

THE EXACT RANK OF SPARSE RANDOM GRAPHS

MARGALIT GLASGOW, MATTHEW KWAN, ASHWIN SAH, AND MEHTAAB SAWHNEY

ABSTRACT. Two landmark results in combinatorial random matrix theory, due to Komlós and Costello–Tao–Vu, show that discrete random matrices and symmetric discrete random matrices are typically nonsingular. In particular, in the language of graph theory, when p is a fixed constant, the biadjacency matrix of a random Erdős–Rényi bipartite graph $\mathbb{G}(n, n, p)$ and the adjacency matrix of an Erdős–Rényi random graph $\mathbb{G}(n, p)$ are both nonsingular with high probability. However, very sparse random graphs (i.e., where p is allowed to decay rapidly with n) are typically singular, due to the presence of “local” dependencies such as isolated vertices and pairs of degree-1 vertices with the same neighbour.

In this paper we give a combinatorial description of the rank of a sparse random graph $\mathbb{G}(n, n, c/n)$ or $\mathbb{G}(n, c/n)$ in terms of such local dependencies, for all constants $c \neq e$ (and we present some evidence that the situation is very different for $c = e$). This gives an essentially complete answer to a question raised by Vu at the 2014 International Congress of Mathematicians.

As applications of our main theorem and its proof, we also determine the asymptotic singularity probability of the 2-core of a sparse random graph, we show that the rank of a sparse random graph is extremely well-approximated by its matching number, and we deduce a central limit theorem for the rank of $\mathbb{G}(n, c/n)$.

1. INTRODUCTION

A foundational theorem in combinatorial random matrix theory, due to Komlós [47, 48], says that discrete random matrices with i.i.d. entries are typically nonsingular (over \mathbb{R}). In particular, let B be an $n \times n$ random matrix with i.i.d. Bernoulli(p) entries (meaning that each entry B_{ij} satisfies $\Pr[B_{ij} = 1] = p$ and $\Pr[B_{ij} = 0] = 1 - p$). For any constant $p \in (0, 1)$, if we take $n \rightarrow \infty$ then such a random matrix is nonsingular *with high probability* (“whp” for short): that is, $\lim_{n \rightarrow \infty} \Pr[B \text{ is singular}] = 0$.

A huge number of strengthenings and variations of Komlós’ theorem have been considered over the years. Two particular highlights include a result of Tikhomirov [62] that for constant $0 < p \leq 1/2$, the singularity probability is $(1 - p + o(1))^n$, and a result of Costello, Tao, and Vu [26] that *symmetric* discrete random matrices are also nonsingular whp. A symmetric binary matrix can be interpreted as the adjacency matrix of a graph, so the Costello–Tao–Vu theorem has an interpretation in terms of random graphs: for constant $p \in (0, 1)$, an Erdős–Rényi random graph $G \sim \mathbb{G}(n, p)$ has nonsingular adjacency matrix whp¹. Actually, Komlós’ theorem can be interpreted in graph-theoretic terms as well: the random matrix B described above can be interpreted as the *biadjacency* matrix of a *bipartite* Erdős–Rényi random graph $G \sim \mathbb{G}(n, n, p)$ (where one of the parts corresponds to the rows of the matrix, and the other part corresponds to the columns).

If p decays too rapidly with n (in particular, if $p \leq (1 - \varepsilon) \log n/n$ for some constant $\varepsilon > 0$), then for reasons related to the coupon collector problem, a typical outcome of $G \sim \mathbb{G}(n, p)$ (respectively, $G \sim \mathbb{G}(n, n, p)$) has isolated vertices, meaning that its adjacency matrix (respectively, biadjacency matrix) has all-zero rows and is therefore singular. In fact, $\log n/n$ is a *sharp threshold* for singularity, in the sense that if $p \geq (1 + \varepsilon) \log n/n$ (and p is bounded away from 1) then a typical $G \sim \mathbb{G}(n, p)$ (respectively, $G \sim \mathbb{G}(n, n, p)$) has nonsingular adjacency matrix (respectively, nonsingular biadjacency matrix). This seems to have been first observed by Costello and Vu [28]², and refinements and generalisations were proved by Basak and Rudelson [7] and Addario-Berry and Eslava [2]. In particular, the latter authors proved a sharp *hitting time* type result: if we consider the random graph *process* where we start with the empty graph on n vertices (or the empty bipartite graph with $n + n$ vertices) and add random edges one-by-one (respecting our bipartition, in the bipartite case), then whp at the very same moment where the last isolated vertex disappears our graph becomes nonsingular.

¹There is a slight difference between a random symmetric Bernoulli matrix and the adjacency matrix of a random graph: namely, the adjacency matrix of any graph has zeroes on the diagonal. However, the same techniques usually apply to both settings, and we will not further concern ourselves with this detail.

²The Costello–Vu proof was only written for $\mathbb{G}(n, p)$, but it can be easily adapted to $\mathbb{G}(n, n, p)$; alternatively, see [34] for a very simple proof in the $\mathbb{G}(n, n, p)$ case.

Naïvely, it is quite surprising that the property of being singular (which is intrinsically an algebraic property of a matrix) can be predicted so effectively by the simple combinatorial property of having an isolated vertex. It is natural to ask whether this theme continues below the singularity threshold $\log n/n$: even when a random matrix is likely to be singular, can we describe the *rank* via “local combinatorial dependencies” such as isolated vertices? In their aforementioned paper, Costello and Vu [28] actually proved the first result along these lines: for $p \geq (1/2 + \varepsilon) \log n/n$, whp the rank of $\mathbb{G}(n, p)$ is precisely n minus the number of isolated vertices. In follow-up work, Costello and Vu [27] considered the more general regime where $p \geq c \log n/n$ for a constant $c > 0$. They found that in this regime it is still possible to give a combinatorial description of the rank, though one must consider more sophisticated types of “local dependencies” than isolated vertices. For example, the next simplest type of dependency is a *cherry*: a pair of degree-1 vertices with the same neighbour. More recently, DeMichele, the first author, and Moreira [30] gave a combinatorial description of the rank of $G \sim \mathbb{G}(n, p)$ and $G \sim \mathbb{G}(n, n, p)$, in terms of a procedure that iteratively deletes local dependencies, which holds whp whenever $\lim_{n \rightarrow \infty} np = \infty$ (i.e., when p asymptotically dominates $1/n$).

The most challenging regime is where $p = c/n$ for constant c . An asymptotic for the typical rank of $\mathbb{G}(n, c/n)$ was conjectured by Bauer and Golinelli [9] (motivated by statistical physics considerations), and this asymptotic was later proved by Bordenave, Lelarge, and Salez [18] via analytic techniques. In his lecture at the 2014 International Congress of Mathematicians [67] (also in [66]), Vu asked whether one can also give a precise combinatorial characterisation of the rank in this regime.

The main purpose of this paper is to provide an answer to Vu’s question, and the analogous question for $\mathbb{G}(n, n, c/n)$, exactly characterising the rank of sparse random graphs (and in the process, providing a linear-time algorithm to compute the rank).

At a high level, we show that whp all linear dependencies arise from two different types of combinatorial structures. First, we need to account for “tree-like” structures generalising isolated vertices and cherries (previously identified in the work of Costello and Vu [27] and DeMichele, the first author, and Moreira [30]), which can be “peeled off” by an iterative process called *Karp–Sipser leaf removal* (defined in Definition 1.4). Second, we need to account for certain short cycles (which we call “special cycles”, defined in Definition 1.5), which cause linear dependencies for a different reason.

Our proof of this characterisation involves a wide range of tools and ideas, both original and adapted from existing work. This includes analysis of degree-constrained random graphs and of the *Karp–Sipser leaf-removal algorithm*, robust analysis of random walks, spectral convergence machinery for locally convergent graphs, a “rank-boosting” technique, and some special-purpose notions of matrix pseudoinverses and “minimal kernel vectors” (all of which we describe further in Section 2). To try to give a rough impression of the most fundamental difficulty compared to previous work: note that the rank of a matrix can be interpreted as the size of its largest nonsingular submatrix. In the setting of most previous work, maximum nonsingular submatrices are in some sense “robustly” nonsingular (in particular, the corresponding subgraphs have good *expansion* properties), which makes it possible to rule out certain types of kernel vectors via lossy union bounds. However, in our situation the largest nonsingular submatrices are in some sense “only barely nonsingular”, with essentially the weakest possible expansion a nonsingular submatrix can have, and there is almost no room to make any kind of lossy approximation.

In any case, once one has a characterisation of the rank in terms of explicit combinatorial structures, it becomes possible to prove further results about the rank via combinatorial tools. Indeed, as corollaries of our main theorem and its proof, we are able to show a number of additional theorems: we compute the asymptotic singularity probability of the 2-core, we obtain a very strong bound on the difference between the matching number and the rank, and we prove a central limit theorem for the rank of $\mathbb{G}(n, c/n)$. Since the statement of our main result (Theorem 1.7) is somewhat technical, we take a moment to discuss these corollaries before presenting the precise statement of our main result.

1.1. Nonsingularity of the 2-core. The k -core $\text{core}_k(G)$ of a graph G is the subgraph obtained by iteratively deleting vertices with degree less than k (in any order). Equivalently, it is the largest induced subgraph with minimum degree at least k . This notion was first introduced in 1984 by Bollobás [16], and k -cores have since become fundamental objects of study in random graph theory.

In the context of combinatorial random matrix theory, an important reason to study k -cores is that all of the most obvious types of “local dependencies” involve vertices of low degree. For example, recall that isolated vertices and cherries are abundant types of local dependencies, and it turns out that all of the “tree-like” local dependencies mentioned earlier in this introduction contain a vertex of degree 1. Another example of a local dependency, which has non-negligible probability of appearing in the regime

$p = c/n$, is a pair of vertices of degree 2 with the same neighbourhood (i.e., a cycle of length 4, in which a pair of opposite vertices have degree exactly 2).

Resolving a conjecture of Vu, it was recently proved by Ferber and the last three authors [33] (see also [30]) that for constants $k \geq 3$ and $c > 0$, the k -core of $\mathbb{G}(n, c/n)$ is nonsingular whp. That is to say, trimming low-degree vertices typically removes any singularity present in the graph (foreshadowing the main result of this paper, that whp the *only* dependencies are “tree-like” or “cycle-like”).

While the assumption $k \geq 3$ is necessary for a “whp” result due to the possibility of “cycle-like” dependencies, these types of dependencies seem to be rare (e.g., the expected number of 4-cycles is only $O(1)$, and with non-negligible probability there are no 4-cycles at all). So, it is natural to ask whether one can still prove meaningful theorems about nonsingularity in the case $k = 2$. We prove such a theorem: roughly speaking, the 2-core is “right on the edge of singularity”, being neither singular whp nor nonsingular whp. (In retrospect, one can see that in the case $k \geq 3$, the k -core is actually “quite far from being singular” with respect to natural local dependencies, and this “wobble room” played a crucial role in the proofs in [30, 33]).

Theorem 1.1. *Fix a constant $c > 1$, let $G \sim \mathbb{G}(n, c/n)$, and let A be the adjacency matrix of the 2-core of the largest component of G (which is unique whp). Then*

$$\lim_{n \rightarrow \infty} \Pr[A \text{ is nonsingular}] = \left(\frac{1 - \left(\frac{\lambda_2}{e^{\lambda_2/2} - e^{-\lambda_2/2}} \right)^4}{1 - \left(\frac{\lambda_2}{e^{\lambda_2} - 1} \right)^4} \right)^{1/4} > 0,$$

where $\lambda_2 = \lambda_2(c) > 0$ is the unique solution to $\lambda_2/(1 - e^{-\lambda_2}) = c$. Moreover, the corank³ of A has an asymptotic Poisson(μ) distribution, where μ is chosen such that the above probability is asymptotic to $e^{-\mu}$ (and in particular, the corank is bounded in probability⁴).

Note that the assumption $c > 1$ corresponds to the celebrated *phase transition* of the Erdős–Rényi random graph. Indeed, if $c < 1$ (the “subcritical” regime), whp all the connected components of $\mathbb{G}(n, c/n)$ have size $O(\log n)$ (and each of them is a tree or is *unicyclic*, having exactly one cycle); thus, in this regime the 2-core is rather trivial, consisting only of a very small number of isolated cycles. On the other hand, if $c > 1$ (the “supercritical” regime), then whp $\mathbb{G}(n, c/n)$ has a unique “giant component” with nontrivial structure (whose number of vertices is of order n), in addition to some trees and unicyclic components of size $O(\log n)$. See for example the monographs [17, 36, 44] for more details about the component phase transition of the Erdős–Rényi random graph, and see [31] for a precise description of the “anatomy” of a supercritical random graph in terms of its 2-core.

Remark. The statement of Theorem 1.1 is only about the giant component in the supercritical regime, but one may also wish to consider the entire 2-core (including any small-cycle components), in which case it makes sense to consider all $c > 0$ (not just $c > 1$). With the methods in this paper (and some results about critical random graphs [1]) it is possible to show that for $G \sim \mathbb{G}(n, c/n)$ we have $\lim_{n \rightarrow \infty} \Pr[\text{the 2-core of } G \text{ is nonsingular}] > 0$ if and only if $c \neq 1$ (when $c = 1$ there are too many nontrivial components, each of which is reasonably likely to be singular). We omit the details. (Also, note that the asymptotic nonsingularity probability in Theorem 1.1 tends to zero as $c \rightarrow 1$.)

1.2. Comparing the rank and the matching number. In a graph G , a *matching* is a collection of disjoint edges. The *matching number* $\nu(G)$ is the maximum number of edges in a matching in G . If G is bipartite, then $\nu(G)$ can be interpreted as the size of the largest permutation matrix “contained” in the biadjacency matrix $B(G)$ of G , where our notion of matrix containment allows deleting rows and columns, and changing 1-entries to 0-entries. Recalling the permutation definition of the determinant, $\nu(G)$ is a trivial upper bound for $\text{rank } B(G)$. Confirming a conjecture of Lelarge [50] motivated by statistical physics considerations, it was proved by Coja-Oghlan, Ergür, Gao, Hetterich, and Rolvien [24] that this trivial bound is nearly best-possible⁵ for sparse random bipartite graphs: for $G \sim \mathbb{G}(n, n, c/n)$ we have $\nu(G) - \text{rank } B(G) = o(n)$ whp.

In the non-bipartite case, there is no general inequality relating the rank of the adjacency matrix $\text{rank } A(G)$ of a graph G with its matching number $\nu(G)$, but a theorem of Bordenave, Lelarge, and

³The *corank* of a matrix is the dimension of its kernel.

⁴A sequence of random variables $(X_n)_{n=1}^\infty$ is said to be *bounded in probability* if for all $\varepsilon > 0$, there are N, M such that $\Pr[X_n \geq M] \leq \varepsilon$ for all $n \geq N$.

⁵Actually, they proved this for a much more general class of random matrices and for rank over any field.

Salez [18] (mentioned earlier in the introduction) shows that nonetheless for a sparse random graph G we have $\text{rank } A(G) = 2\nu(G) + o(n)$ whp (see also [19]). It is also natural to consider an alternative parameter $\sigma(G)$, defined to be the size of the largest permutation matrix “contained” in $A(G)$. This parameter $\sigma(G)$ has a combinatorial interpretation as the maximum number of vertices in a union of vertex-disjoint cycles and edges in G . Note that $2\nu(G)$ and $\text{rank } A(G)$ are both at most $\sigma(G)$.

As our second result (a corollary of our main result [Theorem 1.7](#), to come), we dramatically improve the $o(n)$ error terms in the results described above, showing that $\nu(G)$ is an *extremely* good approximation for $\text{rank } B(G)$, and $2\nu(G)$ and $\sigma(G)$ are extremely good approximations for $\text{rank } A(G)$, away from a “critical point” $p = e/n$. (The significance of this rather mysterious-sounding critical point will be explained later in this introduction; for now we just remark that this point also happens to be critical for several other spectral phenomena in Erdős–Rényi random graphs [22, 25]).

Theorem 1.2. *Fix a constant $c \neq e$.*

- (A) *Let $G \sim \mathbb{G}(n, c/n)$.*
 - (1) *$|\text{rank } A(G) - 2\nu(G)|$ is bounded in probability.*
 - (2) *$|\text{rank } A(G) - \sigma(G)|$ is bounded in probability.*
- (B) *For $G \sim \mathbb{G}(n, n, c/n)$, $|\text{rank } B(G) - \nu(G)|$ is bounded in probability.*

Remark. Given [Theorem 1.2](#), one may wonder whether $\nu(G)$ (in the setting of (B), and $2\nu(G), \sigma(G)$ in the setting of (A)) in fact *perfectly* describe the rank. For example, could it be true that in the setting of (B) we have $\text{rank } B(G) = \nu(G)$ whp? As will become clear when we discuss our main theorem, this is too much to hope for. We believe that in the setting of (B), the asymptotic distribution of $\nu(G) - \text{rank } B(G)$ is Poisson (with a certain explicit parameter), and in the setting of (A), both $2\nu(G) - \text{rank } A(G)$ and $\sigma(G) - \text{rank } A(G)$ have somewhat more complicated “Poisson-like” distributions. However, rigorous proofs of these facts would require adaptations of certain highly nontrivial graph-theoretic results (to characterise $\nu(G)$ and $\sigma(G)$). We believe that these adaptations are possible, but pursuing this direction would be outside the scope of the present paper. See [Section 13](#) for details.

1.3. The asymptotic distribution of the rank. Aronson, Frieze, and Pittel [5] conjectured that for a constant c , the matching number $\nu(G)$ of a random graph $G \sim \mathbb{G}(n, c/n)$ satisfies a central limit theorem. This was proved for $c < 1$ by Pittel [60], and for $c > e$ by Kreačić [49, Theorem 19]. Since [Theorem 1.2\(A1\)](#) tells us that $\text{rank } A(G)$ is extremely well approximated by $2\nu(G)$, it is easy to deduce a corresponding central limit theorem for the rank.

Corollary 1.3. *Let $G \sim \mathbb{G}(n, c/n)$ for a constant $c < 1$ or $c > e$, let $G \sim \mathbb{G}(n, c/n)$, and let X be the rank of the adjacency matrix of G . Then $(X - \mathbb{E}X)/\sqrt{\text{Var } X} \xrightarrow{d} \mathcal{N}(0, 1)$.*

Actually, in an upcoming paper together with Goldschmidt and Kreačić [38], we are able to close the gap between 1 and e in [Corollary 1.3](#). Specifically, the regime $c \leq e$ is rather different in nature than the regime $c > e$, and when $G \sim \mathbb{G}(n, c/n)$ for $c \leq e$, we are able to give a unified proof that the rank and matching number of G both satisfy a central limit theorem (without going through [Theorem 1.2\(A1\)](#)).

Remark. [60] and [49] provide explicit formulas for the asymptotic values of $\mathbb{E}X$ and $\text{Var } X$, though these are a bit too complicated to describe here. It is worth remarking that the asymptotic formula for $\text{Var } X$ is the single place in this paper where there is a material difference between the “binomial” model of Erdős–Rényi random graphs (where each edge is present with probability p independently) and the “uniform” model of Erdős–Rényi random graphs (where we choose a random subset of exactly m edges, for say $m = \lfloor p\binom{n}{2} \rfloor$). Indeed, the variance of the matching number (and therefore the variance of the rank) differs by a constant factor between these two settings; see [49, 60] for details. For all the other results in the paper (which are all stated for the binomial model), one can make trivial changes to the proofs to obtain exactly the same result in the uniform model.

Remark. We believe that a central limit theorem does *not* hold for the rank of $\mathbb{G}(n, n, c/n)$; see [Section 1.6](#).

1.4. Exactly characterising the rank. In this subsection we finally state our main theorem, giving an exact combinatorial characterisation of the rank of a sparse random matrix. First, we need to introduce the *Karp–Sipser leaf removal algorithm*, which was introduced in 1981 by Karp and Sipser [46] as a tool to study matchings in random graphs (in a paper which kickstarted the *differential equations method* for random graph processes; see [70]), but is now also of great importance in statistical physics, theoretical computer science, and random matrix theory (see for example [8, 9, 18, 22, 57]).

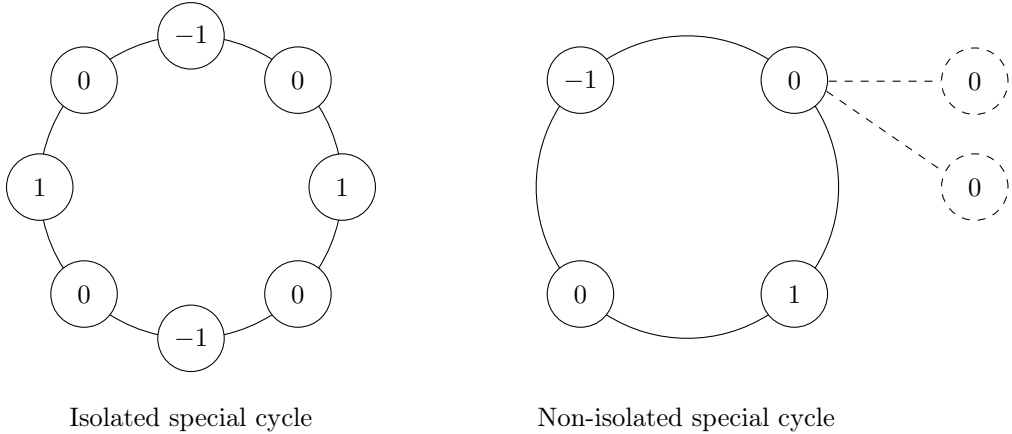


Figure 1. On the left is an isolated special cycle with 8 vertices. On the right is a non-isolated special cycle with 4 vertices. In both pictures, we depict the entries of a kernel vector of the corresponding adjacency matrix. Note that for the isolated cycle, one can obtain an additional linearly independent kernel vector by shifting each entry one edge clockwise around the cycle.

Definition 1.4 (Karp–Sipser leaf removal). Starting from a graph G , choose an arbitrary degree-1 vertex and delete it together with its neighbour. Repeat this “leaf-deletion” until no further degree-1 vertices remain. Let $i(G)$ be the number of isolated vertices in the resulting graph. If G is bipartite, let $i_1(G)$ and $i_2(G)$ be the number of isolated vertices on the two sides of the bipartition $V_1 \cup V_2$. Let $\text{core}_{\text{KS}}(G)$ be the graph of remaining non-isolated vertices (the *Karp–Sipser core*). One can check that $i(G)$ and $\text{core}_{\text{KS}}(G)$ (and $i_1(G), i_2(G)$, if G is bipartite) do not depend on the order that the leaf-deletions are performed (see for example the appendix of [8]).

It is easy to check (see [Lemma 5.1](#)) that a single step of leaf-removal decreases $\text{rank } A(G)$ by exactly 2, and if G is bipartite, decreases $\text{rank } B(G)$ by exactly 1. It is then easy to deduce (see [Corollary 5.2](#)) that $\text{rank } A(G) \leq n - i(G)$ for any n -vertex graph G (i.e., $\text{corank } A(G) \geq i(G)$), and $\text{rank } B(G) \leq n - \max(i_1(G), i_2(G))$ for any $(n+n)$ -vertex bipartite graph G (i.e., $\text{corank } B(G) \geq \max(i_1(G), i_2(G))$). We will refer to these two bounds as the *Karp–Sipser bounds* for the rank of $A(G)$ and $B(G)$, respectively. We remark that there is a one-sided version of the Karp–Sipser bound for $B(G)$ (where leaves are only removed from one of the two sides of our bipartite graph), sometimes called the *2-core bound* in the computer science and statistical physics literature [3, 24, 29] (here “2-core” refers to a certain hypergraph notion of a 2-core, not to be confused with the notion in [Theorem 1.1](#)).

The Karp–Sipser process takes care of “tree-like” local dependencies. In random graphs $\mathbb{G}(n, p)$ or $\mathbb{G}(n, n, p)$ with $np \rightarrow \infty$, these are whp the only types of dependencies that exist (see [27, 30]); that is, the Karp–Sipser core is nonsingular, so the Karp–Sipser bound is sharp. However, in the case $p = O(1/n)$, there may be “cycle-like” local dependencies in the Karp–Sipser core, such as pairs of degree-2 vertices with the same neighbourhood. We capture dependencies of this type in the following definition, depicted in [Figure 1](#).

Definition 1.5 (Special cycles). Say an induced cycle in a graph G is *special* if its length is divisible by 4, and if every second vertex has degree 2 in G . In particular, an *isolated cycle* is a cycle in which every vertex has degree exactly 2 (i.e., it is its own connected component), so isolated cycles with length divisible by 4 are special “in two different ways”. Let $s(G)$ be the number of special cycles in G , where we count each isolated cycle twice.

If G is bipartite, say an induced cycle in G is 1-special (respectively, 2-special) if its length is divisible by 4, and every vertex in V_1 (respectively, every vertex in V_2) has degree 2. Let $s_1(G)$ and $s_2(G)$ be the numbers of 1-special and 2-special cycles in G , respectively.

To see that a special cycle indeed constitutes a dependency, note that we can construct a kernel vector by “alternating ± 1 entries around a special cycle”, as follows.

Fact 1.6. Let G be a graph on the vertex set V . Let u_1, \dots, u_{4k} (in order) be the vertices of a special cycle, where u_2, u_4, \dots, u_{4k} have degree 2. Define $\mathbf{v} \in \{-1, 0, 1\}^V$ by setting the entries indexed by $u_2, u_6, \dots, u_{4k-2}$ to 1 and setting the entries indexed by u_4, u_8, \dots, u_{4k} to -1 , and setting all other

entries to zero. (That is to say, we go around the cycle, alternating 1 and -1 on our degree-2 vertices). Then, \mathbf{v} is a kernel vector of $A(G)$.

When G is a bipartite graph with bipartition $V_1 \cup V_2$, an analogous construction gives a left kernel vector of $B(G)$ if $u_2, u_4, \dots, u_{4k} \in V_1$, and a right kernel vector of $B(G)$ if $u_2, u_4, \dots, u_{4k} \in V_2$.

Our main theorem says that for $c \neq e$, the rank of a sparse random graph $\mathbb{G}(n, c/n)$ or $\mathbb{G}(n, n, c/n)$ can be described in terms of the Karp–Sipser bound and the special cycles within the Karp–Sipser core.

Theorem 1.7. *Fix a constant $c \neq e$.*

(A) *Let $G \sim \mathbb{G}(n, c/n)$. Then whp $\text{corank } A(G) = i(G) + s(\text{core}_{\text{KS}}(G))$.*

(B) *Let $G \sim \mathbb{G}(n, n, c/n)$. Then whp*

$$\text{corank } B(G) = \max(i_1(G) + s_1(\text{core}_{\text{KS}}(G)), i_2(G) + s_2(\text{core}_{\text{KS}}(G))).$$

Remark. If we fix a vertex and consider an exploration process to find a special cycle containing that vertex, it is not hard to show that this process is *subcritical* and explores only $O(1)$ vertices in expectation. Via a standard concentration inequality, it follows that whp we can find all the special cycles in the Karp–Sipser core in time $O(n)$. The Karp–Sipser leaf removal process also completes in time $O(n)$, so [Theorem 1.7](#) actually gives a linear-time algorithm for computing the rank of a sparse random graph⁶.

We can also describe the asymptotic distribution of the “defect” in the Karp–Sipser bound; for this we need to define some Poisson parameters.

Definition 1.8 (Poisson parameters). For $0 \leq c < e$, let $\eta = \eta(c) \in [0, 1]$ be the unique solution to the equation $c = \eta e^\eta$. For $c \geq 0$, define $\Phi_c: [0, 1] \rightarrow [0, 1]$ by $\alpha \mapsto 1 - \exp(-c \exp(-c(1 - \alpha)))$. If $c > e$ then Φ_c has multiple fixed points (see for example [\[22\]](#)); let $\alpha_* = \alpha_*(c)$ and $\alpha^* = \alpha^*(c)$ be the smallest and largest of these fixed points, respectively, and let $\lambda_{\text{KS}}(c) = c(\alpha^* - \alpha_*)$. For $\lambda \geq 0$ let

$$\gamma(\lambda) = -\frac{1}{4} \log \left(1 - \left(\frac{\lambda}{e^{\lambda/2} - e^{-\lambda/2}} \right)^4 \right), \quad \gamma^\dagger(\lambda) = -\frac{1}{8} \log \left(1 - \left(\frac{\lambda}{e^\lambda - 1} \right)^4 \right)$$

Then, for $c \in [0, e) \cup (e, \infty)$ let

$$\gamma_{\text{B}} = \gamma_{\text{B}}(c) = \begin{cases} -\frac{1}{4} \log(1 - \eta^4) & \text{if } c < e, \\ \gamma(\lambda_{\text{KS}}(c)) & \text{if } c > e, \end{cases}$$

$$\gamma_{\text{A}}^\dagger = \gamma_{\text{A}}^\dagger(c) = \begin{cases} \gamma_{\text{B}}/2 & \text{if } c < e, \\ \gamma^\dagger(\lambda_{\text{KS}}(c)) & \text{if } c > e, \end{cases}, \quad \gamma_{\text{A}} = \gamma_{\text{A}}(c) = \begin{cases} 0 & \text{if } c < e, \\ \gamma_{\text{B}} - 2\gamma_{\text{A}}^\dagger & \text{if } c > e. \end{cases}$$

Theorem 1.9. *Fix a constant $c \neq e$.*

(A) *Let $G \sim \mathbb{G}(n, c/n)$. Then*

$$\text{corank } A(G) - i(G) \xrightarrow{d} Y + 2Y^\dagger,$$

where Y, Y^\dagger are independent Poisson random variables with means $\gamma_{\text{A}}(c)$ and $\gamma_{\text{A}}^\dagger(c)$, respectively.

(B) *Let $G \sim \mathbb{G}(n, n, c/n)$. Then*

$$\text{corank } B(G) - \max(i_1(G), i_2(G)) \xrightarrow{d} Y,$$

where Y is Poisson with mean $\gamma_{\text{B}}(c)$.

Remark. As written, our proof is not strong enough to estimate the *expected* defect in the Karp–Sipser bound, but it does seem to be possible to prove such estimates by taking more care with quantitative aspects (which we do not pursue in this paper, in the interests of keeping our proofs as simple as possible). Specifically, one expects $\lim_{n \rightarrow \infty} \mathbb{E}[\text{corank } A(G) - i(G)] = \gamma_{\text{A}}(c) + 2\gamma_{\text{A}}^\dagger(c)$ in the setting of (A) and $\lim_{n \rightarrow \infty} \mathbb{E}[\text{corank } B(G) - \max(i_1(G), i_2(G))] = \gamma_{\text{B}}(c)$ in the setting of (B).

The reader is overdue an explanation for the significance of the “critical point” $c = e$. It turns out that this point amounts to a “phase transition” for the Karp–Sipser process. Namely (in the settings of both $\mathbb{G}(n, n, c/n)$ and $\mathbb{G}(n, c/n)$), for $c < e$, the Karp–Sipser core whp consists of a tiny number of vertex-disjoint cycles, whereas for $c > e$ the Karp–Sipser core whp has a single giant component with nontrivial structure, in addition to a tiny number of vertex-disjoint cycles. This situation parallels the

⁶To be precise, we obtain a linear-time algorithm to compute a quantity that agrees with the rank whp.

phase transition (at $c = 1$) of the components of a random graph, and suggests that when $c = e$, the Karp–Sipser core may have similar structure to the 2-core of a critical random graph $\mathbb{G}(n, 1/n)$ or $\mathbb{G}(n, n, 1/n)$. Unfortunately, it is very challenging to study the Karp–Sipser core in this critical regime, and essentially nothing has been rigorously proved (though see the very recent work of Budzinski, Contat, and Curien [20] on a simpler model of random graphs, and the numerical simulations of Bauer and Golinelli [8]).

Although our understanding of the critical Karp–Sipser process is not sufficient to prove or disprove [Theorem 1.7](#) at the critical point $c = e$, we are at least able to show (as a consequence of [Theorem 1.9](#)) that the defect in the Karp–Sipser bound is unbounded in probability for p near e/n , strongly suggesting that the situation is rather different at the critical point.

Theorem 1.10. *There is a sequence $(p_n)_{n=1}^\infty$ with $np_n \rightarrow e$, such that:*

- (A) *For $G \sim \mathbb{G}(n, p_n)$, we have $\text{corank } A(G) - i(G) \xrightarrow{P} \infty$.*
- (B) *For $G \sim \mathbb{G}(n, n, p_n)$, we have $\text{corank } B(G) - \max(i_1(G), i_2(G)) \xrightarrow{P} \infty$.*

Remark. With some more work, it seems that it would be possible to prove that for $p_n = e/n$ (or any $(p_n)_{n=1}^\infty$ for which np_n converges sufficiently rapidly to e), in the settings of both (A) and (B), the defect in the Karp–Sipser bound is whp at least of order $\log n$. See [Remark 5.4](#).

We discuss the critical regime $c = e$ further in [Section 1.6](#).

1.5. Degree-constrained random graphs. Both the Karp–Sipser core and the 2-core have minimum degree at least 2. In fact, more is true: for each of these types of cores, if we condition on the vertex set of the core, and its number of edges, then it is a *uniformly random* graph on the conditioned vertex set, with the conditioned number of edges, subject to the constraint of having minimum degree at least 2 (as we will see in [Section 4](#)).

Definition 1.11. For a set V and a positive integer $m \geq |V|$, let $\mathcal{K}(V, m, 2)$ be the uniform distribution on graphs with vertex set V , exactly m edges, and minimum degree at least 2. For a pair of sets V_1, V_2 and a positive integer $m \geq 2 \max(|V_1|, |V_2|)$, let $\mathcal{K}(V_1, V_2, m, 2)$ be the uniform distribution on bipartite graphs with vertex set $V_1 \cup V_2$, exactly m edges, and minimum degree at least 2. We write $\mathcal{K}(n, m, 2) = \mathcal{K}(\{1, \dots, n\}, m, 2)$ and $\mathcal{K}(n_1, n_2, m, 2) = \mathcal{K}(\{1, \dots, n_1\}, \{n_1 + 1, \dots, n_1 + n_2\}, m, 2)$.

The main engine driving the proofs of [Theorems 1.1](#) and [1.7](#) is the following theorem on the rank of $\mathcal{K}(n, m, 2)$ and $\mathcal{K}(n_1, n_2, m, 2)$, which may be of independent interest.

Theorem 1.12. *Fix $\varepsilon > 0$. Recall the definitions of $\gamma(\lambda), \gamma^\dagger(\lambda)$ from [Definition 1.8](#).*

- (A) *Suppose $(1 + \varepsilon)n \leq m \leq n/\varepsilon$ and let $G \sim \mathcal{K}(n, m, 2)$.*
 - (1) *Whp $\text{rank } A(G) = n - s(G)$.*
 - (2) *Suppose $2m/n$ converges to a constant $\alpha > 2$. Choose $\lambda > 0$ such that if $Z \sim \text{Poisson}(\lambda)$, then $\alpha = \mathbb{E}[Z|Z \geq 2]$. Then $s(G) \xrightarrow{d} Y + 2Y^\dagger$, where Y, Y^\dagger are independent Poisson with means $\gamma(\lambda) - 2\gamma^\dagger(\lambda)$ and $\gamma^\dagger(\lambda)$ respectively. (Here Y^\dagger captures the isolated cycles with length divisible by 4, and Y captures the other special cycles.)*
- (B) *Suppose $n_1 - n_2 \rightarrow \infty$, $n_1/n_2 \rightarrow 1$ and $(1 + \varepsilon)(n_1 + n_2) \leq m \leq (n_1 + n_2)/\varepsilon$, and let $G \sim \mathcal{K}(n_1, n_2, m, 2)$.*
 - (1) *Whp $\text{rank } B(G) = n_2 - s_2(G)$.*
 - (2) *Suppose $2m/(n_1 + n_2)$ converges to a constant $\alpha > 2$. Choose $\lambda > 0$ such that if $Z \sim \text{Poisson}(\lambda)$, then $\alpha = \mathbb{E}[Z|Z \geq 2]$. Then $s_1(G) \xrightarrow{d} Y$ and $s_2(G) \xrightarrow{d} Y$, where Y is Poisson with mean $\gamma(\lambda)$.*

1.6. Further directions. The theory of random Bernoulli matrices (i.e., adjacency matrices of $\mathbb{G}(n, p)$, biadjacency matrices of $\mathbb{G}(n, n, p)$, and closely related random matrix models) is very rich, and there are a large number of conjectures and open problems. See for example the surveys of Guionnet [40] and Vu [64, 65]. Below we mention some directions which are especially closely related to the present paper.

1.6.1. The critical regime. Perhaps the most obvious direction for further research is to improve our understanding in the critical case $c = e$. Unfortunately, our understanding of the critical Karp–Sipser core is very poor; even its typical number of vertices is unknown (though conjectures motivated by numerical simulations have been made by Bauer and Golinelli [8], and a rigorous result was recently obtained by Budzinski, Contat, and Curien [20] for a simpler model of random graphs). Also, we suspect that in the critical case, the Karp–Sipser core typically has nearly as many vertices as edges (i.e., the average degree is very close to 2), so [Theorem 1.12](#) does not apply, motivating the following question.

Question 1.13. What can we say about the typical rank of the adjacency matrix of $G \sim \mathcal{K}(n, n+t, 2)$, for $t = o(n)$? What can we say about the typical rank of the biadjacency matrix of $G \sim \mathcal{K}(n_1, n_2, 2n_1+t, 2)$, for $n_1 \geq n_2$ with $n_1 = (1 + o(1))n_2$ and $t = o(n_1)$?

We remark that if $t = O(1)$ then the special cycles may intersect each other, and the combinatorial description of the rank in [Theorem 1.12\(1\)](#) no longer holds whp. In this case there may simply not exist a description of the rank that holds whp and which can be reasonably described as “combinatorial”.

1.6.2. *The asymptotic distribution of the rank.* In [Corollary 1.3](#) we proved a central limit theorem for the rank of $\mathbb{G}(n, c/n)$ for $c < 1$ or $c > e$, complemented by upcoming work with Goldschmidt and Kreačić [\[38\]](#) in which we handle the regime $c \leq e$. We wonder whether it may also be possible to prove a *local* central limit theorem for the rank. Indeed, it seems plausible that (at least in the regime $c > e$) the techniques in [\[23\]](#) might be helpful to prove a local central limit theorem for the Karp–Sipser bound $n - i(G)$; we suspect that it would then be possible to adapt the methods in this paper to deduce a local central limit theorem for the rank of $\mathbb{G}(n, c/n)$.

However, we do *not* believe that even a coarse central limit theorem holds for the rank of $G \sim \mathbb{G}(n, n, c/n)$. Recall that $\text{corank } B(G)$ is approximately $\max(i_1(G), i_2(G))$; we believe that the asymptotic joint distribution of $i_1(G)$ and $i_2(G)$ is a nontrivial bivariate Gaussian, in which case the limiting distribution of $\text{rank } B(G)$ would be expressible in terms of the maximum of two Gaussians.

1.6.3. *Other sparse random matrix distributions.* One may wish to study more general types of sparse random matrices than $\mathbb{G}(n, p)$ and $\mathbb{G}(n, n, p)$. For example, we could fix a distribution \mathcal{L} for the nonzero entries (instead of having every nonzero entry be exactly 1). The methods in this paper are quite robust, and we believe it should be possible to handle random matrices of this type, though the notion of “special cycle” would have to be adapted accordingly (the defect in the Karp–Sipser bound would still be controlled by short cycles, but it would be more complicated to describe exactly which short cycles are relevant).

However, the methods in this paper do have some limitations: they are only suitable when an approximate rank result is available (e.g., recall that Bordenave, Lelarge and Salez [\[18\]](#) found a formula for the rank of $\mathbb{G}(n, p)$ up to $o(n)$ additive error). Our methods also do not apply to graphs with bounded degree (e.g. random regular graphs, which were recently shown to have full rank whp by Huang [\[41\]](#) and Mészáros [\[56\]](#), in breakthrough works using completely different methods to the present paper).

1.6.4. *Rank over other fields.* One may wish to study rank over fields other than \mathbb{R} (e.g., rank over \mathbb{F}_2). We do not believe that an exact combinatorial characterisation of the rank is actually possible over finite fields, because in general dependencies need not be “local” (even a dense random matrix has a nontrivial probability of being singular over \mathbb{F}_2). However, we do believe that there are typically very few “non-local dependencies”, and in particular it should still be true that the defect in the Karp–Sipser bound (for both $\mathbb{G}(n, n, c/n)$ and $\mathbb{G}(n, c/n)$, with $c \neq e$) is bounded in probability.

To prove this would require a number of modifications to our proof (for example, one should incorporate some of the techniques in [\[32\]](#), which build on ideas introduced in [\[53\]](#)). The most significant obstacle is that our proof uses spectral convergence machinery due to Bordenave, Lelarge and Salez [\[18\]](#) which is fundamentally only suitable for real rank. In the bipartite setting (i.e., for $\mathbb{G}(n, n, p)$) one can substitute machinery due to Coja-Oghlan, Ergür, Gao, Hetterich, and Rolvien [\[24\]](#), which provides asymptotic formulas for the rank of a broad class of random matrices over arbitrary fields. In the non-bipartite setting, such machinery is not yet available in appropriate generality, but an exciting first step in this direction was very recently made by van der Hofstad, Müller, and Zhu [\[63\]](#).

1.7. **Notation.** We use the notation $\delta \ll \varepsilon$ to indicate that δ is sufficiently small in terms of ε (so $1/M \ll \varepsilon$ means that M is sufficiently *large* in terms of ε , and $\varepsilon \ll 1$ means that ε is sufficiently small in absolute terms).

We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \lesssim g$ to mean that there is a constant C such that $|f(n)| \leq C|g(n)|$ for sufficiently large n . Similarly, we write $f = \Omega(g)$ or $f \gtrsim g$ to mean that there is a constant $c > 0$ such that $|f(n)| \geq c|g(n)|$ for sufficiently large n . Finally, we write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \lesssim g$ and $g \lesssim f$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f(n)/g(n) \rightarrow 0$ as $n \rightarrow \infty$. Subscripts on asymptotic notation indicate quantities that should be treated as constants.

We also use standard graph-theoretic notation. In particular, $V(G)$ and $E(G)$ denote the vertex set of a graph G , and $v(G) = |V(G)|$ and $e(G) = |E(G)|$ denote the numbers of vertices and edges. We write $G[U]$ to denote the subgraph *induced* by a set of vertices $U \subseteq V(G)$. For a vertex $v \in V(G)$,

its neighborhood (i.e., the set of vertices adjacent to v) is denoted by $N_G(v)$, and its degree is denoted $\deg_G(v) = |N_G(v)|$ (the subscript G will be omitted when it is clear from context). We also write $N_U(v) = U \cap N(v)$ and $\deg_U(v) = |N_U(v)|$ to denote the degree of v into a vertex set U .

Somewhat less standardly, in this paper all bipartite graphs will have parts indexed by 1 and 2. We write $V_1(G), V_2(G)$ for the two parts of a bipartite graph G , and write $v_1(G), v_2(G)$ for the number of vertices in each part. For a set of vectors S , we write $\dim S$ for the dimension of the span of S , and we write $\text{supp}(S)$ for the union of supports of vectors in S .

We define the *double factorial* $n!!$ to be the product of all integers from 1 to n which have the same parity as n , and we define the *falling factorial* $(n)_k = n!/(n-k)!$. For a real number x , the floor and ceiling functions are denoted $\lfloor x \rfloor = \max(i \in \mathbb{Z} : i \leq x)$ and $\lceil x \rceil = \min(i \in \mathbb{Z} : i \geq x)$. We will however sometimes omit floor and ceiling symbols and assume large numbers are integers, wherever divisibility considerations are not important. All logarithms in this paper without an explicit base are to base e , and the set of natural numbers \mathbb{N} includes zero.

1.8. Acknowledgments. We would like to thank Noga Alon for suggesting that our main result gives a linear-time algorithm for computing the rank.

2. OVERVIEW OF THE PAPER AND PROOFS

Most of the paper (all of [Sections 7 to 12](#)) is devoted to [Theorem 1.12\(1\)](#), characterising the corank of a degree-constrained random graph. Before discussing its proof, we briefly outline the reductions for the other theorems:

- For the asymptotic distribution of the corank ([Theorem 1.12\(2\)](#)): we simply need to understand the asymptotic distribution of the number of special cycles in a degree-constrained random graph ($\mathcal{K}(n, m, 2)$ or $\mathcal{K}(n_1, n_2, m, 2)$). This can be done with standard techniques (namely, we perform a method-of-moments calculation in the so-called *configuration model* for random graphs with a given degree sequence, after using standard Poisson approximation techniques to study the typical degree sequence of $\mathcal{K}(n, m, 2)$ and $\mathcal{K}(n_1, n_2, m, 2)$). The details appear in [Section 13](#).
- Regarding the 2-core ([Theorem 1.1](#)): for $G \sim \mathbb{G}(n, c/n)$ with $c > 1$, it is easy to estimate the typical number of vertices and edges in the 2-core of G (in particular, there are whp $\Omega(n)$ vertices and the average degree is $2 + \Omega(1)$). So, [Theorem 1.1](#) follows directly from [Theorem 1.12](#). The details appear in [Section 4](#).
- Regarding our main rank characterisation theorems ([Theorems 1.7 and 1.9](#)): for $G \sim \mathbb{G}(n, c/n)$ with $c > e$, the typical number of vertices and edges in the Karp–Sipser core of G were already studied in the seminal work of Karp and Sipser (again, there are whp $\Omega(n)$ vertices and the average degree is $2 + \Omega(1)$). So, in this case the conclusions of [Theorem 1.7\(A\)](#) and [Theorem 1.9\(A\)](#) again follow directly from [Theorem 1.12](#). The case $c < e$ is actually much simpler, and does not require [Theorem 1.12](#): it was shown by Aronson, Frieze and Pittel [[5](#)] that for $G \sim \mathbb{G}(n, c/n)$ with $c < e$, the Karp–Sipser core of G whp consists purely of vertex-disjoint cycles ([Lemma 4.4](#)), so the conclusions of [Theorem 1.7\(A\)](#) and [Theorem 1.9\(A\)](#) then follow simply by reasoning about the rank of adjacency matrices of cycles. In all cases, the bipartite setting (for [Theorem 1.7\(B\)](#) and [Theorem 1.9\(B\)](#)) can be handled similarly.
- [Theorem 1.10](#) (regarding the critical regime $np_n \rightarrow e$) follows from [Theorem 1.9](#), and the observation that the Poisson parameters defined in [Definition 1.8](#) blow up as $c \rightarrow e$.
- For the comparison between rank and matching number ([Theorem 1.2](#)): It turns out that [Theorem 1.2\(A2\)](#) and (B) follow directly from [Theorem 1.7](#), via certain trivial inequalities concerning $\nu(G)$ and $\sigma(G)$. For [Theorem 1.2\(A1\)](#), we combine [Theorem 1.7](#) with an exact description of the matching number of $\mathcal{K}(n, m, 2)$ due to Frieze and Pittel [[37](#)].
- [Corollary 1.3](#) (the central limit theorem for the rank) is an essentially immediate deduction from central limit theorems for the matching number due to Pittel [[60](#)] and Kreačić [[49](#)]. The details appear in [Section 13](#).

The deductions of [Theorems 1.2, 1.7, 1.9, and 1.10](#) all appear in [Section 5](#), together with various facts about the Karp–Sipser process. We also remark that [Section 3](#) contains a few basic preliminary facts that will be used throughout the paper, [Section 4](#) contains some basic facts about 2-cores and Karp–Sipser cores, and [Section 6](#) contains some general facts about degree-constrained random graphs.

Now we discuss the tools and ideas in the proof of [Theorem 1.12\(1\)](#) (restricting our attention to (A1), which is the slightly more difficult of the two settings; (B1) is proved in essentially the same way, but certain minor simplifications are possible, sketched in [Section 12](#)).

2.1. Spectral convergence. Qualitatively, [Theorem 1.12](#)(1) says that degree-constrained random graphs are very nearly nonsingular (the only obstructions to singularity are a small number of special cycles). One can obtain a much weaker result in a similar spirit using spectral convergence machinery of Bordeave, Lelarge, and Salez [18]. Specifically, there is a notion of *local weak convergence* of graphs, introduced independently by Benjamini and Schramm [12] and by Aldous and Steele [4]. In [18], it is shown that when a sequence of graphs locally weakly converges to a Galton–Watson tree, then the spectrum also converges, and one can estimate the limiting rank via a generating function associated with the Galton–Watson tree. (The fact that spectral information can be deduced from a local limit is not surprising, in light of the fact that the t -th moment of the empirical spectral distribution is precisely equal to the number of closed walks of length t .) It can be shown that if $\lim_{n \rightarrow \infty} nm_n$ converges to a limit, then the local weak limit of $\mathcal{K}(n, m_n, 2)$ is a Galton–Watson tree whose offspring distribution has an explicit (“truncated Poisson”) distribution. With a simple calculation concerning generating functions associated with truncated Poisson distributions, and a compactness argument, one can use the machinery of [18] to show that in the setting of [Theorem 1.12](#)(A1), we have $\text{corank}(G) = o(n)$ whp.

Of course, the above result is far weaker than the statement of [Theorem 1.12](#) (we hope to prove that $\text{corank}(G)$ is bounded in probability, not just that $\text{corank}(G) = o(n)$). We will make up the difference using a “rank-boosting” strategy, using tools that are traditionally used to study singularity of random matrices (in particular, tools related to the *Littlewood–Offord problem*).

2.2. The evolving rank, and the Littlewood–Offord problem. In this subsection we very briefly explain the techniques in the seminal paper of Costello, Tao, and Vu [26] (building on the original ideas of Komlós [47, 48]), who proved that *dense* random graphs are nonsingular. Roughly speaking, their approach was to reveal a random graph (say $\mathbb{G}(n, 1/2)$) in a vertex-by-vertex fashion, at each step studying how the addition of a new vertex affects the rank. They proved that if, at a given step, the corank is nonzero, then at the next step the corank will typically decrease by one. On the other hand, if the corank is already zero, then at the next step the corank will typically stay at zero. In this way, they could view the evolution of the corank as a random walk that heavily trends towards zero, and show that such random walks almost always end at zero.

In order to implement this strategy, it is necessary to understand how the rank changes when we add a new vertex. For example, if we add a new vertex v to a graph H to obtain a graph $H + v$ (and let \mathbf{x} be a random zero-one vector describing the presence of edges between v and the vertices of H), then the determinant of $A(H + v)$ can be expressed as a quadratic polynomial in \mathbf{x} (with coefficients depending on H). So, showing that $H + v$ is full-rank is tantamount to showing that a certain quadratic polynomial is nonzero. Correspondingly, an important ingredient in [26] was the fact that certain quadratic polynomials of independent random variables are unlikely to be zero.

The *Littlewood–Offord problem* studies the point probabilities of sums of independent discrete random variables. In particular, the fundamental theorem in this field is the *Erdős–Littlewood–Offord theorem*, which was used in Komlós’ foundational papers [47, 48] on discrete random matrices. To study the evolving rank of a random graph, Costello, Tao, and Vu initiated the study of the *quadratic Littlewood–Offord problem*: specifically, they proved that if an N -variable real quadratic polynomial f has $\Omega(N^2)$ nonzero coefficients, and $\mathbf{x} \in \{0, 1\}^N$ is a uniform random binary vector, then $\Pr[f(\mathbf{x}) = 0] \leq N^{-1/8}$. It turns out that in order to fully understand the evolution of the rank one needs Littlewood–Offord-type theorems of both linear and quadratic type: in order to show that the corank typically decreases when it is nonzero, one considers a linear Littlewood–Offord problem, and in order to show that the corank typically stays zero when it is zero, one considers a quadratic Littlewood–Offord problem.

A key reason for the difficulty of studying sparse random matrices is that Littlewood–Offord theorems break down in very sparse settings: if $\mathbf{x} \in \{0, 1\}^N$ is a random binary vector in which every entry is 1 with probability only c/N , it is simply not in general true that the event $f(\mathbf{x}) = 0$ is unlikely. For example, if $f(\mathbf{x}) = x_1 + \cdots + x_N$ or $f(\mathbf{x}) = (x_1 + \cdots + x_N)^2$ (in the linear and quadratic cases, respectively), then the asymptotic distribution of $f(\mathbf{x})$ is $\text{Poisson}(c)$, or the square of a $\text{Poisson}(c)$ distribution, and the point probabilities of $f(\mathbf{x})$ are of the form $\Omega(1)$. Roughly speaking, the problem is that in this very sparse regime there is “not enough randomness” in \mathbf{x} .

2.3. Rank-boosting. The key insight to overcome this issue is as follows. In the setting of [Theorem 1.12](#), while the *average* degree of G is typically only $O(1)$, whp there are at least a few vertices with much higher degree. Indeed, the maximum of n independent Poisson random variables is typically about $\log n / \log \log n$, and correspondingly it turns out that G typically has at least a few vertices of that degree. More qualitatively, for any $\beta = o(1)$, the βn highest-degree vertices all have degree $\omega(1)$.

In [33], Ferber and the last three authors leveraged this observation together with the techniques discussed in the last two subsections, to prove that the k -core of a random graph (for $k \geq 3$) is nonsingular whp. Specifically, for a random n -vertex graph constrained to have minimum degree at least k , they designed a procedure to identify βn vertices of high degree *without actually revealing the neighbours of these vertices*. They then showed that the graph induced by the remaining $(1 - \beta)n$ vertices locally weakly converges to a Galton–Watson tree, and used the machinery in [18] to prove that the corank of this graph is at most say $\beta n/2$ whp. Now, adding back the βn high-degree vertices one-by-one, and studying the evolution of the rank, at each step there is quite a lot of randomness, because each of these vertices has high degree and its neighbourhood has not yet been revealed. So, with a random walk argument together with a quadratic Littlewood–Offord theorem, they could show that at the end of this vertex-adding process the corank has decreased from $\beta n/2$ to zero whp.

At a high level, the approach in this paper is to apply the same rank-boosting strategy to $\mathcal{K}(n, m, 2)$ to prove [Theorem 1.12](#). However, the situation is far more delicate, for reasons we discuss in the following subsections.

2.4. The small-support kernel, minimal kernel vectors, and stalks. The above rank-boosting strategy cannot succeed as written, because it is simply not true that $\mathcal{K}(n, m, 2)$ is nonsingular whp (due to the possible existence of special cycles). This is due to an issue we have so far neglected to mention: for any Littlewood–Offord-type approach (in which we study the rank via events of the form $f(\mathbf{x}) = 0$), it is necessary to establish “non-degeneracy” conditions for f . For instance, we need to ensure that f has many nonzero coefficients (to see that something like this is necessary, note that if f were the zero polynomial, we would have $f(\mathbf{x}) = 0$ with probability 1, no matter how dense of a random vector \mathbf{x} is).

The polynomials f that we need to consider are defined in terms of the evolving random graph G (as we add vertices one-by-one). It turns out that if, at some point in the process, f has few nonzero coefficients, this essentially corresponds to $A(G)$ having a kernel vector with small *support* (i.e., with few nonzero entries)⁷. Therefore, an essential part of the Littlewood–Offord-based proofs mentioned so far [26, 33, 47, 48] is to prove that $A(G)$ has no small-support kernel vectors.

Crucially, this can be accomplished by purely *combinatorial* means: for example, if \mathbf{v} is a kernel vector of an adjacency matrix $A(G)$ (such that the nonzero entries of \mathbf{v} correspond to a set of vertices R , say), then when a vertex has a neighbour in R , it must in fact have at least two neighbours in R (in order for there to be a cancellation yielding zero in the corresponding entry of $A(G)\mathbf{v}$). In the settings of [26, 33, 47, 48], one can simply use a crude combinatorial union bound calculation to show that whp there are no small sets R with this property (for example, in [33], one can simply use that R and its neighbours would comprise an atypically dense set, which is unlikely to appear in a sparse random graph). Specifically, union bounds of this type can be made to work as long as $R \leq \eta n$ for some small constant η . We remark that when studying the k -core (for $k \geq 3$) in [33], it was possible to engineer the high-degree-vertex extraction in such a way that (crude union bounds show that) whp no short kernel vectors ever appear during the entire vertex-adding process.

Unfortunately, in the setting of [Theorem 1.12](#), small-support kernel vectors seem to be unavoidable: special cycles may exist in $\mathcal{K}(n, m, 2)$ itself, and since we are no longer assuming $k \geq 3$ it seems to be impossible to engineer our high-degree vertex extraction to avoid the emergence of many small-support kernel vectors during our vertex-adding process. Instead, we need to perform a very delicate calculation to upper-bound the numbers of various types of small-support kernel vectors (and in particular to show that at the end of the vertex-adding process, whp the *only* short kernel vectors are those corresponding to special cycles, and linear combinations thereof). It turns out that a naïve union bound does not suffice here, and we need to consider a notion of *minimal* kernel vectors (essentially, kernel vectors which cannot be broken down into kernel vectors with smaller support). This notion was first considered by DeMichele, the first author, and Moreira in [30]. A large part of the paper ([Section 11](#)) is spent very carefully studying the expected number of combinatorial configurations corresponding to minimal kernel vectors (called *stalks*), with support size at most ηn , in degree-constrained random graphs. (Unlike in [33], we cannot merely consider the density of a stalk; we need to very carefully consider its structure.)

2.5. Boosting the large-support kernel, and a special-purpose pseudoinverse. Due to the existence of small-support kernel vectors, the evolution of the rank no longer has such a simple description

⁷To be precise, recall that we may need to consider either linear or quadratic f , depending on the situation. In the linear case, the coefficients of f correspond precisely to a kernel vector, and in the quadratic case there is a correspondence between coefficients of f and “almost kernel vectors” of $A(G)$ (i.e., vectors \mathbf{v} such that $A(G)\mathbf{v}$ has only two nonzero entries). So, in much of what follows, we really need to consider almost kernel vectors as well as kernel vectors.

as in [26, 33]. Instead of showing that the corank drifts towards zero, we show that the corank “drifts towards the dimension of the small-support kernel”. Specifically, we prove that if there is a kernel vector with large support, then the corank decreases whp, and in any case the corank whp does not increase. It turns out that it is still possible to control the probabilities of these events via events of the form $f(\mathbf{x}) = 0$ for some linear or quadratic polynomials f , but unlike in [26, 33], we cannot define f in terms of a determinant (because if there is any kernel vector the determinant is always zero). Instead, our polynomial f is defined in terms of a special-purpose “pseudoinverse”, first (implicitly) considered in [30]. We state and prove a general purpose rank-boosting lemma, summarising a one-step application of a linear and quadratic Littlewood–Offord theorem, in Section 8.

2.6. Robust analysis of a random walk. Summing up, our approach is as follows. After extracting high-degree vertices and showing that the resulting graph has small corank (executed in Section 9), we add back the high-degree vertices one-by-one, and consider the evolution of the rank of this random graph process. Letting $\dim K_t^{(\eta)}$ be the dimension of the span of small-support kernel vectors at time t (which is a lower bound for the corank at time t), we prove an upper bound on $\mathbb{E} \dim K_t^{(\eta)}$ (in terms of t) via direct combinatorial means, and prove using Littlewood–Offord theorems that the corank trends towards $\dim K_t^{(\eta)}$. We wish to combine all these ingredients to prove that at the end of the process, whp the corank is exactly equal to $\dim K_t^{(\eta)}$ (which we then show is equal to $s(G)$).

In order to execute this plan, we need a more robust random walk analysis than in [26, 33]. The main issue is that because we no longer have the “wiggle room” afforded by the assumption $k \geq 3$, it is much harder to prove bounds that hold whp for all steps t (e.g., our bounds on $\mathbb{E} \dim K_t^{(\eta)}$, together with Markov’s inequality, do not provide strong enough probabilistic bounds for a union bound over all t). Instead, we have estimates that hold for *most* steps t , and we need a more robust analysis of random walks that can tolerate a small number of “bad steps” (as long as they are not clustered near the end of the process). We present a general lemma along these lines in Section 7, which we hope will be useful for other applications. In Section 10 we put everything together, completing the proof of Theorem 1.12(A1).

3. PRELIMINARIES

In this section we collect some basic facts that will be used throughout the paper. First, to unify the proofs for the bipartite and nonbipartite cases to the greatest extent possible, we observe that for a bipartite graph G , the rank of its biadjacency matrix is related to the rank of its adjacency matrix.

Fact 3.1. *If G is bipartite, then $\text{rank } A(G) = 2 \text{rank } B(G)$*

Proof. This follows immediately from the fact that (given an appropriate ordering of the vertices) $A(G)$ has the block representation

$$\begin{pmatrix} 0 & B(G)^\top \\ B(G) & 0 \end{pmatrix}. \quad \square$$

We will also need a Chernoff bound for binomial and hypergeometric distributions (see for example [44, Theorems 2.1 and 2.10]). Recall that the hypergeometric distribution $\text{Hyp}(N, K, n)$ is the distribution of $|Z \cap U|$, for fixed sets $U \subseteq S$ with $|S| = N$ and $|U| = K$ and a uniformly random size- n subset $Z \subseteq S$.

Lemma 3.2 (Chernoff bound). *Let X be either:*

- *a sum of independent random variables, each of which take values in $\{0, 1\}$, or*
- *hypergeometrically distributed (with any parameters).*

Then for any $\delta > 0$ we have

$$\Pr[X \leq (1 - \delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/2), \quad \Pr[X \geq (1 + \delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/(2 + \delta)).$$

Finally, we will need a consequence of the Azuma–Hoeffding inequality (see [44, Theorem 2.25]).

Lemma 3.3. *Let Z be a random variable defined in terms of a sequence of random variables X_0, \dots, X_n , such that modifying any individual X_k changes Z by at most c_k . Then*

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_k c_k^2}\right).$$

4. STRUCTURE OF CORES

In this section we collect some standard results on the 2-core and the Karp–Sipser core of a sparse random graph.

4.1. The 2-core. First, the following description of the component structure of the supercritical 2-core follows immediately from, e.g., [44, Theorem 5.12] and the main result of [31].

Lemma 4.1. *Fix a constant $c > 1$ and let $G \sim \mathbb{G}(n, c/n)$. Then whp the following holds. The giant component of G has a 2-core which is connected, has $\Omega(n)$ vertices, and has average degree $2 + \Omega(1)$. Also, all components outside the giant either have empty 2-core or their 2-core is a cycle.*

Second, the following lemma concerns the edge and vertex statistics of the supercritical 2-core (there are whp $\Omega(n)$ vertices, and the average degree is whp $2 + \Omega(1)$). It follows from, e.g., [36, Lemma 2.16].

Lemma 4.2. *Fix $c > 1$, and let $\lambda_2 = \lambda_2(c) > 0$ be the unique solution to $\lambda_2/(1 - e^{-\lambda_2}) = c$. There is $\beta = \beta(c) > 0$ such that the following holds. Let Z be a Poisson random variable with mean λ_2 , and let $G \sim \mathbb{G}(n, c/n)$. Then*

$$\frac{v(\text{core}_2(G))}{n} \xrightarrow{p} \beta, \quad \frac{2e(\text{core}_2(G))}{\beta n} \xrightarrow{p} \mathbb{E}[Z|Z \geq 2] > 2.$$

Note that [Theorem 1.1](#) directly follows from [Theorem 1.12\(A\)](#) given [Lemmas 4.1](#) and [4.2](#), as follows.

Proof of [Theorem 1.1](#). For $G \sim \mathbb{G}(n, c/n)$, it is well-known (see for example [33, Lemma 6.1]) that given the vertex set W of the 2-core and its number of edges m , we have $\text{core}_2(G) \sim \mathcal{K}(W, m, 2)$. So, by [Theorem 1.12\(A\)](#) and [Lemma 4.2](#), after deleting isolated special cycles, the asymptotic distribution of the corank is Poisson with mean $\gamma(\lambda_2) - 2\gamma^\dagger(\lambda_2)$. Considering the probability that such a Poisson random variable is equal to zero, and recalling the structural description in [Lemma 4.1](#) (whp the giant component is obtained precisely by deleting isolated cycles), the desired result follows. \square

4.2. The Karp–Sipser core. For the Karp–Sipser core, we need some results for both the supercritical ($c > e$) and subcritical ($c < e$) cases. First, for the supercritical Karp–Sipser core, we need a counterpart of [Lemma 4.2](#), and in the bipartite setting, we need the fact that the two sides of the Karp–Sipser core have quite different sizes.

Lemma 4.3. *Fix a constant $c > e$. Let $\lambda_{\text{KS}} = \lambda_{\text{KS}}(c)$ be as in [Definition 1.8](#), and let Z be a Poisson random variable with mean λ_{KS} . There is $\beta = \beta(c) > 0$ such that the following holds.*

(A) *If $G \sim \mathbb{G}(n, c/n)$ then*

$$\frac{v(\text{core}_{\text{KS}}(G))}{n} \xrightarrow{p} \beta, \quad \frac{2e(\text{core}_{\text{KS}}(G))}{\beta n} \xrightarrow{p} \mathbb{E}[Z|Z \geq 2] > 2.$$

(B) *Let $G \sim \mathbb{G}(n, n, c/n)$.*

(1) *For each $i \in \{1, 2\}$,*

$$\frac{v_i(\text{core}_{\text{KS}}(G))}{n} \xrightarrow{p} \beta, \quad \frac{e(\text{core}_{\text{KS}}(G))}{\beta n} \xrightarrow{p} \mathbb{E}[Z|Z \geq 2] > 2.$$

(2) *$|v_1(\text{core}_{\text{KS}}(G)) - v_2(\text{core}_{\text{KS}}(G))| \xrightarrow{p} \infty$.*

[Lemma 4.3\(B2\)](#) appears as [22, Lemma 7.1]. There are multiple ways to prove [Lemma 4.3\(A\)](#) and [Lemma 4.3\(B1\)](#). One classical way is to use the so-called *differential equations method* to study the likely trajectories of certain statistics associated with the Karp–Sipser process. In the setting of (A), this was done by Karp and Sipser [46] in one of the first applications of the differential equations method. Their analysis was later refined by Aronson, Frieze, and Pittel [5] (see also the discussion in [49]). A more modern approach (which arguably yields simpler proofs, though often with worse quantitative aspects) is to express the relevant statistics in terms of fixed points of a certain *warning propagation* operator. This was done by Coja–Oghlan, Cooley, Kang, Lee, and Ravelomanana [22] in the setting of (B) (specifically, [Lemma 4.3\(B1\)](#) appears as [22, Proposition 2.6]). Both approaches work equally well in the settings of (A) and (B), with very minor alterations to the proofs.

Second, in the subcritical case ($c < e$), we need the fact that the Karp–Sipser core consists only of vertex-disjoint cycles, and the numbers of cycles of each length are asymptotically jointly Poisson distributed.

Lemma 4.4. *Fix a constant $c < e$. Let $\eta \in [0, 1]$ be the unique solution to $c = \eta e^\eta$.*

(A) *If $G \sim \mathbb{G}(n, c/n)$ then whp $\text{core}_{\text{KS}}(G)$ is a collection of vertex-disjoint cycles. Let N_ℓ be the number of such cycles of length ℓ , and let $(Z_\ell)_{\ell=3}^\infty$ be a sequence of independent Poisson random variables with $\mathbb{E}Z_\ell = \eta^\ell/(2\ell)$. Then $v(\text{core}_{\text{KS}}(G))$ is bounded in probability and*

$$(N_\ell)_{\ell=3}^\infty \xrightarrow{d} (Z_\ell)_{\ell=3}^\infty.$$

(B) If $G \sim \mathbb{G}(n, n, c/n)$ then whp $\text{core}_{\text{KS}}(G)$ is a collection of vertex-disjoint cycles. Let N_ℓ be the number of such cycles of length ℓ , and let $(Z_{2k})_{k=2}^\infty$ be a sequence of independent Poisson random variables with $\mathbb{E}Z_{2k} = \eta^{2k}/(2k)$. Then $v(\text{core}_{\text{KS}}(G))$ is bounded in probability and

$$(N_{2k})_{k=2}^\infty \xrightarrow{d} (Z_{2k})_{k=2}^\infty.$$

Lemma 4.4(A) is implicit in the proof of [5, Theorem 2], and Lemma 4.4(B) can be proved in essentially the same way (as the bipartite case of Lemma 4.4 has not explicitly appeared in the literature before, we provide a brief sketch in Appendix A).

It turns out that the Karp–Sipser core enjoys the same symmetry property as the 2-core: if we condition on the vertex set W of the Karp–Sipser core, and its number of edges m , then $\text{core}_{\text{KS}}(G)$ is distributed as $\mathcal{K}(W, m, 2)$ (or $\mathcal{K}(W_1, W_2, m, 2)$, in the bipartite case, where $W = W_1 \cup W_2$).

Lemma 4.5. Consider any $0 \leq p \leq 1$.

- (A) Let $G \sim \mathbb{G}(n, p)$, let W be the vertex set of $\text{core}_{\text{KS}}(G)$ and let m be the number of edges in $\text{core}_{\text{KS}}(G)$. Then the conditional distribution of $\text{core}_{\text{KS}}(G)$ is $\mathcal{K}(W, m, 2)$.
- (B) Let $G \sim \mathbb{G}(n_1, n_2, p)$, let $W_1 \cup W_2$ be the vertex set of $\text{core}_{\text{KS}}(G)$ and let m be the number of edges in $\text{core}_{\text{KS}}(G)$. Then the conditional distribution of $\text{core}_{\text{KS}}(G)$ is $\mathcal{K}(W_1, W_2, m, 2)$.

Proof. We prove (A); the proof of (B) is similar. Consider any two graphs H, H' on the vertex set W with m edges and minimum degree at least 2. For any outcome of G yielding $\text{core}_{\text{KS}}(G) = G[W] = H$, we can simply replace $G[W]$ with H' to obtain an outcome of G yielding $\text{core}_{\text{KS}}(G) = H'$ (iterated leaf removal yields $G[W]$ in both cases). This implies that H and H' are equally likely to occur as $\text{core}_{\text{KS}}(G)$. \square

In much the same way that we were able to deduce Theorem 1.1 (on the 2-core) from the lemmas in Section 4.1 together with Theorem 1.12, we will be able to deduce Theorems 1.2 and 5.3 from the lemmas in this subsection together with Theorem 1.12. However, the deductions are not quite as immediate, so we save them for the next section.

5. KARP–SIPSER LEAF REMOVAL

In this section we make some basic observations about the Karp–Sipser leaf-removal process, and show how to deduce Theorems 1.2, 1.7, 1.9, and 1.10 from these observations together with Theorem 1.12.

It is a simple fact (first observed by Karp and Sipser [46]) that in any graph G , removing a degree-1 vertex and its neighbour reduces the matching number by exactly 1. This leaf removal also has a predictable effect on $\text{rank } A(G)$ and $\sigma(G)$, and on $\text{rank } B(G)$, if G is bipartite. (Recall that $\sigma(G)$ is the size of the largest permutation matrix “contained” in $A(G)$, where our notion of matrix containment allows deleting rows and columns, and changing 1-entries to 0-entries.)

Lemma 5.1. Fix any graph G , and delete a leaf v and its neighbour w to obtain a graph G' .

- (A) $\text{rank } A(G') = \text{rank } A(G) - 2$ and $\sigma(G') = \sigma(G) - 2$.
- (B) If G is bipartite then $\text{rank } B(G') = \text{rank } B(G) - 1$.

Proof. For (B), without loss of generality we can assume that v corresponds to the first row and w corresponds to the first column. Then, observe that

$$B(G) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ x_2 & & & \\ \vdots & & B(G') & \\ x_n & & & \end{pmatrix}$$

for some $x_2, \dots, x_n \in \{0, 1\}$. Since $B(G)$ comes from adding a zero row to the top of $B(G')$ and then adding a column, clearly $\text{rank } B(G) \leq \text{rank } B(G') + 1$. Furthermore, for every full-rank submatrix of $B(G')$, we can add the first row and column of $B(G)$ to obtain a full-rank submatrix of $B(G)$, so $\text{rank } B(G) \geq \text{rank } B(G') + 1$. The result follows.

Similarly, for (A), without loss of generality we can assume that v corresponds to the first row and column, and w corresponds to the second row and column. Then,

$$A(G) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & y_3 & \cdots & y_n \\ 0 & y_3 & & & \\ \vdots & \vdots & & A(G') & \\ 0 & y_n & & & \end{pmatrix}$$

for some $y_3, \dots, y_n \in \{0, 1\}$. For every full-rank submatrix of $A(G')$, we can add the first two rows and columns of $A(G)$ to obtain a full-rank submatrix of $A(G)$, and we similarly deduce $\text{rank } A(G) = \text{rank } A(G') + 2$. For σ , any permutation submatrix contained in $A(G')$ gives rise to a permutation submatrix in $A(G)$ with two more rows and columns, so we deduce $\sigma(G) = \sigma(G') + 2$. \square

Now we formally state the Karp–Sipser bounds on rank and matching number that were mentioned in the introduction.

Corollary 5.2. *Fix any graph G .*

- (A) $\max(\text{rank } A(G), 2\nu(G)) \leq \sigma(G) \leq v(G) - i(G)$.
- (B) *If G bipartite with vertex set $V_1 \cup V_2$ then*

$$\text{rank } B(G) \leq \nu(G) \leq \min(v_1(G) - i_1(G), v_2(G) - i_2(G)).$$

Proof. First we prove (A). Recall from the introduction (Section 1.2) that $\text{rank } A(G) \leq \sigma(G)$ and $2\nu(G) \leq \sigma(G)$. So, it suffices to prove $\sigma(G) \leq n - i(G)$. Let $v^c = v(\text{core}_{\text{KS}}(G))$. The number of leaf-removal steps in the Karp–Sipser process is $(v(G) - v^c - i(G))/2$, so by Lemma 5.1(A), we have $\sigma(G) = v(G) - v^c - i(G) + \sigma(\text{core}_{\text{KS}}(G))$. The desired result follows from the fact that $\sigma(\text{core}_{\text{KS}}(G)) \leq v^c$.

Now we prove (B). Recall from the introduction that $\text{rank } B(G) \leq \nu(G)$, so it suffices to prove that $\nu(G) \leq \min(v_1(G) - i_1(G), v_2(G) - i_2(G))$. For $i \in \{1, 2\}$, let $v_i^c = v_i(\text{core}_{\text{KS}}(G))$. The number of leaf-removal steps is $v_1(G) - v_1^c - i_1(G) = v_2(G) - v_2^c - i_2(G)$; since each leaf-removal reduces the matching number by exactly 1 we have

$$\nu(G) = v_1(G) - v_1^c - i_1(G) + \nu(\text{core}_{\text{KS}}(G)) = v_2(G) - v_2^c - i_2(G) + \nu(\text{core}_{\text{KS}}(G)).$$

The desired result then follows from the fact that $\nu(\text{core}_{\text{KS}}(G)) \leq \min(v_1^c, v_2^c)$. \square

5.1. Deductions. We now show how to deduce Theorems 1.2, 1.7, 1.9, and 1.10.

Proof of Theorems 1.7 and 1.9. First we prove (A). Let $v = v(\text{core}_{\text{KS}}(G))$ and $m = e(\text{core}_{\text{KS}}(G))$. The number of leaf-removal steps is $(n - v - i(G))/2$, so

$$\text{rank } A(G) = n - v - i(G) + \text{rank } A(\text{core}_{\text{KS}}(G))$$

by Lemma 5.1(A). For Theorem 1.7 we need to prove that $\text{rank } A(\text{core}_{\text{KS}}(G)) = v - s(\text{core}_{\text{KS}}(G))$ whp.

- If $c < e$ then by Lemma 4.4(A), whp $\text{core}_{\text{KS}}(G)$ is a vertex-disjoint union of cycles (and the number of cycles of length ℓ is asymptotically Poisson, with parameter $\eta^\ell/(2\ell)$, where η is the unique solution to the equation $c = \eta e^\eta$). It is easy to compute (see for example [59, Example 7.8]) that for a length- ℓ cycle C_ℓ we have

$$\text{rank } A(C_\ell) = \begin{cases} \ell - 2 & \text{if } \ell \text{ is divisible by 4,} \\ \ell & \text{otherwise.} \end{cases}$$

So, whp $\text{rank } A(\text{core}_{\text{KS}}(G)) = v - s(\text{core}_{\text{KS}}(G))$, proving Theorem 1.7(A). For Theorem 1.9(A), note that the defect in the Karp–Sipser bound is exactly twice the number of 4-divisible cycles $\sum_{k=1}^{\infty} N_{4k}$ (with notation as in Lemma 4.4). Recall that any sum of independent Poisson random variables is itself Poisson, and that v is bounded whp. Hence $\sum_{k=1}^{\infty} N_{4k}$ is Poisson with parameter

$$\sum_{k=1}^{\infty} \frac{\eta^{4k}}{8k} = -\frac{1}{8} \log(1 - \eta^4),$$

and the result follows.

- If $c > e$ then by Lemma 4.3(A) we have $v = \Omega(n)$ and $m/v = 1 + \Omega(1)$ and $m/v = O(1)$ whp. Conditioning on such an outcome of v, m , by Lemma 4.5 and Theorem 1.12(A) we have $\text{rank } A(\text{core}_{\text{KS}}(G)) = v - s(\text{core}_{\text{KS}}(G))$ whp, and the defect $s(\text{core}_{\text{KS}}(G))$ in the Karp–Sipser bound has the required asymptotic distribution.

Next we prove (B). For $i \in \{1, 2\}$, let $v_i = v_i(\text{core}_{\text{KS}}(G))$ and let $m = v(\text{core}_{\text{KS}}(G))$. The number of leaf-removal steps is $n - v_1 - i_1(G) = n - v_2 - i_2(G)$, so

$$\text{rank } B(G) = n - v_1 - i_1(G) + \text{rank } B(\text{core}_{\text{KS}}(G)) = n - v_2 - i_2(G) + \text{rank } B(\text{core}_{\text{KS}}(G))$$

by Lemma 5.1(B). For Theorem 1.7 we need to prove that

$$\text{rank } B(\text{core}_{\text{KS}}(G)) = \min(v_1 - s_1(\text{core}_{\text{KS}}(G)), v_2 - s_2(\text{core}_{\text{KS}}(G)))$$

whp.

- If $c < e$ then by [Lemma 4.4\(B\)](#), whp $\text{core}_{\text{KS}}(G)$ is a vertex-disjoint union of even cycles (and therefore $v_1 = v_2$). Using [Fact 3.1](#) for $G = C_{2\ell}$ we see that

$$\text{rank } B(C_{2\ell}) = \begin{cases} \ell - 1 & \text{if } \ell \text{ is divisible by } 2, \\ \ell & \text{otherwise.} \end{cases}$$

So, whp $\text{rank } B(\text{core}_{\text{KS}}(G)) = v_1 - s_1(\text{core}_{\text{KS}}(G)) = v_2 - s_2(\text{core}_{\text{KS}}(G))$, as desired. Then, [Theorem 1.9\(B\)](#) follows in essentially the same way as for [Theorem 1.9\(A\)](#), using [Lemma 4.4\(B\)](#) for the joint cycle count distribution.

- If $c > e$ then by [Lemma 4.3\(B\)](#) we have $v_1, v_2 = \Omega(n)$ and $v_1 = v_2 + o(n)$ and $|v_1 - v_2| = \omega(1)$ and $1 + \Omega(1) \leq m/(v_1 + v_2) \leq O(1)$ whp. Conditioning on such an outcome of v_1, v_2, m , the desired result follows from [Lemma 4.5](#) and [Theorem 1.12\(B\)](#) (for the i minimising v_i , the defect in the Karp–Sipser bound is exactly $s_i(\text{core}_{\text{KS}}(G))$). \square

For [Theorem 1.2\(A1\)](#) we also need a counterpart of [Theorem 1.7](#) for the matching number. Specifically, we need to know that $n - i(G) - 2\nu(G)$ is bounded in probability; this follows from an exact characterisation of $\nu(G)$ essentially due to Frieze and Pittel [\[37\]](#), as follows. For a graph G , let $q(G)$ be its number of isolated odd cycles.

Theorem 5.3. *Fix a constant $c \neq e$. For $G \sim \mathbb{G}(n, c/n)$, whp*

$$\nu(G) = \left\lfloor \frac{n - i(G) - q(\text{core}_{\text{KS}}(G))}{2} \right\rfloor.$$

Moreover, $q(\text{core}_{\text{KS}}(G))$ is bounded in probability.

Proof. Let $v = v(\text{core}_{\text{KS}}(G))$ and $m = e(\text{core}_{\text{KS}}(G))$. The number of leaf-removal steps is $(n - v - i(G))/2$, so

$$\nu(G) = (n - v - i(G))/2 + \nu(\text{core}_{\text{KS}}(G)).$$

We need to prove that $\nu(\text{core}_{\text{KS}}(G)) = \lfloor (v - q(\text{core}_{\text{KS}}(G)))/2 \rfloor$ whp.

- If $c < e$ then by [Lemma 4.4\(A\)](#), whp $\text{core}_{\text{KS}}(G)$ is a vertex-disjoint union of cycles; let N_ℓ be the number of such cycles of length ℓ . Note that $\nu(C_\ell) = \lfloor \ell/2 \rfloor$ so whp

$$\nu(\text{core}_{\text{KS}}(G)) = \sum_{\ell=3}^{\infty} N_\ell \lfloor \ell/2 \rfloor,$$

from which the desired result follows.

- If $c > e$ then by [Lemma 4.3\(A\)](#) we have $v = \Omega(n)$ and $m/v = 1 + \Omega(1)$ and $m/v = O(1)$ whp. Condition on such an outcome of v, m , and note that by [Lemma 4.5](#) we have $\text{core}_{\text{KS}}(G) \sim \mathcal{K}(v, m, 2)$. A result of Frieze and Pittel [\[37, Theorem 2\]](#), which characterises the matching number of such random graphs whp, then implies the desired result. \square

Proof of Theorem 1.2. [Theorem 1.2\(B\)](#) and (A2) follow directly from [Theorem 1.9](#), given [Corollary 5.2](#). For (A1), we simply compare the formulas in [Theorems 1.7](#) and [5.3](#). \square

Finally, we deduce [Theorem 1.10](#) from [Theorem 1.9](#) (more or less, we just need to observe that $\gamma_B(c), \gamma_A(c) \rightarrow \infty$ as $c \rightarrow e$).

Proof of Theorem 1.10. A direct computation shows that $\gamma_B(c), \gamma_A(c) \rightarrow \infty$ as $c \rightarrow e$ (specifically, $\eta \rightarrow 1$ as $c \rightarrow e$ from below, and $\lambda_{\text{KS}}(c) \rightarrow 0$ as $c \rightarrow e$ from above). Also, by Chebyshev's inequality, for $Z \sim \text{Poisson}(\gamma)$ we have $\Pr[Z \leq \gamma/2] \leq 4/\gamma$. So, for each c , if n is sufficiently large (say $n \geq n_c$), in the setting of (A) we have

$$\Pr[n - i(G) - \text{rank } A(G) < \gamma_A(c)/2] < \frac{5}{\gamma_A(c)}$$

and in the setting of (B) we have

$$\Pr[n - \max(i_1(G), i_2(G)) - \text{rank } B(G) < \gamma_B(c)/2] < \frac{5}{\gamma_B(c)},$$

by [Theorem 1.9](#). Letting $c_k = c + 1/k$ for $k \geq 1$ and $k_n = \max(k: n \geq n_{c_k})$ for all $n \geq n_{c_1}$, the desired result follows by taking $p_n = c_{k_n}/n$ for $n \geq n_{c_1}$ (and say $p_n = 0$ for $n < n_{c_1}$). \square

Remark 5.4. With more work, it seems to be possible to give an alternative (and more direct) proof of [Theorem 1.10](#) with stronger quantitative aspects. Indeed, if the leaf-removals in the Karp–Sipser process are performed one-by-one in a random order, then we obtain a randomly evolving “partial Karp–Sipser core” (which gradually shrinks over time until the final Karp–Sipser core is reached). If np_n converges sufficiently rapidly to e , then using the differential equations method as in [5], we believe that one can track the evolution of the partial Karp–Sipser core until a point where almost all vertices in the partial core have degree 2 (in the strong sense that the sum of degrees different from 2 is an $n^{-\Omega(1)}$ -fraction of the total degree sum). Then, it is not hard to see that the partial core is uniform over all graphs with its degree sequence, and it should follow from a standard configuration-model calculation that there are $\Omega(\log n)$ isolated special cycles (which will end up as isolated special cycles in the final Karp–Sipser core, and will therefore each contribute to the defect in the Karp–Sipser bound).

6. DEGREE-CONSTRAINED RANDOM GRAPHS

Most of the rest of the paper will be spent proving [Theorem 1.12](#), on the rank of degree-constrained random graphs of the form $\mathcal{K}(V, m, 2)$ and $\mathcal{K}(V_1, V_2, m, 2)$. In this section we first prove some basic properties about the degree sequence and edge distribution of such graphs.

First, a key observation is that both $\mathcal{K}(V, m, 2)$ and $\mathcal{K}(V_1, V_2, m, 2)$ are uniform given their degree sequence.

Lemma 6.1.

- (A) Consider $G \sim \mathcal{K}(V, m, 2)$ for any V, m . If we condition on an outcome of $(\deg_G(v))_{v \in V}$, then conditionally G is a uniformly random graph with this degree sequence.
- (B) Consider $G \sim \mathcal{K}(V_1, V_2, m, 2)$ for any V_1, V_2, m . If we condition on outcomes of $(\deg_G(v_1))_{v_1 \in V_1}$ and $(\deg_G(v_2))_{v_2 \in V_2}$, then conditionally G is a uniformly random bipartite graph with this pair of degree sequences.

Proof. For (A), recall that $\mathcal{K}(V, m, 2)$ is a uniform distribution on graphs satisfying certain constraints on their degrees. So, if we condition on a particular degree sequence, the resulting distribution is uniform over all graphs with that degree sequence. Similar reasoning yields (B). \square

With [Lemma 6.1](#) in hand, we can prove certain properties about $\mathcal{K}(V, m, 2)$ and $\mathcal{K}(V_1, V_2, m, 2)$ by first studying their degree sequence, then studying random graphs with given degree sequences. First, we can obtain a precise statistical understanding of the degree sequence using methods due to Cain and Wormald [21]: roughly speaking, the degree statistics can be approximated in terms of *truncated Poisson* random variables, where we take a Poisson random variable Z and condition on the event $Z \geq 2$.

Lemma 6.2. Fix a constant $\varepsilon > 0$.

- (A) For some m, n satisfying $1 + \varepsilon \leq m/n \leq 1/\varepsilon$, let $G \sim \mathcal{K}(n, m, 2)$, and choose $\lambda > 0$ such that if $Z \sim \text{Poisson}(\lambda)$, then $2m/n = \mathbb{E}[Z|Z \geq 2]$. Then the following hold with probability at least $1 - n^{-\omega(1)}$.
 - (1) For all $t \geq 2$, the number of vertices v with $\deg(v) = t$ is $\rho_t n + O_\varepsilon(\sqrt{n} \log n)$, where $\rho_t = \Pr[Z = t|Z \geq 2]$.
 - (2) $\sum_v \binom{\deg(v)}{2} = (E_2 + o_\varepsilon(1))n$, where $E_2 = \mathbb{E}\left[\binom{Z}{2} \middle| Z \geq 2\right]$ for $Z \sim \text{Poisson}(\lambda)$.
 - (3) For any $j \geq 3$, $\sum_v \binom{\deg(v)}{j} \leq e^{O_\varepsilon(j)} n$.
 - (4) For any set S of s vertices we have $\sum_{v \in S} \deg(v) \lesssim_\varepsilon s \log(2n/s)$.
- (B) For some m, n_1, n_2 satisfying $2 + \varepsilon \leq m/n_1, m/n_2 \leq 1/\varepsilon$, let $G \sim \mathcal{K}(n_1, n_2, m, 2)$ (with parts V_1, V_2), and choose $\lambda_1, \lambda_2 > 0$ such that, for $i \in \{1, 2\}$, if $Z_i \sim \text{Poisson}(\lambda_i)$, then $m/n_i = \mathbb{E}[Z_i|Z_i \geq 2]$. Then, writing $n = n_1 + n_2$, the following hold with probability at least $1 - n^{-\omega(1)}$.
 - (1) For all $i \in \{1, 2\}$ and $t \geq 2$, the number of vertices $v \in V_i$ with $\deg(v) = t$ is $\rho_t^{(i)} n_i + O_\varepsilon(\sqrt{n} \log n)$, where $\rho_t^{(i)} = \Pr[Z_i = t|Z_i \geq 2]$.
 - (2) For $i \in \{1, 2\}$ we have $\sum_{v \in V_i} \binom{\deg(v)}{2} = (E_{2;i} + o_\varepsilon(1))n_i$, where $E_{2;i} = \mathbb{E}\left[\binom{Z_i}{2} \middle| Z_i \geq 2\right]$ for $Z_i \sim \text{Poisson}(\lambda_i)$.
 - (3) For any $j \geq 3$, $\sum_v \binom{\deg(v)}{j} \leq e^{O_\varepsilon(j)} n$.

(4) For any set S of s vertices we have $\sum_{v \in S} \deg(v) \lesssim_\varepsilon s \log(2n/s)$.

To prove [Lemma 6.2](#) we need the following observation (due to Bollobás, Cooper, Fenner, and Frieze [\[14\]](#) in the non-bipartite case), that the degree sequences in the setting of [Lemma 6.2](#) can be effectively approximated by a sequence of independent truncated Poisson random variables. (We write $\text{Poisson}_{\geq 2}(\lambda)$ for the conditional distribution of $Z \sim \text{Poisson}(\lambda)$ given $Z \geq 2$.)

Lemma 6.3. Fix $\varepsilon > 0$.

(A) Consider m, n satisfying $1 + \varepsilon \leq m/n \leq 1/\varepsilon$, and choose $\lambda > 0$ such that if $Z \sim \text{Poisson}(\lambda)$ then $2m/n = \mathbb{E}[Z|Z \geq 2]$. For $V = \{1, \dots, n\}$, let $G = \mathcal{K}(V, m, 2)$ and let $(T_v)_{v \in V}$ be a sequence of independent $\text{Poisson}_{\geq 2}(\lambda)$ random variables. Then for any $\mathbf{t} \in \mathbb{N}^V$ we have

$$\Pr[(\deg_G(v))_{v \in V} = \mathbf{t}] \lesssim_\varepsilon \sqrt{n} \Pr[(T_v)_{v \in V} = \mathbf{t}].$$

(B) Consider m, n_1, n_2 satisfying $2 + \varepsilon \leq m/n_1, m/n_2 \leq 1/\varepsilon$, and choose $\lambda_1, \lambda_2 > 0$ such that, for $i \in \{1, 2\}$, if $Z_i \sim \text{Poisson}(\lambda_i)$ then $m/n_i = \mathbb{E}[Z_i|Z_i \geq 2]$. For a partition $V = V_1 \cup V_2$ into two parts of sizes n_1, n_2 , let $G \sim \mathcal{K}(V_1, V_2, m, 2)$, and let $(T_v)_{v \in V}$ be a sequence of independent truncated Poisson random variables with $T_v \in \text{Poisson}_{\geq 2}(\lambda_i)$ whenever $v \in V_i$. Then for any $\mathbf{t} \in \mathbb{N}^V$ we have

$$\Pr[(\deg_G(v))_{v \in V} = \mathbf{t}] \lesssim_\varepsilon n \Pr[(T_v)_{v \in V} = \mathbf{t}].$$

Proof sketch. Part (A) appears as [\[14, Lemma 1\]](#). It is proved by considering a random multigraph distribution (called $\mathcal{M}(n, 2m, 2)$ in [\[14\]](#); the edges are just a sequence of m independent random pairs of vertices sampled with replacement, conditioned on all degrees being at least 2) and observing that the following hold.

- If one conditions on the event that this random multigraph is simple (which occurs with probability $\Omega_\varepsilon(1)$), then one obtains the graph distribution $\mathcal{K}(V, m, 2)$.
- The degree sequence $(\deg_G(v))_{v \in V}$ of this random multigraph has precisely the conditional distribution of $(T_v)_{v \in V}$ given $\sum_v T_v = 2m$ (in [\[14\]](#) this conditional distribution is called $\mathcal{O}(n, \lambda, 2)$, and the unconditional distribution is called $\mathcal{P}(n, \lambda, 2)$). Moreover, the event $\sum_v T_v = 2m$ occurs with probability $\Omega_\varepsilon(1/\sqrt{n})$. (Roughly speaking, this is because $\sum_v T_v$ has standard deviation $O_\varepsilon(\sqrt{n})$, and is more-or-less uniform over integers within standard-deviation-range of the mean.)

For part (B), we can consider an analogous *bipartite* random multigraph distribution on the vertex set $V_1 \cup V_2$: consider m independent random edges between V_1 and V_2 , conditioned on all degrees being at least 2. Then, we analogously observe that the following hold.

- If one conditions on the event that this random multigraph is simple (which occurs with probability $\Omega_\varepsilon(1)$), then one obtains the graph distribution $\mathcal{K}(V_1, V_2, m, 2)$.
- The degree sequence $(\deg_G(v))_{v \in V_1 \cup V_2}$ of this random multigraph has precisely the conditional distribution of $(T_v)_{v \in V_1 \cup V_2}$ given $\sum_{v \in V_1} T_v = \sum_{v \in V_2} T_v = m$. Moreover, the event $\sum_{v \in V_1} T_v = \sum_{v \in V_2} T_v = m$ occurs with probability $\Omega_\varepsilon((1/\sqrt{n}) \cdot (1/\sqrt{n})) = \Omega_\varepsilon(1/n)$. \square

Now we prove [Lemma 6.2](#).

Proof of Lemma 6.2. We just prove (A); the proof of (B) is essentially identical. Let $V = \{1, \dots, n\}$ and let $(T_v)_{v \in V}$ be a sequence of independent $\text{Poisson}_{\geq 2}(\lambda)$ random variables. By [Lemma 6.3](#), to prove that a property of the degree sequence $(\deg_G(v))_{v \in V}$ holds with probability $1 - n^{-\omega(1)}$, it suffices to prove that $(T_v)_{v \in V}$ satisfies this property with probability $1 - n^{-\omega(1)}$. So, we work only with $(T_v)_{v \in V}$.

First, for each t , we have $\Pr[T_v = t] = \rho_t$ for each v independently. So, (1) holds with the desired probability, by a Chernoff bound and a union bound. Also, by a Chernoff bound for the Poisson distribution (see for example [\[58, Theorem 5.4\]](#)), for each v and $t \geq \lambda$ we have

$$\Pr[T_v \geq t] \leq (\Omega_\varepsilon(t))^{-t}. \tag{6.1}$$

This (together with the union bound) implies that with probability $1 - n^{-\omega(1)}$ we have say

$$T_v \leq \log n \text{ for all } v \in V. \tag{6.2}$$

If (1) and (6.2) hold, then trivially (4) holds whenever say $s \leq n^{0.9}$. For the case $s > n^{0.9}$, note that (when (1) and (6.2) hold) for any $t \geq \lambda$ we have

$$\sum_{d_v \geq t} d_v \leq \sum_{j=t}^{\log n} j((\Omega_\varepsilon(j))^{-j} n + \log n \cdot O_\varepsilon(\sqrt{n})) \leq (\Omega_\varepsilon(t))^{-t} n + \sqrt{n}(\log n)^4.$$

Taking $t = C \log(2n/s)$ for $1/C \ll \varepsilon$, we have that

$$\sum_{v \in S} d_v \leq ts + \sum_{d_v \geq t} d_v \leq ts + (s/n)n + \sqrt{n}(\log n)^4 \leq O_\varepsilon(s \log(n/s))$$

so (4) holds in this case too. Next, note that

$$\mathbb{E} \left[\sum_v \binom{T_v}{2} \right] = nE_2,$$

but (2) does not immediately follow from an off-the-shelf concentration inequality, since $\text{Poisson}_{\geq 2}(\lambda)$ is a distribution with unbounded support. Let

$$P = \sum_v \binom{\min(\log n, T_v)}{2}.$$

Recalling (6.2), it suffices to prove that (with probability $1 - n^{-\omega(1)}$) P satisfies the estimate in (2). To this end, note that

$$\begin{aligned} \mathbb{E} \left[\binom{T_v}{2} - \binom{\min(\log n, T_v)}{2} \right] &\leq \sum_{t=\lceil \log n \rceil}^{\infty} O(t^2) \Pr[T_v \geq t] \\ &\leq \sum_{t=\lceil \log n \rceil}^{\infty} O(t^2)(\Omega_\varepsilon(t))^{-t} = o_\varepsilon(1), \end{aligned}$$

by (6.1), so $\mathbb{E}P = (E_2 + o(1))n$. Also, note that changing some T_v changes P by at most $(\log n)^2$, so by the Azuma–Hoeffding inequality (see Lemma 3.3), we have $|P - \mathbb{E}P| \leq n^{1/2+o(1)}$ with probability $1 - n^{-\omega(1)}$, from which (2) follows.

Finally, we prove (3). By (6.2), it suffices to consider the case $3 \leq j \leq \log n$. Then, when (1) and (6.2) hold, using (6.1), we have

$$\begin{aligned} \sum_v T_v^j &= \int_0^\infty |\{v: T_v^j \geq t\}| dt = \int_0^\infty |\{v: T_v \geq s\}| j s^{j-1} ds \\ &= \int_0^{\log n} ((\Omega_\varepsilon(s))^{-s} n + \sqrt{n}(\log n)^3) j s^{j-1} ds + O_\varepsilon(n) \\ &= n \int_0^\infty j s^{j-1} (\Omega_\varepsilon(s))^{-s} ds + O_\varepsilon(n + \sqrt{n}(\log n)^3 \cdot (\log n)^j). \end{aligned}$$

Now, for any $c > 0$ we have

$$\int_0^\infty j s^{j-1} (cs)^{-s} ds = (O_c(j))^j.$$

Also note that for any $x > 0$ we have $(x/j)^j \leq \exp(x/e)$. Taking $x = \log n/100$, it follows that $(\log n)^j \leq n^{1/3}(O(j))^j$. We deduce $\sum_v T_v^j \leq n(O_\varepsilon(j))^j$. Since $j! \geq (\Omega(j))^j$ by Stirling’s approximation, we then deduce $\sum_v T_v^j/j! \leq e^{O_\varepsilon(j)}n$, from which (3) follows. \square

For disjoint sets V_1, V_2 and sequences $\mathbf{d}^1 \in \mathbb{N}^{V_1}, \mathbf{d}^2 \in \mathbb{N}^{V_2}$, write $\mathbb{G}(\mathbf{d}^1, \mathbf{d}^2)$ to denote the uniform distribution on bipartite graphs with degree sequence specified by $(\mathbf{d}^1, \mathbf{d}^2)$. For a set V and a sequence $\mathbf{d} \in \mathbb{N}^V$, write $\mathbb{G}(\mathbf{d})$ to denote the uniform distribution on graphs with degree sequence \mathbf{d} . Now, to work with random graphs of the form $\mathbb{G}(\mathbf{d}), \mathbb{G}(\mathbf{d}^1, \mathbf{d}^2)$, we use an auxiliary random graph model called the *configuration model*. This model was first explicitly considered in 1980 by Bollobás [15] (though similar ideas were considered earlier by various authors [10, 11, 68]), and has since become an indispensable tool in random graph theory.

Definition 6.4. For a degree sequence $\mathbf{d} = (d_1, \dots, d_n)$, consider a set of $r = d_1 + \dots + d_n$ “stubs”, grouped into n labelled “buckets” of sizes d_1, \dots, d_n . A *configuration* is a perfect matching on the r stubs, consisting of $r/2$ disjoint edges. Given a configuration, contracting each of the buckets to a single vertex gives rise to a multigraph with degree sequence d_1, \dots, d_n (where we use the convention that loops contribute 2 to the degree of a vertex).

- (A) For a set V and a degree sequence $\mathbf{d} \in \mathbb{N}^V$, let $\mathbb{G}^*(\mathbf{d})$ be the random multigraph distribution obtained by contracting a uniformly random configuration.

- (B) For disjoint sets V_1, V_2 and a pair of sequences $\mathbf{d}^1 \in \mathbb{N}^{V_1}, \mathbf{d}^2 \in \mathbb{N}^{V_2}$, let $\mathbb{G}^*(\mathbf{d}^1, \mathbf{d}^2)$ be the random bipartite multigraph distribution obtained by contracting a uniformly random configuration in which we only allow edges between the buckets corresponding to V_1 and the buckets corresponding to V_2 .

The uniform models $\mathbb{G}(\mathbf{d}), \mathbb{G}(\mathbf{d}^1, \mathbf{d}^2)$ can be closely compared with their configuration models, as follows.

Lemma 6.5. *Fix $C > 0$.*

- (A) (1) For any set V and sequence $\mathbf{d} \in \mathbb{N}^V$, if we consider $G^* \in \mathbb{G}^*(\mathbf{d})$ and condition on G^* being a simple graph, then we recover the distribution $\mathbb{G}(\mathbf{d})$.
(2) If the sum of squares of entries of \mathbf{d} is at most Cn , then the probability that G^* is simple is $\Omega_C(1)$.
- (B) (1) For any disjoint sets V_1, V_2 and sequences $\mathbf{d}^1 \in \mathbb{N}^{V_1}, \mathbf{d}^2 \in \mathbb{N}^{V_2}$, if we consider $G^* \in \mathbb{G}^*(\mathbf{d}^1, \mathbf{d}^2)$ and condition on G^* being a simple graph, then we recover the distribution $\mathbb{G}(\mathbf{d}^1, \mathbf{d}^2)$.
(2) If the sum of squares of entries of \mathbf{d} is at most Cn , then the probability that G^* is simple is $\Omega_C(1)$.

Parts (A1) and (B1) follow from the (easy) fact that each simple graph corresponds to the same number of configurations. Parts (A2) and (B2) of Lemma 6.5 were first proved by Janson [42] and Blanchet and Stauffer [13], respectively. We remark that Janson [43] later gave a simplified proof for both (A2) and (B2), and that many authors previously proved various special cases (see for example [10, 11, 15, 17, 39, 54, 54, 55]). Several of these special cases are sufficient for the applications in this paper.

The advantage of the configuration model is that it has much more independence than a uniformly random graph with a given degree sequence, and is therefore much easier to study.

We finish this section with a simple expansion estimate for random graphs with given degree sequences.

Lemma 6.6. *Fix $\varepsilon > 0$, and consider one of the following two situations.*

- (A) Suppose m, n satisfy $1 + \varepsilon \leq m/n \leq 1/\varepsilon$, and let $G \sim \mathcal{K}(n, m, 2)$.
(B) Suppose m, n, n_1, n_2 satisfy $2 + \varepsilon \leq m/n_1, m/n_2 \leq 1/\varepsilon$, let $G \sim \mathcal{K}(n_1, n_2, m, 2)$, and let $n = n_1 + n_2$.

In both situations, the following properties hold for n large.

- (1) There exists $C = C_{6.6}(\varepsilon) > 0$ such that the following holds (for large enough n). With probability at least $1 - O(1/\sqrt{n})$: for all s , every subgraph of G with s vertices has at most $s + \lfloor Cs/\sqrt{\log(2n/s)} \rfloor$ edges.
(2) Whp, the number of cycles of length less than $\log \log n$ is at most $\exp((\log \log n)^3)$.

We did not attempt to prove the absolute best bounds possible; for example, with more care, in the setting of (1) it seems one can prove an upper bound of roughly $s + s/\log(n/s)$ when s is not too large.

Remark 6.7. In practice, we will apply (1) in the case where s/n is small with respect to ε . So, $\lfloor Cs/\sqrt{\log(n/s)} \rfloor$ can be thought of as a ‘‘lower order term’’ relative to s . In particular, when (say) $s < \sqrt{\log n}/(2C)$, we have $\lfloor Cs/\sqrt{\log(n/s)} \rfloor = 0$, meaning that no set of s vertices has more than s edges (i.e., we cannot have anything denser than a cycle).

Proof. We only prove (A); the proof of (B) is essentially identical. We handle (1) and (2) together, considering what happens more generally for a set of size s with at least $s + t$ edges.

Let $\mathbf{d} = (d_1, \dots, d_n)$ be the degree sequence of G , and condition on any outcome of \mathbf{d} satisfying the conclusion of Lemma 6.2(A). By Lemmas 6.1 and 6.5, it suffices to prove the desired result for $G \sim \mathbb{G}^*(\mathbf{d})$ (note that Lemma 6.2(A2) implies that $d_1^2 + \dots + d_n^2 = O_\varepsilon(n)$).

Consider any $t \leq 2s$ with say $s \leq n/6$. For a set S of s vertices, the probability that $G[S]$ contains at least $s + t$ edges is at most

$$\binom{\sum_{v \in S} d_v}{2s + 2t} \left(\frac{6s}{n} \right)^{s+t}.$$

Indeed, we are considering the probability of the event that there is some set of $2s + 2t$ stubs from the buckets corresponding to vertices in S , which all pair among themselves (in our random configuration). For any set of $2s + 2t$ stubs from S , the probability that they all pair among themselves is at most $(6s/n)^{s+t}$, since $2s + 2t \leq 6s$ and $2m - (2s + 2t) \geq 2n - 6s \geq n$.

So, the expected number of sets of s vertices with at least $s + t$ edges is at most

$$\begin{aligned} \sum_{\substack{S \subseteq \{1, \dots, n\}: \\ |S|=s}} \left(\frac{\sum_{v \in S} d_v}{2s + 2t} \right) \left(\frac{6s}{n} \right)^{s+t} &\leq \left(\frac{n}{s} \right)^s \left(\frac{s \log(2n/s)}{s+t} \right)^{2s+2t} \left(\frac{s}{n} \right)^{s+t} e^{O_\varepsilon(s)} \\ &\leq \left(O_\varepsilon \left(\log \left(\frac{n}{s} \right) \right) \right)^{O(s)} \left(\frac{s}{n} \right)^t, \end{aligned} \quad (6.3)$$

where in the first inequality we used that $\binom{n}{s} \leq (en/s)^s$ and we used [Lemma 6.2\(A4\)](#) (which says that $\sum_{v \in S} d_v \lesssim_\varepsilon s \log(2n/s)$). We immediately deduce (2), taking $t = 0$ and summing over $s < \log \log n$.

For (1), let $t = \lfloor s/\sqrt{\log(2n/s)} \rfloor + 1$; we will prove that for sufficiently small $c = c(\varepsilon) > 0$, every subgraph with $s \leq cn$ vertices has fewer than $s + t$ edges. The desired result will then follow, noting that (1) trivially holds for subgraphs with $s > cn$ vertices (taking C large in terms of c).

So, we sum the estimate in (6.3) over all $s \leq cn$. The contribution from say $s \leq (\log n)^{2/3}$ is $n^{-1+o_\varepsilon(1)} \leq 1/(2\sqrt{n})$, and the contribution from $s > 2C\sqrt{\log n}$ is at most

$$\begin{aligned} &\sum_{s=\lfloor (\log n)^{2/3} \rfloor + 1}^{\lfloor cn \rfloor} \left(O_\varepsilon \left(\log \left(\frac{n}{s} \right) \right) \right)^s \left(\frac{s}{n} \right)^{\Omega(s/\sqrt{\log(n/s)})} \\ &\leq n \max_{cn \geq s \geq (\log n)^{2/3}} \exp \left(s \left(O_\varepsilon(1) + O \left(\log \log \left(\frac{n}{s} \right) \right) - \Omega \left(\sqrt{\log \left(\frac{n}{s} \right)} \right) \right) \right) \\ &\leq n \max_{cn \geq s \geq (\log n)^{2/3}} \exp \left(-\Omega_\varepsilon(s\sqrt{\log(n/s)}) \right) \leq \frac{1}{2\sqrt{n}} \end{aligned}$$

for small enough c . □

7. RANDOM WALK ANALYSIS

The following lemma is a slight adaptation of [[33](#), Lemma 5.2] (which is itself a variation on [[26](#), Lemma 2.9]). Roughly speaking, it says that certain negatively biased random walks typically end up at a nonpositive value.

Lemma 7.1. *Fix $C, \delta, \varepsilon > 0$. Let X_N, \dots, X_0 be a sequence of real random variables satisfying the following conditions for some $p \in (0, 1)$.*

W1' $X_N \leq (1 - \varepsilon)\delta N$ (with probability 1)

W2' $X_t \leq X_{t+1} + C$ for all $t \leq N$ (with probability 1).

W3' For any $t \leq N - 1$ and any x_N, \dots, x_{t+1} :

(a) if $x_{t+1} > 0$ then $\Pr[X_t \leq x_{t+1} - \delta \mid X_N = x_N, \dots, X_{t+1} = x_{t+1}] \geq 1 - p$.

(b) if $x_{t+1} \leq 0$ then $\Pr[X_t \leq 0 \mid X_N = x_N, \dots, X_{t+1} = x_{t+1}] \geq 1 - p$.

Then

$$\Pr[X_0 > 0] \leq O_{C, \delta, \varepsilon}(p).$$

Informally, condition **W3'** says that our random walk “wants to be nonpositive”: when we are positive we tend to go down at the next step, and when we are nonpositive we tend to stay nonpositive at the next step.

Proof. First, note that the statement is trivial if say $N \leq 10$. Then, note that we can reduce to the case where $\delta = 1$ and each X_t is a nonnegative integer. Indeed, define $X'_t = \max(0, \lfloor X_t/\delta \rfloor)$. Note that conditions **W1'** to **W3'** are still satisfied for X'_N, \dots, X'_0 (with “1” in place of “ δ ”, with “ $\lceil C/\delta + 1 \rceil$ ” in place of “ C ”, and with say $\varepsilon/2$ in place of ε), and note that $X_0 > 0$ implies $X'_0 > 0$. After this reduction, the lemma statement is a slight variant of [[33](#), Lemma 5.2] (and can be proved in the same way). □

We will need a generalisation of [Lemma 7.1](#) permitting a small number of “bad steps” in which we have no control over the behaviour of our random walk. Crucially, the set of bad steps is allowed to depend on the trajectory of the random walk; we only assume that the bad steps are unlikely to concentrate near the end of the walk.

Theorem 7.2. *Fix $C, \delta, \varepsilon > 0$ with $C/\delta \geq 1$. Let X_N, \dots, X_0 be a sequence of real random variables, and let $R \subseteq \{0, \dots, N\}$ be a random set of “bad steps”, satisfying the following conditions for some $p \in (0, 1)$.*

W0 There is an underlying sequence of random elements G_N, \dots, G_0 , such that X_t and the event $\{t \in R\}$ are both determined by G_t (for all $t \leq N$).

- W1** $X_N \leq (1 - \varepsilon)\delta N$ (with probability 1)
W2 $X_t \leq X_{t+1} + C$ for all $t \leq N$ (with probability 1).
W3 For any $t \leq N - 1$ and any $x_{t+1} \in \mathbb{R}$, and any outcomes g_N, \dots, g_{t+1} of G_N, \dots, G_{t+1} satisfying $X_{t+1} = x_{t+1}$:
(a) if $x_{t+1} > 0$ then $\Pr[X_t \leq x_{t+1} - \delta \text{ or } t \in R \mid G_N = g_N, \dots, G_{t+1} = g_{t+1}] \geq 1 - p$.
(b) if $x_{t+1} \leq 0$ then $\Pr[X_t \leq 0 \text{ or } t \in R \mid G_N = g_N, \dots, G_{t+1} = g_{t+1}] \geq 1 - p$.
W4 With probability at least $1 - p$, for each $t \leq N$ we have $|R \cap \{0, \dots, t\}| \leq \delta t / (32C)$.

Then

$$\Pr[X_0 > 0] \leq O_{C, \delta, \varepsilon}(p).$$

Proof. As with [Lemma 7.1](#), we can assume that $\delta = 1$, each X_t is a nonnegative integer, and C is a nonnegative integer, if we prove the result under a slight weakening of condition **W4** that $|R \cap \{0, \dots, t\}| \leq \delta t / (8C) = t / (8C)$. To see this, take $X'_t = \max(0, \lceil X_t / \delta \rceil)$ and then note that conditions **W0** to **W3** and the weaker version of **W4** are still satisfied for X'_N, \dots, X'_0 (with “1” in place of “ δ ”, with “ $\lceil C / \delta + 1 \rceil$ ” in place of “ C ”, and with say $\varepsilon/2$ in place of ε). We will deduce [Theorem 7.2](#) from [Lemma 7.1](#).

Inductively define a sequence Y_N, \dots, Y_0 by taking $Y_N = 0$ and

$$Y_t = Y_{t+1} - (C + 1)\mathbb{1}_{t \in R} + (1/2)\mathbb{1}_{Y_{t+1} < 0}$$

for $t < N$. The reader may wish to imagine a “cost” of $C + 1$ being incurred at every bad step, and that this is repaid over the future of the process (specifically, $1/2$ is repaid per step, until all debts are repaid). Define a modified sequence X'_N, \dots, X'_1 by $X'_t = X_t + Y_t$. The idea is that Y_t “compensates” if t is a bad step. Notice that Y is half-integral, and non-positive.

Note that $X'_N = X_N$ and $X'_t \leq X'_{t+1} + C + 1/2$ for all $t \leq N$ (so our modified sequence still satisfies a version of properties **W1** and **W2** in [Lemma 7.1](#)). We now verify that X'_t satisfies **W3** in [Lemma 7.1](#) with $\delta = 1/2$. Note that if $t \in R$, this is immediate, since $Y_t \leq Y_{t+1} - C - 1/2$ and $X_t \leq X_{t+1} + C$. If $t \notin R$ and $X_{t+1} > 0$, then with probability at least $1 - p$, we have $X_t \leq X_{t+1} - 1$, and deterministically, $Y_t \leq Y_{t+1} + 1/2$. Finally, if $t \notin R$ and $X_{t+1} = 0$, then it must be the case that $X'_{t+1} \leq 0$, since Y is non-positive. Then with probability at least $1 - p$, X_t stays 0, and thus $X'_t \leq 0$.

Applying [Lemma 7.1](#), we see that $\Pr[X'_0 > 0] \leq O_{C, \delta, \varepsilon}(p)$, and it suffices to prove that $Y_0 = 0$ (i.e., that $X_0 = X'_0$) with probability at least $1 - p$.

To see this, for each $i = 0, 1, \dots, \lfloor \log_2(N + 2) \rfloor$ let Q_i be the number of bad steps $t \in R$ in the range $[2^i - 1, 2^{i+1} - 1)$. Note that if the inequality $(C + 1)Q_i < (1/2)2^{i-1}$ holds for all i , then $Y_0 \geq 0$: for each $i \geq 1$, the “cost” incurred in the interval $[2^i - 1, 2^{i+1} - 1)$ is “repaid” in the 2^{i-1} steps in the interval $[2^{i-1} - 1, 2^i - 1)$. The above is guaranteed by our weaker version of **W4**, taking $t = 2^{i+1} - 2$ for each i . \square

8. BOOSTING THE RANK

In this section we prove some general lemmas studying how the rank of a matrix changes when a random row/column is added to it. The lemmas in this section represent the main difference between parts (A) and (B) of [Theorem 1.12](#): there are certain additional dependencies involved when dealing with symmetric random matrices.

First, the following simple lemma will be used for [Theorem 1.12\(B\)](#).

Definition 8.1. The λ -level set of a vector is the set of all entries equal to λ . Say that a vector $\mathbf{v} \in \mathbb{R}^n$ is η -balanced if all of the level sets of \mathbf{v} have size at most $n(1 - \eta)$.

Lemma 8.2. Fix $0 < \eta < 1/2$, let $d \geq 1$, and consider a matrix $B \in \mathbb{R}^{n_1 \times n_2}$. Consider a subset $E \subseteq \{1, \dots, n_1\}$ with size at least $n_1(1 - \eta/3)$, and let $\mathbf{x} = (x_1, \dots, x_{n_1})$ be a random zero-one vector, such that the restriction \mathbf{x}_E to the entries indexed by E is a uniformly random zero-one vector with exactly d ones (and the restriction $\mathbf{x}_{\bar{E}}$ to entries not indexed by E is deterministic). Add \mathbf{x} as a new column of B to obtain a new matrix B' .

If B^\top has an η -balanced kernel vector, then $\text{rank}(B') \geq \text{rank}(B) + 1$ with probability $1 - O_\eta(1/\sqrt{d})$.

[Lemma 8.2](#) concerns the addition of a new random column, but of course it symmetrically applies to the addition of a new random row (we can simply consider the transpose of B). In fact, in our proof of [Theorem 1.12\(B\)](#) we will use [Lemma 8.2](#) to show that when a random row and a random column are independently added, the rank increases by 2.

The following more sophisticated lemma will be used for [Theorem 1.12\(A\)](#).

Definition 8.3. Say that a symmetric matrix $A \in \mathbb{R}^{n \times n}$ is η -unstructured if there are at least ηn^2 pairs of distinct entries $(i, i') \in \{1, \dots, n\}^2$ such that if $\text{supp}(A\mathbf{v}) = \{i, i'\}$ then \mathbf{v} is η -balanced.

Lemma 8.4. Fix $0 < \eta < 1/2$ and let $d \geq 1$. Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ and a subset $E \subseteq \{1, \dots, n\}$ of size at least $n(1 - \eta/3)$. Let $\mathbf{x} = (x_1, \dots, x_n)$ be a random zero-one vector, such that the restriction \mathbf{x}_E to the entries indexed by E is a uniformly random zero-one vector with exactly d ones (and the restriction $\mathbf{x}_{\bar{E}}$ to entries not indexed by E is deterministic). Add \mathbf{x} as a new row and column of A (and put a zero in the new diagonal entry) to obtain a new symmetric matrix A' .

- (a) If A has a η -balanced kernel vector, then $\text{rank}(A') = \text{rank}(A) + 2$ with probability $1 - O_\eta(1/\sqrt{d})$.
- (b) Let $S = \text{supp}(\ker A)$. If A is η -unstructured and $|S| \leq \eta^2 n/32$ and $S \cap E = \emptyset$ and $\mathbf{x}_S = \mathbf{0}$ then $\text{rank}(A') \geq \text{rank}(A) + 1$ with probability at least $1 - O_\eta((\log 2d)^{O(1)}/\sqrt{d})$.

First, Lemma 8.2 and Lemma 8.4(a) will be simple consequences of an anti-concentration inequality for linear polynomials on the ‘‘Boolean slice’’ (i.e., for linear polynomials of uniformly random binary vectors with a prescribed number of 1s). The following lemma is a direct consequence of [34, Lemma 4.2] (a similar inequality also appears in [52]), and is proved using the Erdős–Littlewood–Offord theorem (see for example [61, Chapter 7]).

Lemma 8.5. Let $1 \leq d \leq n/2$, $\kappa > 0$ and let $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ be a κ -balanced vector. Let $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$ be a random vector, uniformly selected from the zero-one vectors with exactly d ones, and consider any $y \in \mathbb{R}$. Then

$$\Pr[\mathbf{v}^\top \mathbf{x} = y] = O((\kappa d)^{-1/2}).$$

Now we prove Lemma 8.2 and Lemma 8.4(a).

Proof of Lemma 8.2. Let $\mathbf{v} \in \mathbb{R}^{n_1}$ be an η -balanced kernel vector of B^\top (note that this means that \mathbf{v} lies in the orthogonal complement of the column space of B). Since \mathbf{v} is η -balanced and $E \geq (1 - \eta/3)n_1$, every level set of \mathbf{v}_E has size at most $((1 - \eta)/(1 - \eta/3))|E|$, implying that \mathbf{v}_E is $\Omega(\eta)$ -balanced. Therefore it follows by Lemma 8.5 that $\mathbb{P}[\mathbf{v}^\top \mathbf{x} = 0] = \mathbb{P}[\mathbf{v}_E^\top \mathbf{x}_E = -\mathbf{v}_{\bar{E}}^\top \mathbf{x}_{\bar{E}}] = O_\eta(1/\sqrt{d})$. But note that if $\mathbf{v}^\top \mathbf{x} \neq 0$, then \mathbf{x} does not lie in the column space of A , and its addition as a new column increases the rank. \square

Proof of Lemma 8.4(a). Let \mathbf{v} be an η -balanced kernel vector of A (or equivalently, of A^\top), so \mathbf{v} lies in the orthogonal complement of the row space of A (or equivalently, the orthogonal complement of the column space). As above, \mathbf{v}_E is $\Omega(\eta)$ -balanced, so by Lemma 8.5 we have $\mathbb{P}[\mathbf{v}^\top \mathbf{x} = 0] = O_\eta(1/\sqrt{d})$. If $\mathbf{v}^\top \mathbf{x} \neq 0$, then \mathbf{x} does not lie in the row space or column space of A , so adding \mathbf{x} as a new row and column increases the rank twice. The desired result follows. \square

For Lemma 8.4(b) we need an anti-concentration inequality for quadratic polynomials of random vectors $(x_1, \dots, x_n) \in \{0, 1\}^n$ on the Boolean slice. The following lemma appears as [33, Proposition 3.4], and is proved using an inequality of Kane [45].

Lemma 8.6. Let $M = (m_{ij})_{i,j}$ be an $n \times n$ symmetric matrix for which there are $\Omega(n^4)$ different 4-tuples (i, i', j, j') with $m_{ij} - m_{i'j} - m_{ij'} + m_{i'j'} \neq 0$. Let $\mathbf{x} = (x_1, \dots, x_n)$ be a random zero-one vector, uniformly selected from the zero-one vectors with exactly $d \leq n/2$ ones. Then for any vector $\mathbf{v} \in \mathbb{R}^n$ and any $x \in \mathbb{R}$ we have

$$\Pr[\mathbf{x}^\top M \mathbf{x} + \mathbf{v}^\top \mathbf{x} = x] \leq O((\log 2d)^{O(1)}/\sqrt{d}).$$

We also need the following lemma implicit in the work of DeMichele, the first author, and Moreira [30], on the existence of a certain kind of ‘‘pseudoinverse’’.

Lemma 8.7. Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$, let $S = \text{supp}(\ker(A))$ and let $\bar{S} = \{1, \dots, n\} \setminus S$. Let $P \in \mathbb{R}^{n \times n}$ be the projection matrix that projects onto the coordinates indexed by \bar{S} (that is, P is a diagonal matrix with ‘‘1’’ in the diagonal entries indexed by \bar{S} , and ‘‘0’’ in the entries indexed by S). Then there is a matrix $B \in \mathbb{R}^{n \times n}$ such that $AB = P$. Further, PBP is symmetric.

Proof. If $i \notin S$, then there is no kernel vector of A with nonzero i -coordinate, which means that the i th row of A cannot be expressed as a linear combination of the other rows of A . This means there is some vector \mathbf{b}_i which is orthogonal to every row of A except the i th (and by rescaling we can assume that the inner product of \mathbf{b}_i with the i th row of A is exactly 1). Let B be the matrix whose i th column is \mathbf{b}_i , for $i \notin S$, and whose columns indexed by S are all-zero. Writing \mathbf{e}_i for the i th standard basis vector, we have $\mathbf{e}_i^\top AB = \mathbf{e}_i$ for $i \notin S$, and $\mathbf{e}_i^\top AB = \mathbf{0}$ for $i \in S$, from which it follows that $AB = P$.

Finally, since A is symmetric, transposition yields $B^\top A = P$. We then have $PB = (B^\top A)B = B^\top (AB) = B^\top P = (PB)^\top$, so $PB = PBP$ is symmetric. \square

A key property of this notion of pseudoinverse is that it provides a sufficient condition for the rank to increase when we add a row and column to a matrix, as follows.

Lemma 8.8. *Consider a matrix $A \in \mathbb{R}^{n \times n}$, and let $S = \text{supp}(\ker(A))$, let $\bar{S} = \text{supp}(\ker(A))$, and let B be a “pseudoinverse” as guaranteed in Lemma 8.7. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, let A' be the matrix obtained from A by appending the new column \mathbf{x} , the new row \mathbf{y} and the new diagonal entry c . If $\mathbf{x}_S = \mathbf{0}$ and $\mathbf{y}^\top B \mathbf{x} \neq c$ then $\text{rank } A' \geq \text{rank } A + 1$.*

Proof. Let A'' be the matrix obtained by appending the column \mathbf{x} (but not appending a new row), and let \mathbf{y}' be the last row of A' (obtained by appending the entry “ c ” to the end of the vector \mathbf{y}). Also, let $\mathbf{w} \in \mathbb{R}^{n+1}$ be the vector obtained by appending the entry “ -1 ” to the end of the vector $B \mathbf{x}$.

Since $\mathbf{x}_S = \mathbf{0}$, we have $AB \mathbf{x} = \mathbf{x}$ by the defining property of B , which implies that \mathbf{w} is a kernel vector of A'' (i.e., it lies in the orthogonal complement of the row space of A''). So, if $\mathbf{w} \cdot \mathbf{y}' \neq 0$, then \mathbf{y}' does not lie in the row space of A'' , meaning that $\text{rank } A' \geq \text{rank } A'' + 1 \geq \text{rank } A + 1$. The desired result follows, noting that $\mathbf{w} \cdot \mathbf{y}' = \mathbf{x}^\top B^\top \mathbf{y} - c$. \square

Now we are ready to prove Lemma 8.4(b).

Proof of Lemma 8.4(b). Let B, P be as in Lemma 8.7, and consider the symmetrisation $B' = (B + B^\top)/2$ of B . Since PBP is symmetric, B and B' have the same entries in positions indexed by $\bar{S} \times \bar{S}$. Recall that $\mathbf{x}_S = \mathbf{0}$, so $\mathbf{x}^\top B \mathbf{x} = \mathbf{x}^\top B' \mathbf{x}$.

We claim that there are at least $\eta^2 n^4/4$ different tuples (i, i', j, j') with $B_{i,j} - B_{i,j'} + B_{i',j} - B_{i',j'} \neq 0$. Since the symmetrisation of B to B' only affects at most $4|S|n^3 \leq \eta^2 n^4/8$ of these tuples, it will follow from Lemma 8.6 that $\mathbf{x}^\top B' \mathbf{x} = \mathbf{x}^\top B \mathbf{x} \neq 0$ with probability at least $1 - O((\log 2d)^{O_\eta(1)}/\sqrt{d})$, in which case $\text{rank } A' \geq \text{rank } A + 1$ by Lemma 8.8, as desired.

To prove the claim, let \mathbf{b}_i be the i th row of B and let \mathbf{e}_i be the i th standard unit vector. For $i, i' \in \bar{S}$, let $\mathbf{w}_{i,i'} = \mathbf{b}_i - \mathbf{b}_{i'}$, so $A \mathbf{w}_{i,i'} = \mathbf{e}_i - \mathbf{e}_{i'}$. Recall that A being η -unstructured means that there are ηn^2 pairs of indices (i, i') for which all \mathbf{w} satisfying $\text{supp}(A \mathbf{w}) = \{i, i'\}$ are η -balanced. So, there are at least $\eta n^2 - 2|S|n \geq \eta n^2/2$ pairs $(i, i') \in \bar{S}^2$ for which $\mathbf{w}_{i,i'}$ is η -balanced. For each such (i, i') , there are at least $\eta(1 - \eta)n^2 \geq \eta n^2/2$ pairs (j, j') for which the j th and j' th entry of $\mathbf{w}_{i,i'}$ differ, in which case $B_{i,j} - B_{i,j'} + B_{i',j} - B_{i',j'} \neq 0$. \square

9. EXTRACTING HIGH-DEGREE VERTICES

As outlined, for the proof of Theorem 1.12 we need to “extract” high degree vertices from our random graph with minimum degree at least 2, without revealing too much about the neighbourhoods of the extracted vertices. We will need certain information about the graph that remains after this extraction; most notably we need control over its corank, and we need to know that most of its vertices still have degree at least 2.

Lemma 9.1. *Fix $\varepsilon, \alpha, \Delta > 0$ such that $1/\Delta \ll \alpha \ll \varepsilon$.*

(A) *Consider sets $S \subseteq V$ and an integer m such that*

- $|V| = n$,
- $\varepsilon n \leq m - n \leq n/\varepsilon$, and
- $|S| = \lfloor \alpha n \rfloor$.

Let $G \sim \mathcal{K}(V, m, 2)$, and let $T = \{v \in S : \deg_G(v) \geq \Delta\}$ be the set of vertices in S with degree at least Δ . Then, whp:

- (1) $\text{corank } A(G[V \setminus T]) \leq |T|/8$,
- (2) *Let W be the set of vertices in $G[V \setminus T]$ which have degree at least 2 (with respect to $G[V \setminus T]$). Then $|(V \setminus T) \setminus W| = O_\varepsilon(\alpha n)$.*
- (3) $|T| \leq \alpha \exp(-\Omega_\varepsilon(\Delta))n$,
- (4) *Whp all but $|T|/\Delta$ vertices $v \in T$ satisfy $\deg_W(v) \geq \sqrt{\Delta}$.*

(B) *Consider disjoint sets $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$, and an integer m , such that*

- $|V_1|, |V_2| = n + o(n)$,
- $\varepsilon n \leq m - |V_1| - |V_2| \leq n/\varepsilon$, and
- $|S_1| = |S_2| = \lfloor \alpha n \rfloor$.

Let $G \sim \mathcal{K}(V_1, V_2, m, 2)$, and for $i \in \{1, 2\}$ let $T_i = \{v \in S_i : \deg_G(v) \geq \Delta\}$ be the set of vertices in S_i with degree at least Δ . Let $V = V_1 \cup V_2$ and $T = T_1 \cup T_2$. Then, whp:

- (1) $\text{corank } B(G[V \setminus T]) \leq |T|/16$,

- (2) Let W be the set of vertices in $G[V \setminus T]$ which have degree at least 2 (with respect to $G[V \setminus T]$). Then $|(V \setminus T) \setminus W| = O_\varepsilon(\alpha n)$.
- (3) $|T_1| = |T_2| + o_{\varepsilon, \alpha, \Delta}(n) \leq \alpha \exp(-\Omega_\varepsilon(\Delta))n$,
- (4) Whp all but $|T|/\Delta$ vertices $v \in T$ satisfy $\deg_W(v) \geq \sqrt{\Delta}$.

All parts of [Lemma 9.1](#) follow from similar calculations to those that were performed in [\[33\]](#). In particular, in the setting of (A), [\[33, Lemma 6.6\]](#) gives asymptotic formulas (in terms of m/n , α and Δ) for $|T|$ and the number of vertices in $G[V \setminus T]$ with a given degree⁸, from which (A2) and (A3) follow. Then, (A4) may be proved with a simple configuration-model calculation; such a calculation appears in the proof⁹ of [\[33, Lemma 8.1\(3\)\]](#) (recalling E from the notation of that lemma, we have $E \subseteq W$ when $k = 2$).

Essentially the same calculations can be performed in the bipartite setting to prove (B2–4) (in fact, the relevant asymptotic formulas are nearly identical, though one needs to consider separate parameters for each side of our random bipartite graph).

(A1) and (B1) warrant a bit more explanation. [Lemma 9.1\(A1\)](#) is basically the same as [\[33, Lemma 7.3\]](#) (which concerned random graphs constrained to have minimum degree at least $k \geq 3$, while we need to handle random graphs constrained to have minimum degree at least 2). Roughly speaking, the idea is to show that our random graph G locally weakly converges (in the sense of Aldous–Steele [\[4\]](#) and Benjamini–Schramm [\[12\]](#)) to a certain Galton–Watson tree. Spectral convergence machinery of Bordenave, Lelarge, and Salez [\[18\]](#) then can be used to bound the corank of $A(G)$ in terms of a certain probability generating function associated with that Galton–Watson tree. The proof of [\[33, Lemma 7.3\]](#) does use the assumption $k \geq 3$, but it was written in a slightly inefficient way; it is possible to slightly modify the proof to overcome this assumption (as we sketch momentarily).

It turns out that essentially the same proof can also be used for [Lemma 9.1\(B1\)](#), because local weak convergence does not “see” whether a graph is bipartite or not.

Proof sketch of [Lemma 9.1\(A1\)](#) and (B1). First, for the reader’s convenience, we outline the proof of [\[33, Lemma 7.3\]](#). We then discuss the minor changes that are necessary to prove [Lemma 9.1\(A1\)](#).

We may assume that $2m/|V|$ converges to some $g \in [2 + 2\varepsilon, 2/\varepsilon]$. Indeed, if the desired property did not hold whp, then for some $\tau > 0$ there would be an infinite sequence of integers n (and accompanying sets $V^{(n)}$) along which for each n the corresponding property fails to hold with probability at least τ . By compactness there would then be an infinite violating subsequence along which $2m/|V^{(n)}|$ converges to a limit.

As proved in [\[33, Lemma 6.10\]](#), G locally converges to a Galton–Watson tree with a certain offspring distribution μ , and by [\[33, Lemma 6.6\]](#), whp $|T| = \beta n + o(n)$ for some explicit $\beta > 0$ depending on g (and the other parameters). As discussed in [\[33, Section 7\]](#), by results of Bordenave, Lelarge, and Salez (specifically [\[18, Theorem 13 and Eq. \(19\)\]](#)), we have $\text{corank } A(G) \leq \max_{x \in [0, 1]} M_\mu(x) + o(n)$ for a certain function $M_\mu: [0, 1] \rightarrow \mathbb{R}$ depending on μ , so it suffices to prove that $\max_{x \in [0, 1]} M_\mu(x) \leq \beta/16$. This is essentially what is proved in [\[33, Lemma 7.5\]](#), but there was one point where the assumption $k \geq 3$ was used: namely, for certain $\gamma, \lambda > 0$ (which depend only on g and α) and a function $\phi: [0, (1 - \gamma)\lambda] \rightarrow \mathbb{R}$ defined by

$$\phi(x) = \sum_{t=k-1}^{\infty} \frac{x^t}{t!} ((\mathbb{1}_{t \geq k} - \alpha \mathbb{1}_{t \geq \Delta}) + \gamma \lambda (\mathbb{1}_{t+1 \geq k} - \alpha \mathbb{1}_{t+1 \geq \Delta})),$$

it is necessary to prove that ϕ'' is log-concave, and the proof in [\[33, Lemma 7.5\]](#) uses the assumption $k \geq 3$. We give an alternative proof for the log-concavity of ϕ'' in the case $k = 2$, as follows.

In the case $k = 2$, one can compute

$$\phi''(x) = \sum_{t=0}^{\infty} \frac{x^t}{t!} ((1 + \gamma \lambda) - \alpha (\mathbb{1}_{t \geq \Delta-2} + \gamma \lambda \mathbb{1}_{t \geq \Delta-3})) = (1 + \gamma \lambda)(e^x - h(x)),$$

where

$$h(x) = \sum_{t=0}^{\infty} \frac{x^t}{t!} (1 + \gamma \lambda)^{-1} \alpha (\mathbb{1}_{t \geq \Delta-2} + \gamma \lambda \mathbb{1}_{t \geq \Delta-3}).$$

⁸The lemmas in [\[33, Section 6\]](#) are stated with an assumption $k \geq 3$, but this is completely unnecessary.

⁹Again, [\[33, Lemma 8.1\]](#) is stated with an assumption $k \geq 3$, but this is unnecessary.

To verify that $\phi''(x)$ is log-concave it suffices to verify that

$$\frac{d^2}{dx^2} \log(e^x - h(x)) = -\frac{e^x(h(x) + h''(x) - 2h'(x)) + (h'(x))^2 - h(x)h''(x)}{(e^x - h(x))^2}$$

is non-positive. First we compute

$$\begin{aligned} h(x) + h''(x) - 2h'(x) &= \sum_{t=0}^{\infty} \left(\frac{x^t}{t!} - \frac{2x^{t-1}}{(t-1)!} + \frac{x^{t-2}}{(t-2)!} \right) (1 + \gamma\lambda)^{-1} \alpha (\mathbb{1}_{t \geq \Delta-2} + \gamma\lambda \mathbb{1}_{t \geq \Delta-3}) \\ &= \sum_{t=0}^{\infty} \frac{x^{t-2}(x^2 - 2tx + t(t-1))}{t!} (1 + \gamma\lambda)^{-1} \alpha (\mathbb{1}_{t \geq \Delta-2} + \gamma\lambda \mathbb{1}_{t \geq \Delta-3}). \end{aligned}$$

For sufficiently large Δ in terms of λ and γ , and $x \in [0, (1 - \gamma)\lambda]$, each term in this expression is nonnegative, because $x^2 - 2tx + t(t-1) \geq 0$ for $x \leq t - \sqrt{t}$. So, it suffices to verify that $h'(x)^2 - h(x)h''(x) \geq 0$, which is equivalent to h being log-concave.

To prove that h is log-concave, we use the well-known fact (see for example [6, Lemma 3]) that if a nonnegative function f is log-concave on an interval $[a, b]$, then its antiderivative $x \mapsto \int_a^x f(y)dy$ is also log-concave on that same interval. Given this, it suffices to prove that the $(\Delta - 3)$ -fold derivative $h^{(\Delta-3)}$ is log-concave. We compute

$$h^{(\Delta-3)}(x) = (1 + \gamma\lambda)^{-1} \alpha ((1 + \gamma\lambda)e^x - \gamma\lambda),$$

so we may now finish the proof by direct differentiation. Indeed, for any $r \in [0, 1]$ we compute $\frac{d^2}{dx^2} \log(e^x - r) = -re^x/(e^x - r)^2 \leq 0$, which implies the desired result.

For (B1), essentially the same proofs as in [33, Section 6] show that G still locally converges to a Galton–Watson tree with offspring distribution μ , and whp $|T| = \beta(|V_1| + |V_2|) + o(n)$, so the same proof as above shows that $\text{corank } A(G[V \setminus T]) \leq |T|/8$. So, by Fact 3.1 we have $\text{corank } B(G[V \setminus T]) \leq |T|/16$. \square

10. ANALYSING A CORANK-WALK

In this section we explain how to prove Theorem 1.12(A1) using the tools from Sections 7 to 9. The proof of Theorem 1.12(B1) is very similar (actually, it is slightly easier), and we briefly sketch the necessary changes for that proof in Section 12.

Fix $\varepsilon > 0$, let $1 + \varepsilon \leq m/n \leq n/\varepsilon$, let $G \sim \mathcal{K}(n, m, 2)$ and write $V = \{1, \dots, n\}$ for its vertex set. Our objective is to prove that whp $\text{corank } A(G) = s(G)$.

Fix $\alpha, \eta, \Delta > 0$ such that $1/\Delta \ll \alpha \ll \eta \ll \varepsilon$. At the end of the proof we will take $\Delta \rightarrow \infty$ as $n \rightarrow \infty$, but for now we view Δ as a constant. (In particular, we assume n is large in terms of α, η, Δ .)

As in Lemma 9.1, consider a set S of $\lfloor \alpha n \rfloor$ vertices (say $S = \{1, \dots, \lfloor \alpha n \rfloor\}$), and let $T = \{v \in S : \deg_G(v) \geq \Delta\}$ be the set of vertices with degree at least Δ . When we take $\Delta \rightarrow \infty$ at the end of the proof, we will have $|T| = o(n)$ whp, but until then the reader should think of $|T|$ as having order n .

Given the information in Lemma 9.1 about $G[V \setminus T]$, our strategy is to study the evolution of the corank as we add back the vertices in T (in a random order), using Theorem 7.2 and Lemma 8.4. Let $N = |T|$, consider a uniform random ordering v_{N-1}, \dots, v_0 of the elements of T , and let $G_t = G[V \setminus \{v_{t-1}, \dots, v_0\}]$ (so $G_N = G[V \setminus T]$ and $G_0 = G$).

Let W be the set of vertices in G_N which have degree at least 2 (with respect to G_N). The idea is that the vertices in W already satisfy their degree constraints, so all the vertices in W are equally likely to be neighbours of vertices in T . The following lemma makes this precise.

Claim 10.1. *Reveal an outcome of T , reveal all the edges of G not between T and W , and reveal $\deg_W(v)$ for each $v \in T$. Then, conditionally, the neighbourhoods $(N_W(v))_{v \in T}$ are independent uniformly random subsets of W with sizes $(\deg_W(v))_{v \in T}$.*

Proof. Let $G[T, W]$ be the bipartite graph of edges between T and W . Let H_1, H_2 be bipartite graphs with the same bipartition $X \cup Y$, such that every vertex in X has the same degree in H_1 as it does in H_2 . Then, for any outcome of G such that $G[T, W] = H_1$, we can swap $G[T, W]$ with H_2 to obtain an outcome of G such that $G[T, W] = H_2$. So, H_1 and H_2 are equally likely to occur as $G[T, W]$. (It is important that this swap can never change the sets W or T , and can never cause the degree of any vertex to drop below 2.) \square

Recall that, by Lemma 9.1(A4), almost all vertices v_1, \dots, v_N have degree at least $\sqrt{\Delta}$ into W . So, the upshot of Claim 10.1 is that when we add the vertices v_1, \dots, v_N back to G_N , at most of these steps we

are essentially adding a new random row and column with many “1”-entries, which puts us in a position to apply [Lemma 8.4](#).

Recall that [Lemma 8.4\(a\)](#) has an assumption that the matrix under consideration has a balanced kernel vector: a kernel vector which is not dominated by a single level set. We therefore need some estimates about kernel vectors of the adjacency matrices $A(G_t)$.

First, say that a level set of a vector \mathbf{v} is a *nonzero level set* if it is the λ -level set of \mathbf{v} for some $\lambda \neq 0$. With some crude estimates it is not hard to show that there are unlikely to ever be kernel vectors which are dominated by a nonzero level set.

Claim 10.2. *With probability $1 - n^{-\omega(1)}$, no $A(G_t)$ has a kernel vector with a nonzero level set larger than $(1 - \eta)(n - t)$ (provided $\eta \ll \varepsilon$ and $\alpha \leq \eta$).*

We defer the simple proof of [Claim 10.2](#) to [Section 11](#). It is much more delicate to deal with kernel vectors which are dominated by their zero level set, i.e., kernel vectors with small *support*. Indeed, the special cycles counted by $s(G)$ each give rise to a kernel vector with small support, so we certainly cannot rule these out entirely. For each $t \leq N$, define the set of “small-support” kernel vectors

$$K_t^{(\eta)} = \{\mathbf{v} \in \ker A(G_t) : |\text{supp}(\mathbf{v})| \leq \eta(n - t)\}.$$

The following lemma shows that while there may be some vectors in $K_t^{(\eta)}$, typically these vectors are collectively supported on a small subset of indices.

Claim 10.3. *For each t ,*

$$\Pr[t \leq N \text{ and } \dim(K_t^{(\eta)}) \geq t/4] \lesssim_\eta \frac{1}{t} + \left(\frac{t}{n}\right)^{1/4}.$$

Remark. Note that N is random, and we are not conditioning on it at this stage.

We also need much more precise control for the last few steps of our random walk, essentially characterising each $K_t^{(\eta)}$ in terms of the special cycles of G_t . For an h -vertex graph H , let $s^{(\eta)}(H)$ be the number of special cycles of length at most $2\eta h$ in H , counting isolated special cycles twice. Let $C_t^{\text{spec}(\eta)}$ be the set of special cycles of length at most $2\eta n$ in G_t and let $V_t^{\text{spec}(\eta)}$ be the set of degree-2 vertices in these special cycles.

Claim 10.4. *The following hold together with probability at least $1 - 1/\Delta$.*

- (a) $\text{supp}(K_t^{(\eta)}) = V_t^{\text{spec}(\eta)}$ for each $t \leq \Delta$.
- (b) $\dim(K_t^{(\eta)}) = s^{(\eta)}(G_t)$ for each $t \leq \Delta$.
- (c) $(C_\Delta^{\text{spec}(\eta)}, V_\Delta^{\text{spec}(\eta)}) = \dots = (C_0^{\text{spec}(\eta)}, V_0^{\text{spec}(\eta)})$.
- (d) $|V_t^{\text{spec}(\eta)}| \leq \eta^2 n / 50$ for each $t \leq \Delta$.
- (e) v_t has no neighbor in $V_{t+1}^{\text{spec}(\eta)}$, for each $t < \Delta$.

(At the end of the proof we will take $\Delta \rightarrow \infty$, meaning that [Claim 10.4](#) will become a with-high-probability statement.)

The proofs of [Claims 10.3](#) and [10.4](#) are very delicate; they proceed by considering a linear-algebraic notion of *minimal kernel vectors*, and studying the combinatorial consequences of this notion. We defer the proofs to [Section 11](#).

For the last few steps of our random walk (for $t < \Delta$) we need to use [Lemma 8.4\(b\)](#), so we also need to know that $A(G_t)$ is likely to be η -unstructured for such t , in the sense of [Definition 8.3](#).

Claim 10.5. *For $t \in \{0, \dots, \Delta\}$, $A(G_t)$ is η -unstructured with probability $1 - (\log n)^{-\omega(1)}$.*

It turns out that [Claim 10.5](#) can be proved within the same general framework as [Claims 10.3](#) and [10.4](#). We defer this proof of [Claim 10.5](#) to [Section 11](#).

We are now ready to define our random walk X_N, \dots, X_0 . Say an index $t \leq N$ is *good* if all of the following hold.

- G1** $\deg_W(v_t) \geq \sqrt{\Delta}$, and
- G2** if $t \geq \Delta/2$, then $\dim(K_{t+1}^{(\eta)}) < t/4$, and
- G3** $A(G_{t+1})$ has no kernel vector with a nonzero level set larger than $(1 - \eta)(n - t - 1)$, and
- G4** if $t < \Delta$, then v_t has no neighbour in $\text{supp}(K_{t+1}^{(\eta)})$, and
- G5** if $t < \Delta$, then $\dim(K_{t+1}^{(\eta)}) = \dim(K_\Delta^{(\eta)})$, and

G6 if $t < \Delta$, then $A(G_{t+1})$ is η -unstructured, and

G7 if $t < \Delta$, then $|\text{supp}(K_{t+1}^{(\eta)})| < \eta^2(n-t-1)/32$.

Let R be the set of indices that are not good (i.e., bad). Then, we have the following consequence of [Claims 10.2 to 10.5](#) (and the properties in [Lemma 9.1](#)).

Claim 10.6. *With probability $1 - o_{\Delta \rightarrow \infty}(1)$, for all $t \leq N$ we have $|R \cap \{1, \dots, t\}| \leq t/100$.*

Remark. Here we use notation of the form $f = o_{\Delta \rightarrow \infty}(n)$ to mean that f/n can be made arbitrarily small by taking sufficiently large Δ . Recall that for the moment we are viewing Δ as a constant, but later on we will take $\Delta \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. By [Lemma 9.1\(A3\)](#), whp $N \leq \beta n$ for some $\beta = o_{\Delta \rightarrow \infty}(n)$. For each $i = 0, 1, \dots, \lfloor \log_2(\beta n + 2) \rfloor$, let Q_i be the number of bad steps $t \in R$ with $t \leq \beta n$ in the range $[2^i - 1, 2^{i+1} - 1)$. It suffices to show that with probability $1 - o_{\Delta \rightarrow \infty}(1)$, for each i we have $Q_i \leq 2^i/800$.

Let $Q_i^{\mathbf{G1}}, \dots, Q_i^{\mathbf{G7}}$ be the contribution to Q_i from failure of each of **G1** to **G7** (so $Q_i \leq Q_i^{\mathbf{G1}} + \dots + Q_i^{\mathbf{G7}}$). We will show that with probability $1 - o_{\Delta \rightarrow \infty}(1)$ we have $Q_i^j \leq 2^i/5600$ for each $j \in \{\mathbf{G1}, \dots, \mathbf{G7}\}$.

First, the cases $j \in \{\mathbf{G3}, \dots, \mathbf{G7}\}$ are easy to handle with [Claims 10.2, 10.4, and 10.5](#) and the union bound. For $j = \mathbf{G1}$, say that a vertex $v \in T$ is “degree-bad” if $\deg_W(v) < \sqrt{\Delta}$. By [Lemma 9.1\(A4\)](#), whp the fraction of degree-bad vertices in T is at most $1/\Delta$. We can condition on such an outcome of these degree-bad vertices without revealing any information about the ordering v_1, \dots, v_N of the vertices in T . So, by a Chernoff bound for the hypergeometric distribution ([Lemma 3.2](#)), we have $\Pr[Q_i^{\mathbf{G1}} > 2^i/5600] \leq e^{-\Omega(2^i)}$, and by Markov’s inequality we have $\Pr[Q_i^{\mathbf{G1}} > 2^i/5600] \leq O(1/\Delta)$. Using the former inequality for say $i \geq \log \Delta$ and the latter inequality for $i < \log \Delta$, the desired result follows by a union bound.

It remains to consider the case $j = \mathbf{G2}$. By Markov’s inequality and [Claim 10.3](#), we have

$$\begin{aligned} \sum_{i=0}^{\lfloor \log_2(\beta n + 2) \rfloor} \Pr[Q_i^{\mathbf{G2}} > \frac{2^i}{5600}] &\leq \sum_{i=0}^{\lfloor \log_2(\beta n + 2) \rfloor} \frac{\mathbb{E}Q_i^{\mathbf{G2}}}{\Omega(2^i)} = \sum_{i=0}^{\infty} \sum_{\substack{t \in [2^i - 1, 2^{i+1} - 1) \\ \Delta/2 \leq t \leq \beta n}} \frac{\Pr[t \leq N \text{ and } \dim(K_{t+1}^{(\eta)}) \geq t/4]}{\Omega(2^i)} \\ &\lesssim_{\eta} \sum_{t=\lceil \Delta/2 \rceil}^{\beta n} \frac{1/t + (t/n)^{1/4}}{t} \lesssim_{\eta} \frac{1}{\Delta} + \beta^{1/4} = o_{\Delta \rightarrow \infty}(1). \quad \square \end{aligned}$$

Now, we are ready to complete the proof of [Theorem 1.12\(A1\)](#) (using the notation and claims from throughout this section).

Proof of [Theorem 1.12\(A1\)](#). Let $\mathcal{I} = (G_N, (\deg_W(v))_{v \in T})$. That is to say, \mathcal{I} specifies G_N (which determines T and W), and the degrees from T into W . Let \mathcal{E} be the event that $|R \cap \{1, \dots, t\}| > t/100$ for some $t \leq N$, so $\Pr[\mathcal{E}] \leq h(\Delta)$ for some h satisfying $h(\Delta) \rightarrow 0$ as $\Delta \rightarrow \infty$ by [Claim 10.6](#).

Note that $\Pr[\mathcal{E}] = \mathbb{E}[\Pr[\mathcal{E} | \mathcal{I}]]$, so applying Markov’s inequality to $\Pr[\mathcal{E} | \mathcal{I}]$, and applying [Lemma 9.1](#) for $\alpha \ll \varepsilon$, we see that \mathcal{I} satisfies

- (i) $\text{corank } A(G_N) \leq |T|/8$,
- (ii) $G[V_N]$ has at most $\eta n/10$ vertices with degree less than 2,
- (iii) $\Pr[\mathcal{E} | \mathcal{I}] \geq 1 - h(\Delta)^{1/2}$.

with probability at least $1 - 2h(\Delta)^{1/2}$. For the rest of the proof, we condition on such an outcome of \mathcal{I} (so, for example, we treat T , N , and G_N as deterministic objects).

Now, for $t \leq N$, let

$$X_t = \text{corank } A(G_t) - \mathbb{1}_{t < \Delta} \dim K_{\Delta}^{(\eta)} - t/4.$$

We claim that the sequence X_N, \dots, X_0 and the “bad set” R satisfy the conditions of [Theorem 7.2](#) (with $\varepsilon = 1/2$, $C = 5/4$, $\delta = 3/4$ and $p = o_{\Delta \rightarrow \infty}(1)$). Conditions **W0** and **W2** are immediate, condition **W1** follows from (i) above, and condition **W4** follows from (iii). So, we just need to verify **W3**.

To this end, condition on any outcome of G_{t+1} (which determines X_{t+1}). We will study how X_t differs from X_{t+1} , in this conditional probability space.

In addition to information revealed so far, reveal v_t and its degree $\deg_W(v_t)$ into W . If $t < \Delta$, also reveal the neighbourhood of v_t in $\text{supp}(K_{t+1}^{(\eta)})$ (this is enough information to see whether step t is bad). Condition on an outcome of the revealed information; we need to show that if step t is not bad, then with probability at least $1 - o_{\Delta \rightarrow \infty}(1)$: if $X_{t+1} > 0$ then $X_t \leq X_{t+1} - 3/4$, and if $X_{t+1} \leq 0$ then $X_t \leq 0$.

We assume that the revealed information is such that step t is not bad (otherwise there is nothing to prove).

Let $d = \deg_W(v_t) \geq \sqrt{\Delta}$ (by **G1**). If $t \geq \Delta$ let $E = W$ and if $t < \Delta$ let $E = W \setminus \text{supp}(K_{t+1}^{(\eta)})$. In either case, we have $|E| \geq (1 - \alpha)n - \eta n/10 - \Delta \geq (1 - \eta/3)n$ (using (ii) above and **G2** and **G5**). By **Claim 10.1** and **G4**, the neighbourhood $N_E(v_t)$ of v_t in E is a uniformly random size- d subset of E . Now, we use one of the two parts of **Lemma 8.4**, as follows.

- *Case 1:* $\text{corank}(A(G_{t+1})) > \dim(K_{t+1}^{(\eta)})$. In this case, there is some kernel vector \mathbf{v} of $A(G_{t+1})$ with $|\text{supp}(\mathbf{v})| \geq \eta(n - t - 1)$. By **G3**, this kernel vector is η -balanced, so **Lemma 8.4(a)** yields that $\text{rank}(A(G_t)) \geq \text{rank}(A(G_{t+1})) + 2$ (and hence $X_t \leq X_{t+1} - 3/4$) with probability at least $1 - O_\eta(\Delta^{-1/4}) \geq 1 - \Delta^{-1/8}$.
- *Case 2:* $\text{corank}(A(G_{t+1})) = \dim(K_{t+1}^{(\eta)})$. By **G2** and **G5**, this case can only happen if $X_{t+1} < 0$. Thus if $\text{corank}(A(G_t)) = \text{corank}(A(G_{t+1}))$, we have $X_t \leq 0$ since the walk is quarter-integral and increases by at most $1/4$. By **G6** and **G7**, the conditions for **Lemma 8.4(b)** are met, and thus with probability at least $1 - \Delta^{-1/8}$, we have $\text{corank}(A(G_t)) = \text{corank}(A(G_{t+1}))$. The desired result follows.

Now, having verified conditions **W0** to **W4**, the conclusion of **Theorem 7.2** is that $X_0 = \text{corank } A(G_0) - \dim K_\Delta^{(\eta)} \leq 0$ with probability at least $1 - O(\min(h(\Delta)^{1/2}, \Delta^{-1/8}))$. By **Claim 10.4**, it follows that $\text{corank } A(G_0) = \dim K_0^{(\eta)} = s^{(\eta)}(G)$ with probability at least $1 - O(\min(h(\Delta)^{1/2}, \Delta^{-1/8}))$. We deduce that $\text{corank } A(G_0) = \dim K_0^{(\eta)} = s^{(\eta)}(G)$ whp, taking $\Delta \rightarrow \infty$ sufficiently slowly. It now just suffices to observe that when $\text{corank } A(G_0) = \dim K_0^{(\eta)}$ there are no special cycles longer than $2\eta n$. Indeed, such a special cycle would give rise to a kernel vector with support larger than ηn , by **Fact 1.6**. \square

11. KERNEL VECTORS AND STALKS

In this section we prove **Claims 10.2** to **10.5**, which are the remaining ingredients in our proof of **Theorem 1.12(A1)**.

Claims 10.2 to **10.4** concern kernel vectors; note that **Claim 10.5** can be interpreted as a claim about *almost* kernel vectors. Indeed, say that a vector $\mathbf{v} \in \mathbb{R}^n$ is an ℓ -almost kernel vector of a matrix $A \in \mathbb{R}^{n \times n}$ if $|\text{supp}(A\mathbf{v})| = \ell$ (so a kernel vector is a 0-almost kernel vector, and the definition of η -unstructuredness in **Definition 8.3** concerns 2-almost kernel vectors). We will therefore be able to prove each of **Claims 10.2** to **10.5** by carefully studying almost-kernel vectors in degree-constrained random graphs.

First, it is easy to show that kernel vectors which are dominated by a nonzero level set are unlikely: the following lemma immediately implies **Claim 10.2**, and is a simple consequence of **Lemma 6.2(A4)**.

Lemma 11.1. *Let $G = G_0, \dots, G_N$ be as in **Section 10**. Suppose $\eta \ll \varepsilon$ and $\alpha \leq \eta$. Then with probability $1 - n^{-\omega(1)}$ no $A(G_t)$ has an ℓ -almost kernel vector with a nonzero level set larger than $(1 - \eta)(n - t)$, for any $\ell \in \{0, 1, 2\}$.*

Proof. We show that the desired result follows whenever G satisfies the conclusion of **Lemma 6.2(A4)**.

Consider any vector $\mathbf{x} \in \mathbb{R}^{V(G_t)}$ with λ -level set $U \subseteq V(G_t) \subseteq V(G)$ larger than $(1 - \eta)(n - t) \geq (1 - \eta)(n - \alpha n) \geq (1 - 2\eta)n$, for some $\lambda \neq 0$, and suppose without loss of generality that $\lambda > 0$. We will show that \mathbf{x} cannot be an ℓ -almost kernel vector of $A(G_t)$, for any $\ell \leq 2$.

Let \bar{U} be the complement of U in $V(G)$, and recall that we defined $G = G_0, \dots, G_N$ by deleting some vertices from a special vertex subset S with $|S| = \lfloor \alpha n \rfloor$. Note that $\bar{U} \cup S$ has at most $O_\varepsilon((\eta n) \log(1/\eta)) \leq n - t - 3$ neighbors in G . This implies that there are at least 3 vertices $v \in V(G_t)$ which have at least 2 neighbours in G_t , all of which are in U . So, the v -coordinate of $A(G_t)\mathbf{x}$ is $\deg_{G_t}(v)\lambda \geq 2\lambda > 0$. The desired result follows. \square

The above lemma handles almost-kernel vectors that are dominated by a nonzero level set, but we also need to handle almost-kernel vectors that are dominated by their zero level set (i.e., almost-kernel vectors with small support). To this end we need the notion of a *minimal* vector (previously appearing in work of DeMichele, the first author, and Moreira [30]).

Definition 11.2. Say that a vector \mathbf{v} is *minimal* if for any $\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with $\text{supp}(\mathbf{w}) \subsetneq \text{supp}(\mathbf{v})$, we have $\text{supp}(A\mathbf{w}) \not\subseteq \text{supp}(A\mathbf{v})$.

It is clear from the above definition that for any vector \mathbf{v} , there is a minimal vector whose support is contained in \mathbf{v} . In fact more is true: every ℓ -almost kernel vector can be written as a sum of such minimal almost kernel vectors.

Lemma 11.3. For any matrix $A \in \mathbb{R}^{n \times n}$ and vector $\mathbf{v} \in \mathbb{R}^n$, we can write \mathbf{v} as a sum of minimal vectors \mathbf{w} satisfying $\text{supp}(A\mathbf{w}) \subseteq \text{supp}(A\mathbf{v})$.

Proof. Suppose for the purpose of contradiction that the lemma statement is false, and let $\mathbf{v} = (v_1, \dots, v_n)$ be a minimal-support counterexample (i.e., with $|\text{supp}(\mathbf{v})|$ as small as possible). We are assuming there is no way to represent \mathbf{v} as a sum of minimal vectors \mathbf{w} which satisfy $\text{supp}(A\mathbf{w}) \subseteq \text{supp}(A\mathbf{v})$. We say a vector \mathbf{w} is *properly contained* in \mathbf{v} if $\text{supp}(\mathbf{w}) \subsetneq \text{supp}(\mathbf{v})$ and $\text{supp}(A\mathbf{w}) \subseteq \text{supp}(A\mathbf{v})$.

By assumption, \mathbf{v} is not itself minimal, meaning that there is a vector $\mathbf{w} = (w_1, \dots, w_n) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ that is properly contained in \mathbf{v} . Fix any $i \in \text{supp}(\mathbf{w})$, let $\lambda = v_i/w_i$ and let $\mathbf{w}' = \mathbf{v} - \lambda\mathbf{w}$. Then \mathbf{w}' is properly contained in \mathbf{v} as well (note that $i \notin \text{supp}(\mathbf{w}')$).

Since \mathbf{v} is a minimal-support counterexample, we can write $\lambda\mathbf{w}$ (respectively \mathbf{w}') as a sum of vectors that are properly contained in $\lambda\mathbf{w}$ (respectively, properly contained in \mathbf{w}'). Note that proper containment is transitive; since $\mathbf{v} = \mathbf{w}' + \lambda\mathbf{w}$, we can now write \mathbf{v} as a sum of vectors that are properly contained in \mathbf{v} , which is a contradiction. \square

Minimal almost-kernel vectors enjoy certain combinatorial properties, which we capture in the notion of a *stalk*. For a graph G and a vertex set R , write $N(R)$ for the union of neighbourhoods of vertices in R (so $N(R)$ may intersect R).

Definition 11.4. Given a graph G , call a set of vertices $R \subseteq V(G)$ an (r, s, ℓ) -stalk for G if:

- S0** $|R| = r$ and $|N(R)| = s$,
- S1** $s \geq r - 1 + \ell$,
- S2** R cannot be split into two nonempty sets with disjoint neighborhoods, and
- S3** All but exactly ℓ vertices $v \in N(R)$ have at least two neighbours in R .

The ℓ vertices $v \in N(R)$ with $|N(\{v\}) \cap R| \leq 1$ are called the *exceptional vertices* for the stalk R . Also, we use the shorthand “ (r, ℓ) -stalk” to describe an (r, s, ℓ) stalk for any $s \geq r - 1 + \ell$, and the shorthand “ $(\leq q, \ell)$ -stalk” to describe an (r, ℓ) -stalk for any $r \leq q$.

Lemma 11.5. Consider an n -vertex graph G and a minimal ℓ -almost kernel vector \mathbf{v} of its adjacency matrix $A(G)$. Then $R = \text{supp}(\mathbf{v})$ is a $(|R|, \ell')$ -stalk, for some $\ell' \leq \ell$.

Proof. Let $Q = \text{supp}(A(G_t)\mathbf{v})$ (so $|Q| = \ell$). For **S1**, suppose for the purpose of contradiction that $|N(R)| < |R| + \ell - 1$. Let R' be obtained by removing an arbitrary vertex of R , and let A' be the $R' \times (N(R) \setminus Q)$ submatrix of $A(G)$. Then A' has $|R| - 1$ rows and $|N(R)| - \ell < |R| - 1$ columns, so has a nonzero left kernel vector \mathbf{w} . Padding this vector with zeroes gives a nonzero vector $\mathbf{w}' \in \ker(A(G))$ with $\text{supp}(\mathbf{w}') \subsetneq \text{supp}(\mathbf{v})$ and $\text{supp}(A(G)\mathbf{w}') \subseteq \text{supp}(A(G)\mathbf{v})$, contradicting the minimality of \mathbf{v} .

For **S2**, suppose for the purpose of contradiction that R can be split into two nonempty sets R_1, R_2 with distinct neighbourhoods. Let $\mathbf{v}_1 \in \mathbb{R}^n$ be the vector obtained from \mathbf{v}_1 by setting all entries not indexed by R_1 to zero. Then \mathbf{v}_1 contradicts the minimality of \mathbf{v} .

For **S3**, suppose for the purpose of contradiction that some $u \in N(R) \setminus Q$ has exactly one neighbour in R (call that neighbour w). But then the u -entry of $A(G)\mathbf{v}$ is the same as the w -entry of \mathbf{v} , which is impossible (recall that $w \in R = \text{supp}(\mathbf{v})$ while $u \notin Q = \text{supp}(A(G)\mathbf{v})$). \square

We also need the following lemma deducing the precise corank of the adjacency matrix from information about its stalks and special cycles. Recall that a special cycle in G is an induced cycle with length divisible by 4, such that every second vertex has degree 2 in G . Recall that $s(G)$ is the number of special cycles, counting isolated special cycles twice.

Lemma 11.6. Let G be an n -vertex graph, and let

$$K^{(\eta)} = \{\mathbf{v} \in \ker A(G) : |\text{supp}(x)| \leq \eta n.\}$$

Suppose that every $(\leq \eta n, 0)$ -stalk R is an $(|R|, |R|, 0)$ -stalk which satisfies $e(R \cup N_G(R)) \leq 2|R|$ and $\deg_G(v) \geq 2$ for all $v \in R$. Also, suppose that all special cycles in G have length at most ηn and are vertex-disjoint from each other.

- (1) $\text{supp} K^{(\eta)}$ is precisely the set of degree-2 vertices in special cycles of G .
- (2) $\dim K^{(\eta)} = s(G)$.

We emphasise that **Lemma 11.6** is a non-probabilistic statement about general graphs G (though we will eventually apply it to the random graphs G_t defined in **Section 10**).

Proof. Let V^{spec} be the set of degree-2 vertices in special cycles of G , and suppose the vertex set of G is $\{1, \dots, n\}$. Recall from [Fact 1.6](#) that special cycles give rise to kernel vectors, so $\text{supp } K^{(\eta)} \supseteq V^{\text{spec}}$. Also, since the special cycles in G are vertex-disjoint, each of the $s(G)$ kernel vectors obtained in this way have disjoint supports, so are linearly independent. This shows that $\dim K^{(\eta)} \geq s(G)$.

Recalling [Lemma 11.3](#), to prove that $\dim K^{(\eta)} = s(G)$ it now suffices to show that every nonzero minimal kernel vector of $A(G)$ whose support size is at most ηn is a multiple of one of the explicit kernel vectors arising from special cycles via [Fact 1.6](#). To this end, consider a nonzero minimal kernel vector $\mathbf{v} = (v_1, \dots, v_n)$, let $R = \text{supp}(\mathbf{v})$, and suppose $|R| \leq \eta n$. Then, by [Lemma 11.5](#) and the assumption in the lemma, $R = \text{supp}(\mathbf{v})$ is an $(|R|, |R|, 0)$ -stalk consisting of vertices with degree at least 2, such that $e(R \cup N_G(R)) \leq 2|R|$. This is only possible if $R \cup N_G(R)$ is an induced cycle in which the vertices in R have degree exactly 2. Write u_2, u_4, \dots, u_{2q} (in order) for the vertices of this cycle, where $u_2, u_4, \dots, u_{2q} \in R$.

For each $u_i \in N_G(R)$ (with i odd), the u_i -coordinate of $A\mathbf{v}$ is precisely $v_{u_{i-1}} + v_{u_{i+1}}$. Since \mathbf{v} is a kernel vector, each of u_2, u_4, \dots, u_{2q} must have the same absolute value, and cyclically alternate their signs. This is only possible if q is even (i.e., if $R \cup N_G(R)$ induces a special cycle), and implies that \mathbf{v} is a multiple of one of the explicit kernel vectors arising from [Fact 1.6](#). \square

Next, the following lemma shows how to establish the η -unstructuredness property in [Definition 8.3](#) using information about stalks.

Lemma 11.7. *For $0 < \eta < 1/2$, consider a graph G with at most $n/10$ different $(\leq \eta n, 1)$ -stalks and at most $n^2/10$ different $(\leq \eta n, 2)$ -stalks, where additionally, $A(G)$ has no 2-almost kernel vector with a nonzero level set of size at least $(1 - \eta)n$. Then the adjacency matrix $A(G)$ is η -unstructured.*

Proof. Since we have assumed $A(G)$ has no 2-almost kernel vector with a nonzero level set of size at least $(1 - \eta)n$, we only need to consider unbalanced almost-kernel vectors with a large zero level set.

We first claim that whenever there is a vector $\mathbf{v} \in \mathbb{R}^n$ with $\text{supp}(A\mathbf{v}) = \{i, i'\}$, there is a *minimal* vector $\mathbf{w} \in \mathbb{R}^n$ with $\text{supp}(A\mathbf{w}) \in \{\{i\}, \{i'\}, \{i, i'\}\}$ and $\text{supp}(\mathbf{w}) \subseteq \text{supp}(\mathbf{v})$. Indeed, consider the decomposition into minimal vectors given by [Lemma 11.3](#). For all of these vectors \mathbf{w} we have $\text{supp}(A\mathbf{w}) \subseteq \{i, i'\}$, and it cannot be the case that all of these vectors are kernel vectors of A (otherwise \mathbf{v} would be a kernel vector as well). Since $\text{supp}(\mathbf{w}) \subseteq \text{supp}(\mathbf{v})$, if \mathbf{v} is non- η -balanced with a large zero level set, then \mathbf{w} is also non- η -balanced.

Then, using [Lemma 11.5](#), the assumptions in the lemma imply that there are at most $n/10$ different i for which there is a non- η -balanced vector $\mathbf{v} \in \mathbb{R}^n$ with $\text{supp}(A\mathbf{v}) = \{i\}$, and there are at most $n^2/10$ different pairs $\{i, i'\}$ for which there is a non- η -balanced vector $\mathbf{v} \in \mathbb{R}^n$ with $\text{supp}(A\mathbf{v}) = \{i, i'\}$. It follows that there are at least $n(n-1) - 2n(n/10) - 2(n^2/10) \geq \eta n^2$ pairs of distinct indices (i, i') for which every \mathbf{v} with $\text{supp}(A\mathbf{v}) = \{i, i'\}$ is η -balanced, meaning that $A(G)$ is η -unstructured. \square

Essentially all that remains is to carefully analyse the stalks that exist in the random graphs G_N, \dots, G_0 defined in [Section 10](#).

11.1. Estimates on stalks. Recall the definitions of the random graphs G_N, \dots, G_0 from [Section 10](#): to obtain G_N we looked at the degrees of the first $\lfloor \alpha n \rfloor$ vertices of a random graph $G \sim \mathcal{K}(n, m, 2)$ and deleted the vertices with degree at least Δ , then we added back these vertices in a random order to obtain G_{N-1}, \dots, G_0 .

Crucially, a similar proof as for [Lemma 6.1](#) shows that for each $t \leq \alpha n$, if we condition on the degree sequence of G_t (more precisely, we condition on the event $t \leq N$ and then further condition on the degree sequence), then G_t is distributed like a uniformly random graph with that degree sequence. So, we perform various calculations after conditioning on properties of the degree sequence of G_t . Specifically, the properties we need are as follows.

Definition 11.8. Consider integers $m > n, t$ and some $\kappa > 0$. Choose $\lambda > 0$ such that if $Z \sim \text{Poisson}(\lambda)$, then $2m/n = \mathbb{E}[Z|Z \geq 2]$. A sequence $\mathbf{d} = (d_1, \dots, d_{n-t}) \in \mathbb{R}^{n-t}$ is (n, m, t, κ) -*typical* if it satisfies the following properties.

- T1** $d_v = 0$ for at most $(t/n)^{3/2}n$ different v .
- T2** $d_v = 1$ for at most $\kappa^{-1}t \log(n/t)$ different v . (Here we use the convention $0 \log \infty = 0$ for the case $t = 0$.)
- T3** $d_v = 2$ for at most $(\Pr[Z = 2|Z \geq 2] + \kappa)n$ different v .
- T4** $2(1 - \kappa)m \leq d_1 + \dots + d_{n-t} \leq 2m$.
- T5** For any $U \subseteq \{1, \dots, n-t\}$ with $|U| = u$, we have $\sum_{v \in U} d_v \leq \kappa^{-1}u \log(2n/u)$.
- T6** $\sum_v \binom{d_v}{2} \leq \left(E \left[\binom{Z}{2} \middle| Z \geq 2 \right] + \kappa \right) n$, and $\sum_v \binom{d_v}{j} \leq \kappa^{-j}n$ for $j \geq 3$.

We remark that most of these bounds are essentially sharp, for a typical outcome of the degree sequence of G_t . The exception is **T1**: the number of isolated vertices is typically about $(t \log(n/t)/n)^2 n$ (but we will not need such a strong estimate).

Lemma 11.9. *Fix $\varepsilon, \alpha, \Delta, \kappa > 0$ such that $1/\Delta \ll \alpha \ll \kappa \ll \varepsilon$. Recall the definitions of G_N, \dots, G_0 (in terms of α, Δ) from [Section 10](#). Then, for each $t \leq \alpha n$: with probability at least $1 - (t/n)^{1/4}$, if $t \leq N$ then the degree sequence of G_t is (n, m, t, κ) -typical.*

Recall that $N = o_{\Delta \rightarrow \infty}(n)$ whp, so when we take $\Delta \rightarrow \infty$, [Lemma 11.9](#) becomes a with-high-probability statement.

Proof. First, note that by [Lemma 6.2\(A4\)](#), with probability at least $1 - n^{-\omega(1)}$ we have $\sum_{v \in S} \deg_G(v) = O_\varepsilon(s \log(2n/s))$ for each size- s subset $S \subseteq V(G)$. This directly yields **T5** since the degrees in G_t are at most those in G . Further, [Lemma 6.2\(A4\)](#) yields that the sum of degrees of the vertices in $V(G) \setminus V(G_t)$ (i.e., the vertices v_{t-1}, \dots, v_0) is $O_\kappa(t \log(2n/t)) \leq \kappa n \leq \kappa m$ for $\alpha \ll \kappa$ (recalling that $t \leq \alpha n$). This yields **T4** and additionally **T2**, since the number of degree 1 vertices is at most the number of edges from $V(G) \setminus V(G_t)$. Similarly **T3** and **T6** hold with probability $1 - n^{-\omega(1)}$, by [Lemma 6.2\(A1\)](#) and (A2,A3) respectively.

For **T1**, we need a simple calculation in the configuration model. Condition on an outcome of the degree sequence \mathbf{d} of G , satisfying the conclusions of [Lemma 6.2](#). This determines T ; also condition on an outcome of $V(G) \setminus V(G_t) = \{v_{t-1}, \dots, v_0\}$, and let $T_t = T \cap (V(G) \setminus V(G_t))$ (so $|T_t| = t$). By [Lemma 6.1](#), after our conditioning, we have $G \sim \mathbb{G}(\mathbf{d})$. By [Lemma 6.5\(A\)](#), it suffices to prove the desired result for $G \sim \mathbb{G}^*(\mathbf{d})$ (i.e., we may work in the configuration model).

A vertex v can only be isolated in G_t if it has at least two neighbours in T_t . The number of stubs corresponding to the vertices in T_t is $O_\kappa(t \log(2n/t))$, so the probability that this happens is $O_\kappa(t \log(2n/t)/n)^2 \leq (t/n)^{1.9}$ for $\alpha \ll \kappa$. That is to say, the expected number of isolated vertices is at most $(t/n)^{1.9} n$, so **T1** holds with probability at least $1 - (t/n)^{1/4}$ by Markov's inequality. \square

Now, the following definition captures the stalks which are not handled by [Lemma 6.6](#).

Definition 11.10. Say a stalk R is ε -sparse if there is no subset of vertices in $R \cup N(R)$ of any size u which spans more than $u + \lfloor C_{6.6}(\varepsilon)u/\sqrt{\log(2n/u)} \rfloor$ edges.

The following lemma encapsulates a careful analysis of small sparse stalks in random graphs with a given typical degree sequence.

Lemma 11.11. *Fix $\eta, \varepsilon, \kappa, \alpha$ such that $\alpha \ll \eta \ll \kappa \ll \varepsilon \ll 1$. Choose n, m, t with $1 + \varepsilon \leq m/n \leq 1/\varepsilon$ and $t \leq \alpha n$, let $\mathbf{d} \in \mathbb{R}^{n-t}$ be a (n, m, t, κ) -typical sequence, and let $G_t \sim \mathbb{G}(\mathbf{d})$.*

For $s \geq r - 1 + \ell$ and $1 \leq r \leq \eta n$, let $X_{r,s,\ell}$ be the number of ε -sparse (r, s, ℓ) -stalks in G_t , and let $X'_{r,s,\ell}$ be the number of such stalks R for which there is a vertex $v \in R$ with $d_v = 1$. Then we have the following estimates.

- (1) *If $s \geq r + 1$, then $\mathbb{E}X_{r,s,0} \lesssim_\kappa e^{-\Omega_\varepsilon(r)}/n$.*
- (2) *$\mathbb{E}X_{r,r,0} \lesssim_\kappa e^{-\Omega_\varepsilon(r)}$.*
- (3) *If $t \leq n^{1/8}$ then $\mathbb{E}X'_{r,r,0} \lesssim_\kappa e^{-\Omega_\varepsilon(r)} n^{-3/4}$.*
- (4) *If $s = r - 1$, then $\mathbb{E}X_{r,s,0} \lesssim_\kappa e^{-\Omega_\varepsilon(r)} (t/n)^{3/2} n$.*
- (5) *For any $\ell \leq 2$, we have $\mathbb{E}X_{r,s,\ell} \lesssim_\kappa e^{-\Omega_\varepsilon(r)} n$.*
- (6) *If $t \leq \log n$, then $\mathbb{E}X_{r,s,1} \lesssim_\kappa e^{-\Omega_\varepsilon(r)} (\log n)^2$.*

We emphasise that the above estimates are only for $r \leq \eta n$ (i.e., for those stalks that correspond to small-support kernel vectors of $A(G_t)$).

We remark that our notion of a stalk has some resemblance to the notion of a *flipper* in [\[22, Section 8\]](#), and [\[22, Lemma 8.1\]](#) plays a similar role to [Lemma 11.11](#). However, in our setting we need much more precision, and the details are much more involved.

For our proof of [Lemma 11.11](#) we collect some elementary estimates. First, we will need to estimate products of factorials.

Lemma 11.12. *If $(k_i)_{i=1}^r$ is a sequence of nonnegative integers with $\sum_{i=1}^r k_i = a$ and $\sum_{i=1}^r ik_i = b$ then*

$$\prod_{i=1}^r k_i! \geq e^{-b} a!$$

Proof. By the multinomial theorem we have

$$\binom{a}{k_1, \dots, k_r} k_1^{k_1} \dots k_r^{k_r} \leq (k_1 + \dots + k_r)^a = a^a.$$

So,

$$\frac{a!}{\prod_{j=1}^r k_j!} = \binom{a}{k_1, \dots, k_r} \leq \prod_j \binom{a}{k_j}^{k_j} = \exp\left(a \sum_j \frac{k_j}{a} \log\left(\frac{a}{k_j}\right)\right).$$

We can interpret the right-hand side as $e^{aH(Y)}$, where $H(Y)$ is the (base- e) entropy of a random variable Y satisfying $\Pr[Y = i] = k_i/a$ for each $i \in \{1, \dots, r\}$. Note that $\mathbb{E}Y = b/a$; among positive integer random variables with this mean, the maximum possible entropy is attained by a geometric random variable with parameter $p := a/b$ (see for example [51]). The entropy of such a geometric random variable is $(-p \log p - (1-p) \log(1-p))/p \leq 1 - \log p$. So,

$$\frac{a!}{\prod_{j=1}^r k_j!} \leq e^{aH(Y)} \leq e^{a(1-\log p)} = (e/p)^{pb} \leq e^b,$$

using the inequality $(e/p)^p \leq e$ (which holds for all $0 \leq p \leq 1$). \square

We also need the following general-purpose inequality to bound various binomial coefficients.

Lemma 11.13. *For any $z > 0$, and any $a, b \in \mathbb{N}$, we have $\binom{a}{b} \leq (1+z)^a e^{O_z(b)}$.*

Proof. If $4b \geq z^2 a$, we have $\binom{a}{b} \leq (ae/b)^b \leq (4e/z^2)^b = e^{O_z(b)}$. Otherwise, if $4b < z^2 a$, writing $x = b/a$, we have $\binom{a}{b} \leq (ae/b)^b = ((e/x)^x)^a \leq (1+2\sqrt{x})^a \leq (1+z)^a$. (Here we used the inequality $(e/x)^x \leq 1+2\sqrt{x}$, which holds for all $x \geq 0$.) \square

We are now ready to prove [Lemma 11.11](#).

Proof of Lemma 11.11. In this proof we think of ε as being a constant (without explicitly writing ε as a subscript on asymptotic notation), and we simply write “sparse” instead of “ ε -sparse”. Also, throughout this proof we let $f(s) = \lfloor C_{6.6}(\varepsilon)s/\sqrt{\log(2n/s)} \rfloor$. Note that f is essentially sub-linear, in the sense that $f(a+b) \leq f(a) + f(b) + 1$.

First, we briefly note that isolated vertices are $(1, 0, 0)$ -stalks (and by [S2](#), isolated vertices are not contained in any other types of stalks). By [T1](#), the number of isolated vertices is at most $(t/n)^{3/2}n \lesssim n$, which handles the $(r, s) = (1, 0)$ cases of (4) and (5). For the rest of the proof we can restrict our attention to (r, s, ℓ) -stalks which do not contain any isolated vertices.

The reader may find it helpful to think of two basic examples of sparse stalks that may occur in graphs with minimum degree at least 2. First, for any even cycle in which every second vertex has degree 2, we can take those degree-2 vertices as a $(r, r, 0)$ -stalk. Second, for any odd cycle in which every vertex has degree 2, we can take the entire vertex set of the cycle as an $(r, r, 0)$ -stalk. It is not hard to estimate the expected number of these types of cycles using the configuration model.

Roughly speaking, the proof strategy is as follows. First, we prove a sequence of inequalities ([Claim 11.14](#)) showing that *every* stalk approximately resembles a union of copies of these two examples. Then, we do an explicit configuration-model calculation that parallels the cycle calculation mentioned above.

For this entire proof we will work with the configuration model $G_t \sim \mathbb{G}^*(\mathbf{d})$, taking $X_{r,s,\ell} = X'_{r,s,\ell} = 0$ whenever G_t is not simple. (By [Lemma 6.5](#), it suffices to prove the desired estimates in this setting, noting that [T6](#) implies that $\sum_v d_v^2 = O(n)$.)

Step 1: Parameters of stalks. Fix r, s, ℓ with $s \geq r - 1 + \ell$ and $\ell \leq 2$. We define a number of parameters of a sparse (r, s, ℓ) -stalk R in G_t . We will later study the contribution to $\mathbb{E}X_{r,s,\ell}$ and $\mathbb{E}X'_{r,s,\ell}$ from each choice of these parameters.

- Let $S = N(R)$, let $S_1 = R_1 = S \cap R$, let $S_2 = S \setminus R$ and let $R_2 = R \setminus S$. For each $i \in \{1, 2\}$, let $r_i = |R_i|$ and $s_i = |S_i|$. (So, $s = s_1 + s_2$ and $r = r_1 + r_2$ and $r_1 = s_1$.)
- For each $i \in \{1, 2\}$, let ℓ_i be the number of exceptional vertices in S_i . (So, $\ell = \ell_1 + \ell_2$.)
- Let x be the number of $v \in R_2$ which have $\deg_{G_t}(v) = 1$.
- Let m_1 be the number of edges in R_1 , let $m_{1,2}$ be the number of edges between R_1 and S_2 , and let $m_{2,2}$ be the number of edges between R_2 and S_2 .
- For $i \geq 1$, let k_i be the number of vertices in S_2 which have exactly i neighbours in R (so in particular $k_1 = \ell_2$).

There are a number of simple inequalities that must hold between our parameters. First, by **S1**, we have

$$s_2 - r_2 = s - r \geq \ell - 1. \quad (11.1)$$

Second, by ε -sparsity, we have

$$m_1 \leq r_1 + f(r_1), \quad m_{2,2} \leq r_2 + s_2 + f(r_2 + s_2), \quad m_1 + m_{1,2} + m_{2,2} \leq r_1 + r_2 + s_2 + f(r_1 + r_2 + s_2) \quad (11.2)$$

Third, recall from **S3** that all non-exceptional vertices in S have at least two neighbours in R . By the considerations at the start of the proof, we are assuming R contains no isolated vertices, and by definition R has exactly x vertices with degree 1 into S . So, summing over degrees in R_1, R_2, S_2 , we obtain

$$m_1 \geq \lceil (2r_1 - \ell_1)/2 \rceil = r_1 - \lfloor \ell_1/2 \rfloor, \quad m_{1,2} + m_{2,2} \geq 2s_2 - \ell_2, \quad m_{2,2} \geq 2r_2 - x. \quad (11.3)$$

Finally, using **T5**, we have

$$m_{1,2} \leq O_\kappa(s_2 \log(2n/s_2)). \quad (11.4)$$

(Other similar inequalities can also be obtained via **T5**, but we will not need them.)

Step 2: The structure of sparse stalks. We now combine the above inequalities, to prove the following claim about the parameters of a sparse (r, s, ℓ) -stalk R . Roughly speaking, the claim says that if we consider two disjoint copies of R and S and a bipartite graph of the edges between the two, then almost all vertices have degree 2, and almost all edges are inside $R_1 = S_1$ or between R_2 and S_2 . (Note that there are no edges between R_2 and $S_1 = R_1$, by the definition of R_1 .)

Claim 11.14. *Consider a sparse (r, s, ℓ) -stalk R , with parameters as defined as in Step 1.*

- (1) $s_2 = r_2 + O(1 + f(r))$ (i.e., S and R have roughly the same size).
- (2) All but $O(1 + x + f(r))$ vertices in R have degree exactly 2.
- (3) $k_2 = s_2 + O(1 + x + f(r))$ (i.e., almost all vertices in S_2 have degree exactly 2 into R).
- (4) $m_{1,2} = O(1 + x + f(r))$ (i.e., there are few edges between R_1 and S_2).
- (5) $2s_2 + O(1 + x + f(r)) \leq m_{2,2} \leq 2s_2 + O(1 + f(s_2))$ (i.e., the number of edges between S_2 and R_2 is not much more than $2s_2$, which by (1) is roughly the same as $2r_2$).
- (6) $m_1 = s_1 + O(1 + f(r_1))$ (i.e., the number of edges inside $S_1 = R_1$, which is half its degree sum, is not much more than $s_1 = (2s_1)/2$).

In light of **Remark 6.7** and since $r/n \leq \eta$, if $\eta \ll \kappa$ we have $f(q) \leq \kappa q$ for $q \in \{s_1, s_2, r\}$ (we are also using **Claim 11.14**(1) here to show s/n is small). In particular, throughout the rest of the proof, terms of the form $f(q)$ can be viewed as being “lower order” than q .

Proof. First, (6) follows from the first inequalities in (11.2) and (11.3).

Next, by combining (6), the last inequality in (11.2), and the second inequality in (11.3), we obtain $s_2 \leq r_2 + O(1 + f(r + s_2))$. Together with (11.1), this nearly gives us (1), but we need to do a little more work to replace the error term “ $O(1 + f(r + s_2))$ ” with the desired error term “ $O(1 + f(r))$ ”. Specifically, to show that these error terms are equivalent, we need to prove that $s_2 \lesssim r$. By **T5** we have $s \leq \sum_{v \in R} d_v \leq \kappa^{-1} r \log(2n/r) \lesssim \eta^{1/2} n$ (assuming $\eta \ll \kappa$), so $f(r + s_2) \lesssim (r + s_2)/\sqrt{\log(1/\eta)}$ and thus our initial inequality implies $s_2 \lesssim r$, as desired.

Then, (5) follows from (11.1), the second inequality in (11.2), (1), and the last inequality in (11.3). After this, we can deduce (4) from (1), (5), (6), and the last inequality in (11.2).

Finally, by (1,4,5,6), note that

$$\sum_{v \in R} (\deg(v) - 2) = 2m_1 + m_{1,2} + m_{2,2} - 2r \lesssim 1 + x + f(r), \quad \sum_{v \in S_2} (\deg_R(v) - 2) = m_{1,2} + m_{2,2} - 2s_2 \lesssim 1 + x + f(r).$$

Recall that at most $x \lesssim 1 + x + f(r)$ vertices in R have degree less than 2, and at most $\ell \lesssim 1 + x + f(r)$ vertices in S have fewer than 2 neighbours in R . So, (2) and (3) follow. \square

Step 3: Breaking down the expectation. For a vector of parameters

$$\mathbf{p} = (r_1, r_2, s_1, s_2, \ell_1, \ell_2, x, m_1, m_{1,2}, m_{2,2}, (k_i)_{i=1}^r),$$

we now consider the contribution to $\mathbb{E}X_{r,s,\ell}$ from sparse stalks with these parameters. We will eventually sum over all possible \mathbf{p} . (For $\mathbb{E}X'_{r,r,0}$, we simply sum over all \mathbf{p} with $x > 0$.)

Recall that we are working in the configuration model $\mathbb{G}^*(\mathbf{d})$, for a particular (n, m, t, κ) -typical degree sequence $\mathbf{d} = (d_1, \dots, d_{n-t})$ (so we have $n - t$ buckets corresponding to vertices, and within the bucket corresponding to a vertex v , there are d_v stubs).

First, we define $N_{\mathbf{p}}$ to be “the number of possible places that a stalk may appear”. Specifically, $N_{\mathbf{p}}$ is the number of ways to choose disjoint vertex sets R_1, R_2, S_2 , and to colour all the stubs from R_1 blue and yellow, and to colour some stubs from S_2 red, and all the stubs from R_2 green, such that the following hold.

- $|R_1| = r_1, |R_2| = r_2, |S_2| = s$.
- There are $2m_1 + m_{1,2}$ stubs coming from R_1 . Exactly $2m_1$ are blue and exactly $m_{1,2}$ are yellow.
- There are exactly $m_{2,2}$ stubs coming from R_2 , all coloured green.
- Among the stubs from S_2 , exactly $m_{1,2} + m_{2,2}$ are red.
- Each vertex in S_2 has at least two red stubs, except exactly ℓ_2 which have one red stub.
- Exactly x of the vertices in R_2 have degree 1.
- Exactly ℓ_1 of the vertices in R_1 have degree 1, and none have degree 0.

Then, for each of the choices of R_1, R_2, S_2 and red/blue/yellow colourings as above, we consider the probability that

- the $2m_1$ blue stubs (from R_1) pair with each other, and
- the $m_{1,2}$ yellow stubs (from R_1) pair with red stubs (from S_2), and
- the $m_{2,2}$ stubs from R_2 pair with red stubs (from S_2).

This probability only depends on \mathbf{p} ; denote it by $P_{\mathbf{p}}$. Observe that $\mathbb{E}X_{r,s,\ell} \leq \sum_{\mathbf{p}} N_{\mathbf{p}} P_{\mathbf{p}}$.

Step 4: Estimating combinatorial quantities. Let Q be the number of possibilities for \mathbf{p} . We now give upper bounds for $Q, N_{\mathbf{p}}$, and $P_{\mathbf{p}}$. We will very often want to use the expression “ $(1 + O(\kappa))^r e^{O_\kappa(1+x+f(r))}$ ” as a multiplicative error term, so we introduce the shorthand “ $O^*(1)$ ” for a term of this form.

Claim 11.15. $Q = O^*(1)$.

Proof. Recall the definitions of the various parameters from Step 1, and recall from Claim 11.14 that $s, m_1, m_{1,2}, m_{2,2} \lesssim r$. It is easy to see that there are at most $(\ell + 1)(r + 1)^2(s + 1) \lesssim r^3$ choices for $r_1, r_2, s_1, s_2, \ell_1, \ell_2$, at most $O(r^3)$ choices for $m_1, m_{1,2}, m_{2,2}$, and at most $r + 1$ choices for x . Then, note that $\sum_{i=1}^r ik_i = m_{1,2} + m_{2,2}$, so $(k_i)_{i=1}^r$ encodes an integer partition of $m_{1,2} + m_{2,2}$ (k_i is the number of parts of size i). For each $m_{1,2}, m_{2,2}$, the number of such partitions is $e^{O(\sqrt{m_{1,2} + m_{2,2}})} = e^{O(\sqrt{r})}$. \square

Claim 11.16. For any \mathbf{p} we have

$$P_{\mathbf{p}} \leq O^*(1) \frac{r_1^{m_1} s_2^{m_{1,2} + m_{2,2}}}{e^{r_1 + 2r_2} \eta^{m_1 + m_{1,2} + m_{2,2}}}.$$

Proof. Let $d_\Sigma = d_1 + \dots + d_{n-t} = (2 + O(\kappa))m$. First note that

$$P_{\mathbf{p}} \leq \frac{(2m_1)!!(d_\Sigma - 2m_1)!!}{d_\Sigma!!} \cdot \binom{m_{1,2} + m_{2,2}}{m_{1,2}} \cdot \frac{m_{1,2}! m_{2,2}!}{(d_\Sigma - 2m_1)_{m_{1,2} + m_{2,2}}}.$$

Indeed, the first term accounts for the probability that the $2m_1$ blue stubs pair with each other, the second term is the number of ways to choose which of the $m_{1,2} + m_{2,2}$ red stubs will pair with yellow stubs and which will pair with green stubs, and the last two terms bound the probability that the red stubs do indeed pair with the yellow and green stubs in this way.

We now just need to manipulate the above expression using the inequalities in Claim 11.14. Throughout, we will use the crude bounds that $m_1, m_{1,2}, m_{2,2} \lesssim r$ without further remark.

First, we have $\binom{m_{1,2} + m_{2,2}}{m_{1,2}} \leq (1 + \kappa)^{m_{1,2} + m_{2,2}} e^{O_\kappa(m_{1,2})}$ by Lemma 11.13, and

$$\frac{(2m_1)!!(d_\Sigma - 2m_1)!!}{d_\Sigma!!} = \frac{\binom{d_\Sigma/2}{m_1}}{\binom{d_\Sigma}{2m_1}} \leq (1 + O(\kappa))^{m_1} \frac{\binom{m}{2m_1}}{\binom{m}{2m_1}} \leq (1 + O(\kappa))^r \left(\frac{m_1}{em}\right)^{m_1},$$

provided $\eta \ll \kappa$ (recall that $m_1 \lesssim r \leq \eta m$, while $m \geq n$). Then (again with $\eta \ll \kappa$) we have $(d_\Sigma - 2m_1)_{m_{1,2}} = (d_\Sigma - O(r))^{m_{1,2} + m_{2,2}} = (1 + O(\kappa))^{O(r)} (d_\Sigma)^{m_{1,2} + m_{2,2}}$. Using Stirling’s formula, we therefore have

$$P_{\mathbf{p}} \leq (1 + O(\kappa))^{O(r)} e^{O_\kappa(m_{1,2})} \sqrt{m_{1,2} m_{2,2}} \left(\frac{m_1}{em}\right)^{m_1} \left(\frac{m_{1,2}}{2em}\right)^{m_{1,2}} \left(\frac{m_{2,2}}{2em}\right)^{m_{2,2}}.$$

Next, using Claim 11.14(4,5,6), we have $m_1 = r_1 + O(1 + f(r_1)) = (1 + O(\kappa))r_1 + O(1)$, $m_{1,2} = O(1 + x + f(r))$, and $m_{2,2} \leq 2s_2 + O(1 + f(s_2)) = (1 + O(\kappa))2s_2 + O(1)$. So, we deduce

$$P_{\mathbf{p}} \leq O^*(1) \frac{r_1^{m_1} m_{1,2}^{m_{1,2}} s_2^{m_{2,2}}}{e^{r_1 + 2r_2} \eta^{m_1 + m_{1,2} + m_{2,2}}}.$$

Now, to finish the proof of the claim it suffices to show that $(m_{1,2}/s_2)^{m_{1,2}} \leq O^*(1)$. We distinguish cases.

- *Case 1:* $f(r) + x \leq 1$. We have $m_{1,2} \lesssim 1$ by [Claim 11.14\(4\)](#), so $(m_{1,2}/s_2)^{m_{1,2}} \lesssim 1$.
- *Case 2:* $x \geq f(r) + 2$. Recall that x counts degree-1 vertices in R_2 , so $x \leq r_2$, and recall from [\(11.1\)](#) that $s_2 \geq r_2 - 1$. Also, [Claim 11.14\(4\)](#) implies that $m_{1,2} \lesssim x$. So, $s_2 \geq r_2 - 1 \geq x - 1 \geq x/2 \gtrsim m_{1,2}$, meaning that $(m_{1,2}/s_2)^{m_{1,2}} \leq e^{O(x)}$.
- *Case 3:* $f(r) + x \geq 2$ and $x \leq f(r) + 1$. In this remaining case, note that $f(r) \gtrsim 1 + x$. Since $s_2/n \leq 1/2$, from [\(11.4\)](#) we have $m_{1,2} \leq O_\kappa(s_2 \log(n/s_2))$, which implies that

$$\frac{m_{1,2}}{\log(n/m_{1,2})} \leq \frac{O_\kappa(s_2 \log(n/s_2))}{\log(n/O_\kappa(s_2 \log(n/s_2)))} = s_2 \frac{O_\kappa(\log(n/s_2))}{\log((n/s_2)/O_\kappa(\log(n/s_2)))} \leq O_\kappa(s_2).$$

Here in the last inequality we have used the fact that $s_2/n \lesssim r/n \leq \eta$ is small relative to κ (meaning that the denominator $\log((n/s_2)/O_\kappa(\log(n/s_2)))$ is at least say $\log(n/s_2)/2$). Now

$$\begin{aligned} \log\left(\left(\frac{m_{1,2}}{s_2}\right)^{m_{1,2}}\right) &= m_{1,2} \log(O_\kappa(\log(n/m_{1,2}))) \lesssim m_{1,2} \log \log\left(\frac{n}{m_{1,2}}\right) \\ &\lesssim f(r) \log \log\left(\frac{n}{f(r)}\right) \lesssim r \left(\frac{\log \log(n/r)}{\sqrt{\log(n/r)}}\right) \leq r \log(1 + \kappa), \end{aligned}$$

provided $\eta \ll \kappa$. Exponentiating yields the desired result. \square

Claim 11.17. For any \mathbf{p} we have

$$N_{\mathbf{p}} \leq O^*(1) \frac{n^{r_1+r_2+s_2}}{r_1^{r_1} s_2^{r_2+s_2}} \cdot e^{r_1+2r_2} \rho_2^{r_1+r_2} \left(\frac{\log(n/t)}{n/t}\right)^x E_2^{r_2},$$

where $\rho_2 = \Pr[Z = t | Z \geq 2]$ and $E_2 = \mathbb{E}\left[\binom{Z}{2} \mid Z \geq 2\right]$ for $Z \sim \text{Poisson}(\lambda)$, where λ is such that $2m/n = \mathbb{E}[Z | Z \geq 2]$.

Proof. Recall that our degree sequence is (n, m, t, κ) -typical. First, we bound the number of choices of R_1 . Recall from [Claim 11.14\(2\)](#) that there is some $i = O(1 + x + f(r))$ such that at least $r_1 - i$ vertices in R_1 have degree exactly 2. The number of ways to choose a sequence of r_1 vertices, for which the first $r_1 - i$ have degree exactly 2, is at most $(\rho_2 + \kappa)^{r_1 - i} n^{r_1} = (\rho_2 + \kappa)^{r_1} (O(1))^i n^{r_1}$. Also, there are $\binom{2m_1 + m_{1,2}}{m_{1,2}}$ ways to choose a blue/yellow colouring of the stubs from R_1 . In total, the number of choices of R_1 and a suitable blue/yellow colouring of its stubs is at most

$$\binom{r_1}{i} \cdot \frac{(\rho_2 + \kappa)^{r_1} (O(1))^i n^{r_1}}{r_1!} \cdot \binom{2m_1 + m_{1,2}}{m_{1,2}} \leq O^*(1) \left(\frac{e n \rho_2}{r_1}\right)^{r_1} \quad (11.5)$$

where we used [Lemma 11.13](#) twice (with $z = \kappa$) and Stirling's inequality.

Second, we bound the number of choices for R_2 . Recall that x of the vertices in R_2 have degree exactly 1, and there is some $j = O(1 + x + f(r))$ such that at least $r_2 - x - j$ of the other vertices in R_2 have degree exactly 2. By [T2](#) and [T3](#), the number of ways to choose a sequence of r_2 vertices, of which the first x have degree exactly 1, and the next $r_2 - x - j$ have degree exactly 2, is at most $O^*(1) ((t \log(n/t)/n)^x (\rho_2 + \kappa)^{r_2 - x - j} n^{r_2})$. The number of choices of R_2 is therefore at most

$$\binom{r_2}{x} \binom{r_2}{j} \cdot \frac{((t \log(n/t)/n)^x (\rho_2 + \kappa)^{r_2 - x - j} (O(1))^j n^{r_2})}{r_2!} = O^*(1) \left(\frac{e \rho_2 n}{r_2}\right)^{r_2} \left(\frac{t \log(n/t)}{n}\right)^x, \quad (11.6)$$

where we used [Lemma 11.13](#) twice and Stirling's inequality.

Third, the number of ways to choose S_2 , and to choose which of its stubs are red, is at most

$$\prod_{i=1}^r \frac{1}{k_i!} \left(\sum_{v \in V(G_i)} \binom{d_v}{i} \right)^{k_i} \leq \frac{1}{k_2!} (E_2 + \kappa)^{k_2} \frac{e^{m_{1,2} + m_{2,2} - 2k_2 - \ell_2}}{(s_2 - k_2 - \ell_2)!} (1/\kappa)^{m_{1,2} + m_{2,2} - 2k_2 - \ell_2} n^{s_2}.$$

Here we used that $\sum_v \binom{d_v}{1} \leq m \lesssim n$ by [T4](#), we used [T6](#), and we used [Lemma 11.12](#) applied to $(k_i)_{i \geq 2}$, noting that $\sum_{i=1}^r k_i = s_2$, $\sum_{i=1}^r i k_i = m_{1,2} + m_{2,2}$ and $k_1 = \ell_2$.

Now, by [Claim 11.14\(1,2,5\)](#), we have $m_{1,2} + m_{2,2} - 2k_2 - \ell_2 = O(1 + x + f(r))$, so the above expression is bounded by

$$O^*(1) \frac{(E_2 n)^{s_2}}{k_2! (s_2 - k_2 - \ell_2)!} = O^*(1) \binom{s_2}{k_2} \frac{(E_2 n)^{s_2}}{s_2!} \leq O^*(1) \left(\frac{e E_2 n}{s_2}\right)^{s_2}. \quad (11.7)$$

(we have used the fact that $s_2 = k_2 + O(1+x+r)$, [Lemma 11.13](#), and Stirling's inequality).

Multiplying the expressions in [\(11.5\)](#) to [\(11.7\)](#) (counting the number of ways to choose R_1, R_2, S_2 , and their stub-colourings) shows that

$$N_{\mathbf{p}} \leq O^*(1) \frac{n^{r_1+r_2+s_2}}{r_1^{r_1} r_2^{r_2} s_2^{s_2}} \cdot e^{s_2} (e\rho_2)^{r_1+r_2} \left(\frac{\log(n/t)}{n/t} \right)^x E_2^{s_2}.$$

The desired result follows, noting that $s_2 = r_2 + O(1+x+f(r))$ by [Claim 11.14\(1\)](#) and hence $(s_2/r_2)^{r_2} = (1 + O(1+x+f(r))/r_2)^{r_2} = e^{O(1+x+f(r))} = O^*(1)$. \square

Step 5: Putting everything together. Let $E_{\mathbf{p}} = QN_{\mathbf{p}}P_{\mathbf{p}}$, so

$$\mathbb{E}X_{r,s,\ell} \leq \max_{\mathbf{p}} E_{\mathbf{p}}, \quad \mathbb{E}X'_{r,s,\ell} \leq \max_{\mathbf{p}:x>0} E_{\mathbf{p}}.$$

Combining [Claims 11.15](#) to [11.17](#), we have

$$E_{\mathbf{p}} \leq O^*(1) \left(\frac{\log(n/t)}{n/t} \right)^x \rho_2^{r_1} (\rho_2 E_2)^{r_2} \frac{n^{r_1+r_2+s_2} r_1^{m_1-r_1} s_2^{m_{1,2}+m_{2,2}-r_2-s_2}}{m^{m_1+m_{1,2}+m_{2,2}}}.$$

Now, $r_1, s_1 \leq 2r$, and $m_1 - r_1, m_{1,2} + m_{2,2} - r_2 - s_2 \lesssim 1 + x + f(r)$, by [Claim 11.14\(1,4,5,6\)](#). So, if $m_1 - r_1$ (respectively, $m_{1,2} + m_{2,2} - r_2 - s_2$) is nonnegative, then $\left(\frac{r_1}{r}\right)^{m_1-r_1} = O^*(1)$ (respectively, $\left(\frac{s_2}{r}\right)^{m_{1,2}+m_{2,2}-r_2-s_2} = O^*(1)$). By [\(11.1\)](#) and [\(11.3\)](#), $m_1 - r_1$ and $m_{1,2} + m_{2,2} - r_2 - s_2$ can only be very slightly negative (i.e., if either is negative, it is $O(1)$). In such a case, we again have $\left(\frac{r_1}{r}\right)^{m_1-r_1} = O^*(1)$ or $\left(\frac{s_2}{r}\right)^{m_{1,2}+m_{2,2}-r_2-s_2} = O^*(1)$, respectively. Also, $(m/n)^{s_2-r_2} = e^{O(1+x+f(r))}$. Putting all this together, we further bound

$$E_{\mathbf{p}} \leq O^*(1) \left(\frac{\log(n/t)}{n/t} \right)^x \left(\frac{\rho_2 n}{m} \right)^{r_1} \left(\frac{\rho_2 E_2 n^2}{m^2} \right)^{r_2} \left(\frac{r}{m} \right)^{m_1+m_{1,2}+m_{2,2}-r_1-r_2-s_2}.$$

Recalling the definitions of ρ_2 and E_2 in terms of a Poisson random variable Z , and recalling the choice of the Poisson parameter λ , we compute

$$\frac{\rho_2 n}{m} = \frac{2 \Pr[Z=2|Z \geq 2]}{\mathbb{E}[Z|Z \geq 2]} = \frac{2 \Pr[Z=2]}{\mathbb{E}[Z \mathbb{1}_{Z \geq 2}]} = \frac{2(\lambda^2 e^{-\lambda}/2)}{\lambda - \lambda e^{-\lambda}} = \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}} = 1 - \Omega_{\varepsilon}(1),$$

and similarly

$$\frac{\rho_2 E_2 n^2}{m^2} = \frac{4 \Pr[Z=2] \mathbb{E}\left(\frac{Z}{2}\right)}{\mathbb{E}[Z \mathbb{1}_{Z \geq 2}]^2} = \frac{4(\lambda^2 e^{-\lambda}/2)(\lambda^2/2)}{(\lambda - \lambda e^{-\lambda})^2} = \frac{\lambda^2 e^{-\lambda}}{(1 - e^{-\lambda})^2} = 1 - \Omega_{\varepsilon}(1).$$

Also, combining the inequalities in [\(11.3\)](#) in different ways, we can obtain $m_1 + m_{1,2} + m_{2,2} \geq s_1 + 2s_2 - \ell = r_1 + r_2 + s_2 - \ell + (s-r)$ and $m_1 + m_{1,2} + m_{2,2} \geq r_1 + 2r_2 - \lfloor \ell_1/2 \rfloor - x = r_1 + r_2 + s_2 - \lfloor \ell_1/2 \rfloor - x + (r-s)$. We deduce

$$E_{\mathbf{p}} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} \left(O_{\kappa}(1) \frac{\log(n/t)}{n/t} \right)^x \left(\frac{r}{m} \right)^{\max(s-r-\ell, r-s-\lfloor \ell_1/2 \rfloor - x)}$$

for $\kappa \ll \varepsilon$, recalling the definition $O^*(1) := (1 + O(\kappa))^r e^{O_{\kappa}(1+x+f(r))} \leq e^{O(\kappa r)} O_{\kappa}(1) O_{\kappa}(1)^x$ (for the inequality, we are using that $\eta \ll \kappa$, so the “ $f(r)$ ” in the exponent is sufficiently small compared to r).

We finally break into cases to prove the six different parts of [Lemma 11.11](#). Observe that since we are assuming $t \leq \alpha n$ for $\alpha \ll \kappa$, we have that $(O_{\kappa}(1) \log(n/t)/(n/t))^x \leq 1$.

- (1) If $s \geq r+1$ and $\ell = 0$, then taking the first term in the “max” in the exponent yields $\mathbb{E}X_{r,s,0} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (r/m) = e^{-\Omega_{\varepsilon}(r)}/n$.
- (2) If $s = r$ and $\ell = 0$, taking the first term in the “max”, we see $\mathbb{E}X_{r,r,0} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)}$.
- (3) When $s = r$, $\ell = 0$, $x > 0$, and $t \leq n^{1/8}$, notice that we have $O_{\kappa}(1)t \log(n/t)/n \leq n^{-3/4}$. Thus taking the first term in the “max” yields $\mathbb{E}X'_{r,r,0} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} n^{-3/4}$.
- (4) Suppose $s = r-1$ and $\ell = 0$. If $x \geq 2$ then take the first term in the “max”, and if $x < 2$ take the second term. Thus for $x \geq 2$, we have $E_{\mathbf{p}} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (t \log(n/t)/n)^2 \cdot n \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (t/n)^{3/2} n$. For $x = 1$, we have $E_{\mathbf{p}} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (t \log(n/t)/n) \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (t/n)^{3/2} n$ and for $x = 0$, we have $E_{\mathbf{p}} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} \frac{r}{m} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)}/m \lesssim e^{-\Omega_{\varepsilon}(r)} (t/n)^{3/2} n$.
- (5) For any ℓ , by [S1](#) we have $s - r - \ell \geq -1$, so (taking the first term in the “max”) we have $\mathbb{E}X_{r,s,\ell} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} n$.
- (6) If $\ell = 1$ and $t \leq \log(n)$, notice that $t \log(n/t)/n \leq (\log n)^2/n$. If $x > 0$ then take the first term in the “max”, which is at most -1 by [S1](#). If $x = 0$, then $\lfloor \ell_1/2 \rfloor = 0$ means the “max” term must evaluate to at least 0. Together these two cases yield $\mathbb{E}X_{r,s,1} \lesssim_{\kappa} e^{-\Omega_{\varepsilon}(r)} (\log n)^2$. \square

11.2. Deductions. We now deduce [Claims 10.3](#) to [10.5](#).

Proof of Claim 10.5. By [Lemma 11.7](#) and [Lemma 11.1](#), for $t \leq \Delta \leq \log n$ it suffices to prove that G_t has at most $(n-t)/10$ different $(\leq \eta(n-t), 1)$ -stalks and at most $(n-t)^2/10$ different $(\leq \eta(n-t), 2)$ -stalks, with probability $1 - (\log n)^{-\omega(1)}$. By [Lemma 6.6](#), we only need to consider ε -sparse stalks. Also, by [Lemma 11.9](#), with probability $1 - (t/n)^{1/4} \geq 1 - (\log n)^{-\omega(1)}$, the degree sequence of G_t is (n, m, t, κ) -typical. Thus it suffices to prove the result conditional on a particular such degree sequence.

Conditioning on a typical degree sequence, by [Lemma 11.11](#)(6) and (5) respectively, we have the expected numbers of ε -sparse $(\leq \eta(n-t), 1)$ -stalks and $(\leq \eta(n-t), 2)$ -stalks are at most

$$\sum_{r=1}^{\infty} r e^{-\Omega(r)} O_{\kappa}(\log n)^2 \lesssim_{\kappa} (\log n)^2 \quad \text{and} \quad \sum_{r=1}^{\infty} r e^{-\Omega(r)} O_{\kappa}(n) \lesssim_{\kappa} n,$$

respectively. The desired result follows from Markov's inequality. \square

Proof of Claim 10.3. Fix t . We would like to prove that with probability at least $1 - O_{\kappa}(1/t + (n/t)^{1/4})$, we either have $t > N$ (i.e., t is outside our range of consideration), or $\dim(K_t^{(\eta)}) < t/4$.

Notice that (if $t \leq N$) we have $\dim(K_t^{(\eta)}) \leq |\text{supp}(K_t^{(\eta)})|$, which by [Lemma 11.5](#) is at most the number of vertices in $(\leq \eta(n-t), 0)$ -stalks. By [Lemma 6.6](#) we only need to worry about ε -sparse stalks, and by [Lemma 11.9](#), it suffices to prove the result conditioned on a particular (n, m, t, κ) -typical degree sequence for G_t (note that the degree sequence of G_t determines whether $t \leq N$).

Conditioning on a typical degree sequence, by [Lemma 11.11](#)(1,2,4), the expected number of vertices in ε -sparse $(\leq \eta(n-t), 0)$ -stalks is at most

$$\sum_{r=1}^{\infty} O_{\kappa}(r e^{-\Omega(r)} + r e^{-\Omega(r)} (t/n)^{3/2} n) = O_{\kappa}(1 + (t/n)^{3/2} n).$$

By Markov's inequality, the probability this number is greater than $t/4$ is $O_{\kappa}((t/n)^{1/2} + 1/t)$. \square

Proof of Claim 10.4. In this proof we only consider $t \leq \Delta$ (so, for example, "all t " should be read as "all $t \leq \Delta$ "). We prove that each of (a,b,c,d,e) hold with probability at least $1 - 1/(5\Delta)$. Say a "special stalk" is an $(r, r, 0)$ -stalk for some $r \leq \eta(n-t)$.

Let V_t^* be the set of vertices contained in an ε -sparse special stalk. By [Lemma 11.11](#)(2), in the setting of [Lemma 11.11](#) (conditioning on a particular typical degree sequence for G_t), we have

$$\mathbb{E}|V_t^*| \lesssim_{\kappa} \sum_{r=1}^{\infty} r e^{-\Omega(r)} = O_{\kappa}(1).$$

By Markov's inequality and [Lemma 11.9](#), and a union bound over t , with probability at least say $1 - 1/(10\Delta)$ each $|V_t^*| \leq O_{\kappa, \Delta}(1)$. So, by [Lemma 6.6](#), with probability at least say $1 - 1/(9\Delta)$, for each t there are at most $O_{\kappa, \Delta}(1)$ vertices in special stalks. For any special cycle of length $4k \leq 2\varepsilon(n-t)$, there is a $(2k, 2k, 0)$ -stalk containing half its vertices (i.e., a special stalk), so this takes care of (d).

Similarly, by [Lemma 11.11](#)(1,2,3,4) together with Markov's inequality and [Lemma 11.9](#), with probability $1/(10\Delta)$ the only 0-stalks in any G_t are $(r, r, 0)$ -stalks which do not contain any degree-1 vertices, for some $|R| = O_{\kappa, \Delta}(1)$. The union of any two non-disjoint cycles has strictly more edges than vertices (since in such a union every vertex has degree at least 2, and some vertex has degree strictly greater than 2). So, given the above event, if two of the special cycles in G_t were not vertex disjoint, they would provide a set of $O_{\kappa, \Delta}(1) = o_{\kappa, \Delta}(\sqrt{\log n})$ vertices contradicting [Lemma 6.6](#) (which holds with probability $1 - n^{-\omega(1)}$). So, [Lemma 11.6](#) takes care of (a,b).

For (c,e), note that by [Lemma 6.6](#)(A2) and [Lemma 6.2](#)(A4), there are at most say $\exp((\log \log n)^4)$ edges (and thus, vertices) in G which are in a cycle of length at most $\log \log n$ or adjacent to such a cycle. We can reveal these "dangerous" vertices in G without revealing the random ordering v_{N-1}, \dots, v_0 of the vertices in T (recall that the vertices of T are deleted then added back in some random order to form our sequence of graphs G_N, \dots, G_0). With probability $1 - 1/(10\Delta)$, none of the vertices v_{Δ}, \dots, v_0 is dangerous (indeed, the expected number of such dangerous vertices is $\Delta n^{-1+o(1)} \leq 1/(10\Delta^2)$, so Markov's inequality yields this). This handles (e). If (e) holds, the special cycles of length at most $\log \log n$ are completely unaffected by the vertex additions defining the sequence G_{Δ}, \dots, G_0 ; this takes care of (c), recalling that with probability at least $1 - 1/(10\Delta)$ each $|V_t^{\text{spec}}| \leq O_{\kappa, \Delta}(1) \leq \log \log n$. \square

12. THE BIPARTITE CASE

Having just proved [Theorem 1.12\(A1\)](#), we now sketch the changes that must be made for a proof of (B1). The proof strategy is extremely similar, but there are some minor simplifications and complications. The most notable simplification is that we can use [Lemma 8.2](#) instead of the more sophisticated [Lemma 8.4](#), and the primary complication is that a small amount of extra notation and bookkeeping becomes necessary, due to the fact that we need to pay attention to both right and left kernels (i.e., the kernels of our matrix B and its transpose B^\top).

Recall that in the setting of [Theorem 1.12\(A1\)](#), we had a set T of high-degree vertices (coming from [Lemma 9.1\(A\)](#)). We “extracted” this set (and used [Lemma 9.1\(A1\)](#) to control the rank of the resulting matrix), then added back these vertices one-by-one in a random order (each such addition corresponds to the addition of a new row and column), studying how the rank changes during this process.

In the setting of [Theorem 1.12\(B1\)](#), after sampling $G \sim \mathcal{K}(n_1, n_2, m, 2)$ and using the setup of [Lemma 9.1\(B\)](#) we will now have *two* sets T_1, T_2 of high-degree vertices (whose sizes are almost the same). We let $\mathbf{T} \subseteq T_1 \times T_2$ be a set of $\min(|T_1|, |T_2|)$ disjoint *pairs* of vertices from $T_1 \times T_2$, which will play the role of T above. Indeed, let $G[V \setminus \mathbf{T}]$ be the (balanced) bipartite graph obtained from G by removing the vertices in the pairs in \mathbf{T} , so by [Lemma 9.1\(B1, B3\)](#) we have $\text{corank } G[V \setminus \mathbf{T}] \leq |\mathbf{T}|/15$. The plan is then to add back the pairs in \mathbf{T} pair-by-pair in a random order (each such addition corresponds to the addition of a new row for the T_1 -vertex and the addition of a new column for the T_2 -vertex).

Similarly to [Section 10](#), we define $N = |\mathbf{T}|$, and let G_t be the graph that results after t of the pairs in \mathbf{T} have been added back. Instead of just defining the “small-support kernel” $K_t^{(\eta)}$, we now need both a right and left version:

$$K_t^{(\eta)} = \{\mathbf{v} \in \ker B(G_t) : |\text{supp}(\mathbf{v})| \leq \eta(n_2 - t)\}, \quad Q_t^{(\eta)} = \{\mathbf{v} \in \ker B(G_t)^\top : |\text{supp}(\mathbf{v})| \leq \eta(n_1 - t)\}.$$

Then, it is straightforward to prove bipartite analogues to [Claims 10.1](#) to [10.4](#) (we have no need for an analogue of [Claim 10.5](#), because the corank-boosting part is now simpler). Specifically, our analogue of [Claim 10.1](#) should say that after appropriate revelations the neighbourhoods of vertices in T_1 are uniformly random subsets of $W \cap V_2$, and the neighbourhoods of the vertices in T_2 are uniformly random subsets of $W \cap V_1$ (of the appropriate sizes), all independent of each other. Our analogue of [Claim 10.2](#) should hold for both $B(G_t)$ and $B(G_t)^\top$, and our analogues of [Claims 10.3](#) and [10.4](#) should hold for both $K_t^{(\eta)}$ and $Q_t^{(\eta)}$ (for [Claim 10.4](#), $K_t^{(\eta)}$ should be described in terms of 2-special cycles, and $Q_t^{(\eta)}$ should be described in terms of 1-special cycles). There are no additional difficulties in the proofs of any of these claims. Actually, things are slightly simpler: we remark that the bipartite analogue of a *stalk* in [Section 11](#) should be defined to be a set of vertices S contained on just one side of our bipartite graph, so there can be no intersection between S and its neighbourhood $N(S)$; this simplifies the calculations in [Claim 11.14](#).

Now, recall that in the proof of [Theorem 1.12\(A1\)](#) we considered a random walk defined by random variables of the form $\dim \ker A(G_t) - \mathbb{1}_{t \leq \Delta} \dim K_\Delta^{(\eta)} - t/4$. For [Theorem 1.12\(B1\)](#) we need a similar definition that takes both sides of our bipartite graph into account: let

$$X_t = \min\left(\dim \ker B(G_t) - \mathbb{1}_{t \leq \Delta} \dim K_\Delta^{(\eta)}, \dim \ker B(G_t)^\top - \mathbb{1}_{t \leq \Delta} \dim Q_\Delta^{(\eta)}\right) - t/4.$$

Actually, it turns out that only the first term of the “min” is really necessary: recall that we are assuming $n_1 - n_2 \rightarrow \infty$, and note that $n_2 - \dim \ker B(G_t) = \text{rank } B(G_t) = n_1 - \dim \ker B(G_t)^\top$. Also, by (a bipartite analogue of) [Claim 10.3](#), whp $\dim K_\Delta^{(\eta)}$ and $\dim Q_\Delta^{(\eta)}$ are of the form $o(n_1 - n_2)$. So whp we actually have

$$X_t = \dim \ker B(G_t) - \mathbb{1}_{t \leq \Delta} \dim K_\Delta^{(\eta)} - t/4.$$

for all t . The above reasoning also shows that whp for all $t \leq \Delta$, we have $\dim \ker B(G_t)^\top - \dim Q_\Delta^{(\eta)} > 0$, i.e., $B(G_t)^\top$ has a kernel vector $\mathbf{v} \in \mathbb{R}^{n_2}$ with $|\text{supp}(\mathbf{v})| \geq \eta n_1$.

Now, we apply [Theorem 7.2](#) in basically the same way as for the proof of [Theorem 1.12\(A1\)](#). We say an index is *good* if it satisfies the natural analogues of [G1](#), [G2](#), [G3](#), [G4](#), [G5](#), [G7](#) (where [G3](#) needs to hold for both $B(G_t)$ and $B(G_t)^\top$, and [G2](#), [G4](#), [G5](#) and [G7](#) need to hold for both $K_t^{(\eta)}$ and $Q_t^{(\eta)}$), and if the following property holds (c.f., the discussion in the previous paragraph):

G8 If $t \leq \Delta$, then $B(G_t)^\top$ has a kernel vector $\mathbf{v} \in \mathbb{R}^{n_2}$ with $|\text{supp}(\mathbf{v})| \geq \eta n_1$.

Now, the rest of the proof of [Theorem 1.12\(A1\)](#) basically translates directly into a proof of [Theorem 1.12\(A1\)](#), with the exception that we need to replace the applications of [Lemma 8.4](#) with applications of [Lemma 8.2](#). Specifically:

- In “Case 1”: instead of [Lemma 8.4\(a\)](#) we apply [Lemma 8.2](#) twice to increase the rank by 2 (first we add our new column to obtain a matrix B' , then we view our new-row-addition as adding a column to $(B')^\top$).
- In “Case 2”: instead of [Lemma 8.4\(b\)](#) we add our new column and apply [Lemma 8.2](#) (using [G8](#)), then note that adding an additional row cannot decrease the rank.

13. ASYMPTOTIC DISTRIBUTIONS

In this section we prove the central limit theorem in [Corollary 1.3](#), and discuss the (Poisson-type) asymptotic distributions of various quantities in [Theorems 1.2](#) and [1.12](#).

First, we prove [Corollary 1.3](#).

Proof of [Corollary 1.3](#). For $p = c/n$ with $c < 1$ or $c > e$, the matching number $\nu(G)$ of a random graph $G \sim \mathbb{G}(n, c/n)$ is known to satisfy a central limit theorem: there are $\mu = \mu(c, n)$ and $\sigma = \sigma(c, n)$ (where μ and σ^2 both have order of magnitude n) such that $(\nu(G) - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$. For $c < 1$ this is due to Pittel [\[60\]](#), and for $c > e$ this is due to Kreačić [\[49, Theorem 19\]](#).

By [Theorem 1.2\(A1\)](#), we have $2\nu(G) - \text{rank } A(G) = o(\sqrt{n})$ whp, which implies that $X = \text{rank } A(G)$ satisfies the same central limit theorem as $\nu(G)$.

Strictly speaking, it remains to show that we also have $(\text{rank } A(G) - \mathbb{E}X)/\sqrt{\text{Var } X} \rightarrow \mathcal{N}(0, 1)$. Indeed, *a priori*, there may be no connection between μ and $\mathbb{E}X$ or between σ^2 and $\text{Var } X$, if the mean or variance of X is dominated by the effect of outliers. To rule out such pathological behaviour, we need the well-known observation (easily proved with the Azuma–Hoeffding martingale concentration inequality; see for example the appendix of [\[18\]](#)) that the rank of a random matrix is subgaussian with “variance proxy” $O(n)$ (and thus the tails have negligible contribution to the mean and variance). \square

Next we prove [Theorem 1.12\(A2\)](#). We omit the proof of [Theorem 1.12\(B2\)](#), as it follows from an easier version of the same argument.

First, we need expressions for certain infinite sums, which can both be obtained by manipulating the Taylor series $\log(1 - x) = -\sum_{k=1}^{\infty} x^k/k$.

Lemma 13.1. *Let $Z \sim \text{Poisson}(\lambda)$.*

(A) *If $Q \in \mathbb{R}$ satisfies $|Q|\lambda < e^{\lambda/2} - e^{-\lambda/2}$ then*

$$\sum_{k=1}^{\infty} \frac{1}{4k} \cdot \frac{(2Q^2 \Pr[Z = 2|Z \geq 2] \mathbb{E}[Z(Z-1)|Z \geq 2])^{2k}}{\mathbb{E}[Z|Z \geq 2]^{4k}} = -\frac{1}{4} \log \left(1 - \left(\frac{Q\lambda}{e^{\lambda/2} - e^{-\lambda/2}} \right)^4 \right).$$

Also, for all $\lambda > 0$ we have $\lambda < e^{\lambda/2} - e^{-\lambda/2}$ (i.e., the above holds for Q sufficiently close to 1).

(B) *If $Q \in \mathbb{R}$ satisfies $|Q|\lambda < e^\lambda - 1$ then*

$$\sum_{k=1}^{\infty} \frac{1}{8k} \left(\frac{2Q \Pr[Z = 2|Z \geq 2]}{\mathbb{E}[Z|Z \geq 2]} \right)^{4k} = -\frac{1}{8} \log \left(1 - \left(\frac{Q\lambda}{e^\lambda - 1} \right)^4 \right).$$

Also, for all $\lambda > 0$ we have $\lambda < e^\lambda - 1$.

Lemma 13.2. *In the setting of [Theorem 1.12\(A\)](#), for any $M \rightarrow \infty$, whp there are no special cycles of length at least $4M$.*

Proof. First, we need to separately rule out extremely long special cycles. One could perform a configuration model calculation, but it is convenient to borrow from the proof of [Theorem 1.12](#): right at the end of [Section 10](#) (at the end of the proof of [Theorem 1.12\(A1\)](#)): we proved that, for an arbitrarily small constant η , whp there is no special cycle longer than $2\eta n$ (in the notation of that section, $s(G) = s^{(\eta)}(G)$). Taking $\eta \rightarrow 0$ sufficiently slowly, it now suffices to consider special cycles of length $o(n)$.

The remaining long cycles of length $o(n)$ can actually also be handled by borrowing from the proof of [Theorem 1.12](#) (specifically, from [Lemmas 11.9](#) and [11.11\(2\)](#)). However, as a warm-up to more involved calculations that will appear later in the proof of [Theorem 1.12\(A2\)](#), we perform an explicit configuration model calculation.

Condition on a degree sequence satisfying the properties in [Lemma 6.2\(A\)](#). We compute with the configuration model (which suffices, by [Lemma 6.5](#)). Let $n' = \Pr[Z = 2|Z \geq 2]n + o(n)$ be the number

of degree-2 vertices (using [Lemma 6.2\(A1\)](#)). Using [Lemma 6.2\(A2\)](#), the number of configurations of $4k$ pairs that correspond to a special cycle of length $4k$ is

$$\begin{aligned} \frac{(n')_{2k} 2^{2k}}{4k} \sum_{\substack{v_1, \dots, v_{2k} \\ \text{distinct}}} \prod_{i=1}^{2k} d_{v_i} (d_{v_i} - 1) &\leq \frac{1}{4k} \left(4n' \sum_v \binom{d_v}{2} \right)^{2k} \\ &\leq \frac{1}{4k} \left((2 + o(1))n \Pr[Z = 2 | Z \geq 2] \mathbb{E}[Z(Z-1) | Z \geq 2] \right)^{2k} \end{aligned}$$

(this counts isolated cycles twice). For $k = o(n)$, the probability such a configuration actually appears is

$$\frac{(2m - 8k - 1)!!}{(2m - 1)!!} = \frac{(m)_{4k} 2^{4k}}{(2m)_{8k}} = \left(\frac{1 + o(1)}{\mathbb{E}[Z | Z \geq 2]} \right)^{4k}.$$

We compute that the expected number of special cycles with length at least $4M$ (and at most $o(n)$) is at most

$$\sum_{k=M}^{o(n)} \frac{((2 + o(1)) \Pr[Z = 2 | Z \geq 2] \mathbb{E}[Z(Z-1) | Z \geq 2])^{2k}}{4k (\mathbb{E}[Z | Z \geq 2])^{4k}}.$$

By [Lemma 13.1](#), this tends to zero as $M \rightarrow \infty$, so the desired result follows from Markov's inequality. \square

We are now ready to prove [Theorem 1.12\(A2\)](#).

Proof of [Theorem 1.12\(A2\)](#). Recall that an induced cycle is special if its length is divisible by 4 and every second vertex has degree 2. Say that the cycle is *weakly special* if it is not isolated (i.e., if it has at least one vertex whose degree is not 2).

We work in the configuration model, conditioning on a degree sequence satisfying the properties in [Lemma 6.2](#). Again, let $n' = \Pr[Z = 2 | Z \geq 2]n + o(n)$ be the number of degree-2 vertices. Let N_k be the number of weakly special cycles of length $4k$, and let N_k^\dagger be the number of isolated cycles of length $4k$. Let A_1 and A_2 be the numbers of loops and 2-cycles (so the random multigraph produced by the configuration model is simple if and only if $A_1 = A_2 = 0$). Also, let

$$\begin{aligned} \gamma_k^\dagger &= \frac{1}{8k} \left(\frac{2 \Pr[Z = 2 | Z \geq 2]}{\mathbb{E}[Z | Z \geq 2]} \right)^{4k}, \quad \gamma_k = \frac{1}{4k} \cdot \frac{(2 \Pr[Z = 2 | Z \geq 2] \mathbb{E}[Z(Z-1) | Z \geq 2])^{2k}}{\mathbb{E}[Z | Z \geq 2]^{4k}} - 2\gamma_k^\dagger, \\ \eta_1 &= \frac{1}{2} \cdot \frac{\mathbb{E}[Z(Z-1) | Z \geq 2]}{\mathbb{E}[Z | Z \geq 2]}, \quad \eta_2 = \frac{1}{4} \left(\frac{\mathbb{E}[Z(Z-1) | Z \geq 2]}{\mathbb{E}[Z | Z \geq 2]} \right)^2. \end{aligned}$$

Let (W_1, W_2) , $(Y_k)_{k=1}^\infty$ and $(Y_k^\dagger)_{k=1}^\infty$ be independent sequences of independent Poisson random variables with $\mathbb{E}W_i = \eta_i$, $\mathbb{E}Y_k = \gamma_k$, and $\mathbb{E}Y_k^\dagger = \gamma_k^\dagger$.

Abusing notation, we write \cup for concatenation of sequences. We claim that for any M , we have

$$(A_1, A_2) \cup (N_k)_{k=1}^M \cup (N_k^\dagger)_{k=1}^M \xrightarrow{d} (W_1, W_2) \cup (Y_k)_{k=1}^M \cup (Y_k^\dagger)_{k=1}^M.$$

(where here n goes to infinity while M is fixed). In combination with [Lemma 13.2](#) and [Lemma 6.5](#), this suffices to prove [Theorem 1.12\(A2\)](#). Indeed, [Lemma 13.1](#) then shows that the number of weakly special cycles converges in distribution to a Poisson random variable with parameter $\gamma(c)$, and the number of isolated cycles converges to a Poisson random variable with parameter $\gamma^\dagger(c)$.

To this end, by the method of moments (see for example [\[69, Lemma 2.8\]](#)) it suffices to prove that for any $(s_1, s_2) \cup (r_k)_{k=1}^M \cup (r_k^\dagger)_{k=1}^M$ (which we treat as fixed, while $n \rightarrow \infty$), we have

$$\mathbb{E} \left[(A_1)_{s_1} (A_2)_{s_2} \prod_{k=1}^M (N_k)_{r_k} (N_k^\dagger)_{r_k^\dagger} \right] \rightarrow \eta_1^{s_1} \eta_2^{s_2} \prod_{k=1}^M \gamma_k^{r_k} (\gamma_k^\dagger)^{r_k^\dagger}.$$

Note that $(A_1)_{s_1} (A_2)_{s_2} \prod_{k=1}^M (N_k)_{r_k} (N_k^\dagger)_{r_k^\dagger}$ is the number of (ordered) collections of distinct cycles, containing s_1 loops, s_2 2-cycles, r_k weakly special cycles of each length $4k$ and r_k^\dagger isolated cycles of each length $4k$.

If such a collection does not consist of vertex-disjoint cycles, then the union of this collection has strictly more edges than vertices (because every vertex has degree at least 2, and some vertex has degree strictly greater than 2). The expected number of such collections is therefore $O(1/n)$. So, it suffices to consider the contribution from collections of vertex-disjoint cycles.

In such a collection of disjoint cycles, let $x = s_1 + 2s_2$ be the number of vertices that should be in loops and 2-cycles (which can have any degree), let $y = \sum_{k=1}^M (2kr_k + 4kr_k^\dagger)$ be the number of vertices

that should have degree 2 in special cycles, and let $t = x + y + \sum_{k=1}^M 2kr_k$ be the total number of vertices (recall that in weakly special cycles, not all vertices have degree 2, so there is a mild restriction on the degrees of the vertices not counted by x and y). The number of ways to choose configurations forming an appropriate collection of disjoint cycles is

$$\frac{(n')_y 2^y}{2^{s_1 4^{s_2}} \prod_{k=1}^M (4k)^{r_k} (8k)^{r_k^\dagger}} \left(\sum_{v_1, \dots, v_{t-y}}^* \prod_{i=1}^{t-y} d_{v_i} (d_{v_i} - 1) \right)$$

where we fix a set V^* of y degree-2 vertices, and the sum with a “*” is over all tuples of distinct vertices $v_1, \dots, v_{t-y} \notin V^*$, satisfying the following condition. After the first x vertices, if we group the vertices into consecutive blocks of lengths $2, \dots, 2, 4, \dots, 4, 6, \dots, 6, \dots, 2M, \dots, 2M$ (where there are r_k blocks of each length $2k$), then each block has at least one non-degree-2 vertex (and therefore its corresponding special cycle is only weakly special). Using [Lemma 6.2\(A4\)](#) with $s = 1$, we see that the distinctness restriction on the v_1, \dots, v_{t-y} makes essentially no difference (and similarly with the condition that they do not lie in V^*). Thus, observe that

$$\sum_{v_1, \dots, v_{t-y}}^* \prod_{i=1}^{t-y} d_{v_i} (d_{v_i} - 1) = \left(\sum_{v \in V(G)} d_v (d_v - 1) \right)^x \prod_{k=1}^M \left(\left(\sum_{v \in V(G)} d_v (d_v - 1) \right)^{2k} - (n')^{2k} 2^{2k} \right)^{r_k} + o(n^{t-y}).$$

We may estimate the above sum using the property in [Lemma 6.2\(A2\)](#). Now, for any of our configurations of disjoint cycles, the probability such a configuration actually appears is

$$\frac{(2m - t)!!}{(2m)!!} = \frac{1 + o(1)}{(n \mathbb{E}[Z | Z \geq 2])^t}.$$

The desired result follows, using our expressions for n' and $\sum_{v \in V(G)} d_v (d_v - 1)$ and simplifying. \square

13.1. Further comments on asymptotic distributions. We finish this section with some discussion of the asymptotic distributions of $\nu(G) - \text{rank } B(G)$ in the setting of [Theorem 1.2\(B\)](#), and $2\nu(G) - \text{rank } A(G)$, $\sigma(G) - \text{rank } A(G)$ in the setting of [Theorem 1.2\(A\)](#).

First, we believe that in the setting of [Theorem 1.2\(B\)](#), whp $\nu(G)$ attains the Karp–Sipser bound in [Corollary 5.2\(B\)](#) exactly (meaning that $\nu(G) = n - \max(i_1(G), i_2(G))$). To prove this in the same way as [Theorem 5.3](#), one needs to prove that in the setting of [Theorem 1.12\(B\)](#), we have $\nu(G) = n_2$ whp (i.e., there is a matching saturating the smaller side of the bipartite graph). A very similar statement was proved in a difficult paper of Frieze [\[35\]](#), and we believe that the ideas in his paper are also applicable to our setting. If this were true, then the asymptotic distribution of $\nu(G) - \text{rank } B(G)$ would be precisely as described in [Theorem 1.9\(B\)](#).

Similarly, we believe that in the setting of [Theorem 1.2\(A\)](#), whp $\sigma(G)$ attains the Karp–Sipser bound in [Corollary 5.2\(A2\)](#) exactly (meaning that $\sigma(G) = n - i(G)$). To prove this in the same way as [Theorem 5.3](#), one needs to prove that in the setting of [Theorem 1.12\(A\)](#), we have $\sigma(G) = n$ (i.e., there is a collection of vertex-disjoint cycles and edges covering the entire graph). We believe that such a collection can almost entirely consist of edges (i.e., it is essentially a matching), but odd cycles must be included if G has isolated odd cycles, and an additional odd cycle may be necessary for parity reasons. We believe that the ideas by Frieze and Pittel [\[37\]](#) on matchings in degree-constrained random graphs should be suitable to prove this (in fact, it may be possible to deduce the desired statement from the main result of [\[37\]](#) in a black-box manner). If this were true, then the asymptotic distribution of $\sigma(G) - \text{rank } A(G)$ would be precisely as described in [Theorem 1.9\(A\)](#).

In the setting of [Theorem 1.2\(A\)](#), we already have a characterisation of $\nu(G)$ from [Theorem 5.3](#). A routine calculation in the configuration model (similar to the one used to prove [Theorem 1.12\(A2\)](#) earlier in this section) shows that the number of isolated odd cycles in $\text{core}_{\text{KS}}(G)$ is asymptotically independent from $s(G)$. If $c > e$, it is not hard to see that the parity of the size of the giant component of $\text{core}_{\text{KS}}(G)$ is asymptotically independent from these two quantities. We believe that the ideas in [\[22, Section 7\]](#) can be used to prove that this parity is asymptotically equidistributed, in which case the asymptotic distribution of $2\nu(G) - \text{rank } A(G)$ would be $Y + 2Y^\dagger - W - \mathbb{1}_{c > e} U$, where Y, Y^\dagger are as in [Theorem 1.9\(A\)](#), and, independently, U is uniform on $\{0, 1\}$ and W is Poisson with mean

$$\sum_{k=1}^{\infty} \frac{1}{2(2k-1)} \cdot \left(\frac{2 \Pr[Z = 2 | Z \geq 2]}{\mathbb{E}[Z | Z \geq 2]} \right)^{2k-1} = \frac{1}{4} \left(\log \left(1 + \frac{\lambda}{e^\lambda - 1} \right) - \log \left(1 - \frac{\lambda}{e^\lambda - 1} \right) \right).$$

REFERENCES

- [1] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *Critical random graphs: limiting constructions and distributional properties*, Electron. J. Probab. **15** (2010), no. 25, 741–775. [↑3](#)
- [2] Louigi Addario-Berry and Laura Eslava, *Hitting time theorems for random matrices*, Combin. Probab. Comput. **23** (2014), 635–669. [↑1](#)
- [3] R. C. Alamino and D. Saad, *Typical kernel size and number of sparse random matrices over Galois fields: a statistical physics approach*, Phys. Rev. E (3) **77** (2008), 061123, 12. [↑5](#)
- [4] David Aldous and J. Michael Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. [↑10](#), [↑25](#)
- [5] Jonathan Aronson, Alan Frieze, and Boris G. Pittel, *Maximum matchings in sparse random graphs: Karp-Sipser revisited*, Random Structures Algorithms **12** (1998), 111–177. [↑4](#), [↑9](#), [↑13](#), [↑14](#), [↑17](#), [↑45](#)
- [6] Mark Bagnoli and Ted Bergstrom, *Log-concave probability and its applications*, Economic theory **26** (2005), 445–469. [↑26](#)
- [7] Anirban Basak and Mark Rudelson, *Sharp transition of the invertibility of the adjacency matrices of sparse random graphs*, arXiv:1809.08454. [↑1](#)
- [8] M. Bauer and O. Golinelli, *Core percolation in random graphs: a critical phenomena analysis*, The European Physical Journal B **24** (2001), 339–352. [↑4](#), [↑5](#), [↑7](#)
- [9] M. Bauer and O. Golinelli, *Exactly solvable model with two conductor-insulator transitions driven by impurities*, Physical Review Letters **86** (2001), 2621–2624. [↑2](#), [↑4](#)
- [10] A. Békéssy, P. Békéssy, and J. Komlós, *Asymptotic enumeration of regular matrices*, Studia Sci. Math. Hungar. **7** (1972), 343–353. [↑19](#), [↑20](#)
- [11] Edward A. Bender and E. Rodney Canfield, *The asymptotic number of labeled graphs with given degree sequences*, J. Combinatorial Theory Ser. A **24** (1978), 296–307. [↑19](#), [↑20](#)
- [12] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. [↑10](#), [↑25](#)
- [13] Jose Blanchet and Alexandre Stauffer, *Characterizing optimal sampling of binary contingency tables via the configuration model*, Random Structures Algorithms **42** (2013), 159–184. [↑20](#)
- [14] B. Bollobás, C. Cooper, T. I. Fenner, and A. M. Frieze, *Edge disjoint Hamilton cycles in sparse random graphs of minimum degree at least k* , J. Graph Theory **34** (2000), 42–59. [↑18](#)
- [15] Béla Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), 311–316. [↑19](#), [↑20](#)
- [16] Béla Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57. [↑2](#)
- [17] Béla Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. [↑3](#), [↑20](#)
- [18] Charles Bordenave, Marc Lelarge, and Justin Salez, *The rank of diluted random graphs*, Ann. Probab. **39** (2011), 1097–1121. [↑2](#), [↑4](#), [↑8](#), [↑10](#), [↑11](#), [↑25](#), [↑40](#)
- [19] Charles Bordenave, Marc Lelarge, and Justin Salez, *Matchings on infinite graphs*, Probab. Theory Related Fields **157** (2013), 183–208. [↑4](#)
- [20] Thomas Budzinski, Alice Contat, and Nicolas Curien, *The critical Karp-Sipser core of random graphs*, arXiv:2212.02463. [↑7](#)
- [21] Julie Cain and Nicholas Wormald, *Encores on cores*, Electron. J. Combin. **13** (2006), Research Paper 81, 13. [↑17](#)
- [22] Amin Coja-Oghlan, Oliver Cooley, Mihyun Kang, Joon Lee, and Jean Bernoulli Ravelomanana, *The sparse parity matrix*, arXiv:2107.06123. [↑4](#), [↑6](#), [↑13](#), [↑32](#), [↑42](#)
- [23] Amin Coja-Oghlan, Oliver Cooley, Mihyun Kang, and Kathrin Skubch, *Core forging and local limit theorems for the k -core of random graphs*, J. Combin. Theory Ser. B **137** (2019), 178–231. [↑8](#)
- [24] Amin Coja-Oghlan, Alperen A. Ergür, Pu Gao, Samuel Hetterich, and Maurice Rolvien, *The rank of sparse random matrices*, Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2020, pp. 579–591. [↑3](#), [↑5](#), [↑8](#)
- [25] Simon Coste and Justin Salez, *Emergence of extended states at zero in the spectrum of sparse random graphs*, Ann. Probab. **49** (2021), 2012–2030. [↑4](#)
- [26] Kevin P. Costello, Terence Tao, and Van Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), 395–413. [↑1](#), [↑10](#), [↑11](#), [↑12](#), [↑21](#)
- [27] Kevin P. Costello and Van Vu, *On the rank of random sparse matrices*, Combin. Probab. Comput. **19** (2010), 321–342. [↑2](#), [↑5](#)
- [28] Kevin P. Costello and Van H. Vu, *The rank of random graphs*, Random Structures Algorithms **33** (2008), 269–285. [↑1](#), [↑2](#)
- [29] Amir Dembo and Andrea Montanari, *Finite size scaling for the core of large random hypergraphs*, Ann. Appl. Probab. **18** (2008), 1993–2040. [↑5](#)
- [30] Patrick DeMichele, Margalit Glasgow, and Alexander Moreira, *On the rank, kernel, and core of sparse random graphs*, arXiv:2105.11718. [↑2](#), [↑3](#), [↑5](#), [↑11](#), [↑12](#), [↑23](#), [↑29](#)
- [31] Jian Ding, Eyal Lubetzky, and Yuval Peres, *Anatomy of the giant component: the strictly supercritical regime*, European J. Combin. **35** (2014), 155–168. [↑3](#), [↑13](#)
- [32] Asaf Ferber, Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Random symmetric matrices: rank distribution and irreducibility of the characteristic polynomial*, arXiv:2106.04049. [↑8](#)
- [33] Asaf Ferber, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *Singularity of the k -core of a random graph*, arXiv:2106.05719. [↑3](#), [↑11](#), [↑12](#), [↑13](#), [↑21](#), [↑23](#), [↑25](#), [↑26](#)

- [34] Asaf Ferber, Matthew Kwan, and Lisa Saueremann, *Singularity of sparse random matrices: simple proofs*, arXiv:2011.01291. [↑1](#), [↑23](#)
- [35] Alan Frieze, *Perfect matchings in random bipartite graphs with minimal degree at least 2*, Random Structures Algorithms **26** (2005), 319–358. [↑42](#)
- [36] Alan Frieze and Michał Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016. [↑3](#), [↑13](#)
- [37] Alan Frieze and Boris Pittel, *Perfect matchings in random graphs with prescribed minimal degree*, Mathematics and computer science. III, Trends Math., Birkhäuser, Basel, 2004, pp. 95–132. [↑9](#), [↑16](#), [↑42](#)
- [38] Margalit Glasgow, Christina Goldschmidt, Eleonora Kreačić, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *A central limit theorem for the matching number of a sparse random graph*, in preparation. [↑4](#), [↑8](#)
- [39] Catherine Greenhill, Brendan D. McKay, and Xiaoji Wang, *Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums*, J. Combin. Theory Ser. A **113** (2006), 291–324. [↑20](#)
- [40] Alice Guionnet, *Bernoulli random matrices*, Proceedings of the 8th European Congress in Mathematics, to appear, arXiv:2112.05506. [↑7](#)
- [41] Jiaoyang Huang, *Invertibility of adjacency matrices for random d -regular graphs*, arXiv:1807.06465. [↑8](#)
- [42] Svante Janson, *The probability that a random multigraph is simple*, Combin. Probab. Comput. **18** (2009), 205–225. [↑20](#)
- [43] Svante Janson, *The probability that a random multigraph is simple. II*, J. Appl. Probab. **51A** (2014), 123–137. [↑20](#)
- [44] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. [↑3](#), [↑12](#), [↑13](#)
- [45] Daniel M. Kane, *The correct exponent for the Gotsman-Linial conjecture*, Comput. Complexity **23** (2014), 151–175. [↑23](#)
- [46] R. M. Karp and M. Sipser, *Maximum matching in sparse random graphs*, 22nd Annual Symposium on Foundations of Computer Science (sfcs 1981), IEEE, October 1981. [↑4](#), [↑13](#), [↑14](#)
- [47] J. Komlós, *On the determinant of $(0, 1)$ matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21. [↑1](#), [↑10](#), [↑11](#)
- [48] J. Komlós, *On the determinant of random matrices*, Studia Sci. Math. Hungar. **3** (1968), 387–399. [↑1](#), [↑10](#), [↑11](#)
- [49] Eleonora Kreačić, *Some problems related to the Karp-Sipser algorithm on random graphs*, Ph.D. thesis, University of Oxford, 2017. [↑4](#), [↑9](#), [↑13](#), [↑40](#)
- [50] Marc Lelarge, *Bypassing correlation decay for matchings with an application to XORSAT*, 2013 IEEE Information Theory Workshop (ITW), IEEE, sep 2013. [↑3](#)
- [51] JHC Lisman and MCA Van Zuylen, *Note on the generation of most probable frequency distributions*, Statistica Neerlandica **26** (1972), 19–23. [↑33](#)
- [52] Alexander E. Litvak, Anna Lytova, Konstantin Tikhomirov, Nicole Tomczak-Jaegermann, and Pierre Youssef, *Adjacency matrices of random digraphs: singularity and anti-concentration*, J. Math. Anal. Appl. **445** (2017), 1447–1491. [↑23](#)
- [53] Kenneth Maples, *Symmetric random matrices over finite fields announcement*, unpublished note, <http://user.math.uzh.ch/maples/maples.symma.pdf>, 2009. [↑8](#)
- [54] Brendan D. McKay, *Asymptotics for symmetric 0-1 matrices with prescribed row sums*, Ars Combin. **19** (1985), 15–25. [↑20](#)
- [55] Brendan D. McKay and Nicholas C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$* , Combinatorica **11** (1991), 369–382. [↑20](#)
- [56] András Mészáros, *The distribution of sandpile groups of random regular graphs*, Trans. Amer. Math. Soc. **373** (2020), 6529–6594. [↑8](#)
- [57] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, *Two solutions to diluted p -spin models and XORSAT problems*, J. Statist. Phys. **111** (2003), 505–533. [↑4](#)
- [58] Michael Mitzenmacher and Eli Upfal, *Probability and computing*, second ed., Cambridge University Press, Cambridge, 2017. [↑18](#)
- [59] Bogdan Nica, *A brief introduction to spectral graph theory*, EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich, 2018. [↑15](#)
- [60] Boris Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures Algorithms **1** (1990), 311–342. [↑4](#), [↑9](#), [↑40](#)
- [61] T. Tao and V. H. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010. [↑23](#)
- [62] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), 593–634. [↑1](#)
- [63] Remco van der Hofstad, Noela Müller, and Haodong Zhu, *The rank of sparse symmetric matrices over arbitrary fields*, arXiv:2301.12978. [↑8](#)
- [64] Van Vu, *Recent progress in combinatorial random matrix theory*, arXiv:2005.02797. [↑7](#)
- [65] Van Vu, *Random discrete matrices*, Horizons of combinatorics, Bolyai Soc. Math. Stud., vol. 17, Springer, Berlin, 2008, pp. 257–280. [↑7](#)
- [66] Van Vu, *Some recent results on random matrices*, lecture at a workshop on Probabilistic Techniques and Applications, hosted by the Institute for Pure & Applied Mathematics (IPAM), <http://www.ipam.ucla.edu/abstract/?tid=8303&pcode=CMAWS1>, 2009. [↑2](#)
- [67] Van Vu, *Combinatorial problems in random matrix theory*, invited lecture at the International Congress of Mathematicians (ICM) in Seoul, <https://www.youtube.com/watch?v=9REUYqf8EWA>, 2014. [↑2](#)
- [68] N C Wormald, *Some problems in the enumeration of labelled graphs*, 1978, PhD thesis, University of Newcastle. [↑19](#)
- [69] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298. [↑41](#)

APPENDIX A. ANALYSING THE SUBCRITICAL KARP–SIPSER PROCESS

Here we briefly sketch the analysis in [5] used to prove Lemma 4.4(A): if $c < e$, then the Karp–Sipser core whp consists of a collection of vertex-disjoint cycles, and the numbers of cycles of each length are asymptotically jointly Poisson distributed. As will become clear, one can prove Lemma 4.4(B) (i.e., the bipartite case of the same fact) with essentially the same analysis.

The main part of the proof is an analysis of the Karp–Sipser leaf-removal process. This analysis is slightly simpler on a random multigraph than a random graph: instead of $\mathbb{G}(n, c/n)$, we consider a random multigraph whose edges correspond to a sequence of exactly $\lfloor cn/2 \rfloor$ pairs of vertices, sampled uniformly at random with replacement (results about such random multigraphs can be transferred to random graphs, as observed in [5, Lemma 1]). At each step of the leaf-removal process, we consider the number v_0 of isolated vertices, the number v_1 of degree-1 vertices, the number v of vertices of degree at least 2, and the number m of edges remaining. These statistics (v_0, v_1, v, m) can be shown to evolve as a Markov chain ([5, Lemma 3]).

The authors study the typical trajectory of these statistics as the process evolves, using the *differential equations method*. Namely, they first study the expected change in each of v_0, v_1, v, m after a single step of the leaf-removal process, in terms of the statistics v_0, v_1, v, m themselves ([5, Lemmas 6 and 7]). These expected one-step changes approximately correspond to a system of differential equations (solved in [5, Lemma 8]), and it can then be shown that whp the trajectories of the evolving statistics v_0, v_1, v, m are well-approximated by the solution to this system of differential equations ([5, Lemma 11]).

Specifically, to study the expected change after a single step of leaf-removal, the authors use the fact that at any time t , the distribution of the remaining multigraph is uniform among all multigraphs with statistics (v_0, v_1, v, m) ([5, Lemma 2]). Apart from the $v_0 + v_1$ vertices of degree 0 and 1, the degrees of the remaining vertices are then shown to be well-approximated by a sequence of truncated Poisson random variables (with a particular Poisson parameter z defined in terms of v_0, v_1, v, m ; see [5, Lemmas 4 and 5]), and this degree information can be used to estimate the expected 1-step changes in the various statistics (in the leaf-removal process, if we delete a leaf x with neighbour y , then the change to v_0, v_1, v_2, m can be described in terms of the degrees of the neighbours of y).

As the process continues, the Poisson parameter z evolves with v_0, v_1, v, m . Differential equation heuristics suggest that if $c < e$ then z converges to zero as the process reaches completion. Actually, it is convenient to parameterise the process by z : [5, Lemma 11] allows one to control the trajectories of v_0, v_1, v, m (showing that they are well-approximated by differential equation heuristics) until say $z < n^{-0.1}$. At this point, almost all of the v vertices with degree at least 2 in fact have degree exactly 2. The number of degree-1 vertices v_1 is about $n(1 - \eta)z^2/c$ and the number of degree-2 vertices is about $n\eta z^2/(2c)$, where η is the solution to the equation $c = \eta e^\eta$ (see [5, Eqs. (79)–(90)]).

Although we are still some way from the end of the process (there are still a lot of degree-1 vertices remaining), the key observation is that it is already possible to see what the final Karp–Sipser core will end up looking like. Indeed, since there are so few vertices with degree 3 or greater, by a configuration-model calculation, it is easy to see that whp there are no “heavy cycles” containing a degree-3 vertex (see [5, Eqs. (91)–(93)] and the following discussion). So, at this stage, the connected components are trees and isolated cycles, meaning that the Karp–Sipser core will consist precisely of the (disjoint) cycles which still exist at this stage. In the rest of [5, Section 5.1], the authors then use the method of moments (in a similar way to the calculations in the proof of Theorem 1.12(A2) in Section 13) to show that the number of cycles of length k is asymptotically Poisson with mean $\eta^k/(2k)$ (independently for each k). Roughly speaking, given a typical outcome of the degree sequence (with $N \approx z^2 n/c$ stubs in the corresponding configuration model), the number of possible sets of k configuration-edges corresponding to a k -cycle is about $(\eta N)^k/(2k)$, and the probability a given k -cycle is present is about $1/N^k$.

In the bipartite case, we can perform essentially exactly the same differential-equations-method calculation to track the evolution of the number of edges m , the numbers of isolated vertices $v_0^{(1)}, v_0^{(2)}$ on each side, the numbers of degree-1 vertices $v_1^{(1)}, v_1^{(2)}$ on each side, and the numbers $v^{(1)}, v^{(2)}$ of vertices of degree at least 2 on each side. Actually, the differential equations are exactly the same (where $v_0^{(1)}, v_0^{(2)}$ both take essentially the same value v_0 , and $v_1^{(1)}, v_1^{(2)}$ take essentially the same value v_1 , and $v^{(1)}, v^{(2)}$ take essentially the same value v), because $\mathbb{G}(n, c/n)$ and $\mathbb{G}(n, n, c/n)$ locally “look the same” (locally, they both look like a $\text{Poisson}(c)$ Galton–Watson tree). Near the end of the process, the number of degree-1

vertices on each side is about $n(1 - \eta)z^2/c$ and the number of degree-2 vertices on each side is about $n\eta z^2/(2c)$. Given such a degree sequence, for even k the number of possible k -cycles is about $(\eta N)^k/k$, where $N \approx z^2 n/c$ is the approximate number of stubs on each side in the corresponding configuration model, and the probability a given k -cycle is present is about $1/N^k$. So, the expected number of k -cycles is η^k/k .

DEPARTMENT OF COMPUTER SCIENCE, STANFORD UNIVERSITY, STANFORD, CA.
Email address: `mglasgow@stanford.edu`

INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA (ISTA).
Email address: `matthew.kwan@ist.ac.at`

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA
Email address: `{asah,msawhney}@mit.edu`