RANKS OF RANDOM MATRICES AND GRAPHS

BY KEVIN COSTELLO

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Mathematics

Written under the direction of
Van Vu
and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey
October, 2007

ABSTRACT OF THE DISSERTATION

Ranks of Random Matrices and Graphs

by Kevin Costello

Dissertation Director: Van Vu

Let Q_n be a random symmetric matrix whose entries on and above the main diagonal are independent random variables (e.g. the adjacency matrix of an Erdős-Rényi random graph). In this thesis we study the behavior of the rank of the matrix in terms of the probability distribution of the entries. One main result is that if the entries of the matrix are not too concentrated on any particular value, then the matrix will with high probability (probability tending to 1 as the size of the matrix tends to infinity) have full rank.

Other results concern the case where Q_n is sparse (each entry is 0 with high probability), and in particular on the adjacency matrix of sparse random graphs. In this case, we show that if the matrix is not too sparse than with high probability any dependency among the rows of Q_n will come from a dependency involving very few rows.

Acknowledgements

I have been fortunate to have been surrounded by supportive people both before and during my graduate career. As an undergraduate, I got my introduction to research during summer programs under the tutelage of Professors Paul Garrett and the University of Minnesota and Richard Wilson at Caltech. Their mathematical advice from both was illuminating, and their advice on the research process was even more so.

My first introduction to random graphs and the probabilistic method came from Professor Fan Chung Graham at UCSD Her insight and mentorship were invaluable to helping me mature as a mathematician.

Both at UCSD and Rutgers I had the benefit of being around other graduate students who inspired me both through fascinating mathematical discussions and just through their example. To all of you, thanks.

I also want to thank my committee members for both their time and their patience putting up with my sometimes rather poor planning and organizational skills. I would like to thank Professor Saks in particular for his efforts in helping ease my transition in and out of Rutgers.

Finally and foremost, I would like to thank Professor Van Vu for his guidance, patience, and insights over the past three years. Working with him has been an immensely rewarding experience, and so much of what I have gained my time in graduate school has come from his efforts.

Dedication

To my parents. Thanks for everything.

Table of Contents

Chapter 1

Preliminaries

1.1 Random Matrices

The matrices we will consider in this thesis will be those characterized by a probability distribution ξ_{ij} that is assumed to independent for each entry on or above the main diagonal, with the entries below the main diagonal either being taken as also being independent or as satisfying $\xi_{ij} = \xi_{ji}$ (to create a symmetric matrix). Although there are other models of random matrices (e.g. distributions only taking values on the class of unitary matrices), we will be focusing exclusively on the entrywise independent model here.

Typically, the goal is to study the limiting spectral behavior of the matrix (e.g. the number of real eigenvalues, the size of the largest eigenvalue, or the number of 0 eigenvalues) as the size of the matrix tends to infinity. A standard result along this line is the following, a rescaling of a result due to Wigner [34]

Theorem 1.1.1 *(Wigner's Semicircle law): Let ξ_{ij} be a doubly infinite symmetric array of real-valued random variables, satisfying the conditions:*

- *The ξ_{ij} with $i \leq j$ are all independent*
- *The distribution of each ξ_{ij} is symmetric about 0*
- *Each ξ_{ij} has variance equal to 1*
- *For each $k \geq 2$ there is a uniform bound C_k on the k^{th} moment of each ξ_{ij}*

Let A_n denote the $n \times n$ matrix whose (i, j) entry is $\xi_{i,j}$, and for $\alpha < \beta$ let $E_n(\alpha, \beta)$ denote the number of eigenvalues of A_n lying between $\alpha\sqrt{n}$ and $\beta\sqrt{n}$. Then for any α

and β, with probability 1 we have

$$\lim_{n\to\infty} \frac{E_n(\alpha,\beta)}{n} = \int_\alpha^\beta \sqrt{4-x^2}dx$$

In other words, once certain assumptions are made on the distribution of each entry, the limiting global distribution of the spectrum of the matrix becomes clear.

Another result in this vein is the following, due to Komlós [22]

Theorem 1.1.2 *Let ξ_{ij} be a doubly infinite array of independent, identically distributed, non-degenerate random variables, and let A_n be the $n \times n$ matrix whose (i,j) entry is $\xi_{i,j}$. Then*

$$\lim_{n\to\infty} \mathbf{P}(\text{rank } (A_n) = n) = 1.$$

This is in some sense a *local* eigenvalue result as opposed to the global result of Wigner. While Wigner's law gives that, for example, the number of 0 eigenvalues will be $o(n)$ for sufficiently large n, Komlós's result gives that with high probability there will not be a single eigenvalue which is exactly 0. In this thesis we will be focusing on such local results, and in particularly will be focusing on the behavior of the rank of the matrix (i.e. the number of 0 eigenvalues) for large n.

Besides the global vs. local dichotomy, there are two other divisions that characterize the study of random matrices:

Symmetric vs. Asymmetric Matrices: In a random symmetric matrix, the entries on or above the main diagonal are viewed as independent random variables, but those below the main diagonal are chosen to satisfy $a_{ij} = a_{ji}$; in a random asymmetric matrix, all n^2 entries are viewed as being independent. The symmetric matrix has the advantage of having real eigenvalues, which makes the use of the moment method much simpler in determining the global spectrum. Hence, for example, the proof of Wigner's theorem is much simpler than obtaining the corresponding spectral density in the nonsymmetric case (Girko's so called "Circular Law", see [17, 2]) However, in general it seems harder to analyze many of the local properties in symmetric models rather than the corresponding asymmetric models.

The main advantage to working in an asymmetric model is the independence of the rows of A. Many of the important properties of a matrix (e.g. the determinant or the

smallest singular value) can be estimated in terms of the distance from each of its rows to the subspace spanned by some other collection of its rows. In an asymmetric matrix, the row and the subspace will be independent of one another. In the symmetric case, however, the last row will be almost completely determined by the remaining $n-1$ rows, which makes attaining such distance bounds much more difficult. We will be focusing mostly on the symmetric case here.

Continuous (Gaussian) vs. Discrete Entries: A further dichotomy arises when considering the distribution of the entries. If the entries are independent Gaussians with nonzero variance then some questions (the expected rank, for instance) become trivial, while others become much easier. This is in part due to the presence of a known joint singular value density for Gaussian matrices (see, for example, [26]), and in part because Gaussian vectors are rotationally symmetric. This greatly simplifies the row-by row analysis of properties such as the determinant, as the distance from a random vector to a fixed subspace is independent of the subspace in question.

In the case of more general entries (especially discrete entries), we lose both of these advantages. In particular, the distance from a random vector to a fixed subspace can depend greatly on the structure on the subspace. Some of these difficulties can be overcome using the so-called Littlewood-Offord Lemma [12], which by bounding the probability that a random linear form is small in absolute value gives an upper bound on the probability that this distance is small for certain subspaces. One of the key lemmas which will be needed to deal with symmetric matrices is a generalization of this lemma to quadratic forms, which will be discussed in Chapter 2.

1.2 Random Graphs

There are several different models of random graphs, though we will mainly be focussing on a single model here. The original $G(n, M)$ model due to Erdős and Rényi [13] focused on graphs chosen uniformly from the collection of all graphs with n vertices and M edges. In practice, however, this model can prove difficult to work with, as the individual edges in the graph are no longer independent. Instead, it is more common

to work with the $G(n,p)$ model introduced by Gilbert [16], where each edge is included uniformly in the graph with probability p.

In general, any graph referred to as an Erdős-Rényi graph in this thesis will be drawn from the $G(n,p)$ model, as the independence of the edges is essential to our result. The lack of independence is also an obstacle here in analyzing random $d-$regular graphs (graphs chosen uniformly from the collection of all $d-$regular graphs), so the coverage of these graphs in this thesis will be limited to a few conjectures near the end.

To each graph in G we associate its adjacency matrix $A(G)$, the matrix which is 1 in the (i,j) position if and only if i is connected to j. We denote the adjacency matrix of a graph in $G(n,p)$ as $Q(n,p)$, which can be thought of as a random symmetric matrix with 0's on the main diagonal whose entries above the main diagonal are each 1 with probability p, 0 with probability $1-p$. The above-diagonal entries in this matrix are all independent (this would not hold for the $G(n,m)$ or $d-$regular model). We will also abuse notation slightly be referring to the rank of $A(G)$ as the "rank of G", and to the rank of $Q(n,p)$ as the "rank of $G(n,p)$.

Random graphs are often analyzed in the limit as n tends to infinity, with $p = p(n)$ being a function allowed to depend on n. We say that a property P holds for "almost all" graphs $G(n,p)$ if the limit as n tends to infinity of the probability that $G(n,p)$ satisfies P tends to 1.

We say that a function $f(n)$ is a **threshold function** for the property P if

$$\lim_{n\to\infty} \mathbf{P}(G(n,p) \text{ satisfies } P) = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \frac{p}{f(n)} = 0; \\ 1 & \text{if } \lim_{n\to\infty} \frac{p}{f(n)} = \infty. \end{cases},$$

One result of chapter 4 of this thesis will be to determine the exact threshold function for the property of a full rank adjacency matrix. Interestingly, it is not obvious that any threshold should exist in the first place. Although thresholds are known to exist for all monotone graph properties (properties which are closed under the addition of edges) [8], the rank of a graph is in general not a monotone property. For example, adding an edge to complete the path of length 4 to the 4-cycle decreases the rank from 4 to 2. This lack of monotonicity also implies that the results obtained here for the rank of $G(n,p)$ will not extend automatically to results for $G(n,M)$.

A major theme in analyzing the rank near the threshold will be the following intuition: failure to satisfy the property of full rank is caused mainly by local obstructions (dependencies involving only a few rows) rather than by global ones (dependencies involving many rows). This same theme shows up in the behavior of random graphs near the thresholds for other properties as well. Consider, for example, the question of whether $G(n,p)$ is connected. On one hand, it is clear that any graph with an isolated vertex cannot be connected, and direct calculation shows that

$$\mathbf{P}(G(n,\frac{\ln n+c}{n}) \text{ has an isolated vertex}) = (1+o(1))(1-e^{-e^{-c}}). \tag{1.1}$$

In [16], Gilbert showed that this natural lower bound on the disconnectivity probability was an upper bound as well, so that

$$\mathbf{P}(G(n,\frac{\ln n+c}{n}) \text{ is not connected}) = (1+o(1))(1-e^{-e^{-c}}).$$

Furthermore, if the graph were not connected in this probability range, then it would almost surely consist of a single connected component along with some number of isolated vertices.

These results can be thought of as follows: There is a clear "local" obstruction to connectedness in the behavior at any one vertex. However, this obstruction is the only obstacle to connectedness in two distinct senses:

(1) For any fixed c, almost every graph with edge probability $\frac{\ln n+c}{n}$ either contains an isolated vertex or is connected.

(2) If a graph with probability in this range does contain isolated vertices, it can almost surely be made into a connected graph by removing those vertices.

Analogues of (1) also hold for many other combinatorial properties of the random graph $G(n,p)$. For example, the threshold for a graph in $G(n,p)$ having a perfect matching corresponds to the local obstruction of the graph having an isolated vertex [13], and the threshold for having a Hamiltonian cycle corresponds to the local obstruction of a vertex having degree less than 2 [23]. The results in Chapter 4 of this thesis can be thought of as an analogue of these results for the question of singularity of random matrices: Again the main causes of singularity will be local (dependencies involving few rows) rather than global (dependencies involving many rows).

1.3 Prior Work on Discrete Models

The ranks of many families of matrices (e.g. those with independent entries drawn from a continuous distribution with nonzero variance) are trivial to analyze, as the singular matrices form a set of measure 0 for any finite n. The question becomes more complicated when the entries are allowed to be drawn from a discrete distribution. The first nontrivial case analyzed was that of a non-symmetric matrix where the entries are equally likely to be 0 and 1. In 1967, Komlós [21] proved that the limit as n tended to infinity of the singularity probability of such matrices was 0.

A second paper one year later by Komlós [22] allowed the entries to be drawn from a more general distribution ξ, which was assumed to be identical for each entry and independent of the size of the matrix. Again, the result was that the singularity probability tended to 0 as n increased, assuming ξ was nondegenerate. Komlós later refined his argument to give a singularity probability of $O(n^{-1/2})$ [6] in the specific case of 0/1 matrices (matrices whose entries are equally likely to be 0 or 1).

The next key breakthrough was due to Kahn, Komlós , and Szemerédi, who showed in 1995 that the singularity probability in the ± 1 case (where each entry is equally likely to be 1 or -1) is exponentially small [19], giving an upper bound of $O(0.999^n)$ on this probability (this result also implies that the singularity probability in the 0/1 case is exponentially small).

This upper bound was later improved by Tao and Vu [30, 31], with the current best known upper bound being $(\frac{3}{4} + o(1))^n$. Interestingly, the strongest results seem to require tools from additive combinatorics, including a variant of Freiman's theorem that helps provide a classification of all subspaces which contain an unusually large number of vertices of the standard hypercube.

However, the exact singularity probability is still at this point unknown. The best known lower bound is $O(\frac{n^2}{2^n})$, which comes from the probability that the matrix has two equal rows. This bound is also conjectured to be sharp.

The case of symmetric matrices, and in particular that of the adjacency matrix of $G(n, p)$, has been much less studied. Bauer and Golinelli examined the rank of a

random tree in [5], where they used a leaf removal argument to obtain both exact and asymptotic expressions for the expected rank. The critical fact used in their argument was that if v is a vertex in a graph G which has exactly one neighbor w then removal of v and w decreases the rank of G by exactly 2. For a tree, one can continue removing vertices in this fashion until all edges have been removed from G, at which point the rank is clearly 0.

In [3] Bauer and Golinelli performed a detailed analysis of the leaf removal process for $G(n,p)$ in the case $p = \frac{c}{n}$. They showed that if $c < e$, then successive removal of leaves will almost surely result in a graph consisting of isolated vertices and a "core" graph on $o(n)$ vertices without any leaves. In particular, they obtained an asymptotic (up to a multiplicative factor of $1+o(1)$) expression for the rank of G using this method. For $c > e$ the core contains a positive fraction of the vertices of G, so the effect of the core is no longer negligible on the rank of G, and Bauer and Golinelli could only obtain lower and upper bounds on the rank. Although they conjectured [4] that the core had full rank, they were unable to analyze it rigorously. Furthermore, as $np \to \infty$, the core contains almost all of the vertices of G, so Bauer and Golinelli's bounds do not give much information about the rank.

1.4 Our Results

Our results can be thought of as expanding the prior matrix results in two distinct directions. In Chapter 3 of this thesis, we will extend the results of Komlós to cover the case of random *symmetric* matrices where no entry concentrates too highly on any particular value. In particular, we will show the following generalization of a result obtained by the author with Tao and Vu in [9]:

Theorem 1.4.1 *Let $0 < \alpha < 1/2$ be any fixed constant. Let Q_n be an $n \times n$ matrix whose upper diagonal entries $x_{ij}, 1 \leq i < j \leq n$, are independent random variables satisfying*

$$\sup_{c \in C} \mathbf{P}(x_{ij} = c) \leq 1 - n^{-\alpha}, \tag{1.2}$$

and whose lower diagonal entries are fixed according to the relationship $x_{ij} = x_{ji}$.

Then for any $\epsilon > 0$ the probability that Q_n is singular is $O(n^{\frac{1}{4}(2\alpha-1+\epsilon)})$, where the implied constant in the O notation is dependent only on ϵ and α.

In particular, this covers the case of the adjacency matrix of $G(n,p)$ for $n^{-1/2} < p < 1 - n^{-1/2}$.

Recall that $Q(n,p)$ denotes the adjacency matrix of $G(n,p)$. In chapter 4, we will examine the rank of $Q(n,p)$, where p is allowed to go as far down as $\frac{c \ln n}{n}$ for any c. One main result here will be the following exact expression for the rank of G in terms of the structure of G, which holds for almost all graphs:

Theorem 1.4.2 *If $p = \Omega\left(\frac{\ln n}{n}\right)$, then the rank of G will almost surely be exactly*

$$\min_{S \subseteq V(G)} n - |S| + |N(S)|,$$

where $N(S)$ denotes the number of vertices in G having at least one neighbor in S.

Our second result will be a complete characterization of the dependent sets of rows of the adjacency matrix:

Theorem 1.4.3 *Let s be any positive integer. If $p > \frac{\ln n}{sn}$, then almost every graph in $G(n,p)$ satisfies the following property:*

A set of rows S of the adjacency matrix of G is dependent if and only if S contains a subset S' with $|S'| \leq s-1$ such that fewer than $|S'|$ vertices of G have a neighbor in S'.

In other words, almost surely every dependency that comes about in the adjacency matrix of $G(n,p)$ comes about due to the presence of a set of at most $s-1$ vertices which doesn't contain enough neighbors. The case $s = 2$ of this theorem was obtained by the author and Vu in [10]. It can be thought of as saying that the isolated vertices provide an obstacle to non-singularity in the same way that they provide one to connectedness in both of the senses mentioned before. This means that

(1) For any fixed c, almost every graph with edge probability $\frac{\ln n + c}{n}$ either contains an isolated vertex or has a nonsingular adjacency matrix.

(2) If a graph with probability in this range does contain isolated vertices, it can almost surely be "fixed" into a graph with nonsingular adjacency matrix by removing those vertices.

1.5 Our General Method

The proofs of the exponential results in [19], [30], and [31], as well as the revised proof of Komlós [6], are all in some sense based on a row by row exposure of the matrix in question. Unfortunately, this approach does not carry over well to symmetric matrices, as exposing $n-1$ rows of a symmetric $n \times n$ matrix makes the remaining row almost entirely deterministic. Instead, we will take the approach of Komlós's original 1967 paper [21] in which we build up our matrix minor by minor, exposing a new row and new column *s*imultaneously. In other words, we view our $n \times n$ matrix as being embedded into a larger sequence $\{Q_1, Q_2, \ldots, Q_m, \ldots\}$, where each element of the sequence satisfies the block matrix relationship

$$Q_m = \begin{bmatrix} Q_{m-1} & v_m \\ v_m^T & x_{mm} \end{bmatrix}.$$

Here v_m is a vector of length $(m-1)$ whose entries are independent and random and x_{mm} is independent from v_m and Q_{m-1} (though in some cases x_{mm} may be fixed rather than random). If we think of Q_n as the adjacency matrix of a graph G on n vertices, then Q_m corresponds to the adjacency matrix of the induced subgraph on m vertices, and the building up of our matrix corresponds exactly to the vertex exposure process of G.

The advantage of working with this augmentation process is that it is very well behaved with respect to the ranks of the matrices in the sequence. Those ranks satisfy

$$\text{rank } (Q_{m-1}) \leq \text{rank } (Q_m) \leq \min\{\text{rank } (Q_{m-1}) + 2, m\}, \tag{1.3}$$

as the new row and new column individually can increase the rank of Q_m by at most 1 each.

When trying to show that Q_n almost surely has full rank, we will follow the rough heuristic that the second inequality in (1.3) will usually be an equality. This means that

if Q_{m-1} is conditioned to be any fixed full rank matrix, then almost surely Q_m will also have full rank. If, on the other hand, Q_{m-1} is conditioned to be any fixed matrix that falls short of full rank, then the augmentation will almost surely have increased the rank of Q_m by 2. This can also be thought of in terms of the function $X_m = m - \text{rank } (Q_m)$. By the inequalities in (1.3), X_m is a random walk on the non-negative integers that moves by at most one unit in each step. Our heuristic is that this random walk has a very strong bias towards 0. If the walk is already at 0, then it tends to stay there. Otherwise, it will almost surely move one unit to the left. Therefore, it will probably finish at 0 if n is large enough.

In other cases, we will be trying to show that the rank of Q_m is almost surely equal to some other logical upper bound $f(Q_m)$ on the rank (for example, $f(Q_m)$ could be the number of nonzero rows of Q_m). In this case, the X_m in our heuristic can be replaced with $X'_m = f(Q_m) - \text{rank } (Q_m)$, and for the right choice of f we might hope that the heuristic still holds.

Unfortunately, the heuristics above do not always hold. There are possible choices of Q_{m-1} which do not augment well at all, and for which the second inequality in (1.3) almost surely does not hold with equality. For example, if our matrix is singular due to having a row which is entirely 0, then adding a single new column may well not fix this problem (especially when we move to the case of $G(n,p)$ for p close to 0). What we will do to account for this is to divide the dependencies of Q_m into two classes based on the size of the support of a vector of the nullspace. We will show in each case that there is an m_0 and a k such that

(1) If Q_m has a dependency involving at least k rows, then augmentation will almost surely increase the rank of Q_m by 2

(2)Almost surely there will be no $m > m_0$ such that Q_m has a dependency among fewer than k rows which is not covered by Theorem 1.4.3.

A similar pair of results will enable us to show that it is rare for a full-rank matrix to augment to one which is not of full rank, so our heuristic will usually hold once we get sufficiently far in the augmentation process.

1.6 Some Useful Inequalities

There are a few standard inequalities that we will make use of throughout.

First, for any real x we have that

$$e^x \leq 1 + x. \tag{1.4}$$

We will also make extensive use of the following upper bound on the binomial coefficient

$$\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b. \tag{1.5}$$

We will also need the following three large-deviation bounds on probability distributions (all of which can be found in [1] and [32])

Theorem 1.6.1 *(Markov s Inequality) Let X be a non-negative random variable with finite expectation. Then for any c,*

$$\mathbf{P}(X > c) \leq \frac{c}{\mathbf{E}(X)}$$

Theorem 1.6.2 *(Chebyshev's inequality) Let X be a non-negative variable with finite expectation and with variance σ^2. Then*

$$\mathbf{P}(|X - \mathbf{E}(X)| > c\sigma) \leq \frac{1}{c^2}$$

Theorem 1.6.3 *(Chernoff's bound) Let $X_1, X_2, \ldots X_n$ be independent random variables satisfying $|X_i| \leq 1$ for all i, and let $X = \sum_{i=1}^n X_i$, and let σ^2 be the variance of X. Then for any positive λ,*

$$\mathbf{P}(|X - \mathbf{E}(X)| > \lambda\sigma) \leq 2\max(e^{-\lambda^2/4}, e^{-\lambda\sigma/2})$$

Chapter 2

Littlewood-Offord Type Problems

2.1 Linear Forms

In order to effectively analyze the augmentation problem described in the previous chapter, we need to be able to answer the question of whether an augmentation is or is not likely to increase the rank of the matrix being augmented. This can in turn be thought of as the question of whether a random vector chosen from some probability distribution is orthogonal to a given (fixed) vector (one lying in the orthogonal complement of the space spanned by the existing rows or columns), thus reducing the question to estimating the probability that some form

$$f(x_1, x_2, \ldots x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n \tag{2.1}$$

is equal to 0.

This problem was originally studied in the case where each x_i is equally likely to be 1 or -1. For this case, Erdős [12], improving a result of Littlewood and Offord [25], obtained the following result, which is the best possible without imposing some sort of further restriction on the a_i.

Theorem 2.1.1 *Let $a_1, a_2, \ldots a_n$ be real numbers, all of which are at least 1 in absolute value. Then the number of sums of the form*

$$\pm a_1 \pm a_2 \pm \cdots \pm a_n \tag{2.2}$$

which lie in any open interval of length 2 is at most $\binom{n}{\lfloor \frac{n}{2} \rfloor}$.

The same bound was later shown to hold for complex a_i by Kleitman [20]. Using Stirling's approximation to bound the central binomial coefficient and rescaling leads to the following probabilistic restatement of Kleitman's result:

Theorem 2.1.2 *Let $a_1, a_2, \dots a_n$ be complex numbers, at least q of which are nonzero. Let x_i be independent random variables which are equal to 1 or -1, each occurring with probability $\frac{1}{2}$. Then for any interval I of length at most 2,*

$$\mathbf{P}(\sum_{i=1}^{n} a_i x_i \in I) = O(\frac{1}{\sqrt{q}}),$$

where the implied constant is absolute.

Remark 2.1.3 *Although Erdős' result was originally stated for ± 1 random variables, his argument works equally well for variables which are equally likely to be 1 or 0. Additionally, the result can be extended to intervals of length longer than 2, although the constant in the O notation is of course dependent on the length of the interval.*

Since the determinant of a matrix with independent entries is linear in each entry, one might hope that Theorem 2.1.2 could be combined with a result stating that almost surely many of the coefficients of the form given by the determinant are nonzero to obtain that the determinant of an augmented matrix is almost surely non-zero. However, Theorem 2.1.2 is not quite adequate for our purposes for two reasons. Firstly, the matrices we will consider will often be symmetric and thus not quite have independent entries. Indeed, the determinant of a symmetric matrix is a *quadratic* form in its entries rather than a linear one, and in the next section we will revise Theorem 2.1.2 to handle quadratic forms. The second difficulty is that in many cases (e.g. random graphs with edge probability not equal to 0.5) the entries will not have the distribution corresponding to that in Theorem 2.1.2.

This second problem can be handled by replacing the Littlewood-Offord Theorem with the following corollary of a result due to Halász [18]:

Theorem 2.1.4 *Let $0 < \rho < 1$ and let x_i be independent random variables such that there are at least q different i for which*

$$\sup_{c \in \mathbf{C}} \mathbf{P}(|a_i x_i - c| < \frac{1}{2}) \leq 1 - \rho.$$

Then for any interval I of length at most 2,

$$\mathbf{P}(\sum_{i=1}^{n} a_i x_i \in I) = O(\frac{1}{\sqrt{q\rho}}),$$

where the implied constant is absolute.

If the variables instead satisfy

$$\sup_{c \in \mathbf{C}} \mathbf{P}(a_i x_i = c) \leq 1 - \rho$$

for q distinct i, then

$$\sup_{c \in \mathbf{C}} \mathbf{P}(\sum_{i=1}^{n} a_i x_i = c) = O(\frac{1}{\sqrt{q\rho}}).$$

The proof of Theorem 2.1.4 is complicated and relies on Fourier analysis. When working with random graphs, however, we only need the following special case of Theorem 2.1.4 where the x_i are Bernoulli variables.

Theorem 2.1.5 *Suppose that $a_1, \ldots a_n$ are constants, at least q of which are at least 1 in absolute value, and that $x_1, \ldots x_n$ are independent Bernoulli random variables each equal to 1 with probability p. Then for any interval I of length at most 2,*

$$\mathbf{P}(\sum_{i=1}^{n} a_i x_i \in I) = O(\frac{1}{\sqrt{q\rho}}), \tag{2.3}$$

where $\rho = \min\{p, 1-p\}$ and the implied constant is absolute.

Theorem 2.1.5 has a short proof that is not reliant on the Fourier Analytic methods of [18].

Proof (of Theorem 2.1.5): We may assume without loss of generality that the first q variables are at least 1 in absolute value. We will first examine the case where $p \leq 1/2$.

By conditioning on $x_{q+1}, \ldots x_n$ and replacing I by $I - \sum_{i=q+1}^{n} a_i x_i$, it suffices to show the result is true in the case where $x_{q+1} = \cdots = x_n = 0$.

Let $r_1, r_2, \ldots r_q$ be independent, identically distributed Bernoulli variables which are equal to 1 with probability $2p$. Let $s_1, s_2, \ldots s_q$ be independent, identically distributed Bernoulli variables each equal to 1 with probability $1/2$. As $r_i s_i$ has the same distribution as x_i, it suffices to show

$$\mathbf{P}(\sum_{i=1}^{q} a_i r_i s_i \in I) = O(\frac{1}{\sqrt{qp}}).$$

But by Bayes' inequality we have

$$\mathbf{P}(\sum_{i=1}^{q} (a_i r_i) s_i \in I) \leq \mathbf{P}(\sum_{i=1}^{q} (a_i r_i) s_i \in I | \sum_{i=1}^{q} r_i \geq qp) + \mathbf{P}(\sum_{i=1}^{q} r_i < qp)$$

The sum of the r_i has expectation equal to $2qp$ and variance equal to $2qp(1-2p) \leq 2qp$. By Chebyshev's inequality, therefore,

$$\mathbf{P}(\sum_{i=1}^{q} r_i < qp) \leq \mathbf{P}(|\sum_{i=1}^{q} r_i - 2qp| \geq qp) = O(\frac{1}{\sqrt{qp}}).$$

By using $a_i r_i$ as the coefficients and applying Theorem 2.1.2 (more precisely, the 0/1 version of the theorem described in the remark following that Theorem), we see that the first term is $O(\frac{1}{\sqrt{qp}})$ as well, so we are done.

Now let us assume $p > 1/2$. Letting $p = \sum_{i=1}^{n} a_i$, we have

$$\mathbf{P}(\sum_{i=1}^{n} a_i x_i \in [b_1, b_2]) = \mathbf{P}(\sum_{i=1}^{n} a_i(1-x_i) \in [a-b_2, a-b_1]),$$

so the result follows immediately from the $p < 1/2$ case applied to the variables $1 - x_i$.

■

Remark 2.1.6 *A similar argument shows that the same result holds when the x_i are random variables which are equal to 1 with probability p, -1 with probability p, and 0 with probability $1-2p$. The only difference is that the ± 1 version of the original Littlewood-Offord result is used instead of the 0/1 version.*

This result is (up to a constant factor) the best possible, as can be seen from the case where $q/2$ of the a_i are equal to 1 and $q/2$ are equal to -1. However, it is possible to obtain stronger results by restricting the structure of the a_i. For example, if all of the nonzero a_i are distinct, then the probability drops to $O((qp)^{-3/2})$. This result can be obtained by following the argument of the proof of Theorem 2.1.5, replacing Chebyshev's inequality by the Chernoff bound and the Littlewood Offord Lemma with a stronger result due to Sárközy and Szemerédi [28]. Alternatively, it can also be obtained as a consequence of Theorem 2 from [18]. However, in our arguments we generally will not have control over the values of the a_i, so we cannot use this stronger result.

2.2 The Quadratic Case

Although the solution to the Littlewood-Offord problem for linear forms described in the previous section is sufficient to answer the singularity question for asymmetric 0/1

matrices in [21], it is not enough to answer the corresponding question for adjacency matrices of random graphs, or indeed for symmetric matrices in general. The difficulty arises in the augmentation of symmetric matrices, where the determinant is a quadratic form rather than a linear form in the entries of the newly added row and column. To overcome this difficulty, what we need is a version of Theorem 2.1.5: If a quadratic form has a large number of nonzero coefficients, then the form should probably not be equal to any fixed value, nor should it lie in any small interval. Such a result was first proved in [9]. The following strengthened version of that result was originally proved in [10].

Theorem 2.2.1 *Let $x_1, x_2, \ldots x_n$ be independent, Bernoulli variables, each of which is 1 with probability $p < 1/2$. Let a_{ij} be a symmetric $n \times n$ array of constants. Assume that there is a partition $\{1, 2, \ldots n\} = S_1 \bigsqcup S_2$ such that for each $j \in S_2$ there are at least q different i in S_1 for which $|a_{ij}| > 1$. Let*

$$Q(x_1, x_2, \ldots x_n) = \sum_{1 \leq i,j, \leq n} a_{ij} x_i x_j$$

be the quadratic form whose coefficients are the a_{ij}. Then for any interval I of length 1,

$$\mathbf{P}(Q(x_1, x_2, \ldots x_n) \in I) = O((p \min(|S_1|, q))^{-1/4}),$$

where the implied constant is absolute.

Proof (of Theorem 2.2.1): Ideally, what we would like to do is to view Q as a pair of nested linear forms. We can write $Q = \sum x_i \eta_i$, where $\eta_i = \sum a_{ij} x_j$, and would like to apply Theorem 2.1.5 twice: Once to show that most of the η_i will not be 0 and then again to show that Q is therefore nonzero. Unfortunately, we can not do this directly, as the η_i and x_i are not independent. We therefore make use of the following decoupling inequality:

Lemma 2.2.2 *Let X and Y be independent random variables, and let $F(X, Y)$ be an event dependent on X and Y. Let X' be a variable independent of X and Y, but having the same distribution as X. Then*

$$\mathbf{P}(F(X, Y))^2 \leq P(F(X, Y) \wedge F(X', Y))$$

Remark 2.2.3 *This is actually a special case of a much more general conjecture of Sidorenko [29], which states that if H is any bipartite graph, then*

$$\mathbf{P}(F(X,Y))^{|E(H)|} \leq \mathbf{P}(\bigwedge_{(X_i,Y_j)\in E(H)} F(X_i,Y_j)).$$

Sidorenko showed the lemma to be true for complete bipartite graphs, and Lemma 2.2.2 is the case $H = K_{2,1}$.

Proof (of Lemma 2.2.2) We will first prove the case where X and Y have finite support. Letting $x_1, x_2, \ldots x_m$ be the support of X and $y_1, y_2, \ldots y_r$ be the support of Y, we have by Bayes' theorem and the independence of X and X' that

$$\begin{aligned}
\mathbf{P}(F(X,Y) \wedge F(X',Y)) &= \sum_{j=1}^{r} [\mathbf{P}(F(X,Y)|Y=y_i)]^2 \mathbf{P}(Y=y_i) \\
&\geq (\sum_{j=1}^{r} \mathbf{P}(F(X,Y)|Y=y_i)\mathbf{P}(Y=y_i))^2 \\
&= [\mathbf{P}(F(X,Y))]^2,
\end{aligned}$$

where the middle line follows from the Cauchy-Schwartz inequality.

The case where the support of X or Y is infinite now follows from discretization or by replacing the summations from the finite case with integrals. ■

In our case, we take X to be those x_i with $i \in S_1$ and Y to be those x_i with $i \in S_2$. $F(X,Y)$ now becomes the event

$$Q(x_1, x_2, \ldots x_n) \in I,$$

while $F(X',Y)$ becomes the event,

$$Q(x'_1, x'_2, x'_{|S_1|}, x_{|S_1|+1}, \ldots x_n) \in I,$$

where the x'_i are independent copies of the x_i. Lemma 2.2.2 now states that

$$\begin{aligned}
\mathbf{P}(F(X,Y) \in I)^2 &\leq \mathbf{P}(F(X,Y) \in I \wedge F(X',Y) \in I) &(2.4)\\
&\leq \mathbf{P}(F(X,Y) - F(X',Y) \in [-1,1]). &(2.5)
\end{aligned}$$

It therefore suffices to show the probability of this last event is $O((p\min(q,|S_1|))^{-1/2})$. Note that we have

$$\begin{aligned} F(X,Y) - F(X',Y) &= g(X,X') + 2\sum_{j=|S_1|+1}^{n} x_j \sum_{i=1}^{|S_1|} a_{ij}(x_i - x_i') \\ &:= g(X,X') + 2\sum_{j=|S|+1}^{n} x_j \eta_j, \end{aligned}$$

where $g(X,X')$ is some function of the variables in X and X' and η_i is defined to be equal to the inner sum in the above equation.

Let ν be the number of η_j which are at most one in absolute value. By Bayes' theorem we have

$$\begin{aligned} \mathbf{P}(F(X,Y) - F(X',Y) \in [-2,2]) \leq \mathbf{P}(F(X,Y) - F(X',Y) \in [-2,2] \,|\, \nu \leq \frac{|S_1|}{2}) \\ + \mathbf{P}(\nu > \frac{|S_1|}{2}). \end{aligned} \tag{2.6}$$

We bound these terms in reverse order. For the second term, we use Chebyshev's inequality. We can think of ν as the sum of $|S_1|$ indicator variables, each corresponding to the event that a given η_j is at most one in absolute value. Since each η_j is a linear form with (by hypothesis) at least q nonzero coefficients, Theorem 2.1.5 (see Remark 2.1.6) guarantees that each η_i is 0 with probability at most $O((qp)^{-1/2})$. In particular, the expectation of ν is at most $O(|S_1|(qp)^{-1/2})$. By Markov's inequality, it follows that

$$\mathbf{P}(\nu > \frac{|S_1|}{2}) = O((qp)^{-1/2}). \tag{2.7}$$

We bound the first term by viewing our process as exposing X and X' before Y. For any x and x' taken on by X and X' and satisfying $\nu \leq \frac{|S_1|}{2}$, we have by 2.1.5 that

$$\mathbf{P}(F(X,Y) - F(X',Y) \in [-2,2] \,|\, X = x \wedge X' = x') = O((p|S_1|)^{-1/2}),$$

as our assumption that $\nu \leq \frac{|S_1|}{2}$ guarantees at least $\frac{|S_1|}{2}$ nonzero η_i.

Summing over all x and x' satisfying this property and using Bayes' theorem, it follows that

$$\mathbf{P}(F(X,Y) - F(X',Y) \in [-2,2] \,|\, \nu \leq \frac{|S_1|}{2}) = O((p|S_1|)^{-1/2}). \tag{2.8}$$

Combining inequalities (2.6), (2.8), and (2.7) completes the proof. ■

In the remainder of this section, we will give several simple corollaries which place Theorem 2.2.1 in forms more suitable for application. The first two corollaries aim to remove the need for an explicit partition of the variables.

Corollary 2.2.4 *Let $x_1, x_2, \dots x_n$ be independent Bernoulli variables, each of which is 1 with probability $p < 1/2$. Let a_{ij} be a symmetric $n \times n$ array of constants. Assume that there are at least d different i, for each of which there are at least d distinct j such that $|a_{ij}| > 1$. Let*

$$Q(x_1, x_2, \dots x_n) = \sum_{1 \le i,j, \le n} a_{ij} x_i x_j$$

be the quadratic form whose coefficients are the a_{ij}. Then for any interval I of length 1,

$$\mathbf{P}(Q(x_1, x_2, \dots x_n) \in I) = O((pd)^{-1/4}),$$

where the implied constant is absolute.

Proof Let S_2 be a arbitrary collection of $\frac{d}{2}$ distinct i, each having at least d distinct j for which a_{ij} is at least 1 in absolute value. The corresponding partition satisfies the hypotheses of Theorem 2.2.1 with $q = \frac{d}{2}$ and $|S_1| \ge \frac{d}{2}$. ■

Corollary 2.2.5 *Let the x_i be as in Corollary 2.2.4, and let a_{ij} be a symmetric $n \times n$ array of constants, at least mn of which are greater than 1 in absolute value. Then*

$$\mathbf{P}(Q(x_1, x_2, \dots x_n) \in I) = 0) = O((pm)^{-1/4}),$$

where the implied constant is absolute.

Proof Let S consist of those i for which there are more than $\frac{m}{2}$ distinct j for which c_{ij} are at least 1 in absolute value. We count the number total number of coefficients c_{ij} which are at least one in absolute value in two different ways. On the one hand, there are at least mn of them by assumption. On the other hand, we know that there are at most

$$n|S| + \frac{m}{2}(n - |S|) \le n(|S| + \frac{m}{2})$$

such coefficients. It follows that $|S| \geq \frac{m}{2}$, so we can take $d = \frac{m}{2}$ in Corollary 2.2.4. ■

Actually, these results are somewhat stronger than what we need for our random matrix problems. Typically the quadratic forms we consider will be the determinants of these matrices, and we will only attempt to prove that these determinants do not concentrate on a single point, rather than to show they are not likely to all lie in entire interval. In this case, we can use a simple scaling argument to relax the hypothesis that many of the coefficients are at least 1 in absolute value, instead only requiring them to be nonzero.

Corollary 2.2.6 *Let $x_1, x_2, \ldots x_n$ be independent Bernoulli variables, each of which is 1 with probability $p < 1/2$. Let a_{ij} be a symmetric $n \times n$ array of constants. Assume that there are at least d different i, for each of which there are at least d distinct j for which a_{ij} is nonzero. Let*

$$Q(x_1, x_2, \ldots x_n) = \sum_{1 \leq i,j, \leq n} a_{ij} x_i x_j$$

be the quadratic form whose coefficients are the a_{ij}. Then

$$\sup_{c \in \mathbf{C}} \mathbf{P}(Q(x_1, x_2, \ldots x_n) = c) = O((pd)^{-1/4}),$$

where the implied constant is absolute.

Proof Let r be the minimum absolute value among all the nonzero a_{ij}. We have

$$\begin{aligned} \mathbf{P}(Q(x_1, x_2, \ldots x_n) = c) &= \mathbf{P}(\frac{1}{2r} Q(x_1, x_2, \ldots x_n) = \frac{c}{2r}) \\ &\leq \mathbf{P}(\frac{1}{r} Q(x_1, x_2, \ldots x_n) \in [\frac{c}{2r} - \frac{1}{2}, \frac{c}{2r} + \frac{1}{2}]) \\ &= O((pd)^{-1/4}), \end{aligned}$$

where the final step comes from an application of Corollary 2.2.4 to the quadratic form $\frac{1}{2r} Q(x_1, \ldots x_n)$ (which has all nonzero coefficients at least 2 in absolute value by construction). ■

By replacing Theorem 2.1.5 with the more general Theorem 2.1.4 of Halász, we can obtain versions of many of the results in this section which apply to a much wider class of random variables. A typical example is the following Theorem:

Theorem 2.2.7 *Let $0 < \rho < 1$, and let x_i be independent random variables each satisfying*

$$\sup_{c \in \mathbf{C}} \mathbf{P}(x_i = c) \leq 1 - \rho.$$

Let a_{ij} be a symmetric array of coefficients satisfying the hypotheses of either Corollary 2.2.6 or Corollary 2.2.5, and let Q be the corresponding quadratic form. Then

$$\sup_{c \in \mathbf{C}} \mathbf{P}(Q(x_1, x_2, \ldots x_n) = c) = O((q\rho)^{-1/4}),$$

where the implied constant is absolute.

It is likely that these results are not the best possible, as there is no reason to expect equality to hold in both (2.4) and (2.5). These inequalities (and the decoupling process as a whole) had the effect of replacing the $-1/2$ exponent of the original Littlewood-Offord bound with a $-1/4$, and we conjecture that in fact all of the exponents in this section could have been replaced by $-1/2$. This result would be essentially the best possible, as in the $p = 1/2$ case we have for any α that

$$\mathbf{P}(2(x_1 + x_2 + \cdots + x_n)(x_1 + x_2 + \cdots + x_{\alpha n}) = 0) = \Theta((\alpha n)^{-1/2}),$$

, since this is the probability that the second factor is 0. On the other hand, this form has αn^2 coefficients which are at least 2 in absolute value.

2.3 Higher Degrees

It is natural to try to extend the results of the previous section to forms of degree larger than two. Again, the intuition is that if a higher dimensional form has many nonzero coefficients, then the form will with high probability not lie in any fixed interval. Here again the intuition turns out to be correct. In this section we will present the following generalized version of Corollary 2.2.5.

Theorem 2.3.1 *Let $0 < \rho < 1$, and let $x_1, x_2 \ldots x_n$ be random variables satisfying*

$$\sup_{c \in \mathbf{C}} \mathbf{P}(|x_i - c| \leq \frac{1}{2}) \leq 1 - \rho$$

for each i. Let

$$f(x_1, x_2, \ldots x_n) = \sum_{1 \leq i_1 \leq i_2 \cdots \leq i_k \leq n} a_{i_1 i_2 \ldots i_k} x_{i_1} x_{i_2} \ldots x_{i_k}$$

be a polynomial of degree $k \geq 2$ in n variables. Suppose furthermore that at least there are at least mn^{k-1} terms of degree k whose coefficients have absolute value at least 1, where $m \geq 4^k$.

Then for any interval I of length 1,

$$\sup_{c \in \mathbf{C}} \mathbf{P}(f(x_1, x_2, \ldots x_n) \in I) = O((m\rho)^{-2^{-(k^2+k-2)/2}}),$$

where the constant in the O notation depends only on k.

Remark 2.3.2 *As in the quadratic case, this result can be extended by scaling to a result giving the same bound for the probability that $f = 0$ in the case where f has at least mn^{k-1} non-zero coefficients.*

Remark 2.3.3 *It seems unlikely that the exponent in the bound here is the best possible. More likely would be a bound of $O((m\rho)^{-1/2})$, corresponding to the polynomial*

$$f(x_1, \ldots x_n) = 2(x_1 + \cdots + x_m)(x_1 + \ldots x_n)^{k-1}.$$

Proof We induct on k. The base case $k = 2$ is Corollary 2.2.5.

Now assume the result is true for polynomials of degree $k-1$ and let f be a polynomial of degree k. Let T denote the collection of all terms of S which both contain k distinct variables and have coefficient at least one in absolute value. As there are at most kn^{k-1} terms of f which are degree k and do not contain at least k distinct variables,

$$|T| \geq (m-k)n^{k-1}.$$

If the variables were partitioned into k sets uniformly at random, the expected number of members of T having exactly one variable in each element of the partition would be $\frac{|T|k!}{k^k}$. Therefore there must exist a specific partition

$$\{x_1, x_2, \ldots x_n\} = U_1 \bigsqcup U_2 \cdots \bigsqcup U_k$$

for which at least $\frac{|T|k!}{k^k}$ elements of T have this property. Call such terms "balanced".

As in the quadratic case, our goal will be in some sense to "decouple" f, so that we can view it as the composition of a linear form and a form of degree $k-1$. In this case, however, our decoupling will be slightly more complex.

Lemma 2.3.4 *Let $X_1, \ldots, X_k$ be random variables and let $E = E(X_1, \ldots, X_k)$ be an event depending on the X_i. For each i let X_i' be an independent copy of X_i. Then*

$$\mathbf{P}(E(X_1, \ldots, X_k)) \leq \mathbf{P}(\bigwedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots, X_k^S))^{1/2^k}$$

where for each S we have $X_i^S = X_i$ if $i \in S$ and $X_i^S = X_i'$, if $i \notin S$.

Proof We proceed by induction on k. The $k = 1$ result follows immediately from the independence of X and X'. Now assume that the result is true for k variables. Discretizing as in Lemma 2.2.2, we have

$$\mathbf{P}(E(X_1, \ldots, X_{k+1})) = \sum_{i=1}^{r} \mathbf{P}(E(X_1, \ldots, X_k, a_j))\mathbf{P}(X_{k+1} = a_j)$$

Applying the induction hypothesis to each $E(X_1, \ldots, X_k, a_j)$, we get

$$\mathbf{P}(E(X_1, \ldots, X_{k+1})) \leq \sum_{i=1}^{n} \mathbf{P}(\bigwedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots, X_k^S, a_j))^{1/2^k} \mathbf{P}(X_{k+1} = a_j).$$

Jensen's inequality applied to the random variable $\mathbf{P}(\wedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots, X_k^S, a_j))$ and the concave functional $f(z) = z^{1/2^k}$ gives that this is at most

$$\left(\sum_j \mathbf{P}(\bigwedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots, X_k^S, a_j))\mathbf{P}(X_{k+1} = a_j) \right)^{1/2^k}$$

$$= \left(\mathbf{P}(\bigwedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots X_k^S, X_{k+1})) \right)^{1/2^k}.$$

The result now follows from application of Lemma 2.2.2 with

$$\begin{aligned} X &= X_{k+1}, \\ Y &= (X_1, X_2, \ldots, X_k, X_1', X_2', \ldots X_k'), \\ E &= (\wedge_{S \subset \{1,\ldots,k\}} E(X_1^S, \ldots, X_k^S, X_{k+1})). \end{aligned}$$

■

In our case the X_i will correspond to the variables in each U_i. Let X_i^S be as in Lemma 2.3.4. By that Lemma we have

$$\mathbf{P}(f(X_1, \dots X_n) \in I) \leq \mathbf{P}(\bigwedge_S f(X_1^S, \dots X_n^S) \in I)^{1/2^k}.$$

Define the new random variable

$$R := \sum_{S \in \{1,2,\dots k\}} (-1)^{|S|} f(X_1^S, X_2^S, \dots X_k^S).$$

If each $f(X_1^S \dots X_n^S)$ is contained in I, it must be the case that

$$R \in 2^{k-1}I - 2^{k-1}I \in [-2^{k-1}, 2^{k-1}].$$

It therefore suffices to show that

$$\mathbf{P}(R \in [-2^{k-1}, 2^{k-1}] = O((m\rho)^{-2^{-(k^2-k-2)/2}}).$$

The advantage to working with R is that the terms of f which are not balanced cancel. Any unbalanced term of f has some U_i which contains no variable from the term, and the summands in R for which $i \in S$ and $i \notin S$ cancel. In fact, direct calculation shows that R can be factored as

$$\begin{aligned} R &= \sum_{i_1 \in U_1} \sum_{i_2 \in U_2} \cdots \sum_{i_k \in U_k} a_{i_1 \dots i_k} (x_{i_1} - x'_{i_1})(x_{i_2} - x'_{i_2}) \dots (x_{i_k} - x'_{i_k}) \\ &= \sum_{i \in U_1} w_i R_i, \end{aligned}$$

where $w_i = x_i - x'_i$ and

$$R_i = \sum_{i_2 \in U_2} \cdots \sum_{i_k \in U_k} a_{i, i_2, \dots, i_n} \prod_{j=2}^{k} w_j.$$

Note that the R_i and w_i are now independent.

Let $Y \subseteq U_1$ be those variables in U_1 which are contained in at least $\frac{|S|k!}{2k^k}$ members of S. As any particular element of U_1 is contained in at most n^{k-1} balanced terms, we have

$$\frac{|T|k!}{k^k} \leq |Y|n^{k-1} + (|U_1| - |Y|)\frac{|T|k!}{2k^k} \leq |Y|n^{k-1} + n\frac{|T|k!}{2k^k},$$

so

$$|Y| \geq \frac{k!|T|}{2k^k n^{k-1}} \geq \frac{k!(m-k)}{2k^k}.$$

Let Y_2 be those $i \in Y$ which satisfy $|R_i| \geq 1$. By Bayes' inequality, we have

$$\mathbf{P}(|R| \leq 2^{k-1}) \leq \mathbf{P}(|R| \leq 2^{k-1} || Y_2| \geq \frac{|Y|}{2}) + \mathbf{P}(|Y_2| \leq \frac{|Y|}{2}).$$

It follows from Theorem 2.1.5 and Bayes' Theorem, along with our lower bound on $|Y_2|$, that the first term is $O(|Y_2|^{-1/2}) = O(m^{-1/2})$.

To bound the other term, we first note that if $i \in Y$, then R_i is a polynomial of degree $k-1$ containing (by our definition of $|Y|$ and S) at least $O(mn^{k-1})$ nonzero terms. It follows by our inductive hypothesis that for any particular $i \in Y$,

$$\mathbf{P}(|R_i| \leq 1) = O((m\rho)^{-2^{-(k^2-k-2)/2}}).$$

By Markov's inequality, it therefore follows that

$$\mathbf{P}(|Y_2| \leq \frac{|Y|}{2}) = O((m\rho)^{-2^{-(k^2-k-2)/2}}),$$

and thus that

$$\mathbf{P}(|R| \leq 2^{k-1}) = O((m\rho)^{-2^{-(k^2-k-2)/2}}),$$

so we are finished. ■

Chapter 3

The Rank of Dense Random Matrices

3.1 Introduction and Statement of Main Result

In this chapter, we will focus on symmetric random matrices whose off-diagonal entries are individually not too concentrated on any particular value. In particular, this class of matrices will include the adjacency matrices of random graphs $G(n,p)$ where the edge probability p either remains fixed or tends slowly towards 0 as n tends to infinity. The adjacency matrices of these graphs are random symmetric matrices whose entries above the main diagonal are independent Bernoulli variables which are 1 with probability p and 0 with probability $1-p$, with the entries below the main diagonal determined by symmetry. We will consider the case where p tends to 0 more quickly in the following chapter. Our main result is the following:

Theorem 3.1.1 *Let $0 < \alpha < 1/2$ be any fixed constant. Let Q_n be an $n \times n$ matrix whose upper diagonal entries $x_{ij}, 1 \leq i < j \leq n$, are independent random variables satisfying*

$$\sup_{c \in C} \mathbf{P}(x_{ij} = c) \leq 1 - n^{-\alpha}, \tag{3.1}$$

and whose lower diagonal entries are fixed according to the relationship $x_{ij} = x_{ji}$.

Then for any $\epsilon > 0$, the probability that Q_n is singular is $O(n^{\frac{1}{4}(2\alpha - 1 + \epsilon)})$, where the implied constant in the O notation is dependent only on ϵ and α.

Note that this theorem does not place any restrictions as to the distribution of the diagonal entries, which can even be entirely deterministic without changing the bound in the theorem. By applying Theorem 3.1.1 to a matrix with off-diagonal Bernoulli entries and equal, deterministic diagonal entries, we obtain the following corollary:

Corollary 3.1.2 *Let $0 < \alpha < 1/2$ and $\epsilon > 0$ be fixed. Let $p = p(n)$ satisfy $n^{-\alpha} < p < n^{1-\alpha}$. Then for any $c \in C$, the probability that c is an eigenvalue of the adjacency matrix of $G(n,p)$ is $O(n^{2\alpha-1+\epsilon})$, where the constant in the O notation is dependent on α and ϵ, but is independent of c. In particular, the adjacency matrix is singular with probability $O(n^{\frac{1}{4}(2\alpha-1)+\epsilon})$.*

Letting α tend to 0, we see that in particular the case where p is fixed between 0 and 1 leads to a singularity probability which is $O(n^{-1/4+\epsilon})$. Thus the adjacency matrix of $G(n,p)$ is almost surely nonsingular in this range.

3.2 Outline of Proof and Statement of Lemmas

As mentioned in chapter 1, our proof will be based on exposing the matrix minor by minor. Letting Q_m denote the upper left $m \times m$ minor of our matrix, we will divide the possible dependencies among the row vectors of Q_m into two classes depending on the number of vectors involved in the combination.

Definition 3.2.1 *Given m vectors $\{x_1, x_2, \ldots x_m\}$, a **linear combination** of the v_i is a vector of the form $v = c_1x_1 + c_2x_2 + \cdots + c_mx_m$, where the c_i are real numbers. The **degree** of a linear combination is the number of nonzero c_i.*

Let $M(m) = m^{1-\alpha-\epsilon}$. Our claim will be that the singular $m \times m$ matrices without vanishing linear combinations of degree at most M tend to behave well under augmentation, while matrices with vanishing linear combinations of degree that small will tend to be rare. This is made precise in the following definition and pair of lemmas.

Definition 3.2.2 *A symmetric singular $m \times m$ matrix is **normal** if its rows do not admit a vanishing linear combination of degree at most M. Otherwise, we call the matrix abnormal.*

Remark 3.2.3 *The terms normal and abnormal are used here only to refer to singular matrices, and there is no corresponding term for non-singular matrices.*

In section 3.4.1, we will prove that singular abnormal matrices are rare.

Lemma 3.2.4 *For any m between 0 and n, the probability that Q_m is both singular and abnormal is $o(\frac{1}{m^3})$.*

In section 3.5.1, we shall prove that singular normal matrices tend to augment well.

Lemma 3.2.5 *Let A be any fixed $m \times m$ singular, normal matrix. Then*

$$\mathbf{P}(\text{rank } (Q_{m+1}) - \text{rank } (Q_m) < 2 | Q_m = A) = O(M^{-1/2}) \tag{3.2}$$

We next turn to the augmentation of nonsingular matrices. We now classify them based on the degrees of the vanishing combinations of their row-deleted submatrices.

Definition 3.2.6 *A row r of a symmetric non-singular $m \times m$ matrix A is **good** if deleting r from A leads to a $(n-1) \times n$ matrix whose columns do not admit a linear combination of degree at most M.*

*An $m \times m$ symmetric non-singular matrix A is **perfect** if all of its rows are good. Otherwise, it is imperfect*

Remark 3.2.7 *The terms perfect and imperfect are used here only to refer to non-singular matrices, and there is no corresponding term for singular matrices.*

The next two lemmas can be thought of as the nonsingular analogues of Lemmas 3.2.4 and 3.2.5. Together, they will state that it is rare the have a nonsingular matrix that fails to augment well.

Lemma 3.2.8 *For any m between 0 and n, the probability that Q_m is both non-singular and imperfect is $o(\frac{1}{m^2})$.*

Lemma 3.2.9 *Let A be any fixed $m \times m$ non-singular, perfect matrix. Then*

$$\mathbf{P}(\text{rank } (Q_{m+1}) = m + 1 | Q_m = A) = O(M^{-1/4}). \tag{3.3}$$

In the next section, we will prove Theorem 3.1.1 assuming that all of the lemmas from this section are true. The remainder of this chapter will consist of the proofs of those lemmas.

3.3 Proof of Theorem 3.1.1 Assuming the Lemmas

Let us now assume that all lemmas from the previous section are true. We will use a variant of an argument from [9] to show that Q_n almost surely has full rank.

Let B_1 be the event that every Q_m with $\sqrt{n} \leq j \leq n$ is either both singular and normal or both non-singular and perfect. By Bayes' inequality, we have

$$\mathbf{P}(\text{rank } Q_n < n) \leq \mathbf{P}(\text{rank } Q_n < n \wedge B_1) + \mathbf{P}(\neg B_1). \tag{3.4}$$

However, for each individual m between $\sqrt{n}$ and n, we know from Lemmas 3.2.4 and 3.2.8 that the probability that Q_m is either abnormal or imperfect is $O(\frac{1}{m^3})$. By the union bound, then, it follows that

$$\mathbf{P}(\neg B_1) = O\left(\sum_{i=\sqrt{n}}^{n} \frac{1}{m^3}\right) = O\left(\frac{1}{n}\right).$$

It therefore suffices to show that the first term is $O(n^{\frac{1}{4}(1-2\alpha)})$.

Let $Y_m = m - \text{rank } (Q_m)$. We aim to show that Y_m is almost surely 0. To do so, we define a second auxiliary variable X_m as follows:

- $X_m = 4^{Y_m} = 4^{m-\text{rank } (Q_m)}$ if $Y_m > 0$ and every Q_j with $\sqrt{n} \leq j \leq n$ is either singular and normal or non-singular and perfect.
- $X_m = 0$ otherwise.

Note that X_m is nonzero precisely when Q_n is singular and B_1 also holds. We will show that X_m is almost surely 0 by showing that the expectation of X_m tends to 0. To obtain this bound on the expectation we will need the following lemma:

Lemma 3.3.1 *For any sequence $\mathcal{Q}_m = \{Q_{\sqrt{n}}, Q_{\sqrt{n}+1}, \ldots, Q_m\}$ encountered in the augmentation process,*

$$\mathbf{E}(X_{m+1}|\mathcal{Q}_m) \leq \frac{3}{5}X_m + O(n^{\frac{1}{4}(1-2c)}).$$

Proof (of Lemma 3.3.1): If any matrix in the sequence $\mathcal{Q}_m$ is either singular and abnormal or non-singular and imperfect, X_{m+1} equals 0 by definition, and we are done.

If every matrix in the sequence is either perfect or normal, and Q_m is of full rank (and thus perfect), then X_{m+1} will be either 0 or 4, with the latter occurring with probability $O(M^{-1/4})$ by Lemma 3.2.9. Thus we have

$$\mathbf{E}(X_{m+1}|\mathcal{Q}_m) = O(M^{-1/4}).$$

If on the other hand all matrices in the sequence are perfect or normal and Q_m has rank $m-j$ for some positive j, then by Lemma 3.2.5 X_{m+1} will be 4^{j-1} with probability $O(m^{-1/2})$, and will otherwise be at most 4^{j+1}. We therefore in this case have

$$\mathbf{E}(X_{m+1}|\mathcal{Q}_m) \leq 4^{j-1} + 4^{j+1}O(M^{-1/2}) \leq \frac{3}{5}4^j + O(M^{-1/2}) = \frac{3}{5}X_j + O(M^{-1/2}).$$

■

By Lemma 3.3.1, along with Bayes' Theorem, we have for any m between $\sqrt{n}$ and n and any $Q_{m'}$ that

$$\mathbf{E}(X_{m+1}|Q_{\sqrt{n}}) \leq \frac{3}{5}\mathbf{E}(X_m|Q_{\sqrt{n}}) + O(n^{\frac{1}{4}(1-2\alpha)}).$$

By induction on $m_2 - m_1$, we now have that for any $m_2 \geq m_1 \geq \sqrt{n}$ that

$$\mathbf{E}(X_{m_2}|Q_{\sqrt{n}}) \leq \left(\frac{3}{5}\right)^{m_2-m_1} \mathbf{E}(X_{m_1}|Q_{\sqrt{n}}) + O(n^{\frac{1}{4}(1-2\alpha)}).$$

In particular, we can take $m_2 = n$ and $m_1 = n'$ in the above inequality, yielding

$$\mathbf{E}(X_n|Q_{\sqrt{n}}) \leq \left(\frac{3}{5}\right)^{n-\sqrt{n}} X_{\sqrt{n}} + O(n^{\frac{1}{4}(1-2\alpha)}).$$

On the other hand, we know that $Y_{\sqrt{n}} \leq \sqrt{n}$, as the rank of $Q_{\sqrt{n}}$ is nonnegative. It follows that for any $Q_{\sqrt{n}}$ we have

$$\mathbf{E}(X_n|Q_{\sqrt{n}}) \leq \left(\frac{3}{5}\right)^{n-\sqrt{n}} 4^{\sqrt{n}} + O(n^{\frac{1}{4}(1-2c)}) = O(n^{\frac{1}{4}(1-2\alpha)}).$$

Adding up over all possible $Q_{\sqrt{n}}$ using Bayes' Theorem, we have that

$$\mathbf{E}(X_n) = O(n^{\frac{1}{4}(1-2\alpha)}).$$

Since X_n will always either be equal to 0 or at least 4, it follows from Markov's inequality that the probability that X_n is 0 is also $O(n^{\frac{1}{4}(1-2c)})$. This bound on the first term of the right hand side of (3.4) completes the proof.

3.4 Proof of Lemmas 3.2.4 and 3.2.8

3.4.1 Proof of Lemma 3.2.4:

Let $g(m,d)$ denote the probability that the d^{th} row of Q_m lies in the span of the first $d-1$ rows. We will first bound $g(m,M)$, then use the union bound to extend this to a bound on the probability that Q_m is abnormal.

The method of bounding $g(m,M)$ is a variation on one first used in [21] and [19]: As the first $M-1$ rows span a space of dimension at most $M-1$, there must be a collection of at most $M-1$ columns which parameterize this subspace. In other words, once some set of at most $M-1$ entries of the M^{th} row is exposed, the remaining entries are forced by our requirement that that row lie in the span of the first $M-1$ rows. However, these entries are all independent and random, with the exception of the (at most) $M-1$ entries which are already exposed from our knowledge of the first $M-1$ rows and the symmetry of Q_m and the entry in that row on the main diagonal.

Thus we have at least $m-2M+1$ random matrix entries, each of which must take on one specific value. Since by (3.1) each entry can take on its required value with probability at most $1-n^{-\alpha}$, it follows that

$$g(m,M) \leq (1-n^{-\alpha})^{m-2M+1}.$$

By symmetry, the same holds true for the probability that any given row lies in the span of any given subset of $M-1$ rows. Since there are $M\binom{m}{M}$ choices for the row and subset in question, we have by the union bound over all subsets of size M that

$$\begin{aligned}
\mathbf{P}(Q_m \text{ is abnormal}) &\leq M\binom{m}{M}g(m,M) \\
&\leq M\binom{m}{M}(1-n^{-\alpha})^{m-2M+1} \\
&\leq M\left(\frac{me}{M}\right)^M e^{-n^{-\alpha}(m-2M)},
\end{aligned}$$

where to obtain the second inequality we used (1.5) and (1.4). Substituting for M (and using the fact that $M = o(m)$), we have

$$\begin{aligned}
\mathbf{P}(Q_m \text{ is abnormal}) &\leq M\left(\frac{me}{M}\right)^M e^{-n^{-\alpha}m(1+o(1))} \\
&\leq m^{1-\alpha-\epsilon}m^{(\alpha+\epsilon+o(1))m^{1-\alpha-\epsilon}}e^{-m^{1-\alpha}}.
\end{aligned}$$

It follows that

$$\begin{aligned}\ln(\mathbf{P}(Q_m \text{ is abnormal})) &\leq -m^{1-\alpha} + (\alpha + \epsilon + o(1))m^{1-\alpha-\epsilon}\ln m \\ &= -m^{1-\alpha}(1+o(1)).\end{aligned}$$

In particular, the probability that Q_m is abnormal is $o(m^{-5})$.

3.4.2 Proof of Lemma 3.2.8:

By symmetry and the union bound, the probability that Q_m fails to be perfect is at most $mb(m)$, where $b(m)$ is the probability that the last row of Q_m fails to be good. However, $b(m)$ is in turn at most m times the probability that the upper left $(m-1)\times(m-1)$ minor Q_{m-1} is abnormal, which by Lemma 3.2.4 is $o(\frac{1}{(m-1)^5})$. Thus Q_m fails to be perfect with probability $o(\frac{1}{m^3})$.

Alternatively, $b(m)$ can be bounded directly via an identical argument to that in the previous section.

3.5 Proof of Lemmas 3.2.5 and 3.2.9

3.5.1 Proof of Lemma 3.2.5:

Let A be the given singular, normal matrix, and let the rank of A be $D < m$. Without loss of generality, we can assume that the first D rows $x_1, x_2, \ldots x_D$ of A that are linearly independent, so that the last row x_m can be written as a linear combination

$$x_m = \sum_{i=1}^{D} a_i x_i$$

of these rows in a unique way. Let D' be the number of a_i which are nonzero. Because A is normal by hypothesis, we know that $D' \geq M$.

We next examine the new column $(y_1, y_2, \ldots y_m, y_{m+1})$ added in the augmentation. If this column does not satisfy

$$y_m = \sum_{i=1}^{D} a_i y_i, \tag{3.5}$$

then the addition of this column must have increased the rank by 1. By Theorem 2.1.4, the probability that this fails to happen is $O(M^{-1/2})$.

It follows by symmetry that if the addition of the new column increased the rank of A by 1, the addition of its transpose as a new row must have further increased the rank by 1 (as the new row killed the same dependency in the columns that the new column killed in the rows). Thus the probability that the rank of A fails to increase by 2 after being augmented is $O(M^{-1/2})$ as well.

3.5.2 Proof of Lemma 3.2.9

We now assume that $A = Q_m$ has full rank and is perfect. Let $(y_1, y_2, \ldots y_{m+1})$ be the new column which, along with its transpose, is added in the augmentation of A to form Q_{m+1}. Let Q be the determinant of the augmented matrix Q_{m+1}. Expanding via cofactors in the last column and row, we obtain

$$Q = (\det\ A)y_{m+1} + \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}y_iy_j, \tag{3.6}$$

where c_{ij} denotes the (i, j) cofactor of A. We wish to bound the probability that $Q = 0$, so by Corollary 2.2.6 it suffices to show that many of the c_{ij} are nonzero.

As A is a nonsingular matrix, dropping any particular column of A will lead to an $m \times m-1$ matrix whose rows admit (up to scaling) precisely one nontrivial linear combination. Furthermore, since A is assumed to be perfect, there must be at least M rows involved in this combination. After removing any row involved in that combination from A, we are left with a nonsingular $m-1 \times m-1$ matrix. It thus follows that for any column of A, there are at least M rows that have a nonzero cofactor in that column, and thus at least M different i for which c_{ij} is equal to 0.

For any fixed y_{m+1}, therefore, it follows by Corollary 2.2.6 that

$$\mathbf{P}(\sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}y_iy_j = -y_{m+1}(\det\ A)) = O(M^{-1/4}).$$

Using Bayes' Theorem and integrating over y_{m+1}, we obtain the desired result.

3.6 Further Conjectures

There are several ways in which Theorem 3.1.1 seems unlikely to be sharp.

- The resulting bound is only $o(1)$ when $\alpha < 1/2$, so the theorem does not give any meaningful information in the case where some off-diagonal entries are highly concentrated (with probability at least $1 - n^{-1/2}$) on a given value. Although some sort of concentration bound on the individual entries must be present (or else the matrix could be effectively deterministic), it seems likely that the given restrictions are too strong. A more natural bound would be something along the order of

$$\sup_{i,j,c} \mathbf{P}(a_{ij} = c) \leq 1 - \frac{(1+\epsilon)\ln n}{n}, \tag{3.7}$$

since this is as weak a restriction as is possible without allowing some rows to become effectively non-random. Some support to this bound is given in the next chapter, where we allow the entries to concentrate highly on 0, but the matrix remains almost surely non-singular until (3.7) is violated.

- Even when α is close to 0, the bound of $O(n^{2\alpha-1+\epsilon})$ on the singularity probability seems unlikely to be optimal. In the nonsymmetric case, it was shown [19, 31] that the singularity probability was exponentially small if each entry was 1 with probability $\frac{1}{2}$ and 0 otherwise. However, the corresponding bound in the symmetric case obtained by Corollary 3.1.2 is only $n^{-1/4+\epsilon}$.

  More generally, let $Q(n,p)$ denote the adjacency matrix of $G(n,p)$. A possible conjecture is that the singularity probability is dominated by the probability that there is a dependency involving at most 2 rows. By taking the maximum of the probability that a single row is 0 and the probability that two rows are equal, we obtain the following conjecture:

  Conjecture 3.6.1 *The probability that the adjacency matrix of $Q(n,p)$ is singular is*

$$O(\max\{n(1-p)^n, n^2(1-p)(p^2+(1-p)^2)^{n-1}\}).$$

  This corresponds with the decades-old conjecture that the probability a random (non-symmetric) Bernoulli matrix is singular is $(1/2 + o(1))^n$.

- Consider the case of a random symmetric matrix $Q(n)$ whose entries are each 1 or -1, each with probability 1/2. Theorem 3.1.1 states that the matrix almost surely has nonzero determinant, and parity arguments can give that it has determinant at least 2^{n-1}. However, it should be possible to get some sort of stronger bound. In [30] it was shown that the corresponding non-symmetric matrix (still with each entry equally likely to be ± 1) almost surely has determinant $n^{(\frac{1}{2}+o(1))n}$ (corresponding to the upper bound given by Hadamard's inequality), and it seems a similar result should hold for $Q(n)$.

- The *condition number* $||A|| \cdot ||A^{-1}||$ of a matrix A arises frequently in numerical analysis. The spectral norm $||Q(n,p)||$ is well understood even for p very close to 0, due to results of Füredi and Komlós [15] and Krivelevich and Sudakov [24]. In order to estimate the condition number of $Q(n,p)$, therefore, it would suffice to estimate the smallest (in absolute value) eigenvalue of $Q(n,p)$. Even for the case of $p = 1/2$, all Theorem 3.1.1 states is that this eigenvalue is with high probability not 0.

  In this case we know by the Wigner semicircle law that almost all of the n eigenvalues of $Q(n,1/2)$ lie in an interval whose length is of order $\sqrt{n}$, so a natural conjecture is that with high probability the closest eigenvalue to 0 is at a distance on the order of $n^{-1/2}$. This would correspond with the known results for random asymmetric Gaussian matrices due to Edelman [11], and more general classes of random matrices due to Rudelson-Vershynin [27] and Tao-Vu [33]

Chapter 4

The Rank of Sparse Random Graphs

4.1 Introduction and Statement of Main Results

As before, let $Q(n,p)$ denote the adjacency matrix of a graph in $G(n,p)$. In other words, $Q(n,p)$ is a random symmetric matrix which is 0 on the main diagonal and has independent entries above the main diagonal, each of which is 0 with probability p and 1 with probability $1-p$.

We know from the previous chapter that if p is at least $n^{1/2+\epsilon}$, then $Q(n,p)$ will almost surely be non-singular. On the other hand, it is easy to see that $Q(n,p)$ will almost surely *not* be of full rank if n is too close to 0. If p is smaller than $\frac{\ln n}{n}$, then the graph $G(n,p)$ will almost surely contain isolated vertices, which in turn correspond to rows in the adjacency matrix which are entirely zero, and thus to a singular adjacency matrix. In this chapter our goal will be to analyze what happens to the rank of this matrix as p drops to and below $\frac{\ln n}{n}$.

For a set S in a graph G, let $N(S)$ denote the neighborhood of S, that is, the set of vertices adjacent to S. (We allow $N(S)$ to contain elements of S.) A set S of vertices of a graph G is **nonexpanding** if $N(S)$, the neighborhood of S in G, satisfies $|N(S)| < |S|$. S is **minimally nonexpanding** if it contains no proper nonexpanding subset.

Examples. An isolated vertex forms a nonexpanding set. A set of two vertices of degree one sharing a common neighbor forms a minimal nonexpanding set of size 2.

A set T of rows of a matrix A is nonexpanding if at most $|T|-1$ columns of A have at least one nonzero entry in T. Note that a set of vertices of a graph is nonexpanding if and only if the corresponding rows of its adjacency matrix are also nonexpanding.

Furthermore, if S is a nonexpanding subset of G, then the corresponding rows in the adjacency matrix of G are dependent, since they span a space of dimension at most $|N(S)|$. It is clear that for any set S of vertices

$$\text{rank } (A(G)) \leq n - |S| + |N(S)|.$$

Here and later $A(G)$ denotes the adjacency matrix of the graph G (with vertex set V). It follows that

$$\text{rank } (A(G)) \leq \min_{S \subset V}(|V| - |S| + |N(S)|).$$

Our first main result shows that for a random graph $G(n,p)$, this upper bound is *tight*.

Theorem 4.1.1 *Assume that* $1/2 > p = \Omega(\frac{\ln n}{n})$. *Then almost surely*

$$\text{rank } (Q(n,p)) = (\text{rank } A(G(n,p)) = \min_{S \subset V} n - |S| + |N(S)|) \tag{4.1}$$

Remark 4.1.2 *Both this result and Theorem 4.1.3 can be extended to cover matrices whose independent entries have the form* $\nu_{ij}\xi_{ij}$, *where* ν_{ij} *is a Bernoulli random variable taking on 1 with probability* p, *and* ξ_{ij} *has any distribution, so long as* $\mathbf{P}(\xi_{ij} = 0) = 0$. *The* ξ_{ij} *can even be different for different entries in the matrix, though it is critical to our argument that all entries have the same probability of being 0. In this case, the underlying graph should be thought of as the graph whose adjacency matrix is given by the* ν_{ij}.

It can be checked that for any fixed positive integer s and $\epsilon > 0$, if $p < \frac{(1-\epsilon)\ln n}{sn}$ then $G(n,p)$ almost surely contains minimal nonexpanding sets of all sizes up to and including s. Each of these nonexpanding sets generates a set of dependent rows in the adjacency matrix. Our second main result shows that this is the *only* reason for dependency.

Theorem 4.1.3 *Let $\frac{c \ln n}{n} < p < 1/2$, where $c > 1/s$ for some positive integer s. Then with probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, $G(n,p)$ has the property that any set of dependent rows of its adjacency matrix contains a nonexpanding set of size at most $s-1$.*

Remark 4.1.4 *Although this theorem holds for $p < 1/2$ (and can easily be extended to $p < 1 - \epsilon$ for any ϵ), it is superseded by 3.1.1 in the case where $p > n^{-1/2+\epsilon}$.*

This theorem implies that if $\frac{c \ln n}{n} < p < 1/2$, where $c > 1/s$ for some positive integer s, then any minimal dependent set of rows has size at most $s-1$. The case where $s \leq 2$ was examined by the author and Vu in [10]. Those cases can be stated in the following corollary.

Corollary 4.1.5 *Let ϵ be any positive constant. If $\frac{(1+\epsilon) \ln n}{2n} \leq p \leq \frac{1}{2}$, then the rank of $G(n,p)$ will almost surely be equal to the number of nonisolated vertices. In particular, if $\frac{(1+\epsilon) \ln n}{n} \leq p \leq \frac{1}{2}$, then the adjacency matrix of $G(n,p)$ will almost surely be nonsingular.*

Theorem 4.1.3 also implies the following corollary.

Definition 4.1.6 *A set S of vertices of a graph G is **s-unobstructed** if it contains no nonexpanding subset of size at most s. S is **unobstructed** if it contains no nonexpanding subset.*

Corollary 4.1.7 *Let $p = \Omega(\frac{\ln n}{n})$. With probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, the rank of the random graph $G(n,p)$ equals the size of its largest unobstructed set.*

In the next section we will state a slightly weaker version of Theorem 4.1.3, along with some lemmas which will be used in the proof of that result. The next several sections will be given over to proven those lemmas and the weaker theorem, after which Theorems 4.1.1 and 4.1.3 will be deduced from the weaker theorem along with some of the lemmas. The last sections will be devoted to several consequences of the two main theorems, along with a few final conjectures.

4.2 The Idea of the Proofs and Some Lemmas

Instead of proving Theorems 4.1.1 and 4.1.3 directly, we are going to prove the following theorem (which is somewhat weaker than Theorem 4.1.3). The proof of this theorem, combined with some lemmas will imply Theorems 4.1.1 and 4.1.3.

We say that a graph G is s-**saturated** if the rank of $A(G)$ equals the size of the largest s-unobstructed set.

Theorem 4.2.1 *Let $\frac{c \ln n}{n} < p < 1/2$, where $c > 1/s$ for some positive integer s. With probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, the random graph $G(n,p)$ is $(s-1)$-saturated.*

As in the previous chapter, we are going to expose $Q(n,p)$ minor by minor, so that Q_{m+1} is formed by taking Q_m and augmenting by a column whose entries are chosen independently, along with the column's transpose. Denote by G_m the graph whose adjacency matrix is Q_m. In graph theoretic terms, we are considering a vertex exposure process of $G(n,p)$.

Our starting observation is that when a good portion of the vertices have been exposed, the rank of the matrix is close to its size.

Recall that $p \geq c \ln n/n$ for a constant $c > 1/s$. Let $0 < \delta < 1$ be a constant such that $1/s < \delta c < 1/(s-1)$. Define $n' := \delta n$.

Lemma 4.2.2 *For any constant $\epsilon > 0$ there exists a constant $\gamma > 0$ such that*

$$\mathbf{P}(\mathit{rank}(Q_{n'}) < (1-\epsilon)n') = o(e^{-\gamma n \ln n})$$

Remark 4.2.3 *This lemma is needed since our remaining bounds will only be able to control the behavior of the matrix for $m > n'$. Since we may only be able to control the behavior of the rank for a small (positive) fraction of the augmentation, we need to be sure that the rank isn't too far away from where we want it to be by the time we gain that control*

Our plan is to show that the addition of the remaining $n - n'$ rows/columns is enough to remove all the linear dependencies from $Q_{n'}$ except those corresponding to nonexpanding subsets of at most $s-1$ vertices.

The next lemmas provide some properties of the (random) graph G_m for $n' \leq m \leq n$.

Definition 4.2.4 *A graph G is* **well-separated** *if the following two conditions hold:*

W1. Any connected subgraph of G on at most $5s$ vertices contains at most $s-1$ vertices with degree at most $\ln \ln n$.

W2. No cycle of length at most $12s$ in G contains a vertex of degree at most $\ln \ln n$.

Lemma 4.2.5 *For any constant $\epsilon > 0$, the probability that there is an m between n' and n for which G_m is not well separated is $O(n^{-sc\delta+1+\epsilon})$.*

Note that by our choice of δ this probability will be $o(1)$ for sufficiently small ϵ.

Definition 4.2.6 *A graph G is a* **small set expander** *if every subset S of the vertices of G with $|S| \leq \frac{n}{\ln^{3/2} n}$ either has at least $|S|$ edges connecting S to $\bar{S}$, its complement, or has a subset $S' \subset S$ with $|S'| \leq s-1$ and at most $|S'|-1$ edges connecting S' to $\bar{S}'$.*

Lemma 4.2.7 *For any $m > n'$ the probability that G_m is well separated but is not a small set expander is $O(n^{-4})$.*

Definition 4.2.8 *A set S of the vertices of a graph G is* **nice** *if there are at least two vertices of G each is adjacent to exactly one vertex in S.*

A set S of the vertices of a graph G is **nearly nice** *if there is at least one vertex in G which has exactly one neighbor in S.*

Set $k := \frac{\ln \ln n}{2p}$. We will next define a class of 'good' matrices which behave well under augmentation.

Definition 4.2.9 *A graph G is* **good** *if the following four properties hold:*

1. Every minimal non-nice subset of the vertices of G either has size at least $k+1$ or contains a non-expanding subset of size at most $s-1$.

2. Every minimal non-nearly nice subset of the vertices of G either has size at least $k+1$ or is a non-expanding set of size at most $s-1$.

3. At most $\frac{1}{p \ln n}$ vertices of G have degree less than s.

4. G is well separated.

A symmetric (0,1) matrix A is **good** *if the graph for which it is an adjacency matrix is good.*

The next lemma states that in the augmentation process we will likely run only into good matrices.

Lemma 4.2.10 *Let ϵ be a positive constant. Then with probability $1 - O(n^{1-sc\delta+\epsilon})$, Q_m is good for every m between n' and n .*

We now consider the effect of augmentation on the rank of A when A is a good matrix.

Definition 4.2.11 *A pair (G, G') of graphs is called* **normal** *if the following properties hold:*

1. G is an induced subgraph on $|G'| - 1$ vertices of G'.

2. The new vertex added to G is not adjacent to any vertex which was part of a non-nearly nice subset in G'.

A pair (A, A') of $(0, 1)$ symmetric matrices is normal if the pair of graphs for which they are the adjacency matrices are normal.

Lemma 4.2.12 *Let A be any fixed, good $m \times m$ matrix which is not $s - 1$-saturated. Then*

$$\mathbf{P}(\ rank(Q_{m+1}) - \ rank(Q_m) < 2 | (Q_m, Q_{m+1}) \text{ is normal} \wedge Q_m = A) = O((kp)^{-1/2}).$$

What the above lemma says is that if Q_m is good but not $s - 1$-saturated, then augmenting it will tend to remove some of the dependencies among the rows of Q_m (note that kp is tending to infinity by assumption). If (Q_{m+1}, Q_m) is normal, then the dependencies removed can't correspond to small nonexpanding subsets of the rows of Q_m. This implies that in some sense Q_{m+1} is a little closer to being saturated then Q_m was.

Now suppose on the other hand that Q_m is both good and already $s - 1$-saturated. We are going to show that (again assuming no change in the nonexpanding subsets) with high probability Q_m does not gain any new dependencies by being augmented.

Lemma 4.2.13 *Let A be any fixed, good $m \times m$ matrix which is also $s-1$-saturated. Then*

$$\mathbf{P}(\ \mathit{rank}(Q_{m+1}) - \ \mathit{rank}(Q_m) < 2 | (Q_m, Q_{m+1})\ \mathit{is\ normal} \wedge Q_m = A) = O((kp)^{-1/4}).$$

ln the next section, we prove Theorem 4.2.1 assuming these lemmas. The proofs of the lemmas will be presented in the sections that follow.

4.3 Proof of Theorem 4.2.1 Assuming All Lemmas

In this section, we assume all lemmas from the previous section are true. We are going to use a variant of the argument from section 3.3. Let B_0 be the event that G_n is $(s-1)-$saturated. Let B_1 be the event that the rank of $Q_{n'}$ is at least $n'(1-\frac{1-\delta}{4\delta})$. Let B_2 be the event that Q_m is good for all $n' \leq m < n$. By Bayes' theorem we have

$$\mathbf{P}(B_0) \leq \mathbf{P}(B_0 \wedge B_2 | B_1) + \mathbf{P}(\neg B_1) + \mathbf{P}(\neg B_2)$$

By Lemma 4.2.2 we have that $\mathbf{P}(\neg B_1) = o(e^{-\gamma n \ln n})$ and by Lemma 4.2.10 we have that $\mathbf{P}(\neg B_2) = O(n^{1-sc\delta+\epsilon})$. Both of these probabilities are much smaller than the bound $O((\ln \ln n)^{-1/4})$ which we are trying to prove, so it only remains to bound the first term.

Let U_m denote the size of the largest $(s-1)-$unobstructed subset of the vertices of G_m.

Let $Y_m = U_m - \ \mathrm{rank}(Q_m)$. Our goal is now to prove that Y_n is almost surely 0. Define a random variable X_m as follows:

- $X_m = 4^{Y_m}$ if $Y_m > 0$ and every Q_j with $n' \leq j \leq m$ is good;
- $X_m = 0$ otherwise.

The core of the proof is the following bound on the expectation of X_{m+1} given any fixed sequence $\mathcal{Q}_m$ of matrices $\{Q_{n'}, Q_{n'+1}, \ldots, Q_m\}$ encountered in the augmentation process.

Lemma 4.3.1 *For any sequence $\mathcal{Q}_m = \{Q_{n'}, Q_{n'+1}, \ldots, Q_m\}$ encountered in the augmentation process,*

$$\mathbf{E}(X_{m+1} | \mathcal{Q}_m) < \frac{3}{5} X_m + O((\ln \ln n)^{-1/4}).$$

Let us (for now) assume Lemma 4.3.1 to be true. This lemma together with Bayes theorem shows that for $n' < m$ we have

$$\mathbf{E}(X_{m+1}|Q_{n'}) < \frac{3}{5}\mathbf{E}(X_m|Q_{n'}) + O((\ln\ln n)^{-1/4}).$$

By induction on $m_2 - m_1$ we now have that for any $m_2 \geq m_1 \geq n'$

$$\mathbf{E}(X_{m_2}|Q_{n'}) < (\frac{3}{5})^{m_2-m_1}\mathbf{E}(X_{m_1}|Q_{n'}) + O((\ln\ln n)^{-1/4}).$$

In particular, by taking $m_2 = n$ and $m_1 = n'$ we get that

$$\mathbf{E}(X_n|Q_{n'}) < (\frac{3}{5})^{n-n'}X_{n'} + O((\ln\ln n)^{-1/4}).$$

If $Q_{n'}$ satisfies B_1, we automatically have $X_{n'} \leq 4^{\frac{(1-\delta)n'}{4\delta}} = (\sqrt{2})^{n-n'}$, so

$$\mathbf{E}(X_n|Q_{n'}) < (\frac{3\sqrt{2}}{5})^{n-n'} + O((\ln\ln n)^{-1/4}) = O((\ln\ln n)^{-1/4}).$$

By Markov's inequality, for any $Q_{n'}$ satisfying B_1

$$\mathbf{P}(X_n > 3|Q_{n'}) = O((\ln\ln n)^{-1/4})$$

On the other hand, by definition $X_n \geq 4$ if G_n is not $(s-1)-$saturated and B_2 holds. It thus follows by summing over all $Q_{n'}$ satisfying B_1 that

$$\mathbf{P}(B_0 \wedge B_2|B_1) = O((\ln\ln n)^{-1/4}),$$

proving the theorem.

It remains to prove Lemma 4.3.1. If a matrix in the sequence $\{Q_{n'}, Q_{n'+1}, \ldots, Q_m\}$ is not good, then $X_{m+1} = 0$ by definition and there is nothing to prove. Thus, from now on we can assume that all matrices in the sequence are good. Let Z_m denote the number of vertices of degree at most s in Q_m adjacent to the $m+1^{st}$ vertex of G.

Claim: $U_{m+1} - U_m \leq Z_m + 1$.

Proof (of claim): Let S_{m+1} denote a $s-$unobstructed subset of the vertices of G_{m+1} such that $|S_{m+1}| = U_{m+1}$. Let S'_m denote the set formed by removing the $m+1^{st}$ vertex from S, as well as any vertices of degree at most s adjacent to that new vertex.

S'_m is $s-$unobstructed since each subset of S'_m of size at most s either contains a vertex of degree at least $s+1$ (in which case it clearly expands) or has the same

neighborhood in G_m as in G_{m+1}. Since at most $Z_m + 1$ vertices were removed to go from S_{m+1} to S'_m, the claim follows. ∎

By the above claim, if Z_m is positive, then augmenting the matrix will increase Y_m by at most $Z_m + 1$ (U_m increases by at most $Z_m + 1$ and the rank does not decrease). Furthermore, $Z_m = 0$ if and only if (Q_m, Q_{m+1}) is normal. By Bayes' theorem, we have

$$\begin{aligned}
&\mathbf{E}(X_{m+1}|\mathcal{Q}_m) \\
&\quad = \mathbf{E}(X_{m+1}\chi(Z_m > 0)|\mathcal{Q}_m) \\
&\qquad\quad +\mathbf{E}(X_{m+1}|\mathcal{Q}_m \wedge (Q_m, Q_{m+1}) \text{ is normal})\mathbf{P}((Q_m, Q_{m+1}) \text{ is normal}|\mathcal{Q}_m) \\
&\quad \leq \mathbf{E}(X_{m+1}\,\chi(Z_m > 0)|\mathcal{Q}_m) + \mathbf{E}(X_{m+1}|\mathcal{Q}_m \wedge (Q_m, Q_{m+1}) \text{ is normal}) \\
&\quad = \mathbf{E}(4^{Z_m+1+Y_m}\chi(Z_m > 0)|\mathcal{Q}_m) + \mathbf{E}(X_{m+1}|\mathcal{Q}_m \wedge (Q_m, Q_{m+1}) \text{ is normal}).
\end{aligned}$$

Since Q_m is good, G_m has at most $\frac{1}{p \ln n}$ vertices which have degree at most s. Thus, we can bound Z_m by the sum of $\frac{1}{p \ln n}$ random Bernoulli variables, each of which is 1 with probability p. It follows that

$$\mathbf{P}(Z_m = i) \leq \binom{(p \ln n)^{-1}}{i} p^i \leq (\ln n)^{-i}.$$

Adding up over all i, we have

$$\mathbf{E}(4^{Z_m+1}\chi(Z_m > 0)|\mathcal{Q}_m) \leq \sum_{i=1}^{\infty} 4^{i+1}(\ln n)^{-i} = O((\ln n)^{-1}).$$

If $Y_m = 0$ and (Q_m, Q_{m+1}) is normal, then by Lemma 4.2.13 (which applies since Q_m is good) X_{m+1} is either 0 or 4, with the probability of the latter being $O((\ln \ln n)^{-1/4})$. Therefore we have for any sequence $\mathcal{Q}_m = \{Q_{n'}, \ldots Q_m\}$ of good matrices with $Y_m = 0$ that

$$\mathbf{E}(X_{m+1}|\mathcal{Q}_m) = O((\ln \ln n)^{-1/4} + (\ln n)^{-1}) = O((\ln \ln n)^{-1/4}). \tag{4.2}$$

If $Y_m = j > 0$ and (Q_m, Q_{m+1}) is normal, then Y_{m+1} is $j - 1$ with probability $1 - O((\ln \ln n)^{-1/2})$ by Lemma 4.2.12, and otherwise is at most $j + 1$. Combining this with the bound on $\mathbf{E}(4^{Z_m+1}\chi(Z_m > 0)|\mathcal{Q}_m)$ we have

$$\mathbf{E}(X_{m+1}|\mathcal{Q}_m) = 4^{j-1} + 4^{j+1}O((\ln \ln n)^{-1/2}) + 4^j O((\ln n)^{-1}) \leq \frac{3}{5}4^j \tag{4.3}$$

The lemma now follows immediately from (4.2) and (4.3).

4.4 Proof of Lemma 4.2.2

By symmetry and the union bound

$$\mathbf{P}(\text{rank}(Q_{n'}) < (1-\epsilon)n') \leq \binom{n'}{\epsilon n'} \times \mathbf{P}(B_1^*),$$

where B_1^* denotes the event that the last $\epsilon n'$ columns of Q'_n are contained in the span of the remaining columns.

We view $Q_{n'}$ as a block matrix,

$$Q_{n'} = \left[\begin{array}{c|c} A & B \\ \hline B^T & C \end{array}\right],$$

where A is the upper left $(1-\epsilon)n' \times (1-\epsilon)n'$ sub-matrix and C has dimension $\epsilon n' \times \epsilon n'$. We obtain an upper bound on $\mathbf{P}(B_1^*)$ by bounding the probability of B_1^* conditioned on any fixed A and B (treating C as random).

B_1^* cannot hold unless the columns of B are contained in the span of those of A, meaning the equation $B = AF$ holds for some matrix F. If this is the case, then B_1^* will hold only when we also have $C = B^T F$. This means that each entry of C is forced by our choice of A, B and our assumption that B_1^* holds.

However, C is still random, and the probability that any given entry takes on its forced value is at most $1-p$. The entries are not all independent (due to the symmetry of C), but those on or above the main diagonal are. Therefore the probability that B_1^* holds for any fixed A and B is at most $(1-p)^{\frac{(\epsilon n')^2}{2}}$.

We therefore have

$$\begin{aligned} \mathbf{P}(\text{rank}(Q_{n'}) < (1-\epsilon)n') &\leq \binom{n'}{\epsilon n'}((1-p)^{\frac{(\epsilon n')^2}{2}} \\ &\leq (\frac{n'e}{\epsilon n'})^{\epsilon n'} e^{\frac{-p(\epsilon n')^2}{2}} \\ &\leq {c_2}^n e^{-c_1 n \ln n}. \end{aligned}$$

where c_1 and c_2 are positive constants depending on ϵ, δ, and c (but independent of n).

4.5 Proof of Lemma 4.2.5

If $p \geq \frac{(\ln n)^2}{n}$, then $G(n,p)$ will with probability at least $1 - o(n^{-3})$ have no vertices whose degree is at most $\ln \ln n$, in which case G is trivially well-separated. We will therefore assume $p \leq \frac{(\ln n)^2}{n}$ for the remainder of this section.

If G_m fails to be well-separated for some m between n' and n, there must be a first m_0 with this property. We are going to bound the probability that a fixed m is this m_0.

Case 1: $m_0 = n'$. We can bound the probability $G_{n'}$ fails condition W1 above by the union bound over all sets of at most $5s$ vertices of the probability that those vertices form a connected subgraph with at least s small-degree vertices.

The probability that any single vertex has sufficiently small degree is at most

$$\sum_{i=0}^{\ln \ln n} \binom{n'-1}{i} p^i (1-p)^{n'-i} \leq (1+o(1)) \sum_{i=0}^{\ln \ln n} (n'p)^i (1-p)^{n'} \leq \frac{(\ln n)^{2 \ln \ln n}}{n^{c\delta}}, \tag{4.4}$$

so the probability that a set of size i contains at least s such vertices is at most

$$\binom{i}{s} n^{-sc\delta+\epsilon}.$$

The probability that a set of size i is connected is (by the union bound over all spanning trees) at most

$$i^{i-2} p^{i-1} \leq \frac{(\ln n)^{2i}}{(n')^{i-1}}.$$

By the FKG inequality [14], these two events are negatively correlated (as one is monotone increasing under edge inclusion, while the other is monotone decreasing), so the probability that some subset fails the first well-separation criterion is at most

$$\sum_{i=s}^{5s} \binom{n'}{i} \binom{i}{s} n^{-sc\delta+\epsilon} \frac{(\ln n)^{2i}}{(n')^{i-1}} = O(n^{1-sc\delta+\epsilon}). \tag{4.5}$$

Similarly, for each $3 \leq j \leq 12s$ the probability that a given set of size j contains a spanning cycle is by the union bound at most

$$\frac{(i-1)!}{2} p^i \leq \frac{(i-1)!(\ln n)^{2i}}{n^i},$$

and again by the FKG inequality this event is negatively correlated with the set containing a vertex of degree at most $\ln \ln n$ in G. Therefore the probability some set fails

W2 is at most

$$\sum_{i=3}^{12s} \binom{n'}{i} \frac{(i-1)!(\ln n)^{2i}}{n^i} \frac{(\ln n)^{2\ln\ln n}}{n^{c\delta}} = O(n^{-c\delta+\epsilon}) \tag{4.6}$$

Case 2: $m_0 = m > n'$. In this case, we can bound the probability that $m_0 = m$ above by the probability that G_{m-1} is well separated but G_m fails to be well separated. As in the previous case, we can take a union bound over all sets of at most $5s$ vertices, but now we need only consider sets which contain the vertex newly added to create G_m (all other sets are covered by our assumption that G_{m-1} is well separated). This means that the probability of failure of either requirement for any particular m in this range is at most $12s/n$ times the corresponding union bound in (4.5) and (4.6), which is $O(n^{-sc\delta+\epsilon})$.

By the union bound, the property that G_m is not well -separated for some m is at most

$$O(n^{1-sc\delta+\epsilon}) + O(n^{-c\delta+\epsilon}) + n \times O(n^{-sc\delta+\epsilon}) = O(n^{1-sc\delta+\epsilon}),$$

completing the proof.

4.6 Proof of Lemma 4.2.7

In order to prove the edge expansion property we first show that almost surely no small subgraphs of $G(n, \frac{c\ln n}{n})$ will have too many edges.

Definition 4.6.1 *A graph G is* **locally sparse** *if every subgraph on at most $\frac{n}{\ln^{3/2} n}$ vertices has average degree less than 8.*

Lemma 4.6.2 *For fixed c the probability that $G(n, \frac{c\ln n}{n})$ is not locally sparse is $O(n^{-4})$.*

Proof (of Lemma 4.6.2) Let q_j be the probability that a subset of size j has at least $4j$ edges. By the union bound, this is at most $\binom{n}{j}$ times the probability of a particular subset having at least $4j$ edges, so

$$\begin{aligned} q_j &\leq \binom{n}{j}\binom{j^2/2}{4j}p^{4j} \\ &\leq (\frac{ne}{j})^j(\frac{ejc\ln n}{8n})^{4j} \\ &\leq (\frac{c^4e^5j^3\ln^4 n}{n^3})^j. \end{aligned}$$

For $j < n^{1/4}$ this gives $q_j \leq n^{-2j}$, while for $j > n^{1/4}$ we have (using our upper bound on j) $q_j \leq (\ln n)^{-j/2} = o(n^{-5})$. By summing over all j at least 2, we conclude that the failure probability is $o(n^{-4})$, completing the proof. ■

Armed with this lemma we can now prove Lemma 4.2.7. We do so in two cases based on the size of p.

Case 1: $p < \frac{12\ln n}{n}$:

If G_m fails to expand edgewise there must be a minimal subset S_0 which both fails to expand and contains no nonexpanding subset of size at most $s-1$.

We claim that the subgraph formed by the vertices of S_0 must have average degree at least 8. For the sake of contradiction, let us suppose that this were not the case. Because fewer than $|S_0|$ edges leave S_0, it follows that the average degree of the vertices of S in G is at most 9, meaning that at most $\frac{9|S_0|}{\ln\ln n}$ vertices in S_0 have degree at least $\ln\ln n$ in G.

We next consider the connected components of the induced subgraph of G on the vertices in S_0. Unless G fails to satisfy condition W1 (in which case we are done), at most $\frac{9(s-1)|S_0|}{\ln\ln n}$ of the vertices with degree at most $\ln\ln n$ can be in the same component as a vertex with degree at least $\ln\ln n$.

Furthermore, the remaining vertices must be in components of size at most $s-1$ (again due to W1). Let T be one of those components. By assumption, T has at least $|T|$ edges leaving T, and each of these edges must also leave S_0. But this implies that $S_0 - T$ is a smaller edgewise nonexpanding set, a contradiction.

What we have actually shown in the above is the following *deterministic* statement: any locally sparse, well-separated graph must also be a small set expander. We can

therefore bound the probability of the existence of a minimal S_0 by the probability of G failing to be locally sparse, which by Lemma 4.6.2 is $O(n^{-4})$.

Case 2: $p \geq \frac{12 \ln n}{n}$:

In this range we will estimate the probability directly via the union bound over all subsets of size at most $\frac{n}{(\ln n)^{3/2}}$. The probability that some set fails to expand can be bounded above by

$$\begin{aligned} \sum_{i=1}^{n \ln^{-3/2} n} \binom{n}{i} ini - 1(1-p)^{i(n-i)-(i-1)} &\leq n^i (en)^{i-1} e^{-inp(1+o(1)} \\ &= \frac{1}{ne} \sum_{i=1}^{n \ln^{-3/2} n} (n^2 e^{-np(1+o(1)})^i. \end{aligned}$$

The lower bound we have on p in this case implies that each term in this last bound is $O(n^{-(4+o(1))i})$, so the sum is $o(n^{-4})$.

4.7 Proof of Lemma 4.2.10

Let C_0 be the event that G_m is good for every m between n' and n. Let C_1 be the event that G_m has at most $\frac{1}{p \ln n}$ vertices of degree less than s for every m between n' and n, C_2 be the event that G_m has maximum degree at most $5knp$ for each m, and C_3 be the event that G_m is well separated, locally sparse, and a small set expander for every m between n' and n. We have

$$\mathbf{P}(\neg C_0) \leq \mathbf{P}(\neg C_0 \wedge C_1 \wedge C_2 \wedge C_3) + \mathbf{P}(\neg C_1) + \mathbf{P}(\neg C_2) + \mathbf{P}(\neg C_3).$$

We are going to bound each term on the right hand side separately, in reverse order.

Lemmas 4.2.7, 4.2.5, and 4.6.2 together show that $\mathbf{P}(\neg C_3) = O(n^{1-sc\delta+\epsilon})$.

$\mathbf{P}(\neg C_2)$ is at most the expected number of vertices of degree at least $5knp$ in G_n, which is at most

$$n \binom{n}{5knp} p^{5knp} \leq n(e/5knp)^{5knp} p^{5knp} \leq ne^{-5knp} = o(n^{-4}).$$

To bound $\mathbf{P}(\neg C_1)$, we note that the probability that some G_m contains a set of vertices of size $t = p^{-1} \ln^{-1} n$, all of whose degrees are less than s, is bounded from above by the probability that at least t vertices in G_n each have at most s neighbors

amongst the vertices of $G_{n'}$. This probability is clearly decreasing in p, and for $p = \frac{\ln n}{sn}$ it is by Markov's inequality at most

$$\begin{aligned}\frac{n}{t}\sum_{i=1}^{s}\binom{n'}{i}p^i(1-p)^{n'-i} &< n\frac{(\ln n)^2}{sn}\sum_{i=1}^{s}n^i\frac{(\ln n)^i}{(sn)^i}e^{-(1+o(1))\ln n'/s} \\ &= O((\ln n)^2 n^{-1/s}\sum_{i=1}^{s}(\ln n)^i) = n^{-1/s+o(1)}.\end{aligned}$$

It remains to estimate the first term, which we will do by the union bound over all m. Since property C_1 implies G_m has few vertices of small degree, it suffices to estimate the probability that G_m contains a non-nice set while still satisfying properties C_1, C_2, and C_3. Let p_j be the probability that conditions C_1, C_2, and C_3 hold but some subset of j vertices causes G_m to fail to be good. Symmetry and the union bound give that p_j is at most $\binom{m}{j}$ times the probability that the three conditions hold and some fixed set S of j vertices causes the graph to fail condition C_0. We will bound this in three cases depending on the size of j.

Case 1: $\frac{1}{p\sqrt{\ln n}} \le j \le k$.

We will show that there are almost surely no non-nice subsets at all in this range, minimal or otherwise. Direct computation of the probability that a fixed set of j vertices has either 0 or 1 vertices adjacent to exactly one vertex in the set gives:

$$\begin{aligned}p_j &\le \binom{m}{j}((1-jp(1-p)^{j-1})^m + mjp(1-p)^{j-1}(1-jp(1-p)^{j-1})^{m-1}) \\ &\le (mep\sqrt{\ln n})^j((1-jp(1-p)^{j-1})^m + mjp(1-p)^{j-1}(1-jp(1-p)^{j-1})^{m-1}) \\ &\le (mep\sqrt{\ln n})^j((1-jpe^{-jp(1+o(1))})^m + mjpe^{-jp(1+o(1))}(1-jpe^{-jp(1+o(1))})^{m-1}) \\ &\le (mep\sqrt{\ln n})^j(e^{-mjp(1+o(1))e^{-jp(1+o(1))}})(1+mjpe^{-jp(1+o(1))}).\end{aligned}$$

It follows from our bounds on j and p that $mjpe^{-jp}$ tends to infinity, so the second half dominates the last term of the above sum and we have

$$p_j \le (mep\sqrt{\ln n})^j(e^{-mjp(1+o(1))e^{-jp(1+o(1))}})(2mjpe^{-jp(1+o(1))}).$$

Taking logs and using $\delta n \leq m \leq n$ gives:

$$\begin{aligned}\ln(p_j) &\leq (1+o(1))j(\ln(enp\sqrt{\ln n}) - \delta np e^{-jp(1+o(1))} - p + \frac{\ln(2njp)}{j}) \\ &\leq (1+o(1))j(4\ln(np) - \delta np e^{-kp(1+o(1))}) \\ &= (1+o(1))j(4\ln(np) - \frac{\delta np}{(\ln n)^{\frac{1}{2}+o(1)}}).\end{aligned}$$

Since $np > \frac{\ln n}{s}$, taking n large gives that the probability of failure for any particular j in this range is $o(1/n^4)$, and adding up over all j and m gives that the probability of a failure in this range is $o(1/n^2)$.

Case 2: $1 \leq j \leq \frac{1}{p\sqrt{\ln n}}$.

Let b be the number of vertices outside S adjacent to at least one vertex in S, and let a be the number of edges between S and the vertices of G outside S. For $j \geq s$ let E_j be the event that the graph satisfies conditions C_1 through C_3 but some fixed set S of j vertices fails to be nice.

By Bayes' theorem, we have

$$p_j \leq \binom{m}{j} \sum_{w=0}^{nj} \mathbf{P}(E_j | a = w)\mathbf{P}(a = w).$$

We bound the terms in two subcases depending on the size of w relative to j.

Case2a: $w < 10j$. The claim here is that in this range it is impossible for E_j to occur.

Let G^2 denote a graph with on the same vertex set as G, but with i connected to j in G^2 if and only if i and j are of distance at most 2 in G.

As in Lemma 4.2.7, if G is locally sparse then S must have at most $\frac{18j}{\ln\ln n}$ vertices of degree at most $\ln\ln n$. Condition W1 now implies that at most $\frac{18sj}{\ln\ln n}$ vertices of S can lie in the same component of the induced subgraph of G^2 on S as a vertex of degree at least $\ln\ln n$.

Thus the induced subgraph of G^2 must contain a component S_1 which does not contain a vertex having degree in G at least $\ln\ln n$. S_1 must have size at most $s-1$ by condition W1. Note that this component shares no neighbors in G with the rest of S.

Suppose that every vertex in G adjacent to S_1 had two neighbors in S_1. It would

then follow that the induced subgraph of G on $S_1 \cup N(S_1)$ contained at least

$$2|N(S_1)| - |S_1 \cap N(S_1)|$$

edges. On the other hand, condition W2 implies that this induced subgraph is a forest, so has at most

$$|S_1 \cup N(S_1)| - 1 = |S_1| + |N(S_1)| - |S_1 \cap N(S_1)| - 1$$

edges. Combining these two inequalities would yield $|N(S_1)| \leq |S| - 1$.

Thus either S_1 is a nonexpanding subset of S of size at most $s - 1$, or there is a vertex adjacent to exactly one vertex in S_1 (and thus in S). This fulfills the second requirement for G to be good.

If $|S| \geq s$, then we can find a second component S_2 satisfying the same conditions as S_1. Unless either S_1 or S_2 fails to expand, there will be a vertex in G adjacent to exactly one vertex in S_1 and another vertex adjacent to exactly one vertex in S_2, so S must be nice. Thus the first requirement for G to be good is also satisfied.

Case 2b: $w \geq 10j$. If S is not nice, then at least $b - 1$ of the neighbors of S must be adjacent to at least two vertices in S. This implies that $b \leq \frac{a+1}{2}$. It follows that we can bound $\mathbf{P}(E_j)$ by $\mathbf{P}(b \leq \frac{w+1}{2} | a = w)$. To do this, we fix a set of $\frac{w+1}{2}$ vertices and bound the probability that w vertices randomly selected from $V(G) \backslash S$ are in that set. Using the union bound over all possible sets of $\frac{w+1}{2}$ vertices, we obtain

$$\begin{aligned}
\mathbf{P}(b \leq \frac{w+1}{2} | a = w) &\leq \binom{m-j}{\frac{w+1}{2}} (\frac{w+1}{2(m-j)})^w \\
&\leq (\frac{2e(m-j)}{w-1})^{\frac{w+1}{2}} (\frac{w+1}{2(m-j)})^w \\
&\leq (\frac{4w}{m})^{\frac{w-1}{2}}.
\end{aligned}$$

This bound is decreasing in w for the entire range under consideration (our bounds on j guarantee w is at most $\frac{10n}{\sqrt{\ln n}}$). Therefore, we can bound $\mathbf{P}(q_j | a = w)$ by the probability given $a = 10j$, giving

$$\begin{aligned}
p_j &\leq 3\sqrt{n} \binom{m}{j} (\frac{40j}{m})^{5j} \\
&\leq 3\sqrt{n} (\frac{me}{j})^j (\frac{40j}{m})^{5j} \\
&\leq 3\sqrt{n} (\frac{130j}{n'})^{4j}.
\end{aligned}$$

This bound is decreasing in j in the range under consideration, and substituting $j = s$ gives that $p_j = o(1/n^4)$ for each j and m in the range. By the union bound the probability of failure in this range is $o(1/n^2)$.

4.8 Proofs of Lemmas 4.2.12 and 4.2.13

4.8.1 Proof of Lemma 4.2.12

Let A be a fixed nice matrix which is not $(s-1)$-saturated. Since A is not saturated, there is a vector $v := (v_1, v_2, \ldots v_m)^T$ in the nullspace of A whose support does not contain the coordinates corresponding to any nonexpanding subset of at most $s-1$ rows. Let D be the number of nonzero coordinates in v. If D were at most k, then the goodness of A would guarantee that the support of v would be nearly nice, meaning that some vertex i had only one neighbor in the support of v. But this is a contradiction, as the product of row i with v would then be v_i, which is nonzero by assumption since i is in the support of v.

We may therefore assume that $D > k$ and, without loss of generality, that it is the first D coordinates of v which are nonzero. We now consider the linear equation

$$\sum_{i=1}^{D} v_i x_i = 0. \tag{4.7}$$

If the new column x does not satisfy this equation, then augmenting A by this column will increase the rank by 1 and (by the symmetry of A) augmenting A simultaneously by the x and its transpose will increase the rank by 2. Therefore it suffices to bound the probability that (4.7) is satisfied. Although the x_i aren't all random in our case (our conditioning on the normality of (G_m, G_{m+1}) guarantees that all variables corresponding to vertices in non-nice subsets of G are 0), most of them will be random. Since we are assuming G to be good, the number of non-random x_i is bounded above by the number of vertices of degree at most s, which is in turn bounded above by $\frac{1}{p \ln n} = o(D)$.

After removing the x_i which are forced to be 0 we are left with $D(1 + o(1))$ independent variables with nonzero coefficients. By Theorem 2.1.5, the probability (4.7) is

satisfied is therefore at most $O((Dp)^{-1/2}) = O((kp)^{-1/2})$.

4.8.2 Proof of Lemma 4.2.13

Just as in the proof of Lemma 4.2.12, the goal will be to eventually use a Littlewood-Offord Lemma to show that a certain expression is almost surely not equal to 0. In this case, the expression in question will be the determinant of an appropriately chosen submatrix of the augmentation of A. Before we apply it, however, we first attempt to obtain additional information on the structure of A.

Let T be the union of all minimal nonexpanding subsets which contain at most $s-1$ vertices. We begin by proving two lemmas on the structure of T under the assumption that G is well-separated.

Lemma 4.8.3 *Let T be defined as above. Then $N(T) \cap T = \emptyset$.*

Proof Assume to the contrary that there are two vertices v and w in T which are adjacent in G. Let T_v and T_w be minimal nonexpanding subsets of G such that T_v contains v, T_w contains w, and $|T_v|$ and $|T_w|$ are both at most $s-1$ (T_v and T_w may in fact be equal here). By the minimality of T_v, we have $|N(T_v - v)| \geq |T_v| - 1 \geq |N(T_v)|$, so it follows that w must also be in $N(T_v - v)$. A similar argument shows that any vertex in T_v with a neighbor in T_w has at least one other neighbor in T_w, and vice versa.

We now consider the graph on vertex set $T_v \cup T_w$ whose edges correspond to edges in G with one vertex lying in T_v and the other vertex in T_w (one or both of these vertices may be in $T_v \cap T_w$). By the above argument, no vertex in this graph has degree exactly 1. However, the well-separation condition W2 guarantees that this graph is a forest (since any vertex in T has degree at most $s-1$ in G), which implies that the graph must in fact be empty. ■

Lemma 4.8.4 *Let T be defined as above. Then there is a $T_1 \subset T$ with $|T_1| = |N(T)|$ such that G has exactly one matching between T_1 and $N(T)$.*

Proof Consider the graph on $T \cup N(T)$ which includes all edges with at least one endpoint in T. We perform the following algorithm to construct T_1: At each step we

pick an unused vertex in T with exactly one unused neighbor in $N(T)$. We then add that vertex to T_1 and consider both it and its neighbor as used.

Assuming this algorithm eventually matches every vertex in $N(T)$ with a vertex in T, we are done, since the uniqueness of the matching is clear from our matching process. Showing the algorithm does not terminate prematurely is equivalent to showing that after each step either every vertex in $N(T)$ is used or there is an unused vertex in T with exactly one unused neighbor in $N(T)$.

To do this, we first note that any unused vertex in $N(T)$ has at least two unused neighbors in T (the argument in the previous lemma shows that it has at least two neighbors in T, and by construction our algorithm marks a vertex in $N(T)$ as used as soon as its first neighbor in T is used). Furthermore, by well separation the induced subgraph on the unused vertices of T and $N(T)$ is a forest, which is nonempty unless all vertices of $N(T)$ have been used. It therefore must have a vertex of degree one, which must be in T since every unused vertex in $N(T)$ has degree at least two. This allows us to continue the algorithm until all of $N(T)$ is matched. ■

Without loss of generality we can now view $A(G)$ as the block matrix below

$$\begin{pmatrix} A(G\backslash(T\cup N(T))) & A(G\backslash(T\cup N(T)),N(T)) & 0 & 0 \\ A(N(T),G\backslash(T\cup T_1)) & A(N(T)) & A(N(T),T_1) & A(N(T),T\backslash T_1) \\ 0 & A(T_1,N(T)) & 0 & 0 \\ 0 & A(T\backslash T_1,N(T)) & 0 & 0 \end{pmatrix},$$

where $A(G,H)$ denotes the adjacency matrix of the induced bipartite subgraph between G and H. The blocks in the lower right are 0 because of Lemma 4.8.3. By construction the fourth row of blocks is contained in the span of the third row.

Let B be the matrix formed by the first three rows and columns of blocks of A (note that rank (B) = rank (A)). To prove Lemma 4.2.13 it suffices to show that augmentation will almost surely increase the rank of B. Our assumption of normality guarantees we can think of the augmentation as

$$B' = \begin{pmatrix} A(G\backslash(T \cup N(T))) & A(G\backslash(T \cup N(T)), N(T)) & 0 & x \\ A(N(T), G\backslash(T \cup T_1)) & A(N(T)) & A(N(T), T_1) & y \\ 0 & A(T_1, N(T)) & 0 & 0 \\ x^T & y^T & 0 & 0 \end{pmatrix},$$

where x and y are random vectors each of whose entries are 1 with probability p and 0 otherwise. We now expand $\det(B')$ by minors simultaneously along all rows and columns in the third row and column of blocks. By Lemma 4.8.4, only one nonzero term remains, so we are left with

$$\det(B') = \pm \det \begin{pmatrix} A(G\backslash(T \cup N(T))) & x \\ x^T & 0 \end{pmatrix} = \pm \sum_{i=1}^{m} \sum_{j=1}^{m} A(i,j) x_i x_j,$$

where $A(i,j)$ denotes the (i,j) cofactor of $A(G\backslash(T \cup N(T)))$.

This is the expression that we are aiming to show is almost surely non-zero. Our goal will be to show that probably enough of these cofactors are nonzero that we can apply Lemma 2.2.6 to say that the determinant of B' is probably not zero. To do so, we first establish some properties of the matrix $C := A(G\backslash(T \cup N(T)))$.

Lemma 4.8.5 *C is nonsingular.*

Proof Any subset of T_1 must be expanding due to the matching between T_1 and $N(T)$. This implies that the first three rows of blocks of A cannot contain any nonexpanding subsets of size at most $s-1$ (such a subset would have to be in T by the definition of T, but could not be entirely within T_1 since T_1 expands. Since A is $(s-1)-$saturated, it follows that the rank of A is at least $n - |(T\backslash T_1)|$.

On the other hand, we know the third row of blocks is independent (the uniqueness of the matching in Lemma 4.8.4 implies $A(T_1, N(T))$ has determinant ± 1) and contains the fourth row of blocks in its span. Since this already accounts for the entire nullspace of A, the first three rows of blocks of A must be independent.

This implies we can perform row reduction to eliminate the $A(G\backslash(T \cup N(T)), N(T))$ block of A, and that the rows of the reduced matrix (including the rows of C) are still independent. ∎

Lemma 4.8.6 *$G\backslash(T \cup N(T))$ is "almost good" in the following sense:*

(1) Every minimal non-nice subset of the vertices of $G\backslash(T \cup N(T))$ has size either at most $s-1$ or at least $k - \frac{1}{p \ln n}$.

(2) Every minimal non-nearly nice subset of the vertices of $G\backslash(T \cup N(T))$ has size at least $k - \frac{1}{p \ln n}$.

(3) At most $\frac{1}{p \ln n}$ vertices of $G\backslash(T \cup N(T))$ have degree less than s.

Proof Let S be a subset of the vertices of $G\backslash(T \cup N(T))$ of size at most $k - \frac{1}{p \ln n}$, and let S_1 denote those vertices in $N(T)$ which have exactly one neighbor in S. We now perform the following algorithm: So long as S_1 remains nonempty, we choose a vertex v in S_1, choose a vertex of T adjacent to v which is not already in s, add that vertex in T to S, and update S_1 accordingly. Since each vertex in $N(T)$ has at least two neighbors in T, we will always be able to continue this process so long as S_1 remains nonempty. In particular, the process must terminate by the time we have added all the vertices in T to S.

Let S' be the set which results once the algorithm terminates. We first note that S' can have size at most k (the vertices in T which we add to S always have degree at most $s-1$, and the number of such vertices is bounded by our assumption that G is good). Furthermore, there is a natural matching between $S' \cap T$ and $N(T)$ given by matching each vertex of $S' \cap T$ with the v in $N(T)$ which caused it to be added to S'. This implies that $S' \cap T$ (and thus S') does not contain any nonexpanding subset of size at most $s-1$. Since G is good, this implies that S' must be nearly-nice, meaning there is some w with only one neighbor in S'. This w can't be in $N(T)$ by construction, and it can't be in T since S' contains no vertices from $N(T)$. It follows that w's neighbor must be in S, so S is also nearly-nice.

A similar argument gives that all minimal non-nice subsets in $G\backslash(T \cup N(T))$ either have size at least $k - \frac{p}{\ln n}$ or at most $s-1$. To show there aren't many vertices of degree at most s in $G\backslash(T \cup N(T))$, we first note that by condition W_1 of well separation a given vertex can only have $s-1$ neighbors in $N(T)$. It follows that any vertex of degree at most s in $G\backslash(T \cup N(T))$ had degree at most $2s$ in G, and this can be bounded by

the same argument as in (4.4). ■

Since C has full rank, dropping any of the columns of C will lead to a $m \times m-1$ matrix whose rows admit (up to scaling) precisely one nontrivial linear combination equal to 0. If any of the rows in that combination are dropped, we will be left with an $m-1 \times m-1$ nonsingular matrix, i.e. a nonzero cofactor.

As in Lemma 4.2.12, the rows with nonzero coefficients in this linear combination must form a non-nearly nice subset of the rows of the column deleted matrix of C. By Lemma 4.8.6 the only way that the combination can involve fewer than $k - \frac{1}{p \ln n}$ rows is if the rows involved formed a non-nice subset of $G \backslash (T \cup N(T))$, one of whose neighbors was the removed column. We can upper bound the number of columns whose removal could possibly cause this difficulty by the number of vertices in $G \backslash (T \cup N(T))$ that have at least one neighbor with degree at most s in $G \backslash (T \cup N(T))$, which by Lemma 4.8.6 is $O(\frac{s}{p \ln n}) = o(n)$.

Dropping any other column will lead to many nonzero cofactors, so we can apply the Quadratic Littlewood Offord Lemma with $q = k - \frac{1}{p \ln n} = k(1 - o(1))$ to bound the probability that the determinant is 0, proving Lemma 4.2.13.

4.9 Proofs of Theorem 4.1.3 and 4.1.1

Proof of Theorem 4.1.3: What we will show here is that every $(s-1)-$saturated good matrix satisfies the conclusion of the theorem. This is sufficient since by Theorem 4.2.1 and 4.2.10 the graph will be both $(s-1)-$saturated and good with probability $1 - O(\ln \ln n^{-1/4})$. We will prove this result by contradiction.

Suppose that some subset W of the rows of Q_n is dependent but does not contain a nonexpanding subset of size at most $s-1$. Since G is good, it much be true that $|W| > \frac{\ln \ln n}{p}$.

Since G is $(s-1)-$saturated, there must be a subset $S_0 \in W$ which is both independent and maximal subject to not containing any nonexpanding sets of size at most $s-1$. In particular, any row in W which is not in S_0 would create a nonexpanding set when added to S_0. Since nonexpanding sets are dependent, any row in $W \backslash S_0$ can be

written as a linear combination of at most $s-1$ rows of S_0.

Now by assumption the rows of W satisfy some linear relationship

$$\sum_{i \in W} a_i v_i = 0. \tag{4.8}$$

For each row which is in W but not in S_0, we substitute the corresponding linear combination of at most $s-1$ rows in S_0 which equals it into (4.8). This yields a linear relationship between the rows of the independent set S_0, which must therefore have zero coefficients. There were initially at least $\frac{\ln \ln n}{p}$ nonzero coefficients in (4.8), and each substitution can change at most s of them to zero. It follows that

$$|W \backslash S_0| \geq \frac{\ln \ln n}{sp}.$$

Each vertex in $W \backslash S_0$ is part of a nonexpanding set of size at most $s-1$, and thus has degree at most $s-2$. However, G is by assumption good, so has at most $\frac{1}{p \ln n}$ vertices of degree this small. This is a contradiction, so Theorem 4.1.3 is proved.

Proof of Theorem 4.1.1: Again, we will show that any $(s-1)$-saturated, good matrix satisfies the conclusion of the theorem.

On one hand, it is clear that, for any S, the expression $n - |S| + |N(S)|$ is an upper bound for the rank of $A(G)$. Thus it suffices to exhibit some set S for which that expression is at most as large to the rank. To do so, we return to the block decomposition of the proof of Lemma 4.2.13. Note that by the proof of Lemma 4.8.5 the rows of the first three blocks of A in this decomposition are independent, so we have

$$\text{rank } (A) \geq n - |T \backslash T_1|.$$

Conversely, if we take $S = T$ then we have

$$n - |S| + |N(S)| = n - |T| + |T_1| = n - |T \backslash T_1|,$$

where the first equality comes from the matching between $N(S)$ and T_1. Combining these two equations yields the desired result.

4.10 Further Results on the Nullspace of $Q(n,p)$

In this section we will use Theorem 4.1.3 and Lemma 4.2.10 to give a description of the nullspace of $Q(n,p)$ which holds for almost every graph. We begin by presenting a series of facts about the minimal dependent subsets in any well-separated graph which satisfies the conclusions of Theorem 4.1.3

Lemma 4.10.1 *Let $s > 0$ be fixed. G be a graph which is both well-separated and satisfies the conclusions of 4.1.3, and let S be a minimal dependent subset of $V(G)$. Then for sufficiently large $|G|$,*

(1) $N(S)$ is disjoint from S

(2) $|N(S)| = |S| - 1$

(3) The induced subgraph of G on $S \cup N(S)$ is a tree.

(4) Each vertex in $N(S)$ is adjacent to exactly 2 vertices in S.

Proof Condition (1) is just a restatement of Lemma 4.8.3. For (2), we know from the dependency of S and the conclusions of Theorem 4.1.3 that S contains a nonexpanding subset. By the minimality of S, that subset must be S itself, so $|N(S)| < |S|$. On the other hand, if $|N(S)|$ were at most $|S| - 2$, then removing a vertex from S would lead to a smaller nonexpanding set which would therefore be dependent.

Since S is by assumption nonexpanding and of size at most s, each vertex of S must have degree at most s, so it follows from W2 that the induced subgraph $S \cup N(S)$ is a forest. If it were disconnected, then one of the components would correspond to a non-expanding subset of S, which would contradict minimality.

We know from (2) and (3) that the induced graph on $S \cup N(S)$ has $2|S| - 2$ edges, and that the average number of neighbors in S of a vertex in $N(S)$ is therefore at most 2. If any vertex in $N(S)$ has exactly one neighbor in S, then we could remove that neighbor from S and get a smaller non-expanding subset, contradicting minimality. It follows that each vertex in $N(S)$ must therefore have exactly 2 neighbors in S. ■

It follows from Lemma 4.10.1 that for any minimal nonexpanding set S with $|S| = r$ there is a corresponding tree H on r vertices such that each vertex in H corresponds

to a vertex in S, and two vertices i and j in the tree are adjacent if and only if their corresponding vertices in S share a common neighbor in G. We refer to such a correspondence as a **neighborhood embedding** of H in G. It can easily be checked that if $p < \frac{(1-\epsilon)\ln n}{rn}$, then almost surely every such tree on r vertices will have a neighborhood embedding in $G(n,p)$.

For any tree H, let $f : H \to \pm 1$ be the unique (up to sign) two-coloring of H. For any S which is a neighborhood embedding of H, the vector v_S which is $f(H)$ on S and 0 on all other vertices of G is in the nullspace of $A(G)$, as each vertex in $N(S)$ corresponds to an edge adjacent to exactly one vertex of each color. Conversely, this v_S must be (up to scaling) the only vector in the nullspace whose support is equal to S, due to the minimality of S. Combining these observations with 4.2.5 and 4.1.3, we have therefore established

Theorem 4.10.2 *Let $\frac{c\ln n}{n} < p < \frac{1}{2}$, with $c > \frac{1}{s}$. Then almost surely the every minimally dependent collection of rows in $Q(n,p)$ corresponds to a neighborhood embedding of a tree on at most $s-1$ vertices, and the corresponding v_S will almost surely span the nullspace of $Q(n,p)$.*

4.11 A Few Further Conjectures and Avenues for Research

While the results here give a description of the behavior of the rank of $G(n,p)$ for p down to and close to $\frac{\ln n}{n}$, there remain several open questions for other values of p and other models of random graphs.

- Is there a characterization of the dependent sets similar to Theorem 4.1.3 in the case where $p = o(\frac{\ln n}{n})$? It seems likely that Theorem 4.1.3 can be extended to an s which grows sufficiently slowly with n. However, that theorem is no longer true in the case $s = \Theta(\frac{1}{n})$, as $G(n, \frac{c}{n})$ will with positive probability contain an isolated 4-cycle, which would be an expanding dependent set.

- An argument similar to Lemma 4.2.2 shows that $Q(n,p)$ will almost surely be of asymptotically full rank (having rank equal to $n(1-o(1))$ as long as $np \to \infty$.

On the other hand, $G(n, \frac{c}{n})$ almost surely has a positive fraction of its vertices isolated, so will not have close to full rank. It seems likely that

$$f(y) := \lim_{n\to\infty} \frac{\mathbf{E}(\text{rank } (Q(n, \frac{c}{n}))}{n}$$

exists, and it would be of interest to determine this limit. Note that since the rank changes by at most two in each step of the vertex exposure process, Azuma's Inequality immediately implies that the rank is highly concentrated around this expectation, wherever it may be.

- Conversely, what is the behavior of the rank of $G(n, 1-p)$ as p approaches 0? We know from Theorem 3.1.1 that the matrix will almost surely be of full rank for $n^{-1/2+\epsilon} < p < \frac{1}{2}$, but the case of smaller p seems less clear. If we assume that our graph contains no self-loops, then one natural obstruction to non-singularity is the presence of an isolated edge in the complement of G (which would correspond to a pair of identical rows in G). It seems likely that this is the primary obstruction, and that we thus have

  Conjecture 4.11.1 *$\frac{\ln n}{2n}$ is a sharp threshold for the non-singularity of $G(n, 1-p)$.*

  By this we mean that if $p = \frac{c \ln n}{2n}$ with $c > 1$, then $G(n, 1-p)$ will almost surely have full rank (as noted above, this will almost surely not be the case if $c < 1$).

  If we allowed self-loops (by treating the diagonal entries as being identically distributed to those above the diagonal), then the conjectured threshold would be $\frac{\ln n}{n}$, corresponding to the presence of multiple rows in Q_n which are entirely 1.

- Let $d > 0$ be fixed, and let $G_{n,d}$ be a graph uniformly chosen from the collection of $d-$regular graphs on n vertices. Let $Q_{n,d}$ be the corresponding random adjacency matrix. It is known that for $d \geq 3$ that $G_{n,d}$ will almost surely be a good expander, at least locally. Since in the case of $G(n, p)$ non-expansion was the main source of singularity, this seems to suggest:

  Conjecture 4.11.2 *For any $d \geq 3$, $Q_{n,d}$ is almost surely nonsingular.*

This conjecture is false in the case $d = 2$, as in that case $G_{n,d}$ will almost surely have components which are cycles of length $4k$ for some k [7, 35]. Direct computation gives that any graph containing such a component has singular adjacency matrix.

References

[1] N. Alon and J. Spencer, The Probabilistic Method, John Wiley & Sons Inc., 2000.

[2] Bai, Z.D. Circular Law, *Ann. Prob.* **25** (1997) 494-529

[3] M. Bauer and O. Golinelli, Core percolation in random graphs: a critical phenomena analysis, *Eur. Phys. J. B* **24** (2001), 339-352

[4] M. Bauer and O. Golinelli, Exactly solvable model with two conductor-insulator transitions driven by impurities, *Phys. Rev. Lett* **86** (2001), 2621-2624

[5] M. Bauer and O. Golinelli, On the kernel of tree incidence matrices, *Journal of Integer Sequences* **3** (2000), no. 1, article 00.1.4

[6] B. Bollobás, Random Graphs, Academic Press, New York, 1985.

[7] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. combin* **1** (1980), 311-316

[8] B. Bollobás and A. Thomason, Threshold Functions, *Combinatorica* **7** (1987), 35-38

[9] K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely non-singular, *Duke Math J.* **135** (2006), no. 2, 395-413

[10] K. Costello and V. Vu, The rank of random graphs, *submitted*

[11] A. Edelman, Eigenvalues and condition numbers of random matrices, *Siam J. Matrix Anal. Appl* **9** (1988), no. 4, 543-560.

[12] P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898-902.

[13] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci* **5** (1960), 17-61

[14] C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets, *Comm. Math. Phys* **22** (1971), no. 2, 89-103

[15] Z. Füredi and J. Komlós, The Eigenvalues of Random Symmetric Matrices, *Combinatorica*, **1** (1981), no. 3, 233-241

[16] E. N. Gilbert, Random Graphs, *Annals of Mathematical Statistics* **30** (1959), 1141-1144

[17] V. L. Girko, Circular Law *Theory Probab. Appl.* **29** (1984) 694-706

[18] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, *Period. Math. Hungar.* **8** (1977), no. 3-4, 197-211.

[19] J. Kahn, J. Komlós, E. Szemerédi, On the probability a random ±1 matrix is singular, *J. Amer. Math Soc.* **8** (1995), 223-240

[20] D. Kleitman, On a lemma of Littlewood and Offord on the distribution of certain sums, *Math. Z.* **90** (1965), 251-259

[21] J. Komlós, On the determinant of (0,1) matrices, *Studia Sci. Math. Hungar.* **2** (1967), 7-22.

[22] J. Komlós, On the determinant of random matrices, *Studia Sci. Math. Hungar.* **3** (1968), 387–399.

[23] J. Komlós and E. Szemerédi, Limit Distributions for the Existence of Hamiltonian Circuits in a Random Graph, *Discrete Math.* **43** (1983), 55-63.

[24] M. Krivelevich and B. Sudakov, The largest eigenvalue of sparse random graphs, *Combin. Probab. Comput.* **12** (2003), no. 1, 61-72

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

[26] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons Inc., 1982

[27] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and the condition number of random matrices, *preprint*

[28] A. Sárközy and E. Szemerédi, Über ein Problem von Erdös und Moser, *Acta Arith.* **11** (1965), 205-208

[29] A. Sidorenko, A correlation inequality for bipartite graphs, *Graphs Combin.* **9** (1991), no. 2, 201-204

[30] T. Tao and V. Vu, On random ±1 matrices: Singularity and Determinant, *Random Structures and Algorithms* **28** (2006), 1-23

[31] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, *J. Amer. Math. Soc* **20** (2007), 603-628

[32] T. Tao and V. Vu, Additive Combinatorics, Cambridge Studies in Advanced Math **105**, Cambridge University Press, Cambridge, 2006

[33] T. Tao and V. Vu, Random Matrices, The Circular Law, *preprint*

[34] E. Wigner, On the distribution of the roots of certain symmetric matrices, *Ann. Math.* (2) **67** (1958), 325-327

[35] N. Wormald, The asymptotic distribution of short cycles in random regular graphs, *J. Combin. Theory Ser. B* **31** (1981), 168-182

Vita

Kevin Costello

2006-2007 Ph. D. in Mathematics, Rutgers University

2003-2006 University of California, San Diego

1999-2003 B. Sc. in Mathematics and Economics with Honors from California Institute of Technology

1999 Graduated from Illinois Mathematics and Science Academy

2006-2007 Teaching assistant, Department of Mathematics, Rutgers University

2004-2006 Teaching assistant, Department of Mathematics, University of California, San Diego

2003-2006 NSF Graduate Research Fellow, Department of Mathematics, University of California San Diego