The rank of symmetric random matrices via a graph process

Laura C. Eslava Fernández

Master of Science

Department of Mathematics and Statistics

McGill University

Montreal,Quebec

2012-12-01

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master in Science

DEDICATION

To my parents. Without them, this could not be possible.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Professor Louigi Addario-Berry, who set me in the road. I am thankful for his encouragement and insight. The quality of my thesis would be substantially lower without his patience and careful reading.
I would like to thank my professors Bruce Reed, Luc Devroye and Hamed Hatami for their cheerful mentorship and for initiating me in the field of probabilistic combinatorics.
I would also like to thank my graduate fellows, specially Jon Noel and Xing Shi Cai, without whom my time in the master program would not have been as enjoyable and interesting.
Finally and foremost, I would like to thank my fiancé, Guillermo, whose support goes beyond frontiers and whose wise words always help me overcome any difficulties.

ABSTRACT

Random matrix theory comprises a broad range of topics and avenues of research, one of them being to understand the probability of singularity for discrete random matrices. This is a fundamental, basic question about discrete matrices. Although is been proven that for random symmetric Bernoulli matrices the probability of singularity decays at least polynomially in the size of the matrix, it is conjectured that the right order of decay is exponential.

We are interested in the adjacency matrix $Q_{n,p}$ of the Erdős-Réyni random graph $G_{n,p}$ and we study the statistics of the rank of $Q_{n,p}$ as a means of understanding the probability of singularity of $Q_{n,p}$.

We take a stochastic process perspective, looking at the family $\{Q_{n,p}\}_{p\in(0,1)}$ as an increasing family of random matrices. We then investigate the structure of $Q_{n,p}$ at the moment that it becomes non-singular and prove that, similar to some monotone properties of random graphs, the property of being non-singular obeys a so-called 'hitting time theorem'. Broadly speaking, this means that all-zero rows, which are a 'local' property of the matrix, are the only obstruction for non-singularity. This fact, which is the main novel contribution to the thesis, extends previous work by Costello and Vu.

ABRÉGÉ

La théorie des matrices aléatoires a un large éventail de sujets et de pistes de recherche, l'un d'entre eux étant de comprendre la probabilité de la singularité des matrices aléatoires discrètes. Ça a été prouvé que pour des matrices aléatoires de Bernoulli symétriques la probabilité de singularité a des bornes polynomiales, mais la conjecture est que le bon ordre de décroissance est exponentiel.

Nous sommes intéressés par la matrice d'adjacence $Q_{n,p}$ du graphe aléatoire d'Erdős et Réyni $G_{n,p}$ et nous étudions les statistiques du rang de $Q_{n,p}$ comme un moyen de comprende la probabilité de singularité de $Q_{n,p}$. Nous proposons maintenant une perspective de processus stochastique.

Dans ce mémoire, nous considérons la famille $\{Q_{n,p}\}_{p \in (0,1)}$ comme une famille croissante de matrices aléatoires et nous étudions la structure de $Q_{n,p}$ au moment où il devient non singulière et nous prouvons de la même façon pour certaines propriétés monotones des graphes aléatoires, la propriété d'être non singulière obéit à soi-disant 'théorème de temps d'arrêt'. D'une manière globale, cela signifie que les lignes remplies de zéros, qui sont une propriété locale de la matrice, sont la seule obstruction pour la non-singularité. Ce fait, qui est la nouvelle contribution principale de ce mémoire, élargie les résultats antérieurs de Costello et Vu.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
Random Matrices: A brief overview

The study of random matrices has been developed widely over the last decades due to practical and theoretical motivations. Since the work of Ginibre and Wigner [46], and up to that of Erdős, Schlein, Yau [13, 14], Tao, Vu [38, 40] , Rudelson, Vershinyn [34] and Tran, Vu, Wang [44] (to name a few), random matrix theory has been a source of interesting questions and conjectures, many of which remain unsettled. One major goal has been to understand the universal behaviour of high dimensional systems. This universality was first noticed in works on mechanical statistics, where models involve a great number of components so that statistical considerations can be applied to them. Physicists introduced random matrices, where each of the entries is itself random and independent. In spite of the simplicity of this assumption (compared with the actual complexity of the object of study), random matrices seems to yield good approximations for many physical phenomena involving a large number of elements. As a consequence of this, and for their inherent mathematical interest, random matrices have been widely studied in different areas such as mathematical physics, theoretical computer science and combinatorics, to name a few.

Broadly speaking, the *universality phenomenon* for matrix ensembles is the observation that as the size of a random square matrix tends to infinity, the behaviour of many natural matrix-theoretic and spectral properties is determined just by the first moments of the distribution of its elements, regardless of the actual distribution. This parallels the universality of sums of random variables given by the central limit theorem.

In general, we say that a matrix $A = \{\xi_{ij}\}_{1\leq i,j\leq n}$ is random if its entries are random variables. Random matrices are classified into different ensembles, depending on the general structure of the matrix and the distribution of the ξ_{ij}'s. The principal ensembles, and by far the most widely studied, are the Gaussian ensembles. These possess a well established theory largely because they are endowed with an algebraic structure (see, for example [1], [36]), while discrete ensembles have been more difficult to analyse. Among the latter are the Bernoulli ensembles (where each ξ_{ij} takes the values ± 1 with equal probability), and the adjacency matrices of random graphs.

Random matrices are divided also into substantially different types depending on the structure of the matrix. One of the principal dichotomies is between symmetric and non-symmetric matrices. In non-symmetric ensembles, the main characteristic is that entries of the matrix are independent. The canonical examples are those where ξ_{ij} are iid On the other hand, a wide class of symmetric matrices is given by the *Wigner* matrices, which includes self-/home/tpks/Documents/Latex/adjoint matrices when the support of ξ_{ij} is complex. Wigner matrices are conditioned to $\xi_{ij} = \overline{\xi_{ji}}$ for each entry in the matrix, and to have independent upper diagonal entries with mean zero and identical second moments. The canonical continuous Wigner ensembles are the Gaussian orthogonal ensemble (GOE) and the Gaussian unitary ensemble (GUE), where the ξ_{ij}'s are real-valued and complex-valued Gaussians respectively.

In this thesis we focus on a particular type of symmetric sparse ensembles: the adjacency matrix $Q_{n,p}$ of the Erdős-Rényi graph $G_{n,p}$ defined in Chapter 3, below. The off-diagonal elements in these matrices take value 1 with probability p and zero otherwise. Thus, the ξ_{ij}'s are sparse and have non-zero mean; these hypothesis do not fit the Wigner ensemble and so the approach to these matrices is slightly different.

A more difficult constraint in a matrix is the correlation between entries (other than symmetry). This is the case for the adjacency matrix $Q_{n,d}$ of a random d-regular graph, where each row of the matrix contains exactly d non-zero entries. The latter ensemble is quite interesting but quite poorly understood from a theoretic perspective; we briefly return to this model at the end of the chapter.

We are interested in the statistics of the rank and determinant as a means of understanding the probability of singularity of $Q_{n,p}$. We propose a stochastic process perspective, looking at the family $\{Q_{n,p}\}_{p\in(0,1)}$ as an increasing family of random matrices. From this point of view, new questions naturally present themselves. In this thesis, we investigate the structure of $Q_{n,p}$ at the moment that it becomes non-singular and prove that, similar to some monotone properties of random graphs, the property of being non-singular obeys a so-called 'hitting time theorem'. Broadly speaking, this means that all-zero rows, which are a 'local' property of the matrix, in that they correspond to isolated vertices in $G_{n,p}$, are the only obstruction for non-singularity. This fact, which is the main novel contribution to the thesis, extends previous work by Costello and Vu [11].

In the next section, we briefly pause to present some basic notation used throughout the thesis. In the remainder of the introduction we give some heuristics concerning the universality of the empirical distribution of Wigner and iid ensembles; then we see that the ensembles coming from random graphs also present the same phenomenological properties as the classical Gaussian orthogonal ensemble. We dedicate Chapter 2 to present some history and current results on the probability of singularity for discrete random matrices, along with Costello and Vu's proof that the rank of $Q_{n,p}$ equals the number of non-zero rows in the matrix, for $p \geq \frac{\ln n}{2n}$, [11]. In Chapter 3 we introduce a graph process approach for the family $\{Q_{n,p}\}_{p\in(0,1)}$ and we prove that not only does the property of non-singularity have a threshold function, but its hitting time also coincides with that of connectivity with high probability. Chapter 4 contains the details regarding the graph-theoretic part of our proofs; this builds upon work of Costello and Vu. We conclude the thesis with some other related questions and further avenues of research.

1.1 Notation

We write $Q_{n,p}$ and $Q_{n,d}$ for the adjacency matrix of the Erdős-Rényi graph $G_{n,p}$, and the d-regular random graph, respectively.

Let $\{E_n\}_{n\in\mathbb{N}}$ be a sequence of events. We say that $\{E_n\}_{n\in\mathbb{N}}$ holds asymptotically almost surely (or a.a.s.) if

$$\liminf_{n\to\infty} \mathbf{P}\{E_n\} = 1.$$

If $\{X_n\}_{n\in N}$ is a sequence of random variables, we say that $\{X_n\}_{n\in\mathbb{N}}$ converges almost surely towards X if

$$\mathbf{P}\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

We use a_i, a_{ij} to denote fixed real numbers and x_i to denote random variables. For any positive integers l, m, we denote the set $\{1,\dots,l\}$ by $[l]$ and we denote the set $\{l, l+1,\dots,m\}$ by $[l,m]$; depending on the context, $[m]$ will refer to a set of vertices or a range of indices. We refer to the rows of an $n\times n$ matrix Q by $\mathbf{r}_1,\dots,\mathbf{r}_n$ and to its columns by $\mathbf{c}_1,\dots,\mathbf{c}_n$. For any vector $\mathbf{v} = (v_1,\dots,v_n)$, we define $\mathbf{v}^0 = (v_1,\dots,v_n,0)$. For $Q_{n,p}$ we define $i(Q_{n,p}) = |\{i\in[n];\ \mathbf{r}_i = \mathbf{0}\}|$.

Given a graph $G = (V,E)$ and $v\in V$, we denote the *neighbourhood of* v by $N_G(v)$: this is the set of vertices adjacent to v in G. The number of isolated vertices is $i(G) = |\{v\in V;\ N_G(v) = \emptyset\}|$, and for any sets $A, B\subset V$ we denote by $E_G(A,B)$ the set of edges connecting vertices of A and B. The *minimum degree* of a graph G is $\delta(G) := \min_{i\in V}\{|N_G(i)|\}$. If the graph has vertex set $V = [n]$, then for any $l < m \le n$ let us write $G[m]$ to denote the subgraph of G induced by the set of vertices $\{1,\dots,m\}$ and $G[m\setminus l]$ to denote the subgraph of G induced by the set of vertices $\{l+1,\dots,m\}$.

Finally, we occasionally omit floors and ceilings when these are not essential.

1.2 Universality of spectral statistics

Here we will present the global spectral statistics for both Wigner and iid ensembles, along with some comments about the Lindeberg principle, which is an important tool used to obtain universality results. The principal statistics in a matrix are the eigenvalues and singular values (see [2] for definitions and basic notions), as other matrix statistics can be derived from them, for example, the determinant, the norm of the matrix and the condition number.

Consider an $n \times n$ matrix A_n with entries ξ_{ij}. The eigenvalues of A_n are complex-valued in general; however, if A_n is self-adjoint, then its eigenvalues are real and so we can denote them by $\lambda_1 \leq \ldots \leq \lambda_n$. On the other hand, singular values are always real, as they are the square roots of the eigenvalues of the self-adjoint matrix $A_n^* A_n$, and the latter matrix is always positive semidefinite. Let us denote the singular values of A_n by $0 \leq \sigma_n \leq \ldots \leq \sigma_1$.

We define the *empirical spectral distribution (ESD) of a self-adjoint matrix A_n* as the normalized cumulative number of its eigenvalues below a threshold x:

$$W_n(x) := \frac{1}{n} \left| \{ i;\, \lambda_i \leq x \} \right| . \tag{1.1}$$

When the eigenvalues are complex we instead get a measure on the complex plane. We define the *empirical spectral distribution of an iid matrix A_n* as

$$\mu_n(x,y) := \frac{1}{n} \left| \{ i;\, Re(\lambda_i) \leq x, Im(\lambda_i) \leq y \} \right| . \tag{1.2}$$

Simulations of such distributions indicate that limiting distributions for these empirical distributions exist, see Figures 1–1 and 1–2.

The first rigorous result, called Wigner's semicircular law, was established by Wigner in 1958 for symmetric ensembles where the ξ_{ij}'s have common variance and bounded higher moments, [47]. Its proof is based on the observation that the trace of a matrix equals the sum of its eigenvalues, and so

$$trace(A_n^k) = \sum_{i=1}^{n} {\lambda_i}^k, \tag{1.3}$$

for any $k \in \mathbb{N}$; the expected value of the right-hand side is n times the k-moment of a uniformly chosen eigenvalue λ_i. Thus, we transform the problem to understanding the distribution of the trace of A_n^k.

The semicircular law is defined as

$$W(x) := \frac{2}{\pi}\sqrt{4 - x^2},$$

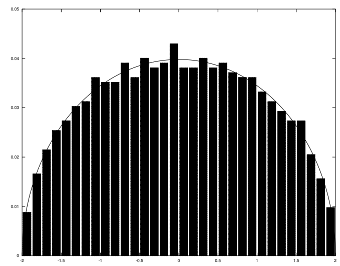

Figure 1–1: The semicircular law and the ESD of a symmetric random matrix $\frac{1}{\sqrt{n}}A_n$ where $A_{ij} \sim N(0,1)$ and $n = 1024$.

for $|x| \leq 2$ and $W(x) := 0$ otherwise. The following version of Wigner's semicircular law is due to Pastur [33], and is presented in the monograph of Bai and Silverstein [3], see Figure 1–1.

Theorem 1.2.1 (Semicircular law [3]). *Let A_n be the an $n \times n$ Wigner matrix whose upper diagonal entries are iid complex random variables with mean 0 and variance 1. Then the ESD of $\frac{1}{\sqrt{n}}A_n$ converges almost surely to the semicircular law.*

Approaches to the study of the spectrum using observation (1.3) are collectively called the *trace method*. The trace method cannot be applied when the eigenvalues are complex, in which case other methods are used. The limiting distribution for Gaussian ensembles was established by Mehta [31] in 1967, using Ginibre's formula for the joint density function of the eigenvalues of A_n, see Figure 1–2. It was later extended to ensembles with different restrictions on the distribution of ξ_{ij}; breakthroughs were due to Girko and Bai and more recently Tao and Vu have established the circular law for any ensemble with iid ξ_{ij} with mean zero and variance one (see [39] and the references therein).

Theorem 1.2.2 (Circular Law [39]). *Let A_n be an $n \times n$ random matrix whose entries are iid complex random variables with mean 0 and variance 1. Then the ESD of $\frac{1}{\sqrt{n}}A_n$ converges (both in probability and in the almost sure sense) to the uniform distribution on the unit disk.*

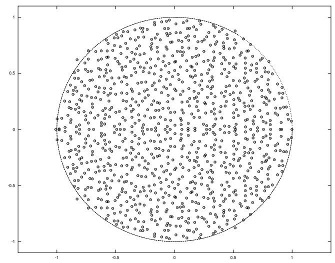

Figure 1–2: The circular law and the ESD of a random matrix $\frac{1}{\sqrt{n}}A_n$ where $A_{ij} \sim N(0,1)$ and $n = 1024$.

The Gaussian ensembles are the best understood random matrices because they are endowed with a strong group structure: Gaussian random variables are closed under linear combinations, and the GOE and the GUE are closed under orthonormal and unitary transformations, respectively. Thus, the joint distribution of the eigenvalues can be explicity written and so other important quantities can be derived. To the contrary, other ensembles lack this algebraic structure and consequently other techniques were developed to study non-Gaussian ensembles.

Relatively recently, Chatterjee introduced the Lindeberg swapping method to random matrix theory to treat Wigner matrix with exchangeable entries [9]. This method is based on Lindeberg's idea for a proof of the central limit theorem. Lindeberg proved that the limit $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ is common for any sequence of independent random variables $X_1, X_2, \ldots$ with mean zero and variance one [28]; in which case it is enough to compute such a limit for normalized Gaussian random variables.

In the context of random matrices, the Lindeberg swapping method is applied by showing that a certain function of the variables, not necessarily linear, does not vary too much if we replace one or two variables with a Gaussian with the same mean and variance. If it is proven that the error becomes negligible in the limit, then any ensemble with normalized elements shares the same limit (of the function) with the GOE or GUE. Then, we can use the algebraic properties in GOE and GUE to compute the actual limit.

This method has been used for some other spectral properties such as delocalization of eigenvalues and eigenvectors and the condition is that the ensembles have four matching moments. Recent developments concerning Wigner matrices cover the spacing between eigenvalues and the theory of Wigner matrices is almost complete, see [12] and [41], though some results get weaker convergence than almost sure convergence and, for some results, exponential moments for the elements are still required.

The picture for the universality phenomenon in the case of Wigner ensembles is clear now, and, so it is in a less extent for some generalizations of Wigner matrices. However, for such universality results it is of great importance that the elements in the upper triangle are independent and that they have mean zero. The case of adjacency matrices of graphs therefore requires a different analysis. A particular problem for $Q_{n,p}$ is that the elements can be heavily concentrated around zero. In the next section we present the main spectral statistics of sparse matrices such as $Q_{n,p}$ and $Q_{n,d}$, and some recent breakthroughs.

1.3 Sparse Matrices and Graph ensembles

Among the symmetric models that do not fit the conditions of Wigner ensembles there are, for example: self-adjoint matrices whose elements have different variance; 'band' matrices, such as tridiagonal matrices; symmetric matrices where randomly chosen entries are set to zero; and adjacency matrices of different classes of random graphs. We are interested particularly in $Q_{n,p}$, the adjacency matrix of an Erdős-Réyni graph. For fixed $p \in (0, 1)$, the Erdős-Rényi graph $G_{n,p}$ is the random graph on vertices $\{1, \ldots, n\}$, in which each possible edge is independently present with probability p.

Another interesting ensemble is $Q_{n,d}$, with d fixed or varying with n, and corresponding to the d-regular random graph $G_{n,d}$, which is uniformily chosen among the d-regular graphs on n vertices. The latter is important because it is a simple model where the elements in the upper triangle are not independent.

In the case of adjacency matrices, the largest eigenvalue is not near the rest of the eigenvalues. If $\lambda_1 \leq \ldots, \leq \lambda_n$ are the eigenvalues of $Q_{n,p}$ (resp. $Q_{n,d}$) and p is not too small, then λ_n is of the order of np (resp. $\lambda_n = d$). However, after a proper normalization, the rest of the eigenvalues follow the semicircular law if p is not so small (resp. $d \to \infty$).

Theorem 1.3.1 ([20]). *There is a constant c such that the following holds. If $p = p(n) \geq \frac{\ln^c n}{n}$ then the ESD of $\frac{1}{\sqrt{np(1-p)}} Q_{n,p}$ converges to the semicircular law with probability one.*

Theorem 1.3.2 ([44]). *Let $d = d(n) \to \infty$. Let $\hat{Q}_{n,d} = Q_{n,d} - \frac{d}{n}\mathbf{1}$, where $\mathbf{1}$ denotes the all-ones matrix. Then the ESD of $\frac{1}{\sqrt{nd(n-d)}} \hat{Q}_{n,d}$ converges to the semicircular law with probability one.*

Theorem 1.3.1 is a result of Furedi and Kolmós from 1981 and Theorem 1.3.2 was recently settled by Tran, Vu and Wang in 2010. The proof of the latter theorem uses the fact that adding a matrix of rank 1 can not overly perturb the spectrum.

One of the problems in handling sparse matrices, is precisely the large number of zeros in the matrix. This is essentially the reason why we have the condition $p \geq \frac{\ln^c n}{n}$ for Theorem 1.3.1. It is remarkable that for the sparse d-regular graphs, when d fixed, a limiting distribution also exists and this in turn approaches to the semicircular law as $d \to \infty$. Let

$$W_d(x) := \frac{d^2 - d}{d^2 - 4(d-1)x^2} W(x).$$

Theorem 1.3.3 (Kesten-McKay's law [25] [30]). *For any fixed positive integer d, the ESD of $\frac{1}{\sqrt{d-1}} Q_{n,d}$ converges to W_d with probability one.*

The Kesten-McKay's law was established over 30 years before Theorem 1.3.2.

Some other statistics of $Q_{n,p}$ and $Q_{n,d}$ are known to behave as predicted by the classical matrix ensembles. In particular, in $Q_{n,p}$ both eigenvalues and eigenvectors are delocalized [44].

Now consider $Q_{n,d}$ and let

$$\lambda = \sup_{\|\mathbf{v}\|=1, \mathbf{v}\cdot\mathbf{1}=0} |Av| = \max\{|\lambda_{n-1}|, |\lambda_1|\}.$$

This parameter corresponds to the spectral norm of the normalized matrix and is useful in that whenever λ is significantly less than d, then $Q_{n,d}$ is in a certain sense 'distributed' as a random graph with edge density d/n (see [22] for more details). The difference $d - \lambda$ is known as the spectral gap of $Q_{n,d}$.

We finish this section with a conjecture on the spectral gap. A d-regular graph G is called a Ramanujan graph if $\lambda(G) \leq 2\sqrt{d-1}$. The only explicit constructions of Ramanujan graphs are based on deep results in number theory; however, it is believed that a positive proportion of regular graphs are in fact Ramanujan. The most famous result in this direction is Friedman's theorem (Alon's conjecture), which we state below.

Theorem 1.3.4 (Alon's conjecture [19])**.** *For any $\varepsilon > 0$ and any fixed $d \geq 3$, a.a.s*

$$\lambda(Q_{n,d}) = (2+\varepsilon)\sqrt{d-1}.$$

It is plausible that, in fact, $Q_{n,d}$ is Ramanujan with positive constant probability, for d fixed and n tending to infinity. However, at this point, it is not even known that for every d there are infinitely many d-regular Ramanujan graphs. Also, the analogue of Alon's conjecture in the case that d grows with n, has yet to be tackled. One reasonable conjecture posed by Vu in [45] is the following.

Conjecture 1. *Assume that $d \leq \frac{n}{2}$ and both d and n tend to infinity. Then a.s.*

$$\lambda(Q_{n,d}) = (2 + o(1))\sqrt{d\left(1 - \frac{d}{n}\right)}.$$

CHAPTER 2
Rank and Determinant of discrete matrices

In contrast to the global spectral statistics of random matrices where both continuous and discrete ensembles have the same behaviour, the question of whether a matrix is singular becomes non-trivial only in the discrete setting. Breakthroughs in this direction have been possible due to the use of additive combinatorics tools such as Littlewood-Offord results, which studies the concentration of linear combinations of random variables; and Freiman's inverse theorem, which gives conditions for a set of integers to have a strong additive structure.

In the case of $Q_{n,p}$, Costello and Vu have gone beyond the study of the probability of singularity to the study of the dynamics of the rank of $Q_{n,p}$, showing that with high probability $rank(Q_{n,p})$ equals the number of non-zero rows of $Q_{n,p}$ when $p \geq \frac{\ln n}{2n}$ [CostelloVu10]. They also extended the study of the rank to smaller values of p, see [11].

The Littlewood-Offord problem is part of the study of additive combinatorics and has been one of the cornerstones in the study of the singularity probability and to estimate the determinant and rank of both symmetric and non-symmetric random matrices. As Littlewood-Offord results are of independent interest, and the basic results have simple and elegant proofs, we start by presenting some of these results. Then we return to survey the results and conjectures concerning the singularity probability of iid Bernoulli ensembles, and of the $Q_{n,p}$ and $Q_{n,d}$ ensembles. Since we build upon the work on [11], we present their work in full detail in the remainder of the chapter.

2.1 Littlewood-Offord results

The Littlewood-Offord problem is part of the additive combinatorics theory. It studies the concentration of linear, quadratic and, more generally, polynomial forms of random variables. From its most simple form (Theorem 2.1.3) to the most recent and sophisticated version of the inverse problem due to Tao and Vu [40], these concentration inequalities have played a key role in the study of random matrices. Komlós used Theorem 2.1.3 to bound the probability of singularity for Bernoulli random matrices and more sophisticated version are used to bound the probability of singularity and to estimate the determinant of discrete random matrices.

These inequalities are named after Littlewood and Offord, as they first posed the question in a study of the roots of complex polynomials [29]. They were looking for an upper bound on the number of different sums

$$\sum_{i=1}^{n} a_i \eta_i$$

that lie in a circle of radius 1, where η_i are signs ± 1 and a_i are non-zero complex numbers with norm at least 1. In 1945, Erdős showed [15] that the number of different sums in any such circle is no more than $\binom{n}{\lfloor n/2 \rfloor}$, which is best possible if we set all a_i to be equal. The argument is simple and elegant, and uses Sperner's theorem on the size of antichains.

Definition 2.1.1. *Given a set X, an antichain $\mathcal{A}$ is a collection of subsets of X, in which for any two subsets $A, B \in \mathcal{A}$ neither of them is properly contained in the other.*

Clearly, for the set $[n] = \{1, 2, \ldots, n\}$, the collection of all subsets of size k form an antichain with $\binom{n}{k}$ elements and so we maximize its size by letting $k = \lfloor n/2 \rfloor$. Sperner's theorem asserts that the size of an antichain can not be larger than that. There are several proofs to this theorem, here we present one with a probabilistic point of view.

Theorem 2.1.2 (Sperner's theorem [42])**.** *Let $\mathcal{A}$ be an antichain on $[n]$. Then*

$$|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Proof. Consider a random permutation $\sigma : [n] \to [n]$ uniformly chosen among the $n!$ possible permutations. We now compute the probability of the auxiliary event

$$\bigcup_{A \in \mathcal{A}} \{\sigma(A) = [|A|]\},$$

that is, the probability that at least one of the sets A in the antichain is mapped exactly into the first $|A|$ elements in $[n]$. We claim the above union is a union of disjoint events. To see this, fix distinct $A, B \in \mathcal{A}$ with $|A| \leq |B|$. If $\sigma(A) = [|A|]$ and $\sigma(B) = [|B|]$ then necessarily $A \subset B$, contradicting the fact that $\mathcal{A}$ is an antichain. On the other hand,

$$\mathbf{P}\{\sigma(A) = [|A|]\} = \binom{n}{|A|}^{-1}$$

for any set $A \in [n]$. Thus,

$$\mathbf{P}\left\{\bigcup_{A \in \mathcal{A}} \sigma(A) = [|A|]\right\} = \sum_{A \in \mathcal{A}} \binom{n}{|A|}^{-1} \leq 1.$$

The result follows by replacing each of the binomial coefficients $\binom{n}{|A|}$ by the largest binomial coefficient $\binom{n}{\lfloor n/2 \rfloor}$ to obtain

$$|\mathcal{A}| \cdot \binom{n}{\lfloor n/2 \rfloor}^{-1} \leq \sum_{A \in \mathcal{A}} \binom{n}{\lfloor n/2 \rfloor}^{-1} \leq \sum_{A \in \mathcal{A}} \binom{n}{|A|}^{-1} \leq 1.$$

□

Erdős's observation was that a set of signed sums all lying in a sufficiently small disc in the complex plane can be used to define an antichain of subsets of $[n]$, at which point Sperner's theorem can be applied. For our purposes it suffices to consider the case where the $a_1, \ldots, a_n$ are real rather than complex and work with intervals rather than circles. However, as we will not restrict the norm of the a_i, the length of the interval for which Sperner's theorem can be applied will depend on the smallest of $|a_i|$, $i \in [n]$. In the next theorem we present the result in probability concentration terms as it will be more convenient to look at the probability a random sum takes specific values.

Theorem 2.1.3 ([15])**.** *Let $a_1, \ldots, a_n$ be real non-zero coefficients, and let $x_1, \ldots x_n$ be independent random variables taking values -1 and 1 with equal probability. Then*

$$\sup_{c \in \mathbb{R}} \mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \frac{1}{\sqrt{n}}.$$

Proof. Write $r = \min\{|a_i|;\ 1 \leq i \leq n\}$. Fix $c \in \mathbb{R}$ and write $I = (c - r, c + r)$, note that I is not empty because $r > 0$. We then have

$$\mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \mathbf{P}\left\{\sum a_i x_i \in I\right\}.$$

We can assume that the coefficients $a_1, \ldots a_n$ are positive because the variables x_i are symmetric. Write $A = \{i \in [n];\ x_i = 1\}$, we have

$$\sum_{i=1}^{n} a_i x_i = \sum_{i \in A} a_i - \sum_{i \notin A} a_i.$$

We now claim that the collection

$$\mathcal{A} = \{S \in [n];\ \sum_{i \in S} a_i - \sum_{i \notin S} a_i \in I\}$$

forms an antichain. To see this, consider two different subsets $S \subsetneq T \subset [n]$, the difference between their associated sums is at least

$$\left(\sum_{i \in T} a_i - \sum_{i \notin T} a_i\right) - \left(\sum_{i \in S} a_i - \sum_{i \notin S} a_i\right) \geq 2 \sum_{i \in T \setminus S} a_i \geq 2r$$

and so the sums can not simultaneously lie in the interval I. By Sperner's theorem it follows that $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$. Therefore, for any $c \in \mathbb{R}$ we have

$$\mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \mathbf{P}\left\{\sum a_i x_i \in I\right\} \leq \binom{n}{\lfloor n/2 \rfloor} 2^{-n} \leq \frac{1}{\sqrt{n}}.$$

The last inequality follows from a nice manipulation of the closed form

$$\binom{n}{\lfloor n/2 \rfloor} 2^{-n} = \prod_{m=1}^{\lceil n/2 \rceil} \frac{2m-1}{2m},$$

which is developed in the appendix. $\square$

Remark 2.1.4. *The bound of Theorem 2.1.3 also holds if we consider x_i as independent random varibles taking values 0 or 1 independently with equal probability. That is*

$$\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i = c\right\} = \frac{1}{\sqrt{n}}.$$

Proof. Indeed, we can define $x_i' = 2x_i - 1$ so x_i' is 1 whenever $x_i = 1$ and x_i' is -1 whenever $x_i = 0$. Then rescaling the coefficients to $a_i' = a_i/2$ and shifting the tarjet value to $c' = c - \sum a_i$ gives

$$\mathbf{P}\left\{\sum a_i x_i = c\right\} = \mathbf{P}\left\{\sum a_i' x_i' = c'\right\} \leq \frac{1}{\sqrt{n}},$$

for any $c \in \mathbb{R}$. $\square$

Costello and Vu extended the results above to apply them to sparse random matrices such as $Q_{n,p}$ with $p < \frac{1}{2}$. The next two theorems give analogous bounds for random variables taking values 1 with probability p and zero otherwise, and for random variables taking values ± 1 with probability $p/2$ and zero with probability $1 - p$.

Theorem 2.1.5 ([11])**.** *Let $a_1, \ldots, a_n$ be real non-zero coefficients and $x_1, \ldots x_n$ independent random variables taking values* 1 *with probability* p *and* 0 *otherwise. Then*

$$\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \frac{2}{\sqrt{np}}.$$

Proof. First note that the supremum is trivially bounded by 1, so we can assume $np > 1$. We transform the random variables by letting $x_i = y_i z_i$, where the y_i and z_i are independent 0-1 random variables, taking value 1 with probability $2p$ and $1/2$ respectively. Then, for any $c \in \mathbb{R}$, we can condition on the number of non-zero variables y_i to bound $\mathbf{P}\left\{\sum a_i x_i = c\right\}$ with

$$\mathbf{P}\left\{\sum y_i < np\right\} + \sum_{k\geq np} \mathbf{P}\left\{\sum_{i;\, y_i=1} (a_i y_i) z_i = c \;\middle|\; \sum y_i = k\right\} \mathbf{P}\left\{\sum y_i = k\right\}. \qquad (2.1)$$

The first term can be bounded by Chebyshev's inequality. Clearly, $\sum y_i$ has a binomial distribution and thus

$$\mathbf{P}\left\{\sum y_i < np\right\} \leq \mathbf{P}\left\{\left|\sum y_i - 2np\right| > np\right\} \leq \frac{2np(1-2p)}{(np)^2}.$$

As for the second term, Remark 2.1.4 gives that for each fixed $k \geq np$,

$$\mathbf{P}\left\{\sum_{i;\, y_i=1} (a_i y_i) z_i = c \;\middle|\; \sum y_i = k\right\} \leq \frac{1}{\sqrt{k}}.$$

Finally, plugging the above bounds in (2.1) allows us to conclude that

$$\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \frac{1}{np} + \frac{1}{\sqrt{np}} \leq \frac{2}{\sqrt{np}}.$$

$\square$

Theorem 2.1.6. *Let $a_1, \dots, a_n$ be real non-zero coefficients and $x_1, \dots x_n$ independent random variables taking values 1 or -1 each with probability $p < 1/2$ and 0 otherwise. Then*

$$\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i = c\right\} \leq \frac{2}{\sqrt{np}}.$$

Proof. Following the ideas in the proof above, let $x_i = y_i z_i$, where the y_i and z_i are independent random variables, y_i takes values 1 with probability $2p$ and 0 otherwise, but now z_i is 1 or -1 with equal probability. We get the same conclusion by conditioning on the number of non-zero variables y_i. $\square$

It is perhaps worth recalling here that in the linear combinations we encounter during the study of random matrices we assume no detailed information about the coefficients. Nevertheless, a lower bound on the number of non-zero coefficients is sufficient to obtain a worse, still useful, bound for the concentration of such linear combinations.

As we see in more detail in the next section, linear concentration bounds are useful to study the probability of singularity of iid Bernoulli matrices. In contrast, the same probability for symmetric matrices involves a quadratic function of their entries. To this end, Costello, Tao and Vu formulated the Litllewood-Offord problem for quadratic forms [10] which is a key step in the proof that $Q_{n,1/2}$ is a.s. non-singular (Theorem 2.2.4). Here we present the quadratic Littlewood-Offord inequality developed in [11].

Theorem 2.1.7 ([11]). *Let $A = \{a_{ij}\}$ be a $n \times n$ real-valued symmetric matrix in which there are at at least $2k$ rows having at least $2k$ non-zero entries each. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a column vector in which each x_i is 1 independently with probability $p < 1/2$ and 0 otherwise. Then, letting $D(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ we have*

$$\sup_{c \in \mathbb{R}} \mathbf{P}\{D(\mathbf{x}) = c\} \leq 3(kp)^{-1/4}.$$

The proof of this theorem uses a decoupling lemma to separate some of the variables, say $\{x_i;\, 1 \leq i \leq k\}$, and extract from $D(\mathbf{x})$ linear functions of the rest of the variables, $\{x_i;\, k < i \leq n\}$, for which we can apply Theorem 2.1.6, thus reducing the complexity of the problem. The decoupling lemma, which is a particular case of a conjecture of Sidorenko, consists of a simple application of the Cauchy-Schwartz inequality.

Lemma 2.1.8. *Let X and Y be random variables, and let $E(X,Y)$ be an event depending on X and Y. Let Y' be an independent copy of Y, then*

$$\mathbf{P}\{E(X,Y)\}^2 \leq \mathbf{P}\{E(X,Y), E(X,Y')\}.$$

Proof. For clarity we assume that X takes a finite number of values $x_1, \ldots x_n$. Note that

$$\mathbf{P}\{E(X,Y)\} = \sum_{i=1}^n \mathbf{P}\{E(X,Y) \mid X = x_i\} \mathbf{P}\{X = x_i\}.$$

By the Cauchy-Schwartz inequality we obtain

$$\begin{aligned}\mathbf{P}\{E(X,Y)\}^2 &\leq \left(\sum_{i=1}^n \mathbf{P}\{E(X,Y) \mid X = x_i\}^2 \mathbf{P}\{X = x_i\}\right)\left(\sum_{i=1}^n \mathbf{P}\{X = x_i\}\right)\\ &= \mathbf{P}\{E(X,Y), E(X,Y')\}\end{aligned}$$

where the last equality is due to the independence of Y and Y'. $\square$

Proof of Theorem 2.1.7. Let $x_1, \dots, x_k, y_{k+1}, \dots, y_n$ be independent random variables taking value 1 with probability p and zero otherwise. We combine these variables to form two independent random vectors X and Y; let $X = (x_1, \dots, x_k)$ and $Y = (y_{k+1}, \dots, y_n)$. Denote by

$$D(X,Y) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_ix_j + \sum_{i=1}^{k}\sum_{j=k+1}^{n} (a_{ij}+a_{ji})x_iy_j + \sum_{i=k+1}^{n}\sum_{j=k+1}^{n} a_{ij}y_iy_j.$$

Then $D(X,Y)$ is distributed as $D(\mathbf{x})$. Now, for a fixed $c \in \mathbb{R}$ we use Lemma 2.1.8 to estimate the probability of $E(X,Y)$, the event that $D(X,Y) = c$. Let $Y' = (y'_{k+1}, \dots, y'_n)$ be an independent copy of Y, then

$$\mathbf{P}\left\{D(X,Y) = c\right\} \le \mathbf{P}\left\{D(X,Y) = D(X,Y') = c\right\}^{1/2};$$

thus, it suffices to bound the probability of the event

$$\{D(X,Y) - D(X,Y') = 0\} \supset \{D(X,Y) = D(X,Y') = c\}.$$

We manipulate the quadratic form $D(X,Y) - D(X,Y')$ to obtain a linear function of X.

$$\begin{aligned} D(X,Y) - D(X,Y') =& \sum_{i=1}^{k}\sum_{j=k+1}^{n} (a_{ij}+a_{ji})(y_j - y'_j)x_i \\ &+ \sum_{i=k+1}^{n}\sum_{j=k+1}^{n} a_{ij}(y_iy_j - y'_iy'_j). \end{aligned}$$

Recalling that A is symmetric, and thus $a_{ij} = a_{ji}$, we get

$$D(X,Y) - D(X,Y') = f(Y,Y') + \sum_{i\le k} W_i(Y,Y')x_i, \tag{2.2}$$

where

$$f(Y,Y') = \sum_{i,j>k} a_{ij}(y_i y_j - y'_i y'_j),$$

$$W_i(Y,Y') = \sum_{j>k} 2a_{ij}(y_j - y'_j).$$

Note that W_i is a linear form in the variables $(x_j - x'_j)$, each of which takes values ± 1 with probability $p(1-p)$ each.

We can assume, without loss of generality, that the first k rows of A contain at least $2k$ non-zero entries each. Thus, for any index $i \in [k]$ we can guarantee that at least k of the coefficients a_{ij} with $j \in [n] \setminus [k]$ are non-zero. As a consequence the random sum W_i has at least k non-zero coefficients a_{ij}, so by Theorem 2.1.6

$$\mathbf{P}\left\{W_i = 0\right\} \leq \frac{2}{\sqrt{kp(1-p)}}. \tag{2.3}$$

Furthermore, conditioning on the number of non-zero W_i, the probability that $D(X,Y) - D(X,Y') = 0$ is at most

$$\mathbf{P}\left\{\sum W_i x_i = -f(Y,Y') \;\middle|\; \sum \mathbf{1}_{W_i \neq 0} \geq k/2\right\} + \mathbf{P}\left\{\sum \mathbf{1}_{W_i = 0} > k/2\right\},$$

where the sums are over $i \in [k]$. The second term in the expression above is bounded by Markov's inequality, using (2.3) we obtain

$$\mathbf{P}\left\{\sum \mathbf{1}_{W_i=0} > \frac{k}{2}\right\} \leq \frac{2k\mathbf{P}\left\{W_i = 0\right\}}{k} \leq \frac{4}{\sqrt{kp(1-p)}}.$$

For the second term we use Theorem 2.1.5 to bound

$$\begin{aligned}&\mathbf{P}\left\{\sum W_i x_i = -f(Y,Y') \;\middle|\; \sum \mathbf{1}_{W_i \neq 0} \geq k/2\right\} \\ \leq &\sup_{c' \in \mathbb{R}} \mathbf{P}\left\{\sum W_i x_i = c' \;\middle|\; \sum \mathbf{1}_{W_i \neq 0} \geq k/2\right\} \leq 2\left(\frac{kp}{2}\right)^{-1/2}\end{aligned}$$

Putting these two bounds together and using that $p < 1/2$ we get

$$\mathbf{P}\left\{D(X,Y) - D(X,Y') = 0\right\} \leq \frac{2\sqrt{2}}{\sqrt{kp}} + \frac{4\sqrt{2}}{\sqrt{kp}} \leq \frac{9}{\sqrt{kp}},$$

therefore, by Lemma 2.1.8

$$\mathbf{P}\left\{D(X,Y) = c\right\} \leq \mathbf{P}\left\{D(X,Y) - D(X,Y') = 0\right\}^{1/2} \leq 3(kp)^{-1/4}.$$

This holds for all $c \in \mathbb{R}$, so the result follows. $\square$

An inverse question arises naturally: if $\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i\right\}$ is large, what can be said about the multiset $\{a_i\}_{i\in[n]}$? Tao and Vu proved the first inverse Littlewood-Offord concentration bound in [43] and used it to study the condition number of a random matrix. Essentially, they proved that if $\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i\right\}$ is large, then $\{a_i\}_{i\in[n]}$ has a strong additive structure.

Previous results related to this phenomenon are in [35], where Sárközy and Szemerédi showed that if the coefficients a_i are all different, then the bound can be improved from $n^{-1/2}$ to $n^{-3/2}$. The latter bound is tight, for example, when the set of coefficients are $a_k = k$. More generally, if $\sup_{c\in\mathbb{R}} \mathbf{P}\left\{\sum a_i x_i\right\}$ is large then the coefficients are contained in a generalized arithmetic progression (for explicit statements of this kind see [32]).

This area of additive combinatorics has been a useful tool for the study of discrete random matrices where not only the norm of the vectors is important but their direction in the space (or structure). In the next section we cover some of the history of the study of the probability of singularity, together with some estimates for the order of the determinant.

2.2 Singularity probability and Determinant

Recall that an $n \times n$ matrix A_n with row vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$ is singular if and only if $det(A_n) = 0$ and note that the determinant of a matrix can be express as the volume of the parallelepiped spanned by its row vectors. Thus,

$$det(A_n) = \prod_{k=1}^{n} d(\mathbf{r}_k, V_k), \tag{2.4}$$

where V_k is the subspace of $\mathbb{R}^n$ spanned by the row vectors $\mathbf{r}_1, \ldots, \mathbf{r}_{k-1}$. If A_n is random, it follows that

$$\mathbf{P}\left\{A_n \text{ is singular}\right\} = \mathbf{P}\left\{\mathbf{r}_k \in V_k, \text{ for some } k \in [n]\right\}.$$

If A_n has iid entries and the distribution of the entries is continuous, then the probability that a row $\mathbf{r}_k$ is contained in the span of the previous rows is zero. However, if the entries of A_n have a discrete distribution, then the distance $d(\mathbf{r}_k, V_k)$ becomes less trivial: it depends on the structure of the subspace V_k.

Singularity can also be expressed in terms of the rank. An $n \times n$ matrix A_n is singular if and only if $rank(Q) < n$. Here the intuition is that

$$\textit{`Singularity should come from small dependencies'} \tag{2.5}$$

To explain this intuition, we consider the random matrix M_n, whose entries are iid taking values 1 or -1 with equal probability. The smallest set of linearly dependent vectors in $\{-1, 1\}^n$ has size 2. As a consequence,

$$\mathbf{P}\left\{det(M_n) = 0\right\} \geq \binom{n}{2} \left(\frac{1}{2}\right)^n.$$

This follows from a union bound of all distinct pairs of rows vectors and pairs of column vectors in $Q_{n,p}$; the probability that any such pair is equal (up to sign) is $2^{-(n-1)}$. It has been conjectured that having two equal rows is essentially the only way to have singularity.

Conjecture 2. *For the random matrix M_n, and for n sufficiently large*

$$\mathbf{P}\left\{det(M_n) = 0\right\} = \left(\frac{1}{2} + o(1)\right)^n.$$

The first breakthrough towards this conjecture was give by Komlós in [26].

Theorem 2.2.1 ([26]). *For the random matrix M_n, we have that for n sufficiently large*

$$\mathbf{P}\left\{det(M_n) = 0\right\} \leq \frac{1}{\sqrt{n}}.$$

The next step was due to Kahn, Komlós and Szemerédi in [24], where they established an exponential bound: $(0.999^n + o(1))^n$. Later, Tao and Vu introduced results of the type of Freiman's inverse theorem to improve the constant to $\frac{3}{4}$ in [38]. To date, the strongest result on this conjecture is due to Bourgain, Vu and Wood.

Theorem 2.2.2 ([8]). *For the random matrix M_n, we have that for n sufficiently large*

$$\mathbf{P}\left\{det(M_n) = 0\right\} \leq (1/\sqrt{2} + o(1))^n$$

Once we have established that $\det(M_n) \neq 0$ a.a.s., we can also think about the expected value of $det(M_n)$. A straightforward bound is given by Hadamard's inequality which states that the determinant of a matrix is upper-bounded by the product of the length of its row vectors. So for M_n we have

$$det(M_n) \leq (\sqrt{n})^n.$$

It is conjectured that the above inequality is tight a.a.s. (up to terms of smaller order). A step towards this conjecture was made by Tao and Vu in [37].

Theorem 2.2.3 ([37]). *The random matrix M_n satisfy a.a.s.*

$$|det(A_n)| \geq \sqrt{n!}\exp(-29\sqrt{n\ln n}).$$

The proof of this theorem used the approximation of $det(M_n)$ via the distance of $\mathbf{r}_k$ to V_k as in (2.4) and Littlewood-Offord bounds are used for the case where k is close to n. As we have seen, results on $det(M_n)$ use mainly Theorem 2.1.3, which estimates the concentration of linear forms of a random vector.

A new obstacle to estimate $\mathbf{P}\{det(Q_{n,p}) = 0\}$ is that the entries of the matrix are no longer independent. As a consequence, *a)* estimating $\mathbf{P}\{d(\mathbf{r}_k, V_k)\}$ is not trivial, and *b)* the determinant of $Q_{n,p}$ is a quadratic function of its entries. The singularity probability $\mathbf{P}\{det(Q_{n,p}) = 0\}$ is conjectured to have the same exponential order as $\mathbf{P}\{det(M_n) = 0\}$. However, in contrast to the already exponential bounds for the probability of singularity for the iid ensemble, the best known bound for the singularity of $Q_{n,p}$ are still polynomial. And the question of whether $\mathbf{P}\{det(Q_{n,d}) = 0\}$ tends to zero for either d fixed or varying with n remains unsettled.

From now on, we only focus on the $Q_{n,p}$ ensemble. The study of the probability of singularity for symmetric matrices required the development of new techniques and tools. The breakthrough in this direction is due to Costello, Tao and Vu which established a.a.s. non-singularity for $Q_{n,1/2}$.

Theorem 2.2.4 ([10])**.** *The symmetric random matrix $Q_{n,1/2}$ is a.a.s. non-singular. More precisely*

$$\mathbf{P}\left\{Q_{n,1/2} \text{ is singular}\right\} = O(n^{-1/8+\alpha}),$$

for any positive constant α (the implicit constant in the $O(\cdot)$ notation of course is allowed to depend on α).

The current best upper bound for random symmetric matrices (with entries ± 1) is due to Nguyen.

Theorem 2.2.5 ([32])**.** *The symmetric random matrix A_n, with entries taking values 1 or -1 with equal probability satisfy*

$$\mathbf{P}\{A_n \text{ is singular}\} = O(n^{-C}).$$

for any positive constant C.

To prove Theorem 2.2.4, Costello, Tao and Vu developed a quadratic version of the Littlewood-Offord problem. Later, Costello and Vu introduced the vertex exposure method used in random graph theory to study $Q_{n,p}$ with $p = \frac{c \ln n}{n} < \frac{1}{2}$, with $c > \frac{1}{2}$, see Theorem 2.3.2 below.

In the case of symmetric matrices $Q_{n,p}$, the simplest cause of singularity is the presence of all-zeros rows. This is, in fact, a general difficulty when studying sparse matrices: their large amount of null entries.

For the case of $Q_{n,p}$, the probability that $Q_{n,p}$ has an all-zero rows tends to 0 if $p \geq \frac{\ln n}{n}$. However, the intuition in (2.5) goes beyond the existence of all-zero rows in $Q_{n,p}$. More can be said about small dependencies of $Q_{n,p}$ for $p < \frac{1}{2}$.

To do so, we need to study a finer property than that of singularity. In the following section we study in more detail the function $rank(Q_{n,p})$ which, interestingly, equals the number of non-zero rows in $Q_{n,p}$ for p fixed with high probability.

2.3 Rank of symmetric sparse matrices

Recall that each matrix $Q_{n,p}$ represents a random graph $G_{n,p}$ and zero-row vectors in $G_{n,p}$ are in bijective correspondence with isolated vertices in the graph. Recall that $i(Q_{n,p})$ denotes the number of zero-rows in $Q_{n,p}$. As we have seen in the previous section, all-zeros rows in $Q_{n,p}$ leads to singularity of $Q_{n,p}$.

In this section, we consider the rank of $Q_{n,p}$, which is equal to the dimension of the space spanned by its row vectors, $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \{0,1\}^n$. A set of *linearly dependent* vectors spans a space of strictly lower dimension than the number of such vectors. Thus, the rank is full precisely if $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$ are linearly independent.

The zero vector $\mathbf{0}$ has dimension zero and so we always have

$$rank(Q_{n,p}) \leq n - i(Q_{n,p}).$$

We are interested in understanding when equality holds above.

Definition 2.3.1. *A set of vectors* $\{\mathbf{r}_i\}_{i \in S}$ *is a* non-trivial dependency *or equivalently, a* vanishing set *if* $\mathbf{r}_i \neq \mathbf{0}$ *for all* $i \in S$ *and there exist non-zero coefficients* $\{a_i\}_{i \in S}$ *such* *that*

$$\sum_{i\in S} a_i \mathbf{r}_i = \mathbf{0},$$

We say that $\{\mathbf{r}_i\}_{i\in S}$ *has a* trivial dependency *if* $\mathbf{r}_i$ *is the zero vector for some* $i \in S$.

The simplest set of linearly dependent row vectors $\mathbf{r}_i$ are pairs of equal rows. These pairs correspond in the graph model to pairs of non-adjacent vertices with the same neighbourhood. One case of this is the endpoints of a path of length two. The probability of having a 2-path in $G_{n,p}$ tends to zero if $p \geq \frac{\ln n}{2n}$ (see [23], Chapter 5). Surprisingly, for $p \geq \frac{\ln n}{2n}$ the probability of having a non-trivial dependency in $Q_{n,p}$ tends to zero as n tends to infinity.

Theorem 2.3.2 ([11])**.** *For* $Q_{n,p}$ *with* $p = \frac{c \ln n}{n} < \frac{1}{2}$ *and* $c > 1/2$ *fixed, with probability* $1 - O((\ln \ln n)^{-1/4})$

$$rank(Q_{n,p}) = n - i(Q_{n,p}).$$

Consequently, for p in this range, the matrix is invertible if the graph $G_{n,p}$ has no isolated vertices, and the giant component has full rank.

As a warm up, we show that, with high probability the rank is close to n if $p = \frac{\ln n}{n}$.

Lemma 2.3.3. *Consider the random matrix* $Q_{n,p}$*, with* $p = \frac{\ln n}{n}$*. For any* $\varepsilon > 0$*,*

$$\mathbf{P}\left\{rank(Q_{n,p}) \leq (1-\varepsilon)n\right\} \leq c_2^n e^{-c_1 n \ln n},$$

where $c_1 = \frac{\varepsilon^2}{4}$ *and* $c_2 = \left(\frac{e}{\varepsilon}\right)^{\varepsilon}$*.*

Proof. Let A_0 be the event that the last εn rows of $Q_{n,p}$ are contained in the span of the first $(1-\varepsilon)n$ rows. By symmetry and a union bound

$$\mathbf{P}\left\{rank(Q_{n,p}) \leq (1-\varepsilon)n\right\} \leq \binom{n}{\varepsilon n} \mathbf{P}\left\{A_0\right\}.$$

We view the matrix $Q_{n,p}$ as the following block matrix

$$Q_{n,p} = \begin{pmatrix} B & C^T \\ C & D \end{pmatrix},$$

where B is the upper-leftmost $(1-\varepsilon)n \times (1-\varepsilon)n$ submatrix of $Q_{n,p}$ and D is the remaining $\varepsilon n \times \varepsilon n$ symmetric matrix. If A_0 holds, then necessarily exists a matrix F (not necessarily unique) satisfying $C = FB$ and

$$D = FC^T. \tag{2.6}$$

Note that the above equation holds for any such F. To estimate the probability of A_0 we condition on any fixed matrices B and C. In this case, if A_0 holds, then D is determined by (2.6). The probability that an element of D, coincides with FC^T is at most $1-p$, $(p < 1/2 < 1-p)$. The independence of the above diagonal entries yields

$$\mathbf{P}\left\{A_0 | B, C, C = FB\right\} = \mathbf{P}\left\{D = FC^T \;\middle|\; B, C, C = FB\right\} \leq (1-p)^{\binom{\varepsilon n}{2}}.$$

This holds for any fixed B and C so we obtain $\mathbf{P}\left\{A_0\right\} \leq (1-p)^{\binom{\varepsilon n}{2}}$ and therefore,

$$\begin{aligned}
\mathbf{P}\left\{rank(Q_{n,p}) \leq (1-\varepsilon)n\right\} &\leq \binom{n}{\varepsilon n}(1-p)^{\binom{\varepsilon n}{2}} \\
&\leq \left(\frac{e}{\varepsilon}\right)^{\varepsilon n} e^{-\frac{\varepsilon^2 n^2 p}{4}} \\
&\leq c_2^n e^{-c_1 n \ln n}.
\end{aligned}$$

$\square$

This holds even if we look at a graph with αn vertices for any fixed $\alpha \in (0,1)$ and we let $p = p(n) \in (\frac{\ln n}{2n}, \frac{1}{2})$.

Corollary 2.3.4 ([11])**.** *Fix $\alpha > 0$ and $n' = \lfloor \alpha n \rfloor$. Let Q be the adjacency matrix of $G_{n',p}$ with $p = \frac{c \ln n}{n} < \frac{1}{2}$ and fixed $c > \frac{1}{2}$. For any $\varepsilon > 0$,*

$$\mathbf{P}\left\{rank(Q) \leq (1-\varepsilon)n'\right\} \leq c_2^n e^{-c_1 n \ln n},$$

where $c_1 = \frac{\varepsilon^2 \alpha^2}{8}$ and $c_2 = \left(\frac{e}{\varepsilon}\right)^{\varepsilon\alpha}$.

In the remainder of the chapter we present the proof of Theorem 2.3.2 due to Costello and Vu. Later, in Chapter 3, we will generalize this proof to a new model of random graphs for which certain edges are not random.

From now on, we consider only sets of rows containing only non-zero vectors. To prove Theorem 2.3.2, we need to understand other forms of dependency.

Observe that if $\{\mathbf{r}_i\}_{i \in S}$ is a vanishing set of row vectors in $\{0,1\}^n$, then for every coordinate $k \in [n]$, necessarily either $\mathbf{r}_{ik} = 0$ for all $i \in S$ or else there are distinct $i, j \in S$ with $\mathbf{r}_{ik} = 1$ and $\mathbf{r}_{jk} = 1$. Translated to graph terms, a set of vertices S corresponding to a non-trivial dependency in $Q_{n,p}$ necessarily satisfies that every vertex in V is either adjacent to at least two vertices in S or to none at all. This simple observation is a key in the proof of Theorem 2.5.8, which roughly states that with high probability, any non-trivial dependency in $Q_{n,p}$ has large size.

An important technique is the *vertex exposure method*, introduced for symmetric matrices in [10]. The vertex exposure method has been used in random graph theory before: it consists of adding vertices one by one and revealing just the edges connecting to the vertices already exposed. This augmentation process is similar to that used in the estimation of the determinant for non-symmetric matrices [26].

In the case of symmetric matrices, the dependence among the entry values modify the way new information is added. The analysis starts with the symmetric submatrix generated by the first αn rows and columns in $Q_{n,p}$, where $\alpha \in (0,1)$ is chosen carefully. From then on, at each step an independent row and its transpose are added, preserving the symmetry of the matrix.

We now define the matrices of the exposure method through their corresponding graphs. Denote by $Q_{n,p}[m]$ the adjacency matrix of $G_{n,p}[m]$, which we define as the subgraph of $G_{n,p}$ induced by $[m]$. Moreover, note that $Q_{n,p}[m]$ is the upper-leftmost $m \times m$ submatrix of $Q_{n,p}$. Writing $n' = \lceil \alpha n \rceil$ for an appropriately chosen $\alpha > 0$, we will consider the sequence of matrices $\mathcal{Q}_{n,p}[(n', n)] = (Q_{n,p}[n'], Q_{n,p}[n'+1], \ldots, Q_{n,p}[n])$.

Costello and Vu used Corollary 2.3.4 together with the vertex exposure method to prove Theorem 2.3.2. We close this section by briefly sketching our approach to the proof of Theorem 2.3.2; the detailed proof can be found in Section 2.5.

We first apply Corollary 2.3.4 to show that with high probability $Q_{n,p}[n']$ has rank close to n'. In Section 2.4 we explain how the addition of a new row $\mathbf{x}^0 = (x_1, \ldots, x_m, 0)$ together with its transpose can 'optimally' increase the rank of $Q_{n,p}[m+1]$ in terms of the rank of $Q_{n,p}[m]$, and obtain functions of the new row that determine whether the increase of the rank is optimal or not. Additive combinatorics plays an important role here as such functions are linear or quadratic in x_i, so Littlewood-Offord bounds obtained in Section 2.1 can be applied. The technical part of the proof consists of showing that $\mathcal{Q}_{n,p}[(n', n)]$ is endowed with a 'robust' structure that allows us to apply the Littlewood-Offord bounds. The structure, as we will show in Section 2.5, is closely related to blocking small dependencies. Fortunately, such robust structure occurs naturally in $G_{n,p}$ (with probability tending to one). As a consequence, matrices obtained in the process will increase their rank optimally with probability large enough to guarantee that, by the end of the process, all non-trivial dependencies will be removed.

2.4 Growing a symmetric matrix

In this section we study how the addition of a new row and column modifies the rank of a symmetric matrix. Throughout this section, the randomness of the matrix is irrelevant. We write simply Q for an $m \times m$ matrix and define a growth operation which takes a symmetric matrix Q and a vector $\mathbf{x} \in \{0,1\}^m$ and builds Q' by adding to Q the row vector $\mathbf{x}$ and the transpose of $\mathbf{x}^0$.

Definition 2.4.1. *The* growth operation $\Gamma(Q, \mathbf{x}) \to Q'$ *maps an* $m \times m$ *symmetric matrix* Q *with row vectors* $\mathbf{r}_1, \ldots, \mathbf{r}_m$ *and* $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$ *to* Q' *with row* *vectors* $\mathbf{r}'_1, \dots \mathbf{r}'_{m+1}$ *defined as*

$$\mathbf{r}'_i = \begin{cases} (r_{i1}, \dots, r_{im}, x_i) & \text{if } i \in [m], \\ (x_1, \dots, x_m, 0) & \text{if } i = m+1. \end{cases}$$

The matrices here used are deterministic, and we assume throughout this section that Q contains no zero row vectors. It follows then, that no set of rows in Q has a trivial dependency.

The first observation we can make about the growth operation is that

$$rank(Q) \leq rank(Q') \leq rank(Q) + 2. \tag{2.7}$$

To see this, recall that the rank of a matrix equals the dimension of the span of its rows. It is clear that the rows of Q' span at least the same space in the first m coordinates as the rows in Q, so the rank cannot decrease. To see the second inequality, we now take a closer look at the resulting matrix Q'. We claim that the rows of Q' are in the span of

$$\mathcal{R} = \{\mathbf{r}^0_1, \dots, \mathbf{r}^0_m, \mathbf{x}^0, \mathbf{e}_{m+1}\}, \tag{2.8}$$

where $\mathbf{e}_{m+1} = (0, 0, \dots, 0, 1) \in \{0,1\}^{m+1}$ is the canonical vector of the $m+1$-th coordinate. This holds because the rows in $\mathbf{r}'_i, i \in [m]$ are of the form

$$\mathbf{r}'_i = \mathbf{r}^0_i + x_i \mathbf{e}_{m+1}.$$

The second inequality in (2.7) follows from the observation that $\mathcal{R}$ spans a space of dimension at most $rank(Q) + 2$.

So, the rank of Q does not increase by more than 2 after the growth operation. It is clear that if the matrix is already invertible, then the best increase we can obtain by adding a new row/column is 1. The next lemma states the condition under which $rank(Q') = rank(Q) + 2$ is achieved.

Lemma 2.4.2. *Let Q be an $m \times m$ symmetric matrix and let $\mathbf{x} = (x_1, \ldots, x_m)$. Let $Q' = \Gamma(Q, \mathbf{x})$, then*

$$rank(Q') = rank(Q) + 2$$

if and only if $\mathbf{x}$ is independent of the rows of Q.

Proof. Recall that the rank of a matrix is also defined as the dimension spanned by its columns. Suppose $\mathbf{x}$ is independent of the rows in Q. Then, in particular, $rank(Q) < m$. Furthermore, adding $\mathbf{x}$ to Q will increase its rank by one. We claim that the transpose of $\mathbf{x}^0$ is independent of the remaining columns of Q' and so the rank increases further by 1. Otherwise, by the symmetry of Q', $\mathbf{x}^0$ is in the span of $\mathbf{r}'_1, \ldots, \mathbf{r}'_m$ and, in particular $\mathbf{x}$ is in the span of $\mathbf{r}_1, \ldots, \mathbf{r}_m$. This is a contradiction; hence the rank of Q increases by 2 after the growth operation.

Now, suppose $\mathbf{x}$ is not independent of the rows in Q. Recall that the rank of Q' is at most the dimension of the span of $\mathcal{R}$, defined in (2.8). Since $\mathbf{x}$ is in the span of the rows in Q, it follows that

$$span(\{\mathbf{r}_1^0, \ldots, \mathbf{r}_m^0, \mathbf{x}^0, \mathbf{e}_{m+1}\}) = span(\{\mathbf{r}_1^0, \ldots, \mathbf{r}_m^0, \mathbf{e}_{m+1}\}).$$

Hence, the rank of Q increases at most 1. $\square$

Motivated by Lemma 2.4.2, we next define ‘optimal’ rank increase and state a lemma describing linear and quadratic polynomials that determine whether optimal rank increase occurs. When we move to considering random matrices, we will need the coefficients of the polynomials to be independent of $\mathbf{x}$ to apply the Littlewood-Offord concentration bounds.

Definition 2.4.3. *Let Q be an $m \times m$ symmetric matrix and let $\mathbf{x} = (x_1, \ldots, x_m)$. We say that the rank of Q* increases optimally *after the growth operation, if writing*

$Q' = \Gamma(Q, \mathbf{x})$, *we have*

$$rank(Q') = \begin{cases} rank(Q) + 1 & \text{if } rank(Q) = m, \\ rank(Q) + 2 & \text{if } rank(Q) < m. \end{cases}$$

Lemma 2.4.4. *Fix an* $m \times m$ *matrix* Q*, a vector* $\mathbf{x} = (x_1, \ldots, x_m)$ *and let* $Q' = \Gamma(Q, \mathbf{x})$*. If* Q *is singular, then there exists a linear polynomial* f *such that if* Q *does not increase optimally, then* $f(\mathbf{x}) = 0$*. The polynomial is defined as*

$$f(\mathbf{x}) = x_m - \sum_{i \in S} a_i x_i, \tag{2.9}$$

where the coefficients $a_i \neq 0$ *for all* $i \in S$ *and the choice of* $\{a_i\}_{i \in S}$ *depend exclusively on* Q*.*

Proof. Write $r = rank(Q)$ and assume without loss of generality that the columns $\mathbf{c}_1, \ldots, \mathbf{c}_r$ of Q are a maximal set of independent column vectors. We consider $\mathbf{c}_m$, the last column of Q; since $r < m$, it follows that there exist unique integers $a_1, \ldots a_n$ such that $\mathbf{c}_m = \sum_{i=1}^{r} a_i \mathbf{c}_i$. Let $S = \{i \leq r;\ a_i \neq 0\}$ so that in fact,

$$\mathbf{c}_m = \sum_{i \in S} a_i \mathbf{c}_i.$$

Denote by $\{\mathbf{c}'_i\}_{i \in [m]}$ the set of columns of the matrix obtained after adding $\mathbf{x}$. Note, that $\mathbf{c}'_i$ is obtained by appending x_i to the vector $\mathbf{c}_i$. Now, suppose Q does not increase optimally. By Lemma 2.4.2, the vector x is dependent of the set of rows of Q and so the addition of $\mathbf{x}$ does not increase the rank of the matrix. Thus,

$$\mathbf{c}'_m = \sum_{i \in S} a_i \mathbf{c}'_i.$$

In particular, the last coordinates satisfy

$$x_m = \sum_{i \in S} a_i x_i,$$

where all a_i with $i \in S$ are non zero and $\{\mathbf{c}_i\}_{i \in S}$ is a linearly independent set. The result follows. $\square$

For the case $rank(Q) = m$, Q spans the space $\{0,1\}^m$ and so there exist coefficients such that $\mathbf{x} = \sum_{i=1}^{m} a_i \mathbf{r}_i$ and, consequently, the last coordinates of $\mathbf{r}'_1 \ldots, \mathbf{r}'_m$ have to satisfy as well

$$0 = x_{m+1} = \sum_{i=1}^{m} a_i x_i.$$

However, we require another function as the coefficients a_i depend on the vector $\mathbf{x}$ and not exclusively on Q.

Lemma 2.4.5. *Fix an $m \times m$ matrix Q, a vector $\mathbf{x} = (x_1, \ldots, x_m)$ and let $Q' = \Gamma(Q, \mathbf{x})$. If Q is invertible, then there exists a quadratic polynomial g such that Q increases optimally if and only if $g(\mathbf{x}) \neq 0$. The polynomial is defined as*

$$g(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j, \tag{2.10}$$

$g(\mathbf{x})$ is the determinant of Q' and the coefficients a_{ij} depend exclusively on Q.

Proof. By Definition 2.4.3, the rank of Q increases optimally if and only if the determinant of Q' is non-zero. There is a clever way to compute the determinant of Q' as a quadratic function of $\mathbf{x}$. The determinant of Q' is expressed as

$$det(Q') = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j,$$

where $a_{ij} = (-1)^{i+j+1} det(Q^{(i,j)})$ and $Q^{(i,j)}$ is the matrix obtained by deletion of the i-th row and j-th column of Q.

To see this, we make a double application of the cofactor expansion of the determinant. The cofactor C_{ij} of an $m \times m$ matrix M is defined as

$$C_{ij} = (-1)^{i+j} det(M^{(i,j)}),$$

where $M^{(i,j)}$ is the matrix obtained by deletion of the i-th row and j-th column of M. Then the determinant of M is defined either by its j-th column cofactor expansion,

$$det(M) = \sum_{i=1}^{m} C_{ij} M_{ij};$$

or by its i-th row cofactor expansion,

$$det(M) = \sum_{j=1}^{m} C_{ij} M_{ij}.$$

To get the desired expression we write the determinant of Q' with its $m+1$-th column cofactor expansion; let $x_{m+1} = 0$, then

$$det(Q') = \sum_{i=1}^{m+1} (-1)^{i+m+1} det(Q'^{(i,m+1)}) x_i. \tag{2.11}$$

In the same way, for each $i \in [m]$ we can express $det(Q'^{(i,m+1)})$, with its m-th row cofactor expansion. Note that after deleting the j-th column and m-th row of $Q'^{(i,m+1)}$ we get exactly $Q^{(i,j)}$, thus

$$\det(Q'^{(i,m+1)}) = \sum_{j=1}^{m} (-1)^{m+j} det(Q^{(i,j)}) x_j.$$

The result follows by plugging these equalities into (2.11). □

In the next section we define conditions on the structure of $Q_{n,p}$ which guarantee that, with high probability, the functions in Lemmas 2.4.4 and 2.4.5 have a minimum number of non-zero coefficients and so we can apply efficiently the Littlewood-Offord concentration bounds of Section 2.1 to prove Theorem 2.3.2.

2.5 Robust sequences and the proof of: $rank(Q_{n,p}) = n - i(Q_{n,p})$

In this section we describe the structure that $G_{n,p}$ must possess in order for the vertex exposure method to establish that $rank(Q_{n,p}) = n - i(Q_{n,p})$ (Theorem 2.3.2). This structure guarantees, with high probability, that the difference between the rank and the number of non-isolated vertices decreases optimally through the augmentation process; we formalize this intuition in Theorem 2.5.7. We close the section with the proof of Theorem 2.3.2.

The following definitions are equivalent to those in [11].

Definition 2.5.1. *Fix a set of vectors* $\{\mathbf{r}_i\}_{i\in S}$*, where* $\mathbf{r}_i = (\mathbf{r}_{i1}, \ldots, \mathbf{r}_{in})$ *for all* $i \in S$*. We say that the* j*-th coordinate is a* blocking column *in* S *if and only if there exist exactly 1 non-zero value among* $\mathbf{r}_{ij}$*,* $i \in S$*.*

Definition 2.5.2. *A set of non-zero vectors* $\{\mathbf{r}_i\}_{i\in S}$*,* $S \subset [n]$*, is blocked if it has at least two blocking columns.*

We will say that a matrix Q *is* k*-blocked if every set* $\{\mathbf{r}_i;\, i \in S\}$ *of non-zero rows in* Q *with* $2 \leq |S| \leq k$ *is blocked.*

Remark 2.5.3. *If* Q *is* k*-blocked then any non-trivial dependency in* Q *has cardinality greater than* k*. Moreover, this property still holds in* Q *after deleting any given column.*

The remark follows from the observation that a non-trivial dependency cannot have blocked columns. The condition that a blocked set has two blocked columns ensures that even after deleting a column in Q, blocked sets still have at least one blocked column.

Recall that we denote by $Q_{n,p}[m]$ the adjacency matrix of G_m, which we take to be the subgraph of $G_{n,p}$ induced by $[m]$. Note that the vertex exposure method is equivalent to the growth operation defined in Section 2.4. This is so because, given the matrix $Q_{n,p}[m]$, we can write $Q_{n,p}[m+1] = \Gamma(Q_{n,p}[m], \mathbf{x})$, where $\mathbf{x}$ is a random vector. Lemmas 2.4.4 and 2.4.5 provide functions which determine when the growth operation increases the rank optimally. Moreover, these functions are closely related to sets of independent rows in $Q_{n,p}[m]$ and sets of non-trivial dependencies.

The exclusion of small non-trivial dependencies allows us to obtain bounds, via the Littlewood-Offord inequalities, on the probability that the rank increases optimally under assumption that we explain below. In the following definition we insist that the necessary conditions to apply the Littlewood-Offord bounds are fulfilled in all the matrices of $\mathcal{Q}_{n,p}[n',m] := (Q_{n,p}[n'],\ldots,Q_{n,p}[m])$.

Definition 2.5.4. *Let $k = \frac{\ln\ln n}{2p}$. We say that the sequence $\mathcal{Q}_{n,p}[n',m]$ is* robust *if for all $n' \leq j \leq m$, $Q_{n,p}[j]$ is k-blocked, and additionally $Q_{n,p}[j]$ has at most $(p\ln n)^{-1}$ rows which have at most 2 non-zero entries.*

Let $\mathcal{Q}_{n,p} = \mathcal{Q}_{n,p}[n',n]$ and set $n' = \lceil \alpha n \rceil$; the value of $\alpha \in (0,1)$ is carefully chosen so that Lemma 2.5.8 below holds. However, apart from its role in Lemma 2.5.8, the actual value of α is not relevant throughout the rest of the section.

Let Z_m be the number of all-zero rows in $Q_{n,p}[m]$ which contain a non-zero coordinate in $Q_{n,p}[m+1]$. The following two lemmas bound from above the probability that the rank does not increase optimally, under the assumption that the sequence is robust and $Z_m = 0$. We remark that the bounds achieved are directly related to the value of k in the definition of robust sequence.

Lemma 2.5.5. *Fix $n' \leq m \leq n$ and consider $Q_{n,p}[m]$ and $Q_{n,p}[m+1] = \Gamma(Q_{n,p}[m], \mathbf{x})$, with $\mathbf{x}$ a random vector whose entries are distributed independently as Bernouilli with parameter p. Conditioning on the event that $\mathcal{Q}_{n,p}[n',m]$ is robust and $Z_m = 0$; if $rank(Q_{n,p}[m]) + i(Q_{n,p}[m]) < m$, then with probability at most $2^{3/2}(\ln\ln n)^{-1/2}$*

$$rank(Q_{n,p}[m+1]) - rank(Q_{n,p}[m]) < 2.$$

Proof. If $Z_m = 0$, then the rows that contain only zero in $Q_{n,p}[m]$ do so in $Q_{n,p}[m+1]$ as well. In this case, removing those rows and their corresponding columns will not change the rank of $Q_{n,p}[m]$ and $Q_{n,p}[m+1]$. As a consequence, by conditioning on $Z_m = 0$ and given the independence of the coordinates of $\mathbf{x}$ we can assume that $Q_{n,p}[m]$ has no all-zero rows.

By Lemma 2.4.4, if $rank(Q_{n,p}[m]) < m$, there exist $S \subset [m]$ and non-zero coefficients a_i, $i \in S$ such that if the rank does not increase optimally, then

$$x_m - \sum_{i \in S} a_i x_i = 0. \tag{2.12}$$

Furthermore, by construction, the set $\{\mathbf{r}_i\}_{i \in S}$ is linearly independent; and $\{\mathbf{r}_m\} \cup \{\mathbf{r}_i\}_{i \in S}$ is a non-trivial dependency of size $|S| + 1$. Since $Q_{n,p}[m]$ is blocked, it is the case that $|S| \geq k$. Thus, the linear function in (2.12) has at least k non-zero coefficients. Therefore, by the Littlewood-Offord concentration bound (Lemma 2.1.5)

$$\mathbf{P}\left\{x_k - \sum_{i \in S} a_i x_i = 0\right\} \leq \frac{2}{\sqrt{kp}} \leq \frac{2^{3/2}}{(\ln \ln n)^{1/2}}.$$

$\square$

Lemma 2.5.6. *Fix $n' \leq m \leq n$ and consider $Q_{n,p}[m]$ and $Q_{n,p}[m+1] = \Gamma(Q_{n,p}[m], \mathbf{x})$, with $\mathbf{x}$ a random vector whose entries are distributed independently as Bernouilli with parameter p. Conditioning on the event that $\mathcal{Q}_{n,p}[n', m]$ is robust and $Z_m = 0$; if $rank(Q_{n,p}[m]) + i(Q_{n,p}[m]) = m$, then with probability at most $5(\ln \ln n)^{-1/4}$*

$$rank(Q_{n,p}[m+1]) - rank(Q_{n,p}[m]) < 1.$$

Proof. As explained in the previous proof, conditioning on $Z_m = 0$ we can to assume that $Q_{n,p}[m]$ has no all-zero rows and $rank(Q_{n,p}[m]) = m$.

By Lemma 2.4.5, if the rank fails to increase then

$$det(Q_{n,p}[m+1]) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j = 0, \tag{2.13}$$

where $a_{ij} = 0$ if and only if the minor $Q_{n,p}[m]^{(ij)}$ is singular. To apply the quadratic Littlewood-Offord concentration we need to bound the number of non-zero coefficients a_{ij}.

Fix $j \in [m]$. We claim that the number of non-zero coefficients a_{ij} is greater than k if the deletion of the j-th column does not create all-zero rows in the matrix. To see this, note that if the deletion of the j-th column does not create a zero row, the resulting $m \times m-1$ matrix has rank $m-1$ and so there exists (up to scaling factors) a unique vanishing linear combination

$$\sum_{i=1}^{m} a_i \mathbf{r}_i^j = 0,$$

where $\mathbf{r}_i^j$ is obtained from $\mathbf{r}_i$ by deleting the j-th coordinate. Now, the matrix becomes invertible again if and only if we delete a row $\mathbf{r}_i^j$ with $a_i \neq 0$. The set of rows with $a_i \neq 0$ is, by definition, a non-trivial dependency, so it holds that its size is greater than k.

Finally, since $\mathcal{Q}_{n,p}[n', m]$ is robust, no more than $(p \ln n)^{-1}$ rows in $Q_{n,p}[m]$ have less than 2 non-zero entries. Consequently, at most $(p \ln n)^{-1}$ columns leave zero rows after their removal. It is clear then that at least k indices j have at least k indices i such that $a_{ij} \neq 0$. Therefore, by Lemma 2.1.7, equation (2.13) then holds with probability at most $3\left(\frac{kp}{2}\right)^{-1/4}$. The result follows. □

We now define an auxiliary variable which allows us to use a martingale-like approach. Let $Y_m = m - rank(Q_{n,p}[m]) - i(Q_{n,p}[m])$ be the decrease in rank due to non trivial dependencies in $Q_{n,p}[m]$. Let

$$X_m = \begin{cases} 4^{Y_m} & \text{if } Y_m > 0 \text{ and } \mathcal{Q}_{n,p}[n', m] \text{ is robust,} \\ 0 & \text{otherwise.} \end{cases}$$

The following theorem formalizes the intuition that the variable Y_m tends to decrease during the augmentation process.

Theorem 2.5.7 ([11]). *Given a graph $G_{n,p}$ with $p = \frac{c \ln n}{n} < \frac{1}{2}$ and $c > \frac{1}{2}$. For any sequence $\mathcal{Q}_{n,p}[n', m] = \{Q_{n,p}[n'], \ldots, Q_{n,p}[m]\}$ as defined above,*

$$\mathbf{E}\left[X_{m+1} \mid \mathcal{Q}_{n,p}[n', m]\right] \leq \frac{3}{5} X_m + 20(\ln \ln n)^{-1/4}. \tag{2.14}$$

Furthermore,

$$\mathbf{E}\left[X_n \mid \mathcal{Q}_{n,p}[n']\right] \le \left(\frac{3}{5}\right)^{n-n'} X_{n'} + 50(\ln\ln n)^{-1/4}. \tag{2.15}$$

Proof. Assume $\mathcal{Q}_{n,p}[m]$ is robust, otherwise $X_{m+1} = 0$ and the inequality holds trivially. We analyse X_{m+1} by conditioning according to the events $\{Z_m = i\}$, $0 \le i \le m$.

$$\mathbf{E}\left[X_{m+1} \mid \mathcal{Q}_{n,p}[m]\right] = \sum_{i=0}^{m} \mathbf{E}\left[X_{m+1} \mid \mathcal{Q}_{n,p}[m], Z_m = i\right] \mathbf{P}\left\{Z_m = i \mid \mathcal{Q}_{n,p}[m]\right\}.$$

If $Z_m = i > 0$ then, regardless of the increase in the rank,

$$Y_{m+1} \le Y_m + Z_m + 1.$$

Since $\mathcal{Q}_{n,p}[n', m]$ is robust, there are at most $\lfloor (p\ln n)^{-1} \rfloor$ all-zero rows in $Q_{n,p}[m]$. The distribution of Z_m is binomial, thus $\mathbf{P}\left\{Z_m = i \mid \mathcal{Q}_{n,p}[n', m]\right\}$ is bounded by

$$\begin{aligned}
\mathbf{P}\left\{Z_m \ge i \mid \mathcal{Q}_{n,p}[n', m]\right\} &\le \binom{\lfloor (p\ln n)^{-1} \rfloor}{i} p^i \\
&\le \lfloor (p\ln n)^{-1} \rfloor^i p^i \\
&= (\ln n)^{-i}.
\end{aligned}$$

Summing over i, it follows that for n large enough (we need that $\ln n \ge 5$, say),

$$\begin{aligned}
\mathbf{E}\left[X_{m+1}\mathbf{1}_{[Z_m > 0]} \mid \mathcal{Q}_{n,p}[n', m]\right] &\le \sum_{i=1}^{m} 4^{Y_m + i + 1} (\ln n)^{-i} \\
&\le 4X_m \sum_{i=1}^{\infty} \frac{4^i}{(\ln n)^i} \\
&\le \frac{20X_m}{\ln n}.
\end{aligned}$$

If $Z_m = 0$, then we consider separately the cases $Y_m = 0$ and $Y_m > 0$. If $Y_m = 0$, then $X_m = 0$. In this case, $X_{m+1} = 0$ if the rank increases optimally, and otherwise $X_{m+1} = 4$. Lemma 2.5.6 then yields

$$\mathbf{E}\left[X_{m+1}\mathbf{1}_{[Y_m = 0]} \mid \mathcal{Q}_{n,p}[n', m], Z_m = 0\right] \le \frac{4 \cdot 5}{(\ln\ln n)^{1/4}}.$$

Likewise; if $Y_m > 0$, then $X_{m+1} = \frac{X_m}{4}$ if the rank increases optimally, and otherwise $X_{m+1} \le 4X_m$. By Lemma 2.5.6 we then have

$$\mathbf{E}\left[X_{m+1}\mathbf{1}_{[Y_m>0]} \mid \mathcal{Q}_{n,p}[n',m], Z_m = 0\right] \le \frac{X_m}{4} + \frac{2^{3/2}\cdot 4X_m}{(\ln\ln n)^{1/2}}.$$

Using that $\mathbf{P}\left\{Z_m = 0 \mid \mathcal{Q}_{n,p}[n',m]\right\} \le 1$; all the cases together yield, for n large enough,

$$\begin{aligned}\mathbf{E}\left[X_{m+1} \mid \mathcal{Q}_{n,p}[n',m]\right] &\le X_m\left(20\ln n + \frac{1}{4} + \frac{16}{(\ln\ln n)^{1/2}}\right) + \frac{20}{(\ln\ln n)^{1/4}}\\ &\le \frac{3}{5}X_m + 20(\ln\ln n)^{-1/4}.\end{aligned}$$

Next, we prove by induction that for all integer $0 < k \le n - n'$

$$\mathbf{E}\left[X_{n'+k} \mid \mathcal{Q}_{n,p}[n']\right] \le \left(\frac{3}{5}\right)^k X_{n'} + \sum_{i=1}^{k}\left(\frac{3}{5}\right)^{i-1}\frac{20}{(\ln\ln n)^{1/4}}. \tag{2.16}$$

The base of the induction follows from the previous inequality. We assume now that (2.16) holds for a fixed $k < n - n'$ and prove that it holds for $k+1$ as well. Using the Tower Law for conditional expectations, we have

$$\begin{aligned}\mathbf{E}\left[X_{n'+k+1} \mid \mathcal{Q}_{n,p}[n']\right] &= \mathbf{E}\left[\mathbf{E}\left[X_{n'+k+1} \mid \mathcal{Q}_{n,p}[n', n'+k]\right] \mid \mathcal{Q}_{n,p}[n']\right]\\ &\le \mathbf{E}\left[\frac{3}{5}X_{m+k} + \frac{20}{(\ln\ln n)^{1/4}} \mid \mathcal{Q}_{n,p}[n']\right]\\ &\le \frac{3}{5}\left(\left(\frac{3}{5}\right)^k X_{n'} + \sum_{i=1}^{k}\left(\frac{3}{5}\right)^{i-1}\frac{20}{(\ln\ln n)^{1/4}}\right) + \frac{20}{\ln\ln n)^{1/4}}\\ &= \left(\frac{3}{5}\right)^{k+1} X_{n'} + \sum_{i=1}^{k+1}\left(\frac{3}{5}\right)^{i-1}\frac{20}{(\ln\ln n)^{1/4}}.\end{aligned}$$

The first inequality follows from (2.14), whereas the second inequality follows by the induction hypothesis. Let $k = n - n'$, this yields

$$\begin{aligned}\mathbf{E}\left[X_n \mid \mathcal{Q}_{n,p}[n']\right] &\le \left(\frac{3}{5}\right)^{n-n'} X_{n'} + \sum_{i=1}^{n-n'}\left(\frac{3}{5}\right)^{i-1}\frac{20}{(\ln\ln n)^{1/4}}\\ &\le \left(\frac{3}{5}\right)^{n-n'} X_{n'} + \frac{50}{(\ln\ln n)^{1/4}}.\end{aligned}$$

□

We now present the proof of Theorem 2.3.2 assuming that the robust structure of $\mathcal{Q}_{n,p}[n',n]$ occurs with probability tending to 1, for $p = \frac{c\ln n}{n}$ and $c > \frac{1}{2}$ fixed.

Lemma 2.5.8 ([11])**.** *Given a random matrix $Q_{n,p}$ with $p = \frac{c\ln n}{n} < \frac{1}{2}$, $c > \frac{1}{2}$. For $\alpha \in (0,1)$ with $\frac{1}{2} < \alpha c < \frac{3}{5}$, consider the matrix sequence $\mathcal{Q}_{n,p}[n',n]$ with $n' = \alpha n$. Then for any fixed $\varepsilon' > 0$, the probability that $\mathcal{Q}_{n,p}[n',n]$ is not robust is $O(n^{1-2c\alpha+\varepsilon'})$.*

Its somewhat technical proof is omitted as we prove a more general version in Chapter 4.

Proof of Theorem 2.3.2. Write $\mathcal{Q}_{n,p}$ for $\mathcal{Q}_{n,p}[n',n]$. We show that $rank(Q_{n,p}) = n - i(Q_{n,p})$ with probability $1 - O((\ln\ln n)^{-1/4})$. By definition, $X_n < 1$ if and only if $rank(Q_{n,p}) = n - i(Q_{n,p})$ or the sequence $\mathcal{Q}_{n,p}$ is not robust. Hence,

$$\mathbf{P}\left\{rank(Q_{n,p}) + i(Q_{n,p}) < n\right\} \leq \mathbf{P}\left\{X_n \geq 1\right\} + \mathbf{P}\left\{\mathcal{Q}_{n,p} \text{ is not robust}\right\}.$$

Let B_1 be the event that $rank(Q_{n,p}[n']) \geq \left(1 - \frac{1-\alpha}{4\alpha}\right) n'$. Then the probability that $rank(Q_{n,p}) < n - i(Q_{n,p})$ is at most

$$\mathbf{P}\left\{X_n \geq 1, B_1\right\} + \mathbf{P}\left\{\overline{B}_1\right\} + \mathbf{P}\left\{\mathcal{Q}_{n,p} \text{ is not robust}\right\}. \tag{2.17}$$

So it suffices to show that each term above is at most, say $60(\ln\ln n)^{-1/4}$. The last two terms in (2.17) decay faster than $60(\ln\ln n)^{-1/4}$ for n large enough. To see this, first, Lemma 2.5.8 yields $\mathbf{P}\left\{\mathcal{Q}_{n,p} \text{ is not robust}\right\} \leq n^{1-2c\alpha+\varepsilon'}$, where we choose ε' small enough that $1 - 2c\alpha + \varepsilon' < 0$.

Next, setting $\varepsilon = \frac{1-\alpha}{4\alpha}$ in Lemma 2.3.4 yields that there exists constants $c_1, c_2 > 0$ such that with probability at most $c_2^n e^{-c_1 n\ln n}$ the event B_1 does not hold.

We use the Tower Law for conditional expectation to bound the first term in (2.17),

$$\mathbf{P}\left\{X_n \mathbf{1}_{[B_1]} \geq 1\right\} \leq \mathbf{E}\left[X_n \mathbf{1}_{[B_1]}\right] = \mathbf{E}\left[\mathbf{E}\left[X_n \mathbf{1}_{[B_1]} \mid \mathcal{Q}_{n,p}[n']\right]\right].$$

Now, B_1 is measurable with respect to $Q_{n,p}[n']$, thus

$$\mathbf{E}\left[\mathbf{E}\left[X_n \mathbf{1}_{[B_1]} \mid \mathcal{Q}_{n,p}[n']\right]\right] = \mathbf{E}\left[\mathbf{E}\left[X_n \mid \mathcal{Q}_{n,p}[n']\right] \mathbf{1}_{[B_1]}\right].$$

We are in condition to use (2.15) in Theorem 2.5.7, the conditional expectation above is at most

$$\mathbf{E}\left[\left((3/5)^{n-n'} X_{n'} + 50(\ln\ln n)^{-1/4}\right) \mathbf{1}_{[B_1]}\right] \leq \left(\frac{3\sqrt{2}}{5}\right)^{n-n'} + 50(\ln\ln n)^{-1/4}$$

The last inequality is obtained since the variable is zero except on the event B_1, where we have that $X_{n'} \leq 2^{(1-\alpha)\frac{n}{2}}$. Thus we concluded that

$$\mathbf{P}\left\{X_n > 1, B_1\right\} \leq \left(\frac{3\sqrt{2}}{5}\right)^{(1-\alpha)n} + 50(\ln\ln n)^{-1/4} \leq 60(\ln\ln n)^{-1/4}.$$

The last inequality holds for n large enough as $\frac{3\sqrt{2}}{5} < 1$. Combining the three bounds we obtain the desired result. $\square$

Theorem 2.3.2 currently yields a bound that is not strong enough to give useful bounds for $\mathbf{P}\left\{rank(Q_{n,p}) = 0\right\}$ with p being a random function. In the next chapter we extend Theorem 2.3.2 to matrices which possess a bounded number of deterministic rather than random entries. The latter extension allows us to handle the particular case of Q_{n,p^*} where p^* is the hitting time corresponding to not having all-zero rows in $Q_{n,p}$.

CHAPTER 3
A Graph Process approach for the rank of $Q_{n,p}$

We want to bring the techniques used in random graph theory to analyse the behaviour of sparse random matrices. In particular, in this thesis we study the invertibility of $Q_{n,p}$ via a careful investigation of suitable structural properties of the corresponding graph $G_{n,p}$.

Random graphs were introduced by Erdős and Rényi back in 1959 [16] and have been widely studied ever since (see, for example, the books by Bollobás [5] and Janson, Luczak and Rucinski [23]).

One useful perspective on Erdős-Rényi random graphs is given by coupling the family $\{G_{n,p}\}_{p\in(0,1)}$ to form a stochastic process. Formally, the graph process $\{G_{n,p}\}_{p\in(0,1)}$ is built from a set of variables $\{U_{ij};\, i<j\in[n]\}$ which are independent and uniform on $(0,1)$. Then for $p\in(0,1)$, we define $G_{n,p}$ as the graph with vertex set $[n]$ and edge set $\{e=ij;\, U_e\le p\}$. Hence, the process $\{G_{n,p}\}_{p\in(0,1)}$ starts with an empty graph and adds edges e one by one according to its time of arrival U_e. We see that the process $\{G_{n,p}\}_{p\in(0,1)}$ is increasing, in the sense that for $p<p'$, with probability one $G_{n,p}$ is a subgraph of $G_{n,p'}$.

A *graph property* $\mathcal{P}$ is a class of graphs closed under isomorphism. We say that a graph property is *increasing* (or respectively decreasing) if it closed under addition (respectively, deletion) of edges. The graph process is widely used to study how a given property of the graph $G_{n,p}$ evolves as p goes from zero to one.

In the following sections we define 'hitting times' and comment on some classic results of random graph theory that show a relation between the hitting times of monotone graph properties and the minimum degree of $G_{n,p}$. We then present the novel contribution of the thesis (Theorem 3.2.1), which states that in $\{Q_{n,p}\}_{p\in(0,1)}$, the hitting time for $rank(Q_{n,p}) = 0$ equals the hitting time for $\delta(G_{n,p}) \geq 1$ a.a.s. The last two sections of this chapter are dedicated to proving Theorem 3.2.1 assuming Theorem 3.3.5 below, which itself is proven in Chapter 4.

3.1 The minimum degree condition for hitting times

One of the most important and well known monotone properties is that of being connected. In fact, in their first paper on random graphs, Erdős and Rényi investigated the probability that a graph on n vertices and about $n \ln n$ edges uniformly chosen at random is connected. They proved the following theorems (although not directly for graphs $G_{n,p}$).

Theorem 3.1.1. *Let $G_{n,p}$ be a random graph with $p = \frac{\log n + x}{n}$ and $x \in \mathbb{R}$. Then,*

$$\lim_{n\to\infty} \mathbf{P}\left\{G_{n,p} \textit{ is connected}\right\} = e^{-e^{-x}}.$$

Theorem 3.1.2. *Let $G_{n,p}$ be a random graph with $p = \frac{\log n + x}{n}$ and $x \in \mathbb{R}$. Let $i(G_{n,p})$ denote the number of isolated vertices in $G_{n,p}$. Then for any $k \in N$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{i(G_{n,p}) = k\right\} = \frac{e^{-xk}e^{-e^{-x}}}{k!}.$$

That is, $i(G_{n,p})$ converges in distribution to a Poisson variable with mean $\lambda = e^{-x}$.

We can observe that the probability that $G_{n,p}$ is connected is (asymptotically) equal to the probability that $G_{n,p}$ has no isolated vertices. It is clear that the latter property is necessary for connectivity. In [17], Erdős and Rényi made the link between connectedness and minimum degree, showing that for p near the connectivity threshold, the number of vertices one must delete in order to disconnect the graph is with high probability given by $\delta(G_{n,p})$. In particular, if $i(G_{n,p}) = 0$ then the graph is connected a.a.s. Such a result holds for any fix p, however, a stronger version of this fact holds for $\{G_{n,p}\}_{p\in(0,1)}$. To state this we need the notion of a hitting time.

Definition 3.1.3. *In the graph process* $\{G_{n,p}\}_{p\in(0,1)}$*, let* $\mathcal{P}$ *be a graph property, then the hitting time of* $\mathcal{P}$ *is defined as the random variable*

$$\tau_{\mathcal{P}} = \tau_{\mathcal{P}}(n) := \inf\{p \in (0,1);\ \mathcal{P} \text{ holds in } G_{n,p}\}.$$

We will set the infimum of an empty set as ∞*.*

We define p^* to be the hitting time for $\delta(G_{n,p}) \geq 1$ and τ_C to be the hitting time for connectivity. Then $p^* \leq \tau_C$ and moreover, equality holds a.a.s.

Theorem 3.1.4 ([7])**.** *In the graph process* $\{G_{n,p}\}_{p\in(0,1)}$*,*

$$\mathbf{P}\left\{\tau_C(n) = p^*(n)\right\} \to 1,\ \text{ as } n \to \infty.$$

Results such as Theorem 3.1.4 are often called hitting time theorems. For many monotone graph properties $\mathcal{P}$, a minimum degree requirement is an obvious necessary condition. Interestingly, for the random graph process, it is also often the case that $\tau_{\mathcal{P}}(n)$ is a.a.s. equal to the first time the minimum degree condition is satisfied.

Bollobás and Frieze established one of the famous early hitting time theorems, on perfect matchings and Hamiltonian cycles [6]. The former events require $\delta(G)$ to be at least 1 and $2k$, respectively. Bollobás and Frieze showed that if n is even then, the hitting time for a perfect matching is equal a.a.s. to p^* and the hitting time of $\delta = 2k$ coincides with the hitting time of having k disjoint Hamiltonian cycles. These results extended previous work of Komlós and Szemerédi [27], which discussed the case of a single Hamiltonian cycle.

One of the key steps of the proof in [6] is a general lemma about 'minimum-degree'-type hitting time theorems. The lemma essentially states that for monotone graph properties, suitably precise estimates such as those in Theorems 3.1.1 and 3.1.2 can be transformed into hitting time theorems. It is proved by use of the so-called 'edge-exposure martingale', and an 'edge rejection procedure'.

Although the property of having full rank is not monotone itself, the monotonicity of $G_{n,p}$ is crucial. We will use, for example, the FKG inequality [18] which establishes positive correlation between increasing functions in certain probability spaces. The FKG inequality is an extension of Harris's lemma [21], which already implies that any two increasing graph properties in $G_{n,p}$ are positively correlated.

Lemma 3.1.5. *Let $\mathcal{A}(n)$ and $\mathcal{B}(n)$ be increasing graph properties of $G_{n,p}$, then*

$$\mathbf{P}\left\{\mathcal{A}(n), \mathcal{B}(n)\right\} \geq \mathbf{P}\left\{\mathcal{A}\right\}\mathbf{P}\left\{\mathcal{B}\right\}.$$

If $\mathcal{A}(n)$ is an increasing graph property and $\mathcal{B}(n)$ is a decreasing graph property, then

$$\mathbf{P}\left\{\mathcal{A}(n), \mathcal{B}(n)\right\} \leq \mathbf{P}\left\{\mathcal{A}\right\}\mathbf{P}\left\{\mathcal{B}\right\}.$$

See the appendix for a proof of the lemma above.

In the following section we make some remarks about $rank(Q_{n,p})$ and present the novel result of the thesis, which is a hitting time theorem for full rank in the random graph process.

3.2 The hitting time for full rank

We study the rank of $Q_{n,p}$ as a graph property as a means of understanding the structure of the dependencies that exist between the rows of $Q_{n,p}$ (see also, [11]). A main observation is that the rank of a graph is not monotone under addition of edges (e.g., the path on 4 vertices has full rank, but the cycle on 4 vertices has rank 2). Therefore, the machinery for monotone properties can not be directly applied to the event of non-singularity, that is $rank(Q_{n,p}) = n$.

In this thesis we focus on the rank of $Q_{n,p}$ with p in a small interval around $\frac{\ln n}{n}$, which is the threshold for p^*. In particular, we are interested in the event $Q_{n,p}$ has full rank, $rank(Q_{n,p}) = n$, and its hitting time τ_F.

In spite of the lack of monotonicity, the rank of $Q_{n,p}$ has a direct relation with the number of isolated vertices, as has been proven by Costello and Vu [11]. Theorem 2.3.2

implies that the giant component of $G_{n,p}$ has full rank if p is fixed and lies in a small interval of around $\frac{\ln n}{n}$. This suggests that the hitting times τ_F and p^* coincide, and this is indeed the case

Theorem 3.2.1. *For the matrix process* $\{Q_{n,p}\}_{p\in(0,1)}$*,*

$$\mathbf{P}\left\{\tau_F(n) = p^*(n)\right\} \to 1, \text{ as } n \to \infty.$$

Clearly, a matrix with full rank contains no all-zeros row so $p^*(n) \leq \tau_F(n)$. Therefore, the following proposition is equivalent to Theorem 3.2.1.

Proposition 3.2.2. *In the matrix process* $\{Q_{n,p}\}_{p\in(0,1)}$ *we have*

$$\mathbf{P}\left\{rank(Q_{n,p^*}) = n\right\} \to 1, \text{ as } n \to \infty.$$

This proposition is similar to Theorem 2.3.2, however, note that Theorem 2.3.2 is valid for fixed p, so it can not be extended directly to a random point such as p^*. Instead, we condition on the 'last stage of the evolution' before connectivity. In the following paragraph, we sketch the proof of Proposition 3.2.2.

To prove Proposition 3.2.2, we fix p_1 small enough that $\mathbf{P}\{p^* > p_1\}$ is high, and partition the probability space according to the set M of isolated vertices in G_{n,p_1} and on the neighbourhoods of these vertices in G_{n,p^*}. By symmetry, we may assume that $M = \{1, \ldots, |M|\}$, and write $S_1, \ldots, S_{|M|}$ for the neighbourhoods of these vertices in G_{n,p^*}. For fixed $l \geq 1$, consider working on the event that $M = [l]$ and that $N_{G_{n,p^*}}(i) = S_i$ for each $i \in [l]$. In this case we may replace G_{n,p^*} by another graph, later denoted $G^{\mathcal{T}}_{n,p^*}$, in which the neighbourhoods of vertices $\{1, \ldots, l\}$ are deterministic rather than random, without changing the event under consideration. Theorem 3.3.5, below, generalizes Costello and Vu's Theorem (2.3.2) to the case where a subset of the vertices have deterministic rather than random neighbourhoods. Applying this theorem will then allow us to prove Proposition 3.2.2. We now proceed with the details.

3.3 The graph $G^{\mathcal{T}}_{n,p}$ and its relation with G_{n,p^*}

In this section we derive an expression for the probability that the adjacency matrix of G_{n,p^*} has full rank. To do so we establish suitable conditional independence between events depending on different edges of the graph process. This follows from the basic observation that, given two independent σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$, for every $A_1, B_1 \in \mathcal{F}_1$ and $A_2, B_2 \in \mathcal{F}_2$, we have

$$\mathbf{P}\left\{A_1, A_2 \mid B_1, B_2\right\} = \mathbf{P}\left\{A_1 \mid B_1\right\}\mathbf{P}\left\{A_2 \mid B_2\right\}.$$

To see this, it suffices to express the term on the left-hand side of the equation using the definition of conditional probability and then use the independence of $\mathcal{F}_1$ and $\mathcal{F}_2$ to separate the terms accordingly.

For any positive integer $l < n$, let

$$\mathcal{F}_l := \sigma(\{U_{ij};\, i \in [l], j \in [n]\}), \qquad \mathcal{F}_{>l} := \sigma(\{U_{ij};\, i, j \in [n] \setminus [l]\}).$$

For any positive integer $l < n$ and for any real $p \in (0,1)$, we define the following two events; let $A_l = A_l(p)$ be the event that $G_{n,p}[n \setminus l]$ has minimum degree at least 1 and let $B_l = B_l(p)$ be the event that each vertex $i \in [l]$ is isolated in $G_{n,p}$.

Additionally, let τ_l be the hitting time of the event that all vertices in $[l]$ are not isolated,

$$\tau_l := \min\{p \in (0,1);\, \forall i \in [l],\, U_{ij} \le p \text{ for some } j \in [n]\};$$

and for any sequence $\mathcal{T} = (S_i)_{i\in[l]}$, let $C_{\mathcal{T}}$ be the event that $N_{G_{n,\tau_l}}(i) = S_i$ for each vertex $i \in [l]$.

Definition 3.3.1. *For any positive integers n, l, L with $1 \le l \le n$, let $\mathcal{M}^n(l, L)$ be the collection of ordered sequences of subsets of $[n]$, $\mathcal{T} = (S_i)_{i\in[l]}$ of pairwise disjoint sets such that for each $i \in [l]$: $S_i \subset [n] \setminus [l]$ and $1 \le |S_i| \le L$.*

Definition 3.3.2. *For any $p \in (0,1)$, positive integers n, l, L with $1 \leq l \leq n$, and $\mathcal{T} \in \mathcal{M}^n(l, L)$, we write $G^{\mathcal{T}}_{n,p}$ for the graph with vertex set $[n]$ obtained from $G_{n,p}$ by replacing the neighbourhood of each vertex $i \in [l]$ by the set S_i, so that $N_{G^{\mathcal{T}}_{n,p}}(i) = S_i$.*

We write A_l, B_l instead of $A_l(p), B_l(p)$ when the dependence on p and l is clear from the context. We observe that B_l, $C_{\mathcal{T}}$ and τ_l are measurable with respect to $\mathcal{F}_l$, whereas A_l is measurable with respect to $\mathcal{F}_{>l}$. Furthermore, the edges that are random in $G^{\mathcal{T}}_{n,p}$ are precisely those corresponding to the random variables generated by $\mathcal{F}_{>l}$. The next lemma uses this fact to split the conditional probability of $Y(G_{n,p^*}) = 0$ given $A_l(p), B_l(p)$ and $C_{\mathcal{T}}$ in such a way that the random value p^* is replaced with an integral. Later this allows us to use uniform bounds of the type of Theorem 2.3.2. To shorten some coming formulas, for a $n \times n$ symmetric matrix Q, let

$$Y(Q) := n - rank(Q) - i(Q).$$

Lemma 3.3.3. *For positive integers l, L, K and fixed $p \in (0,1)$, let $A_l = A_l(p)$ and $B_l = B_l(p)$. If $\mathcal{T} \in \mathcal{M}^n(l, L)$, then the difference*

$$\left| \mathbf{P}\left\{Y(G_{n,p^*}) = 0 \mid A_l, B_l, C_{\mathcal{T}}\right\} - \sum_{i=0}^{K-1} \mathbf{P}\left\{Y(G^{\mathcal{T}}_{n,\frac{i+1}{K}}) = 0 \mid A_l\right\} \mathbf{P}\left\{\tau_l \in [\tfrac{i}{K}, \tfrac{i+1}{K}) \mid B_l, C_{\mathcal{T}}\right\} \right|$$

is at most

$$\frac{n^4}{K\mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}\right\}}.$$

Proof. If A_l, B_l and $C_{\mathcal{T}}$ all occur, then necessarily $rank(G_{n,p^*}) = rank(G^{\mathcal{T}}_{n,p^*})$ and $\tau_l = p^*$. For any $K \in \mathbb{N}$, $\mathbf{P}\left\{Y(G_{n,p^*}) = 0 \mid A_l, B_l, C_{\mathcal{T}}\right\}$ is equal to

$$\sum_{i=0}^{K-1} \mathbf{P}\left\{Y(G^{\mathcal{T}}_{n,\tau_l}) = 0, \tau_l \in [\tfrac{i}{K}, \tfrac{i+1}{K}) \mid A_l, B_l, C_{\mathcal{T}}\right\}. \tag{3.1}$$

We can replace $G^{\mathcal{T}}_{n,\tau_l}$ with $G^{\mathcal{T}}_{n,\frac{i+1}{K}}$ if $\frac{i}{K} \leq \tau_l \leq \frac{i+1}{K}$ and no edges come in the interval $(\tau_l, \frac{i+1}{K})$ for each $i \in [0, K-1]$. Let D be the event that a pair of distinct edges arrive within an interval of length $\frac{1}{K}$. Thus, if $G^{\mathcal{T}}_{n,\tau_l} \neq G^{\mathcal{T}}_{n,\tau_l + \frac{1}{K}}$ then at least two edges arrived in the interval $[\tau_l, \tau_l + \frac{1}{K}]$ and D holds. Thus, it follows that

$$\left| \mathbf{P}\left\{ Y(G_{n,p^*}) = 0 \mid A_l, B_l, C_{\mathcal{T}} \right\} - \sum_{i=0}^{K-1} \mathbf{P}\left\{ Y(G^{\mathcal{T}}_{n,\frac{i+1}{K}}), \tau_l \in [\tfrac{i}{K}, \tfrac{i+1}{K}) \mid A_l, B_l, C_{\mathcal{T}} \right\} \right|$$

is at most

$$\mathbf{P}\left\{ G^{\mathcal{T}}_{n,\tau_l} \neq G^{\mathcal{T}}_{n,\tau_l+\frac{1}{K}} \mid A_l, B_l, C_{\mathcal{T}} \right\} \leq \frac{\mathbf{P}\left\{D\right\}}{\mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}\right\}}.$$

We apply a union bound over all possible pairs of distinct edges e and e', and use that $\mathbf{P}\left\{|U_e - U_{e'}| \leq \frac{1}{K}\right\} \leq \frac{2}{K}$. Then $\mathbf{P}\left\{D\right\} \leq \binom{n}{2}^2 \frac{2}{K}$, and so

$$\mathbf{P}\left\{ G^{\mathcal{T}}_{n,\tau_l} \neq G^{\mathcal{T}}_{n,\tau_l+\frac{1}{K}} \mid A_l, B_l, C_{\mathcal{T}} \right\} \leq \frac{n^4}{K\mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}\right\}}.$$

Finally, we claim that each term in (3.1) is equal to

$$\mathbf{P}\left\{ Y(G^{\mathcal{T}}_{n,\frac{i+1}{K}}) \mid A_l \right\} \mathbf{P}\left\{ \tau_l \in [\tfrac{i}{K}, \tfrac{i+1}{K}) \mid B_l, C_{\mathcal{T}} \right\}. \tag{3.2}$$

This follows from the previous observation about conditional independence as $\mathcal{F}_l$ and $\mathcal{F}_{>l}$ are independent. □

Taking $K \to \infty$, the following corollary is immediate

Corollary 3.3.4. *For positive integers l, L and fixed $p \in (0,1)$, let $A_l = A_l(p)$ and $B_l = B_l(p)$. If $\mathcal{T} \in \mathcal{M}^n(l, L)$, then*

$$\mathbf{P}\left\{ Y(G_{n,p^*}) = 0 \mid A_l, B_l, C_{\mathcal{T}} \right\} = \int_0^1 \mathbf{P}\left\{ Y(G^{\mathcal{T}}_{n,t}) = 0 \,\middle|\, A_l \right\} f(t \mid B_l, C_{\mathcal{T}}),$$

where $f(\cdot \mid B_l, C_{\mathcal{T}})$ is the conditional density of τ_l given B_l and $C_{\mathcal{T}}$.

The following theorem is a generalization of Theorem 2.3.2 to $Q^{\mathcal{T}}_{n,p}$, the adjacency matrix of $G^{\mathcal{T}}_{n,p}$. It is one of the novel results of the thesis; its proof appears in Chapter 4.

Theorem 3.3.5. *Fix positive integers l, L. Then there exists $n_0 = n_0(l, L)$ such that for any $n \geq n_0$, $\mathcal{T} = (S_i)_{i \in [l]} \in \mathcal{M}^n(l, L)$ and any $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$, we have*

$$\mathbf{P}\left\{ Y(G^{\mathcal{T}}_{n,p}) = 0 \right\} \geq 1 - \frac{100}{(\ln \ln n)^{\frac{1}{4}}}.$$

We are now ready to present the proof of Proposition 3.2.2.

3.4 The graph G_{n,p^*} has full rank

Theorem 3.3.5 is the key to the proof of Proposition 3.2.2. It can be applied to the expression of Corollary 3.3.4 in which certain events are assumed to hold. For suitable choices of p and range of l, the events $A_l(p), B_l(p)$ and $C_{\mathcal{T}}$ occur with high probability up to relabelling of the vertices.

Definition 3.4.1. *Fix positive integers k, K. Let $\mathcal{D}_{k,K}(p)$ be the event that there exists a set $M \subset [n]$ with $1 \le |M| \le k$ such that M is the set of isolated vertices in $G_{n,p}$, M is a stable set in G_{n,p^*} and the sets $\{N_{G_{n,p^*}}(i)\}_{i\in M}$ are pairwise disjoint and of size at most K.*

Lemma 3.4.2. *For any real $a > 0$. Let $p_1 = p_1(a) = \frac{\ln n - a}{n}$, $p_2 = p_2(a) = \frac{\ln n + a}{n}$, $l_0 = l_0(a) = \lfloor 2e^a \rfloor$ and $L = L(a) = \lfloor 8a \rfloor$. Fix $\varepsilon > 0$, then there exists $a = a(\varepsilon)$ and $n_0 = n_0(\varepsilon)$ such that if $n \ge n_0$ we have*

$$\sum_{l=1}^{l_0} \sum_{\mathcal{T}\in\mathcal{M}^n(l,L)} \binom{n}{l} \mathbf{P}\left\{A_l(p_1), B_l(p_1), C_{\mathcal{T}}, p^* \le p_2\right\} = \mathbf{P}\left\{\mathcal{D}_{l_0,L}(p_1), p^* \le p_2\right\} \ge 1-\varepsilon.$$

Proof. Throughout this proof, we write M for the set of isolated vertices in G_{n,p_1}. First,

$$\begin{aligned}\mathbf{P}\left\{\mathcal{D}_{l_0,L}(p_1), p^* \le p_2\right\} &= \sum_{l=1}^{l_0} \mathbf{P}\left\{\mathcal{D}_{l_0,L}(p_1), p^* \le p_2, |M| = l\right\} \\ &= \sum_{l=1}^{l_0} \binom{n}{l} \mathbf{P}\left\{\mathcal{D}_{l_0,L}(p_1), p^* \le p_2, |M| = [l]\right\},\end{aligned}$$

the last inequality holding by symmetry. But we also have

$$\mathcal{D}_{l_0,L}(p_1) \cap \{M = [l]\} = \bigcup_{\mathcal{T}\in\mathcal{M}^n(l,L)} A_l(p_1) \cap B_l(p_1) \cap C_{\mathcal{T}},$$

and this union is disjoint. The equality in Lemma 3.4.2 follows. We next derive a lower bound on the probability that both $\mathcal{D}_{l_0,L}(p_1)$ and $p^* \le p_2$ hold. Let $\mathcal{D}_L(p_1, p_2)$ be the event that

i) $N_{G_{n,p_2}}(i) \subset [n] \setminus M$ for every $i \in M$, *ii)* $|N_{G_{n,p_2}}(i)| \leq L$ for every $i \in M$ and

iii) $N_{G_{n,p_2}}(i) \cap N_{G_{n,p_2}}(j) = \emptyset$, for every $i \neq j \in M$.

Then, $\mathbf{P}\left\{\mathcal{D}_{l_0,L}(p_1), p^* \leq p_2\right\}$ is at least

$$1 - \mathbf{P}\left\{\overline{\mathcal{D}_L}(p_1,p_2) \ \mid 1 \leq |M| \leq l_0\right\} - \mathbf{P}\left\{|M| = 0 \text{ or } |M| > l_0\right\} - \mathbf{P}\left\{p^* > p_2\right\}. \quad (3.3)$$

We analyse first the event $\mathcal{D}_L(p_1,p_2)$ given that $1 \leq |M| \leq l_0$. The probability that an edge e with an endpoint in M is present in G_{n,p_2} is

$$q := \mathbf{P}\left\{U_e \leq p_2 \ \mid U_e > p_1\right\} = \frac{2a}{n(1-p_1)} \leq \frac{4a}{n}$$

for n large enough. Let $q' = \frac{4a}{n}$ and assume $1 \leq |M| \leq l_0$. Under this conditioning, the probability that there is at least one edge in G_{n,p_2} with both endpoints in M is at most

$$\binom{l_0}{2} q' \leq \frac{(l_0)^2 2a}{n}.$$

On the other hand if $i \in M$, then $|N_{G_{n,p_2}}(i)|$ has a binomial distribution with parameters $n-1$ and q. Since $L = 2nq'$,

$$\mathbf{P}\left\{|N_{G_{n,p_2}}(i)| > L \ \mid i \in M\right\} \leq \mathbf{P}\left\{Bin(n,q') \geq 2nq'\right\} \leq e^{-\frac{(2nq')^2}{4}} = e^{-16a^2}.$$

Consequently, the probability that there exists $i \in M$ with $|N_{G_{n,p_2}}(i)| > L$ is at most $l_0 e^{-16a^2} = 2e^{a-16a^2}$. Finally, the probability that distinct $i,j \in M$ share a common neighbour in G_{n,p_2} is at most nq^2, so given that $1 \leq |M| \leq l_0$, the probability that the sets $\{N_{G_{n,p_2}}(i)\}_{i \in M}$ are not pairwise disjoint is at most

$$\binom{l_0}{2} nq^2 \leq \frac{(l_0)^2 8a^2}{n}.$$

Let $a = a(\varepsilon)$ satisfy $\max\{2e^{a-16a^2}, 1-e^{-e^{-a}}, \left(\frac{e}{4}\right)^{e^a}, e^{-e^a}\} < \frac{\varepsilon}{8}$. Then, for n large enough we have

$$\mathbf{P}\left\{\overline{\mathcal{D}}_L(p_1,p_2) \ \mid 1 \leq |M| \leq l_0\right\} \leq \frac{(l_0)^2(2a+8a^2)}{n} + 2e^{a-16a^2} \leq \frac{\varepsilon}{4}.$$

We proceed to bound the remaining two terms in (3.3) using Theorem 3.1.2. The number of isolated vertices in $G_{n,p}$ with $p = \frac{\ln n + x}{n}$ is asymptotically distributed as a Poisson variable with mean e^{-x}. Therefore,

$$\mathbf{P}\left\{p^* > p_2\right\} = \mathbf{P}\left\{i(G_{n,p_2}) \neq 0\right\} + o(1) = 1 - e^{-e^{-a}} + o(1) \leq \frac{\varepsilon}{4},$$

where the inequality holds for n sufficiently large. To estimate the probability that $M = \emptyset$ or $|M| > l_0$, let $\lambda = e^a$, by a Chernoff-like bound (see Lemma 5.0.9 in appendix) we have

$$\mathbf{P}\left\{i(G_{n,p_1}) \geq 2\lambda\right\} + o(1) \leq \left(\frac{e}{4}\right)^{\lambda} + o(1).$$

Thus,

$$\mathbf{P}\left\{p^* \leq p_1 \text{ or } |M| > l_0\right\} \leq e^{-e^a} + \left(\frac{e}{4}\right)^{\lambda} + o(1) \leq \frac{\varepsilon}{2},$$

where the last inequality holds for n sufficiently large. The result follows. □

Proof of Theorem 3.2.2. Fix $\varepsilon > 0$, let l_0, L, p_1, p_2 and $a = a(\varepsilon)$ be as in the proof of Lemma 3.4.2. For any positive integer $l \leq l_0$ let us write $A_l = A_l(p_1)$ and $B_l = B_l(p_1)$. Let $\mathcal{T} \in \mathcal{M}^n(l, L)$, we will show that for n large enough

$$\mathbf{P}\left\{Y(G_{n,p^*}) = 0, A_l, B_l, C_{\mathcal{T}}\right\} \geq \left(1 - \tfrac{200e^{e^a}}{(\ln\ln n)^{1/4}}\right) \mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}, p^* \leq p_2\right\}. \qquad (3.4)$$

If (3.4) holds, then we consider the space where there are $1 \leq l \leq l_0$ isolated vertices in G_{n,p_1} to get a lower bound on $\mathbf{P}\left\{Y(G_{n,p^*}) = 0\right\}$.

$$\begin{aligned}
\mathbf{P}\left\{Y(G_{n,p^*}) = 0\right\} &\geq \sum_{l=1}^{l_0} \binom{n}{l} \mathbf{P}\left\{Y(G_{n,p^*}) = 0, A_l(p_1), B_l(p_1)\right\} \\
&\geq \sum_{l=1}^{l_0} \sum_{\mathcal{T} \in \mathcal{M}^n(l,L)} \binom{n}{l} \mathbf{P}\left\{Y(G_{n,p^*}) = 0, A_l, B_l, C_{\mathcal{T}}\right\} \\
&\geq \sum_{l=1}^{l_0} \sum_{\mathcal{T} \in \mathcal{M}^n(l,L)} \binom{n}{l} \left(1 - \tfrac{100\kappa}{(\ln\ln n)^{1/4}}\right) \mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}, p^* \leq p_2\right\} \\
&\geq \left(1 - \tfrac{200e^{e^a}}{(\ln\ln n)^{1/4}}\right)(1 - \varepsilon),
\end{aligned}$$

where the second last inequality follows if (3.4) holds and the last inequality follows from Lemma 3.4.2. We now proceed to prove (3.4).

Let $N = n - l$, so that $G_{n,p_1}[n \setminus l]$ is distributed as G_{N,p_1}. Write $p_1' = \frac{\ln N - a}{N} \geq p_1$, it follows that $\mathbf{P}\left\{A_l(p_1)\right\}$ is equal to

$$\mathbf{P}\left\{i(G_{N,p_1}) = 0\right\} \geq \mathbf{P}\left\{i(G_{N,p_1'}) = 0\right\} = e^{-e^a} + o(1) \geq \frac{1}{2e^{e^a}},$$

where the last inequality follows from Theorem 3.1.2. Additionally, if $p \in (p_1, p_2)$ and n is large enough, Theorem 3.3.5 can be applied to get

$$\mathbf{P}\left\{Y(G^{\mathcal{T}}_{n,p}) > 0 \mid A_l\right\} \leq \frac{\mathbf{P}\left\{Y(G^{\mathcal{T}}_{n,p}) > 0\right\}}{\mathbf{P}\left\{A_l\right\}} \leq \frac{200e^{e^a}}{(\ln\ln n)^{1/4}}. \tag{3.5}$$

On the other hand using Corollary 3.3.4 and (3.5) we get

$$\begin{aligned}
\mathbf{P}\left\{Y(G_{n,p^*}) = 0 \mid A_l, B_l, C_{\mathcal{T}}\right\} &\geq \int_{p_1}^{p_2} \mathbf{P}\left\{Y(G^{\mathcal{T}}_{n,t}) = 0 \,\middle|\, A_l\right\} f(t \mid B_l, C_{\mathcal{T}}), \\
&\geq \left(1 - \tfrac{200e^{e^a}}{(\ln\ln n)^{1/4}}\right) \int_{p_1}^{p_2} f(t \mid B_l, C_{\mathcal{T}}) \\
&= \left(1 - \tfrac{200e^{e^a}}{(\ln\ln n)^{1/4}}\right) \mathbf{P}\left\{\tau_l \in [p_1, p_2] \mid B_l, C_{\mathcal{T}}\right\}.
\end{aligned}$$

Finally, we multiply by $\mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}\right\}$ on both sides of the inequality above to get (3.4). On the right-hand side of the inequality we use that

$$\begin{aligned}
\mathbf{P}\left\{\tau_l \in [p_1, p_2] \mid B_l, C_{\mathcal{T}}\right\} \mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}\right\} &= \mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}, \tau_l \in [p_1, p_2]\right\} \\
&= \mathbf{P}\left\{A_l, B_l, C_{\mathcal{T}}, p^* \leq p_2\right\};
\end{aligned}$$

this follows as τ_l, B_l and $C_{\mathcal{T}}$ are independent of A_l, and $\tau_l = p^* \geq p_1$ if all A_l, B_l and $C_{\mathcal{T}}$ hold. Therefore,

$$\lim_{n \to \infty} \mathbf{P}\left\{rank(G_{n,p^*}) = n\right\} = 1.$$

$\square$

CHAPTER 4
The rank of $G^{\mathcal{T}}_{n,p}$

In this chapter we consider the graph $G^{\mathcal{T}}_{n,p}$, defined in Section 3.3, with fixed $\mathcal{T} = (S_i)_{i\in[l]} \in \mathcal{M}^n(l, L)$ for positive l and L. This means that the sequence $(S_i)_{i\in[l]}$ satisfies the following properties: the sets are non-empty, pairwise disjoint and, for all $i \in [l]$ we have $S_i \subset [n] \setminus [l]$ and $|S_i| \leq L$. The graph $G^{\mathcal{T}}_{n,p}$ is obtained from $G_{n,p}$ by replacing the neighbourhood $N_{G_{n,p}}(i)$ by S_i for each $i \in [l]$. The aim of this chapter is to complete the proof of Theorem 3.3.5.

Theorem 3.3.5 states that for p in a sufficiently small interval around $\frac{\ln n}{n}$, we have that $rank(G^{\mathcal{T}}_{n,p}) = n - i(G^{\mathcal{T}}_{n,p})$ with high probability; that is, $rank(G^{\mathcal{T}}_{n,p}) = n - i(G^{\mathcal{T}}_{n,p})$. This theorem is an extension of Theorem 2.3.2. In Section 2.5 we present a proof for Theorem 2.3.2 in which the required structural property of $G_{n,p}$ is to be *robust*. In the model of $G^{\mathcal{T}}_{n,p}$, which concerns us now, the structure is random except around the set of vertices $[l]$ where edges are determined by $\mathcal{T}$. By hypothesis, the number of such edges is bounded by a constant, and it is possible to prove that the property of robustness is achieved in this new setting with high probability. Moreover, the bound we obtain is uniform in a small interval around $\frac{\ln n}{n}$.

Once the result concerning the property of robustness is settled (analogous to Theorem 2.5.8), Theorem 3.3.5 is proven with the same line of argumentation as Theorem 2.3.2 with minor changes. (As the changes are almost purely notational we decline to rewrite the proof of Theorem 3.3.5 in its entirety.) We now explain what those changes are.

The vertex exposure method relies on adding at each step, a new vertex which is connected to each of the previous vertices independently with probability p; from the matrix point of view, at each step we add a vector $\mathbf{x} = (x_1, \ldots, x_m)$ and its transpose. At this point we use Littlewood-Offord concentration inequalities on polynomial functions of $\mathbf{x}$. Throughout the proof of Theorem 2.3.2, we assume that all coordinates of $\mathbf{x}$ are i.i.d. Bernoullis with parameter p. The rows of the adjacency matrix of $G^{\mathcal{T}}_{n,p}$ are not completely random as they are determined in their first l coordinates. We overcome this problem by relabelling the vertices $\{l+1,\ldots,n\}$ so that for each $i \in [l]$, $S_i \subset [\lceil \frac{3n}{4} \rceil]$. In this way, if the vertex exposure method starts with the adjacency matrix of the first $n' := \lceil \alpha n \rceil$ vertices, with $\alpha \in (\frac{3}{4}, 1)$, then the remaining vectors to be added are of the form $\mathbf{x} = (0,\ldots,0,x_{l+1},\ldots,x_m)$, where $x_{l+1},\ldots,x_m$ are i.i.d. random variables. Consequently, the first l coordinates have no bearing in the application of the Littlewood-Offord inequalities. After this consideration, the proof follows without other major changes.

In this chapter we present a complete proof that $G^{\mathcal{T}}_{n,p}$ is robust if $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$.

4.1 Robustness in $G^{\mathcal{T}}_{n,p}$

Recall that, for a graph $G=(V,E)$ and a set $S \subset V$, we say that $v \in V$ is a *blocker for* S if $|N_G(v) \cap S| = 1$. (Earlier we stated this definition in terms of the adjacency matrix of G.) The main property of a robust sequence (Section 2.5), is that if $S \subset V$ satisfies certain conditions, then we can find at least 2 vertices that are blockers for S.

For fixed $\mathcal{T} = (S_i)_{i \in [l]} \in \mathcal{M}^n(l,L)$, we relabel the vertices $\{l+1,\ldots,n\}$ so that we can assume $S_i \subset [\lceil \frac{3n}{4} \rceil]$ for each $i \in [l]$. Let $M := [l]$ and $T := \bigcup_{i \in [l]} S_i$. Note that $T \subset [n] \setminus [l]$ and that $G^{\mathcal{T}}_{n,p}[n \setminus l]$ is distributed as $G_{n-l,p}$. The only deterministic entries in the adjacency matrix of $G^{\mathcal{T}}_{n,p}$ appear in rows and columns corresponding to the set $M \cup T$. Therefore, it is natural to expect that for $S \in [n] \setminus (M \cup T)$, the required blocking columns can be found within $[n] \setminus (M \cup T)$. We now extend the definition of a blocked set by insisting that some of the blocking vertices do not lie within a given set of vertices.

Definition 4.1.1. *Fix a graph $G=(V,E)$ and $J \subset V$. We say that $S \subset V$ is J-*blocked *if there exist distinct $v, w \in J$ that are blockers for S.*

Definition 4.1.2. *Fix a graph $G = (V, E)$, a set $J \subset V$ and $s \in [2, |V|]$. We say that G is (J, s)*-blocked *if for all $S \subset V$ of size $|S| \in [2, s]$ and containing no isolated vertices, we have that S is J-blocked.*

This essentially generalizes Defintion 2.5.2 as a graph is (V, k)-blocked if and only if it is k-blocked. We next define robustness, in graph theoretic language.

Definition 4.1.3. *Fix $\alpha \in (\frac{3}{4}, 1)$. Set $k = k(n,p) = \frac{\ln\ln n}{2p}$ and $n' = \lceil \alpha n \rceil$. We say that the sequence $\mathcal{G}^{\mathcal{T}}_{n,p} = (G^{\mathcal{T}}_{n,p}[n'], \ldots, G^{\mathcal{T}}_{n,p}[n])$ is* robust *if for all integer $m \in [n', n]$, $G^{\mathcal{T}}_{n,p}[m]$ is k-blocked, and additionally $G^{\mathcal{T}}_{n,p}[m]$ has at most $(p \ln n)^{-1}$ vertices of degree at most one.*

Theorem 4.1.4. *Fix positive integers l, L. Then there exists $n_0 = n_0(l, L)$ such that for all $n \geq n_0$ and all $\mathcal{T} \in \mathcal{M}^n(l, L)$, for any $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$, we have*

$$\mathbf{P}\left\{\mathcal{G}^{\mathcal{T}}_{n,p} \text{ is robust}\right\} \geq 1 - \frac{8}{n^{1/10}}.$$

We now prove Theorem 4.1.4 assuming the following lemma, which itself is proved in Section 4.2.

Lemma 4.1.5. *Fix positive integers l, L and let $s = s(n) = \frac{2n \ln\ln n}{3 \ln n}$. Then there exists $n_0 = n_0(l, L)$ such that for all $n \geq n_0$ and all $\mathcal{T} = (S_i)_{i \in [l]} \in \mathcal{M}^n(l, L)$, writing $T = \cup_{i \in [l]} S_i$, for any $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$ we have*

$$\mathbf{P}\left\{G^{\mathcal{T}}_{n,p}[m \setminus l] \text{ is } (\overline{T}, s)\text{-blocked for all } m \in [n', n]\right\} \geq 1 - \frac{6}{n^{1/10}}.$$

Proof of Theorem 4.1.4. Recall we are assuming that $T \subset [\lceil \frac{3n}{4} \rceil]$. We define the following events: Let A_1 be the event that for every $m \in [n', n]$, the graph $G^{\mathcal{T}}_{n,p}[m]$ contains at most $(p \ln n)^{-1}$ vertices with at most one neighbour in $G^{\mathcal{T}}_{n,p}[m]$. Let A_2 be the event that $|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n'] \setminus (M \cup T)| \geq 2$ for every vertex $v \in T \cup U$, where $U := N_{G^{\mathcal{T}}_{n,p}}(T) \setminus M$. Finally, let A_3 be the event that $G^{\mathcal{T}}_{n,p}[m \setminus l]$ is $(\overline{T}, s)$-blocked for all $m \in [\lceil \frac{3n}{4} \rceil, n]$.

We first prove that $\mathcal{G}^{\mathcal{T}}_{n,p}$ is robust if events A_1, A_2 and A_3 hold. The event A_1 is one of the conditions of being robust. So, it remains to show that for any $m \in [n', n]$

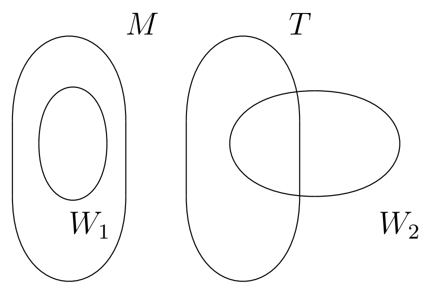

Figure 4–1: The graph $G^{\mathcal{T}}_{n,p}[m]$. The sets $M = [l]$, its neighbourhood $T = \cup_{i\in[l]} S_i$, and $S = W_1 \cup W_2$ with $W_1 \subset M$ and $W_2 \subset T \cup R_{[m]}$, where $R_{[m]} = [m] \setminus (M \cup T)$.

and any set $S \subset [m]$ of vertices in $G^{\mathcal{T}}_{n,p}[m]$ with no isolated vertices and $|S| \in [2, k]$ has two blocking vertices in $[m]$.

We now focus on $G^{\mathcal{T}}_{n,p}[m]$ for fixed $m \in [n', n]$. Let us write for any $v \in [m]$, $N_{[m]}(v) = N_{G^{\mathcal{T}}_{n,p}[m]}(v)$. Let $R_{[m]} = [m] \setminus (M \cup T)$. Consider a set $S \subset [m]$ with no isolated vertices and $|S| \in [2, k]$. Write $S = W_1 \cup W_2$ where $W_1 \subset M$ and $W_2 \subset T \cup R_{[m]}$, see Figure 4.1. We then have 4 cases.

Case 1. If $W_2 = \emptyset$, then every vertex in $N_{[m]}(S)$ is a blocker for S. This follows since $N_{G^{\mathcal{T}}_{n,p}}(S) = \cup_{i\in S} S_i$ and the sets S_i are non-empty and pairwise disjoint. Furthermore, $\lceil \frac{3n}{4} \rceil \le m$ and so $N_{G^{\mathcal{T}}_{n,p}}(S) = N_{[m]}(S)$. Hence, the number of blocking vertices of S is at least $|S| \ge 2$.

Case 2a. If $W_2 = \{v\}$ and $v \notin T \cup U$, then $N_{G^{\mathcal{T}}_{n,p}}(W_1) \cap N_{G^{\mathcal{T}}_{n,p}}(v) = \emptyset$. Consequently, every vertex in $N_{[m]}(S)$ is a blocker for S. Again, the number of blocking vertices of S is at least $|S| \ge 2$.

Case 2b. If $W_2 = \{v\}$ and $v \in T \cup U$, we claim that for $w \in N_{[m]}(v) \cap R_{[m]}$ we have $N_{[m]}(w) \cap S = \{v\}$. This follows since $N_{G^{\mathcal{T}}_{n,p}}(w) \cap M = \emptyset$ by definition of $R_{[m]}$. Hence, the number of blocking vertices of S is at least 2 is $|N_{[m]}(v) \cap R_{[m]}| \ge 2$ by the assumption of A_2 ($N_{[m]}(v) = N_{G^{\mathcal{T}}_{n,p}}(v) \cap [m]$ and $n' \le m$).

Case 3. If $|W_2| \ge 2$, then W_2 is $\overline{T}$-blocked in $G^{\mathcal{T}}_{n,p}[m \setminus l]$ by the assumption of A_3 (and $k \le s$). It follows that there exist $v_1, v_2 \in R_{[m]} = ([m] \setminus [l]) \setminus T$ which are blockers for W_2. Furthermore, vertices in $R_{[m]}$ are not adjacent to any vertex in M. It follows then that v_1 and v_2 are blockers for S and so S is $\overline{T}$-blocked as well.

We now proceed to bound the probability that events A_1, A_2 and A_3 fail to occur. For fixed $v \in [n] \setminus [l]$, $|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n']|$ is $Binomial(n', p)$, thus

$$\begin{aligned}
\mathbf{P}\left\{|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n']| \leq 1\right\} &\leq (1-p)^{n'} + n'p(1-p)^{(n'-1)} \\
&\leq e^{-p(n'-1)}(1-p+n'p) \\
&\leq e^{-\frac{n'p}{2}}(2n'p).
\end{aligned}$$

Since $n' \geq \frac{3n}{4}$ and $p \in (\frac{8\ln n}{9n}, \frac{12\ln n}{11n})$, the above inequality yields

$$\mathbf{P}\left\{|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n']| \leq 1\right\} \leq n^{-\frac{2}{3}}\left(\frac{24\ln n}{11}\right) \leq \frac{1}{n^{1/5}}. \tag{4.1}$$

For the event A_1, let X be the number of vertices $v \in [n]\setminus[l]$ with $|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n']| \leq 1$. We have no information about the degree of vertices in $[l]$, then

$$\mathbf{P}\left\{\overline{A}_1\right\} \leq \mathbf{P}\left\{X \geq (p\ln n)^{-1} - l\right\},$$

an application of Markov's inequality yields

$$\leq \frac{(n-l)\mathbf{P}\left\{|N_{G^{\mathcal{T}}_{n,p}}(v) \cap [n']| \leq 1\right\}}{(p\ln n)^{-1} - l} \leq \frac{3\ln^2 n}{n^{1/5}} < \frac{1}{n^{1/10}},$$

the second to last inequality uses (4.1) and the last inequality holds for n sufficiently large.

Let $t := |T|$. If the event A_2 does not hold, then there exist at least one vertex in $T \cup U$ which has at most one neighbour in $[n'] \setminus (M \cup T)$. Let Y be the number of vertices $v \in T \cup U$ for which $|N_{[n']}(v) \setminus (M \cup T)| \leq 1$. Then

$$\mathbf{P}\left\{\overline{A}_2\right\} \leq \sum_{k=0}^{4t\ln n} \mathbf{P}\left\{Y \geq 1, |U| = k\right\} + \mathbf{P}\left\{|U| > 4t\ln n\right\}.$$

We analyse the two terms separately. For $k < 4t\ln n$ the conditional Markov's inequality yields

$$\mathbf{P}\{Y \geq 1,\, |U| = k\} \leq (t+k)\mathbf{P}\left\{|N_{[n']}(v) \setminus (M \cup T)| \leq 1\right\}\mathbf{P}\{|U| = k\}.$$

In this case, $|N_{[n']}(v) \setminus (M \cup T)|$ is $Binomial(n'-k, p)$ with $k \leq l+t$. Thus, we similarly get

$$\mathbf{P}\left\{|N_{[n']}(v) \setminus (M \cup T)| \leq 1\right\} \leq \frac{1}{n^{1/5}}. \tag{4.2}$$

By (4.2), we get

$$\sum_{k=0}^{4t\ln n} \mathbf{P}\{Y \geq 1,\, |U| = k\} \leq \frac{t(1+4\ln n)}{n^{1/5}}.$$

The distribution of $|U|$ is binomial and

$$\mathbf{P}\{|U| \geq 4t\ln n\} \leq \mathbf{P}\{Bin(n,q) \geq 4t\ln n\},$$

where $q := 1-(1-p)^t$ is the probability that a vertex $v \in [n] \setminus [l]$ is connected to T. It follows that $q \leq tp$ and so $2nq \leq 4t\ln n$. Hence, the Chernoff bound yields

$$\mathbf{P}\{|U| \geq 4t\ln n\} \leq \mathbf{P}\{Bin(n,q) \geq 2nq\} \leq e^{-\frac{nq}{4}} \leq n^{-\frac{2t}{45}};$$

in the last inequality we use that $q \geq \frac{tp}{2}$. Therefore,

$$\mathbf{P}\left\{\overline{A}_2\right\} \leq \frac{t(1+4\ln n)}{n^{1/5}} + \frac{1}{n^{2t/45}} \leq \frac{1}{n^{1/10}},$$

the last inequality holding for n sufficiently large, as $1 \leq t \leq lL$ is bounded. Finally, by Lemma 4.1.5, $\mathbf{P}\left\{\overline{A}_3\right\} \leq 6n^{-1/10}$. Combining the three bounds we get the result. $\square$

At this point, we have proved, assuming Lemma 4.1.5, that the graph $G^{\mathcal{T}}_{n,p}$ has a robust sequence $\mathcal{G}^{\mathcal{T}}_{n,p} = (G^{\mathcal{T}}_{n,p}[n'], \ldots, G^{\mathcal{T}}_{n,p}[n])$ with high probability. The next section is therefore dedicated to proving Lemma 4.1.5.

4.2 (J, s)-blocked graphs in $G_{n,p}$

In Lemma 4.1.5 we consider $G^{\mathcal{T}}_{n,p}[m \setminus l]$ for $m \in [\lceil \alpha n \rceil, n]$ and $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$. These graphs are subgraphs of $G^{\mathcal{T}}_{n,p}[n \setminus l]$ which is distributed as $G_{n-l,p}$. To simplify the notation we therefore work with the graph $G_{n,p}$ and its subgraphs $G_{n,p}[m]$ for $m \in [\lceil \frac{3n}{4} \rceil, n]$ and $p \in (\frac{4 \ln n}{5n}, \frac{6 \ln n}{5n})$; see the comment just before 4.2.6, below. We start by defining two useful properties of $G_{n,p}$ in the range of study.

Definition 4.2.1. *Let G be a graph with vertex set $V = [n]$. We say that a vertex $v \in V$ is a* low degree *vertex if its degree in G is at most $d := \lfloor \ln \ln n \rfloor$; otherwise, we say that v is a* high degree *vertex.*

We say that G is well-separated *if every pair of low degree vertices are at distance at least 3.*

Lemma 4.2.2. *If $p \in (\frac{4 \ln n}{5n}, \frac{6 \ln n}{5n})$ and n is sufficiently large, then the probability that $G_{n,p}[m]$ is well separated for all $m \in [\lceil \frac{3n}{4} \rceil, n]$ is at least $1 - n^{-\frac{1}{10}}$.*

Proof. Let $n_1 = \lceil \frac{3n}{4} \rceil$. For each $m \in [n_1, n]$, we obtain upper bounds on the probability that $G_{n,p}[m]$ is not well separated graph and $G_{n,p}[l]$ is well separated for all $l \in [n_1, m-1]$.

First, consider the graph $G_{n,p}[n_1]$. The event that fixed vertices v_1 and v_2 are at distance at most 2 is a monotone increasing property; while the event that both vertices have low degree is monotone decreasing. By Lemma 3.1.5, these events are negatively correlated and so the probability that both events hold is bounded from above by the product of their probabilities.

The random variable $|N_{G_{n,p}[n_1]}(v_1)|$ in $G_{n,p}[n_1]$ is $Binomial(n_1 - 1, p)$ distributed. It follows that for n sufficiently large and for $d = \lfloor \ln \ln n \rfloor$,the probability that v_1 has degree at most d is

$$\sum_{i=0}^{d}\binom{n_1-1}{i}p^i(1-p)^{n_1-i-1} \le (1-p)^{n_1}\sum_{i=0}^{d}2(2n_1p)^i$$
$$\le e^{-n_1p}(2n_1p)^{d+1}$$
$$\le n^{-3/5}(3\ln n)^{d+1}.$$

Where in the first inequality we use that $1-p \ge \frac{1}{2}$, and in the third we use that $p \in (\frac{4\ln n}{5n}, \frac{6\ln n}{5n})$. The same inequality holds for the probability that v_2 has at most d neighbours excluding v_1. The event that v_1 and v_2 are low degree vertices implies that $|N_{G_{n,p}[n_1]}(v_1)| \le d$ and v_2 has at most d neighbours in $G_{n,p}[n_1]$ excluding v_1. Moreover, the latter conditions are independent. Hence, the probability that two fixed vertices v_1 and v_2 have low degree is

$$\mathbf{P}\left\{|N_{G_{n,p}[n_1]}(v_1)|, |N_{G_{n,p}[n_1]}(v_2)| \le d\right\} \le \frac{(3\ln n)^{2d+2}}{n^{6/5}} \le \frac{\ln^{3d} n}{n^{6/5}};$$

this bound also holds for the probability that $|N_{G_{n,p}[m]}(v_1)|, |N_{G_{n,p}[m]}(v_2)| \le d$ for any $m \in [n_1, n]$.

On the other hand, the probability that v_1 and v_2 are adjacent or have a common neighbour is at most $p + np^2$. By a union bound, the probability that $G_{n,p}[n_1]$ is not well separated is therefore at most

$$\binom{n_1}{2}\frac{\ln^{3d} n(p+np^2)}{n^{6/5}} \le \frac{2\ln^{3d+2} n}{n^{1/5}}.$$

We next consider $G_{n,p}[m]$ with $m > n_1$. Let $z = m$ be the unique vertex of $G_{n,p}[m]$ not in $G_{n,p}[m-1]$. If $G_{n,p}[m]$ is the first graph which is not well separated, then it is the case that either (a) z connects two low degree vertices v_1, v_2 which were at distance at least 3 in $G_{n,p}[m-1]$ or (b) z is itself a low degree vertex at distance 1 or 2 of a second vertex v_0 of low degree. By a union bound over the pairs of low degree vertices in $G_{n,p}[m-1]$ we obtain that (a) occurs with probability at most

$$\binom{m}{2}\frac{p^2 \ln^{3d} n}{n^{6/5}} \leq \frac{\ln^{3d+2} n}{n^{6/5}}.$$

Similarly, a union bound over all vertices of $G_{n,p}[m-1]$ implies that (b) occurs with probability at most

$$\frac{m \ln^{3d} n(p+np^2)}{n^{6/5}} \leq \frac{3\ln^{3d+2} n}{n^{6/5}},$$

this follows since z and any fixed vertex $v_0 \in [m-1]$ are at distance 1 or 2 with probability at most $p+np^2$.

Combining these bounds, we obtain that the probability that there exists $m_0 \in [n_1, n]$ for which $G_{n,p}[m_0]$ is the first not well separated graph in $(G_{n,p}[n_1], \ldots, G_{n,p}[n])$ is at most

$$\frac{2\ln^{3d+2} n}{n^{1/5}} + \frac{n}{4}\left(\frac{7\ln^{3d+2} n}{n^{6/5}}\right) \leq \frac{4\ln^{3d+2} n}{n^{1/5}} \leq \frac{1}{n^{1/10}},$$

where the last inequality holds for n large enough. □

We now generalize the definition of small set expander from [11].

Definition 4.2.3. *A graph $G=(V,E)$ is a* small set expander *if every set of vertices S with $|S| \leq \frac{n}{\ln^{3/2} n}$ and which contains no isolated vertices, we have*

$$|E_G(S, V \setminus S)| \geq |S|.$$

As we are interested in J-blocked sets, we generalize the concept of small set expansion. Here, $N_G(J)$ will play the role of the set of non-isolated vertices in the definition of small set expander.

Definition 4.2.4. *Let $G=(V,E)$ be a graph with vertex set $V=[n]$ and a set $J \subset V$. We say that G is a J-*expander *if for all $S \subset N_G(J)$ with $|S| \leq \frac{n}{\ln^{3/2} n}$, we have*

$$|E_G(S, J \setminus S)| \geq |S|.$$

For $J \not\subseteq V$ we will simply write J-expander instead of $(J \cap V)$-expander. The importance of well separatedness and small set expansion is that they allow us to easily bound the probability of linear dependencies among sets of small size, where $S \in V$ is considered small if $|S| \leq \frac{n}{\ln^{3/2} n}$. In the following lemma we give conditions under which the graphs $G_{n,p}[m]$ are well separated and are J-expanders with high probability.

Lemma 4.2.5. *Fix a positive integer k. Then there exists $n_0 = n_0(k)$ such that for all $n \geq n_0$ and all $J \subset [n]$ with $|J| \geq n - k$, for any $p \in (\frac{4 \ln n}{5n}, \frac{6 \ln n}{5n})$ we have that*

$$\mathbf{P}\left\{G_{n,p}[m] \text{ is well separated and } J\text{-expander for all } m \in [\lceil \tfrac{3n}{4} \rceil, n]\right\} \geq 1 - \frac{2}{n^{1/10}}.$$

Proof. Let $n_1 = \lceil \frac{3n}{4} \rceil$. Let B be the event that $G_{n,p}[m]$ is well separated and is a J-expander for all $m \in [n_1, n]$. If B doesn't hold either one of $G_{n,p}[m]$ is not well separated, or there exists $G_{n,p}[m]$ which is well separated and not a J-expander . The former condition holds with probability at most $n^{-1/10}$ by Lemma 4.2.2. We claim that for any fixed $m \in [n_1, n]$,

$$\mathbf{P}\left\{G_{n,p}[m] \text{ is well separated and is not a } J\text{-expander }\right\} \leq \frac{3}{n^{20}}. \tag{4.3}$$

Assuming (4.3), by a union bound over $m \in [n_1, n]$, we obtain

$$\mathbf{P}\left\{\overline{B}\right\} \leq \frac{1}{n^{1/10}} + \frac{3}{n^{19}} \leq \frac{2}{n^{1/10}},$$

proving the lemma, so we now turn to proving (4.3).

Suppose that $G_{n,p}[m]$ is well separated but is not a J-expander and consider a minimal set $S_0 \subset N_{G_{n,p}[m]}(J)$ of size at most $\frac{n}{\ln^{3/2} n}$ for which

$$|E(S_0, (J \cap [m]) \setminus S_0)| < |S_0|.$$

Write $S_0 = L \cup H$, where L is the set of low degree vertices in S_0 and $H = S \setminus L$. If $G_{n,p}[m]$ is well separated then L is a stable set. We claim that every vertex in L has at least one neighbour in H. Suppose to the contrary that for some $v \in L$ it is the case that $N_{G_{n,p}[m]}(v) \cap H = \emptyset$. Since L is a stable set and $S_0 \subset N_{G_{n,p}[m]}(J)$, it follows that $|N_{G_{n,p}[m]}(v) \cap J \setminus S_0| \geq 1$. Hence, the set $S_1 = S_0 \setminus \{v\}$ satisfies

$$|E(S_1, (J \cap [m]) \setminus S_1)| < |S_0| - 1 = |S_1|.$$

This contradicts the minimality of S_0.

On the other hand, if $G_{n,p}[m]$ is well separated, then no two vertices in L have a common neighbour in H and it follows that $|L| \leq |H|$.

Now, the average degree of $G_{n,p}[S_0]$ is at least

$$\frac{1}{|S_0|}\left(\sum_{v \in H} |N_{G_{n,p}}(v)| - |E(S_0, [m] \setminus S_0)|\right) \geq \frac{\lfloor \ln \ln n \rfloor}{2} - (1 + |[m] \setminus J|).$$

We have that $|[m] \setminus J| \leq k$. Hence, for n large enough, the average degree of $G_{n,p}[S_0]$ is greater than 8. We claim that the probability such a set S_0 exists in $G_{n,p}$ is at most $3n^{-20}$.

Let C be the event that $G_{n,p}$ contains a subgraph of size at most $\frac{n}{\ln^{3/2} n}$ and induced average degree greater than 8 and let p_j be the probability that there exists $S \subset [n]$ with $|S| = j$ and such that $G_{n,p}[S]$ has average degree at least 8. By a union bound over all sets of size j, we have

$$\begin{aligned} p_j \leq \binom{n}{j}\binom{\binom{j}{2}}{4j} p^{4j} &\leq \left(\frac{ne}{j}\right)^j \left(\frac{ej}{8}\right)^{4j} \left(\frac{6 \ln n}{5n}\right)^{4j} \\ &\leq \left(\frac{j^3 \ln^4 n}{n^3}\right)^j \end{aligned}$$

It follows that

$$\mathbf{P}\{C\} \leq \sum_{j=10}^{\lfloor n \ln^{-3/2} n \rfloor} \left(\frac{j^3 \ln^4 n}{n^3}\right)^j. \tag{4.4}$$

We analyse the sum in two parts. If $j \leq n^{1/4}$, then we have that for n large enough,

$$\frac{j^3 \ln^4 n}{n^3} \leq \frac{\ln^4 n}{n^{9/4}} \leq \frac{1}{n^2};$$

whereas for $j \geq n^{1/4}$, the upper bound on j yields

$$\frac{j^3 \ln^4 n}{n^3} \leq \frac{\ln^4 n}{\ln^{9/2} n} = (\ln n)^{-1/2}.$$

Furthermore, if $j \geq n^{1/4}$, then $(\ln n)^{-j/2} \leq n^{-21}$ for n large enough.

Therefore $\mathbf{P}\{C\}$ is at most

$$\sum_{j=10}^{\lfloor n^{1/4} \rfloor} n^{-2j} + \sum_{\lceil n^{1/4} \rceil}^{\lfloor n \ln^{-3/2} n \rfloor} (\ln n)^{-j/2} \leq \sum_{j=10}^{\infty} n^{-2j} + \sum_{j=\lceil n^{1/4} \rceil}^{\lfloor n \ln^{-3/2} n \rfloor} \frac{1}{n^{21}}$$
$$\leq \frac{2}{n^{20}} + \frac{1}{n^{20}}.$$

The result follows. $\square$

In Lemma 4.1.5 we consider $G^{\mathcal{T}}_{n,p}[n \setminus l]$ with $p \in (\frac{8 \ln n}{9n}, \frac{12 \ln n}{11n})$ and its subgraphs $G^{\mathcal{T}}_{n,p}[m \setminus l]$ for $m \in [\lceil \alpha n \rceil, n]$. and . These graphs are distributed as $G_{n,p}$ with $p \in (\frac{4 \ln n}{5n}, \frac{6 \ln n}{5n})$ and the subgraphs $G_{n,p}[m]$ for $m \in [\lceil \frac{3n}{4} \rceil, n]$ respectively. Hence, the following theorem implies Lemma 4.1.5.

Theorem 4.2.6. *Fix a positive integer t and let $s = s(n) = \frac{2n \ln \ln n}{3 \ln n}$. Then there exists $n_0 = n_0(t)$ such that for all $n \geq n_0$ and $T \subset [n]$ with $|T| \leq t$, for any $p \in (\frac{4 \ln n}{5n}, \frac{6 \ln n}{5n})$, we have*

$$\mathbf{P}\left\{G_{n,p}[m] \text{ is } (\overline{T}, s)\text{-blocked for all } m \in [\lceil \tfrac{3n}{4} \rceil, n]\right\} \geq 1 - \frac{6}{n^{1/10}}.$$

Proof. Let $n_1 = \lceil \frac{3n}{4} \rceil$. By relabelling the vertices we may assume $T \subset [n_1]$. Let D be the event that $G_{n,p}[m]$ is $(\overline{T}, s)$-blocked for all $m \in [n_1, n]$ and let D_1 be the event that $|N_{G_{n,p}}(v) \cap [n_1] \setminus T| \geq 2$ for every $v \in T \cup N_{G_{n,p}}(T)$. We define further two auxiliary events; let D_2 be the event that $G_{n,p}$ has maximum degree at most $\lfloor 10np \rfloor$, and let D_3 be the event that $G_{n,p}[m]$ is well separated and is a $\overline{T}$-expander for all $m \in [n_1, n]$.

We bound the probability that there exists $m \in [n_1, n]$ such that $G_{n,p}[m]$ is not $(\overline{T}, s)$-blocked by

$$\mathbf{P}\left\{\overline{D}\right\} \leq \mathbf{P}\left\{\overline{D}, D_1, D_2, D_3\right\} + \mathbf{P}\left\{\overline{D_1}\right\} + \mathbf{P}\left\{\overline{D_2}\right\} + \mathbf{P}\left\{\overline{D_3}\right\}. \tag{4.5}$$

We bound the terms in the right-hand side of the inequality in reverse order.

First, by Lemma 4.2.5, we have that $\mathbf{P}\left\{\overline{D}_3\right\} \leq 2n^{-1/10}$. Second, by Markov's inequality, the probability that $G_{n,p}$ has one vertex with degree at least $\lceil 10np \rceil$ is at most the expected number of such vertices, so

$$\begin{aligned}
\mathbf{P}\left\{\overline{D}_2\right\} &\leq n\binom{n}{\lceil 10np \rceil}p^{\lceil 10np \rceil} \\
&\leq n\left(\frac{nep}{10np}\right)^{\lceil 10np \rceil} \\
&\leq ne^{-\lceil 10np \rceil} \\
&\leq n^{-7}.
\end{aligned}$$

where in the last inequality we used that $p \geq \frac{4\ln n}{5n}$.

Now, if D_1 does not hold, then there exists at least one vertex in $T \cup N_{G_{n,p}}(T)$ which has at most one neighbour in $[n_1] \setminus T$. For fixed $v \in [n]$,

$$\begin{aligned}
\mathbf{P}\left\{|N_{G_{n,p}}(v) \cap [n_1] \setminus T| \leq 1\right\} &\leq (1-p)^{n_1-t-1} + n'p(1-p)^{n_1-t-2} \\
&\leq e^{-p(n_1-t-2)}(1-p+n_1p) \\
&\leq e^{-\frac{n_1p}{2}}(2n_1p).
\end{aligned}$$

Since $n_1 \geq \frac{3n}{4}$ and $p \in (\frac{8\ln n}{9n}, \frac{12\ln n}{11n})$, the above inequality yields

$$\mathbf{P}\left\{|N_{G_{n,p}}(v) \cap [n_1] \setminus T| \leq 1\right\} \leq n^{-\frac{3}{10}}\left(\frac{12\ln n}{5}\right) \leq \frac{1}{n^{1/5}}. \tag{4.6}$$

Let X be the number of vertices $v \in T \cup N_{G_{n,p}}(T)$ for which $|N_{G_{n,p}}(v) \cap [n_1]| \leq 1$. Then,

$$\mathbf{P}\left\{\overline{D}_1\right\} \leq \sum_{k=0}^{4t\ln n} \mathbf{P}\left\{X \geq 1,\, |N_{G_{n,p}}(T)| = k\right\} + \mathbf{P}\left\{|N_{G_{n,p}}(T)| > 4t\ln n\right\},$$

we will analyse the two terms separately. First note that for fixed $v \in [n]$ the events $v \in T \cup N_{G_{n,p}}(v)$ and $\{|N_{G_{n,p}}(v) \cap [n_1] \setminus T| \leq 1\}$ so by (4.6) we have

$$\mathbf{P}\left\{|N_{G_{n,p}}(v) \cap [n_1] \setminus T| \leq 1 \;\middle|\; v \in T \cup N_{G_{n,p}}(v)\right\} \leq \frac{1}{n^{1/5}}.$$

Now, for $k < 4t \ln n$ the conditional Markov's inequality yields

$$\begin{aligned}\mathbf{P}\left\{X \geq 1, \, |N_{G_{n,p}}(T)| = k\right\} &\leq \frac{|T \cup N_{G_{n,p}}(v)|\mathbf{P}\left\{|N_{G_{n,p}}(T)| = k\right\}}{n^{1/5}} \\ &\leq \frac{(t+k)\mathbf{P}\left\{|N_{G_{n,p}}(T)| = k\right\}}{n^{1/5}}.\end{aligned}$$

Summing over k we obtain

$$\sum_{k=0}^{4t \ln n} \mathbf{P}\left\{X \geq 1, \, |N_{G_{n,p}}(T)| = k\right\} \leq \frac{t(1+4\ln n)}{n^{1/5}}.$$

The distribution of $|N_{G_{n,p}}(T)|$ is binomial and

$$\mathbf{P}\left\{|N_{G_{n,p}}(T)| \geq 4t \ln n\right\} \leq \mathbf{P}\left\{Bin(n,q) \geq 4t \ln n\right\},$$

where $q = 1 - (1-p)^t$ is the probability that a vertex is connected to T. we have that $q \leq tp$ and so $2nq \leq 4t \ln n$. Hence,

$$\mathbf{P}\left\{|N_{G_{n,p}}(T)| \geq 4t \ln n\right\} \leq \mathbf{P}\left\{|N_{G_{n,p}}(T)| \geq 2nq\right\} \leq \mathbf{P}\left\{Bin(n,q) \geq 2nq\right\}.$$

Moreover, the Chernoff bound yields

$$\mathbf{P}\left\{Bin(n,q) \geq 2nq\right\} \leq e^{-\frac{nq}{4}} \leq n^{-\frac{t}{10}},$$

where in the last inequality we use that $q \geq \frac{tp}{2}$. Therefore,

$$\mathbf{P}\left\{\overline{D}_1\right\} \leq \frac{t(1+4\ln n)}{n^{1/5}} + \frac{1}{n^{t/10}} \leq \frac{2}{n^{1/10}},$$

the last inequality hold as $1 \leq t$ and t is a constant. This completes our bounds on $\mathbf{P}\left\{\overline{D}_1\right\}, \mathbf{P}\left\{\overline{D}_2\right\}$ and $\mathbf{P}\left\{\overline{D}_3\right\}$, and we now turn to the first term on the right-hand side of (4.5).

Given positive integers $m \in [n_1, n]$ and $j \in [2, s]$, let $D_{m,j}$ be the event that for all $S \subset [m]$ with $|S| = j$ containing no isolated vertices in $G_{n,p}[m]$, the set S is $\overline{T}$-blocked in $G_{n,p}[m]$ so that

$$D = \bigcap_{m=n_1}^{n} \bigcap_{j=2}^{s} D_{m,j}.$$

We claim that for all $m \in [n_1, n]$ and $j \in [2, s]$

$$\mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \leq n^{-21/10}, \tag{4.7}$$

from which a union bound yields

$$\mathbf{P}\left\{\overline{D}, D_1, D_2, D_3\right\} \leq \sum_{m=n_1}^{n} \sum_{j=2}^{\lfloor \frac{n \ln \ln n}{\ln n} \rfloor} \mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \leq \frac{1}{n^{1/10}},$$

completing the proof of Theorem 4.2.6. The remainder of the proof is devoted to proving the bound (4.7).

To warm up, we treat the (simplest) case that $\lceil \frac{n}{\ln^{3/2} n} \rceil \leq j \leq \lfloor \frac{n \ln \ln n}{\ln n} \rfloor$. Fix such a j, and $m \in [n_1, n]$, fix $S \subset [m]$ with $|S| = j$. Let $R_m = R_m(S) = [m] \setminus (S \cup T)$, and let

$$q_m(S) = \mathbf{P}\left\{S \text{ is not } \overline{T}\text{-blocked in } G_{n,p}[m]\right\},$$

we have that

$$q_m(S) = \mathbf{P}\left\{\text{At most one vertex } v \in R_m \text{ has } |N_{G_{n,p}[m]}(v) \cap S| = 1\right\}.$$

If $v \in R_m$, then

$$\mathbf{P}\left\{|N_{G_{n,p}[m]}(v) \cap S| = 1\right\} = jp(1-p)^{j-1}.$$

Note that the events $\{|N_{G_{n,p}[m]}(v) \cap S| = 1\}_{v \in R_m}$ are independent. For the following computations we recall that $|T| \leq t$ and that writing $r = |R_m|$ we have that for n large,

$\frac{2n}{3} \le r' \le n$, as t is constant and $|S| = j \le s = o(n)$. Thus,

$$\begin{aligned} q_m(S) &= \left(1 - jp(1-p)^{j-1}\right)^r + rjp(1-p)^{j-1}\left(1 - jp(1-p)^{j-1}\right)^{r-1} \\ &\le \left(1 - jp(1-p)^{j-1}\right)^{r-1}(1 + rjp(1-p)^{j-1}) \\ &\le \left(1 - jp(1-p)^{j}\right)^{r-1}(1 + 2rjpe^{-jp}). \end{aligned}$$

By assumption, $\frac{n}{\ln^{3/2} n} \le j \le \frac{2n \ln\ln n}{3 \ln n}$, and it follows that $2rjpe^{-jp} \ge n \ln^{-1} n \ge 1$; also, for n large enough, $(1-p)^j \ge e^{-6jp/5}$. Hence

$$\begin{aligned} q_m(S) &\le \left(1 - jpe^{-6jp/5}\right)^{r-1}(3rjpe^{-jp}) \\ &\le e^{-\frac{1}{2}njpe^{-6jp/5}}(3njpe^{-jp}). \end{aligned}$$

We use now a union bound over all sets of size j in $[m]$ to get

$$\mathbf{P}\left\{\overline{D}_{m,j}\right\} \le \binom{m}{j} e^{-\frac{1}{2}njpe^{-6jp/5}}(3njpe^{-jp}).$$

Using that $\binom{m}{j} \le \binom{n}{j} \le \left(\frac{ne}{j}\right)^j$, we obtain that

$$\ln(\mathbf{P}\left\{\overline{D}_{m,j}\right\}) \le j \ln\left(\frac{ne}{j}\right) - \frac{1}{2}njpe^{-6jp/5} + \ln\left(3njpe^{-jp}\right). \tag{4.8}$$

To prove (4.7), it suffices to show that $\ln(\mathbf{P}\left\{\overline{D}_{m,j}\right\}) \le -3\ln n$. To accomplish this, we show that the term $-\frac{1}{2}njpe^{-6jp/5}$ decreases with a higher rate than the rate at which the remaining terms increases. First,

$$\begin{aligned} j \ln\left(\frac{ne}{j}\right) + \ln\left(3njpe^{-jp}\right) &\le j\left(\ln\left(\frac{3n}{j}\right) + \frac{\ln(3njp)}{j}\right) \\ &\le j\left(\ln\left(3\ln^{3/2} n\right) + \frac{\ln(.8n \ln\ln n)\ln^{3/2} n}{n}\right) \\ &\le j\left(2\ln\ln n\right). \end{aligned}$$

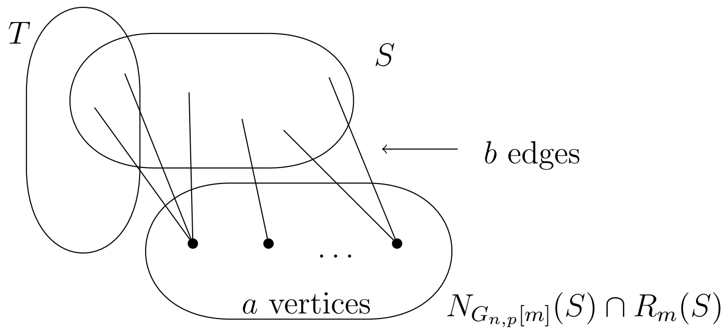

Figure 4–2: The graph $G_{n,p}[m]$. The sets T, S and $N_{G_{n,p}[m]}(S) \cap R_m$ where $R_m(S) = [m] \setminus (S \cup T)$. We let $a = |N_{G_{n,p}[m]}(S) \cap R_m|$ and $b = |E(S, R_m)|$.

Second, using the upper bound on j we get

$$\begin{aligned} j(\tfrac{1}{2}npe^{-6jp/5}) &\geq \tfrac{2}{5}j \ln ne^{-\frac{24}{25}\ln\ln n} \\ &\geq \tfrac{2}{5}j \ln^{1/25} n. \end{aligned}$$

Thus, the right-hand side of (4.8) is at most

$$j\left(2\ln\ln n - \frac{2}{5}j\ln^{1/25} n\right) \leq -\frac{j}{5}\ln^{1/25} n \leq -\frac{n\ln^{1/25} n}{5\ln^{3/2} n}.$$

Hence,

$$\mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \leq \mathbf{P}\left\{\overline{D}_{m,j}\right\} \leq \frac{1}{n^3}.$$

This proves (4.7) in the case $j \geq \lceil \frac{n}{\ln^{3/2} n} \rceil$.

For smaller j we use the auxiliary events D_1, D_2, D_3 to control the number of edges going from S to R_m. Fix $S \subset [m]$ with $|S| = j \leq \frac{n}{\ln^{3/2} n}$ and let

$$q'_m(S) = \mathbf{P}\left\{S \text{ is } \overline{T}\text{-blocked in } G_{n,p}[m], D_1, D_2, D_3\right\}.$$

And let

$$\begin{aligned} a = a_m(S) &:= |N_{G_{n,p}[m]}(S) \cap R_m|, \\ b = b_m(S) &:= |E_{G_{n,p}[m]}(S, R_m)|. \end{aligned}$$

If D_2 and D_3 hold, then

$$j \leq b \leq 10npj.$$

Moreover, if S is not $\overline{T}$-blocked , then there exists at most one vertex $v \in [m] \setminus T$ with $|N_{G_{n,p}[m]}(v) \cap S| = 1$. Since every vertex of $N_{G_{n,p}[m]}(S) \cap R_m$ has at least one neighbour in S, it follows that $b \geq 2a - 1$. Let $F_{a,b}$ be the event that $a \leq \frac{b+1}{2}$. By the preceding remark, if S is not $\overline{T}$-blocked, then $F_{a,b}$ occurs. Conditioning on the value of b, yields

$$q'_m(S) \leq \sum_{w=j}^{\lfloor 10npj \rfloor} \mathbf{P}\left\{F_{a,b}, D_1, D_2, D_3 \mid \ b = w\right\} \mathbf{P}\left\{b = w\right\}.$$

The following definitions are to shorten coming formulas. Let

$$\begin{aligned} p_m(S,w) &:= \mathbf{P}\left\{F_{a,b}, D_1, D_2, D_3 \mid b = w\right\}, \\ f_m^{(1)}(S) &:= \sum_{\substack{S \subset [m] \\ |S| = j}} \sum_{w=j}^{8j-1} \mathbf{P}\left\{b = w\right\} p_m(S,w), \\ f_m^{(2)}(S) &:= \sum_{\substack{S \subset [m] \\ |S| = j}} \sum_{w=8j}^{\lfloor 10npj \rfloor} \mathbf{P}\left\{b = w\right\} p_m(S,w). \end{aligned}$$

With the notation above we have

$$\mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \leq \sum_{\substack{S \subset [m]: \\ |S| = j}} \sum_{w=j}^{\lfloor 10npj \rfloor} q'_m(S) \leq f_m^{(1)}(S) + f_m^{(2)}(S).$$

To bound $f_m^{(1)}$ and $f_m^{(2)}$ we first bound $\mathbf{P}\left\{b \leq 8j\right\}$ and $p_m(S,w)$. We claim that

$$\mathbf{P}\left\{b \leq 8j\right\} \leq n^{-\frac{13j}{25}}. \tag{4.9}$$

Note that b is the sum of jr independent Bernoulli(p) variables. Thus, the exact value of $\mathbf{P}\{b \le 8j\}$ is

$$\begin{aligned}\sum_{i=0}^{8j}\binom{jr}{i}p^i(1-p)^{jr-i} &\le (1-p)^{jr}\sum_{i=0}^{8j}\left(\frac{ejrp}{i(1-p)}\right)^i\\ &\le e^{-jrp}\sum_{i=0}^{8j}\left(\frac{12ej\ln n}{5i}\right)^i\\ &\le n^{-\frac{8j}{15}}\sum_{i=0}^{8j}\left(\frac{12ej\ln n}{5i}\right)^i.\end{aligned}$$

The second inequality holds since $\frac{p}{1-p} \le \frac{12\ln n}{5n} \le \frac{12\ln n}{5r}$ for n sufficiently large, and the last inequality holds since $r \ge \frac{2n}{3}$ and $p \ge \frac{4\ln n}{5n}$. It remains to show that the sum in the last term is at most $n^{\frac{j}{75}}$. To do so, let $K = \frac{12e}{5}$, divide the sum in two parts and use that $j \le \frac{n}{\ln^{3/2} n}$; then

$$\sum_{i=0}^{\lfloor\frac{j}{75}\rfloor-1}\left(\frac{Kj\ln n}{i}\right)^i \le \sum_{i=0}^{\lfloor\frac{j}{75}\rfloor-1}(Kj\ln n)^i \le (Kn\ln^{-1/2} n)^{\lfloor\frac{j}{75}\rfloor}$$

and

$$\sum_{i=\lfloor\frac{j}{75}\rfloor}^{8j}\left(\frac{Kj\ln n}{i}\right)^i \le \sum_{i=\lfloor\frac{j}{75}\rfloor}^{8j}(75K\ln n)^i \le (75K\ln n)^{8j+1},$$

where the last inequality follows from the fact that $i \ge \lfloor\frac{j}{75}\rfloor$. Thus, (4.9) holds for n large enough as $\ln^x n = o(n^y)$ for any $x, y > 0$.

We next bound the probability that $|N_{G_{n,p}[m]}(S) \cap R_m|$ is small (and that D_1, D_2, D_3 hold) given $|E_{G_{n,p}[m]}(S, R_m)|$. Let

$$g(j,w) := \sqrt{\frac{6n}{j}}\left(\frac{18w}{n}\right)^{\frac{w}{2}}.$$

We claim that for $S \subset [m]$ with $|S| = j$,

$$p_m(S) = \mathbf{P}\left\{a \le \frac{w+1}{2}, D_1, D_2, D_3 \;\middle|\; b = w\right\} \le g(j,w). \tag{4.10}$$

To do so, we estimate the number of different vertices that are reached by the edges in $E_{G_{n,p}[m]}(S, R_m)$ as follows. We imagine that each edge in $E_{G_{n,p}[m]}(S, R_m)$ chooses its endpoint in R_m uniformly at random and independently of all other edges in $E_{G_{n,p}[m]}(S, R_m)$. This underestimates the size of $N_{G_{n,p}[m]}(S) \cap R_m$, since if edges have a common endpoint in S, they can not have a second common endpoint. However, this assumption only increases our estimate for the probability that $N_{G_{n,p}[m]}(S) \cap R_m$ contains few vertices.

Applying a union bound over all sets $W \subset R_m$ of size $\lfloor \frac{w+1}{2} \rfloor$, we bound the probability that each edge in $E_{G_{n,p}[m]}(S, R_m)$ has an endpoint in W. This yields the bound

$$p_m(S, w) \leq \binom{r}{\lfloor \frac{w+1}{2} \rfloor} \left(\frac{w+1}{2r} \right)^w .$$

We claim that for n large

$$\binom{r}{\lfloor \frac{w+1}{2} \rfloor} \leq \left(\frac{er}{\lfloor \frac{w+1}{2} \rfloor} \right)^{\lfloor \frac{w+1}{2} \rfloor} \leq \left(\frac{2er}{w} \right)^{\frac{w+1}{2}} .$$

To see this, note that on one hand, $\lfloor \frac{w+1}{2} \rfloor \geq \frac{w}{2}$. On the other hand, we have $w \leq 2er$ because $w \leq 10npj$, $j \leq \frac{n}{\ln^{3/2} n}$ and $r \geq \frac{2n}{3}$. This yields

$$\begin{aligned} p_m(S, w) &\leq \left(\frac{2er}{w} \right)^{\frac{w+1}{2}} \left(\frac{w+1}{2r} \right)^w \\ &\leq \left(\frac{2er}{w} \right)^{\frac{1}{2}} \left(\frac{2er(w+1)}{w(2r)} \right)^{\frac{w}{2}} \left(\frac{w+1}{2r} \right)^{\frac{w}{2}} \end{aligned}$$

now we use, in the first term, that $j \leq w$ and $r \leq n$; for the remaining terms we use that $r \geq \frac{2n}{3}$ and $w + 1 \leq 2w$ to get

$$p_m(S, w) \leq \sqrt{\frac{6n}{j}} \left(\frac{18w}{n} \right)^{\frac{w}{2}} = g(j, w).$$

Furthermore, the function $g(j,w)$ is decreasing in the range of w. To see this, we verify that the derivative is negative. The derivative is

$$\frac{dg(j,w)}{dw} = \sqrt{\frac{6n}{j}}\left(\frac{18w}{n}\right)^{\frac{w}{2}} \cdot \left\{\frac{1}{2}\ln\left(\frac{18w}{n}\right) + \frac{1}{2}\right\},$$

which is negative if $w < \frac{n}{18e}$, and the latter holds for $2 \le w \le 10jnp$ and n large enough, by the assumption that $j = |S| \le \frac{n}{\ln^{3/2} n}$. Now we are ready to obtain further bounds for $f_m^{(1)}$ and $f_m^{(2)}$. We divide the analysis in two cases.

Case 1. Suppose $150 \le j \le \lfloor \frac{n}{\ln^{3/2} n} \rfloor$. Then

$$f_m^{(1)}(S) \le \binom{m}{j} \mathbf{P}\{b \le 8j\} \max_{\substack{S \subset [m]:\\ |S|=j}} \max_{j \le w \le 8j-1} p_m(S,w).$$

Using (4.9) and (4.10) and that $g(j,w)$ is decreasing, we get

$$\begin{aligned} f_m^{(1)}(S) &\le \binom{m}{j} \mathbf{P}\{b \le 8j\}\, g(j,w) \\ &\le \sqrt{\frac{6n}{j}} \left(\frac{me}{jn^{\frac{13}{25}}}\right)^{j} \left(\frac{18j}{n}\right)^{\frac{j}{2}} \\ &\le \sqrt{\frac{6n}{j}} \left(\frac{18e^2}{jn^{\frac{1}{25}}}\right)^{\frac{j}{2}}. \end{aligned}$$

In this case, $j \ge 150$. Hence $f_m^{(1)}(S) \le K_1 n^{-\frac{5}{2}}$, where K_1 is a constant. We proceed similarly for $f_m^{(2)}(S)$, but now we neglect the probability $\mathbf{P}\{b \ge 8j\}$; so

$$\begin{aligned} f_m^{(2)}(S) &\le \binom{m}{j} \max_{\substack{S \subset [m]:\\ |S|=j}} \max_{8j \le w \le \lfloor 10npj \rfloor} g(j,w) \\ &\le \left(\frac{ne}{j}\right)^{j} g(j,8j) \\ &\le \sqrt{\frac{6n}{j}} \left(\frac{144^4 e j^3}{n^3}\right)^{j}, \end{aligned}$$

the last inequality by the definition of g. Let $K_2 = 144^4 e$ and denote by $h(j)$ the term in the right-most term in the inequality above. We note that for n sufficiently large

$h(j)$ is decreasing for $j \in [2, \frac{n}{\ln^{3/2} n}]$, since its derivative,

$$\frac{dh(j)}{dj} = h(j)\left\{\ln\left(\frac{K_2 j^3}{n^3}\right) + 3 - \frac{1}{2j}\right\},$$

is negative in that range. Therefore, $f_m^{(2)}(S) \le h(2) = K_3 n^{-\frac{11}{2}}$, where K_3 is a constant. Combining the bounds on $f_m^{(2)}(S)$ and $f_m^{(2)}(S)$ we have that

$$\mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \le K_1 n^{-\frac{5}{2}} + K_3 n^{-\frac{11}{2}} \le n^{-\frac{21}{10}}.$$

Case 2. Suppose $2 \le j \le 149$.

First, note that if S has no isolated vertices in $G_{n,p}[m]$, then any vertex of $S \setminus (T \cup N_{G_{n,p}}(T))$ has at least a neighbour in $\overline{T} \cap [m]$. If, additionally D_1 holds, then every element of $S \cap (T \cup N_{G_{n,p}}(T))$ has a neighbour in $\overline{T} \cap [m]$, and it follows that $S \subset N_{G_{n,p}}(\overline{T}) \cap [m]$. On the other hand, if $S \subset N_{G_{n,p}}(\overline{T}) \cap [m]$ with $|S| = j$ and $w \le 8j$, then

$$\mathbf{P}\left\{F_{a,b}, D_1, D_2, D_3 \mid b = w\right\} = 0.$$

To see this, note that if D_3 holds, then $G_{n,p}[m]$ is is well separated and is a $\overline{T}$-expander. So $b \le 8j$ implies

$$|E_{G_{n,p}[m]}(S, [m] \setminus S)| \le |S|(|T| + 8) \le 16(t + 8).$$

Thus, vertices in S have bounded degree. It follows that S contains only low degree vertices, which are not adjacent by the well separatedness assumption. Therefore, all vertices in $N_{G_{n,p}[m]}(S)$ have exactly one neighbour in S. On the other hand, there exist at least $|S| \ge 2$ edges going from S to R_m since $G_{n,p}[m]$ is a $\overline{T}$-expander. Therefore S is $\overline{T}$-blocked and consequently, condition $F_{a,b}$ does not hold.

Hence, $f_m^{(1)} = 0$. The bound obtained previously for $f_m^{(2)}$ is valid in this case, so for n large

$$\mathbf{P}\left\{\overline{D}_{m,j}, D_1, D_2, D_3\right\} \le f_m^{(2)} \le K_3 n^{-\frac{11}{2}} \le n^{-\frac{21}{10}}.$$

The proof is complete. $\square$

CHAPTER 5
Conclusion

In this thesis we extended the study of sparse symmetric matrices to the study of the stochastic process for the family $\{Q_{n,p}\}_{p\in(0,1)}$. This allowed us to show that the property of singularity, though not being monotone, still follows a so-called hitting time theorem.

The latter settles one of the questions that naturally arises from this novel stochastic point of view. It would also be of interest to know whether

- a similar phenomenon occurs for the event $rank(Q_{n,p}) + i(Q_{n,p}) = n$, which by Theorem 2.3.2 has a threshold at $p = \Theta(\frac{\ln n}{2n})$;
- there exist an interval $I \subset (0,1)$ of length $\Theta(n^{-1})$ in which $\{rank(Q_{n,p})\}_{p\in I}$ is monotone increasing.

In the direction of Theorem 3.3.5, we are also interested in finding a threshold for the number of vertices whose neighbourhoods can be fixed and whether the result holds if the corresponding fixed neighbourhoods have more edges.

Appendix A

To state Harris's lemma we need the following notation. We write $(\{0,1\}^n, \mathbf{P})$, for the product space of n iid Bernoulli (p) and fixed $p \in (0,1)$. We endow $\{0,1\}^n$ with the following order: for any two elements $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ with $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ we say that $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for all $i \in [n]$.

We say that an event $E \subset \{0,1\}^n$ is an *increasing event* if for all $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{x} \in E$ then $\mathbf{y} \in E$. We say that $E \subset \{0,1\}^n$ is a *decreasing event* if $\{0,1\}^n \setminus E$ is an increasing event. Let

$$E_0 := \{(x_1, \ldots, x_{n-1}) \in \{0,1\}^{n-1}; \, (x_1, \ldots, x_{n-1}, 0) \in E\},$$
$$E_1 := \{(x_1, \ldots, x_{n-1}) \in \{0,1\}^{n-1}; \, (x_1, \ldots, x_{n-1}, 1) \in E\}.$$

Note that, by the independence of each coordinate, we have that

$$\mathbf{P}\{E\} = (1-p)\mathbf{P}\{E_0\} + p\mathbf{P}\{E_1\}.$$

Lemma 5.0.7 (Harris's Lemma [4])**.** *For fixed $p \in (0,1)$ and $(\{0,1\}^n, \mathbf{P})$, let $A, B \subset \{0,1\}^n$. If both A and B are increasing events, then*

$$\mathbf{P}\{A \cap B\} \geq \mathbf{P}\{A\}\mathbf{P}\{B\}. \tag{5.1}$$

If A is an increasing event and B is a decreasing event, then

$$\mathbf{P}\{A \cap B\} \leq \mathbf{P}\{A\}\mathbf{P}\{B\}. \tag{5.2}$$

Proof. We prove (5.1) by induction on n. For $n = 1$ the inequality is trivial. Now, suppose $n \geq 2$ and consider two events $A, B \in \{0,1\}^n$. If A and B are increasing then $A_0 \subset A_1$, $B_0 \subset B_1$; so $(\mathbf{P}\{A_1\} - pA_0)(\mathbf{P}\{B_1\} - \mathbf{P}\{B_0\}) \geq 0$ holds. This in turn, yields

$$\mathbf{P}\{A_0\}\mathbf{P}\{B_1\} + \mathbf{P}\{A_1\}\mathbf{P}\{B_0\} \leq \mathbf{P}\{A_0\}\mathbf{P}\{B_0\} + \mathbf{P}\{A_1\}\mathbf{P}\{B_1\}. \tag{5.3}$$

Now, using that $(A \cap B)_i = A_i \cap B_i$ for $i = 0, 1$; and the induction hypothesis we get

$$\begin{aligned}
\mathbf{P}\{A \cap B\} &= (1-p)\mathbf{P}\{A_0 \cap B_0\} + p\mathbf{P}\{A_1 \cap B_1\} \\
&\geq (1-p)\mathbf{P}\{A_0\}\mathbf{P}\{B_0\} + p\mathbf{P}\{A_1\}\mathbf{P}\{B_1\} \\
&\geq [(1-p)\mathbf{P}\{A_0\} + p\mathbf{P}\{A_1\}]\,[(1-p)\mathbf{P}\{B_0\} + p\mathbf{P}\{B_1\}] \\
&= \mathbf{P}\{A\}\mathbf{P}\{B\},
\end{aligned}$$

where the last inequality is obtained using (5.3). To prove (5.2) we apply (5.1) to the increasing events A and $\overline{B}$. This yields,

$$\begin{aligned}
\mathbf{P}\{A \cap B\} &= \mathbf{P}\{A\} - \mathbf{P}\left\{A \cap \overline{B}\right\} \\
&\leq \mathbf{P}\{A\} - \mathbf{P}\{A\}\mathbf{P}\left\{\overline{B}\right\} \\
&= \mathbf{P}\{A\}(1 - \mathbf{P}\left\{\overline{B}\right\}) \\
&= \mathbf{P}\{A\}\mathbf{P}\{B\}.
\end{aligned}$$

$\square$

Appendix B

Lemma 5.0.8. *For all n, $\binom{n}{\lfloor n/2 \rfloor} 2^{-n} \leq n^{-1/2}$.*

Proof. Let $m = \lfloor n/2 \rfloor$. We will expand the terms and group them conveniently to get a product.

If $n = 2m + 1$,

$$\begin{aligned}
\binom{n}{\lfloor n/2 \rfloor} 2^{-n} &= \frac{1(2 \cdot 1)3(2 \cdot 2)5 \cdots (2 \cdot m)(2m+1)}{m+1!m!2^n} \\
&= \frac{1 \cdot 3 \cdots (2m+1)}{m+1!2^{m+1}} \\
&= \prod_{i=1}^{m+1} \frac{2i-1}{2i}
\end{aligned}$$

If $n = 2m$,

$$\begin{aligned}
\binom{n}{\lfloor n/2 \rfloor} 2^{-n} &= \frac{1(2 \cdot 1)3(2 \cdot 2)5 \cdots (2m-1)(2 \cdot m)}{m!m!2^n} \\
&= \frac{1 \cdot 3 \cdots (2m-1)}{m!2^m} \\
&= \prod_{i=1}^{m} \frac{2i-1}{2i}
\end{aligned}$$

In both cases the product goes from $i = 1$ to $\lceil n/2 \rceil$. Let's deduce now the following equality

$$\prod_{i=1}^{k} \frac{2i-1}{2i} = \frac{1}{\sqrt{2k+1}} \sqrt{\prod_{i=1}^{k} \left(1 - \frac{1}{4i^2}\right)}. \tag{5.4}$$

Denote by K the product in the left-hand side and multiply by $(2k+1)K$, which is a similar product but the factors are $(2i+1)/2i$. We obtain that

$$(2k+1)K^2 = \prod_{i=1}^{k} \frac{(2i-1)(2i+1)}{4i^2} = \prod_{i=1}^{k} \frac{4i^2-1}{4i^2}.$$

It is clear that the second term in the right-hand side of (5.4) is less than 1 and so

$$K \leq \frac{1}{\sqrt{2k+1}}.$$

Letting $k = \lceil n/2 \rceil$ in (5.4) we get

$$\binom{n}{\lfloor n/2 \rfloor} 2^{-n} \leq \frac{1}{\sqrt{2k+1}} \leq \frac{1}{\sqrt{n}}.$$

$\square$

Lemma 5.0.9. *For any $\lambda \geq 0$. If X is a Poisson(λ) random variable, then for any positive integer k*

$$\mathbf{P}\left\{X \geq k\lambda\right\} \leq \left(\frac{e^{k-1}}{k^k}\right)^{\lambda}.$$

Proof. To obtain this, we use a Chernoff type of argument. Let $t, x \geq 0$, then

$$\mathbf{P}\left\{X \geq x\right\} = \mathbf{P}\left\{e^{tX} \geq e^{tx}\right\} \leq \frac{\mathbf{E}\left[e^{tX}\right]}{e^{tx}}. \tag{5.5}$$

Now, we calculate the expected value above,

$$\mathbf{E}\left[e^{tX}\right] = \sum_{m=1}^{\infty} \frac{e^{tm}\lambda^m e^{-\lambda}}{m!} = \frac{e^{\lambda e^t}}{e^{\lambda}} \sum_{m=1}^{\infty} \frac{\left(\lambda e^t\right)^m e^{-\lambda e^t}}{m!} = e^{\left(e^t-1\right)\lambda}.$$

Let $t = \ln k$ and $x = k\lambda$, then by (5.5)

$$\mathbf{P}\left\{X \geq x\right\} \leq \frac{\left(e^{k-1}\right)^{\lambda}}{\left(k^k\right)^{\lambda}}.$$

$\square$

References

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Number 118 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[2] Sheldon Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. New York, NY: Springer. xii, 251 p., 2nd ed edition, 1997.

[3] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[4] B. Bollobás and O. Riordan. *Percolation*. Cambridge University Press, 2006.

[5] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2nd edition edition, 2001.

[6] Béla Bollobás and Alan M. Frieze. On matchings and Hamiltonian cycles in random graphs. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 23–46. North-Holland, Amsterdam, 1985.

[7] Béla Bollobás and Andrew Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[8] Jean Bourgain, Van H. Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[9] Sourav Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34(6):2061–2076, 2006.

[10] Kevin P. Costello, Terence Tao, and Van Vu. Random symmetric matrices are almost surely nonsingular. *Duke Math. J.*, 135(2):395–413, 2006.

[11] Kevin P. Costello and Van Vu. On the rank of random sparse matrices. *Combin. Probab. Comput.*, 19(3):321–342, 2010.

[12] L. Erdësh. Universality of Wigner random matrices: a survey of recent results. *Uspekhi Mat. Nauk*, 66(3(399)):67–198, 2011.

[13] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

[14] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN*, (3):436–479, 2010.

[15] P. Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[16] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[17] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961.

[18] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[19] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[20] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981. 10.1007/BF02579329.

[21] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[22] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[23] Svante Janson, Tomasz Lucsak, and Andrzej Runcinski. *Random Graphs.* Dicrete Mathematics and Optimization. Wiley-Interscience, 2000.

[24] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random 1-matrix is singular. *Journal of the American Mathematical Society*, 8(1):pp. 223–240, 1995.

[25] Harry Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.

[26] J. Komlós. On the determinant of $(0,1)$ matrices. *Studia Sci. Math. Hungar*, 2:7–21, 1967.

[27] János Komlós and Endre Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Math.*, 43(1):55–63, 1983.

[28] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, 15(1):211–225, 1922.

[29] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.

[30] Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40(0):203 – 216, 1981.

[31] M. L. Mehta. *Random matrices and the statistical theory of energy levels.* Academic Press, New York, 1967.

[32] Hoi H. Nguyen. Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Math. J.*, 161(4):545–586, 2012.

[33] L. A. Pastur. The spectrum of random matrices. *Teoret. Mat. Fiz.*, 10(1):102–112, 1972.

[34] Mark Rudelson and Roman Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[35] A. Sárközy and E. Szemerédi. Über ein Problem von Erdős und Moser. *Acta Arithmetica*, 11:205–208, 1965.

[36] Terence Tao. *Topics in random matrix theory*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

[37] Terence Tao and Van Vu. On random ± 1 matrices: singularity and determinant. *Random Structures Algorithms*, 28(1):1–23, 2006.

[38] Terence Tao and Van Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.

[39] Terence Tao and Van Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[40] Terence Tao and Van Vu. A sharp inverse Littlewood-Offord theorem. *Random Structures Algorithms*, 37(4):525–539, 2010.

[41] Terence Tao and Van Vu. Random matrices: The universality phenomenon for wigner ensembles. 2012.

[42] Terence Tao and Van H. Vu. *Additive Combinatorics*. Number 105 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[43] Terence Tao and Van H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.

[44] Linh Tran, Van Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. 2010.

[45] Van Vu. Random discrete matrices. In *Horizons of combinatorics*, volume 17 of *Bolyai Soc. Math. Stud.*, pages 257–280. Springer, Berlin, 2008.

[46] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics*, 62(3):pp. 548–564, 1955.

[47] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *The Annals of Mathematics*, 67(2):pp. 325–327, 1958.