Combinatorics, Probability and Computing (2010) 19, 321–342. © Cambridge University Press 2009
doi:10.1017/S0963548309990447

On the Rank of Random Sparse Matrices

KEVIN P. COSTELLO[1†] and VAN VU[2‡]

[1]Department of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA
(e-mail kcostell@math.gatech.edu)

[2]Department of Mathematics, Rutgers, Piscataway, NJ 08854, USA
(e-mail vanvu@math.rutgers.edu)

Received 28 February 2008; revised 31 July 2009; first published online 12 October 2009

We investigate the rank of random (symmetric) sparse matrices. Our main finding is that with high probability, any dependency that occurs in such a matrix is formed by a set of few rows that contains an overwhelming number of zeros. This allows us to obtain an exact estimate for the co-rank.

1. Introduction and statement of main results

Let $w_{ij}, 1 \leqslant i \leqslant j \leqslant n$ be complex numbers, where the w_{ij} are non-zero for $i < j$ (we place no restriction on the w_{ii}). We denote by W the symmetric matrix where:

- the above-diagonal entries are $w_{ij}, 1 \leqslant i < j \leqslant n$,
- the diagonal entries are w_{ii},
- $w_{ji} = w_{ij}$ for all $1 \leqslant i < j \leqslant n$.

Let p (which may depend on n) be a positive number at most one, and let ξ_{ij}, $1 \leqslant i \leqslant j \leqslant n$ be independent Bernoulli random variables taking on value 1 with probability p and 0 with probability $1 - p$.

We sparsify the matrix W by replacing each entry above or on the diagonal by zero independently with probability $1 - p$ while keeping the matrix symmetric. The resulting symmetric matrix, denoted by $Q := Q(W, p)$, has entries $q_{ij} = w_{ij}\xi_{ij}, 1 \leqslant i \leqslant j \leqslant n$ and $q_{ji} = q_{ij}$.

Q is a very general model of random matrices and the study of its linear algebraic parameters is of considerable interest. In this paper, we will focus on the rank of Q. This seemingly simple parameter was not understood until very recently, when in [1] T. Tao

† Supported in this research by NSF grant DMS-0635607.
‡ Supported by NSF grant DMS-0901216 and DOD grant AFOSAR-FA-9550-09-1-0167.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

and the authors proved (in a slightly less general form) that if $0 < p < 1$ is a constant not depending on n, then Q almost surely has full rank, thus proving a conjecture of Weiss. A refinement of this result in [2] showed that the same conclusion still holds as far as $p > \frac{(1+\epsilon)\ln n}{n}$. This bound is sharp, as for $p < \frac{(1-\epsilon)\ln n}{n}$, by the coupon-collector theorem, there will be an all-zero row with probability tending to 1 as n tends to infinity.

In the current paper, we investigate the case when p can be significantly smaller than $\frac{\ln n}{n}$. In this case, the co-rank of the matrix is almost surely positive (in fact, polynomial in n) and this makes the situation quite different from the full-rank case considered before.

The main intuition that underlies many problems concerning the rank of a random matrix is that:

Dependency should come from small configurations.

The most famous question based on this intuition is the $(1/2)^n$ conjecture, which asserts that the probability that a random Bernoulli matrix of order n (a non-symmetric matrix with entries being i.i.d. Bernoulli random variables, taking values 1 and -1 with probability $1/2$) is singular is $(1/2 + o(1))^n$. The lower bound is trivial, as it comes from the probability that there are two equal rows. Despite several efforts (see [6] for a recent survey), this conjecture is still open. Here and later, the asymptotic notation will be used under the assumption that n, the size of the matrix, tends to infinity.

In this paper, we are able, for the first time, to verify this intuition rigorously in a general setting. We will show that for a certain range of p, any dependency that occurs in the random matrix indeed comes from a small configuration. To be more precise, we will prove that if $p = \Omega(\frac{\ln n}{n})$, then any dependency comes from a collection of at most k rows, for some $k = O(1)$, which between them have at most $k - 1$ non-zero columns. In particular, it will follow that the co-rank of Q depends only on the local structure of an underlying graph formed by the ξ_{ij}, regardless of the values of the w_{ij}. (It is, however, critical that $w_{ij} \neq 0$ for $i \neq j$.)

In order to state our results formally, we will need a few definitions.

Definition. The graph $G(Q)$ of a symmetric matrix Q is the graph whose vertices are the rows of Q, with vertex i connected to vertex j if and only if $q_{ij} \neq 0$.

Note that this graph may contain self-loops if the entries on the main diagonal of Q are non-zero. Furthermore, except for these loops the graph of $Q(W, p)$ is independent of W due to our assumption that $w_{ij} \neq 0$.

For a set S in a graph G, let $N(S)$ denote the neighbourhood 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 *non-expanding* if it satisfies $|N(S)| < |S|$. S is *minimally non-expanding* if it contains no proper non-expanding subset.

Examples. An isolated vertex forms a non-expanding set. A set of two vertices of degree one sharing a common neighbour forms a minimal non-expanding set of size 2. However, a single vertex with a self-loop does not form a non-expanding set (as it is its own neighbourhood).

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

We call a set T of rows of a matrix Q non-expanding if at most $|T|-1$ columns of Q have at least one non-zero entry in T. Note that a set of rows of Q is non-expanding if and only if the corresponding set of vertices in $G(Q)$ is non-expanding. Furthermore, if S is a non-expanding subset of $G(Q)$, then the corresponding rows in Q are dependent (regardless of W), since they span a space of dimension at most $|N(S)|$. It is thus clear that, for any set S of vertices of $G(Q)$ and any weights W,

$$\operatorname{rank}(Q) \leqslant n - |S| + |N(S)|.$$

By considering all sets of vertices in $G(Q)$, we immediately obtain the upper bound

$$\operatorname{rank}(Q) \leqslant \min_{S \subset V(G(Q))} n - |S| + |N(S)|.$$

Our first main result shows that for p which is sufficiently large, this upper bound is *tight*.

Theorem 1.1. *Assume that* $p = \Omega(\frac{\ln n}{n}) < \frac{1}{2}$. *Let* W *be any fixed symmetric matrix whose off-diagonal entries are non-zero. Then, almost surely,*

$$\operatorname{rank}(Q(W,p)) = \min_{S \subset V(G(Q))} (n - |S| + |N(S)|). \tag{1.1}$$

In other words, the rank of Q is determined, almost surely, by the local structure of $G(Q)$, which in turn depends only on the ξ and which diagonal entries of W are non-zero (actually, which diagonal entries are non-zero will also turn out to be with high probability irrelevant to the rank of Q; see Remark 1).

Theorem 1.1 together with earlier results from [1, 2] imply the following corollary.

Corollary 1.2. *Let* $1/2 > p = \Omega(\frac{\ln n}{n})$, *and let* Q *be a random symmetric matrix whose above-diagonal entries are independent random variables* ξ_{ij} *satisfying*

$$\mathbb{P}(\xi_{ij} = 0) = 1 - p,$$

and whose on-diagonal entries are 0 *with probability at least* $1 - p$.

Then, almost surely,

$$\operatorname{rank}(Q) = \min_{S \subset V(G(Q))} (n - |S| + |N(S)|). \tag{1.2}$$

The ξ_{ij} here are not necessarily identically distributed or Bernoulli.

We next turn to the question of classifying the dependency in $Q(W,p)$. It is routine to check that, for any fixed positive integer s and $\epsilon > 0$, if

$$n^{\frac{1}{2s}-1+\epsilon} < p < \frac{(1-\epsilon)\ln n}{sn},$$

then $G(n,p)$ almost surely contains minimal non-expanding sets of all sizes up to and including s. In $Q(W,p)$ such a set of rows is automatically dependent, regardless of the values of the non-zero entries.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

Our second main result shows that, in the range we are considering, this is the *only* reason for dependency.

Theorem 1.3. *Let s be a fixed positive integer and let $c > \frac{1}{s}$ be a positive constant. Assume that $\frac{c \ln n}{n} < p < 1/2$. Then with probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, $Q(W,p)$ has the property that any set of dependent rows of $Q(W,p)$ contains a non-expanding set of size at most $s-1$. The constant implicit in the O notation is independent of W.*

Note that, in particular, the $s = 1$ case of this result gives that $Q(W,p)$ is almost surely non-singular for p sufficiently large. This particular case can also be extended to non-symmetric models (see Section 11). Again, this result can easily be extended to a result for random W in the vein of Corollary 1.2.

The special cases $s = 1, 2$ of this result were proved in an earlier paper [2] for $Q = Q(n,p)$. 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 non-expanding set has size at most $s-1$. It also implies the following corollary.

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

Corollary 1.4. *With probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, the rank of the random matrix $Q(W,p)$ equals the size of its largest unobstructed set in its graph.*

The proofs of the main theorems combine techniques developed in earlier papers [1, 2] with some delicate properties of sparse random graphs. We are going to sketch the main ideas in the next section.

2. The idea of the proofs and some lemmas

Instead of proving Theorems 1.1 and 1.3 directly, we are going to prove the following theorem (which is somewhat weaker than Theorem 1.3). The proof of this theorem combined with some lemmas will imply Theorems 1.1 and 1.3.

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

Theorem 2.1. *Let s be a fixed positive integer and let $c > 1/s$ be a positive constant. Assume $\frac{c \ln n}{n} < p < 1/2$. Then with probability $1 - O(\frac{1}{(\ln \ln n)^{1/4}}) = 1 - o(1)$, the random matrix $Q(W,p)$ is $(s-1)$-saturated.*

Since the case where $p = \Omega(1)$ was already addressed in [1], we will assume that $p = o(1)$.

Following the ideas from [5, 1], we are going to expose $Q(W,p)$ minor by minor. Letting Q_m denote the upper left $m \times m$ minor of $Q(W,p)$, we view Q_{m+1} as being formed by taking Q_m and augmenting by a column whose entries are chosen independently, along

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

with the column's transpose. Let $G_m = G(Q_m)$. In graph-theoretic terms, we can view the sequence of G_m as 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 $p \geqslant c \ln n/n$ 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 2.2. *For any constant $\epsilon > 0$ there exists a constant $\gamma > 0$ such that*

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

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 non-expanding subsets of at most $s-1$ vertices.

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

Definition. 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 (in G) at most $\ln \ln n$.

W2. No cycle of length 1 or between 3 and $12s$ in G contains a vertex of degree (in G) at most $\ln \ln n$.

Lemma 2.3. *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.

Remark 1. One corollary of this lemma and our main results is that whether the diagonal entries of W are non-zero is likely to be irrelevant to the collection of dependent sets in $Q(W,p)$. The minimal dependent sets will with high probability be the non-expanding sets of size at most $s-1$ and therefore correspond to vertices of degree at most $s-1$ in G. On the other hand, a graph satisfying W2 only contains self-loops at vertices of degree at least $\ln \ln n$.

Another way of thinking about this is as follows. If we form a new matrix W' by replacing all of the entries on the diagonal of W by 0, with high probability the graph of $Q(W',p)$ will contain exactly the same non-expanding sets as the graph of $Q(W,p)$ and thus (by Theorem 1.3) the same dependent sets of rows.

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

Lemma 2.4. *For any $m > n'$, the probability that G_m is well-separated but is not a small set expander is $O(n^{-4})$.*

Definition. A set S of the vertices of a graph G is *nice* if there are at least two vertices of G, each 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 neighbour in S.

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

Definition. 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 matrix Q is *good* if the graph $G(Q)$ is good.

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

Lemma 2.5. *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. 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 symmetric matrices is normal if the pair of graphs $(Q(A), Q(A'))$ is normal.

Lemma 2.6. *Let Q be any fixed, good $m \times m$ matrix which is not $(s-1)$-saturated. Then*

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

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

dependencies removed cannot correspond to small non-expanding subsets of the rows of Q_m. This implies that in some sense Q_{m+1} is a little closer to being saturated than 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 non-expanding subsets) with high probability Q_m does not gain any new dependencies by being augmented.

Lemma 2.7. *Let Q be any fixed, good $m \times m$ matrix which is also $(s-1)$-saturated. Then*

$$\mathbb{P}(\operatorname{rank}(Q_{m+1}) - \operatorname{rank}(Q_m) < 1 | (Q_m, Q_{m+1}) \text{ is normal} \wedge Q_m = Q) = O((kp)^{-1/4}).$$

ln the next section, we prove Theorem 2.1 assuming these lemmas. The proofs of the lemmas will be presented in the sections that follow.

3. Proof of Theorem 2.1

In this section, we assume all lemmas from the previous section are true. We are going to use a variant of an argument from [1]. 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 $n' \leqslant m < n$. We therefore have

$$\mathbb{P}(B_0) \leqslant \mathbb{P}(B_0 \wedge B_2 | B_1) + \mathbb{P}(\neg B_1) + \mathbb{P}(\neg B_2).$$

By Lemma 2.2 we have $\mathbb{P}(\neg B_1) = o(e^{-\gamma n \ln n})$, and by Lemma 2.5, $\mathbb{P}(\neg B_2) = O(n^{1-sc\delta+\epsilon})$. 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 $Y_m = U_m - \operatorname{rank}(Q_m)$. 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 $n' \leqslant j \leqslant m$ 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 3.1. *For any sequence $\mathcal{Q}_m = \{Q_{n'}, Q_{n'+1}, \ldots, Q_m\}$ encountered in the augmentation process,*

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

Let us (for now) assume Lemma 3.1 to be true. For $n' < m$ we then have

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

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

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

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

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

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

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

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

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

On the other hand, by definition $X_n \geqslant 4$ 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

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

proving the theorem.

It remains to prove Lemma 3.1. If a matrix in the sequence $\mathcal{Q}_m = \{Q_{n'}, Q_{n'+1}, \ldots, Q_m\}$ 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 G_m adjacent to the $(m+1)$st vertex of G_{m+1}.

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

Proof of claim. Let S_{m+1} denote an 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 neighbourhood 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. Splitting the expectation according to the normality of (Q_m, Q_{m+1}), we have

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

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

probability p. It follows that

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

Adding up over all i, we have

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

If $Y_m = 0$ and (Q_m, Q_{m+1}) is normal, then by Lemma 2.7 (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

$$\mathbb{E}(X_{m+1}|\mathcal{Q}_m) = O((\ln \ln n)^{-1/4} + (\ln n)^{-1}) = O((\ln \ln n)^{-1/4}). \tag{3.1}$$

If $Y_m = j > 0$ and (Q_m, Q_{m+1}) is normal then, with probability $1 - O((\ln \ln n)^{-1/2})$, Y_{m+1} is $j - 1$ by Lemma 2.6, and otherwise is at most $j + 1$. Combining this with the bound on $\mathbb{E}(4^{Z_m+1}\chi(Z_m > 0)|\mathcal{Q}_m)$, we have

$$\mathbb{E}(X_{m+1}|\mathcal{Q}_m) = 4^{j-1} + 4^{j+1}O((\ln \ln n)^{-1/2}) + 4^j O((\ln n)^{-1}) \leqslant \frac{3}{5}4^j. \tag{3.2}$$

The lemma now follows immediately from (3.1) and (3.2).

4. Proof of Lemma 2.2

We use a variant of an argument from [2]. By symmetry and the union bound,

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

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 $\mathbb{P}(B_1^*)$ 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 (not necessarily unique) matrix F. Choosing an arbitrary such F, B_1^* will hold only when we also have $C = B^T F$. In other words, for any fixed A and B, the assumption that B_1^* holds forces each entry in C to take on a certain value.

However, C is still random, and the probability that any given off-diagonal entry takes on its forced value is at most $1 - p$ (this being the probability that the entry is 0). The entries are not all independent (due to the symmetry of C), but those above the main

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

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

$$\begin{aligned}\mathbb{P}(\mathrm{rank}(Q_{n'}) < (1-\epsilon)n') &\leqslant \binom{n'}{\epsilon n'}((1-p)^{\frac{(\epsilon n')^2}{2}}) \\ &\leqslant \left(\frac{n'e}{\epsilon n'}\right)^{\epsilon n'} e^{\frac{-p(\epsilon n')^2}{2}} \\ &\leqslant {c_2}^n e^{-c_1 n \ln n},\end{aligned}$$

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

5. Proof of Lemma 2.3

If p is at least $(\ln n)^2/n$, then G_m will with probability at least $1 - o(1/n^3)$ have no vertices with degree at most $\ln \ln n$, in which case the lemma is trivially true. Therefore we can assume $p \leqslant (\ln n)^2/n$.

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 that $G_{n'}$ fails condition W1 by the union bound over all sets T of at most $5s$ vertices of the probability that T both forms a connected subgraph and contains at least s vertices that each have at most $\ln \ln n$ neighbours outside T.

For a fixed T the probability that any single vertex has sufficiently small degree outside T is at most

$$\sum_{i=0}^{\ln \ln n} \binom{n' - |T|}{i} p^i (1-p)^{n'-|T|-i} \leqslant (1+o(1)) \sum_{i=0}^{\ln \ln n} (n'p)^i (1-p)^{n'} \leqslant \frac{(\ln n)^{2 \ln \ln n}}{n^{c\delta}}, \tag{5.1}$$

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

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

These events are independent (as one depends only on edges within T and the other only on edges with exactly one endpoint in T), so the probability that some subset fails the first well-separation criterion is at most

$$\sum_{i=s}^{5s} \binom{n'}{i} \binom{i}{s} n^{-sc\delta+\epsilon} \frac{(\ln n)^{2i}}{(n')^{i-1}} = O(n^{1-sc\delta+\epsilon}). \tag{5.2}$$

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

Similarly, for each $1 \leqslant j \leqslant 12s$, the probability that a given set of size j contains a spanning cycle is by the union bound at most

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

and this event is independent from the event that some vertex in the set is adjacent to at most $\ln\ln n$ edges outside the set. Therefore the probability that some set fails W2 is at most

$$\sum_{i=1}^{12s}\binom{n'}{i}\frac{(i-1)!(\ln n)^{2i}}{n^i}\frac{(\ln n)^{2\ln\ln n}}{n^{c\delta}} = O(n^{-c\delta+\epsilon}). \tag{5.3}$$

Case 2: $m_0 = m > n'$. In this case we can bound the probability that $m_0 = m$ 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 (5.2) and (5.3), 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.

6. Proof of Lemma 2.4

In order to prove the edge expansion property we first show that almost surely all small subgraphs of $G(n, \frac{c\ln n}{n})$ will not have too many edges.

Definition. 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 6.1. *For fixed c the probability that $G(n, \frac{c\ln n}{n})$ is not locally sparse is $O(n^{-4})$.*

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

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

For $j < n^{1/4}$ this gives $q_j \leqslant n^{-2j}$, while for $j > n^{1/4}$ we have (using our upper bound on j) $q_j \leqslant (\ln n)^{-j/2} = o(n^{-5})$. By adding up 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 2.4, which we do in two cases depending on the value of p.

Case 1: $p \geqslant (12 \ln n)/n$. We estimate the probability that there is a non-expanding small set directly by using the union bound over all sets of size $i < n \ln^{-3/2} n$. The probability in question can be bounded from above by

$$\sum_{i=1}^{n \ln^{-3/2} n} \binom{n}{i} \binom{i(n-i)}{i-1} (1-p)^{i(n-i)-(i-1)} \leqslant \sum_{i=1}^{n \ln^{-3/2} n} n^i (en)^{i-1} e^{-inp(1+o(1))}$$
$$= \frac{1}{en} \sum_{i=1}^{n \ln^{-3/2} n} (n^2 e^{-np(1+o(1))})^i.$$

The lower bound on p guarantees that the summand is $O(n^{-(4+o(1))i})$, so the probability for any p in this range is $o(\frac{1}{n^4})$.

Case 2: $p < (12 \ln n)/n$. If G_m fails to expand edgewise, there must be a minimal subset S_0 which both fails to expand and contains no non-expanding 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_0 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 non-expanding 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 6.1 is $O(n^{-4})$.

7. Proof of Lemma 2.5

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, let C_2 be the event that G_m has maximum degree at most $5knp$ for each m, and let C_3 be the event that G_m is well-separated, locally sparse, and a small set expander for every m

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

between n' and n. We have

$$\mathbb{P}(\neg C_0) \leqslant \mathbb{P}(\neg C_0 \wedge C_1 \wedge C_2 \wedge C_3) + \mathbb{P}(\neg C_1) + \mathbb{P}(\neg C_2) + \mathbb{P}(\neg C_3).$$

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

Lemmas 2.4, 2.3, and 6.1 together show that $\mathbb{P}(\neg C_3) = O(n^{1-sc\delta+\epsilon})$.

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

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

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

Recall that we defined $k = \frac{\ln\ln n}{2p}$. Note that by our lower bound on p we have $k = o(n)$.

Case 1: $1/(p\sqrt{\ln n}) \leqslant j \leqslant k$. We will show that there are almost surely no non-nice subsets at all in this range, minimal or otherwise. We can give an upper bound on the probability that a particular set S of j vertices fails to be nice by the probability that either 0 or 1 vertices outside S have exactly 1 neighbour in S (we restrict our search to vertices outside S so we do not have to worry about the presence or absence of self-loops).

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

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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

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

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

Taking logs and using $\delta n \leqslant m \leqslant n$ gives

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

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 failure in this range is $o(1/n^2)$.

Case 2: $1 \leqslant j \leqslant 1/(p\sqrt{\ln n})$. 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 $j \geqslant s$ let E_j be the event that the graph satisfies conditions C_1 to C_3 but some fixed set S of j vertices fails to be nice.

By considering all possibilities for a, we have

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

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

Case 2a: $w < 10j$. The claim here is that in this range it is impossible for E_j to occur.

Let G^2 denote a graph 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 2.4, 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 $\frac{90sj}{\ln\ln n}$ 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 neighbours in G with the rest of S.

Suppose that every vertex in G adjacent to S_1 had two neighbours 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 $|N(S_1)| \leqslant |S| - 1$.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

Thus either S_1 is a non-expanding 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 fulfils the second requirement for G to be good.

If $|S| \geqslant s$, 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: $w \geqslant 10j$. If S is not nice, then at least $b-1$ of the neighbours of S must be adjacent to at least two vertices in S. This implies that $b \leqslant \frac{a+1}{2}$. It follows that we can bound $\mathbb{P}(E_j)$ by $\mathbb{P}(b \leqslant \frac{w+1}{2} | a = w)$. To do this, we fix a set of $\frac{w+1}{2}$ vertices and bound the probability that w consequentially randomly selected vertices were in that set. Using the union bound over all possible sets of $\frac{w+1}{2}$ vertices, we obtain

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

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 $\mathbb{P}(q_j | a = w)$ by the probability given $a = 10j$, giving

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

This bound is decreasing in j in the range under consideration, and plugging in $j = s$ 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 $o(1/n^2)$.

8. Proofs of Lemmas 2.6 and 2.7

Proof of Lemma 2.6. Let Q be a fixed nice matrix which is not $(s-1)$-saturated. Since Q is not saturated, there is a vector $v := (v_1, v_2, \ldots, v_m)^T$ in the nullspace of Q whose support does not contain the coordinates corresponding to any non-expanding subset of at most $s-1$ rows. Let D be the number of non-zero coordinates in v. If D were at most k, then the goodness of Q would guarantee that the support of v would be nearly nice, meaning that some vertex i has only one neighbour 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 by assumption not 0 since it is in the support of v.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

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

$$\sum_{i=1}^{D} v_i x_i = 0. \tag{8.1}$$

If the new column x does not satisfy this equation, then augmenting Q by this column will increase the rank by 1 and (by the symmetry of Q) augmenting Q simultaneously by x and its transpose will increase the rank by 2. Therefore it suffices to bound the probability that (8.1) is satisfied. To do so, we use the following variation of the classical Littlewood–Offord theorem [3] due to Halász [4].

Theorem 8.1. *Let χ_i be independent random variables satisfying*

$$\max_i \sup_{c \in \mathbb{R}} \mathbb{P}(\chi_i = c) \leqslant 1 - \rho.$$

Then

$$\mathbb{P}\left(\sum_{i=1}^{D} \chi_i = 0\right) = O((D\rho)^{-1/2})$$

Although the x_i are not all random in our case (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 $\frac{1}{p \ln n} = o(D)$.

Thus, after removing the x_i which are forced to be 0, we are left with $D(1 + o(1))$ independent variables with non-zero coefficients. Each $v_i x_i$ is equal to 0 with probability $1 - p$ and $v_i w_i$ with probability p. Applying Theorem 8.1 with $\rho = p$, we see that the probability (8.1) is satisfied is therefore at most $O((Dp)^{-1/2}) = O((kp)^{-1/2})$. □

Proof of Lemma 2.7. Just as in the proof of Lemma 2.6, the goal will be eventually to 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 Q. Before we apply it, however, we first attempt to obtain additional information on the structure of Q, and in particular on $G = G(Q)$.

Let T be the union of all minimal non-expanding subsets of G 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 8.2. *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 that are adjacent. Let T_v and T_w be minimal non-expanding 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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

here). By the minimality of T_v, we have $|N(T_v - v)| \geqslant |T_v| - 1 \geqslant |N(T_v)|$, 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 neighbour in T_w has at least one other neighbour 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 8.3. *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 neighbour in $N(T)$. We then add that vertex to T_1 and consider both it and its neighbour 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 that the algorithm does not terminate prematurely is equivalent to showing that after every step either every vertex in $N(T)$ is used or there is an unused vertex in T with exactly one unused neighbour in $N(T)$.

To do this, we first note that any unused vertex in $N(T)$ has at least two unused neighbours in T (the argument in the previous lemma shows that it has at least two neighbours in T, and by construction our algorithm marks a vertex in $N(T)$ as used as soon as its first neighbour 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 non-empty 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

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

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 8.2. 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 $\text{rank}(B) = \text{rank}(A)$). To prove Lemma 2.7 it suffices to show that augmentation will almost surely increase the rank of B. Our assumption of normality guarantees we can

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

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, 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 8.3, only one non-zero 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. Unlike in the proof of Lemma 2.6, however, this expression is a quadratic form in the variables x_i. Thus, instead of using the original Littlewood–Offord lemma, we will use the following quadratic version proved in [2].

Lemma 8.4. *Let a_{ij} be fixed constants such that there are at least q indices j such that, for each j, there are at least q indices i for which $a_{ij} \neq 0$. Let $z_1, z_2, \dots, z_n$ be as in Lemma 8.1. Then, for any fixed c,*

$$\mathbb{P}\left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} z_i z_j = c \right) = O((q\rho)^{-1/4}), \tag{8.2}$$

where the implied constant is absolute.

Our goal will be to show that probably enough of these cofactors are non-zero to apply this quadratic Littlewood–Offord lemma, 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 8.5. *C is non-singular.*

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 non-expanding 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 8.3 implies that $A(T_1, N(T))$ 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.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

This implies that 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 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 neighbour in S. We now perform the following algorithm. So long as S_1 remains non-empty, 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 neighbours in T, we will always be able to continue this process so long as S_1 remains non-empty. 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 non-expanding 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 neighbour in S'. This w cannot be in $N(T)$ by construction, and it cannot be in T since S' contains no vertices from $N(T)$. It follows that w's neighbour 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 are not many vertices of degree at most s in $G\backslash(T \cup N(T))$, we first note that by condition W1 of well-separation a given vertex can only have $s-1$ neighbours 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 (5.1). □

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

As in Lemma 2.6, the rows with non-zero coefficients in this linear combination must form a non-nearly nice subset of the rows of the column-deleted matrix of C. By Lemma 8.6 the only way that the combination can involve fewer than $k - \frac{1}{p \ln n}$ rows is if

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

the rows involved formed a non-nice subset of $G\backslash(T \cup N(T))$, one of whose neighbours 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 $G\backslash(T \cup N(T))$ which have at least one neighbour that has degree at most s in $G\backslash(T \cup N(T))$, which by Lemma 8.6 is $O(\frac{s}{p \ln n}) = o(n)$.

Dropping any other column will lead to many non-zero cofactors, so we can apply the quadratic Littlewood–Offord lemma with $q = k - \frac{1}{p \ln n} = k(1 - o(1))$ and $\rho = p$ to bound the probability that the determinant is 0, proving Lemma 2.7. □

9. Proofs of Theorems 1.3 and 1.1

Proof of Theorem 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 Theorems 2.1 and 2.5 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 R of the rows of Q_n is minimally dependent but does not contain a non-expanding subset of size at most $s-1$. Since G is good, it must be true that $|R| > \frac{\ln \ln n}{p}$.

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

Now, by assumption, the rows of R satisfy some linear relationship

$$\sum_{i \in W} a_i v_i = 0. \tag{9.1}$$

For each row which is in R but not in S_0, we substitute the corresponding linear combination of at most $s-1$ rows in S_0 into (9.1). This yields a linear relationship between the rows of the independent set S_0, which must therefore have all its coefficients equal to zero. There were initially at least $\frac{\ln \ln n}{p}$ non-zero coefficients in (9.1), and each substitution can change at most s of them to zero. It follows that

$$|R\backslash S_0| \geqslant \frac{\ln \ln n}{sp}.$$

Each vertex in $R\backslash S_0$ is part of a non-expanding 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 1.3 is proved. □

Proof of Theorem 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

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

proof of Lemma 2.7. Note that by the proof of Lemma 8.5 the rows of the first three blocks of A in this decomposition are independent, so we have

$$\operatorname{rank}(A) \geqslant n - |T \setminus T_1|.$$

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

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

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

10. Extensions, open problems and avenues for further research

Although the symmetric model $Q(W, p)$ is natural from a graph-theoretic viewpoint, it is also of interest to consider what happens when the matrix W is no longer symmetric. In the case $s = 1$, we can show, via an argument similar to that in this paper, that the following result holds.

Theorem 10.1. *Let W be any (symmetric or non-symmetric) matrix of weights which is non-zero off the main diagonal. Let $\frac{(1+\epsilon)\ln n}{n} < p < \frac{1}{2}$, and let ξ_{ij} be independent Bernoulli variables which are 1 with probability p and 0 otherwise. Let $A(W, p)$ be the matrix whose entries are given by $a_{ij} = w_{ij}\xi_{ij}$. Then*

$$\mathbb{P}(A(W, p) \text{ is singular}) = O((\ln \ln n)^{-1/2}).$$

It seems more difficult to obtain a workable extension of the full statement of Theorem 1.3 to non-symmetric models. For example, suppose we were to look at a matrix A whose entries were 1 with probability $\frac{0.8 \ln n}{n}$ and 0 otherwise. With high probability this matrix will have rows and columns that are entirely 0, and furthermore the number of non-zero rows and columns of A will likely not be equal. If A has more non-zero rows than columns, then that would imply some sort of dependency among the non-zero rows of A. However, there does not seem to be any obvious way to describe this dependency.

The assumption in Corollary 1.2 that the entries of A have equal probabilities of being 0 seems quite artificial, and it would be of interest to replace that assumption by the assumption that no entry is too likely to take on the value 0. More generally, we have the following conjecture.

Conjecture 10.2. *Let A be a random symmetric matrix whose above-diagonal entries a_{ij} are independent, random variables satisfying*

$$\sup_{i,j} \sup_{c \in \mathbb{C}} \mathbb{P}(a_{ij} = c) \leqslant 1 - \frac{(1+\epsilon)\ln n}{n}.$$

Then A is almost surely non-singular.

The methods of [1] give that this conjecture is true if $\frac{(1+\epsilon)\ln n}{n}$ is replaced by $n^{-1/2+\epsilon}$. Conversely, the methods in this paper and [2] cover the case where the entries are allowed

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press

to concentrate on the particular value of 0 with high (equal) probability. However, there seems to be no natural reason why the singularity probability should be less when the variables concentrate on 0 as opposed to some other value.

Alternatively, we can consider the situation where the matrix W which is being sparsified already has certain entries set equal to 0. Here the intuition is that if W does not have too many entries already set equal to 0, then the results of this paper should also hold for the sparsified matrix. This leads to the following conjecture.

Conjecture 10.3. *Let W be a symmetric matrix containing at least m non-zero entries in each row. Then, if $\frac{(1+\epsilon)\ln n}{m} < p < \frac{1}{2}$, then $Q(W,p)$ is almost surely non-singular.*

This conjecture and a short coupling argument would be sufficient to handle the case of matrices whose entries are 0 with non-equal probability.

To motivate our final conjecture, we again turn to the linkage between expansion of a graph G and dependencies in its adjacency matrix $A(G)$. One way of thinking of our theorems is that, even for our very sparse graphs, non-expansion is the only source of dependency in the adjacency matrix. If every subset of a set S of rows expands well, then S will be independent. Although random regular graphs are very sparse, they will also typically expand very well. Is it then the case that the adjacency matrix of a random d-regular graph will almost surely be non-singular?

This is not the case for $d = 2$ (as the graph will almost surely contain a $4k$-cycle for some k, which implies a singular adjacency matrix), but it seems likely to be true for larger d.

Conjecture 10.4. *For any fixed $d > 2$, the adjacency matrix of a random d-regular graph is almost surely non-singular.*

References

[1] Costello, K., Tao, T. and Vu, V. (2006) Random symmetric matrices are almost surely non-singular. *Duke Math J.* **135** 395–413.
[2] Costello, K. and Vu, V. (2008) The rank of random graphs. *Random Struct. Alg.* **33** 269–285.
[3] Erdős, P. (1945) On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902.
[4] Halász, G. (1977) Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.* **8** 197–211.
[5] Komlós, J. (1967) On the determinant of $(0,1)$ matrices. *Studia Sci. Math. Hungar.* **2** 7–22.
[6] Vu, V. (2007) Random discrete matrices. In *Horizons of Combinatorics*, Springer, Berlin, pp. 257–280.

https://doi.org/10.1017/S0963548309990447 Published online by Cambridge University Press