A CENTRAL LIMIT THEOREM FOR THE MATCHING NUMBER OF A SPARSE RANDOM GRAPH

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

Abstract. In 1981, Karp and Sipser proved a law of large numbers for the matching number of a sparse Erdős–Rényi random graph, in an influential paper pioneering the so-called *differential equation method* for analysis of random graph processes. Strengthening this classical result, and answering a question of Aronson, Frieze and Pittel, we prove a central limit theorem in the same setting: the fluctuations in the matching number of a sparse random graph are asymptotically Gaussian.

Our new contribution is to prove this central limit theorem in the *subcritical* and *critical* regimes, according to a celebrated algorithmic phase transition first observed by Karp and Sipser. Indeed, in the *supercritical* regime, a central limit theorem has recently been proved in the PhD thesis of Kreačić, using a stochastic generalisation of the differential equation method (comparing the so-called *Karp–Sipser process* to a system of stochastic differential equations). Our proof builds on these methods, and introduces new techniques to handle certain degeneracies present in the subcritical and critical cases. Curiously, our new techniques lead to a *non-constructive* result: we are able to characterise the fluctuations of the matching number around its mean, despite these fluctuations being much smaller than the error terms in our best estimates of the mean.

We also prove a central limit theorem for the rank of the adjacency matrix of a sparse random graph.

1. Introduction

One of the foundational theorems in random graph theory, proved by Erdős and Rényi [13] in 1966, characterises the asymptotic probability that a random graph contains a *perfect matching* (i.e., that we can pair up all the vertices of the graph using disjoint edges). In particular, this property has a *sharp threshold*: for any positive constant $\varepsilon > 0$, and a large even integer n, random graphs with n vertices and more than $((1+\varepsilon)\log n)\cdot n/2$ edges are very likely to contain perfect matchings, while random graphs with n vertices and fewer than $((1-\varepsilon)\log n)\cdot n/2$ edges are very likely *not* to contain perfect matchings.

Below the perfect matching threshold, it is typically not possible to pair up *all* the vertices, but it is still natural to ask what fraction of vertices can be paired up. In 1981, Karp and Sipser [22] provided an asymptotic answer to this question, as follows.

Theorem 1.1. *Fix a constant $c > 0$, consider a set of n vertices, and let G be a random graph defined in one of the following two ways*[1]*:*

- *G contains each of the $\binom{n}{2}$ possible edges with probability c/n independently, or*
- *G contains a uniformly random subset of exactly $\lfloor cn/2 \rfloor$ of the possible edges.*

Let $\nu(G)$ be the matching number of G (i.e., the maximum size of a set of disjoint edges in G). Then, for some constant $\alpha_c \in [0,1]$, we have the convergence in probability

$$\frac{\nu(G)}{n/2} \xrightarrow{p} \alpha_c$$

as $n \to \infty$. Specifically, $\alpha_c = \min_{x\in[0,1]} \exp(-c\exp(-c(1-x)))$.

In their proof of Theorem 1.1, Karp and Sipser introduced a number of highly influential ideas. First, they introduced a random graph process (now often called the *Karp–Sipser leaf-removal process*) designed to construct a near-maximum matching in a random graph. This process has since found a number of important applications outside the context it was originally introduced (e.g., in statistical physics, theoretical computer science and random matrix theory [2, 3, 6, 9, 17, 26]). In order to analyse the behaviour of this process, Karp and Sipser identified certain statistics (evolving with the process),

Kwan was supported by ERC Starting Grant "RANDSTRUCT" No. 101076777. Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-2141064. Sah was supported by the PD Soros Fellowship.

[1]The name "Erdős–Rényi random graph" is used to refer to both these notions of a random graph: either we independently include each edge with a given probability, or we choose a uniformly random graph with a given number of edges. These two models are closely related, and the same types of techniques can be used to study both. Here we have chosen the parameters in such a way that, in both models, the average degree is likely to be about c.

such that the random trajectories of these statistics concentrate around a deterministic "limit trajectory", described as the solution to a certain system of differential equations. This is arguably the first application of the so-called *differential equation method* for random graph processes (see the surveys in [12,29]), which has had an enormous impact in combinatorics and theoretical computer science.

Note that Theorem 1.1 can be interpreted as a *law of large numbers*: with high probability, the matching number $\nu(G)$ is close to its expected value $\mathbb{E}\nu(G) = \alpha_c(n/2) + o(n)$. From this point of view, it is natural to wonder whether there is a corresponding *central limit theorem* in the same setting: are the fluctuations of $\nu(G)$ around its mean $\mathbb{E}\nu(G)$ asymptotically Gaussian? This question seems to have been first explicitly asked in a 1998 paper of Aronson, Frieze and Pittel [1]. As our main result, we answer this question, proving a central limit theorem for the matching number.

Theorem 1.2. *Define G and $\nu(G)$ as in Theorem 1.1. Then we have the convergence in distribution*

$$\frac{\nu(G) - \mathbb{E}\nu(G)}{\sqrt{\operatorname{Var}\nu(G)}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to \infty$. Moreover, the asymptotics of $\operatorname{Var}\nu(G)$ can be explicitly described in terms of an integral involving the solution to a certain system of differential equations; see Remark 5.7.

We remark that in the case $c < 1$, the statement of Theorem 1.2 follows from powerful general results of Pittel [27] (when $c < 1$, random graphs have a very simple structure with no large connected components, and this structure can be very precisely characterised). However, we will not need this in our proof.

More significantly, in the case $c > e$, the statement of Theorem 1.2 was recently proved in the PhD thesis of Kreačić [23] via a beautiful *stochastic* generalisation of the differential equation method. This work considers the same statistics of the Karp–Sipser leaf-removal process that were considered in Karp and Sipser's seminal paper, but instead of simply proving that the trajectory of these statistics concentrates around a deterministic limit, it proves that this trajectory converges in distribution to a certain Gaussian process, obtained as the solution to a system of *stochastic* differential equations. This is accomplished via a general limit theorem for Markov chains due to Ethier and Kurtz [14]. (See also the related techniques of Janson [20] to prove central limit theorems for the so-called *k-core problem*).

Unfortunately, this type of analysis breaks down when $c \le e$, due to the notorious *phase transition* of the Karp–Sipser process: when $c > e$, the process remains "macroscopic" until its termination, whereas when $c \le e$ the process becomes more and more degenerate as it reaches completion (and one loses control over all relevant statistics). For the purposes of a law of large numbers (Theorem 1.1) this short period of degenerate behaviour can be ignored (its contribution is trivially $o(n)$), but for the purposes of a central limit theorem (Theorem 1.2) it is not clear how to rule out dangerously large fluctuations during this short degenerate period[2].

A number of additional ideas are therefore required. In particular, we show how to combine Gaussian process approximation with coupling and concentration inequalities (and a careful stability analysis of a certain system of differential equations) to prove our central limit theorem *non-constructively*: we are able to prove a central limit theorem for $\nu(G)$ around its mean $\mathbb{E}\nu(G)$, despite not having any way to actually determine the value of $\mathbb{E}\nu(G)$ (of course, we have the estimate $\mathbb{E}\nu(G) = \alpha_c n + o(n)$ from Theorem 1.1, but the error term here is much larger than the typical fluctuations of $\nu(G)$).

Remark 1.3. There are a number of powerful general techniques to prove central limit theorems in random graphs (see for example [21, Section 6]). In particular, there is a recent general framework due to Cao [8] which is suitable for proving central limit theorems for a broad range of graph parameters defined in terms of optimisation problems with a "long-range independence" property. While a certain form of this property is satisfied for the maximum matching problem, it is not satisfied in a strong enough way[3] to apply the techniques in [8].

1.1. **The rank of a random graph.** Let $\operatorname{rk}(G)$ be the rank of the adjacency matrix of a graph G. It turns out that $\operatorname{rk}(G)$ is very closely related to $\nu(G)$, due to a connection between both of these parameters and the Karp–Sipser leaf-removal process. It was first observed by Bordenave, Lelarge and Salez [6] that

[2]It is difficult to precisely describe the issue without rather a lot of setup; a concrete description of the relevant problem will eventually appear at the end of Section 4.1.

[3]We remark that in the "smoother" setting of *weighted* sparse random graphs (in which a random Exponential(1) weight is assigned to each edge), it was proved by Gamarnik, Nowicki and Swirszcz [16] that the necessary long-range independence property *is* satisfied, so in this setting a central limit theorem (for the maximum weight of a matching) does immediately follow.

the statement of Theorem 1.1 holds with $\mathrm{rk}(G)/2$ in place of $\nu(G)$ (a more general connection between $\mathrm{rk}(G)$ and $\nu(G)$ was subsequently conjectured by Lelarge [24] and proved by Coja-Oghlan, Ergür, Gao, Hetterich, and Rolvien [10]). Using similar techniques as for Theorem 1.2, we are able to prove a central limit theorem for $\mathrm{rk}(G)$.

Theorem 1.4. *The statement of Theorem 1.2 holds with* $\mathrm{rk}(G)$ *in place of* $\nu(G)$.

We remark that we recently proved the $c > e$ case of Theorem 1.4 in [17] as a *corollary* of the $c > e$ case of Theorem 1.2, using a combinatorial description of the rank of a sparse random graph (which was the main result of [17]). Again, our main contribution here is the case $c \le e$.

1.2. **Notation.** We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ to mean that there is a constant C such that $|f(n)| \le C|g(n)|$ for sufficiently large n. Similarly, we write $f = \Omega(g)$ to mean that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for sufficiently large n. We write $f = \Theta(g)$ to mean that $f = O(g)$ and $g = \Omega(f)$, and we write $f = o(g)$ to mean that $f(n)/g(n) \to 0$ as $n \to \infty$. Subscripts on asymptotic notation indicate quantities that should be treated as constants.

1.3. **Acknowledgments.** We would like to thank Christina Goldschmidt and Eleonora Kreačić for insightful discussions and clarifications about their work in the thesis [23].

2. Proof overview

In this section we break down the proof of Theorems 1.2 and 1.4 into some key lemmas, which will be proved in the rest of the paper. This section also serves as an overview for the ideas in the proof and how they come together. Note that given the results in [17, 23], it suffices to prove Theorems 1.2 and 1.4 for $c \le e$, but we will leave remarks throughout describing how the situation changes in the case $c > e$.

2.1. **A random multigraph model.** For most of the proof, instead of dealing directly with random graphs, we will work with a closely related model of random *multigraphs*.

Definition 2.1. Let $\mathbb{G}^*(n, m)$ be a random m-edge multigraph on the vertex set $\{1, \dots, n\}$, whose edges are obtained as a sequence of m uniformly random pairs of (not necessarily distinct) vertices, sampled with replacement.

The advantage of $\mathbb{G}^*(n, m)$ is that if we condition on information about degrees of vertices, the conditional distribution remains tractable (as we will see in Section 2.4). We will need a theorem comparing random graphs and random multigraphs; to state this, we need a notion of graph similarity.

Definition 2.2. The *edit distance* $\mathrm{d_E}(G, G')$ between two multigraphs G, G' (on the same vertex set) is the number of edges that must be added and removed to obtain one from the other.

Now, our comparison theorem is as follows.

Theorem 2.3. *Fix a constant* $c > 0$*, and consider the vertex set* $\{1, \dots, n\}$*. Consider one of the following two situations.*

- *Let G contain a uniformly random subset of exactly $\lfloor cn/2 \rfloor$ of the possible edges. Let $G^* \sim \mathbb{G}^*(n, \lfloor cn/2 \rfloor)$.*
- *Let G be a random graph containing each of the $\binom{n}{2}$ possible edges with probability c/n independently. Let $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$ and let $G^* \sim \mathbb{G}^*(n, M)$.*

Then we can couple G, G^* such that $\mathrm{d_E}(G, G^*)$ is bounded in probability[4].

That is to say, a random graph with $\lfloor cn/2 \rfloor$ edges can be very well approximated by $\mathbb{G}^*(n, \lfloor cn/2 \rfloor)$, and a random graph with edge probability c/n can be very well approximated by first sampling its number of edges $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$, and then considering $\mathbb{G}^*(n, M)$. We prove Theorem 2.3 in Section 3, using a recent theorem of Janson [19].

For our purposes, the significance of the edit distance is that the rank and matching number are both *Lipschitz functions* with respect to this distance: if we add or remove a single edge, we change the rank by at most 2, and the matching number by at most 1. So, writing $\alpha(G) = \nu(G)$ or $\alpha(G) = \mathrm{rk}(G)/2$, we have

$$|\alpha(G) - \alpha(G')| \le \mathrm{d_E}(G, G'). \tag{2.1}$$

[4]Recall that a random variable X is *bounded in probability* if for any ε there is M such that $\Pr[|X| \ge M] \le \varepsilon$.

We remark that there are many other similar theorems comparing different models of random graphs and multigraphs. In particular, to prove a central limit theorem for a "binomial" random graph (where every edge is present with some probability p independently), a common approach would be to prove a central limit theorem for a random graph with a fixed number of edges, and then "integrate" that central limit theorem over possible numbers of edges (this approach appeared for example in the previous work of Pittel [27] and Kreačić [23]). Due to the non-constructiveness of our proof (briefly mentioned in the introduction), this approach is actually not possible for us, and we need the more direct coupling in Theorem 2.3.

2.2. **Karp–Sipser leaf removal.** Now we describe the Karp–Sipser leaf removal algorithm.

Definition 2.4 (Karp–Sipser leaf removal)**.** Starting from a (multi)graph G, repeatedly do the following. As long as there exist degree-1 vertices (*leaves*), choose one uniformly at random, and delete both this vertex and its unique neighbour. Let $G(i)$ be the graph remaining after i steps of this process, minus its isolated vertices (so $G(0)$ consists of G without its isolated vertices).

It is easy to see that a single step of leaf-removal decreases the matching number of G by exactly one, and the rank of G by exactly two. So, for any G and any time i (for which $G(i)$ is defined), writing $\alpha(G) = \nu(G)$ or $\alpha(G) = \mathrm{rk}(G)/2$, we always have

$$\alpha(G) = i + \alpha(G(i)). \tag{2.2}$$

It is natural to continue the leaf-removal process as long as possible, until the point when we run out of leaves (let I be this point in time, so $G(I)$ may or may not be empty, but definitely has no leaves). If G is an Erdős–Rényi random graph, then the final "Karp–Sipser core" $G(I)$ is quite well-behaved: the distribution of $G(I)$ has an explicit description in terms of some simple statistics of $G(I)$, and one can study its rank or matching number $\alpha(G(I))$ directly. So, in order to prove a central limit theorem for $\alpha(G)$, a sensible strategy is to study the joint distribution of I and of certain statistics of $G(I)$, and then study $\alpha(G(I))$ in terms of these statistics and apply (2.2). This was precisely the approach taken in [17, 23] for the regime $c > e$.

The significance of the distinction between $c \le e$ and $c > e$ is that it represents a "phase transition" for the behaviour of the above leaf–removal process up to time I, as follows (see for example [22]).

Theorem 2.5. *Fix a constant $c > 0$, and let G be as in Theorem 1.1. Run the Karp–Sipser leaf-removal process until the time I when $G(I)$ has no leaves remaining; then the number of vertices $v(G(I))$ in $G(I)$ satisfies*

$$\frac{v(G(I))}{n} \xrightarrow{p} \beta_c,$$

where $\beta_c = 0$ for $c \le e$ and $\beta_c > 0$ for $c > e$.

That is to say, if $c > e$ then the Karp–Sipser core $G(I)$ has size comparable to G, whereas if $c \le e$ then $G(I)$ is vanishingly small compared to G (in fact, when $c < e$, the expected number of vertices in $G(I)$ is only $O_c(1)$ [1]). From a certain point of view, this makes the subcritical regime $c < e$ seem *easier* than the supercritical regime $c > e$. Indeed, in the subcritical regime, $\alpha(G(I))$ is trivially almost zero, so recalling (2.2) it suffices to prove a central limit theorem for I. On the other hand, if $c > e$, one must work hard[5] to understand $\alpha(G(I))$, in addition to studying fluctuations related to the Karp–Sipser process[6].

However, in the regime $c \le e$ the leaf-removal process becomes more and more degenerate as we reach the end of the process. As $G(i)$ becomes smaller and smaller, we begin to lose law-of-large-numbers-type effects, and it is very difficult to maintain control over the evolution (and fluctuation) of various statistics. This is the key reason that the $c \le e$ cases of Theorems 1.2 and 1.4 were open until now.

[5]It turns out that $\alpha(G(I))$ is very likely to be very nearly equal to half the number of vertices in the Karp–Sipser core $G(I)$. Given this approximation, for the purpose of proving a central limit theorem, one only needs to understand the joint distribution of I and the number of vertices in the Karp–Sipser core. However, it is a highly nontrivial matter to actually prove this approximation, especially in the case $\alpha = \mathrm{rk}/2$; see [1, 15, 17].

[6]One may also wonder about the Karp–Sipser core $G(I)$ in the *critical* regime $c = e$. This regime is much more difficult to understand, and beyond Theorem 2.5 essentially nothing has been rigorously proved (though see the very recent work of Budzinski, Contat, and Curien [7] on a simpler model of random graphs, and the numerical simulations of Bauer and Golinelli [2]).

2.3. **A stopped central limit theorem.** In the regime $c \le e$ of interest, our approach is instead to stop the process *well before* time I, while it is "not too degenerate". Indeed, if we stop the process at some time I' when $G(I')$ has comparable size to G, then it is straightforward to adapt the techniques in [23] to prove a central limit theorem for I'. The specific result we need is as follows (recall the random multigraph $\mathbb{G}^*(n,m)$ from Definition 2.1).

Theorem 2.6. *Fix constants $0 < c \le e$ and $\delta > 0$, with δ sufficiently small in terms of c. Let G be a random graph defined in one of the following two ways:*

- *let $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$ and let $G \sim \mathbb{G}^*(n,M)$, or*
- *let $G \sim \mathbb{G}^*(n, \lfloor cn/2 \rfloor)$.*

Then, consider the Karp–Sipser leaf-removal process on G, and let I_δ be the first time i for which $G(i)$ has at most δn edges[7] *(let $I_\delta = \infty$ if this never happens, say). For some $\mu_\delta, \sigma_\delta$, we have*

$$\frac{I_\delta - \mu_\delta}{\sigma_\delta} \xrightarrow{d} \mathcal{N}(0,1).$$

We sketch the proof of Theorem 2.6 in Section 4 (it is nearly exactly the same as the proof in [23] for the $c > e$ case of Theorem 1.2[8]).

Recall from (2.2) that $\alpha(G) = I_\delta + \alpha(G(I_\delta))$. So, given Theorem 2.6, in order to prove a central limit theorem for $\alpha(G)$, it suffices to show that the fluctuations in $\alpha(G(I_\delta))$ are small compared to the fluctuations in I_δ. To be more precise, although Theorem 2.6 is stated for constant $\delta > 0$, a compactness argument shows that it also holds if $\delta \to 0$ sufficiently slowly. We want to show that as we send $\delta \to 0$, the fluctuations in $\alpha(G(I_\delta))$ become negligible compared to σ_δ.

It is rather difficult to actually prove (for $c \le e$ and $\delta \to 0$) that the fluctuations in I_δ dominate the fluctuations in $\alpha(G(I_\delta))$ (we need to get into the weeds of a complicated system of differential equations). However, we believe that it is very intuitive, morally speaking, that this should be the case. Indeed, in the case $c \le e$, imagine running the process *backwards* from $I = I_0$; at time I the "Karp–Sipser core" is basically empty and has tiny fluctuations, while we should expect plenty of fluctuation in the stopping time I itself. As we step backward in time, we expect the fluctuations in $\alpha(G(I_\delta))$ should gradually build up, but for small δ we do not expect these fluctuations to suddenly dominate the fluctuation in I_δ (when δ is a constant bounded away from zero, it is easy to see that these two quantities have fluctuations of the same order of magnitude).

Remark 2.7. The above strategy only makes sense for $c \le e$. Indeed, recall from Theorem 2.5 that in the regime $c > e$, the "core" $G(I)$ is large, so for any reasonable stopping time I' one should expect $\alpha(G(I'))$ to have rather large fluctuations.

To execute the above strategy, we need a lower bound on σ_δ, and an upper bound on the fluctuations in $\alpha(G(I_\delta))$. For both these bounds, we will need to carefully study the system of differential equations used to approximate the Karp–Sipser process, and for the latter bound we will also need a coupling argument taking advantage of the "smoothness" of the rank and matching number (as observed in (2.1)).

First, the following lemma records a lower bound on σ_δ.

Lemma 2.8. *Fix constants $c \le e$ and $\delta > 0$ and let $G, I_\delta, \mu_\delta, \sigma_\delta$ be as in Theorem 2.6. Then $\sigma_\delta = \Theta_c(\sqrt{n})$.*

Then, we need to prove an upper bound on the fluctuations of $\alpha(G(I_\delta))$. To this end, the first step is to obtain lower bounds on the fluctuations of its degree statistics, as follows.

Lemma 2.9. *Fix constants $c \le e$ and $\varepsilon > 0$. Then there is $\delta > 0$ such that the following holds for sufficiently large n. Let G, I_δ be as in Theorem 2.6. For each $d \ge 1$, let $X^{(d)}$ be the number of degree-d vertices in $G(I_\delta)$. Then there are $\mu^{(1)}, \dots, \mu^{(n)}$ such that:*

1. *$\sum_d d\mu^{(d)} \le \varepsilon n$, and*
2. *writing $D = \sum_{d=1}^\infty d\,|X^{(d)} - \mu^{(d)}|$, we have $\Pr[D > \varepsilon\sqrt{n}] \le \varepsilon$.*

[7] It would be more natural to define I_δ to be the first time i for which $G(i)$ has at most δn leaves (this time is always finite, even for $c > e$, and is a more natural generalisation of the stopping time I discussed earlier). However, for technical reasons this stopping time is somewhat less convenient to analyse.

[8] To say a bit more: the approach in the regime $c > e$ is to prove a bivariate central limit theorem for $I = I_0$ and the number of vertices $v(G(I))$ in the Karp–Sipser core. One can then deduce a central limit theorem for $\alpha(G)$, using (2.2) and a separate result due to Aronson, Frieze and Pittel [1, Theorem 3] approximating $\alpha(G(I))$ with $v(G(I))/2$.

The proofs of Lemmas 2.8 and 2.9 (which appear in Section 5) use the same Gaussian convergence machinery as the proof of Theorem 2.6, and also involve some rather nontrivial analysis of a system of differential equations associated with the Karp–Sipser process. In particular, Lemma 2.9 (which states that certain fluctuations become negligible as $\delta \to 0$) is especially delicate, because as the process evolves, the degree statistics actually experience larger and larger fluctuations. The key is that the stopping time I_δ itself has rather large fluctuations, and these fluctuations explain almost all the fluctuation in the degree statistics. That is to say, to prove Lemma 2.9, we need to show that near the end of the process, the fluctuations in all the degree statistics are very strongly correlated with the fluctuations in the number of edges. So, if we stop the process at the point where there are δn edges, we have essentially eliminated all fluctuation in all the degree statistics.

To actually prove the necessary correlation estimates on the degree statistics, the key idea is to consider a system of differential equations describing the joint evolution of all degree statistics, and show that, near the end of the process, this can be approximated by a *linear* system of differential equations. We then study the eigenvalues of this linear system, and find that all but one of the eigenvalues are negative. This means that fluctuations are suppressed in all but one direction, so near the end of the process all relevant fluctuations are highly correlated.

2.4. **Coupling configuration models.** Now, crucially, as first observed by Karp and Sipser [22], the degree statistics $X^{(d)}$ are sufficient to describe the distribution of $G(I_\delta)$: the conditional distribution of $G(I_\delta)$ given $(X^{(d)})_{d=1}^{\infty}$ is precisely described by an associated *configuration model*, as follows.

Definition 2.10. 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 $d_1, \dots, d_n$. 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.

Lemma 2.11. *Let G be as in Theorem 2.6. For any i, d, let $V^{(d)}(i)$ be the set of degree-d vertices in $G(i)$.*

For any stopping time I (with respect to the filtration described by the $V^{(d)}(i)$), the distribution of $G(I)$ may be described as follows. Let $V = V^{(1)}(I) \cup \dots \cup V^{(d)}(I)$, and define $\mathbf{d} \in \mathbb{N}^V$ by taking $d_v = d$ when $v \in V^{(d)}$. Then $G(I) \sim \mathbb{G}^*(\mathbf{d})$.

The proof of Lemma 2.11 is fairly immediate (we can view G as a uniformly random sequence of edges, and a random multigraph $G^* \sim \mathbb{G}^*(\mathbf{d})$, with its edges randomly ordered, can be interpreted as a uniformly random sequence of edges constrained to have degree sequence $\mathbf{d}$). Alternatively, Lemma 2.11 follows directly from [1, Lemma 2].

Recall that our goal is to upper-bound the fluctuation in $\alpha(G(I_\delta))$, using the upper bounds on fluctuations of degree statistics in Lemma 2.9. To this end, we need the following coupling lemma for random configurations: if we have two degree sequences $\mathbf{d}, \mathbf{d}'$ which are statistically similar, then we can couple the corresponding configuration models $\mathbb{G}^*(\mathbf{d}), \mathbb{G}^*(\mathbf{d}')$ to be close with respect to edit distance (recall the definition of edit distance from Definition 2.2).

Lemma 2.12. *Fix two degree sequences $\mathbf{d} = (d_1, \dots, d_n)$ and $\mathbf{d}' = (d_1', \dots, d_n')$. Then we can couple $G \sim \mathbb{G}^*(\mathbf{d})$ and $G' \sim \mathbb{G}^*(\mathbf{d}')$ such that $\mathrm{d_E}(G, G') \le \sum_v |d_v - d_v'| + 1$.*

Proof. Let $d_v^{\max} = \max(d_v, d_v')$ for all $v \in [n]$ (we then increase some $d_v^{\max}$ by one, if necessary, to ensure that $\sum_v d_v^{\max}$ is even). Let $\mathbf{d}^{\max} = (d_v^{\max})_v \in \mathbb{N}^n$.

Now, consider $\sum_v d_v^{\max}$ stubs, grouped into buckets of sizes $d_v^{\max}$. For each vertex v, label $d_v^{\max} - d_v$ of its stubs as being "$\mathbf{d}$-bad", and label $d_v^{\max} - d_v'$ of stubs as being "$\mathbf{d}'$-bad". If a stub is not $\mathbf{d}$-bad we say it is "$\mathbf{d}$-good" and if a stub is not $\mathbf{d}'$-bad we say it is "$\mathbf{d}'$-good" So, a perfect matching of all the stubs is a configuration for $\mathbf{d}^{\max}$, a perfect matching of all the $\mathbf{d}$-good stubs is a configuration for $\mathbf{d}$, and a perfect matching of all the $\mathbf{d}'$-good stubs is a configuration for $\mathbf{d}'$.

Starting from a uniformly random configuration $\pi^{\max}$ for $\mathbf{d}^{\max}$, we can obtain a random configuration π for $\mathbf{d}$ as follows. First, delete all matching edges involving a $\mathbf{d}$-bad stub. Some $\mathbf{d}$-good stubs are now unmatched; choose a uniformly random perfect matching of these unmatched $\mathbf{d}$-good stubs to extend our matching to a configuration π for $\mathbf{d}$. By symmetry, π is a uniformly random configuration for $\mathbf{d}$, and $\mathrm{d_E}(\pi^{\max}, \pi)$ is at most the number of $\mathbf{d}$-bad stubs, which is $\sum_v (d_v^{\max} - d_v)$.

In the same way, starting from $\pi^{\max}$ we can obtain a uniformly random configuration π' for $\mathbf{d}'$ with $\mathrm{d_E}(\pi^{\max}, \pi') \le \sum_v (d_v^{\max} - d'_v)$. By the triangle inequality we then have

$$\mathrm{d_E}(\pi, \pi') \le \sum_v (d_v^{\max} - d_v) + \sum_v (d_v^{\max} - d'_v) \le \sum_v |d_v - d'_v| + 1,$$

as desired. □

2.5. **Smoothness of the rank and matching number.** Recall from (2.1) that the rank and matching number are both *Lipschitz functions* in terms of edit distance, that is,

$$|\alpha(G) - \alpha(G')| \le \mathrm{d_E}(G, G').$$

This implies that $\alpha(G)$ is well-concentrated for random G. We need this fact for a few different models of random (multi)graphs G, as follows.

Lemma 2.13. *The following hold:*

1. *Let G be a random graph as in Theorem 1.1. Then $\alpha(G)$ is subgaussian*[9] *with variance proxy $O_c(n)$.*
2. *Let $G \sim \mathbb{G}^*(\mathbf{d})$ for some degree sequence $\mathbf{d} = (d_1, \dots, d_n)$. Then $\alpha(G)$ is subgaussian with variance proxy $O(d_1 + \cdots + d_n)$.*

Proof. First, (1) is easily proved with the Azuma–Hoeffding inequality (see for example the appendix of [6]).

For (2), note that a uniformly random configuration can be defined in terms of a uniformly random permutation σ of length $N := d_1 + \cdots + d_n$. Indeed, consider $d_1 + \cdots + d_n$ stubs divided into buckets of sizes $d_1, \dots, d_n$ and consider the configuration (i.e., perfect matching on the stubs) with edges

$$\sigma(1)\sigma(2),\ \sigma(2)\sigma(3),\ \dots,\ \sigma(N-1)\sigma(N).$$

Modifying σ by a transposition results in a change of at most two edges of our random configuration (which changes $\alpha(G)$ by at most 2), so the desired result follows from a version of the Azuma–Hoeffding inequality for random permutations (see for example [28, Theorem 5.2.6]). □

2.6. **Completing the proof.** We now show how to combine all the relevant ingredients to prove Theorems 1.2 and 1.4.

Proof of Theorems 1.2 and 1.4. Let $\alpha(G)$ be $\nu(G)$ or $\mathrm{rk}(G)/2$. Recall that we only need to handle the case $c \le e$. We fix a constant $c \le e$ throughout this proof (implicit constants in asymptotic notation are allowed to depend on c).

First, a minor technical remark: what we will prove is that there are some μ, σ such that $(\alpha(G) - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$. *A priori*, there may be no connection between μ and $\mathbb{E}[\alpha(G)]$ or between σ^2 and $\mathrm{Var}[\alpha(G)]$, if the mean or variance of $\alpha(G)$ is dominated by the effect of outliers. However, such pathological behaviour is ruled out by Lemma 2.13(1).

Then, observe that, using Theorem 2.3 and (2.1), it suffices to consider G drawn from one of the two random multigraph models in Theorem 2.6 (instead of the two random graph models in Theorem 1.1). Indeed, we can assume that the $O(1)$ edit-distance error arising from Theorem 2.3 and (2.1) is negligible relative to the fluctuation in $\alpha(G)$: note that the conclusion of Theorem 2.6 can only hold for $\sqrt{\mathrm{Var}[\alpha(G)]} \to \infty$, because $\alpha(G)$ only takes values in the lattice $(1/2)\mathbb{Z}$.

Now, recall that convergence in distribution is metrisable: for example, the *Prohorov* metric $\mathrm{d_P}$ on real probability distributions is such that $X_n \xrightarrow{d} Z$ if and only if $\mathrm{d_P}(X_n, Z) \to 0$ (see for example [4, Section 6]). Recall the definitions of I_δ and $D = \sum_{d=1}^{\infty} d\,|X^{(d)} - \mu^{(d)}|$ from Theorem 2.6 and Lemma 2.9, let $\mu_\delta, \sigma_\delta$ be as in Theorem 2.6, let $\delta = \delta(\varepsilon) > 0$ be as in Lemma 2.9, and note that Theorem 2.6 and Lemma 2.9 together imply that for any $\varepsilon > 0$ we have

$$\mathrm{d_P}\left(\frac{I_\delta - \mu_\delta}{\sigma_\delta},\ \mathcal{N}(0,1)\right) \le \varepsilon \text{ and } \Pr[D > \varepsilon\sqrt{n}] \le \varepsilon \tag{2.3}$$

for sufficiently large n (say, $n \ge n_\varepsilon$). Since this holds for any constant $\varepsilon > 0$, one can abstractly show that in fact (2.3) holds for some $\varepsilon = o(1)$ (decaying with n). Indeed, for $n \ge n_1$, let $k_n = \max\{k \colon n \ge n_{1/k}\}$,

[9]We say that a random variable X is *subgaussian* with *variance proxy* ν if $\Pr[|X - \mathbb{E}X| \ge x] \le O(\exp(-x^2/\nu))$ for all $x \in \mathbb{R}$.

so (2.3) holds for $\varepsilon = 1/k_n = o(1)$. We have proved that

$$\frac{I_\delta - \mu_\delta}{\sigma_\delta} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{and} \quad \frac{D}{\sqrt{n}} \xrightarrow{p} 0 \tag{2.4}$$

(where now $\delta = \delta(\varepsilon)$ depends on n via ε).

The second part of (2.4) says that there is some $\rho = o(1)$ such that whp[10] $D \le \rho\sqrt{n}$. That is to say, writing $\mathcal{X}$ for the set of all sequences $(x^{(d)})_{d=1}^\infty \in \mathbb{Z}_{\ge 0}^{\mathbb{N}}$ such that $\sum_{d=1}^\infty d|x^{(d)} - \mu^{(d)}| \le \rho\sqrt{n}$, we have $(|X^{(d)}|)_{d=1}^\infty \in \mathcal{X}$ whp (recall that $X^{(d)}$ is the number of degree-d vertices that remain at time I_δ).

For each $\mathbf{x} = (x^{(d)})_{d=1}^\infty \in \mathcal{X}$, let $\mathcal{E}_\mathbf{x}$ be the event that $(|X^{(d)}|)_{d=1}^\infty = \mathbf{x}$ and let $G_\mathbf{x} \sim \mathbb{G}^*(\mathbf{d})$, where $\mathbf{d}$ is a degree sequence containing $x^{(d)}$ copies of each d. By Lemma 2.11, up to relabelling vertices, the conditional distribution of $G(I_\delta)$ given $\mathcal{E}_\mathbf{x}$ is precisely that of $G_\mathbf{x}$ (so, in particular, $\mathbb{E}[\alpha(G(I_\delta))\,|\,\mathcal{E}_\mathbf{x}] = \mathbb{E}[\alpha(G_\mathbf{x})]$). Note that the number of vertices in $G_\mathbf{x}$ is $\sum_{d=1}^\infty x^{(d)} \le \varepsilon n + \rho\sqrt{n} \le 2\varepsilon n$ by Lemma 2.9(1).

Recalling that $\varepsilon = o(1)$, by Lemma 2.13(2) we have

$$\Pr\Big[\big|\alpha(G(I_\delta)) - \mathbb{E}[\alpha(G_\mathbf{x})]\big| \ge \varepsilon^{1/3}\sqrt{n}\,\Big|\,\mathcal{E}_\mathbf{x}\Big] = O(\exp(-\Omega(\varepsilon^{-1/3}))) = o(1). \tag{2.5}$$

Now, consider $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, with corresponding degree sequences $\mathbf{d}$ and $\mathbf{d}'$. Reorder $\mathbf{d}, \mathbf{d}'$ such that $d_v = d'_v$ for as many vertices as possible, so $\sum_v |d_v - d'_v| \le \sum_d d\,|x^{(d)} - x^{(d')}| \le 2D = o(\sqrt{n})$. Then, Lemma 2.12 tells us that we can couple $G_\mathbf{x}$ and $G_{\mathbf{x}'}$ such that $\mathrm{d_E}(G_\mathbf{x}, G_{\mathbf{x}'}) = o(\sqrt{n})$, meaning that $|\mathbb{E}[\alpha(G_\mathbf{x})] - \mathbb{E}[\alpha(G_{\mathbf{x}'})]| = o(\sqrt{n})$ (recalling (2.1)). Since this is true for each $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, it follows that there is some μ_{res} such that $|\mathbb{E}[\alpha(G_\mathbf{x})] - \mu_{\mathrm{res}}| = o(\sqrt{n})$ for each $\mathbf{x} \in \mathcal{X}$. (We remark that we do not actually know the value of μ_{res}; in this sense our proof is "non-constructive").

Combining this with (2.5), we deduce that

$$\frac{\alpha(G(I_\delta)) - \mu_{\mathrm{res}}}{\sqrt{n}} \xrightarrow{p} 0.$$

Finally, recalling from Lemma 2.8 that $\sigma_\delta = \Theta(n^{1/2})$, and recalling the first part of (2.4), we see that

$$\frac{\alpha(G) - \mu_\delta - \mu_{\mathrm{res}}}{\sigma_\delta} \xrightarrow{d} \mathcal{N}(0,1)$$

since $\alpha(G) = \alpha(G(I_\delta)) + I_\delta$, as desired. □

3. Reduction to Multigraphs

In this section we prove Theorem 2.3. For G, G^* as in Theorem 2.3, note that the distribution of G may be obtained from the distribution of G^* simply by conditioning on the event that G^* is a simple graph (note that every m-edge simple graph is an equally likely outcome of $\mathbb{G}^*(n,m)$). Our proof of Theorem 2.3 has two parts. First, by an estimate of McKay and Wormald [25], conditioning on simplicity does not significantly bias the distribution of the degree sequence. Second, by a result of Janson [19], given a particular degree sequence $\mathbf{d}$, we can efficiently couple the configuration model $\mathbb{G}^*(\mathbf{d})$ with a random graph constrained to have degree sequence $\mathbf{d}$.

We start with the (very simple) fact that in a sparse random graph with average degree about cn, the maximum degree is at most $\log n$, and the second factorial-moment of the degrees is about $c^2 n$.

Definition 3.1. For a constant c, say that a degree sequence $\mathbf{d} = (d_1, \dots, d_v)$ is *(n,c)-good* if

- $\max_v d_v \le \log n$,
- $\left|\sum_v d_v - cn\right| \le n^{3/4}$,
- $\left|\sum_v d_v(d_v - 1) - c^2 n\right| \le n^{3/4}$.

Lemma 3.2. *Fix a constant $c > 0$ and let G, G^* be as in Theorem 2.3. Then whp the degree sequences of G, G^* are both (n,c)-good.*

Proof. It is well-known (see for example [5, Lemma 1]) that the degree sequence of G and G^* can both be obtained by considering a sequence of n independent Poisson(c) random variables, and conditioning on an event that holds with probability $\Omega_c(1/\sqrt{n})$. Then the desired result follows from a Chernoff bound (noting that if $Q \sim \mathrm{Poisson}(c)$ then $\mathbb{E}[Q(Q-1)] = c^2$). □

The following estimate on the simplicity probability follows from, for example, [25, Lemma 5.1].

[10] We say an event holds *with high probability*, or *whp* for short, if it holds with probability $1 - o(1)$.

Lemma 3.3. *Suppose* $\mathbf{d}$ *is an* (n, c)*-good degree sequence, and let* $G^* \sim \mathbb{G}^*(\mathbf{d})$*. Then*

$$\Pr[G^* \text{ is simple}] = \exp(-c/2 - c^2/4) + o(1).$$

Then, we need the following consequence of [19, Theorems 2.1 and 3.2], due to Janson.

Theorem 3.4. *Suppose* $\mathbf{d}$ *is an* (n, c)*-good degree sequence. Let* $G^* \sim \mathbb{G}^*(\mathbf{d})$*, and let* G *be a uniformly random graph on the vertex set* $\{1, \dots, n\}$ *with degree sequence* $\mathbf{d}$*. Then we can couple* G, G^* *such that* $\mathrm{d_E}(G, G^*)$ *is bounded in probability.*

Now we prove Theorem 2.3.

Proof of Theorem 2.3. Let $\mathbf{D}, \mathbf{D}^*$ be the (random) degree sequences of G and G^*, and recall that G can be obtained by conditioning on simplicity of G^*. For any (n, c)-good degree sequence $\mathbf{d}$, using Lemma 3.3 we have

$$\Pr[\mathbf{D} = \mathbf{d}] = \frac{\Pr[\mathbf{D}^* = \mathbf{d}] \Pr[G^* \text{ is simple} \,|\, \mathbf{D}^* = \mathbf{d}]}{\Pr[G^* \text{ is simple}]} = \frac{\exp(-c/2 - c^2/4) + o(1)}{\Pr[G^* \text{ is simple}]} \Pr[\mathbf{D}^* = \mathbf{d}].$$

By Lemma 3.2, the sum of $\Pr[\mathbf{D} = \mathbf{d}]$ over all good $\mathbf{d}$ and the sum of $\Pr[\mathbf{D}^* = \mathbf{d}]$ over all good $\mathbf{d}$ are both $1 - o(1)$. So, the above equation implies that

$$\frac{\exp(-c/2 - c^2/4) + o(1)}{\Pr[G^* \text{ is simple}]} = 1 + o(1),$$

meaning that for each good $\mathbf{d}$ we have $\Pr[\mathbf{D} = \mathbf{d}] = (1 + o(1)) \Pr[\mathbf{D}^* = \mathbf{d}]$.

So, we can couple $\mathbf{D}, \mathbf{D}^*$ to be equal whp. Then, given outcomes of $\mathbf{D}, \mathbf{D}^*$, we have $G^* \sim \mathbb{G}^*(\mathbf{D}^*)$, and G is a uniformly random graph with degree sequence $\mathbf{D}$, so the desired result follows from Theorem 3.4. □

4. Analysis of the Karp–Sipser process

In this section we sketch how to prove Theorem 2.6 using the approach in Kreačić's thesis [23] (a stochastic generalisation of the differential equations method). As we will see, we do not require any change to the proof approach in [23]; we simply need to change the quantity we are interested in estimating (so, the reader may wish to refer to [23] for more details). The concepts and notation in this section will also be important for the proof of Lemma 2.9, which will appear in the next section.

Where convenient, we use the same notation as in [23] (in particular, contrary to the preceding sections, for objects that depend on n we explicitly write a superscript n, and we do not use boldface for vectors). Throughout this section we fix a constant c (implicit constants in asymptotic notation are allowed to depend on c).

4.1. A general framework for distributional approximation of Markov chains.

Before we begin to discuss the details of the proof of Theorem 2.6, we orient the unfamiliar reader with the general framework of Ethier and Kurtz [14, Chapter 11] for approximating sequences of (time- and space-) inhomogeneous Markov Chains by Gaussian processes. We stress that this should be treated merely as an outline/sketch; more details can be found in [23, Section 2.3].

One way to characterise (d-dimensional) Brownian motion is as a scaling limit of a sequence of unbiased random walks on the integer lattice $\mathbb{Z}^d$. In [14, Chapter 11], Ethier and Kurtz situated this in a much more general framework. Given a collection of "rate functions" $\beta_l : \mathbb{R} \to \mathbb{R}_{\geq 0}$ (for $l \in \mathbb{Z}^d$), for each $n \in \mathbb{N}$ we define a space-inhomogeneous random walk $(U^n(t))_{t \geq 0}$ in $\mathbb{Z}^d$: when $U^n(t)$ is located at position $k \in \mathbb{Z}^d$, steps in direction l are taken with rate $n\beta_l(k/n)$. Under appropriate assumptions, Ethier and Kurtz proved that, as $n \to \infty$, a sequence of random walks of this type converges to a certain Gaussian process which can be described as the solution to a stochastic partial differential equation involving the functions β_l.

In more detail: define the "drift" function $F(y) = \sum_{l \in \mathbb{Z}^d} \beta_\ell(y)$, and let[11] $(u(s))_{s \geq 0}$ be the solution to the differential equation $u(s) = U^n(0)/n + \int_0^s F(u(q))\, dq$. Under appropriate assumptions, we expect $U^n(t)$ to be approximately equal to $nu(t/n)$: indeed, the position of our particle at time t is approximately the accumulation of the rate functions until time t, taking into account the changes in these rate functions as the particle moves through space. This function u is sometimes known as the *fluid limit approximation* of our discrete-time process. The method of obtaining this function and showing

[11] Departing slightly from [23], we use the letter s to denote time "in the continuous world" and t to denote time "in the discrete world" (these differ by a factor of n).

that it indeed approximates $U^n(t)$ (in probability) is (an instance of) the *differential equation method* in combinatorics.

Next, for $x \in \mathbb{R}^d$, define the matrix $\partial F(x) \in \mathbb{R}^{d\times d}$ by $(\partial F(x))_{i,j} = \partial F_i(x)/\partial x_j$, and for each $l \in \mathbb{Z}^d$, let $W_l\colon [0,\infty) \to \mathbb{R}$ be a standard Brownian motion. Then, define the (continuous-time) random process $(V(s))_{s\geq 0}$ as the solution to the SPDE

$$V(s) = V(0) + \sum_{l\in\mathbb{Z}^d} W_l\Big(\int_0^s \beta_l(\chi(q))dq\Big)l + \int_0^s \partial F(u(q))V(q)dq, \tag{4.1}$$

where $V(0)$ is a Gaussian random variable approximating the initial fluctuations of $U^n(0)/n$ (we allow for the possibility that the particle starts in a random position). Under appropriate assumptions, we expect $(U^n(ns) - nu(s))/\sqrt{n}$ to be approximately distributed as $V(s)$ (jointly for all s up to any fixed time horizon). Indeed, the first term captures the Gaussian fluctuation remaining after approximating the accumulation of random steps by its fluid limit approximation. The second term captures the fluctuation in the accumulation of the drift function F itself, due to the fact that the drift "should really" be evaluated at $U^n(q)/n$, and not the deterministic vector $u(q)$ (the idea is that $\partial F(u(q))V(q)$ describes the extent to which the fluctuation in $V(q)$ affects this difference). The validity of this approximation is the main content of the Ethier–Kurtz framework.

It is possible to solve the SPDE in (4.1) in terms of a Gaussian process and a function Φ which is itself defined in terms of a deterministic differential equation (i.e., we can describe $V(s)$ "explicitly", instead of in terms of the solution to an SPDE). Specifically, for $s, u \geq 0$, let $\Phi(s,u) \in \mathbb{R}^{d\times d}$ be the matrix solution to the system of differential equations

$$\frac{\partial}{\partial s}\Phi(s,u) = \partial F(\chi(s))\Phi(s,u),$$

with boundary conditions $\Phi(u,u) = I$ for all $u \geq 0$. This matrix function Φ can be thought of as a "temporal correlation function" measuring the extent to which fluctuations at time u influence fluctuations at some later time s. Under appropriate assumptions, the solution to the SPDE in (4.1) is given by

$$V(s) = \Phi(s,0)V(0) + \int_0^s \Phi(s,u)\, d\mathcal{W}(u), \tag{4.2}$$

where

$$\mathcal{W}(s) = \sum_{l\in\mathbb{Z}^4} W_l\Big(\int_0^s \beta_l(\chi(q))dq\Big)l.$$

(Note that $V(s)$ is a Gaussian process since it is an Itô integral with respect to a Gaussian process of a deterministic function.)

Also, we remark that from the description in (4.2) and the Ethier–Kurtz approximation theorem, it is possible to deduce the approximate distribution of U^n at certain stopping times: for example, if for some $\delta \geq 0$ and some coordinate $i \leq d$, we define τ_δ^n to be the (random) first point in time where $U_i^n \leq \delta n$, and s_δ to be the minimum value of s such that $u(s) \leq \delta$, then (under appropriate assumptions) we have

$$\frac{U^n(\tau_\delta^n) - nu(s_\delta)}{\sqrt{n}} \xrightarrow{d} V(s_\delta) - \frac{V_i(s_\delta)}{F_3(u(s_\delta))}F(u(s_\delta)) \tag{4.3}$$

(i.e., we "subtract away" the fluctuations in U^n due to the fluctuations in the stopping time τ_δ^n itself). **This last step is the critical place where degeneracy issues can cause problems:** specifically, we will run into problems if $F_3(u(s_\delta)) = 0$ (in which case, attempting to adjust for the fluctuation in τ_δ^n amounts to an indeterminate division $0/0$).

In the rest of this section, we will sketch the application of the framework described above (with certain small technical modifications) to study the Karp–Sipser process and prove Theorem 2.6. As we will see, the only place where we depart from [23] is that we use a different stopping time τ_δ^n, chosen to avoid the type of degeneracy described above.

4.2. **Setup for the Markov Chain.** To be able to easily apply the Ethier–Kurtz machinery, it is convenient to study the Karp–Sipser process in continuous time, according to a "Poisson clock": at time zero, a leaf-removal takes place, and then after each leaf-removal we wait for an Exponential(1) amount of time before the next leaf-removal. For $t \geq 0$:

- let $X_1^n(t)$ be the number of vertices with degree 1 at time t,
- let $X_2^n(t)$ be the number of vertices of degree at least 2 at time t,
- let $X_3^n(t)$ be the number of edges at time t, and

- let $X_4^n(t)$ be the number of leaf-removal steps up until time t.

Then, let $X^n(t) = (X_1^n(t), X_2^n(t), X_3^n(t), X_4^n(t))$. As observed in [1, Lemma 2] (restated as [23, Lemma 5]; see also [22, Section 3]), $X^n(t)$ is a continuous-time Markov chain. Let $\tau_\delta^n = \inf\{t \geq 0 : X_3^n(t) \leq \delta n\}$ (with $\tau_\delta^n = \infty$ if $X_3^n(t) > \delta n$ for all t); we will study the evolution of X^n until time τ_δ^n.

Note that the random variable I_δ in Theorem 2.6 is precisely $X_4(\tau_\delta^n)$ with this setup (with the convention that $X_4^n(\infty) = \infty$).

4.3. **Rate and drift functions.** To understand the evolution of $X^n(t)$, we need to study the transition probabilities corresponding to a single random leaf-removal.

When we delete a leaf v together with its neighbour w, we also delete all other edges incident to w, and all the neighbours of w get their degree reduced by 1. The effect this has on $X_1^n(t), X_2^n(t), X_3^n(t)$ can be described in terms of the degrees of w and its neighbours (with respect to the multigraph $G^n(t)$ remaining at time t). In order to study the distribution of these quantities, one can explicitly describe the conditional distribution of $G^n(t)$ given $X_1^n(t), X_2^n(t), X_3^n(t)$: this conditional distribution of $G^n(t)$ is the same as the conditional distribution of $\mathbb{G}^*(X_1(t)+X_2(t),\ X_3(t))$, given that there are $X_1(t)$ vertices of degree 1 and $X_2(t)$ vertices of degree at least 2. One can study the typical degree distribution of this random graph, and then use standard techniques for studying random graphs with given degree sequences.

In a bit more detail: it turns out that the degree distribution in such a graph $G^n(t)$ can be described by a truncated Poisson distribution with a certain parameter depending on X_1^n, X_2^n, and X_3^n. Indeed, for $x = (x_1, x_2, x_3, x_4) \in (\mathbb{R}_{\geq 0})^4$ with $x_2 > 0$ and $2x_3 \geq x_1 + 2x_2$, let $z(x) \geq 0$ be the unique solution[12] to

$$\frac{z(x)(e^{z(x)} - 1)}{e^{z(x)} - z(x) - 1} = \frac{2x_3 - x_1}{x_2},$$

and let $Z^n(t) = z(X^n(t))$. Then, consider any time t where the process is "not too degenerate" (in notation to be introduced in Section 4.4, we can take any $t \leq \tau_\delta^n \wedge \zeta_\delta^n$), and condition on a corresponding outcome of $X^n(t)$. For any of the $X_2^n(t)$ vertices with degree at least 2, and any $d \leq \log n$, the probability the degree of that vertex is exactly d is

$$\Pr[Q = d \,|\, Q \geq 2] + o\left(\frac{\log^3 n}{n}\right),$$

where $Q \sim \text{Poisson}(Z^n(t))$. That is to say, the degree distribution is roughly "truncated Poisson" with parameter $Z^n(t)$. (One can show that whp no vertex ever has degree greater than $\log n$, so it suffices to consider $d \leq \log n$). The above fact appears as [23, Lemma 22][13] (deduced from [1, Lemma 5]). Using this fact, the transition probabilities for a single leaf-removal step are computed in [23, Theorem 28]: given that $(X_1^n(t), X_2^n(t), X_3^n(t)) = (q_1, q_2, q_3)$ for some $t \leq \tau_\delta^n \wedge \zeta_\delta^n$, we consider the conditional probability that the next-removed leaf v has a neighbour w which itself has $k_1 \geq 1$ degree-1 neighbours, k_2 degree-2 neighbours and k_3 neighbours of degree at least 3 (which causes $(X_1^n, X_2^n, X_3^n) = (q_1 - k_1 + k_2,\ q_2 - 1 - k_2,\ q_3 - k_1 - k_2 - k_3)$ if $k_1 + k_2 + k_3 \geq 2$, and $(X_1^n, X_2^n, X_3^n) = (q_1 - 2,\ q_2,\ q_3 - 1)$ if $k_1 + k_2 + k_3 = 1$). This conditional probability is shown to be a somewhat complicated formula involving $q_1, q_2, q_3, k_1, k_2, k_3$, plus an additive error term of the form $o(\log^3 n/n)$.

This estimate for the 1-step transition probabilities translates into an estimate for the transition rates of the continuous-time Markov chain $X^n(t)$: specifically, the approximate transition rate from a state $q \in \mathbb{Z}_{\geq 0}^4$ to a state $q + l \in \mathbb{Z}_{\geq 0}^4$ is $n\beta_l(q/n)$, for rate functions β_l defined as follows. Let

$$\mathcal{K} = \{(k_1, k_2, k_3) : k_1 \geq 1,\ k_2, k_3 \geq 0,\ k_1 + k_2 + k_3 \geq 2\}$$

and for $l = (-k_1 + k_2, -1 - k_2, -k_1 - k_2 - k_3, 1)$ with $(k_1, k_2, k_3) \in \mathcal{K}$, let $\beta_l(x)$ be

$$\frac{1}{(k_1 - 1)!k_2!k_3!} \cdot \frac{x_2}{2x_3} \cdot \frac{z(x)}{e^{z(x)} - z(x) - 1} \left(\frac{x_1}{2x_3} z(x)\right)^{k_1 - 1} \left(\frac{x_2 z(x)^2}{2x_3(e^{z(x)} - z(x) - 1)} z(x)\right)^{k_2} \left(\frac{x_2 z(x)}{2x_3} z(x)\right)^{k_3}.$$

Also, let $\beta_{(-2,0,-1,1)}(x) = x_1/(2x_3)$, and let $\beta_l(x) = 0$ for all other l.

[12]In [23], Kreačić writes "z_x"; we have changed the notation for readability.

[13]This is stated for $t \leq \tau^n \wedge \zeta^n$ (for $c > e$) instead of $t \leq \tau_\delta^n \wedge \zeta_\delta^n$, but the proof is exactly the same: the only role ζ^n plays is to ensure that $X_1^n(t), X_2^n(t), X_3^n(t)$ are of size $\Omega(n)$.

Taking a weighted sum of the approximate transition rates $\beta_l(x)$ allows us to estimate the expected infinitesimal change to $X^n(t)$: as in [23, Equation 2.46] we can define the "drift function"

$$F(x) = \sum_{l \in \mathbb{Z}^4} \beta_l(x) l = (F_1(x), F_2(x), F_3(x), F_4(x))$$

and compute

$$F_1(x) = -1 - \frac{x_1}{2x_3} + \frac{x_2^2 z(x)^4 e^{z(x)}}{(2x_3 f(z(x)))^2} - \frac{x_1 x_2 z(x)^2 e^{z(x)}}{(2x_3)^2 f(z(x))} \tag{4.4}$$

$$F_2(x) = -1 + \frac{x_1}{2x_3} - \frac{x_2^2 z(x)^4 e^{z(x)}}{(2x_3 f(z(x)))^2} \tag{4.5}$$

$$F_3(x) = -1 - \frac{x_2 z(x)^2 e^{z(x)}}{2x_3 f(z(x))} \tag{4.6}$$

$$F_4(x) = 1. \tag{4.7}$$

where $f(x) = e^x - x - 1$.

4.4. **Fluid limit approximation.** Given the rate functions β_l, we can now solve a system of differential equations to obtain[14] a *fluid limit approximation* χ for X^n. In this section we record formulas for this fluid limit approximation (and record the theorem that the trajectory of X^n does indeed concentrate around this approximation).

Let $p(u) = e^{-u}(e^u - u - 1)$, let $\beta(u)$ be the unique solution to $\beta(u)e^{c\beta(u)} = e^u$ and implicitly define the function $\vartheta\colon [0, \infty) \to \mathbb{R}$ by

$$s = \frac{1}{c}\left(c(1 - \beta(\vartheta(s))) - \frac{1}{2}\log^2 \beta(\vartheta(s))\right).$$

Then let

$$\chi_1(s) = \frac{1}{c}\left(\vartheta^2(s) - \vartheta(s) \cdot c \cdot \beta(\vartheta(s))(1 - e^{-\vartheta(s)})\right),$$
$$\chi_2(s) = p(\vartheta(s))\beta(\vartheta(s)),$$
$$\chi_3(s) = \frac{1}{2c}\vartheta^2(s)$$
$$\chi_4(s) = s.$$

The following theorem is presented as [23, Theorem 20], as a consequence of estimates in [1]. It holds for all $c > 0$.

Theorem 4.1. *With notation as defined in this section, for $s^* = \inf\{s \geq 0 \colon \chi_1(s) = 0\}$ and any $s < s^*$ (not depending on n) we have*

$$\sup_{u \leq s}\left\|\frac{X^n(nu)}{n} - \chi(u)\right\|_\infty \xrightarrow{p} 0.$$

For $c \leq e$, let $s_\delta = \inf\{s \geq 0 \colon \chi_3(s) \geq \delta\} < s^*$ be the "fluid limit prediction" for τ_δ^n. One can check (e.g., using the series expansions we will compute in Section 5.1) that s_δ is finite and well-defined, and that $\chi_1(s_\delta), \chi_2(s_\delta) > 0$ (i.e., at time τ_δ^n, there are likely to be $\Omega_\delta(n)$ vertices of degree at least 2, and therefore $\Omega_\delta(n)$ edges). Let

$$\zeta_\delta^n = \inf\{t \geq 0 : X_1^n(t) \leq \chi_1(s_\delta)n/2 \text{ or } X_2^n(t) \leq \chi_2(s_\delta)n/2\},$$

so whp $\tau_\delta^n \leq \zeta_\delta^n$ (that is to say, whp $X_1^n(t), X_2^n(t), X_3^n(t)$ are whp of size $\Omega_\delta(n)$ until time τ_δ^n).

Remark 4.2. The stopping times τ_δ^n and ζ_δ^n are the main point of difference between our setting and Kreačić's thesis [23]. In [23], the hitting times $\tau^n = \inf\{t \geq 0 : X_1^n(t) = 0\}$ and $\zeta^n = \inf\{t \geq 0 : X_2^n(t) \leq \chi_2(s^*)/2 \text{ or } X_3^n(t) \leq \chi_3(s^*)/2\}$ are instead considered; when $c > e$, one can check that $\chi_2(s^*), \chi_3(s^*) > 0$, and $\tau^n \leq \zeta^n$ whp.

[14] As described in Section 4.1, it is possible to prove a fluid limit approximation by studying a differential equation involving the drift functions F. However, this is not actually done in Kreačić's thesis [23]: it is more convenient simply to cite the previous work in [1] (which has slightly different notation and a different formulation of the relevant differential equations).

4.5. **Initial fluctuations.** Knowing the transition rates for our Markov chain is essentially tantamount to knowing the full distribution of $(X^n(t))_{t\geq 0}$, except that we also need to estimate the (asymptotically multivariate Gaussian) distribution of the initial state $X^n(0)$. This routine computation is performed in [23, Section 2.3.2] in the case where $G \sim \mathbb{G}^*(n, \lfloor cn/2 \rfloor)$; we also need a similar calculation in the case where $G \sim \mathbb{G}^*(n, M)$ with $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$. This is the only place where the distinction between our two random graph models actually has an impact.

Lemma 4.3. *Fix $c > 0$.*

- *If $G \sim \mathbb{G}^*(n, \lfloor cn/2 \rfloor)$, then*
$$\left(\frac{X_1^n(0) - n\chi_1(0)}{\sqrt{n}}, \frac{X_2^n(0) - n\chi_2(0)}{\sqrt{n}}, \frac{X_3^n(0) - n\chi_3(0)}{\sqrt{n}}, \frac{X_4^n(0) - n\chi_4(0)}{\sqrt{n}}\right)$$
converges in distribution to a multivariate Gaussian distribution with mean zero and covariance matrix
$$\Sigma = \begin{pmatrix} c^2e^{-2c} + ce^{-c} - ce^{-2c} - c^3e^{-2c} & -ce^{-c} + ce^{-2c} + c^3e^{-2c} & 0 & 0 \\ -ce^{-c} + ce^{-2c} + c^3e^{-2c} & (e^{-c} + ce^{-c})(1 - e^{-c} - ce^{-c}) - c^3e^{-2c} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$
- *If $G \sim \mathbb{G}^*(n, M)$ with $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$ then*
$$\left(\frac{X_1^n(0) - n\chi_1(0)}{\sqrt{n}}, \frac{X_2^n(0) - n\chi_2(0)}{\sqrt{n}}, \frac{X_3^n(0) - n\chi_3(0)}{\sqrt{n}}, \frac{X_4^n(0) - n\chi_4(0)}{\sqrt{n}}\right)$$
converges in distribution to a multivariate Gaussian distribution with mean zero and covariance matrix
$$\Sigma = \begin{pmatrix} (c^3 - 3c^2 + c)e^{-2c} + ce^{-c} & (-c^3 + 2c^2 + c)e^{-2c} - ce^{-c} & (c - c^2)e^{-c} & 0 \\ (-c^3 + 2c^2 + c)e^{-2c} - ce^{-c} & (c^3 - c^2 - 2c - 1)e^{-2c} + (c+1)e^{-c} & c^2e^{-c} & 0 \\ (c - c^2)e^{-c} & c^2e^{-c} & c/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We remark in order to prove the central limit theorems in Theorems 1.2 and 1.4 we need to know that initial fluctuations are asymptotically Gaussian, but if we are not interested in knowing the asymptotic variance of the rank or matching number it is not necessary to know the actual value of Σ.

Proof of Lemma 4.3. Since the $\mathbb{G}^*(n, \lfloor cn/2 \rfloor)$ case was already considered in [23, Theorem 39], we just consider the case where $G \sim \mathbb{G}^*(n, M)$ with $M \sim \mathrm{Bin}(\binom{n}{2}, c/n)$. By Theorem 2.3, it actually suffices to consider the number X_1 of degree-1 vertices, the number X_2 of vertices of degree at least 2, and the number X_3 of edges, in an n-vertex Erdős–Rényi random graph with edge probability c/n.

Let $X^{(d)}$ be the number of vertices with degree d in such a random graph. So, writing $a_1^{(d)} = \mathbb{1}_{d=1}$, $a_2^{(d)} = \mathbb{1}_{d\geq 2}$ and $a_3^{(d)} = d/2$, we have $X_j = \sum_{d=0}^{\infty} a_j^{(d)} X^{(d)}$ for $j \in \{1, 2, 3\}$. In [18, Theorem 4.1], Janson finds the asymptotic joint distribution of $(X^{(d)})_{d=0}^{\infty}$: namely,
$$\left(\frac{X^{(d)} - \mathbb{E}X^{(d)}}{\sqrt{n}}\right)_{d=0}^{\infty} \xrightarrow{d} (U^{(d)})_{d=0}^{\infty},$$
where the $U^{(d)}$ are jointly Gaussian with
$$\mathrm{Cov}[U^{(d)}, U^{(d')}] = \pi(d)\pi(d')\left(\frac{(d-c)(d'-c)}{c} - 1\right) + \mathbb{1}_{d=d'}\pi(d),$$
where $\pi(d) = e^{-c}c^d/d!$. As discussed in [18, Theorem 4.1], this convergence in distribution behaves well with respect to (potentially infinite) linear combinations of the $U^{(d)}$, as long as the coefficients do not grow super-exponentially fast: specifically, letting $U_j = \sum_{d=0}^{\infty} a_j^{(d)} U^{(d)}$, we have
$$\left(\frac{X_j - \mathbb{E}X_j}{\sqrt{n}}\right)_{j\in\{1,2,3\}} \xrightarrow{d} (U_1, U_2, U_3),$$
with $\mathrm{Cov}[U_j, U_{j'}] = \sum_{d=0}^{\infty} \sum_{d'=0}^{\infty} a_j^{(d)} a_{j'}^{(d')} \mathrm{Cov}[U^{(d)}, U^{(d')}]$. The desired result then follows from a routine (if tedious) calculation[15]. □

[15]Accompanying the arXiv version of this paper, we include a Mathematica script that performs this calculation.

4.6. **Approximation by a Gaussian process.** Now, using all the estimates we have obtained so far, we can apply the general framework described in Section 4.1, to show that the normalised process

$$\left(\frac{X^n(ns)-n\chi(s)}{\sqrt{n}}\right)_{s\geq 0}$$

is well-approximated by a certain Gaussian process V (defined in terms of the initial covariance matrix Σ in Lemma 4.3 and the approximate transition rates $\beta_l(x)$, via the drift function F).

Consequently, we can show that

$$\frac{X^n(\tau_\delta^n)-n\chi(s_\delta)}{\sqrt{n}}$$

converges to a certain multivariate Gaussian distribution, which proves Theorem 2.6.

In more detail: as in Section 4.1, define the matrix $\partial F(x)\in\mathbb{R}^{4\times 4}$ by $(\partial F(x))_{i,j}=\partial F_i(x)/\partial x_j$. Independently for each $l\in\mathbb{Z}^4$ let $W_l\colon[0,\infty)\to\mathbb{R}$ be a standard Brownian motion, and define the Gaussian processes

$$\mathcal{W}(s)=\sum_{l\in\mathbb{Z}^4}W_l\left(\int_0^s\beta_l(\chi(q))dq\right)l.$$

Let

$$\Phi\colon\{(s,u):0\leq u<s^*,u\leq s<s^*\}\to\mathbb{R}^{4\times 4}$$

be the unique matrix solution to the system of differential equations

$$\frac{\partial}{\partial s}\Phi(s,u)=\partial F(\chi(s))\Phi(s,u),$$

with boundary conditions $\Phi(u,u)=I$ for all $0\leq u<s^*$ (here I is the 4×4 identity matrix).

Let $V(0)$ be a 4-variate Gaussian random variable with covariance matrix Σ as in Lemma 4.3, and define the Gaussian process

$$V(s)=\Phi(s,0)V(0)+\int_0^s\Phi(s,u)\,d\mathcal{W}(u). \tag{4.8}$$

As (informally) described in Section 4.1, one can show that for any $s<s^*$ (not depending on n) we have the convergence in distribution of processes

$$\left(\frac{X^n(ns)-n\chi(s)}{\sqrt{n}}\right)_{0\leq s<t}\xrightarrow{d}(V(s))_{0\leq s<t}$$

and (assuming $c\leq e$) at the stopping time τ_δ^n, we have convergence in distribution

$$\frac{X^n(\tau_\delta^n)-n\chi(s_\delta)}{\sqrt{n}}\xrightarrow{d}V(s_\delta)-\frac{V_3(s_\delta)}{F_3(\chi(s_\delta))}F(\chi(s_\delta)), \tag{4.9}$$

which yields Theorem 2.6. To orient the reader relative to the writeup in [23]: Equation (4.9) is essentially proved in [23, Theorem 45] (with s^* in place of s_δ, in the regime $c>e$)[16].

5. Estimating fluctuations

In this section we prove Lemmas 2.8 and 2.9 via the machinery discussed in Section 4. Throughout this section, we again fix a constant $c\leq e$ (implicit constants in asymptotic notation are again allowed to depend on c).

For Lemma 2.8 we need to study the number of steps I_δ until there are at most δn edges remaining. In the notation of Section 4, this number of steps is precisely $X_4^n(\tau_\delta^n)$.

For Lemma 2.9, we need to study certain degree statistics. These can be studied in terms of the quantities $X_1^n(\tau_\delta^n),X_2^n(\tau_\delta^n),X_3^n(\tau_\delta^n)$ from Section 4 (namely, the number of leaves, the number of vertices of degree at least 2, and the number of edges, at the first point where there are at most δn edges remaining). All of these statistics are small when δ is small (they measure quantities of the remaining graph very close to the end of the process), and we expect them to have small fluctuations.

So, for both Lemmas 2.8 and 2.9, the main challenge is to estimate various parameters of the limiting distribution of $X^n(\tau_\delta^n)$ described in (4.9). Specifically, we will prove the following lemma.

[16]Notice that s_δ,τ_δ^n are defined in terms of χ_3,X_3^n, whereas in [23], s^*,τ^n are defined in terms of χ_1,X_1^n. This explains the role of V_3,F_3 in (4.9) versus V_1,F_1 in [23].

Lemma 5.1. *Let χ be the fluid limit approximation defined in Section 4.4, let s_δ be the fluid limit prediction for τ_δ^n, and let Σ_δ be the covariance matrix of the limiting random vector in (4.9). Let $z(x)$ be defined as in Section 4.3.*

(1) *$\chi_1(s_\delta) = O(\delta)$ and $\chi_j(s_\delta) = \Theta(\delta)$ for $j \in \{2, 3\}$.*
(2) *$z(\chi(s_\delta))) = \Theta(\sqrt{\delta})$.*
(3) *$\lim_{\delta\to 0}(\Sigma_\delta)_{j,j} = 0$ for $j \in \{1, 2, 3\}$.*
(4) *$\liminf_{\delta\to 0}(\Sigma_\delta)_{4,4} > 0$.*

Lemma 5.1(4) directly implies Lemma 2.8, because the quantity σ_δ^2 in Lemma 2.8 is asymptotic to $(\Sigma_\delta)_{4,4}n$. We will deduce Lemma 2.9 from Lemma 5.1(1–3) at the end of this section (in Section 5.5) using some estimates of Aronson, Frieze, and Pittel [1] (we need similar considerations as we informally discussed at the start of Section 4.3). Meanwhile, most of this section will be spent proving Lemma 5.1.

Lemma 5.1(1–2) follow from quite routine computations using the explicit formulas for the fluid limit approximation. The real challenge is to prove (3) and (4) (bounding certain variances). At a very high level, the goal is to prove that near the end of the process, there is very strong correlation between the fluctuations of X_1^n, X_2^n, X_3^n (so if we stop at a time τ_δ^n which fixes the value of X_3^n, then we essentially eliminate the fluctuation in X_1^n, X_2^n), and to prove that there is only weak correlation between the fluctuations of these three statistics and of X_4^n.

To be a bit more specific, the machinery outlined in Section 4 gives us an approximation of our process X^n in terms of a Gaussian process V. Our goal is to prove that (in the $\delta \to 0$ limit) the covariance matrix $\Sigma(s_\delta)$ of $V(s_\delta)$ has rank 2, and its first three rows and columns comprise a rank-1 submatrix. That is to say, $V_1(s_\delta), V_2(s_\delta), V_3(s_\delta)$ are essentially multiples of each other, while $V_4(s_\delta)$ has plenty of additional variance. The correction in (4.9) for the stopping time τ_δ^n then yields the desired conclusions.

Recalling the definitions in Section 4, the covariance matrix of $V(s)$ can be written in terms of (an integral involving) the matrix-valued correlation function Φ (see (4.8)), so our task is to obtain a good understanding of Φ. For example, for (3), we want to show that near the end of the process the first three rows and columns of Φ form a matrix that is nearly rank-1. While Φ is defined in terms of a system of nonlinear partial differential equations (see (5.8)), near the end of the process this system is quite well-approximated in terms of an (inhomogeneously time-dilated) system of *linear* differential equations (basically, we just need to know the limiting value of ∂F as we approach the end of the process). We can then obtain our desired conclusions by studying the limiting eigensystem of ∂F. In particular, for (3), negative eigenvalues in this eigensystem give us exponential decay in all directions but one.

5.1. **Estimating the fluid limit.** First we prove Lemma 5.1(1–2) and some estimates that will be used in the proofs of Lemma 5.1(3–4). For all of these, it is convenient to re-parameterise the fluid limit approximation χ in terms of $z = z(\chi(s))$. Indeed, as observed in [1, Lemma 8], the fluid limit approximation can be re-expressed as

$$\begin{aligned}
\chi_1(s) &= \frac{1}{c}(z^2 - zc\beta(z)(1 - e^{-z})),\\
\chi_2(s) &= (1 - (1 + z)e^{-z})\beta(z),\\
\chi_3(s) &= \frac{1}{2c}z^2,\\
\chi_4(s) &= s = \frac{1}{c}\left(c(1 - \beta(z)) - \frac{1}{2}\log(\beta(z))^2\right).
\end{aligned}$$

(where, as in Section 4.4, β is defined implicitly by $\beta(u)e^{c\beta(u)} = e^u$). Also, we have $z \to 0$ as $s \to s^*$ (this is only true because we are assuming $c \le e$; if $c > e$ then $z \to z^*$ for some explicit $z^* > 0$ computed in [1]).

Recall the Lambert W function, satisfying $W(t)e^{W(t)} = t$, and note that we can rewrite $\beta(z)$ as $W(ce^z)/c$. Note also that $W(t) > 0$ for $t > 0$, and $W(e) = 1$.

A direct computation with the series expansion for W (see for example [11]) shows that

$$\begin{aligned}
\chi_1(s) &= \frac{1 - W(c)}{c}z^2 + O(z^3),\\
\chi_2(s) &= \frac{W(c)}{2c}z^2 + O(z^3),\\
\chi_3(s) &= \frac{1}{2c}z^2,
\end{aligned}$$

$$\chi_4(s) = s = \frac{2c - 2W(c) - W(c)^2}{2c} - \frac{1}{2c(W(c)+1)} z^2 + O(z^3). \tag{5.1}$$

Lemma 5.1(1) and (2) immediately follow: from the definition of s_δ we have that $\chi_3(s_\delta) = \delta = \frac{1}{2c} z^2$ yielding (2); plugging in this value of z yields (1). This computation also yields

$$s^* = \frac{2c - 2W(c) - W(c)^2}{2c}. \tag{5.2}$$

We next collect several estimates on the drift function and its partial derivatives near $z = 0$ that follow from the above equations. The necessary computations, while routine, are a bit tedious; accompanying the arXiv version of this paper, we include a Mathematica script that performs these calculations.

For the rest of this section, we generalise the notation "$O(f)$" to denote any matrix or vector whose entries are all of the form $O(f)$ (so we can write matrix equations with error terms).

We begin by estimating the drift function near its limit at s^*. The following estimate can be obtained by plugging the series expansion of $\chi(s)$ in (5.1) into the formula for F in (4.4):

Fact 5.2.

$$F(\chi(s)) = -(1 + W(c)) \begin{pmatrix} 2 - 2W(c) \\ W(c) \\ 1 \\ -1 - W(c) \end{pmatrix} + O(z).$$

Let $\hat{F} = (F_1, F_2, F_3)$ be the first 3 coordinates of the drift function, and let

$$v_0 = (2 - 2W(c), W(c), 1)^\intercal \tag{5.3}$$

be the direction of the first 3 coordinates of the drift function near time $\chi(s^*)$. We want to show that at time s_δ, (X_1^n, X_2^n, X_3^n) is approximately parallel to v_0, by showing that any fluctuations in (X_1^n, X_2^n, X_3^n) in directions orthogonal to v_0 are suppressed. To do this, we will need to understand the partial derivatives of $\hat{F}$, which govern how the drift function changes due to fluctuations about $\hat{F}(\chi(s))$. Let $\partial\hat{F}(x)$ be the 3×3 matrix defined by $(\partial\hat{F}(x))_{i,j} = \partial\hat{F}_i(x)/\partial x_j$. The following lemma shows that for s near s^*, $\partial\hat{F}(\chi(s))$ is approximately a matrix with one zero eigenvalue, corresponding to the eigenvector v_0, and two negative eigenvalues. This ensures that near s^*, if (in addition to some fluctuation γv_0 parallel to v_0) there is a large fluctuation v orthogonal to v_0, there will be a tendency for that orthogonal fluctuation to be suppressed, because $v^\intercal \hat{F}(X(s)) \approx v^\intercal \hat{F}(\chi(s) + \gamma v_0 + v) \approx v^\intercal \partial\hat{F}(\chi(s)) v < 0$.

Lemma 5.3.

$$\partial\hat{F}(\chi(s)) = \frac{1}{z^2} Q D Q^{-1} + O\left(\frac{1}{z}\right),$$

where

$$D = c(1 + W(c)) \begin{pmatrix} 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad Q = \begin{pmatrix} 2 - 2W(c) & \frac{1}{2}(W(c) - 1)(3W(c) - 1) & -4W(c) \\ W(c) & \frac{1}{2}(-2W(c) - 1)(W(c) - 1) & 2W(c) + 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Lemma 5.3 may be obtained by explicitly computing partial derivatives of the formulas for F in Section 4.3, using implicit differentiation to compute $\partial z(x)/\partial x_i$, then using the formulas in (5.1) for $\chi(s)$ in terms of z, and then computing a series expansion.

In Section 5.3, we will use Fact 5.2 and Lemma 5.3 (plus some elementary estimates) to study the correlation function $\Phi(s, u)$ (for s near s^*). We will also use the following two facts, which are easily verified from the series expansions in (5.1).

Fact 5.4. *There exists a constant $c_1 > 0$ such that for any $c_2 > 0$ small enough and any $s^* - c_1 \le s \le s^* - c_2$, we have $z(\chi(s)), \chi_1(s), \chi_2(s) = \Omega(1)$.*

(Here the constants in asymptotic notation are allowed to depend on c_1, c_2.)

Proof. If s is bounded away from s^*, then from the formulas in Section 4.4, $\vartheta(s)$ is bounded away from 0, and thus so are $\chi_1(s), \chi_2(s)$ and $\chi_3(s)$. Thus from (5.1), $z(\chi(s))$ is also bounded away from 0. □

(Although it will not be necessary for us, we remark that one can directly check using the formulas in Section 4.4 or Section 5.1 that $z(\chi(s)), \chi_1(s), \chi_2(s) = \Omega(1)$ whenever $s - s^* = \Omega(1)$, i.e., the above statement holds for *any* constants $0 < c_2 \le c_1 \le s^*$.)

Fact 5.5. $s^* - s = \Theta(z^2) = \Theta(\chi_3(s))$. *In particular,* $s^* - s_\delta = \Theta(z(\chi(s_\delta))^2) = \Theta(\delta)$.

5.2. **Formulas for the limiting variance.** In this subsection, we show how to reduce the task of proving Lemma 5.1(3–4), concerning the covariance matrix of $X^n(\tau_\delta^n)$, to certain analytic estimates on the functions Φ and β_ℓ (which we will prove in the next section).

Recall the definitions of V and $\mathcal{W}$ from Section 4.6 (in terms of functions Φ, β_l, F and an initial covariance matrix Σ that depends on which of the two models of random graphs in Theorem 1.2 we are considering). Let $\Sigma(s)$ be the covariance matrix of $V(s)$, and compute from (4.8)

$$\Sigma(s) = \Phi(s,0)\,\Sigma\,\Phi(s,0)^\intercal + \int_0^s \sum_{l\in\mathbb{Z}^4} \beta_l(\chi(u))\,(\Phi(s,u)\,l\,l^\intercal\,\Phi(s,u)^\intercal)\,du. \tag{5.4}$$

Remark 5.6. For the reader not familiar with stochastic calculus: it may be helpful to view a Brownian motion as the limit of a discrete-time random walk, so our Itô integral can be approximated as a sum of independent increments (these increments have different sizes given by Φ, and the rate they occur at is given by the β_l). The total variance is then the sum of variances of these increments; taking a limit gives the above integral.

Recalling (4.9) (with the correction for the stopping time τ_δ^n), we have

$$\Sigma_\delta = P_\delta\,\Phi(s_\delta,u)\,\Sigma\,\Phi(s_\delta,0)^\intercal\,P_\delta^\intercal + \int_0^{s_\delta} \sum_{l\in\mathbb{Z}^4} \beta_l(\chi(u))\,(P_\delta\,\Phi(s_\delta,u)\,l\,l^\intercal\,\Phi(s_\delta,u)^\intercal P_\delta^\intercal)\,du, \tag{5.5}$$

where

$$P_\delta = I - \frac{1}{F_3(\chi(s_\delta))}\begin{pmatrix} 0 & 0 & F_1(\chi(s_\delta)) & 0\\ 0 & 0 & F_2(\chi(s_\delta)) & 0\\ 0 & 0 & F_3(\chi(s_\delta)) & 0\\ 0 & 0 & F_4(\chi(s_\delta)) & 0\end{pmatrix} = \begin{pmatrix} 1 & 0 & -F_1(\chi(s_\delta))/F_3(\chi(s_\delta)) & 0\\ 0 & 1 & -F_2(\chi(s_\delta))/F_3(\chi(s_\delta)) & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & -F_4(\chi(s_\delta))/F_3(\chi(s_\delta)) & 1\end{pmatrix}.$$

We can now reduce Lemma 5.1(3–4) to some more explicit estimates on the functions χ, β_l, Φ. Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 0 & 1\end{pmatrix},$$

so $A\Sigma_\delta A^\intercal$ contains the first three rows and columns of Σ_δ and $B\Sigma_\delta B^\intercal$ contains the $(4,4)$-entry of Σ_δ. This means that Lemma 5.1(3–4) is tantamount to the claims that $\lim_{\delta\to 0}\|A\Sigma_\delta A^\intercal\| = 0$ and $\liminf_{\delta\to 0}\|B\Sigma_\delta B^\intercal\| > 0$, for any matrix norm $\|\cdot\|$ (all matrix norms are equivalent).

Recalling that the Frobenius norm[17] $\|\cdot\|_{\mathrm{F}}$ is subadditive and submultiplicative, from (5.5) we estimate

$$\|A\Sigma_\delta A^\intercal\|_{\mathrm{F}} \le \|\Sigma\|_{\mathrm{F}}\|AP_\delta\,\Phi(s_\delta,0)\|_{\mathrm{F}}^2 + \int_0^{s_\delta} \|AP_\delta\,\Phi(s_\delta,u)\|_{\mathrm{F}}^2 \sum_{l\in\mathbb{Z}^4} \beta_l(\chi(u))\,\|l\|_2^2\,du. \tag{5.6}$$

Roughly speaking, to prove Lemma 5.1(3) it now suffices to prove that $\sum_{l\in\mathbb{Z}^4}\beta_l(\chi(u))\|l\|_2^2 = O(1)$ for all $u\in[0,s_\delta]$, and that $\|AP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}}$ is tiny (e.g., decays with δ) unless u is very close to s_δ.

For Lemma 5.1(4), consider any vectors $\mathcal{L} = \{l_1,l_2,l_3,l_4\}$ in $\mathbb{Z}^4$. Let $L\in\mathbb{R}^{n\times n}$ be the matrix with columns l_1,l_2,l_3,l_4, and let σ be the least singular value of L. Then, the single entry of the matrix $\sum_{l\in\mathcal{L}}(BP_\delta\,\Phi(s_\delta,u)\,l\,l^\intercal\,\Phi(s_\delta,u)^\intercal P_\delta^\intercal B^\intercal)$ is $\|BP_\delta\Phi(s_\delta,u)L\|_{\mathrm{F}}^2$, which is at least $\sigma^2\|BP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}}^2$. So, we have

$$\|(B\Sigma_\delta B^\intercal)\|_{\mathrm{F}} \ge \sigma^2\int_0^{s_\delta}\|BP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}}^2 \inf_{l\in\mathcal{L}}\beta_l(\chi(u))\,du. \tag{5.7}$$

Roughly speaking, to prove Lemma 5.1(4) it suffices to prove that, for some specific basis $\mathcal{L}$ of the vector space $\mathbb{R}^4$ (not depending on δ) and for all u in some interval of length $\Omega(1)$, both $\inf_{l\in\mathcal{L}}\beta_l(\chi(u))$ and $\|BP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}}^2$ are of the form $\Omega(1)$.

Remark 5.7. It is worth remarking that the quantities in this section can be used to give asymptotic formulas for the variance of $\alpha(G)$ (in the settings of Theorems 1.2 and 1.4). Indeed, our proof of Theorems 1.2 and 1.4 (in the case $c\le e$) shows that $\alpha(G)$ is asymptotically Gaussian with variance asymptotic to $(\lim_{\delta\to 0}(\Sigma_\delta)_{4,4})n$. The calculations in this section (in particular, the computation of $\lim_{\delta\to 0}P_\delta$ which we will see in Lemma 5.11) can be used to show that $\lim_{\delta\to 0}(\Sigma_\delta)_{4,4}$ is precisely the single entry of the matrix

$$BP\,\Phi(s^*,u)\,\Sigma\,\Phi(s^*,0)^\intercal\,P^\intercal B^\intercal + \int_0^{s^*}\sum_{l\in\mathbb{Z}^4}\beta_l(\chi(u))\,(BP\,\Phi(s^*,u)\,l\,l^\intercal\,\Phi(s^*,u)^\intercal P^\intercal B^\intercal)\,du,$$

[17]The *Frobenius norm* of a matrix M is the square root of the sum of squares of entries of M.

where

$$P = \begin{pmatrix} 1 & 0 & 2W(c)-2 & 0 \\ 0 & 1 & -W(c) & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1+W(c) & 1 \end{pmatrix}.$$

It is not clear whether the asymptotic variance of $\alpha(G)$ can be expressed more explicitly. (The situation for $c > e$, treated in [23], is essentially the same.)

5.3. **Estimating the correlation function.** In this subsection we study Φ via its defining system of differential equations. Roughly speaking, our goal is to show that $\Phi(s_\delta, u)$ is approximately a rank-2 matrix, whose first 3 rows and columns approximately comprise a rank-1 submatrix spanned by v_0 (as defined in (5.3)). From this, we will obtain estimates on $AP_\delta\Phi(s_\delta, u)$ and $BP_\delta\Phi(s_\delta, u)$ as foreshadowed in the previous subsection.

Recall from Section 4.6 that Φ is the solution to the system of differential equations

$$\frac{\partial}{\partial s}\Phi(s,u) = \partial F(\chi(s))\Phi(s,u), \tag{5.8}$$

with boundary conditions $\Phi(u,u) = I$ for each u, where $\partial F(x)$ is the matrix defined by $(\partial F(x))_{i,j} = \partial F_i(x)/\partial x_j$ (and F is the drift function from Section 4.3). Note that $\partial F_4/\partial x_j \equiv 0$ and $\partial F_j/\partial x_4 \equiv 0$ for all $j \in \{1,2,3,4\}$, so we can write

$$\Phi(s,u) = \begin{pmatrix} \hat\Phi(s,u) & 0 \\ 0 & 1 \end{pmatrix}, \qquad \frac{\partial}{\partial s}\hat\Phi(s,u) = \partial\hat F(\chi(s))\hat\Phi(s,u), \qquad \hat\Phi(u,u) = \hat I$$

where $\hat\Phi(s,u) = A\Phi(s,u)A^\intercal$ and $\partial\hat F(x) = A\partial F(x)A^\intercal$ contain just the first three rows and columns of $\Phi(s,u)$ and $\partial F(x)$, respectively, and $\hat I \in \mathbb{R}^{3\times 3}$ is the 3×3 identity matrix. Note that there is never any interaction between different u: we can think of u as a parameter indexing a family of differential equations, and s as the independent variable in each differential equation (ranging from u to s^*). One could write $\hat\Phi_u(s)$ instead of $\hat\Phi(s,u)$ to emphasise this.

Also note that there is no interaction between the three columns of $\hat\Phi(s,u)$: each of the columns $g_u(s)$ of $\hat\Phi(s,u)$ is separately a solution to the system of differential equations $\partial g(s)/\partial s = \partial\hat F(\chi(s))g(s)$ (though, the different columns have different initial conditions). In this section we show that in the limit $s \to s^*$, the direction of $g(s)$ does not actually depend on the initial conditions (i.e., the columns of $\hat\Phi(s,u)$ are roughly proportional to each other). We formalize this in Lemma 5.8 below, which states that for s near s^*, $g_u(s)$ is nearly proportional to $(2-2W(c), W(c), 1)^\intercal$ for all u.

Roughly speaking, the reason this holds is that $\partial\hat F(\chi(s^*))$ has two negative eigenvalues and one zero eigenvalue (with corresponding eigenvector $v_0 := (2-2W(c), W(c), 1)^\intercal$). So, as our differential equation evolves near time s^*, the mass of g_u diminishes in all directions except the direction of v_0. (The eventual mass in direction v_0 depends on the initial conditions, defined in terms of u and the particular column of $\hat\Phi(s,u)$ that we're interested in.)

Lemma 5.8. *For some $u \in [0, s^*]$, let $g\colon [u, s^*] \to \mathbb{R}^3$ be a solution to the system of differential equations*

$$\frac{d}{ds}g(s) = \partial\hat F(\chi(s))g(s)$$

satisfying some initial conditions $g(u)$ with $\|g(u)\|_2 \le 1$. Then, for all $s \in [u, s^*]$ we have

$$g(s) = C\begin{pmatrix} 2-2W(c) \\ W(c) \\ 1 \end{pmatrix} + O\left(\left(\frac{s^*-s}{s^*-u}\right)^{1/4}\right)$$

for some $C \in \mathbb{R}$ (depending on s, u).

(Recall that we have generalised the notation "$O(f)$" so that it may describe a *vector* whose entries are of the form $O(f)$.)

Proof. As in Section 5.1, we work largely in terms of the reparameterisation $z = z(\chi(s))$. First, we recall from Lemma 5.3 the limiting behaviour of $\partial\hat F$ around $z = 0$:

$$\partial\hat F(\chi(s)) = \frac{1}{z^2}QDQ^{-1} + O\left(\frac{1}{z}\right),$$

where

$$D = c(1+W(c))\begin{pmatrix} 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad Q = \begin{pmatrix} 2-2W(c) & \frac{1}{2}(W(c)-1)(3W(c)-1) & -4W(c) \\ W(c) & \frac{1}{2}(-2W(c)-1)(W(c)-1) & 2W(c)+1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Denoting operator norm by $\|\cdot\|_{\mathrm{op}}$, observe that $\|Q\|_{\mathrm{op}}, \|Q^{-1}\|_{\mathrm{op}} = O(1)$ (this amounts to the fact that $\det Q \neq 0$). So, $h = Q^{-1}g$ is a solution to a system of differential equations

$$\frac{d}{ds}h(s) = \left(\frac{1}{z^2}D + O(1/z)\right)h(s)$$

for some initial conditions $h(0)$ satisfying $\|h(u)\|_2 = O(1)$.

Next, we reparameterise the time variable s to obtain a system of differential equations that is (approximately) linear. Namely, we first let $r = s^* - s$, so $ds/dr = -1$ and $1/r = 2c(W(c)+1)/z^2 + O(1/z)$ by (5.1) and (5.2). This means that

$$\frac{d}{dr}h(s) = \left(\frac{-1}{2cr(W(c)+1)}D + O(1/\sqrt{r})\right)h(s)$$

(recall Fact 5.5). Then, we let $q = -\log r$, so $dr/dq = -r$ and we can write

$$\frac{d}{dq}h(s) = \left(\frac{1}{2c(W(c)+1)}D + E(s)\right)h(s),$$

for some matrix-valued function E whose entries are at most $O(\sqrt{r}) = O(e^{-q/2})$. The time interval from $s = u$ to $s = s^*$ corresponds to the interval from $r = s^* - u$ to $r = 0$, which corresponds to the interval from $q = q_u := -\log(s^* - u)$ to $q = \infty$.

Now, we are finally ready to study the evolution of our system of differential equations. First, we prove that $h(s)$ does not blow up as $s \to s^*$.

Claim 5.9. $\sup_{s\in[u,s^*]} \|h(s)\|_2 = O(1)$.

Proof. Let $\Lambda = (1/(2c(W(c)+1)))D$ (which is a diagonal matrix with diagonal entries $(\lambda_1, \lambda_2, \lambda_3) = (0, -3/2, -1)$). We have

$$\begin{aligned}\frac{d}{dq}\|h(s)\|_2^2 = \sum_{j=1}^{3} 2h_j(s)\frac{d}{dq}h_j(s) &= \sum_{j=1}^{3} 2h_j(s)\big(\Lambda h(s) + E(s)h(s)\big)_j \\ &\le \sum_{j=1}^{3}\big(2\lambda_j h_j(s)^2 + 2h(s)_j\|E(s)\|_{\mathrm{op}}\|h(s)\|_2\big) \\ &\le 2\|E(s)\|_{\mathrm{op}}\|h(s)\|_1\|h(s)\|_2 = O(e^{-q/2}\|h(s)\|_2^2).\end{aligned}$$

(For the final line, we recall that each λ_j is nonpositive.) We can rewrite this inequality as

$$\frac{d}{dq}\log\|h(s)\|_2^2 = O(e^{-q/2}).$$

Then, for any s, integration yields

$$\log\|h(s)\|_2^2 \le \log\|h(u)\|_2^2 + \int_{q_u}^{q} O(e^{-p/2})\,dp = O(1).$$ □

Next, using Claim 5.9, we prove that $h_2(s)$ and $h_3(s)$ decay as $s \to s^*$.

Claim 5.10. *For $j \in \{2,3\}$ we have $h_j(s)^2 = O(e^{(q_u - q)/2})$.*

Proof. Using Claim 5.9 and similar calculations as above, for $j \in \{2,3\}$ we have

$$\begin{aligned}\frac{d}{dq}h_j(s)^2 = 2h_j(s)\frac{d}{dq}h_j(s) &\le 2\lambda_j h_j(s)^2 + 2\|E(s)\|_{\mathrm{op}}\|h(s)\|_1\|h(s)\|_2 \\ &\le -2h_j(s)^2 + O(e^{-q/2}\|h(s)\|_2^2) \le -2h_j(s)^2 + O(e^{-q/2}).\end{aligned}$$

(For the last line we used that $\lambda_2, \lambda_3 \le -1$.) We can rewrite this inequality as

$$\frac{d}{dq}(e^{2q}h_j(s)^2) \le e^{2q}\cdot O(e^{-q/2}) = O(e^{3q/2}).$$

Integration then yields

$$e^{2q}h_j(s)^2 - e^{2q_u}h_j(u)^2 \le \int_{q_u}^{q} O(e^{3p/2})\,dp = O(e^{3q/2}),$$

which implies

$$h_j(s)^2 \le O(e^{2(q_u-q)} + e^{-q/2}) = O(e^{(q_u-q)/2}). \qquad \square$$

Together Claims 5.9 and 5.10 imply that

$$h(s) = Q^{-1}g(s) = C_0\begin{pmatrix}1\\0\\0\end{pmatrix} + O(e^{(q_u-q)/4}) = C_0\begin{pmatrix}1\\0\\0\end{pmatrix} + O\left(\left(\frac{s^*-s}{s^*-u}\right)^{1/4}\right)$$

for some $C_0 \in \mathbb{R}$ with $|C_0| = O(1)$. Multiplying by Q yields the desired result. $\square$

We can now deduce some estimates on $P_\delta\Phi(s_\delta, u)$, suitable for application with (5.6) and (5.7).

Lemma 5.11. *Recall the definitions of* A, B, P_δ *from Section 5.2, and write* $f_\delta(u) = ((s^*-s_\delta)/(s^*-u))^{1/4}$.

(1) $\|AP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}} = O(f_\delta(u))$,
(2) $\|BP_\delta\Phi(s_\delta,u)\|_{\mathrm{F}} = \Omega(1) - O(f_\delta(u))$.

Proof. By Lemma 5.8, for each s, u there are real numbers $C_1, C_2, C_3 = O(1)$ such that

$$\Phi(s_\delta, u) = \begin{pmatrix} C_1(2-2W(c)) & C_2(2-2W(c)) & C_3(2-2W(c)) & 0\\ C_1W(c) & C_2W(c) & C_3W(c) & 0\\ C_1 & C_2 & C_3 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} + O(f_\delta(u)).$$

Recall from Fact 5.2 that we have

$$F(\chi(s)) = -(1+W(c))\begin{pmatrix}2-2W(c)\\ W(c)\\ 1\\ -1-W(c)\end{pmatrix} + O(z).$$

By Fact 5.5 we have $z(\chi(s_\delta)) = O(f_\delta(u))$ (for any u), so we deduce (using the definition of P_δ) that

$$P_\delta = \begin{pmatrix}1 & 0 & -2+2W(c) & 0\\ 0 & 1 & -W(c) & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1+W(c) & 1\end{pmatrix} + O(f_\delta(u)).$$

We then compute

$$\begin{aligned} AP_\delta\Phi(s_\delta,u) &= O(f_\delta(u)),\\ BP_\delta\Phi(s_\delta,u) &= \begin{pmatrix}C_1(1+W(c)) & C_2(1+W(c)) & C_3(1+W(c)) & 1\end{pmatrix} + O(f_\delta(u)).\end{aligned}$$

The desired results follow. $\square$

5.4. **Integrating the correlation function.** Now we combine Lemma 5.11 with some estimates on the transition rates $\beta_l(\chi(s))$, to complete the proofs of Lemma 5.1(3–4) via the strategy outlined in Section 5.2.

Lemma 5.12. *Define the basis*

$$\mathcal{L} = \left\{\begin{pmatrix}-2\\0\\-1\\1\end{pmatrix}, \begin{pmatrix}0\\-2\\-3\\1\end{pmatrix}, \begin{pmatrix}0\\-3\\-6\\1\end{pmatrix}, \begin{pmatrix}1\\-3\\-5\\1\end{pmatrix}\right\}$$

of $\mathbb{R}^4$. *For some* $c_1 > 0$ *and any small enough* $c_2 > 0$, *for all* $s^* - c_1 \le s \le s^* - c_2$, *we have* $\inf_{l\in\mathcal{L}} \beta_l(\chi(s)) = \Omega(1)$.

Proof. Recall the notation defined in Section 4.3. The corresponding choices of $(k_1, k_2, k_3) \in \mathcal{K}$ for the four vectors in $\mathcal{L}$ are $(1,0,0), (1,1,1), (2,2,2), (1,2,2)$, respectively. Using the formula for $\beta_l(x)$ in Section 4.3, for constants k_1, k_2, k_3, we have

$$\beta_l(x) = \Omega\left(\left(\frac{x_2}{x_3}\right)^{k_2+k_3+1}\right) \cdot \Omega\left(\left(\frac{z}{e^z - z - 1}\right)^{k_2+1}\right)\left(\frac{x_1}{2x_3}\right)^{k_1-1} \cdot \Omega\big(z^{k_1+2k_2+2k_3-1}\big).$$

Fact 5.4 then shows that for some $c_1 > 0$ and any small enough $c_2 > 0$, we have $z(\chi(s)), \chi_1(s), \chi_2(s) = \Omega(1)$, so $\beta_l(\chi(s)) = \Omega(1)$. □

Lemma 5.13. *For all $0 \le s \le s^*$ we have*

$$\sum_{l \in \mathbb{Z}^4} \|l\|_2^2 \beta_l(\chi(s)) = O(1).$$

Proof. Using the approximations in Section 5.1 and the formulas for $\beta_l(x)$ in Section 4.3, we can see that $\beta_{(-2,0,-1,1)}(\chi(s)) = O(1)$, and for $l = (-k_1 + k_2, -1 - k_2, -k_1 - k_2 - k_3, 1)$ with $(k_1, k_2, k_3) \in \mathcal{K}$,

$$\beta_l(x) = \frac{1}{(k_1-1)!k_2!k_3!} O\left(\frac{1}{z}\right)(O(z))^{k_1-1}(O(z))^{k_2}\left(O(z^2)\right)^{k_3} = \frac{1}{(k_1-1)!k_2!k_3!}(O(z))^{k_1+k_2+k_3-2}.$$

Indeed, we observe from (5.1) that $x_3 = \Omega(z^2)$, and $x_1, x_2 = O(z^2)$, and then use L'Hôpital's rule on the functions of z. Also, for such l we have $\|l\|_2^2 = O(k_1^2 + k_2^2 + k_3^2)$. So,

$$\sum_{l \in \mathbb{Z}^4} \|l\|_2^2 \beta_l(\chi(s)) = O(1) + \sum_{(k_1,k_2,k_3) \in \mathcal{K}} O\left(\frac{k_1^2 + k_2^2 + k_3^2}{(k_1-1)!k_2!k_3!}\right)(O(z))^{k_1+k_2+k_3-2} = O(1),$$

recalling that $k_1 + k_2 + k_3 \ge 2$ for all $(k_1, k_2, k_3) \in \mathcal{K}$. □

We are now ready to complete the proofs of Lemma 5.1(3–4).

Proof of Lemma 5.1(3). Note that $s^* - s_\delta = O(\delta)$ (by Fact 5.5). For $j \in \{1,2,3\}$, substituting Lemma 5.11(1) and Lemma 5.13 into (5.6) yields

$$\begin{aligned}
(\Sigma_\delta)_{j,j} \le \|A\Sigma_\delta A^\intercal\|_{\mathrm{F}} &\le O\Big(((s^* - s_\delta)^{1/4})^2\Big) + \int_0^{s_\delta} O\left(\left(\left(\frac{s^* - s_\delta}{s^* - u}\right)^{1/4}\right)^2\right) du \\
&= O(\delta^{1/2}) + O(\delta^{1/2}) \int_0^{s^* - O(\delta)} \frac{du}{(s^* - u)^{1/2}} \\
&= O(\delta^{1/2}) \cdot (1 + (s^*)^{1/2} - O(\delta)^{1/2}) = O(\delta^{1/2}),
\end{aligned}$$

which tends to zero as $\delta \to 0$. □

Proof of Lemma 5.1(4). Let c_1 and c_2 be the constants in Lemma 5.12, so when u is in the range between $s^* - c_1$ and $s^* - c_2$ we have $\inf_{l \in \mathcal{L}} \beta_l(\chi(s)) = \Omega(1)$.

Also, when u is in this range between $s^* - c_1$ and $s^* - c_2$, we have $f_\delta(u) = O(\delta^{1/4})$ (using Fact 5.5, and using that $c_2 = \Omega(1)$). So, for sufficiently small δ, for u in this range we have $\|BP_\delta \Phi(s_\delta, u)\|_{\mathrm{F}} = \Omega(1)$ by Lemma 5.11(2).

We can then directly substitute the above two estimates into (5.7). We obtain $(\Sigma_\delta)_{4,4} = \|B\Sigma_\delta B^\intercal\|_{\mathrm{F}} = \Omega(1)$ as desired. □

5.5. **Fluctuations of the degree sequence.** Now, having proved Lemma 5.1, it remains to deduce Lemma 2.9 from Lemma 5.1(1–3).

Recall that $X_1^n(\tau_\delta^n), X_2^n(\tau_\delta^n), X_3^n(\tau_\delta^n)$ measure the number of leaves, the number of vertices of degree at least 2, and the number of edges, at the first time τ_δ^n where there are at most δn edges remaining. Lemma 5.1(3) (together with the Gaussian approximation (4.9)) allows us to control the fluctuations of these statistics.

In the statement of Lemma 2.9, $X^{(d)}$ is the number of degree-d vertices at time τ_δ^n (so in particular $X^{(1)} = X_1^n(\tau_\delta^n)$ and $\sum_{d=2}^\infty X^{(d)} = X_2^n(\tau_\delta^n)$). To prove Lemma 2.9, we study the fluctuations of these degree statistics $X^{(d)}$ *conditional on* $X_1^n(\tau_\delta^n), X_2^n(\tau_\delta^n), X_3^n(\tau_\delta^n)$.

To this end, we use estimates of Aronson, Frieze, and Pittel [1], already mentioned informally at the start of Section 4.3. Specifically, the following lemma is a slight strengthening of [1, Lemma 5][18], and

[18] In [1, Lemma 5], the notation "v" is used instead of "k_2", and the notation "X_j" is used for the degree of a vertex j.

follows from the same proof (the estimate for $\Pr[\deg(v) = \deg(v') = d]$ below is just slightly less wasteful than in the statement of [1, Lemma 5]).

Lemma 5.14. *For some $k_1, k_2, k_3 \in \mathbb{N}$, let G be a random multigraph $\mathbb{G}^*(k_1 + k_2, k_3)$ conditioned on the event that the vertices in $V_1 := \{1, \dots, k_1\}$ have degree exactly 1, and the vertices in $V_2 := \{k_1 + 1, \dots, k_1 + k_2\}$ have degree at least 2. Let $f(z) = e^z - z - 1$ and let z be the unique solution to*

$$\frac{z(e^z - 1)}{f(z)} = \frac{2k_3 - k_1}{k_2},$$

(so z is precisely $z(k_1, k_2, k_3, 0)$ in the notation of Section 4.3).

(1) *Suppose that $k_2 z = \Omega(\log^2 n)$. Then for any distinct $v, v' \in V_2$ and any $2 \le d \le \log k_2$ we have*

$$\Pr[\deg(v) = d] = \frac{z^d}{d! f(z)} \left(1 + O\left(\frac{d^2 + 1}{k_2 z}\right)\right)$$

and

$$\Pr[\deg(v) = \deg(v') = d] = \left(\frac{z^d}{d! f(z)}\right)^2 \left(1 + O\left(\frac{d^2 + 1}{k_2 z}\right)\right).$$

(2) *For any $v \in V_2$ and $d \ge 2$, we have the cruder estimate*

$$\Pr[\deg(v) = d] = O\left((k_2 z)^{1/2} \frac{z^d}{d! f(z)}\right).$$

Note that $z^d/(d! f(z)) = \Pr[Q = d \,|\, Q \ge 2]$, for $Q \sim \mathrm{Poisson}(z)$ (i.e., it is a point probability for a truncated Poisson random variable). Very briefly, the proof strategy for Lemma 5.14 is as follows: one can show that the degree sequence of $\mathbb{G}^*(n, m)$ is precisely a sequence of independent Poisson random variables conditioned on their sum being exactly $2m$. So, the proof of Lemma 5.14 essentially comes down to careful estimates on point probabilities of sums of independent truncated Poisson random variables, using standard techniques for proving local limit theorems.

Proof of Lemma 2.9. Recalling the fluid limit approximations in Section 4.4, and the definition of $z(x)$ from Section 4.3, let $z_\delta = z(\chi(s_\delta))$, and let $Q \sim \mathrm{Poisson}(z_\delta)$. Let $\mu^{(1)} = \chi_1(s_\delta) n$ and $\mu^{(d)} = \chi_2(s_\delta) n \Pr[Q = d \,|\, Q \ge 2]$ for $2 \le d \le \log n$, and $\mu^{(d)} = 0$ for $d > \log n$. Then

$$\sum_d d\mu^{(d)} \le (\chi_1(s_\delta) + \chi_2(s_\delta)\, \mathbb{E}[Q \,|\, Q \ge 2]) n.$$

By the definition of $z(\chi(x))$, and using Lemma 5.1(1), we have

$$\mathbb{E}[Q \,|\, Q \ge 2] = O(1),$$

so recalling that $\chi_1(s_\delta), \chi_2(s_\delta) \to 0$ as $\delta \to 0$ (again by Lemma 5.1(1)), we have $\sum_d d\mu^{(d)} \le \varepsilon n$ for sufficiently small δ.

By Lemma 5.1(3) and Chebyshev's inequality, together with the Gaussian approximation (4.9), we see that with probability at least $1 - \varepsilon/2$ we have

$$\sum_{j \in \{1,2,3\}} |X_j^n(\tau_\delta^n) - \chi_j(s_\delta) n| \le h(\delta)\sqrt{n}. \tag{5.9}$$

for some $h(\delta)$ tending to zero as $\delta \to 0$. Recalling the definition $D = \sum_d d\,|X^{(d)} - \mu^{(d)}|$ from the statement of Lemma 2.9, our goal is now to show that $\mathbb{E}[D \,|\, X^n(\tau_\delta^n)]$ is small whenever $X^n(\tau_\delta^n)$ satisfies (5.9) (we will then finish the proof with Markov's inequality). We consider each $|X^{(d)} - \mu^{(d)}|$ separately.

First, note that $X^{(1)} = X_1^n(\tau_\delta^n)$, so when (5.9) holds we have

$$|X^{(1)} - \mu^{(1)}| \le h(\delta)\sqrt{n}. \tag{5.10}$$

For the cases where $d \ge 2$ we will apply Lemma 5.14: note that if we condition on the set V_1 of $X_1^n(\tau_\delta^n)$ degree-1 vertices at time τ_δ^n, and the set V_2 of $X_2^n(\tau_\delta^n)$ vertices of degree at least 2, then (up to relabelling vertices), the remaining graph at time τ_δ^n is distributed as $\mathbb{G}^*(X_1^n(\tau_\delta^n) + X_2^n(\tau_\delta^n), X_3^n(\tau_\delta^n))$. This simple fact appears explicitly as [1, Lemma 2].

In order to apply Lemma 5.14, we first need to show that when (5.9) holds, $Z^n(\tau_\delta^n)$ is well-approximated by its fluid limit approximation z_δ. Indeed, we compute

$$\frac{z(e^z - 1)}{f(z)} = 2 + z/3 + O(z^2). \tag{5.11}$$

For $X^n(\tau_\delta^n)$ satisfying (5.9), using that $\chi_2(s_\delta) = \Omega(\delta)$ (by Lemma 5.1(1)) we have

$$\frac{2X_3^n(\tau_\delta^n) - X_1^n(\tau_\delta^n)}{X_2^n(\tau_\delta^n)} = \frac{2\chi_3(s_\delta) - \chi_1(s_\delta)}{\chi_2(s_\delta)} + O\left(\frac{h(\delta)}{\delta\sqrt{n}}\right),$$

so by (5.11),

$$Z^n(\tau_\delta^n) = z(X^n(\tau_\delta^n)) = z_\delta + O\left(\frac{h(\delta)}{\delta\sqrt{n}}\right). \tag{5.12}$$

Now, we consider $|X^{(d)} - \mu^{(d)}|$ in the case $2 \le d \le \log n$. Condition on outcomes of V_1, V_2 satisfying (5.9), and in the resulting conditional probability space, let $\mathbb{1}_v$ be the indicator random variable for the event $\deg(v) = d$. By Lemma 5.14, for any distinct $v, v' \in V_2$ we have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_v] &= \frac{Z^n(\tau_\delta^n)^d}{d!f(Z^n(\tau_\delta^n))}\left(1 + O\left(\frac{d^2+1}{X_2^n(\tau_\delta^n)Z^n(\tau_\delta^n)}\right)\right)\\
&= \Pr[Q = d \,|\, Q \ge 2] + \frac{1}{d^{\Omega(d)}}O\left(\frac{h(\delta)}{\delta\sqrt{n}}\right),\\
\mathrm{Var}[\mathbb{1}_v] &\le \mathbb{E}[\mathbb{1}_v] \le \frac{1}{d^{\Omega(d)}}O(1),\\
\mathrm{Cov}[\mathbb{1}_v, \mathbb{1}_{v'}] &= \left(\frac{Z^n(\tau_\delta^n)^d}{d!f(Z^n(\tau_\delta^n))}\right)^2 O\left(\frac{d^2+1}{X_2^n(\tau_\delta^n)Z^n(\tau_\delta^n)}\right)\\
&= \frac{1}{d^{\Omega(d)}}O\left(\frac{1}{\delta\sqrt{\delta}n}\right).
\end{aligned}$$

For these estimates we have used (5.12), that $\chi_2(s_\delta) = \Theta(\delta)$ and $z_\delta = \Theta(\sqrt{\delta})$ (from Lemma 5.1(1–2)), that $f'(z) = O(1)$ for all $z \in \mathbb{R}$, and that $d! = d^{\Omega(d)}$ by Stirling's approximation.

So, for $X^n(\tau_\delta^n)$ satisfying (5.9), we have

$$\begin{aligned}
\mathbb{E}[X^{(d)} \,|\, X^n(\tau_\delta^n)] &= X_2^n(\tau_\delta^n)\left(\Pr[Q = d \,|\, Q \ge 2] + \frac{1}{d^{\Omega(d)}}O\left(\frac{h(\delta)}{\delta\sqrt{n}}\right)\right)\\
&= \mu^{(d)} + \frac{1}{d^{\Omega(d)}}O(h(\delta)\sqrt{n}),\\
\mathrm{Var}[X^{(d)} \,|\, X^n(\tau_\delta^n)] &\le X_2^n(\tau_\delta^n)\left(\frac{1}{d^{\Omega(d)}}O(1)\right) + X_2^n(\tau_\delta^n)^2\left(\frac{1}{d^{\Omega(d)}}O\left(\frac{1}{\delta\sqrt{\delta}n}\right)\right)\\
&= \frac{1}{d^{\Omega(d)}}O(\sqrt{\delta}n).
\end{aligned}$$

Using Hölder's inequality, we deduce

$$\begin{aligned}
\mathbb{E}\Big[|X^{(d)} - \mu^{(d)}| \,\Big|\, X^n(\tau_\delta^n)\Big] &\le \Big|\mathbb{E}[X^{(d)} \,|\, X^n(\tau_\delta^n)] - \mu^{(d)}\Big| + \sqrt{\mathrm{Var}[X^{(d)} \,|\, X^n(\tau_\delta^n)]}\\
&\le \frac{1}{d^{\Omega(d)}}O(h(\delta)\sqrt{n} + \delta^{1/4}\sqrt{n}).
\end{aligned} \tag{5.13}$$

Finally we consider $|X^{(d)} - \mu^{(d)}|$ in the case $d > \log n$. By Lemma 5.14(2), again using Lemma 5.1(1–2),

$$\begin{aligned}
\mathbb{E}\Big[|X^{(d)} - \mu^{(d)}| \,\Big|\, X^n(\tau_\delta^n)\Big] = \mathbb{E}[X^{(d)} \,|\, X^n(\tau_\delta^n)] &\le X_2^n(\tau_\delta^n)\, O\left((X_2^n(\tau_\delta^n)Z^n(\tau_\delta^n))^{1/2}\frac{Z^n(\tau_\delta^n)^d}{d!f(Z^n(\tau_\delta^n))}\right)\\
&\le \frac{1}{d^{\Omega(d)}}O(\delta^{7/4}).
\end{aligned} \tag{5.14}$$

Combining (5.10), (5.13), and (5.14), we deduce that whenever $X^n(\tau_\delta^n)$ satisfies (5.9) we have

$$\mathbb{E}[D \,|\, X^n(\tau_\delta^n)] = \sum_d d\,\mathbb{E}\Big[|X^{(d)} - \mu^{(d)}| \,\Big|\, X^n(\tau_\delta^n)\Big] \le O(h(\delta)\sqrt{n} + \delta^{1/4}\sqrt{n}) \le (\varepsilon^2/2)\sqrt{n}$$

for sufficiently small $\delta > 0$. So, by Markov's inequality, if we condition on (5.9) we have $D \le \varepsilon\sqrt{n}$ with probability at least $1 - \varepsilon/2$. The desired result follows, recalling that (5.9) holds with probability at least $1 - \varepsilon/2$. □

References

[1] Jonathan Aronson, Alan Frieze, and Boris G. Pittel, *Maximum matchings in sparse random graphs: Karp-Sipser revisited*, Random Structures Algorithms **12** (1998), 111–177. ↑2, ↑4, ↑5, ↑6, ↑11, ↑12, ↑15, ↑21, ↑22

[2] M. Bauer and O. Golinelli, *Core percolation in random graphs: a critical phenomena analysis*, The European Physical Journal B **24** (2001), 339–352. ↑1, ↑4

[3] M. Bauer and O. Golinelli, *Exactly solvable model with two conductor-insulator transitions driven by impurities*, Physical Review Letters **86** (2001), 2621–2624. ↑1

[4] Patrick Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication. ↑7

[5] 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. ↑8

[6] Charles Bordenave, Marc Lelarge, and Justin Salez, *The rank of diluted random graphs*, Ann. Probab. **39** (2011), 1097–1121. ↑1, ↑2, ↑7

[7] Thomas Budzinski, Alice Contat, and Nicolas Curien, *The critical Karp–Sipser core of random graphs*, arXiv:2212.02463. ↑4

[8] Sky Cao, *Central limit theorems for combinatorial optimization problems on sparse Erdős-Rényi graphs*, Ann. Appl. Probab. **31** (2021), 1687–1723. ↑2

[9] Amin Coja-Oghlan, Oliver Cooley, Mihyun Kang, Joon Lee, and Jean Bernoulli Ravelomanana, *The sparse parity matrix*, arXiv:2107.06123. ↑1

[10] 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

[11] Robert M. Corless, David J. Jeffrey, and Donald E. Knuth, *A sequence of series for the Lambert W function*, Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation (Kihei, HI), ACM, New York, 1997, pp. 197–204. ↑15

[12] R. W. R. Darling and J. R. Norris, *Differential equation approximations for Markov chains*, Probab. Surv. **5** (2008), 37–79. ↑2

[13] P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368. ↑1

[14] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. ↑2, ↑9

[15] 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. ↑4

[16] David Gamarnik, Tomasz Nowicki, and Grzegorz Swirszcz, *Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method*, Random Structures Algorithms **28** (2006), 76–106. ↑2

[17] Margalit Glasgow, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *The exact rank of sparse random graphs*, arXiv:2303.05435. ↑1, ↑3, ↑4

[18] Svante Janson, *Monotonicity, asymptotic normality and vertex degrees in random graphs*, Bernoulli **13** (2007), 952–965. ↑13

[19] Svante Janson, *Random graphs with given vertex degrees and switchings*, Random Structures Algorithms **57** (2020), 3–31. ↑3, ↑8, ↑9

[20] Svante Janson and Malwina J. Luczak, *Asymptotic normality of the k-core in random graphs*, Ann. Appl. Probab. **18** (2008), 1085–1137. ↑2

[21] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. ↑2

[22] 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. ↑1, ↑4, ↑6, ↑11

[23] Eleonora Kreačić, *Some problems related to the Karp-Sipser algorithm on random graphs*, Ph.D. thesis, University of Oxford, 2017. Available online at the *Oxford University Research Archive*, `https://ora.ox.ac.uk/objects/uuid:3b2eb52a-98f5-4af8-9614-e4909b8b9ffa`. ↑2, ↑3, ↑4, ↑5, ↑9, ↑10, ↑11, ↑12, ↑13, ↑14, ↑18

[24] Marc Lelarge, *Bypassing correlation decay for matchings with an application to XORSAT*, 2013 IEEE Information Theory Workshop (ITW), IEEE, September 2013. ↑3

[25] 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. ↑8

[26] 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. ↑1

[27] Boris Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures Algorithms **1** (1990), 311–342. ↑2, ↑4

[28] Roman Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018. ↑7

[29] 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. ↑2

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