ON THE NUMBER OF INTEGRAL GRAPHS

KEVIN P. COSTELLO AND PARKER WILLIAMS

ABSTRACT. We show that at most a $2^{-cn^{3/2}}$ proportion of graphs on n vertices have integral spectrum. This improves on previous results of Ahmadi, Alon, Blake, and Shparlinski (2009), who showed that the proportion of such graphs was exponentially small.

Keywords. Graph Spectrum, Integral Eigenvalue, Random Matrix

AMS Subject Classification: 05C50, 05C80, 15B52

1. INTRODUCTION AND STATEMENT OF MAIN RESULTS

Call a graph **Integral** if all eigenvalues of its adjacency matrix are integers. Examples of integral graphs include the hypercube, the complete graph K_n, the symmetric complete bipartite graph $K_{n,n}$, and the Paley graph on q vertices where q is an odd square prime power.

Integral graphs were first studied by Harary and Schwenk [6], who gave several families of integral graphs, but at the same time described the general question of classifying all integral graphs as "intractable". More recently, there have been several papers noting that integral graphs may be useful in designing quantum spin networks with perfect state transfer (see, for instance, [3, 2]).

It is natural to try and count these graphs, or, equivalently, to give the probability that a random graph (one chosen uniformly at random from the $2^{\binom{n}{2}}$ graphs on n vertices) is integral. The first non-trivial upper bound on this problem was given by Ahmadi, Alon, Blake, and Shparlinski [1]. In probabilistic language, the bound they obtained was

Theorem 1. *The probability that a randomly chosen graph on n vertices is integral is, for sufficiently large n, at most $2^{-n/400}$,*

and they noted in their paper that "we believe our bound is far from being tight and the number of integral graphs is substantially smaller". Our main result confirms this belief, showing that the proportion of integral graphs decays much faster than exponentially.

Theorem 2. *The probability that a randomly chosen graph on n vertices is integral is, for large n, at most $2^{-cn^{3/2}}$, for some absolute constant c.*

Remark 1. This result is likely still not close to being tight. For more on this, see the final section of this paper.

In the next section we will collect a few linear algebraic properties that hold for the adjacency matrix of every graph, deterministic or random. The Proof of Theorem 2 will follow from combining these properties with a counting argument, essentially tracking how the spectrum of the adjacency matrix behaves as the graph grows.

2. A Few Deterministic Observations

We begin with the quick observation that adjacency matrices of graphs cannot have too many large eigenvalues.

Lemma 1. *Let G be an arbitrary graph on n vertices, and let A be the adjacency matrix of G. Then A must contain at least $\frac{3n}{4}$ eigenvalues in the interval $[-2\sqrt{n}, 2\sqrt{n}]$.*

Remark 2. It follows from Wigner's semicircular law [12] together with interlacing that for large n almost every graph on n vertices has fewer than $3n/4$ eigenvalues in the interval $[-1.26\sqrt{n}, 1.26\sqrt{n}]$.

Proof. Let λ_i be the ith eigenvalue of A. We have

$$\sum_{i=1}^{n} \lambda_i^2 = Tr(A^2) = 2E(G) \leq n^2,$$

where $E(G)$ is the number of edges of G. Let T be the multiset of eigenvalues of A that are at least $2\sqrt{n}$ in absolute value. Then we also have

$$\sum_{i=1}^{n} \lambda_i^2 \geq (2\sqrt{n})^2 |T|.$$

Combining the above two bounds, we have $|T| \leq \frac{n}{4}$ so at least $\frac{3n}{4}$ eigenvalues lie inside the interval.

$\square$

In particular, this implies that any integral matrix must have a large amount of multiplicity in its spectrum: The average multiplicity of an integer in $[-2\sqrt{n}, 2\sqrt{n}]$ as an eigenvalue is proportional to $\sqrt{n}$. Tao and Vu [10] have shown that having even a single eigenvalue repeated even twice is (polynomially) unlikely for a random matrix, and the goal will be to show that such a large amount of repetition is far less likely. For this our perspective, drawing on an idea originally due to Komlós [7], will be to think of A as being grown "minor by minor" (equivalently, we will expose the graph G vertex by vertex).

Let A_k be the upper left $k \times k$ minor of A. Then A_{k+1} has the block structure

$$A_{k+1} = \begin{pmatrix} A_k & x_{k+1} \\ x_{k+1}^T & 0 \end{pmatrix}$$

where x_{k+1} is the newly added column. Our second observation (a variation on Lemma 2.1 from [10]) is that whenever an eigenvalue's multiplicity increases from A_k to A_{k+1}, the new column must satisfy certain orthogonality conditions.

Lemma 2. *Let A_k, A_{k+1}, and x_{k+1} be as above. Let λ be any eigenvalue whose multiplicity in the spectrum of A_{k+1} is strictly larger than in A_k, and let v be any eigenvector of A_k corresponding to λ. Then x_{k+1} and v are orthogonal.*

Proof. Let w be an arbitrary eigenvector of A_{k+1}, and write $w = \begin{pmatrix} w' \\ w^{(k+1)} \end{pmatrix}$, where $w' \in \mathbb{R}^k$ and $w^{(k+1)} \in \mathbb{R}$. Note that if $w^{(k+1)} = 0$, then w' is itself an eigenvector of A_k with the same eigenvalue. In particular, since the multiplicity of λ increases from A_k to A_{k+1}, there must be an eigenvector w with $Aw = \lambda w$ and $w^{(k+1)} \neq 0$. Fix such a w.

Now let $v_1, \ldots, v_k$ be an orthonormal eigenbasis for A_k. After changing our basis for $\mathbb{R}^{k+1}$ to $\{v_1, \ldots, v_k, e_{k+1}\}$ (where e_{k+1} is the standard basis vector), we may without loss of generality assume that A_k is diagonal, with the eigenvalues corresponding to λ appearing in the first j coordinates for some j. For any $1 \leq i \leq j$, the ith coordinate of the eigenvalue equation $A_{k+1}w = \lambda w$ now gives

$$\lambda w^{(i)} = \lambda w^{(i)} + w^{(k+1)} x_k^{(i)}$$

where $x_k^{(i)}$ denotes the ith coordinate of x_k. Since by assumption $w^{(k+1)} \neq 0$, we must have $x_k^{(i)} = 0$. Returning to the original basis, this corresponds to x_k being orthogonal to v_i. This is true for every i between 1 and j, so x_k is orthogonal to the entire eigenspace. $\square$

Finally, we will make use of the following observation of Odlyzko [9], whose proof we include here for completeness.

Lemma 3. *Let S be an arbitrary $n-\ell$ dimensional subspace of $\mathbb{R}^n$. Then S contains at most $2^{n-\ell}$ vectors from $(0,1)^n$. Equivalently, S contains at most a $2^{-\ell}$ proportion of all $(0,1)$ vectors.*

Proof. Since S has dimension $n-\ell$, there must be a collection of $n-\ell$ coordinates which parameterize the space, in the sense that those coordinates uniquely determine the remaining ℓ coordinates. There are $2^{n-\ell}$ choices for the values of those coordinates. $\square$

3. The Proof of Theorem 2

The rough idea of our argument will be to track the growth of each integer's multiplicity as an eigenvalue as k increases. On the one hand, we know from Lemma 1 that by the end of the process there must be a large amount of total multiplicity, in the sense that the average multiplicity of an integer in $[-2\sqrt{n}, 2\sqrt{n}]$ as an eigenvalue is large. This will imply there must be many distinct pairs (k, λ) where some eigenvalue λ that already has large multiplicity increases its multiplicity further as A_k is augmented to A_{k+1}. Now Lemma 2 will imply we have many orthogonality relations, which will turn out to be unlikely for a random matrix. We now turn to the details.

Given an $n \times n$ matrix A and an integer i between $-2\sqrt{n}$ and $2\sqrt{n}$, let $\text{Mult}(A_k, i)$ denote the multiplicity of i as an eigenvalue of A_k, and let $\text{Mult}(A, i) = \text{Mult}(A_n, i)$ be its multiplicity in the spectrum of A. By Cauchy Interlacing (e.g. [5] Theorem 4.3.8) we have

$$|\text{Mult}(A_k, i) - \text{Mult}(A_{k+1}, i)| \le 1.$$

In particular, i must attain each multiplicity between 0 and $\text{Mult}(A, i)$ at least once during the augmentation process. With this in mind, we define the vector a_i as follows:

$$a_i(m) = \begin{cases} 0, & \text{if } \text{Mult}(A_m, i) > \text{Mult}(A, i) \\ 0, & \text{if there is some } m' > m \text{ with } \text{Mult}(A_{m'}, i) = \text{Mult}(A_m, i) \\ j, & \text{if } m \text{ is the largest } k \text{ with } \text{Mult}(A_k, i) = j \text{ and } j \le \text{Mult}(A, i) \end{cases}$$

The role of the vectors a_i here is to track how the multiplicity of each eigenvalue of A increases throughout the augmentation process. We make the following observations about the vectors a_i:

- For each $m < n$ with $a_i(m) \ne 0$, the multiplicity of i as an eigenvalue must necessarily increase as A_m is augmented to A_{m+1}, since otherwise m would not be the largest minor with this multiplicity. This is clearly the case if $\text{Mult}(A_{m+1}, i) = \text{Mult}(A_m, i)$, and if $\text{Mult}(A_{m+1}, i) < \text{Mult}(A_m, i)$ then by the interlacing property above the multiplicity of i must reach $\text{Mult}(A_m, i)$ again on the way to $\text{Mult}(A, i)$ (here it is critical that $a_i(m) \ne 0$ only if $\text{Mult}(A_m, i) \le \text{Mult}(A, i)$).
- The value $a_i(n)$ is (by definition) the multiplicity of i as an eigenvalue of A. In particular, by Lemma 1 we have $\sum_i a_i(n) \ge \frac{3n}{4}$.
- As noted above, by interlacing and discrete continuity each multiplicity between 0 and $a_i(n)$ is achieved at some point during the augmentation process. So for each j between 1 and $a_i(n)$, there is a unique m with $a_i(m) = j$, and these m are increasing in j for each i.

We will refer to the collection of sequences a_i corresponding to a matrix A as the **type** of the matrix. Note that the number of possible types for an integral matrix is not too large. There are at most $(n+1)^{4\sqrt{n}+1}$ possible choices for the integral spectrum of A (i.e. the choices $a_i(n)$ for each i), since each $a_i(n)$ is an integer between 0 and n. Once the $a_i(n)$ are chosen, there are at most $n^{a_i(n)}$ possible values for each a_i (choosing where each nonzero value in the vector is). Multiplying over all i, the number of possible types given $a_i(n)$ for all i is at most

$$n^{\sum_i a_i(n)} \le n^n$$

So the total number of distinct possible types for A is at most $(n+1)^{4\sqrt{n}+1} n^n = 2^{o(n^{3/2})}$. This means it is enough to show

Claim 1. *For any fixed type, the probability G is both integral and has that type is at most $2^{-cn^{3/2}}$.*

So let us now consider a fixed type. For $0 \le m \le n-1$, let E_{m+1} be the event that for every eigenvalue i with $a_i(m) \ne 0$, the multiplicity of i as an eigenvalue increases from $a_i(m)$ to $a_i(m)+1$ as we augment A_m to A_{m+1}. These events correspond to the each eigenvalue's multiplicity following the correct track throughout the augmentation process. For a matrix to have the desired type, each

of the E_{m+1} must hold, so we have

$$\mathbf{P}(A \text{ has the given type }) \leq \prod_{m=0}^{n-1} \mathbf{P}\left(E_{m+1}|E_1, \ldots, E_m\right)$$

$$\leq \prod_{m=0}^{n-1} \max_{A_m} \mathbf{P}\left(E_{m+1}|A_m\right),$$

where the last inequality follows since A_m completely determines the events E_1 through E_m.

By lemma 2, we know that for E_{m+1} to hold, x_{m+1} must be orthogonal to the eigenspace of A_m corresponding to each eigenvalue whose multiplicity increases from A_m to A_{m+1}. In particular, it must be simultaneously orthogonal to the eigenspace for every i where $a_i(m) \neq 0$, since by definition that corresponds to an increase in the multiplicity of i (there may be other eigenvalues whose multiplicities increase but for which $a_i(m) = 0$ because m is not maximal for its multiplicity. We ignore them).

Since we are now conditioning on A_m, this corresponds to a fixed subspace in which x_{m+1} must lie for E_{m+1} to hold. That subspace has co-dimension equal to $\sum_i a_i(m)$. Using Lemma 3 the probability x_{m+1} lies in this subspace, we have

$$\mathbf{P}\left(E_{m+1}|A_m\right) \leq 2^{-\sum_i a_i(m)}.$$

Multiplying over all m and taking logarithms, we have

$$-\log_2\left(\mathbf{P}(A \text{ has the given type })\right) \geq \sum_{m=0}^{n-1} \sum_i a_i(m)$$

One way of thinking about this sum is that every time an eigenvalue's multiplicity increases, it contributes to the total co-dimension (and thus to the exponent) an amount equal to its current multiplicity. Recall that for each i and for each $0 \leq j \leq a_i(n)$, there is a unique m satisfying $a_i(m) = j$. So we can rewrite this bound as

$$-\log_2\left(\mathbf{P}(A \text{ has the given type })\right) \geq \sum_{i=-2\sqrt{n}}^{2\sqrt{n}} \sum_{j=0}^{a_i(n)-1} j$$

$$= \frac{1}{2} \sum_{i=-2\sqrt{n}}^{2\sqrt{n}} a_i(n)^2 - \frac{1}{2} \sum_{i=-2\sqrt{n}}^{2\sqrt{n}} a_i(n)$$

$$\geq \frac{1}{2(4\sqrt{n}+1)} \left(\sum_{i=-2\sqrt{n}}^{2\sqrt{n}} a_i(n) \right)^2 - \frac{1}{2} \sum_{i=-2\sqrt{n}}^{2\sqrt{n}} a_i(n)$$

By Lemma 1, we have

$$\frac{3n}{4} \leq \sum_{i=-2\sqrt{n}}^{2\sqrt{n}} a_i(n) \leq n.$$

So it follows that

$$-\log_2\left(\mathbf{P}(A \text{ has the given type })\right) \geq \frac{1}{2(4\sqrt{n}+1)}\left(\frac{3n}{4}\right)^2 - \frac{n}{2} = \left(\frac{9}{128} + o(1)\right)n^{3/2}$$

4. EXTENSIONS AND FURTHER CONJECTURES

The proof of Theorem 2 did not use in any vital way the fact that we were working with matrices with integral eigenvalues. Indeed, an identical argument would show the following.

Theorem 3. *Let S be any subset of the algebraic integers, and suppose there are constants $\alpha < 2$ and c such that for sufficiently large N*

$$|S \cap [-N, N]| \leq cN^\alpha$$

then there is a constant c' such that for sufficiently large n, the proportion of graphs on n vertices having spectrum lying entirely in S is at most $2^{-c'n^{2-\alpha/2}}$

Similarly, although $(0,1)$ matrices were natural to look at due to the graphical motivations, the actual distribution of the entries was not critical here. A similar theorem would hold if the entries were drawn from any other bounded non-degenerate distribution (with the constant now depending on the distribution in question).

It seems unlikely that $3/2$ is the correct exponent in Theorem 2, and indeed we suspect that the probability a graph is integral is $2^{-\left(\frac{1}{2}+o(1)\right)n^2}$ (equivalently, that the number of integral graphs is $2^{o(n^2)}$). However, it seems that improving the exponent beyond $3/2$ will require some significant new idea. The main stumbling block is in a sense estimating the number of graphs for which a fixed eigenvalue appears with large multiplicity. Even in the case $\lambda = 0$, this seems like an interesting problem.

Question 1. *Let Q_n be a random $n \times n$ symmetric matrix where each above diagonal entry is equally likely to be 0 and 1. For a given s (possibly growing with n), what is the probability Q_n has rank at most $n - s$?*

In the case $s = 1$ this corresponds to estimating the singularity probability of Q_n, which is a well-studied problem [4, 8, 11]. The current best upper bound is due to Vershynin, who showed in [11] that for large n the probability is at most $\exp\left(-n^c\right)$ for some $c > 0$. The best known lower bound is $(1 + o(1))\binom{n}{2}2^{-n}$, coming from the probability that some pair of rows of Q_n are equal (and it is a longstanding conjecture that this bound is optimal).

For larger s, the authors of [1] showed an upper bound of $2^{-\frac{s^2-s}{2}}$ in their proof of Theorem 1, and a similar bound showed up in our argument where we estimated the probability of a matrix having a given type (with $a_i(n)$ for a single i playing the role of s). A natural lower bound here would be the probability that Q_n contains at least s zero rows (equivalently, that the corresponding graph has at least s isolated vertices), which for s much smaller than n is $2^{-(1+o(1))ns}$. This bound may well be essentially optimal.

References

[1] O. Ahmadi, N. Alon, I. Blake, and I. Shparlinski. Graphs with integral spectrum. *Linear Alg. Appl.* **430** (2009), 547-552

[2] M. Bavsić, M. Petković, and D. Stevanović. Perfect state transfer in integral circulant graphs. *Appl. Math. Lett.* **22** (2009), 1117-1121

[3] M. Christandl, N. Datta, A. Ekert, and A.J. Landahl. Perfect state transfer in quantum spin networks. *Phys. Rev. Lett.* **92** (2004), 189702.

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

[5] R. Horn and C. Johnson, *Matrix Analysis*, 2nd edition. Cambridge University Press (1999).

[6] F. Harary and A.J. Schwenk. Which Graphs have Integral Spectra? in *Graphs and Combinatorics*. Lecture Notes in Mathematics Vol. 406 (1974) 45-51.

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

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

[9] A. Odlyzko. On subspaces spanned by random selections of ± 1 vectors. *J. Comb. Theory A*, **47** (1988) 124-133.

[10] T. Tao and V. Vu. Random matrices have simple spectrum. *Preprint*, available at arXiv:1412.1438

[11] R. Vershynin. Invertibility of symmetric random matrices, *Random Structures and Algorithms* **44** (2014), 135-182

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

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, RIVERSIDE, 900 UNIVERSITY AVE., RIVERSIDE, CA 92521 USA

E-mail address: costello@math.ucr.edu

E-mail address: parker@math.ucr.edu