e(c)}{n}\right)^{\omega(c)} \le 1$.

To bound I_3, note that for each $c \in \mathcal{C}_3$, from Proposition 4.25:

$$|\mathbb{E}(\prod_{i=1}^{e(c)} (A_{e_i} - \bar{A}_{e_i})^{m_i})| \le \left(\prod_{i=1}^{e(c)} \frac{d_{e_i}}{n} \right) \left(1 + O\left(\frac{\log(n)^2}{n}\right)\right) \left(\frac{e(c)}{n}\right)^{\omega(c)}. \tag{4.39}$$

The notation $\omega(c)$ above is to indicate that the value of ω from Proposition 4.25 depends on the cycle c and the order of the edges $\{e_i\}$.

Note that the right hand side of (4.39) is decreasing in ω. Our strategy will be to show that if $v(c)$ is large then $\omega(c)$ is also large and thus the right hand side in (4.39) is small.

More precisely, let $c \in \mathcal{C}_3$ be a cycle with $v(c) = kl + t$ and denote by $\tilde{e}(c)$ the number of edges that are traversed exactly one in c. We have $e(c) \ge v(c)$. Since $e(c) - \tilde{e}(c)$ edges are traversed at least two times and the length of c is $2kl$ we have:

$$\tilde{e}(c) + 2(e(c) - \tilde{e}(c)) \leq 2kl$$

which implies $\tilde{e}(c) \geq 2t$. By Lemma 4.24 we get:

$$\omega(c) \geq \frac{t}{d_1 + d_2}. \tag{4.40}$$

Combining (4.35), (4.39) and (4.40) we get:

$$I_3 \leq \sum_{v=kl+2}^{2kl} \sum_{e=v}^{2kl} 2(2n)^v[(v+1)^2(l+1)]^{2k(1+e-v+1)} \left(\frac{d_1+d_2}{2n}\right)^{v-1} \left(\frac{(d_1 \vee d_2)}{n}\right)^{e-v+1} \left(\frac{e}{n}\right)^{\frac{v-kl}{d_1+d_2}}.$$

Rewrite the right hand side above as:

$$\sum_{v=kl+2}^{2kl} \sum_{e=v}^{2kl} 4(d_1 \vee d_2)[(v+1)^2(l+1)]^{4k} \left(\frac{(d_1 \vee d_2)[(v+1)^2(l+1)]^{2k}}{n}\right)^{e-v} (d_1+d_2)^{v-1} \left(\frac{e}{n}\right)^{\frac{v-kl}{d_1+d_2}}.$$

We have:

$$\left(\frac{(d_1 \vee d_2)[(v+1)^2(l+1)]^{2k}}{n}\right)^{e-v} \leq 1$$

for n large. Note that the numerator is bounded by some polynomial in $\log(n)$.

Also note that

$$(d_1+d_2)^{v-1} \left(\frac{e}{n}\right)^{\frac{e-kl}{d_1+d_2}} \leq (d_1+d_2)^{kl-1} \left((d_1+d_2)\left(\frac{e}{n}\right)^{\frac{1}{d_1+d_2}}\right)^{v-kl} \leq (d_1+d_2)^{kl-1}.$$

We conclude that

$$\begin{aligned} I_3 &\leq \sum_{v=kl+2}^{2kl} \sum_{e=v}^{2kl} 4(d_1 \vee d_2)[(v+1)^2(l+1)]^{4k}(d_1+d_2)^{kl-1} \\ &\leq 4(kl)^2(d_1 \vee d_2)[(2kl+1)^2(l+1)]^{4k}(d_1+d_2)^{kl-1}. \end{aligned} \tag{4.41}$$

Substitute (4.37), (4.38) and (4.41) in (4.36), and note that the bounds for I_2 and I_3 are negligible compared to the one for I_1, we see that

$$\mathbb{E}(\rho(\Delta^{(l)})^{2k}) \leq 12kl(1+o(1))[(kl+1)^2(l+1)]^{2k} n(d_1+d_2)^{kl}.$$

Finally, given ϵ choose k such that $2k\epsilon > 1$. We can now apply Markov's Inequality and obtain the desired bound on $\rho(\Delta^{(l)})$:

$$\begin{aligned}
\mathbb{P}(\rho(\Delta^{(l)}) \geq n^{\epsilon}(d_1+d_2)^{l/2}) &\leq \frac{\mathbb{E}(\rho(\Delta^{(l)})^{2k})}{n^{2k\epsilon}(d_1+d_2)^{kl}} \\
&\leq 12kl(1+o(1))[(kl+2)^2(l+1)]^{2k} n^{1-2k\epsilon} \\
&= O(l^{6k+1}n^{1-2k\epsilon}) = o(1)\ .
\end{aligned}$$

More generally, the same counting arguments and Markov Inequality can be used to show the following probability bound for $\Delta^{(l-m)}$, for all $m = 1, 2, \dots, l$:

$$\mathbb{P}\left(\rho(\Delta^{(l-m)}) \geq n^{\epsilon}(d_1+d_2)^{l/2}\right) \leq O\left((l-m)^{6k+1}n^{1-2k\epsilon}\right) = o(1)\ . \tag{4.42}$$

Let us now turn our attention to the spectral radii of $\Gamma^{(l,m)}$ (recall (4.31)).

Denote the spectral radio of each such matrix by $\rho(\Gamma^{(l,m)})$. For any positive integer k, we have:

$$\mathbb{E}(\rho(\Gamma^{(l,m)})^{2k}) \leq \mathbb{E}(Tr((\Gamma^{(l,m)})^{2k})) = \sum_{c\in\mathcal{D}} \mathbb{E}(\prod (M_{e_i})^{m_i}).$$

The right hand side is the sum over the set $\mathcal{D}$ of cycles c of length $2kl$ each of which is obtaining by concatenation of $2k$ paths, with each of those paths being a concatenation of two self-avoiding walks of length $l-m$, respectively, $m-1$, and with non-empty intersection. The entries M_{e_i} correspond to either $(A-\bar{A})_{e_i}$, $\bar{A}_{e_i}$ or A_{e_i} and m_i is the number of times the edge e_i is traversed in the cycle c.

We want to bound the number of such cycles. The same representation from [42] we used to count the cycles in $\mathcal{C}$ gives the following bound for the number of such cycles with v vertices and e edges:

$$D_{e,v} := v^{2k}[(v+1)^2(l+1)]^{4k(1+e-v+1)}. \tag{4.43}$$

Note that we have at least $(m \vee l-m+1)$ different vertices, since there are at least two self-avoiding walks of length $(m-1)$, respectively $(l-m)$, in each cycle; there are at most $2k(l-1)$ vertices because each length l path is the concatenation of two self-avoiding walks with non-empty intersection.

Let c be one of these cycles; we need to estimate $\mathbb{E}(\prod (M_{e_i})^{m_i})$.

We know that exactly $2k$ edges of c contributed $\bar{A}_{e_i}$, counting multiplicity. We can bound their contribution by $\left(\frac{(d_1 \vee d_2)}{n}\right)^{2k}$. What is left, for each e_i, has the form $A_{e_i}^{n_i}(A-\bar{A})_{e_i}^{m_i}$.

Here n_i is the number of times the edge e_i is weighted by A_{e_i} and m_i is the number of times the same edge is weighted by $(A-\bar{A})_{e_i}$. If $n_i > 0$, because $A_{e_i}^{n_i} = A_{e_i}$,

$$A_{e_i}^{n_i}(A-\bar{A})_{e_i}^{m_i} = A_{e_i}(1-\bar{A})_{e_i}^{m_i} \ ;$$

hence

$$\mathbb{E}(A_{e_i}^{n_i}(A-\bar{A})_{e_i}^{m_i}) \leq \mathbb{E}(A_{e_i}(1-\bar{A})_{e_i}^{m_i}) \leq \frac{d_{e_i}}{n} \ .$$

If $n_i = 0$ we can use Proposition 4.25 to bound the term directly. Combining these bounds with (4.43) and use Lemma 4.26 to get the following bound:

$$\mathbb{E}(Tr((\Gamma^{(l,m)})^{2k})) \leq 8kl(1+o(1))n[(2k(l-1)+2)^5(l+1)^2]^{2k}\left(\frac{(d_1+d_2)^l}{n}\right)^{2k} . \quad (4.44)$$

We have employed the same considerations here as in (4.37), in particular we noticed that

$$\frac{(d_1+d_2)^l}{n} \leq 1 \ ,$$

because of our choice of l (see Proposition 4.20).

Given ϵ, choose k such that $2k\epsilon > 1$ and use Markov's inequality again to get:

$$\begin{aligned} \mathbb{P}(\rho(\Gamma^{(l,m)}) \geq n^\epsilon) &\leq \frac{\mathbb{E}(\rho(\Gamma^{(l,m)})^{2k})}{n^{2k\epsilon}} \\ &\leq kl(1+o(1))[(2k(l-1)+2)^5(l+1)^2]^{2k} n^{1-2k\epsilon} = o(1) \ . \end{aligned}$$

Remark 4.27. Note that in the case of each spectral bound for $\Delta^{(l-m)}$ of $\Gamma^{(l,m)}$ for $1 \leq m \leq l$ we showed, the probability of the spectral radius being larger than the bound decays roughly like $n^{1-2k\epsilon}$ for k large enough. Since the total number of such bounds is $O(l)$, so logarithmic, we can conclude that all of them happen simultaneously with high probability.

Finally, we have all the tools to prove Lemma 4.23.

Proof of Lemma 4.23. We have shown that $S^{(l)} = \sum_{m=1}^{l} \Delta^{(l-m)}\bar{A}S^{(m-1)} + E$, where E is a small-spectral-radius perturbation (the sum of $\Delta^{(l)}$ and $\Gamma^{(l,m)}$ for $m = 1, \ldots, l$). We will now focus our attention on the remaining (significant) term.

Let $\mathcal{T}_m$ be the set of vertices which m-neighborhood is a tree, we have, for $i \in \mathcal{T}_m$

$$(S^{(m)}e)_i = \sum_{j=1}^{2n} S_{ij}^{(m)} = |\partial B_m(i)| = (d_1+d_2)(d_1+d_2-1)^{m-1}.$$

For $i \notin \mathcal{T}_m$:

$$(S^{(m)}e)_i = \sum_{j=1}^{2n} S_{ij}^{(m)} \leq 2|B_m(i)| \leq 2(d_1+d_2)^m,$$

since the l–tangle–freeness of the graph implies that $S_{ij}^{(m)} \leq 2$ for each i and j.

We then have

$$|e'S^{(m-1)}x| = \left|\sum_{i=1}^{2n} x_i(S^{(m-1)}e)_i\right| \leq \left|\sum_{i\notin\mathcal{T}_m} x_i(S^{(m-1)}e)_i\right| + \left|\sum_{i\in\mathcal{T}_m} x_i(d_1+d_2)(d_1+d_2-1)^{m-1}\right|$$

$$= \left|\sum_{i\notin\mathcal{T}_m} x_i(S^{(m-1)}e)_i\right| + \left|\sum_{i\notin\mathcal{T}_m} x_i(d_1+d_2)(d_1+d_2-1)^{m-1}\right|$$

The last equality uses the fact that $x'e = 0$. Using Lemma 2.3 and Cauchy-Schwarz's inequality, we obtain that

$$|e'S^{(m-1)}x| \leq 3n^{\delta/2}(d_1+d_2)^m \ . \tag{4.45}$$

The proof that

$$|\sigma'S^{(m-1)}x| \leq 3n^{\delta/2}(d_1+d_2)^m \tag{4.46}$$

is analogous.

To prove the inequality of the lemma, namely that

$$||S^{(l)}x||_2 \leq n^{\epsilon}(d_1+d_2)^{l/2}(1+o(1)) \ ,$$

recall the matrix $\bar{A}$ defined in (4.30) and decomposition (4.32). We have shown (see Remark 4.27) that

$$\max\{\rho(\Delta^{(l)}), \rho(\Gamma^{(l,m)})\} \leq n^{\epsilon}(d_1+d_2)^{l/2} \ ,$$

with high probability. Then

$$||\bar{A}S^{(m-1)}x||_2 \leq \frac{d_1}{n}||S^{(m-1)}x||_2 + O\left(n^{-1/2}\left(|e'S^{(m-1)}x| + |\sigma'S^{(m-1)}x|\right)\right) \ .$$

Bound the spectral radii of $S^{(m-1)}$ by $O((d_1+d_2)^{m-1})$ and use (4.45) and (4.46) to get:

$$\begin{aligned} ||\bar{A}S^{(m-1)}x||_2 &\leq O(n^{-1}(d_1+d_2)^{m-1}) + O(n^{-1/2+\delta/2}(d_1+d_2)^m) \\ &= O(n^{-1/2+\delta/2}(d_1+d_2)^m) \\ &= O(n^{-1/2+\delta/2}(d_1+d_2)^l) \\ &= o(1) \ , \end{aligned}$$

as $(d_1 + d_2)^{l/2} = n^\delta$, and $\delta < 1/4$.

Finally, using (4.42) and putting it all together,

$$||S^{(l)}x|| \leq (l+1)n^\epsilon(d_1+d_2)^{l/2}+\sum_{m=1}^{l} n^\epsilon(d_1+d_2)^{(l-m)/2}o(1) = (l+1)n^\epsilon(d_1+d_2)^{l/2}(1+o(1)) \,.$$

With this, the proof of Lemma 4.23 is completed. □

Bibliography

[1] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[2] Daniel A Spielman. Spectral graph theory and its applications. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 29–38. IEEE, 2007.

[3] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[4] Giuliana Davidoff, Peter Sarnak, and Alain Valette. *Elementary number theory, group theory and Ramanujan graphs*, volume 55. Cambridge University Press, 2003.

[5] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[6] Alon Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2): 207–210, 1991.

[7] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[8] Grigorii Aleksandrovich Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy peredachi informatsii*, 24(1):51–60, 1988.

[9] Moshe Morgenstern. Existence and explicit constructions of q+ 1 regular ramanujan graphs for every prime power q. *Journal of Combinatorial Theory, Series B*, 62(1): 44–62, 1994.

[10] Joel Friedman. A proof of alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724. ACM, 2003.

[11] Charles Bordenave. A new proof of friedman's second eigenvalue theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[12] Harry Kesten. Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92(2):336–354, 1959.

[13] Joel Friedman et al. Relative expanders or weakly relatively ramanujan graphs. *Duke Mathematical Journal*, 118(1):19–35, 2003.

[14] Alon Amit, Nathan Linial, Jiří Matoušek, and Eyal Rozenman. Random lifts of graphs. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 883–894. Society for Industrial and Applied Mathematics, 2001.

[15] Bojan Mohar. A strengthening and a multipartite generalization of the alon-boppana-serre theorem. *Proceedings of the American Mathematical Society*, 138 (11):3899–3909, 2010.

[16] Chris D Godsil and Bojan Mohar. Walk generating functions and spectral measures of infinite graphs. *Linear Algebra and its Applications*, 107:191–206, 1988.

[17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *arXiv preprint arXiv:1703.10146*, 2017.

[18] Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V Tran. Recovery and rigidity in a regular stochastic block model. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1589–1601. Society for Industrial and Applied Mathematics, 2016.

[19] Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V Tran. Recovery and rigidity in a regular stochastic block model. *arXiv preprint arXiv:1507.00930*, 2015.

[20] E.A. Bender and E.R. Canfield. The asymptotic number of non-negative integer matrices with given row and column sums. *J. Combinatorial Theory*, 24:296–307, 1978.

[21] Béla Bollobás. *Random Graphs, volume 73 of Cambridge studies in advanced mathematics.* Cambridge University Press, Cambridge, 2001.

[22] Joel Friedman. *A proof of Alon's second eigenvalue conjecture and related problems.* American Mathematical Soc., 2008.

[23] Eyal Lubetzky, Allan Sly, et al. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

[24] Brendan D McKay. Subgraphs of random graphs with specified degrees. 1981.

[25] Doron Puder. Expansion of random graphs: New proofs, new results. *Inventiones mathematicae*, pages 1–64, 2015.

[26] Andrei Broder and Eli Shamir. On the second eigenvalue of random regular graphs. In *Foundations of Computer Science, 1987., 28th Annual Symposium on*, pages 286–294. IEEE, 1987.

[27] Joel Friedman. On the second eigenvalue and random walks in randomd-regular graphs. *Combinatorica*, 11(4):331–362, 1991.

[28] Ki-ichiro Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. *Automorphic forms and geometry of arithmetic varieties.*, pages 211–280, 1989.

[29] Motoko Kotani and Toshikazu Sunada. Zeta functions of finite graphs. 2000.

[30] Hyman Bass. The ihara-selberg zeta function of a tree lattice. *International Journal of Mathematics*, 3(06):717–797, 1992.

[31] Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 367(6):4287–4318, 2015.

[32] Michael R Garey, David S. Johnson, and Larry Stockmeyer. Some simplified np-complete graph problems. *Theoretical computer science*, 1(3):237–267, 1976.

[33] Martin E. Dyer and Alan M. Frieze. The solution of some random np-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[34] Mark Jerrum and Gregory B Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1):155–175, 1998.

[35] Anne Condon and Richard M Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, 18(2):116–140, 2001.

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

[37] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W 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.

[38] Amin Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(02):227–284, 2010.

[39] 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(6):066106, 2011.

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

[41] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[42] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. *arXiv preprint arXiv:1311.3085*, 2013.

[43] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

[44] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *arXiv preprint arXiv:1405.3267*, 2014.

[45] Joel Friedman and David-Emmanuel Kohler. The relativized second eigenvalue conjecture of alon. *arXiv preprint arXiv:1403.3462*, 2014.

[46] Charles Bordenave. A new proof of friedman's second eigenvalue theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[47] Alon Amit, Nathan Linial, Jiří Matoušek, and Eyal Rozenman. Random lifts of graphs. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 883–894. Society for Industrial and Applied Mathematics, 2001.

[48] Brendan D McKay and Xiaoji Wang. Asymptotic enumeration of 0–1 matrices with equal row sums and equal column sums. *Linear algebra and its applications*, 373: 273–287, 2003.

[49] Brendan D McKay and Nicholas C Wormald. Asymptotic enumeration by degree sequence of graphs with degrees o($n^{1/2}$). *Combinatorica*, 11(4):369–382, 1991.

[50] Gregory F Lawler and Alan D Sokal. Bounds on the l^2 spectrum for markov chains and markov processes: a generalization of cheegers inequality. *Transactions of the American mathematical society*, 309(2):557–580, 1988.

[51] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178, 1989.

[52] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[53] Catherine Greenhill, Svante Janson, Jeong Han Kim, and Nicholas C Wormald. Permutation pseudographs and contiguity. *Combinatorics, Probability & Computing*, 11(03):273–298, 2002.

[54] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.