arXiv:2105.14883v1 [math.CO] 31 May 2021

COMPONENT BEHAVIOUR AND EXCESS OF RANDOM BIPARTITE GRAPHS NEAR THE CRITICAL POINT[†]

TUAN ANH DO[‡], JOSHUA ERDE[‡], AND MIHYUN KANG[‡]

ABSTRACT. The binomial random bipartite graph $G(n,n,p)$ is the random graph formed by taking two partition classes of size n and including each edge between them independently with probability p. It is known that this model exhibits a similar phase transition as that of the binomial random graph $G(n,p)$ as p passes the critical point of $\frac{1}{n}$. We study the component structure of this model near to the critical point. We show that, as with $G(n,p)$, for an appropriate range of p there is a unique 'giant' component and we determine asymptotically its order and excess. We also give more precise results for the distribution of the number of components of a fixed order in this range of p. These results rely on new bounds for the number of bipartite graphs with a fixed number of vertices and edges, which we also derive.

1. INTRODUCTION

1.1. **Background and motivation.** It was shown by Erdős and Rényi [10] that a 'phase transition' occurs in the uniform random graph model $G(n,m)$ when m is around $\frac{n}{2}$. Standard arguments on the asymptotic equivalence of the two models imply that a similar phenomenon occurs in the binomial random graph model $G(n,p)$ when p is around $\frac{1}{n}$. More precisely, when $p=\frac{1-\epsilon}{n}$ for a fixed $\epsilon>0$, with high probability[1] (whp for short) every component of $G(n,p)$ has order at most $O(\log n)$; when $p=\frac{1}{n}$ whp the order of the largest component is $\Theta\left(n^{\frac{2}{3}}\right)$; and when $p=\frac{1+\epsilon}{n}$ whp $G(n,p)$ contains a unique 'giant component' $L_1\left(G(n,p)\right)$ of order $\Omega(n)$.

Whilst it may seem at first that the component behaviour of the model $G(n,p)$ exhibits quite a sharp 'jump' at this point, subsequent investigations, notably by Bollobás [4] and Łuczak [17], showed that in fact, if one chooses the correct parameterisation for p, this change can be seen to happen quite smoothly. In particular, Łuczak's work implies the following result in the *weakly supercritical regime*. Throughout the paper let $L_i(G)$ denote the ith largest component of a graph G for $i\in\mathbb{N}$. We use the standard Landau notation for asymptotic orders.

Theorem 1.1 ([17]). *Let $\epsilon=\epsilon(n)>0$ be such that $\epsilon^3 n\to\infty$ and $\epsilon=o(1)$, and let $p=\frac{1+\epsilon}{n}$. Then whp*

$$\left|L_1\left(G(n,p)\right)\right| = (1+o(1))2\epsilon n \qquad \text{and} \qquad \left|L_2\left(G(n,p)\right)\right| \le n^{\frac{2}{3}}.$$

Furthermore, Łuczak's work allowed him to give a precise estimate for the excess of $L_1\left(G(n,p)\right)$ (the *excess* of a graph is the difference between the number of edges and vertices). The excess is in some way a broad measure of the complexity of the giant component, determining its density, which has important consequences, for example in terms of the length of the longest cycle in (see for example [16]), or the genus of the giant component (see for example [9]).

Theorem 1.2 ([17]). *Let $\epsilon=\epsilon(n)>0$ be such that $\epsilon^3 n\to\infty$ and $\epsilon=o(1)$, and let $p=\frac{1+\epsilon}{n}$. Then whp*

$$\text{excess}\left(L_1\left(G(n,p)\right)\right) = (1+o(1))\frac{2}{3}\epsilon^3 n.$$

Łuczak also gave a finer picture of the distribution of the components in $G(n,p)$ in the *weakly subcritical* and *weakly supercritical* regimes.

Theorem 1.3 ([17]). *Let $\epsilon=\epsilon(n)$ be such that $|\epsilon|^3 n\to\infty$ and $\epsilon=o(1)$, let $p=\frac{1+\epsilon}{n}$, let $\delta=\epsilon-\log(1+\epsilon)$, and let $\alpha=\alpha(n)>0$ be an arbitrary function. Then the following hold in $G(n,p)$*

[†]An extended abstract of this paper will appear in the Proceedings of the European Conference on Combinatorics, Graph Theory and Applications (Eurocomb 21).

[‡] Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria, {do, erde,kang}@math.tugraz.at. Supported by Austrian Science Fund (FWF): I3747, W1230.

[1]With probability tending to one as $n\to\infty$.

(i) With probability $1-e^{-\Omega(\alpha)}$ *there are no tree components of order larger than*

$$\frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right)-\frac{5}{2}\log\log\left(|\epsilon|^3 n\right)+\alpha\right).$$

(ii) With probability $1-e^{-\Omega(\alpha)}$ *there are no unicyclic components of order larger than* $\frac{\alpha}{\delta}$*.*
(iii) If $\epsilon<0$*, then whp there are no complex components.*
(iv) If $\epsilon>0$*, then with probability* $1-O\left(\left(\epsilon^3 n\right)^{-1}\right)$ *there are no complex components of order smaller than* $n^{\frac{2}{3}}$*.*

In this paper we investigate similar questions about the component structure of a different random graph model, the binomial random bipartite graph $G(n,n,p)$, near to its critical point. The binomial random bipartite graph $G(n_1,n_2,p)$ is the random graph given by taking two partition classes N_1 and N_2 of sizes n_1 and n_2 respectively and including each edge between N_1 and N_2 independently and with probability p. For simplicity, we restrict our attention to the case where $n_1=n_2$. It is possible that similar techniques will work as long as the ratio $\frac{n_1}{n_2}=\Theta(1)$ is a fixed constant.

As in the case of $G(n,p)$, it is known, see for example [14], that when $p=\frac{1-\epsilon}{n}$ for a fixed $\epsilon>0$, whp every component of $G(n,n,p)$ has order at most $O(\log n)$, and when $p=\frac{1+\epsilon}{n}$, whp $G(n,n,p)$ contains a unique 'giant component' $L_1\left(G(n,n,p)\right)$ of order $\Omega(n)$. Hence, a phase transition occurs at $p=\frac{1}{n}$, as in $G(n,p)$.

There has been some interest in this model recently: Johannson [14] determinined the critical point as described above in the general $G(n_1,n_2,p)$ model, Mohar and Ying [13] determined the genus of $G(n_1,n_2,p)$ in the dense regime, and Do, Erde and Kang [8] determined the genus of $G(n_1,n_2,p)$ in the sparse regime.

This model can also be considered as a special case of the *inhomogeneous random graphs* studied by Bollobás, Janson and Riordan[6], who studied the phase transition in this much broader model. Whilst their results do not apply in the weakly supercritical regime, this regime was studied for a particular model of inhomogenous random graphs, which again generalises the bipartite binomial random graph, namely the *multi-type binomial random graph*, by Kang, Koch and Pachón[15]. In particular, it follows from their work that in the weakly supercritical regime there is a unique giant component, and they determine asymptotically its order.

Theorem 1.4 ([15])**.** *Let* $\epsilon=\epsilon(n)>0$ *be such that* $\epsilon^3 n\to\infty$ *and* $\epsilon=o(1)$*, let* $p=\frac{1+\epsilon}{n}$*, and let* $L_i=L_i\left(G(n,n,p)\right)$ *for* $i=1,2$*. Then whp*

$$|L_1\cap N_1|=(1+o(1))2\epsilon n \quad \text{and} \quad |L_1\cap N_2|=(1+o(1))2\epsilon n.$$

Furthermore, whp $|L_2|=o(\epsilon n)$*.*

In this paper we extend and strengthen the work in [14, 15] on the component structure of $G(n,n,p)$ in the weakly supercritical regime.

1.2. **Main results.** In this paper we prove the following analogues of Theorems 1.1–1.3 in the binomial random bipartite graph model.

Our first main result determines the existence and asymptotic order of the 'giant' component in $G(n,n,p)$ near the critical point.

Theorem 1.5. *Let* $\epsilon=\epsilon(n)>0$ *be such that* $\epsilon^4 n\to\infty$ *and* $\epsilon=o(1)$*, let* $p=\frac{1+\epsilon}{n}$*, and let* $L_i=L_i\left(G(n,n,p)\right)$ *for* $i=1,2$*. Let* ϵ' *be defined as the unique positive solution to* $(1-\epsilon')e^{\epsilon'}=(1+\epsilon)e^{-\epsilon}$*, then with probability* $1-O\left(\left(\epsilon^4 n\right)^{-\frac{1}{6}}\right)$ *we have*

$$\left|L_1-\frac{2(\epsilon+\epsilon')}{1+\epsilon}n\right|<\frac{1}{50}n^{\frac{2}{3}} \qquad \text{and} \qquad |L_2|\le n^{\frac{2}{3}}.$$

Furthermore, with probability $1-O\left(\left(\epsilon^4 n\right)^{-\frac{1}{6}}\right)$ *we have that*

$$|L_1\cap N_1|=(1\pm 2\sqrt{\epsilon})|L_1\cap N_2|.$$

Note that $\epsilon'=\epsilon+\frac{2}{3}\epsilon^2+O(\epsilon^3)$. Hence, Theorem 1.5 gives a more precise bound on the order of L_1 than Theorem 1.4, as well as determining more precisely the distribution of the vertices of L_1 between the partition classes, and giving a better bound on the order of the second largest component. Moreover, we determine asymptotically the excess of L_1.

Theorem 1.6. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^4 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. Then whp*

$$\mathit{excess}\left(L_1\left(G(n,n,p)\right)\right) = (1+o(1))\frac{4}{3}\epsilon^3 n.$$

In addition, we can give a much more precise picture of the component structure of $G(n,n,p)$ near to the critical point in both the weakly subcritical and weakly supercritical regime. In what follows, let us write

$$\delta = \epsilon - \log(1+\epsilon). \tag{1}$$

Firstly, for the tree components, we show that whp there are no tree components of order significantly larger than $\frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right)\right)$. Moreover, we show that the number of tree components of order around this tends to a Poisson distribution.

Theorem 1.7. *Let $\epsilon = \epsilon(n)$ be such that $|\epsilon|^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$.*

(i) Given $r_1, r_2 \in \mathbb{R}^+$ with $r_1 < r_2$ let Y_{r_1,r_2} denote the number of tree components in $G(n,n,p)$ of orders between

$$\frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + r_1\right) \text{ and } \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + r_2\right),$$

where δ is as in (1) and let $\lambda = \lambda(r_1, r_2) := \frac{1}{\sqrt{\pi}}\left(e^{-r_1} - e^{-r_2}\right)$. Then Y_{r_1,r_2} converges in distribution to $Po(\lambda)$.

(ii) With probability $1 - e^{-\Omega(\alpha)}$, $G(n,n,p)$ contains no tree components of order larger than

$$\frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + \alpha\right)$$

for any function $\alpha = \alpha(n) > 0$.

Secondly, for the unicyclic components, we show that whp there are no unicyclic components of order significantly larger than $\frac{1}{\delta}$, and moreover, that the number of unicyclic components of order around this again tends to a Poisson distribution.

Theorem 1.8. *Let $\epsilon = \epsilon(n)$ be such that $|\epsilon|^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$.*

(i) Given $u_1, u_2 \in \mathbb{R}^+$ with $u_1 < u_2$ let Z_{u_1,u_2} denote the number of unicyclic components in $G(n,n,p)$ of orders between

$$\frac{u_1}{\delta} \text{ and } \frac{u_2}{\delta},$$

where δ is as in (1) and let $\nu = \nu(u_1,u_2) := \frac{1}{2}\int_{u_1}^{u_2}\frac{\exp(-t)}{t}\,dt$. Then Z_{u_1,u_2} converges in distribution to $Po(\nu)$.

(ii) With probability $1 - e^{-\Omega(\alpha)}$, $G(n,n,p)$ contains no unicyclic components of order larger than $\frac{\alpha}{\delta}$ for any function $\alpha = \alpha(n) > 1$.

Finally, we show that there are whp no complex components of order great than $n^{\frac{2}{3}}$, and in fact no complex components at all in the weakly subcritical regime.

Theorem 1.9. *Let $\epsilon = \epsilon(n)$ be such that $|\epsilon|^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$.*

(i) If $\epsilon < 0$ then with probability $1 - O\left(\left(|\epsilon|^3 n\right)^{-1}\right)$, $G(n,n,p)$ contains no complex components.

(ii) If in addition $\epsilon^4 n \to 0$, then with probability $1 - O\left(\left(\epsilon^4 n\right)^{-1}\right)$, $G(n,n,p)$ contains no complex components of order smaller than $n^{\frac{2}{3}}$.

Note that, in comparison to Theorems 1.1–1.3 and Theorem 1.4, some of our results require a stronger assumption that $\epsilon^4 n \to \infty$. This deficiency, in terms of the size of the critical window, seems to be an artefact of the proofs, and we expect the correct size of the critical window to be as in $G(n,p)$.

More precisely, our results rely on enumerative estimates for the number of bipartite graphs with a fixed number of vertices and edges. For certain ranges of these parameters we are only able to give a weak bound on the number of such graphs, however there is a natural conjecture to make, motivated by the corresponding bounds in the non-bipartite case, contingent on which our proofs would work under the weaker assumption that $\epsilon^3 n \to \infty$. We will discuss this issue in more detail in Section 6.

1.3. **Key proof ideas.** As opposed to previous results concerning the phase transition in the binomial random bipartite graph, such as [14], [15] and [6], which analyse this model by comparison to branching processes, our approach is at heart based on enumerative methods, following the methods of Bollobás [4] and Łuczak [17]. That is, we first derive estimates for the number of bipartite graphs with a fixed number of vertices and edges, and use these to bound the expectation, and higher order moments, of the number of components of various types in $G(n,n,p)$. This will allow us to describe the distribution of the *small* components in $G(n,n,p)$, in particular Theorems 1.7–1.9 will follow from such considerations. Furthermore, we can also bound quite precisely the number of vertices contained in *large* components in $G(n,n,p)$, those of order at least $n^{\frac{2}{3}}$. It can then be shown using a standard sprinkling argument that whp there is a unique large component L_1 containing all these vertices. Given the order of L_1, we can again use these enumerative estimates to give a weak bound on its excess, which we can then bootstrap to an asymptotically tight bound via a multi-round exposure argument. This turns out to be quite a delicate argument, and in particular we make use of a correlation inequality of Harris. The main difficulties here, as opposed to the case of $G(n,p)$, come from the fact that the components of a fixed order can be split in various different ways across the partition classes, making the combinatorial expressions for the expected number of such components much harder to estimate or evaluate.

In order to derive these estimates for the number of bipartite graphs with a fixed number of vertices and edges we will use some standard enumerative tools such as Prüfer codes, as well as a probabilistic argument based on a result of Łuczak [18]. We will find that it is much easier to count the bipartite graphs whose partition classes have relatively equal sizes, which we call *balanced*, and in this case we obtain effective bounds. Since these enumerative results translate directly into bounds on the moments of the number of components with a fixed number of vertices and edges in $G(n,n,p)$, we will often have to split such calculations into two parts depending on whether these components are balanced or not. In the latter case it is then necessary to obtain tighter probabilistic bounds to account for the weaker enumerative bounds.

The benefit in working directly with these enumerative results is in the increased accuracy, allowing for much finer control over the structure of $G(n,n,p)$ in the weakly supercritical regime. For this reason, these estimates may be useful in order to apply similar methods to study the structure of $G(n,n,p)$ in this regime in more detail. For example, in the case of $G(n,p)$ Łuczak [16] used similar ideas to describe the distribution of cycles in this regime, and more recently, using some of these ideas, Dowden, Kang and Krivelevich [9] were able to determine asymptotically the genus of $G(n,p)$ in this regime. It is possible that similar ideas could be applied to $G(n,n,p)$, for example to study the distribution of cycles, or the length of the longest cycle, or the genus in this model.

1.4. **Overview of the paper.** The rest of the article is organised as follows. In Section 2, we collect some preliminary results which are used later in the paper. In Section 3 we derive bounds for the expected number of components of $G(n,n,p)$ with a fixed order and excess, which form the foundation of many of the calculations in this paper. These bounds depend on good estimates for the number of bipartite graphs with a fixed number of vertices and edges, whose proofs we give in Section 5. In Section 4.1, we use these estimates to study the distribution of components in $G(n,n,p)$ and prove Theorems 1.7–1.9. Then, in Section 4.2, using the previous results, we investigate the size of the largest components and prove Theorem 1.5. Using this, we then determine the excess of the giant component and prove Theorem 1.6 in Section 4.3. Finally, in Section 6, we discuss possible extensions of our results, formulate a conjecture, and give some open problems.

2. Preliminaries

Unless stated otherwise, all the asymptotics in this paper are taken as $n \to \infty$. In particular, we write $f(n) \approx g(n)$ if $f = (1+o(1))g$, $f(n) \lesssim g(n)$ if $f \leq (1+o(1))g$, and $f(n) \gtrsim g(n)$ if $f \geq (1+o(1))g$. Furthermore, we write that $f(n) \gg g(n)$ if $f(n) \geq Cg(n)$ for an implicit large constant C. We write $\mathbb{N}$ for the set of positive integers, so that in particular $0 \notin \mathbb{N}$.

We will often need the following elementary estimates on the size of the falling factorial, which hold for any $i, n \in \mathbb{N}$ with $i \leq n$

$$(n)_i := \prod_{j=0}^{i-1}(n-j) = n^i \exp\left(-\frac{i^2}{2n} - \frac{i^3}{6n^2} + O\left(\frac{i^4}{n^3}\right)\right), \tag{2}$$

and also

$$(n)_i \le n^i \exp\left(-\frac{i^2}{2n} - \frac{i^3}{6n^2}\right) \le n^i \exp\left(-\frac{i^2}{2n}\right). \tag{3}$$

The following result of Spencer [20] is a useful tool for relating integrals and sums.

Lemma 2.1 ([20, Theorem 4.3])**.** *Let $a < b$ be integers, let $f(x)$ be an integrable function in $[a-1, b+1]$, and let $S := \sum_{i=a}^{b} f(i)$ and $I := \int_a^b f(x)dx$. Let M be such that $|f(x)| \le M$ for all $x \in [a-1, b+1]$ and suppose that $[a-1, b+1]$ can be broken into at most r intervals such that $f(x)$ is monotone on each. Then*

$$|S - I| \le 6rM.$$

Often, when calculating certain expected values, we will need an asymptotic expression for sums of the following form, whose proof we relegate to Appendix A.

Lemma 2.2. *Let $m \ge 0$ and $c > 0$ be constants and let $L = L(n)$ and $k = k(n)$ be such that $L + 1 \le k \le n$, $L = \omega(1)$ and $k = o(n)$. Then*

$$S := \sum_{d=-L}^{L} \frac{1}{(k^2 - d^2)^m} \left(\frac{k-d}{k+d}\right)^{cd} \exp\left(-\frac{d^2}{2n}\right) \approx \sqrt{\frac{\pi}{2c}} k^{\frac{1}{2}-2m}.$$

We will use the following Chernoff type bounds on the tail probabilities of the binomial distribution, see e.g. [2, Appendix A].

Lemma 2.3. *Let $n \in \mathbb{N}$, let $p \in [0,1]$, and let $X \sim Bin(n, p)$. Then for every positive a with $a \le \frac{np}{2}$,*

$$\mathbb{P}\left(\left|X - np\right| > a\right) < 2\exp\left(-\frac{a^2}{4np}\right).$$

We will also need to use the following correlation inequality, which follows from an inequality of Harris [11], which is itself a special case of the FKG-inequality, see for example [2, Section, 6].

Lemma 2.4. *If A is an increasing event and B is a decreasing event of bipartite graphs, then in $G(n, n, p)$*

$$\mathbb{P}(A|B) \le \mathbb{P}(A).$$

Finally, we will also need the following lemma, which gives a useful criterion for when a sequence of random variables converges in distribution to a Poisson distribution.

Lemma 2.5 ([12])**.** *If $X_1, X_2, \cdots$ are random variables with finite moments such that $\mathbb{E}((X_n)_k) \to \lambda^k$ as $n \to \infty$ for every positive integer k, where $(X_n)_k$ is the kth factorial moment of X_n and $\lambda \ge 0$ is a constant, then X_n converges in distribution to $Po(\lambda)$.*

3. Component structure of $G(n, n, p)$

One of the main ways in which we derive information about the distribution of the components in $G(n, n, p)$ is by calculating various moments of the number of components with particular properties, and in particular the expected value.

Given $i, j \in \mathbb{N}$ and $\ell \in \mathbb{Z}$, let $X(i, j, \ell)$ denote the number of components in $G(n, n, p)$ with i vertices in N_1, j vertices in N_2, and $i + j + \ell$ edges. Letting $i + j = k$, we have

$$\mathbb{E}\left(X(i, j, \ell)\right) = \binom{n}{i}\binom{n}{j} C(i, j, \ell) p^{k+\ell} (1-p)^{kn-ij-k-\ell}, \tag{4}$$

where $C(i, j, \ell)$ is the number of bipartite graphs with i vertices in one partition class, j in the second, and $i + j + \ell$ many edges. Hence, in order to understand the quantities $\mathbb{E}\left(X(i, j, \ell)\right)$, it is important to know how the quantities $C(i, j, \ell)$ behave.

In this section we state some bounds for $C(i, j, \ell)$, which we will prove later in Section 5, and derive some consequences of these bounds, using (4), for the expected number of tree, unicyclic and complex components, where a tree component has excess $\ell = -1$, a *unicyclic* component has excess $\ell = 0$ (and hence contains a unique cycle), and a *complex* component has excess $\ell \ge 1$.

The following estimates are useful to this end. Using the fact that $1+x=e^{x+O(x^2)}$ for any $x=o(1)$, we see that for any $i+j=k$ and $\epsilon=o(1)$,

$$\left(1-\frac{1+\epsilon}{n}\right)^{kn-2ij}=\exp\left(-(1+\epsilon)k+\frac{(1+\epsilon)2ij}{n}+O\left(\frac{ij}{n^2}\right)+O\left(\frac{k}{n}\right)\right). \tag{5}$$

Similarly, for any $k=o(n)$, $i+j=k$, and $\epsilon=o(1)$

$$\left(1-\frac{1+\epsilon}{n}\right)^{kn-ij+O(k)}=\exp\left(-(1+\epsilon)k+\frac{(1+\epsilon)ij}{n}+O\left(\frac{ij}{n^2}\right)\right). \tag{6}$$

Throughout this section, unless stated otherwise, we let $\epsilon=\epsilon(n)$ be such that $|\epsilon|^3n\to\infty$ and $\epsilon=o(1)$, and let $p=\frac{1+\epsilon}{n}$. We will also refer to δ as defined in (1), i.e.

$$\delta=\epsilon-\log(1+\epsilon)\approx\frac{\epsilon^2}{2}.$$

3.1. **Tree components.** Let us write $\hat{C}(i,\ell)$ for the number of (not-necessarily bipartite) graphs with i vertices and $i+\ell$ many edges. It is a classic result of Cayley that the number of trees on i vertices, in other words $\hat{C}(i,-1)$, is i^{i-2}. The following result of Scoins [19] gives an analogue for bipartite trees.

Theorem 3.1 ([19])**.** *For any $i,j\in\mathbb{N}$ we have $C(i,j,-1)=i^{j-1}j^{i-1}$.*

As a consequence, we can derive an asymptotic formula for the expected number of tree components in $G(n,n,p)$.

Theorem 3.2. *For any $i=i(n)$, $j=j(n)\in\mathbb{N}$ satisfying $k=i+j=o(n)$, we have*

$$\mathbb{E}\left(X(i,j,-1)\right)\approx\frac{n}{2\pi(ij)^{\frac{3}{2}}}e^{-\delta k}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}-\frac{i^3+j^3}{6n^2}+\frac{\epsilon ij}{n}+O\left(\frac{ij}{n^2}\right)+O\left(\frac{i^4+j^4}{n^3}\right)\right). \tag{7}$$

Proof. By Theorem 3.1 and (4), together with Stirling's formula, we have

$$\begin{aligned}\mathbb{E}\left(X(i,j,-1)\right)&=\binom{n}{i}\binom{n}{j}C(i,j,-1)p^{k-1}(1-p)^{kn-ij-k+1}\\&=\frac{(n)_i}{i!}\frac{(n)_j}{j!}i^{j-1}j^{i-1}p^{k-1}(1-p)^{kn-ij-k+1} \qquad (8)\\&\approx\frac{e^k}{2\pi(ij)^{\frac{3}{2}}}\left(\frac{i}{j}\right)^{j-i}\frac{(n)_i(n)_j}{n^{k-1}}(1+\epsilon)^k\left(1-\frac{1+\epsilon}{n}\right)^{kn-ij-k+1}. \qquad (9)\end{aligned}$$

Hence, by (9), (2) and (6), we obtain

$$\mathbb{E}\left(X(i,j,-1)\right)\approx\frac{n}{2\pi(ij)^{\frac{3}{2}}}e^{-\delta k}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}-\frac{i^3+j^3}{6n^2}+\frac{\epsilon ij}{n}+O\left(\frac{ij}{n^2}\right)+O\left(\frac{i^4+j^4}{n^3}\right)\right).$$

□

3.2. **Unicyclic components.** We will derive in Section 5 the following expression for the number of unicyclic bipartite graphs.

Theorem 3.3. *For any $i,j\in\mathbb{N}$ we have*

$$C(i,j,0)=\frac{1}{2}\left(\frac{1}{i}+\frac{1}{j}\right)i^jj^i\sum_{r=2}^{\min\{i,j\}}\frac{(i)_r(j)_r}{i^rj^r},$$

and so in particular, for any $i=i(n)$, $j=j(n)\in\mathbb{N}$ satisfying $i,j\to\infty$ and $\frac{1}{2}\le\frac{i}{j}\le2$ we have

$$C(i,j,0)\approx\sqrt{\frac{\pi}{8}}\sqrt{i+j}\,i^{j-\frac{1}{2}}j^{i-\frac{1}{2}}.$$

We note for comparison that it is known that

$$\hat{C}(i,0) \approx \sqrt{\frac{\pi}{8}}\, i^{i-\frac{1}{2}},$$

see [5, Corollary 5.19]. We can derive as a consequence an asymptotic formula for the expected number of unicyclic components in $G(n,n,p)$ which are appropriately balanced across the partition classes.

Theorem 3.4. *For any $i = i(n), j = j(n) \in \mathbb{N}$ satisfying $i, j \to \infty$ and $\frac{1}{2} \le \frac{i}{j} \le 2$, and letting $k = i + j$, we have*

$$\mathbb{E}\big(X(i,j,0)\big) \approx \frac{\sqrt{k}}{4\sqrt{2\pi}\, ij} e^{-\delta k} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n} - \frac{i^3+j^3}{6n^2} + \frac{\epsilon ij}{n} + +O\left(\frac{ij}{n^2}\right) + O\left(\frac{i^4+j^4}{n^3}\right)\right). \tag{10}$$

Proof. By Theorem 3.3 and (4), together with Stirling's formula, if $\frac{1}{2} \le \frac{i}{j} \le 2$, then

$$\begin{aligned}\mathbb{E}\big(X(i,j,0)\big) &= \binom{n}{i}\binom{n}{j} C(i,j,0) p^k (1-p)^{kn-ij-k} \approx \frac{(n)_i}{i!}\frac{(n)_j}{j!}\sqrt{\frac{\pi}{8}}\sqrt{k}\, i^{j-\frac{1}{2}} j^{i-\frac{1}{2}} p^k (1-p)^{kn-ij-k} \\ &\approx \frac{e^k \sqrt{k}}{4\sqrt{2\pi}\, ij}\left(\frac{i}{j}\right)^{j-i} \frac{(n)_i (n)_j}{n^k} (1+\epsilon)^k \left(1 - \frac{1+\epsilon}{n}\right)^{kn-ij-k}. \end{aligned} \tag{11}$$

Hence, by (11), (2) and (6), we get

$$\mathbb{E}\big(X(i,j,0)\big) \approx \frac{\sqrt{k}}{4\sqrt{2\pi}\, ij} e^{-\delta k} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n} - \frac{i^3+j^3}{6n^2} + \frac{\epsilon ij}{n} + O\left(\frac{ij}{n^2}\right) + O\left(\frac{i^4+j^4}{n^3}\right)\right).$$

□

3.3. **Complex components.** In Section 5 we will also prove the following upper bound on the number of bipartite graphs with a fixed excess which are appropriately balanced across the partition classes.

Theorem 3.5. *There is a constant c such that for any $i, j, \ell \in \mathbb{N}$ with $\ell \le ij - i - j$ and $\frac{1}{2} \le \frac{i}{j} \le 2$,*

$$C(i,j,\ell) \le i^{j+\frac{1}{2}} j^{i+\frac{1}{2}} (i+j)^{\frac{3\ell+1}{2}} \left(\frac{i}{j}\right)^{\frac{i-j}{2}} \left(\frac{c}{\ell}\right)^{\frac{\ell}{2}}.$$

Furthermore, if $\ell \ge i + j$, then

$$C(i,j,\ell) \le i^{j-\frac{1}{2}} j^{i-\frac{1}{2}} (i+j)^{\frac{3\ell+1}{2}} \ell^{-\frac{\ell}{2}}.$$

We note that for small enough ℓ, for example $\ell = O(1)$, the naive bound that follows from

$$C(i,j,\ell) \le C(i,j,0)(ij)^\ell \le \sqrt{i+j}\, i^{j+\ell-\frac{1}{2}} j^{i+\ell-\frac{1}{2}} \tag{12}$$

is more effective than the first part of Theorem 3.5. We also note for comparison that it is known that there is an absolute constant c such that

$$\hat{C}(i,\ell) \le c\ell^{-\frac{\ell}{2}} i^{i+\frac{3\ell-1}{2}},$$

see [5, Corollary 5.21].

As before, using these bounds we can give an upper bound on the expected number of components with a fixed excess which are appropriately balanced across the partition classes.

Theorem 3.6. *For any $i = i(n), j = j(n), \ell = \ell(n) \in \mathbb{N}$ satisfying $\frac{1}{2} \le \frac{i}{j} \le 2$, $\ell \le ij - i - j$, and $k = i + j = o(n)$, we have*

$$\mathbb{E}\big(X(i,j,\ell)\big) \le \sqrt{k}\left(\frac{i}{j}\right)^{\frac{j-i}{2}} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(-\delta k + \frac{\epsilon k^2}{4n} - \frac{(i-j)^2}{2n} + O\left(\frac{ij}{n^2}\right) + \ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right), \tag{13}$$

and for $\ell = O(1)$, we have

$$\mathbb{E}\big(X(i,j,\ell)\big) = O\left(\frac{\sqrt{k}(ij)^{\ell-1}}{n^\ell}\left(\frac{i}{j}\right)^{j-i} \exp\left(-\delta k + \frac{\epsilon k^2}{4n} - \frac{(i-j)^2}{2n} + O\left(\frac{ij}{n^2}\right)\right)\right). \tag{14}$$

Proof. By Theorem 3.5 and (4), together with Stirling's formula, if $\frac{1}{2} \le \frac{i}{j} \le 2$ and $\ell \le ij - i - j$, then there is an absolute constant c such that,

$$\begin{aligned}\mathbb{E}\big(X(i,j,\ell)\big) &= \binom{n}{i}\binom{n}{j} C(i,j,\ell) p^{k+\ell}(1-p)^{kn-ij-k-\ell} \\ &\le \frac{(n)_i}{i!}\frac{(n)_j}{j!} i^{j+\frac{1}{2}} j^{i+\frac{1}{2}} (i+j)^{\frac{3\ell+1}{2}} \left(\frac{i}{j}\right)^{\frac{i-j}{2}} \left(\frac{c}{\ell}\right)^{\frac{\ell}{2}} p^{k+\ell}(1-p)^{kn-ij-k-\ell} \\ &\le \sqrt{k} e^{-\delta k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \frac{(n)_i (n)_j}{n^k} (1+\epsilon)^{k+\ell} \left(1 - \frac{1+\epsilon}{n}\right)^{kn-ij-k-\ell}. \end{aligned} \tag{15}$$

Hence, by (15), (3) and (6), we see that

$$\begin{aligned}\mathbb{E}\big(X(i,j,\ell)\big) &\le \sqrt{k}\left(\frac{i}{j}\right)^{\frac{j-i}{2}} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(-\delta k + \frac{\epsilon ij}{n} - \frac{(i-j)^2}{2n} + O\left(\frac{ij}{n^2}\right) + \ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right) \\ &\le \sqrt{k}\left(\frac{i}{j}\right)^{\frac{j-i}{2}} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(-\delta k + \frac{\epsilon k^2}{4n} - \frac{(i-j)^2}{2n} + O\left(\frac{ij}{n^2}\right) + \ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right).\end{aligned}$$

For $\ell = O(1)$, using (12) instead of Theorem 3.5, we can instead bound

$$\mathbb{E}\big(X(i,j,\ell)\big) = O\left(\frac{\sqrt{k}(ij)^{\ell-1}}{n^\ell}\left(\frac{i}{j}\right)^{j-i} \exp\left(-\delta k + \frac{\epsilon k^2}{4n} - \frac{(i-j)^2}{2n} + O\left(\frac{ij}{n^2}\right)\right)\right).$$

□

3.4. **More about components.** Since we only have good estimates for $C(i,j,\ell)$ when i and j are comparable in size, it will be useful to show that the expected number of components of a given excess and order is dominated by the contribution from those which are 'evenly spread' across the partition classes, and we should perhaps expect by the symmetry in the model that this is the case for most components. For the most part, we are able to get away with considering a relatively weak notion of 'evenly spread'.

We say a component C of $G(n,n,p)$ is *balanced* if $|C \cap N_1| \le 2|C \cap N_2|$ and $|C \cap N_2| \le 2|C \cap N_1|$, and *unbalanced* otherwise. The following lemma will be useful for simplifying certain calculations, which roughly says that we do not expect there to be any large unbalanced components in $G(n,n,p)$.

Lemma 3.7. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, let $p = \frac{1+\epsilon}{n}$, and let $\alpha = \alpha(n) \to \infty$ be an increasing function.*

(i) With probability $1 - O\big(n^{-1}\big)$, $G(n,n,p)$ contains no unbalanced components of order $\ge 2000 \log n$.
(ii) With probability $1 - e^{-\Omega(\alpha)}$, $G(n,n,p)$ contains no unbalanced non-tree components of order $\ge \alpha$.
(iii) With probability $1 - O\big(n^{-1}\big)$, $G(n,n,p)$ contains no unbalanced complex components.

Proof. Every unbalanced component of order k with excess at least ℓ must contain a spanning tree together with $\ell+1$ extra edges which is otherwise disconnected from the rest of the graph and hence $G(n,n,p)$ contains a component of order k and excess at least ℓ if and only if $G(n,n,p)$ contains such a substructure. Let us denote by $Y(k,\ell)$ the number of such substructures. It follows that if $Y(k,\ell) = 0$, then $G(n,n,p)$ contains no components of order k with excess at least ℓ.

In order to count the expected size of $Y(k,\ell)$ we note that we can specify such a substructure S by choosing i vertices in the first partition class and j vertices in the second, such that that $i+j=k$ and either $j \ge 2i$ or $i \ge 2j$, choosing one of the $i^{j-1}j^{i-1}$ possible bipartite spanning trees on these vertices, and then choosing one of the at most $\binom{ij-i-j}{\ell+1}$ possible sets of $\ell+1$ extra edges. Note that the number of non-edges from these k vertices to the other vertices in $G(n,n,p)$ is $i(n-j) + j(n-i) = kn - 2ij$ (see Figure 1).

It follows that we can bound

$$\mathbb{E}(Y(k,\ell)) \le \sum_{(i,j)\in U_k} \binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1} \binom{ij-i-j}{\ell+1} p^{k+\ell}(1-p)^{kn-2ij},$$

where $U_k = \big\{(i,j) \in \mathbb{N}^2 \colon i+j=k \text{ and } i \ge 2j \text{ or } j \ge 2i\big\}$.

Therefore, using (3), (5) and Stirling's approximation we can bound the expected number by

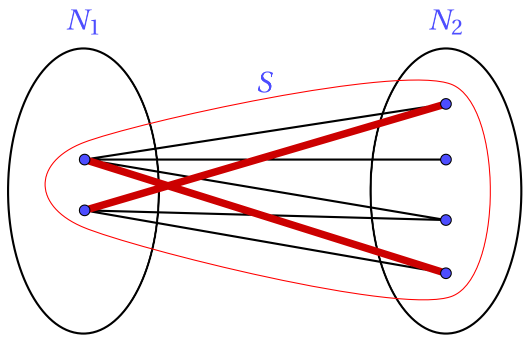

FIGURE 1. A substructure S (in the proof of Lemma 3.7) with $i = 2$ vertices in N_1 and $j = 4$ vertices in N_2 containing a spanning tree (whose edges are drawn with thin edges) and $\ell = 2$ excess edges (which are drawn with thick edges), where none of the $kn - 2ij$ edges from $V(S)$ to the rest of the graph are in $G(n,n,p)$.

$$\begin{aligned}\mathbb{E}(Y(k,\ell)) &\le n^{-\ell} e^{-\delta k} \sum_{(i,j)\in U_k} \frac{(ij)^{\ell-\frac{1}{2}}}{2\pi} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{i^2+j^2}{2n} + \frac{(1+\epsilon)2ij}{n} + O\left(\frac{ij}{n^2}\right) + O\left(\frac{k}{n}\right)\right) \\ &\le n^{-\ell} \sum_{(i,j)\in U_k} (ij)^{\ell-\frac{1}{2}} \left(\frac{i}{j}\right)^{j-i} \exp\left(\frac{ij(1+2\epsilon)}{n} + O(1)\right),\end{aligned}$$

since $e^x < 1$ for $x < 0$, $-i^2 - j^2 + 2ij < 0$ and $i, j \le k \le n$.

However, if $j \ge 2i$ and $i + j = k$, then $j \ge \frac{2k}{3}$ and so $j - i \ge \frac{j}{2} \ge \frac{k}{3}$, and $ij \le \frac{2k^2}{9}$. It follows that $\left(\frac{i}{j}\right)^{j-i} \le \left(\frac{1}{2}\right)^{\frac{k}{3}}$. A similar calculation holds if $i \ge 2j$. Hence the expected number of such substructures is at most

$$\mathbb{E}(Y(k,\ell)) \le n^{-\ell} \exp\left(\frac{2(1+2\epsilon)k^2}{9n} - \frac{k\log 2}{3} + O(1)\right) \sum_{(i,j)\in U_k} (ij)^{\ell-\frac{1}{2}} \le n^{-\ell} e^{-\frac{k}{1000}} \sum_{(i,j)\in U_k} (ij)^{\ell-\frac{1}{2}},$$

since $\frac{2(1+2\epsilon)k^2}{9n} - \frac{k\log 2}{3} + O(1) \le k\left(\frac{2(1+2\epsilon)}{9} - \frac{\log 2}{3}\right) + O(1) \le -\frac{k}{1000}$ when ϵ is sufficiently small.

Hence, if we let $Y_{\ge r}(\ell) = \sum_{k\ge r} Y(k,\ell)$, then with $r = 2000\log n$

$$\mathbb{E}(Y_{\ge r}(-1)) \le n \sum_{k\ge r} e^{-\frac{k}{1000}} \sum_{(i,j)\in U_k} (ij)^{-\frac{3}{2}} \le n \sum_{k\ge r} e^{-\frac{k}{1000}} = O\left(\frac{1}{n}\right).$$

Hence, by Markov's inequality, with probability $1 - O(n^{-1})$, $Y_{\ge r}(-1) = 0$ and in particular there are no unbalanced components of order at least $2000\log n$.

Similarly, if $\alpha = \alpha(n) \to \infty$ is an increasing function, then

$$\mathbb{E}(Y_{\ge\alpha}(0)) \le \sum_{k\ge\alpha} e^{-\frac{k}{1000}} \sum_{(i,j)\in U_k} (ij)^{-\frac{1}{2}} \le \sum_{k\ge\alpha} \sqrt{k} e^{-\frac{k}{1000}} = O\left(e^{-\frac{\alpha}{2000}}\right),$$

and so, again by Markov's inequality, with probability $1 - e^{-\Omega(\alpha)}$, there are no unbalanced components of order at least α with excess greater than zero, and so in particular no unicyclic components of order at least α.

Finally, we see that

$$\mathbb{E}(Y_{\ge 1}(1)) \le \frac{1}{n} \sum_{k\ge 1} e^{-\frac{k}{1000}} \sum_{(i,j)\in U_k} (ij)^{\frac{1}{2}} \le \frac{1}{n} \sum_{k\ge 1} k^2 e^{-\frac{k}{1000}} = O\left(\frac{1}{n}\right),$$

and so as before with probability $1 - O(n^{-1})$ there are no unbalanced complex components. □

For most applications the rather coarse notion of balanced is enough for our purposes, but in one case we will need to restrict our attention to components which are much more evenly distributed over the partition classes. We say a component C of $G(n,n,p)$ is *ϵ-uniform* if $||C \cap N_1| - |C \cap N_2|| < \epsilon^{\frac{1}{4}} \sqrt{n}$.

Lemma 3.8. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. Then with probability $1 - o\left(n^{-1}\right)$, $G(n,n,p)$ contains no non-ϵ-uniform tree components of order at most $n^{\frac{2}{3}}$.*

Proof. As in the previous lemma, let us write $U_k = \left\{(i,j) \in \mathbb{N}^2 : i+j=k, |i-j| \geq \epsilon^{\frac{1}{4}}\sqrt{n}\right\}$ for the pairs (i,j) representing non-ϵ-uniform components. Note that, if $(i,j) \in U_k$ and $k \leq n^{\frac{2}{3}}$, then

$$\left(\frac{i}{j}\right)^{j-i} \leq \left(1 - \frac{\epsilon^{\frac{1}{4}}\sqrt{n}}{n^{\frac{2}{3}}}\right)^{\epsilon^{\frac{1}{4}}\sqrt{n}} \leq e^{-\sqrt{\epsilon} n^{\frac{1}{3}}}.$$

Then, using (7), we can bound the expected number of non-ϵ-uniform tree components of order at most $n^{\frac{2}{3}}$ by

$$\begin{aligned}\sum_{k=1}^{n^{\frac{2}{3}}} \sum_{(i,j)\in U_k} \mathbb{E}\left(X(i,j,-1)\right) &\leq \sum_{k=1}^{n^{\frac{2}{3}}} \sum_{(i,j)\in U_k} \frac{n}{2\pi (ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(\frac{\epsilon i j}{n} + o(1)\right) \\ &\leq \sum_{k=1}^{n^{\frac{2}{3}}} n e^{\frac{\epsilon k^2}{4n} - \sqrt{\epsilon} n^{\frac{1}{3}}} \sum_{(i,j)\in U_k} \frac{1}{(ij)^{\frac{3}{2}}} \leq \sum_{k=1}^{n^{\frac{2}{3}}} \frac{n}{k^{\frac{1}{2}}} e^{\frac{\epsilon k^2}{4n} - \sqrt{\epsilon} n^{\frac{1}{3}}}.\end{aligned}$$

However, since $k \leq n^{\frac{2}{3}}$ and $\epsilon^3 n \to \infty$, it follows that

$$\frac{\epsilon k^2}{4n} - \sqrt{\epsilon} n^{\frac{1}{3}} = -\Omega\left(\sqrt{\epsilon} n^{\frac{1}{3}}\right) \leq -n^{\frac{1}{6}}.$$

It follows that the expected number of non-ϵ-uniform tree components of order at most $n^{\frac{2}{3}}$ is at most

$$n e^{-n^{\frac{1}{6}}} \sum_{k=1}^{n^{\frac{2}{3}}} \frac{1}{k^{\frac{1}{2}}} \leq n^{\frac{4}{3}} e^{-n^{\frac{1}{6}}} = o\left(n^{-1}\right).$$

Hence, the result follows by Markov's inequality. □

It will also be useful to have a bound on the variance of the number of ϵ−uniform tree components with small order, which is given by the following lemma, whose proof is given in Appendix B.

Lemma 3.9. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. Given $\tilde{k}, a \in \mathbb{N}$, set $Z_a = \sum_{k=1}^{\tilde{k}} k^a Z(k)$ where $Z(k)$ is the number of ϵ-uniform tree components of order k in $G(n,n,p)$. If $\tilde{k} \leq n^{\frac{2}{3}}$ and $\frac{3\epsilon \tilde{k}^2}{n} < 1$, then* $\mathrm{Var}(Z_1) = O\left(\frac{n}{\epsilon}\right)$.

4. A finer look at component structure of $G(n,n,p)$

Using the bounds from Section 3 on the expected number of components with a fixed order and excess we can describe more precisely the component structure of $G(n,n,p)$.

4.1. **Distribution of the number of components: proof of Theorems 1.7–1.9.**

Firstly, as indicated in Theorem 1.7, we show that whp there are no tree components in $G(n,n,p)$ whose order is significantly larger than $\frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right)\right)$. Moreover, we show that the number of tree components of order around this tends to a Poisson distribution.

Proof of Theorem 1.7.
Part (i)**:** Let us write $k_i = \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + r_i\right)$ for $i \in \{1,2\}$. Then for all $k_1 \leq k \leq k_2$, we have that $\frac{k^3}{n^2}$, $\frac{k^4}{n^3}$ and $\frac{\epsilon k^2}{n}$ are all $o(1)$. Therefore, it follows from (7) that

$$\mathbb{E}(Y_{r_1,r_2}) \approx \frac{n}{2\pi} \sum_{k=k_1}^{k_2} e^{-\delta k} \sum_{i+j=k} \frac{1}{(ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right) = \frac{4n}{\pi} \sum_{k=k_1}^{k_2} e^{-\delta k} \sum_{d=-k+1}^{k-1} \frac{1}{(k^2-d^2)^{\frac{3}{2}}} \left(\frac{k-d}{k+d}\right)^d \exp\left(-\frac{d^2}{2n}\right),$$

where the last equality holds by reparameterising over $d = j - i$. Hence, by Lemma 2.2, we have

$$\mathbb{E}(Y_{r_1,r_2}) \approx \frac{2\sqrt{2}n}{\sqrt{\pi}} \sum_{k=k_1}^{k_2} \frac{e^{-\delta k}}{k^{\frac{5}{2}}}. \tag{16}$$

Now, for any $\frac{r_1}{\delta} \le a \le \frac{r_2}{\delta}$ and

$$k = \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right)\right) + a,$$

we have that

$$k^{\frac{5}{2}} \approx 4\sqrt{2}|\epsilon|^{-5}\left(\log\left(|\epsilon|^3 n\right)\right)^{\frac{5}{2}},$$

since $\delta \approx \frac{\epsilon^2}{2}$, and hence in this range

$$\frac{e^{-\delta k}}{k^{\frac{5}{2}}} = \frac{\left(\log\left(|\epsilon|^3 n\right)\right)^{\frac{5}{2}} e^{-\delta a}}{|\epsilon^3| n k^{\frac{5}{2}}} \approx \frac{|\epsilon^2| e^{-\delta a}}{4\sqrt{2}n} \approx \frac{\delta e^{-\delta a}}{2\sqrt{2}n}. \tag{17}$$

Hence, substituting (17) into (16) we obtain

$$\mathbb{E}(Y_{r_1,r_2}) \approx \frac{1}{\sqrt{\pi}} \sum_{a=\frac{r_1}{\delta}}^{\frac{r_2}{\delta}} \delta e^{-\delta a} \approx \frac{1}{\sqrt{\pi}} \int_{r_1}^{r_2} e^{-t} dt = \frac{1}{\sqrt{\pi}}\left(e^{-r_1} - e^{-r_2}\right) =: \lambda.$$

Next, we calculate the expected value of $(Y_{r_1,r_2})_2$, i.e. the second factorial moment of Y_{r_1,r_2}, which is the expected number of ordered pairs of tree components whose orders lie between r_1 and r_2. We have that

$$\mathbb{E}\left((Y_{r_1,r_2})_2\right) = \sum_{k=k_1}^{k_2} \sum_{i+j=k} \binom{n}{i}\binom{n}{j} p^{k-1}(1-p)^{kn-ij-k+1} \sum_{k'=k_1}^{k_2} \sum_{r+s=k} \binom{n-i}{r}\binom{n-j}{s} p^{k'-1}(1-p)^{k'n-rs-is-rj-k'+1},$$

and we note that the inner sum is the expected number of tree components of order between k_1 and k_2 in $G(n_1, n_2, p)$, where $n_1 = n-i, n_2 = n-j$. However, since $i, j \le k_2 = o(n)$ the same argument as before shows that this inner sum is asymptotically equal to $\mathbb{E}(Y_{r_1,r_2})$, and hence

$$\mathbb{E}\left((Y_{r_1,r_2})_2\right) \approx \left(\mathbb{E}(Y_{r_1,r_2})\right)^2 \approx \lambda^2.$$

A similar argument shows that the ith factorial moment $\mathbb{E}\left((Y_{r_1,r_2})_i\right) \approx \lambda^i$ for each $i \in \mathbb{N}$, and hence Y_{r_1,r_2} converges in distribution to $\mathrm{Po}(\lambda)$ by Lemma 2.5.

Part (ii)**:** Let us write $k_3 = \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + \alpha\right)$ and $Y_{\ge\alpha}$ for the number of tree components of order at least k_3. From (7), but using (3) instead of (2) to bound the falling factorial term, we can bound the expected value of $Y_{\ge\alpha}$ from above as

$$\mathbb{E}(Y_{\ge\alpha}) \le n \sum_{k=k_3}^{n} e^{-\delta k} \sum_{i+j=k} \frac{1}{2\pi(ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n} - \frac{i^3+j^3}{6n^2} + \frac{\epsilon ij}{n} + O\left(\frac{ij}{n^2}\right)\right).$$

For any $k \le n$ and $i + j = k$ we have that $\frac{ij}{n^2} \le 1/4$, $\frac{\epsilon ij}{n} \le \frac{\epsilon k^2}{4n}$ and also $\frac{i^3+j^3}{6n^2} \ge \frac{k^3}{24n^2}$, and so

$$\mathbb{E}(Y_{\ge\alpha}) \le e^{\frac{1}{4}} n \sum_{k=k_3}^{n} \exp\left(-\delta k + \frac{\epsilon k^2}{4n} - \frac{k^3}{24n^2}\right) \sum_{i+j=k} \frac{1}{2\pi(ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right).$$

Then, reparameterising with $d = j - i$ and applying Lemma 2.2 as before gives us that,

$$\mathbb{E}(Y_{\ge\alpha}) = O\left(n \sum_{k=k_3}^{n} \frac{1}{k^{\frac{5}{2}}} \exp\left(-\delta k - \frac{k^3}{24n^2} + \frac{\epsilon k^2}{4n}\right)\right). \tag{18}$$

Let $s = \epsilon n$, then we are interested in the function

$$-\delta k + \frac{\epsilon k^2}{4n} - \frac{k^3}{24n^2} = \frac{k}{n^2}\left(-\frac{\delta s^2}{\epsilon^2} + \frac{sk}{4} - \frac{k^2}{24}\right).$$

Now, since $-\frac{\delta s^2}{\epsilon^2} + \frac{sk}{4} - \frac{k^2}{24}$ as a function of k is a parabola, whose maximum comes at $k = 3s$, we can bound

$$\frac{k}{n^2}\left(-\frac{\delta s^2}{\epsilon^2} + \frac{sk}{4} - \frac{k^2}{24}\right) \le k\left(-\delta + \frac{3\epsilon^2}{4} - \frac{9\epsilon^2}{24}\right) \le -\frac{\delta k}{5}. \tag{19}$$

Hence, by (18) and (19), we have

$$\mathbb{E}(Y_{\geq\alpha}) = O\left(n\sum_{k=k_3}^{n}\frac{1}{k^{\frac{5}{2}}}\exp\left(-\frac{\delta k}{5}\right)\right).$$

Hence, if $\alpha \geq 10\log\left(|\epsilon|^3 n\right)$, then

$$\begin{aligned}
\mathbb{E}(Y_{\geq\alpha}) &= O\left(n\sum_{k=k_3}^{n}\frac{1}{k^{\frac{5}{2}}}\exp\left(-\frac{\delta k}{5}\right)\right) = O\left(n\sum_{k\geq\alpha/\delta}\frac{1}{k^{\frac{5}{2}}}\exp\left(-\frac{\delta k}{5}\right)\right) = O\left(e^{-\frac{\alpha}{10}}n\sum_{k\geq\alpha/\delta}\frac{1}{k^{\frac{5}{2}}}\exp\left(-\frac{\delta k}{10}\right)\right)\\
&= O\left(e^{-\frac{\alpha}{10}}\frac{n}{(\frac{\alpha}{\delta})^{\frac{5}{2}}}\sum_{k\geq\alpha/\delta}\exp\left(-\frac{\delta k}{10}\right)\right) = O\left(e^{-\frac{\alpha}{10}}\frac{n\delta^{\frac{5}{2}}e^{-\frac{\alpha}{10}}}{\left(\log\left(|\epsilon|^3 n\right)\right)^{\frac{5}{2}}\left(1-e^{-\frac{\delta}{10}}\right)}\right) = O\left(e^{-\frac{\alpha}{10}}\frac{n\delta^{\frac{5}{2}}}{\delta\left(\log\left(|\epsilon|^3 n\right)\right)^{\frac{5}{2}}|\epsilon|^3 n}\right)\\
&= O\left(e^{-\frac{\alpha}{10}}\frac{1}{\left(\log\left(|\epsilon|^3 n\right)\right)^{\frac{5}{2}}}\right) \leq e^{-\Omega(\alpha)}.
\end{aligned}$$

Finally, if $\alpha \leq 10\log\left(|\epsilon|^3 n\right) := \hat{\alpha}$, let $k_4 = \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + \hat{\alpha}\right)$. We can argue as in the first part that

$$\mathbb{E}(Y_{\alpha,\hat{\alpha}}) = e^{-\Omega(\alpha)},$$

since as in (17), as long as $k = \frac{1}{\delta}\left(\log\left(|\epsilon|^3 n\right) - \frac{5}{2}\log\log\left(|\epsilon|^3 n\right) + \alpha\right) = \Theta\left(\frac{\log|\epsilon|^3 n}{\delta}\right)$ we have that $e^{-\delta k}k^{-\frac{5}{2}} = \Theta\left(\delta e^{-\delta\alpha}n^{-1}\right)$. It follows that,

$$\mathbb{E}(Y_{\geq\alpha}) = \mathbb{E}(Y_{\alpha,\hat{\alpha}}) + \mathbb{E}(Y_{\geq\hat{\alpha}}) = e^{-\Omega(\alpha)} + e^{-\Omega(\hat{\alpha})} = e^{-\Omega(\alpha)},$$

and so the result follows from Markov's inequality. ☐

Secondly, as indicated in Theorem 1.8, we show that whp there are no unicyclic components in $G(n,n,p)$ of order significantly larger than $\frac{1}{\delta}$, and moreover, that the number of unicyclic components of order around this tends to a Poisson distribution.

Proof of Theorem 1.8.
Part (i)**:** Let us write $s_i = \frac{u_i}{\delta}$ for $i \in \{1,2\}$. We first note that, by Lemma 3.7, $G(n,n,p)$ contains no unbalanced non-tree components of order $\geq s_1$ with probability $1-e^{-\Omega(s_1)}$, and hence whp $Z_{u_1,u_2} = Z'_{u_1,u_2}$ where Z'_{u_1,u_2} is the number of unicyclic balanced components with order between s_1 and s_2.

Let us write $B_k = \left\{(i,j)\in\mathbb{N}^2\colon\, i+j=k \text{ and } \frac{1}{2}\leq\frac{i}{j}\leq 2\right\}$. Since for $s_1 \leq k \leq s_2$ we have that $\frac{k^3}{n^2}$, $\frac{\epsilon k^2}{n}$, $\frac{k^2}{n^2}$ and $\frac{k^4}{n^3}$ are all $o(1)$, it follows from (10) and Lemma 2.2 that

$$\begin{aligned}
\mathbb{E}(Z'_{u_1,u_2}) = \sum_{k=s_1}^{s_2}\sum_{(i,j)\in B_k}\mathbb{E}\left(X(i,j,0)\right) &\approx \frac{1}{4\sqrt{2\pi}}\sum_{k=s_1}^{s_2}\sqrt{k}e^{-\delta k}\sum_{(i,j)\in B_k}\frac{1}{ij}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right)\\
&\approx \frac{1}{2}\sum_{k=s_1}^{s_2}\frac{1}{k}e^{-\delta k} \approx \frac{1}{2}\int_{u_1}^{u_2}\frac{e^{-t}}{t}dt := \nu. \qquad (20)
\end{aligned}$$

As in Theorem 1.7 (i) a similar argument shows that $\mathbb{E}\left((Z'_{u_1,u_2})_i\right) \approx \nu^i$ for all $i \in \mathbb{N}$ and hence Z'_{u_1,u_2}, and so also Z_{u_1,u_2}, converges in distribution to $\mathrm{Po}(\nu)$.

Part (ii)**:** Let $s_3 = \frac{\alpha}{\delta}$ and let $Z_{\geq\alpha}$ and $Z'_{\geq\alpha}$ be the number of unicyclic components and balanced unicyclic components respectively of order at least s_3. Note that, as before, $Z_{\geq\alpha} = Z'_{\geq\alpha}$ with probability $1-e^{-\Omega(s_3)} = 1-e^{-\Omega(\alpha)}$.

A similar argument as in Theorem 1.7 (ii) shows that for any $i+j=k\leq n$

$$\mathbb{E}\left(Z'_{\geq\alpha}\right) = O\left(\sum_{k=s_3}^{n}\sqrt{k}\exp\left(-\delta k - \frac{k^3}{24n^2} + \frac{\epsilon k^2}{4n}\right)\sum_{i+j=k}\frac{1}{ij}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right)\right) = O\left(\sum_{k=s_3}^{n}\frac{1}{k}e^{-\frac{\delta k}{5}}\right). \qquad (21)$$

On the other hand, it can be shown, see for example [1, Formulas 5.1.1 and 5.1.20], that

$$E_1(x) := \int_x^\infty\frac{e^{-t}}{t}dt \leq e^{-x}\log\left(1+\frac{1}{x}\right)$$

and hence

$$\sum_{k=s_3}^{n}\frac{1}{k}e^{-\frac{\delta k}{5}}\approx\int_{s_3}^{n}\frac{1}{u}e^{-\frac{\delta u}{5}}\,du=\int_{\frac{\alpha}{5}}^{\frac{\delta n}{5}}\frac{e^{-t}}{t}\,dt\le e^{-\frac{\alpha}{5}}\log\left(1+\frac{5}{\alpha}\right)=e^{-\Omega(\alpha)}\tag{22}$$

for $\alpha\ge 5$.

By (21) and (22), it follows that $\mathbb{E}\left(Z'_{\ge\alpha}\right)=e^{-\Omega(\alpha)}$. In the case where $1<\alpha\le 5$ we can use (20) to see

$$\mathbb{E}\left(Z'_{\ge\alpha}\right)=\mathbb{E}\left(Z'_{\alpha,5}\right)+\mathbb{E}\left(Z'_{\ge 5}\right)\le\frac{1}{2}E_1(\alpha)+e^{-\Omega(1)}\le e^{-\Omega(1)}.$$

□

Finally, as indicated in Theorem 1.9, we show that whp there are no large complex components in $G(n,n,p)$, and in fact no complex components at all in the subcritical regime.

Proof of Theorem 1.9.

Part (i)**:** To show the first part, we use the observation that, since each complex component must contain a connected subgraph of excess precisely two, it is sufficient to show that whp $G(n,n,p)$ contains no such subgraphs.

We note that any connected graph of excess precisely two consists of a pair of cycles, which are either joined by a path or whose intersection is a path. Let us denote the number of such subgraphs by Q. The key observation is that any such graph of order k can be built by taking a path on k vertices and adding an edge from each of its endpoints to another vertex in the path. Hence, we can choose such a subgraph on k vertices by first choosing the $i=\left\lfloor\frac{k}{2}\right\rfloor$ vertices of the path lying in one partition class and the $j=\left\lceil\frac{k}{2}\right\rceil$ vertices of the path lying in the other partition class, choosing the order which the vertices appear in the path in at most $i!j!$ many ways and then choosing for each endpoint of the path one of the at most k many edges from this endpoint to another vertex in the path. It follows that

$$\begin{aligned}\mathbb{E}(Q)&\le 2\sum_{k=3}^{n}\binom{n}{k}\binom{n}{k}(k!)^2k^2p^{2k+1}+2\sum_{k=2}^{k_0}\binom{n}{k}\binom{n}{k+1}(k+1)!k!k^2p^{2k+2}\\&\le 2\sum_{k=3}^{n}\frac{n^k}{k!}\frac{n^k}{k!}(k!)^2k^2\left(\frac{1+\epsilon}{n}\right)^{2k+1}+2\sum_{k=2}^{k_0}\frac{n^k}{k!}\frac{n^{k+1}}{(k+1)!}(k+1)!k!k^2\left(\frac{1+\epsilon}{n}\right)^{2k+2}\\&\le 4\sum_{k=2}^{n}\frac{k^2}{n}e^{2\epsilon k}\le\frac{4}{n}\int_0^{\infty}x^2e^{2\epsilon x}\,dx=\frac{1}{|\epsilon|^3n}.\end{aligned}$$

Therefore, by Markov's inequality, with probability at least $1-\frac{1}{|\epsilon|^3n}$ there are no complex components.

Part (ii) **:** As in Theorem 1.8 (i), let $A(k,\ell)$ and $A'(k,\ell)$ be the number of components and balanced components respectively of order k with excess $\ell\ge 1$. If we write $A=\sum_{k=1}^{n^{\frac{2}{3}}}\sum_{\ell\ge 1}A(k,\ell)$ and $A'=\sum_{k=1}^{n^{\frac{2}{3}}}\sum_{\ell\ge 1}A'(k,\ell)$, then by Lemma 3.7 with probability $1-O\left(n^{-1}\right)=1-O\left(\left(\epsilon^4n\right)^{-1}\right)$, $A=A'$.

We split the computation of $\mathbb{E}(A')$ into three cases: when k is small, when k is large and ℓ is small, and when both k and ℓ are large. More explicitly, we write

$$\begin{aligned}\mathbb{E}(A')&=\sum_{k=1}^{n^{\frac{2}{3}}}\sum_{(i,j)\in B_k}\sum_{\ell=1}^{ij-i-j}\mathbb{E}\left(X(i,j,\ell)\right)\\&=\sum_{k=1}^{\epsilon^{-\frac{1}{2}}}\sum_{(i,j)\in B_k}\sum_{\ell=1}^{ij-i-j}\mathbb{E}\left(X(i,j,\ell)\right)+\sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}}\sum_{(i,j)\in B_k}\sum_{\ell=1}^{3}\mathbb{E}\left(X(i,j,\ell)\right)+\sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}}\sum_{(i,j)\in B_k}\sum_{\ell=4}^{ij-i-j}\mathbb{E}\left(X(i,j,\ell)\right)\\&:=S_1+S_2+S_3,\end{aligned}$$

where B_k is as in the proof of Theorem 1.8.

Let us deal with S_3 first. Since $\frac{k^2}{n^2}=o(1)$ for $k\le n^{\frac{2}{3}}$, we see by (13) in Theorem 3.6 that

$$S_3\lesssim\sum_{k=1}^{n^{\frac{2}{3}}}\sqrt{k}\exp\left(-\delta k+\frac{\epsilon k^2}{4n}\right)\sum_{(i,j)\in B_k}\left(\frac{i}{j}\right)^{\frac{j-i}{2}}\exp\left(-\frac{(i-j)^2}{2n}\right)\sum_{\ell=4}^{ij-i-j}\left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}}\exp\left(\ell\log(1+\epsilon)+\frac{\ell(1+\epsilon)}{n}\right).\tag{23}$$

Let us first deal with the innermost sum of (23)

$$\sum_{\ell=4}^{ij-i-j} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(\ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right).$$

The ratio of consecutive terms in the sum is

$$\sqrt{\frac{k^3}{cn^2}} \frac{\ell^{\frac{\ell}{2}}}{(\ell+1)^{\frac{\ell+1}{2}}} \exp\left(\log(1+\epsilon) + \frac{1+\epsilon}{n}\right) < 1,$$

when ℓ is large enough compared to c. However, for any constant $\ell \geq 4$ the individual terms can be seen to have order

$$O\left(\frac{k^3}{n^2}\right)^{\frac{\ell}{2}} = O\left(\frac{k^6}{n^4}\right)$$

since $k \leq n^{\frac{2}{3}}$. It follows that

$$\sum_{\ell=4}^{ij-i-j} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(\ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right) = O\left(\frac{k^6}{n^4}\right). \tag{24}$$

Next, we see that the second sum can be evaluated using Lemma 2.2 to give

$$\sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \exp\left(-\frac{(i-j)^2}{2n}\right) = O\left(\sqrt{k}\right). \tag{25}$$

Hence, by (24) and (25), and using that, since $k \leq n^{\frac{2}{3}} = o(\epsilon n)$ we have $\frac{\epsilon k^2}{4n} = o(\delta k)$, we see that

$$S_3 = O\left(\sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}} e^{-\frac{\delta k}{2}} \frac{k^7}{n^4}\right) = O\left(\frac{1}{n^4} \sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}} k^7 e^{-\frac{\delta k}{2}}\right) = O\left(\frac{1}{n^4} \int_0^\infty x^7 e^{-\frac{\delta x}{2}}\right) = O\left(\frac{1}{(\epsilon^4 n)^4}\right).$$

Next, to bound S_2 we use (14) rather than (13), and see via similar calculations that

$$\begin{aligned} S_2 &= O\left(\sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}} \sqrt{k} \exp\left(-\delta k + \frac{\epsilon k^2}{4n}\right) \sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right) \sum_{\ell=1}^{3} \frac{(ij)^{\ell-1}}{n^\ell}\right) \\ &= O\left(\sum_{\ell=1}^{3} \frac{1}{n^\ell} \sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}} \sqrt{k} e^{-\frac{\delta k}{2}} \left(\frac{k^2}{4}\right)^{\ell-1} \sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right)\right) \\ &= O\left(\sum_{\ell=1}^{3} \frac{1}{n^\ell} \sum_{k=\epsilon^{-\frac{1}{2}}}^{n^{\frac{2}{3}}} k^{2\ell-1} e^{-\frac{\delta k}{2}}\right) = O\left(\sum_{\ell=1}^{3} \frac{1}{(\epsilon^4 n)^\ell}\right). \end{aligned}$$

Finally, let us deal with S_1. As before, since $\frac{k^2}{n^2} = o(1)$ for $k \leq n^{\frac{2}{3}}$, it follows from (13) that

$$S_1 \lesssim \sum_{k=1}^{\epsilon^{-\frac{1}{2}}} \sqrt{k} \exp\left(-\delta k + \frac{\epsilon k^2}{4n}\right) \sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \exp\left(-\frac{(i-j)^2}{2n}\right) \sum_{\ell=1}^{ij-i-j} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(\ell \log(1+\epsilon) + \frac{\ell(1+\epsilon)}{n}\right).$$

Then, since $\left(\frac{i}{j}\right)^{\frac{j-i}{2}} \leq 1$, $\frac{\epsilon k^2}{4n} = o(\delta k)$, and $\ell \leq ij \leq k^2 = o(n)$, we have that

$$\begin{aligned} S_1 &\lesssim \sum_{k=1}^{\epsilon^{-\frac{1}{2}}} \sqrt{k} \sum_{(i,j)\in B_k} \sum_{\ell=1}^{ij-i-j} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \exp\left(\ell \log(1+\epsilon) + o(1)\right) \lesssim \sum_{k=1}^{\epsilon^{-\frac{1}{2}}} \sqrt{k} \sum_{(i,j)\in B_k} \sum_{\ell=1}^{ij-i-j} \left(\frac{ck^3 e^{2\log(1+\epsilon)}}{\ell n^2}\right)^{\frac{\ell}{2}} \\ &= O\left(\sum_{k=1}^{\epsilon^{-\frac{1}{2}}} \sqrt{k} \sum_{(i,j)\in B_k} \frac{k^{\frac{3}{2}}}{n}\right) = O\left(\frac{1}{\epsilon^2 n}\right). \end{aligned}$$

Hence, if $\epsilon^4 n \to \infty$, then

$$\mathbb{E}(A') = S_1 + S_2 + S_3 = O\left(\frac{1}{\epsilon^2 n}\right) + O\left(\sum_{\ell=1}^{3} \frac{1}{(\epsilon^4 n)^\ell}\right) + O\left(\frac{1}{(\epsilon^4 n)^4}\right) = O\left(\frac{1}{\epsilon^4 n}\right),$$

and so the result follows by Markov's inequality. □

4.2. **Largest and second largest components: proof of Theorem 1.5.** In order to show that there is in fact a unique giant component in $G(n,n,p)$ for an appropriate range of ϵ, we follow a relatively standard approach. First, we estimate quite precisely the number of vertices which are contained in small tree or unicyclic components, noting that by the lemmas in the previous section there are whp no large tree or unicyclic components and no small complex components. It follows that whp all the remaining vertices are contained in large complex components, and by a sprinkling argument we are able to show that whp these vertices are in fact all contained in a single component.

Throughout this section, let us consider two quantities related to $\epsilon = \epsilon(n) > 0$ satisfying $\epsilon = o(1)$: firstly δ as defined in (1), i.e.

$$\delta = \epsilon - \log(1+\epsilon) \approx \frac{\epsilon^2}{2},$$

and secondly ϵ' as in Theorem 1.5, which is defined implicitly as the unique positive solution to

$$(1-\epsilon')e^{\epsilon'} = (1+\epsilon)e^{-\epsilon}. \tag{26}$$

We note that $\epsilon' = \epsilon + \frac{2}{3}\epsilon^2 + O(\epsilon^3)$. We also note that ϵ' has the following natural interpretation in terms of branching processes: If we consider a $\mathrm{Po}(1+\epsilon)$ branching process and condition on the event that it does not survive, then it can be shown that this model is distributed as a $\mathrm{Po}\left(1-\epsilon'\right)$ branching process. Whenever we use the terms ϵ' and δ they refer to these quantities for a fixed ϵ, which should be clear from the context.

As indicated in Theorem 1.9, in the weakly subcritical regime whp there are no complex components. However, for our proof it will be necessary to know more, namely that in this regime we do not expect to have many vertices contained in 'large' components. The proof of this fact can be deduced from a standard comparison to a branching process and we defer the details to Appendix C.

Theorem 4.1. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1-\epsilon}{n}$. Then the expected number of vertices in $G(n,n,p)$ in components of order at least $\sqrt{\frac{n}{3\epsilon}}$ is $o\left(\sqrt{\frac{n}{\epsilon}}\right)$.*

Let us begin then, by estimating the number of vertices contained in small tree or unicyclic components.

Lemma 4.2. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \gg \omega \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. Let $Y(-1)$ and $Y(0)$ denote the number of vertices in tree and unicyclic components of order at most $n^{\frac{2}{3}}$ in $G(n,n,p)$ respectively. Then with probability $1 - O\left(\omega^{-1}\right)$, we have*

$$Y(0) \le \frac{4\omega}{\delta},$$

and

$$\left| Y(-1) - \frac{2(1-\epsilon')}{1+\epsilon} n \right| \le \frac{\omega\sqrt{n}}{\sqrt{\epsilon}}.$$

Proof. First, we bound $Y(0)$. As before, we let

$$B_k = \left\{(i,j) \in \mathbb{N}^2 \colon i+j=k \text{ and } i < 2j \text{ and } j \le 2i\right\} \text{ and } U_k = \left\{(i,j) \in \mathbb{N}^2 \colon i+j=k\right\} \setminus B_k.$$

Then we can split the calculation of $\mathbb{E}(Y(0))$ into two parts

$$\mathbb{E}(Y(0)) = \sum_{k \le n^{\frac{2}{3}}} k \sum_{(i,j)\in B_k} \mathbb{E}\left(X(i,j,0)\right) + \sum_{k \le n^{\frac{2}{3}}} k \sum_{(i,j)\in U_k} \mathbb{E}\left(X(i,j,0)\right) := S_1 + S_2.$$

Since if $i+j=k$, then $\frac{\epsilon i j}{n} \le \frac{\epsilon k^2}{4n} = o(\delta k)$ for $k \le n^{\frac{2}{3}}$, it follows from (10) and Lemma 2.2 that

$$S_1 \le \frac{1}{4\sqrt{2\pi}} \sum_{k \le n^{\frac{2}{3}}} k^{\frac{3}{2}} e^{-\delta k} \sum_{(i,j)\in B_k} \frac{1}{ij}\left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n} + \frac{\epsilon i j}{n}\right) \le \frac{1}{2} \sum_{k \le n^{\frac{2}{3}}} e^{-\frac{\delta k}{2}} \le \frac{1}{2}\int_0^\infty e^{-\frac{\delta x}{2}}\, dx \le \frac{1}{\delta}.$$

Furthermore, using the very naive bound that $C(i,j,0) \le ijC(i,j-1) \le i^j j^i$, we can calculate as in (10)

$$\begin{aligned}
S_2 &\le \sum_{k\le n^{\frac{2}{3}}} k \sum_{(i,j)\in U_k} \mathbb{E}(X(i,j,0)) \le \sum_{k\le n^{\frac{2}{3}}} k \sum_{(i,j)\in U_k} \binom{n}{i}\binom{n}{j} i^j j^i p^k (1-p)^{kn-ij-k} \\
&\approx \frac{1}{2\pi} \sum_{k\le n^{\frac{2}{3}}} k e^{-\delta k} \sum_{(i,j)\in U_k} \frac{1}{(ij)^{\frac{1}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n} + \frac{\epsilon i j}{n}\right) \\
&\le \sum_{k\le n^{\frac{2}{3}}} k e^{-\frac{\delta k}{2}} \left(\frac{1}{2}\right)^{\frac{k}{3}} \sum_{(i,j)\in U_k} \frac{1}{(ij)^{\frac{1}{2}}} \le \sum_{k\le n^{\frac{2}{3}}} k^{\frac{3}{2}} \left(\frac{1}{2}\right)^{\frac{k}{3}} = O(1).
\end{aligned}$$

The first part of the lemma then follows by Markov's inequality.

So, let us consider the bound on $Y(-1)$. Firstly, we note that by part (ii) of Theorem 1.7 with $\alpha(n) = \sqrt{\frac{\epsilon^3 n}{16}}$, with probability $1 - o\left(e^{-\sqrt{\frac{\epsilon^3 n}{16}}}\right) = 1 - O\left(\omega^{-1}\right)$ there are no tree components in $G(n,n,p)$ of order at least $\sqrt{\frac{n}{3\epsilon}}$, and by Lemma 3.8 with probability $1 - o\left(n^{-1}\right) = 1 - O\left(\omega^{-1}\right)$ there are no non-ϵ-uniform tree components in $G(n,n,p)$ of order at most $n^{\frac{2}{3}}$. Hence, with probability $1 - O\left(\omega^{-1}\right)$, $Y(-1) = Z_1$ where Z_1 is, as in Lemma 3.9, the number of vertices in ϵ-uniform tree components in $G(n,n,p)$ of order at most $\tilde{k} = \sqrt{\frac{n}{3\epsilon}}$.

Next, following a technique of Bollobás [5, Theorem 6.6], we consider the model $G(n,n,p')$ where $p' = \frac{1-\epsilon'}{n}$. Let us write $Y'(-1)$ and $Y'(0)$ for the number of vertices in tree and unicyclic components in $G(n,n,p')$ of order at most $n^{\frac{2}{3}}$ respectively, and similarly Z_1' for the number of vertices in ϵ-uniform tree components in $G(n,n,p')$ of order at most $\sqrt{\frac{n}{3\epsilon}}$.

We will show that almost every vertex in $G(n,n,p')$ lies in ϵ-uniform tree components of order at most $\sqrt{\frac{n}{3\epsilon}}$, and we are able to calculate the ratio $\mathbb{E}(Z_1)/\mathbb{E}(Z_1')$ quite precisely. Combining this with the bound on the variance of Z_1 from Lemma 3.9 we are able to deduce the second part of the lemma.

Indeed, by Theorem 4.1 the expected number of vertices in components of order greater than $\sqrt{\frac{n}{3\epsilon}}$ in $G(n,n,p')$ is $o\left(\sqrt{\frac{n}{\epsilon}}\right)$. Furthermore, similar calculations as in the first part of this lemma, show that the expected number of vertices in unicyclic and complex components of order at most $n^{\frac{2}{3}}$ in $G(n,n,p')$ is $o\left(\sqrt{\frac{n}{\epsilon}}\right)$.

Indeed, if we let $Y'(\ge 1)$ be the number of vertices in complex components of order at most $n^{\frac{2}{3}}$ in $G(n,n,p')$, then

$$\mathbb{E}\left(Y'(\ge 1)\right) = \sum_{k\le n^{\frac{2}{3}}} k \sum_{(i,j)\in B_k} \sum_{\ell=1}^{ij-i-j} \mathbb{E}\left(X(i,j,\ell)\right) + \sum_{k\le n^{\frac{2}{3}}} k \sum_{(i,j)\in U_k} \sum_{\ell=1}^{ij-i-j} \mathbb{E}\left(X(i,j,\ell)\right) := S_1' + S_2'.$$

One can bound S_2' in a similar fashion as with S_2, since the exponentially small term $\left(\frac{i}{j}\right)^{j-i}$ is the dominating term. For S_1' we use (15), and as in Theorem 1.9 we can argue

$$\begin{aligned}
S_1' &\le \sum_{k\le n^{\frac{2}{3}}} k \sum_{(i,j)\in B_k} \sum_{\ell=1}^{ij-i-j} \sqrt{k} e^{-\delta k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} \frac{(n)_i (n)_j}{n^k} (1-\epsilon')^{k+\ell} \left(1 - \frac{1-\epsilon'}{n}\right)^{kn-ij-k-\ell} \\
&\le \sum_{k\le n^{\frac{2}{3}}} k^{\frac{3}{2}} (1-\epsilon')^k \sum_{(i,j)\in B_k} \sum_{\ell=1}^{ij-i-j} \left(\frac{ck^3}{\ell n^2}\right)^{\frac{\ell}{2}} = O\left(\frac{1}{n} \sum_{k\le n^{\frac{2}{3}}} k^4 e^{-\epsilon' k}\right) = O\left(\frac{1}{\epsilon^5 n}\right) = o\left(\sqrt{\frac{n}{\epsilon}}\right),
\end{aligned}$$

as long as $\epsilon^3 n \to \infty$. The calculations for the expected number of vertices in small unicyclic components are similar. Finally, the expected number of vertices in non-ϵ-uniform components of order at most $n^{\frac{2}{3}}$ in $G(n,n,p')$ is $o(1)$, as follows from the proof of Lemma 3.8. It follows that

$$\mathbb{E}\left(Z_1'\right) = 2n - o\left(\sqrt{\frac{n}{\epsilon}}\right).$$

Let us write $Z_1(k)$ and $Z_1'(k)$ for the number of vertices in ϵ-uniform tree components of order $k \le \sqrt{\frac{n}{3\epsilon}}$ in $G(n,n,p)$ and $G(n,n,p')$ respectively, and let us consider the ratio

$$\frac{\mathbb{E}\left(Z_1(k)\right)}{\mathbb{E}\left(Z_1'(k)\right)} = \frac{\sum_{i+j=k}\binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1} \left(\frac{1+\epsilon}{n}\right)^{k-1}\left(1-\frac{1+\epsilon}{n}\right)^{kn-ij-k+1}}{\sum_{i+j=k}\binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1} \left(\frac{1-\epsilon'}{n}\right)^{k-1}\left(1-\frac{1-\epsilon'}{n}\right)^{kn-ij-k+1}},$$

where the sums run over the ϵ-uniform pairs (i,j).

Note that,

$$\left(1-\frac{1+\epsilon}{n}\right)^{kn-ij-k+1} = \left(\frac{n-1}{n}\right)^{kn-ij-k+1}\left(1-\frac{\epsilon}{n-1}\right)^{kn-ij-k+1} = \left(\frac{n-1}{n}\right)^{kn-ij-k+1}\exp\left(-\epsilon k + \frac{\epsilon i j}{n} + o(1)\right),$$

and similarly

$$\left(1-\frac{1-\epsilon'}{n}\right)^{kn-ij-k+1} = \left(\frac{n-1}{n}\right)^{kn-ij-k+1}\exp\left(\epsilon' k - \frac{\epsilon' i j}{n} + o(1)\right).$$

Hence, we have

$$\frac{\mathbb{E}(Z(k))}{\mathbb{E}(Z'(k))} = \frac{1-\epsilon'}{1+\epsilon}\left(\frac{(1+\epsilon)e^{-\epsilon}}{(1-\epsilon')e^{\epsilon'}}\right)^k \frac{e^{O\left(\frac{\epsilon k^2}{n}\right)}}{e^{-O\left(\frac{\epsilon' k^2}{n}\right)}} \frac{\sum_{i+j=k}\binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1}\left(\frac{n-1}{n}\right)^{kn-ij-k+1}}{\sum_{i+j=k}\binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1}\left(\frac{n-1}{n}\right)^{kn-ij-k+1}} = \frac{1-\epsilon'}{1+\epsilon} e^{O\left(\frac{\epsilon k^2}{n}\right)},$$

since the second term is equal to 1 by the definition of ϵ' in (26).

So, we see that

$$\frac{\mathbb{E}\left(Z_1(k)\right)}{\mathbb{E}\left(Z_1'(k)\right)} = \frac{1-\epsilon'}{1+\epsilon} + O\left(\frac{\epsilon k^2}{n}\right),$$

or, in other words,

$$\mathbb{E}\left(Z_1(k)\right) = \frac{1-\epsilon'}{1+\epsilon}\mathbb{E}\left(Z_1'(k)\right) + O\left(\frac{\epsilon k^2}{n}\right)\mathbb{E}\left(Z_1'(k)\right).$$

Hence, by writing $\mathbb{E}\left(Z'(k)\right) = k\mathbb{E}\left(\hat{Y}(k)\right)$ where $\hat{Y}(k)$ is the number of ϵ-uniform tree components of order k in $G(n,n,p')$ we see that

$$\begin{aligned}\mathbb{E}(Z_1) &= \sum_{k=1}^{\sqrt{\frac{n}{3\epsilon}}}\mathbb{E}(Z_1(k)) = \frac{1-\epsilon'}{1+\epsilon}\sum_{k=1}^{\sqrt{\frac{n}{3\epsilon}}}\mathbb{E}\left(Z_1'(k)\right) + O\left(\frac{\epsilon}{n}\right)\sum_{k=1}^{\sqrt{\frac{n}{3\epsilon}}} k^3\mathbb{E}\left(\hat{Y}(k)\right) = \frac{1-\epsilon'}{1+\epsilon}\mathbb{E}\left(Z_1'\right) + O\left(\frac{\epsilon}{n}\right)\sum_{k=1}^{\sqrt{\frac{n}{3\epsilon}}} k^3\mathbb{E}\left(\hat{Y}(k)\right) \\ &= \frac{1-\epsilon'}{1+\epsilon}\left(2n - o\left(\sqrt{\frac{n}{\epsilon}}\right)\right) + O\left(\frac{\epsilon}{n}\right)\sum_{k=1}^{\sqrt{\frac{n}{3\epsilon}}} k^3\mathbb{E}\left(\hat{Y}(k)\right). \end{aligned} \tag{27}$$

The final sum can be bounded by the corresponding sum over all possible tree components, ϵ-uniform or not. That is, writing $X'(i,j,-1)$ for the number of tree components with i vertices in one partition class and j vertices in the other in $G(n,n,p')$ and noting that $\delta = \epsilon - \log(1+\epsilon) = \log(1-\epsilon') - \epsilon'$, we can bound in a similar manner to (7)

$$\begin{aligned}\sum_{k=1}^{n^{\frac{2}{3}}} k^3\mathbb{E}\left(Y'(k)\right) &\le \sum_{k=1}^{n^{\frac{2}{3}}} k^3 \sum_{i+j=k}\mathbb{E}\left(X'(i,j,-1)\right) \le \sum_{k=1}^{n^{\frac{2}{3}}} k^3 \sum_{i+j=k}\binom{n}{i}\binom{n}{j} i^{j-1} j^{i-1} p'^{k-1}(1-p')^{kn-ij-k+1} \\ &\le \sum_{k=1}^{n^{\frac{2}{3}}} k^3 n e^{-\delta k}\sum_{i+j=k}\frac{1}{(ij)^{\frac{3}{2}}}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right) := S.\end{aligned}$$

To bound S, we split into two cases. Let us take s such that $s^{\frac{7}{2}} = \delta^{-\frac{3}{2}}$, and first consider the case when $k \le s$, where

$$S_1 := n\sum_{k=1}^{s} k^3 e^{-\delta k}\sum_{i+j=k}\frac{1}{(ij)^{\frac{3}{2}}}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right) \le n\sum_{k=1}^{s} k^3 \sum_{i+j=k}\frac{1}{k^{\frac{3}{2}}} \le n\sum_{k=1}^{s} k^{\frac{5}{2}} \le n s^{\frac{7}{2}} = O\left(\frac{n}{\delta^{\frac{3}{2}}}\right) = o\left(\left(\frac{n}{\epsilon}\right)^{\frac{3}{2}}\right).$$

Conversely, when $k \ge s$,we see that, by Lemma 2.2

$$S_2 := n \sum_{k=s}^{\sqrt{\frac{n}{3\epsilon}}} k^3 e^{-\delta k} \sum_{i+j=k} \frac{1}{(ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right) \le n \sum_{k=s}^{\sqrt{\frac{n}{3\epsilon}}} \sqrt{k} e^{-\delta k} \le n \int_0^\infty \sqrt{x} e^{-\delta x} dx = O\left(\frac{n}{\delta^{\frac{3}{2}}}\right) = o\left(\left(\frac{n}{\epsilon}\right)^{\frac{3}{2}}\right).$$

Hence, $S = S_1 + S_2 = o\left(\left(\frac{n}{\epsilon}\right)^{\frac{3}{2}}\right)$ and so by (27)

$$\mathbb{E}(Z_1) = \frac{1-\epsilon'}{1+\epsilon}\left(2n - o\left(\sqrt{\frac{n}{\epsilon}}\right)\right) + O\left(\frac{\epsilon}{n}\right) o\left(\left(\frac{n}{\epsilon}\right)^{\frac{3}{2}}\right) = \frac{2(1-\epsilon')}{1+\epsilon} n + o\left(\sqrt{\frac{n}{\epsilon}}\right).$$

Finally, by Lemma 3.9 with $\tilde{k} = \sqrt{\frac{n}{3\epsilon}}$, which can be seen to satisfy $\frac{3\epsilon\tilde{k}^2}{n} \le 1$, we conclude that $\mathrm{Var}(Z_1) = O\left(\frac{n}{\epsilon}\right)$. Hence, by Chebyshev's inequality, $\left|Z_1 - \frac{2(1-\epsilon')}{1+\epsilon} n\right| \le \omega n^{\frac{1}{2}} \epsilon^{-\frac{1}{2}}$ with probability $1 - O\left(\omega^{-1}\right)$. Thus, with probability $1 - O\left(\omega^{-1}\right)$

$$\left| Y(-1) - \frac{2(1-\epsilon')}{1+\epsilon} n \right| \le \omega n^{\frac{1}{2}} \epsilon^{-\frac{1}{2}}.$$

□

Using Lemma 4.2, we can give a good bound on the number of vertices which are contained in components of order at least $n^{\frac{2}{3}}$ in $G(n,n,p)$. Then, using a sprinkling argument we can deduce that whp all these vertices are contained in a unique 'giant' component, and determine asymptotically its order.

Proof of Theorem 1.5. Let $\mathscr{L}(G)$ denote the set of vertices lying in components of G of order at least $n^{\frac{2}{3}}$, which we call *large*. We first estimate quite precisely the size of $\mathscr{L}(G(n,n,p))$ and then show that there is only one large component in $G(n,n,p)$.

Indeed, if we let $\omega = \left(\epsilon^4 n\right)^{\frac{1}{6}}$, then by part (ii) of Theorem 1.9 with probability $1 - O\left(\left(\epsilon^4 n\right)^{-1}\right) \ge 1 - \omega^{-1}$ there are no small complex components in $G(n,n,p)$. We note that this is the only point in the proof that it is necessary to assume that $\epsilon^4 n \to \infty$, and in what follows we could instead take $\omega = c(\epsilon^3 n)^{\frac{1}{6}}$ for an appropriately small constant c.

Now, by Lemma 4.2, with probability $1 - O\left(\omega^{-1}\right)$ the number of vertices in small unicyclic components is at most $\frac{\omega}{\delta} \ll n^{\frac{2}{3}}$ and the number of vertices in small tree components $Y(-1)$ is such that

$$\frac{2(1-\epsilon')}{1+\epsilon} n - \frac{n^{\frac{2}{3}}}{100} \le \frac{2(1-\epsilon')}{1+\epsilon} n - \frac{\omega\sqrt{n}}{\sqrt{\epsilon}} \le Y(-1) \le \frac{2(1-\epsilon')}{1+\epsilon} n + \frac{\omega\sqrt{n}}{\sqrt{\epsilon}} \le \frac{2(1-\epsilon')}{1+\epsilon} n + \frac{n^{\frac{2}{3}}}{100}.$$

It follows that with probability $1 - O\left(\omega^{-1}\right)$

$$\left| |\mathscr{L}(G(n,n,p))| - 2n\left(1 - \frac{1-\epsilon'}{1+\epsilon}\right) \right| \le \frac{n^{\frac{2}{3}}}{50}.$$

Note that $\epsilon' = \epsilon + O(\epsilon^2)$, and so $|\mathscr{L}(G(n,n,p))| \approx 4\epsilon n$.

In order to show the existence of a unique large component, we use a sprinkling argument. Let

$$p_1 = p - \frac{n^{-\frac{4}{3}}}{10} \quad \text{and} \quad p_2 = \frac{p-p_1}{1-p_1} \ge \frac{n^{-\frac{4}{3}}}{20},$$

and let us write $p_1 = \frac{1+\epsilon_1}{n}$. A standard argument allows us to couple an independent pair $(G(n,n,p_1), G(n,n,p_2))$ with $G(n,n,p)$ so that $G(n,n,p_1) \cup G(n,n,p_2) = G(n,n,p)$.

If we let $\epsilon_1 = \epsilon - \frac{1}{10n^{\frac{1}{3}}}$, then it is clear that $\omega = \left(\epsilon^4 n\right)^{\frac{1}{6}} \approx \left(\epsilon_1^4 n\right)^{\frac{1}{6}}$. Hence, the same argument as before shows that with probability $1 - O\left(\omega^{-1}\right)$

$$\left| |\mathscr{L}(G(n,n,p_1))| - 2n\left(1 - \frac{1-\epsilon_1'}{1+\epsilon_1}\right) \right| \le \frac{n^{\frac{2}{3}}}{50},$$

where ϵ_1' is defined as the solution to $(1-\epsilon_1')e^{\epsilon_1'} = (1+\epsilon_1)e^{-\epsilon_1}$.

Next, some basic analysis tells us that these bounds for $|\mathscr{L}(G(n,n,p))|$ and $|\mathscr{L}(G(n,n,p_1))|$ are not far apart. More precisely, we claim that

$$\frac{1-\epsilon_1'}{1+\epsilon_1} - \frac{1-\epsilon'}{1+\epsilon} \le \frac{4}{10n^{\frac{1}{3}}}.$$

Indeed, consider the function $y = g(x)$ where y is given as the unique positive solution to $(1-y)e^y = (1+x)e^{-x}$. Then, by the derivative of implicit functions formula, $g_x' = \frac{x}{y}e^{-x-y}$. Thus, by Lagrange's theorem there is a $\psi \in [\epsilon_1, \epsilon]$ such that

$$\epsilon' - \epsilon_1' = g(\epsilon) - g(\epsilon_1) = g'(\psi)(\epsilon - \epsilon_1) < 2(\epsilon - \epsilon_1),$$

since $g'(\psi) = \frac{\psi)}{g(\psi)}e^{-\psi - g(\psi)} < \frac{\psi}{g(\psi)} = \frac{\psi}{\psi + O(\psi^2)} \le 2$.

Hence, it follows that

$$\frac{1-\epsilon_1'}{1+\epsilon_1} - \frac{1-\epsilon'}{1+\epsilon} = \frac{\epsilon - \epsilon_1 + \epsilon' - \epsilon_1' - \epsilon_1'\epsilon + \epsilon'\epsilon_1}{(1+\epsilon)(1+\epsilon_1)} \le 3(\epsilon - \epsilon_1) + \epsilon'\epsilon_1 - \epsilon_1'\epsilon \le 3(\epsilon-\epsilon_1) + \epsilon(\epsilon' - \epsilon_1') \le (3+2\epsilon)(\epsilon - \epsilon_1) \le \frac{4}{10n^{\frac{1}{3}}}.$$

Hence with probability $1 - O\left(\omega^{-1}\right)$

$$|\mathscr{L}(G(n,n,p))| - |\mathscr{L}(G(n,n,p_1))| \le \frac{1-\epsilon_1'}{1+\epsilon_1}2n - \frac{1-\epsilon'}{1+\epsilon}2n + \frac{n^{\frac{2}{3}}}{25} \le \frac{4}{5}n^{\frac{2}{3}} + \frac{n^{\frac{2}{3}}}{25} < n^{\frac{2}{3}}.$$

Since, by our coupling, $G(n,n,p_1) \subseteq G(n,n,p)$, it follows that in this event every large component of $G(n,n,p)$ contains a large component of $G(n,n,p_1)$. Hence, in order to show that there is a unique large component in $G(n,n,p)$ it is be sufficient to show that all the large components in $G(n,n,p_1)$ are contained in a single component in $G(n,n,p)$.

By Lemma 3.7, with probability $1 - O\left(n^{-1}\right) \ge 1 - O\left(\omega^{-1}\right)$, each component of order at least $n^{\frac{2}{3}}$ in $G(n,n,p_1)$ is balanced, and so we can partition the vertices in $\mathscr{L}(G(n,n,p_1))$ into subsets $V_1, W_1, V_2, W_2, \dots V_m, W_m$ such that $\frac{n^{\frac{2}{3}}}{3} \le |V_i|, |W_i| \le n^{\frac{2}{3}}$ and V_i and W_i lie in the same component in $G(n,n,p_1)$ for each i, say in a greedy manner.

Now, let us consider the edges in $G(n,n,p_2)$. Either all vertices in $\mathscr{L}(G(n,n,p_1))$ are contained in one component of $G(n,n,p_1) \cup G(n,n,p_2)$, or there is a family $\mathscr{A} = \left\{(V_{i_1}, W_{i_1}), (V_{i_2}, W_{i_2}), \dots, (V_{i_r}, W_{i_r})\right\}$, where $1 \le r \le \frac{m}{2}$ such that there is no edge in $G(n,n,p_2)$ with one end point in $(V_i, W_i) \in \mathscr{A}$ and the other in $(V_j, W_j) \notin \mathscr{A}$ (see Figure 2). Note that, for any such family $\mathscr{A}$, there are at least $\frac{2}{9}r(m-r)n^{\frac{4}{3}}$ many edges with one end point in $(V_i, W_i) \in \mathscr{A}$ and the other in $(V_j, W_j) \notin \mathscr{A}$.

Hence, the probability that such a family $\mathscr{A}$ exists is bounded by

$$\sum_{r=1}^{\frac{m}{2}} \binom{m}{r} (1-p_2)^{\frac{2}{9}r(m-r)n^{\frac{4}{3}}} \le \sum_{r=1}^{\frac{m}{2}} \left(\frac{em}{r}\right)^r \left(1 - \frac{n^{-\frac{4}{3}}}{20}\right)^{\frac{2}{9}r(m-r)n^{\frac{4}{3}}} \le \sum_{r=1}^{\frac{m}{2}} \left(\frac{em}{r}e^{-\frac{m-r}{100}}\right)^r \le \sum_{r=1}^{\frac{m}{2}} \left(\frac{em}{r}e^{-\frac{m}{200}}\right)^r.$$

However, since $|\mathscr{L}(G(n,n,p_1))| \approx 4\epsilon n$, it follows that $m = \Theta\left(\epsilon n^{\frac{1}{3}}\right) = \omega(1)$, and hence

$$\sum_{r=1}^{\frac{m}{2}} \left(\frac{em}{r}e^{-\frac{m}{200}}\right)^r = e^{-\Omega(m)} = O\left(\omega^{-1}\right).$$

It follows that, with probability $1 - O\left(\omega^{-1}\right)$, $\mathscr{L}(G(n,n,p))$ consists of just the vertices in the largest component $L_1\left(G(n,n,p)\right)$, and so the claim follows.

For the last part, since with probability $1 - O\left(\omega^{-1}\right)$,

$$|L_1| \approx \frac{2(\epsilon + \epsilon')}{1+\epsilon}n \approx 4\epsilon n,$$

and by Theorems 1.7 and 1.8, with probability $1 - O\left(\omega^{-1}\right)$, there are no large tree or unicyclic components, it suffices to show that with sufficiently small probability there are no complex components in $G(n,n,p)$ of order around $4\epsilon n$ which are very unbalanced. We shall bound from above the expected number of complex components C of $G(n,n,p)$ with order in the interval $[3\epsilon n, 5\epsilon n]$, which have $|C \cap N_1| \ge (1+2\sqrt{\epsilon})|C \cap N_2|$ or $|C \cap N_2| \ge (1+2\sqrt{\epsilon})|C \cap N_1|$. As in Lemma 3.7 we can bound the expected number of such components by the

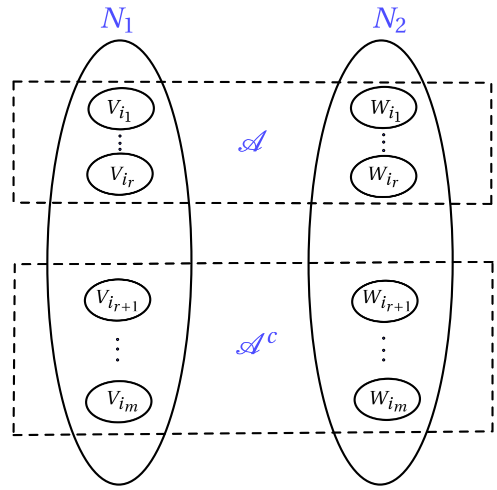

FIGURE 2. A partition of the vertices in $\mathscr{L}(G(n,n,p_1))$ into $\mathscr{A}$ and $\mathscr{A}^c$ with no edges between $V_{i_s} \in \mathscr{A}$ and $W_{i_t} \in \mathscr{A}^c$ or between $V_{i_s} \in \mathscr{A}^c$ and $W_{i_t} \in \mathscr{A}$.

expected number of trees with 2 extra edges, otherwise disconnected from the rest of the graph, which can be bounded as in Lemma 3.7

$$\sum_{k=3\epsilon n}^{5\epsilon n} \sum_{\substack{i+j=k,\\ j\geq(1+2\sqrt{\epsilon})i}} \binom{n}{i}\binom{n}{j} i^{j+1} j^{i+1} p^{k+1}(1-p)^{kn-2ij} \leq \frac{1}{n}\sum_{k=3\epsilon n}^{5\epsilon n} \sum_{\substack{i+j=k,\\ j\geq(1+2\sqrt{\epsilon})i}} (ij)^{\frac{1}{2}}\left(\frac{i}{j}\right)^{j-i} e^{\frac{(1+2\epsilon)ij}{n}}$$
$$\leq \frac{1}{n}\sum_{k=3\epsilon n}^{5\epsilon n}\left(\frac{1}{1+2\sqrt{\epsilon}}\right)^{\frac{\sqrt{\epsilon}}{1+2\sqrt{\epsilon}}k} e^{\frac{(1+2\epsilon)k^2}{4n}} \sum_{\substack{i+j=k,\\ j\geq(1+2\sqrt{\epsilon})i}} (ij)^{\frac{1}{2}} \leq \frac{1}{n}\sum_{k=3\epsilon n}^{5\epsilon n} k^2 \exp\left(-\frac{2\epsilon}{(1+2\sqrt{\epsilon})^2}k + \frac{5(1+2\epsilon)\epsilon}{4}k\right)$$
$$\leq \frac{1}{n}\sum_{k=3\epsilon n}^{5\epsilon n} k^2 e^{-\Omega(\epsilon k)} = O\left(\frac{1}{\epsilon^3 n}\right) = O\left(\frac{1}{\omega}\right).$$

Hence, the result follows from Markov's inequality. □

4.3. **The excess of the giant component: proof of Theorem 1.6.** Using Theorem 1.5, we can quite easily give a bound on the excess of the giant component which is of the correct asymptotic order. Indeed, we can bound the order of the giant component in quite a small interval, and then using Theorem 3.5 we can bound the probability that any component of this order has too large an excess.

This is enough to show that whp the excess of the giant component is $O(\epsilon^3 n)$. Note that, this can be seen to be of the correct order by a simple sprinkling argument: If we take $p_1 = \frac{1+\frac{\epsilon}{2}}{n}$ and $p_2 = \frac{p-p_1}{1-p_1} \geq \frac{\epsilon}{2n}$ then our previous results imply that, for an appropriate range of ϵ, whp there is a giant component of order $\Theta(\epsilon n)$ in $G(n,n,p_1)$ which is equally distributed across the partition classes. However, then whp there are $\Theta\left((\epsilon n)^2 p_2\right) = \Theta(\epsilon^3 n)$ many edges of $G(n,n,p_2)$ on the vertex set of the giant component.

In order to find the *correct leading constant*, we follow an argument of Łuczak [17] and use a multi-round exposure argument, starting with a supercritical p' which is significantly smaller than p. By our weaker bound on the excess we can show that at the start of our process the excess of the giant component in $G(n,n,p')$ is $o(\epsilon^3 n)$, and we can also estimate quite precisely the change in the excess of the giant component between each stage of the multi-round exposure as we increase p' to p, giving us an asymptotically tight bound on the excess of the giant.

So, let us begin by deriving our weak upper bound on the excess of the giant component.

Lemma 4.3. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^4 n \gg \omega \to \infty$ and $\epsilon \leq \frac{1}{\omega}$, and let $p = \frac{1+\epsilon}{n}$. Then with probability $1 - O\left((\epsilon^4 n)^{-\frac{1}{6}}\right)$ the excess of the largest component in $G(n,n,p)$ is $O(\epsilon^3 n)$.*

Proof. We first note that, by Theorem 1.5, with probability $1-O\left(\left(\epsilon^4 n\right)^{-\frac{1}{6}}\right)$, the largest component L_1 of $G(n,n,p)$ is balanced and satisfies

$$\left|L_1 - \frac{2(\epsilon+\epsilon')}{1+\epsilon}n\right| < \frac{n^{\frac{2}{3}}}{50}.$$

Let X be the number of balanced components in $G(n,n,p)$ of order between

$$\frac{2(\epsilon+\epsilon')}{1+\epsilon}n - \frac{n^{\frac{2}{3}}}{50} := k_1 \le k \le k_2 =: \frac{2(\epsilon+\epsilon')}{1+\epsilon}n + \frac{n^{\frac{2}{3}}}{50},$$

which have excess at least $C\epsilon^3 n$, where we will choose C sufficiently large later. Then $\mathbb{E}(X)$ can be bounded above using Theorem 3.6 as

$$\mathbb{E}(X) \lesssim \sum_{k=k_1}^{k_2} \sqrt{k}\exp\left(-\delta k + \frac{\epsilon k^2}{4n}\right) \sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \exp\left(-\frac{(i-j)^2}{2n}\right) \sum_{\ell=C\epsilon^3 n}^{ij-k} \left(\frac{ck^3(1+\epsilon)^2 e^{\frac{1+\epsilon}{2n}}}{\ell n^2}\right)^{\frac{\ell}{2}},$$

where B_k is as before the set of balanced pairs (i,j), since for $k_1 \le k \le k_2$ we have $\frac{k^2}{n^2} = o(1)$.

Let us first deal with the innermost sum. Since $k = \Theta(\epsilon n)$, for large enough C we can bound

$$\sum_{\ell=C\epsilon^3 n}^{ij-k} \left(\frac{ck^3(1+\epsilon)^2 e^{\frac{1+\epsilon}{2n}}}{\ell n^2}\right)^{\frac{\ell}{2}} \le \sum_{\ell=C\epsilon^3 n}^{ij-k} \left(\frac{1}{e^2}\right)^{\frac{\ell}{2}} = O\left(e^{-C\epsilon^3 n}\right).$$

The middle sum can be dealt with by Lemma 2.2 as usual to see that

$$\sum_{(i,j)\in B_k} \left(\frac{i}{j}\right)^{\frac{j-i}{2}} \exp\left(-\frac{(i-j)^2}{2n}\right) = O\left(\sqrt{k}\right).$$

Therefore, we can bound

$$\mathbb{E}(X) = O\left(\sum_{k=k_1}^{k_2} k\exp\left(-\delta k + \frac{\epsilon k^2}{4n} - C\epsilon^3 n\right)\right).$$

However, since $k = \Theta(\epsilon n)$, and so both δk and $\frac{\epsilon k^2}{n}$ are $O(\epsilon^3 n)$, for C large enough

$$\mathbb{E}(X) = O\left((k_2-k_1)\epsilon n e^{-\Omega(\epsilon^3 n)}\right) = O\left(\epsilon n^{\frac{5}{3}} e^{-n^{\frac{1}{4}}}\right) = o(1),$$

where we used that $\epsilon^3 n = \omega\left(n^{\frac{1}{4}}\right)$. □

Using Lemma 4.3, we can then determine asymptotically the excess of the giant component. As previously mentioned, we will argue via a multi-round exposure argument, taking a sequence $p_1 \le p_2 \le \ldots \le p_s$ of probabilities such that p_1 is supercritical, but significantly smaller than p, and $p_s = p$. Via a standard coupling argument, we can think of sampling $G(n,n,p_1)$ and then sampling an independent sequence of bipartite random graphs $G(n,n,p'_i)$ where $p'_i = \frac{p_{i+1}-p_i}{1-p_i}$ so that for each $1 \le i \le s$

$$G(n,n,p_1) \cup \left(\bigcup_{j=1}^{i-1} G(n,n,p'_j)\right) \sim G(n,n,p_i),$$

and so we have the inclusions $G(n,n,p_1) \subseteq G(n,n,p_2) \subseteq \ldots \subseteq G(n,n,p_s)$.

Our choice of p_0, together with Theorem 1.6, guarantees that the excess of $L_1\left(G(n,n,p_1)\right)$ is significantly smaller than $\epsilon^3 n$. We then estimate precisely the change in the excess of the giant component in each of the sprinkling steps. To do so, we bound whp from above and below the number Δ_i of extra excess edges in the giant component when adding each $G(n,n,p'_i)$. Here, it is essential that the probability of failure in each step is small enough that the sum of these probabilities over all $0 \le i \le s$ is still small. Then, we can asymptotically determine the excess of $L_1\left(G(n,n,p_s)\right)$ as a sum of the Δ_i, which we can approximate by an integral.

Theorem 4.4. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^4 n \gg \omega \to \infty$ and $\epsilon \le \frac{1}{\omega}$, and let $p = \frac{1+\epsilon}{n}$. Then with probability $1 - O\left(\omega^{-0.01}\right)$*

$$\textit{excess}\left(L_1\left(G(n,n,p)\right)\right) \approx \frac{4}{3}\epsilon^3 n.$$

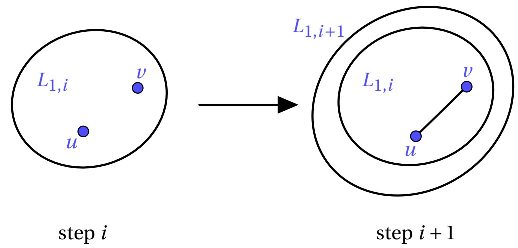

FIGURE 3. In step $i+1$ every edge uv in $G(n,n,p'_i)$ with $u,v \in V\left(L_{1,i}\right)$ contributes to Δ_i.

Proof. For each $i \in \mathbb{N}$, let

$$\epsilon_i = \omega^{0.2} n^{-\frac{1}{4}} \left(1+\omega^{-0.1}\right)^{i-1} \quad \text{and} \quad p_i = \frac{1+\epsilon_i}{n}.$$

Throughout the proof we work under the assumption that i is small enough so that $\epsilon_i = o(1)$.

By a standard coupling argument we can think of moving from $G(n,n,p_i)$ to $G(n,n,p_{i+1})$ via sprinkling. That is, we choose independently for each i a random graph $G(n,n,p'_i)$ where

$$p'_i = \frac{p_{i+1}-p_i}{1-p_i} = \frac{\epsilon_{i+1}-\epsilon_i}{n-1-\epsilon_i},$$

in such a way that $G(n,n,p_{i+1}) = G(n,n,p_i) \cup G(n,n,p'_i)$ for each i. We note that, if we write $L_{1,i}$ for the largest component of $G(n,n,p_i)$ for each i, then by Theorem 1.5,

$$|L_{1,i} \cap N_1| \approx \frac{\epsilon_i+\epsilon'_i}{1+\epsilon_i} n \quad \text{and} \quad |L_{1,i} \cap N_2| \approx \frac{\epsilon_i+\epsilon'_i}{1+\epsilon_i} n, \tag{28}$$

with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$. Furthermore, by Lemma 4.3 with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$,

$$a_i := \text{excess}\left(L_{1,i}\right) = O\left(\epsilon_i^3 n\right). \tag{29}$$

Note that, by (29), with probability $1-O\left(\left(\epsilon_0^4 n\right)^{-\frac{1}{6}}\right)$, $a_1 = O\left(\omega^{0.6} n^{\frac{1}{4}}\right) = o\left(\omega^{\frac{3}{4}} n^{\frac{1}{4}}\right) = o\left(\epsilon^3 n\right)$, and so to begin with we may assume that the excess is much smaller than $\epsilon^3 n$.

We show that we can control quite precisely how the excess of the giant component changes in each sprinkling step. More precisely, we claim that for each i, with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$

$$\Delta_i := a_{i+1} - a_i \approx \frac{(\epsilon_i+\epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i). \tag{30}$$

In order to show (30) we bound from above and below the number of new excess edges added in step $i+1$.

Claim 4.5. *With probability* $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$,

$$\Delta_i \gtrsim \frac{(\epsilon_i+\epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i).$$

Proof of Claim 4.5. We note that every edge in $G(n,n,p'_i)$ which has both ends in $L_{1,i}$ adds at least one to the quantity Δ_i (see Figure 3). Hence, by (28) and (29) with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$, Δ_i stochastically dominates a binomial random variable $\text{Bin}(m,q) = Y$ with parameters

$$m \approx \left(\frac{\epsilon_i+\epsilon'_i}{1+\epsilon_i} n\right)^2 - 2\frac{\epsilon_i+\epsilon'_i}{1+\epsilon_i} n - O(\epsilon_i^3 n) \quad \text{and} \quad q = p'_i.$$

Now, we see that

$$\mathbb{E}(Y) \gtrsim \frac{(\epsilon_i+\epsilon'_i)^2}{(1+\epsilon_i)^2} n^2 \frac{\epsilon_{i+1}-\epsilon_i}{n-1-\epsilon_i} \approx \frac{(\epsilon_i+\epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i),$$

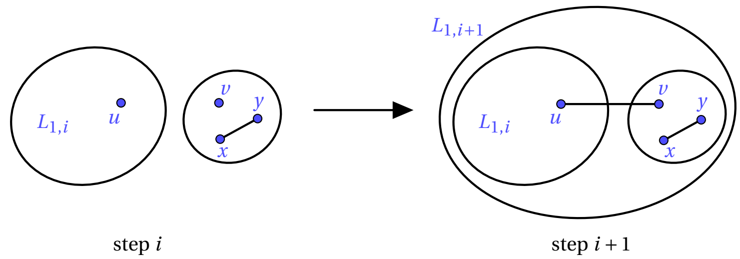

FIGURE 4. In step $i+1$ the only contribution to Δ_i comes from edges uv in $G(n,n,p'_i)$ with $u,v \in V\left(L_{1,i+1}\right)$ or excess edges xy in components of $G(n,n,p_i)$ joined to $L_{1,i}$ by such an edge.

and so $\mathbb{E}(Y) = \Omega(\epsilon_i^3 n)$. Hence, by Lemma 2.3 we obtain

$$\mathbb{P}\left(|\mathbb{E}(Y) - Y| \geq \frac{\mathbb{E}(Y)}{(\epsilon_i^4 n)^{\frac{1}{4}}}\right) \leq \exp\left(-\Omega\left(\frac{\mathbb{E}(Y)}{(\epsilon_i^4 n)^{\frac{1}{2}}}\right)\right) \leq \exp\left(-\left(\epsilon_i^4 n\right)^{\frac{1}{2}}\right) = O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right).$$

Hence, with probability $1 - O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$, we get

$$Y \gtrsim \mathbb{E}(Y) \gtrsim \frac{(\epsilon_i + \epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1} - \epsilon_i),$$

and so with at least this probability

$$\Delta_i \gtrsim \frac{(\epsilon_i + \epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1} - \epsilon_i).$$

□

Claim 4.6. *With probability* $1 - O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$

$$\Delta_i \lesssim \frac{(\epsilon_i + \epsilon'_i)^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1} - \epsilon_i).$$

Proof of Claim 4.6. For an upper bound, we need to be slightly more careful. We note that there are two ways that edges in $G(n,n,p'_i)$ can contribute to Δ_i. Firstly, edges in $G(n,n,p'_i)$ which have both endpoints in $L_{1,i+1}$ adds one to this quantity. However, there are some other edges, specifically excess edges in non-giant components of $G(n,n,p_i)$ which are joined to $L_{1,i+1}$ by an edge of $G(n,n,p'_i)$, which also add to this quantity (see Figure 4).

We first show that the contribution from the former of these is approximately what we expect, and then show that the contribution from the latter is negligible.

For the first of these, let $\mathscr{A}$ be the event that

$$\max\left\{|C \cap N_j| \colon C \text{ a component of } G(n,n,p_{i+1})\} \text{ for } j = 1,2\right\} \lesssim \frac{\epsilon_i + \epsilon'_i}{1+\epsilon_i} n.$$

Note that,

$$\frac{\epsilon_i + \epsilon'_i}{1+\epsilon_i} n \approx \frac{\epsilon_{i+1} + \epsilon'_{i+1}}{1+\epsilon_{i+1}} n.$$

Then, by (28), $\mathbb{P}(\mathscr{A}) \geq 1 - O\left(\left(\epsilon_{i+1}^4 n\right)^{-\frac{1}{6}}\right)$ and $\mathscr{A}$ is a decreasing property. Thus, by Harris' inequality (Lemma 2.4), given any set of edges F the probability that $F \subseteq G(n,n,p'_i)$ conditioned on $\mathscr{A}$ is strictly less than the probability that $F \subseteq G(n,n,p'_i)$. Hence, with probability $1 - O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$ the number of edges added to the vertex set of the new giant component is stochastically dominated by a binomial random variable $\text{Bin}(m', q') := Z$ with parameters

$$m' \approx \left(\frac{\epsilon_i + \epsilon'_i}{1+\epsilon_i} n\right)^2 \text{ and } q' = p'_i.$$

As before, we have

$$\mathbb{E}(Z) \lesssim \frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n^2 . \frac{\epsilon_{i+1}-\epsilon_i}{n-1-\epsilon_i} \approx \frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i),$$

and so again by Lemma 2.3 with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$,

$$Z \lesssim \mathbb{E}(Z) \lesssim \frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i).$$

Now, let us bound the contribution to Δ_i from excess edges in non-giant components of $G(n,n,p_i)$. By Theorem 1.9 with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$ there are no complex components of order smaller than $n^{\frac{2}{3}}$ and by Theorem 1.5 with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$ there are no components apart from the giant component of order at most $n^{\frac{2}{3}}$. Hence, it follows that with at least this probability every non-tree component in $G(n,n,p_i)$ except $L_{1,i}$ is unicyclic, and so the contribution to Δ_i from excess edges in non-giant components of $G(n,n,p_i)$ is equal to the number of unicylic components in $G(n,n,p_i)$ which are joined to $L_{1,i}$ by $G(n,n,p_i')$, which we can bound from above by the number of edges in $G(n,n,p_i')$ which join such components to $L_{1,i}$.

Then, by Lemma 4.2, with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$ the number of vertices in unicyclic components of $G(n,n,p_i)$ is at most

$$O\left(\frac{(\epsilon_i^4 n)^{\frac{1}{6}}}{\epsilon_i^2}\right) = o\left(n^{\frac{2}{3}}\right).$$

Hence, since rather crudely $|V(L_{1,i})| \le 5\epsilon_i n$, the expected number of edges in $G(n,n,p_i')$ which connect unicyclic components in $G(n,n,p_i)$ to $L_{1,i}$ is less than

$$5\epsilon_i n n^{\frac{2}{3}} p_i' = O\left(\epsilon_i(\epsilon_{i+1}-\epsilon_i) n^{\frac{2}{3}}\right).$$

Then, by Markov's inequality, with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$ the number of such edges is at most

$$O\left(\epsilon_i(\epsilon_{i+1}-\epsilon_i) n^{\frac{2}{3}} \left(\epsilon_i^4 n\right)^{\frac{1}{6}}\right) = o\left(\frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i)\right).$$

It follows that, with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$

$$\Delta_i \lesssim \frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i).$$

□

Hence, by Claims 4.5 and 4.6, (30) holds with probability $1-O\left(\left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right)$. Therefore, by a union bound, (30) holds for all $i \in \mathbb{N}$ such that $\epsilon_i = o(1)$ with probability

$$1-O\left(\sum_{i=1}^{\infty} \left(\epsilon_i^4 n\right)^{-\frac{1}{6}}\right) \ge 1-O\left(\omega^{-0.01}\right),$$

which can be seen by noting that the sum is a geometric series.

Let $s \in \mathbb{N}$ be such that $\epsilon_{s-1} \le \epsilon \le \epsilon_s$. Then

$$a_s - a_1 = \sum_{i=1}^{s-1} \Delta_i \approx \sum_{i=1}^{s-1} \frac{(\epsilon_i+\epsilon_i')^2}{(1+\epsilon_i)^2} n(\epsilon_{i+1}-\epsilon_i) \approx n \int_{\epsilon_1}^{\epsilon_s} \frac{(x+y)^2}{(1+x)^2} dx, \tag{31}$$

where $(1-y)e^y = (1+x)e^{-x}$, and we can approximate the sum by the integral since $\epsilon_{i+1}-\epsilon_i = o(1)$.

If we let $F(x,y) = (1-y)e^y - (1+x)e^{-x}$ then $F(x,y) = 0$ when $(1-y)e^y = (1+x)e^{-x}$. By the derivative of implicit functions, we have

$$\frac{dy}{dx} = -\frac{F_x'}{F_y'} = \frac{x}{y}\frac{1-y}{1+x}.$$

This implies that $\frac{d}{dx}\left(\frac{x^2-y^2}{1+x}\right)=\frac{(x+y)^2}{(1+x)^2}$. In this manner, we can conclude from (31) that with probability $1-O\left(\omega^{-0.01}\right)$

$$a_s-a_1\approx n\left(\frac{\epsilon_s^2-\epsilon_s'^2}{1+\epsilon_s}-\frac{\epsilon_1^2-\epsilon_1'^2}{1+\epsilon_1}\right)\approx n\left(\frac{1}{1+\epsilon_s}\left(\frac{4}{3}\epsilon_s^3+O\left(\epsilon_s^4\right)\right)-\frac{\epsilon_1^2-\epsilon_1'^2}{1+\epsilon_1}\right)\approx\frac{4}{3}\epsilon_s^3 n,$$

where we used that $\epsilon_s'=\epsilon_s+\frac{2}{3}\epsilon_s^2+O(\epsilon_s^3)$. Since, as previously mentioned, $a_1=o(\epsilon^3 n)$ it follows that $a_s\approx\frac{4}{3}\epsilon_s^3 n$. A similar argument shows that $a_{s-1}\approx\frac{4}{3}\epsilon_{s-1}^3 n$.

Then, since

$$\epsilon_{s-1}\le\epsilon\le(1+\omega^{-0.1})\epsilon_{s-1}\quad\text{and}\quad\frac{\epsilon_s}{1+\omega^{-0.1}}\le\epsilon\le\epsilon_s,$$

and we can couple the three random bipartite graphs such that $G(n,n,p_{s-1})\subseteq G(n,n,p)\subseteq G(n,n,p_s)$, it follows that

$$\text{excess}\left(L_1\left(G(n,n,p)\right)\right)\approx\frac{4}{3}\epsilon^3 n.$$

□

Proof of Theorem 1.6. The theorem follows directly from Theorem 4.4. □

5. Counting bipartite graphs: proofs of Theorems 3.3 and 3.5

5.1. **Unicyclic bipartite graphs: proof of Theorem 3.3.** Since a unicyclic graph is the union of a cycle and a forest, we are able to deduce Theorem 3.3 from a formula for the number of bipartite forests, which we derive using a standard counting tool known as Prüfer codes.

Lemma 5.1. *Given $i,j,s,t\in\mathbb{N}$ satisfying $s\le i$ and $t\le j$, let $F(i,j,s,t)$ denote the number of bipartite forests with partition classes $I=\{x_1,\dots,x_i\}$ and $J=\{y_1,\dots,y_j\}$ with $s+t$ components where the vertices $x_1,\dots,x_s,y_1,\dots,y_t$ belong to distinct components. Then*

$$F(i,j,s,t)=si^{j-t-1}j^{i-s}+tj^{i-s-1}i^{j-t}=\left(\frac{s}{i}+\frac{t}{j}\right)i^{j-t}j^{i-s}.\tag{32}$$

Proof. Let us fix an arbitrary ordering on $\{0\}\cup I\cup J$ so that the first $s+t$ vertices are 0 followed by $\{x_1,\dots,x_s,y_1,\dots,y_t\}$. Given a bipartite forest F as in the definition of $F(i,j,s,t)$, let us construct a tree $T(F)\supset F$ by adding a new vertex 0 and edges from 0 to each vertex in $\{x_1,\dots,x_s,y_1,\dots,y_t\}$.

We construct a sequence $w\in(I\cup J\cup\{0\})^{i+j}$, normally called a *Prüfer code*, by recursively deleting the largest leaf in $T(F)$ and adding its unique neighbour to the end of w, until just one vertex remains. We note that each time we delete a leaf in I, we add a vertex in $J\cup\{0\}$ to the list, and vice versa. Furthermore, since the last $s+t$ leaves we delete are $\{x_1,\dots,x_s,y_1,\dots,y_t\}$, the final $s+t$ entries in the list are 0, and the one just preceding this is in $\{x_1,\dots,x_s,y_1,\dots,y_t\}$. Hence, w consists of a list $u\in I^{j-t}$ and a list $v\in J^{i-s}$ interleaved in some order, such that either the last entry of u is in $\{x_1,\dots,x_s\}$ or the last entry of v is in $\{y_1,\dots,y_t\}$, followed by $s+t$ many 0s. Let

$$\mathscr{W}=\left\{(u,v)\in I^{j-t}\times J^{i-s}\colon u_{j-t}\in\{x_1,\dots,x_s\}\text{ or }v_{i-s}\in\{y_1,\dots,y_t\}\right\}.$$

It is a simple check that if $F\neq F'$ then $T(F)\neq T(F')$ and so the Prüfer codes w and w' they produce are different. Conversely, we claim that every pair $(u,v)\in\mathscr{W}$ uniquely determines an F such that the Prüfer code for $T(F)$ consists of u and v interleaved in some order followed by $s+t$ many 0s.

Indeed, we first note that the degree of a vertex $z\in I\cup J$ in F can be seen to be one plus the number of times it appears in u or v, and hence from (u,v) we can recover the degree sequence $d: I\cup J\cup\{0\}\to\mathbb{N}$ of F.

We construct the forest F in a recursive manner, keeping track of two integers $t_1(r)+t_2(r)=r$, two lists $u^r\in I^{j-t-t_1(r)}$ and $v^r\in J^{i-s-t_2(r)}$ and a degree sequence $d^r: I\cup J\cup\{0\}\to\mathbb{N}$ where initially we have $t_1=t_2=0$, $u^0=u$, $v^0=v$ and $d^0=d_{T(F)}$ is the degree function of $T(F)$.

For each $0\le r\le i+j-s-t$ we do the following: Let $z=\max\{z'\in I\cup J\colon d^r(z')=1\}$. If $z\in I$ we add the edge (z,v_1^r) to F, we set $t_1(r+1)=t_1(r)$ and $t_2(r+1)=t_2(r)+1$ and we let $u^{r+1}=u^r$ and let v^{r+1} be formed from v^r by deleting v_1^r. Finally, we form d^{r+1} by reducing the degree of z and v_1 by one in d^r. If $z\in J$ we instead add the edge (z,u_1^r) to F, set $t_1(r+1)=t_1(r)+1$ and $t_2(r+1)=t_2(r)$ and we let $v^{r+1}=v^r$ and let u^{r+1} be formed from u^r by deleting u_1^r. Finally we form d^{r+1} by reducing the degree of z and u_1^r by one in d^r. We continue until both u^r and v^r are empty.

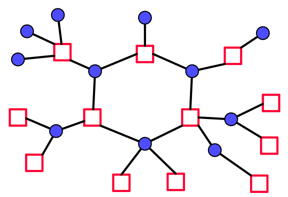

FIGURE 5. Every unicyclic bipartite graph contains an even cycle C whose deletion leaves a bipartite forst with $|V(C)|$ many components.

It is easy to check that if w is the Prüfer code for $T(F)$, then this construction produces F, but only depends on the pair $(u,v)\in\mathscr{W}$ which form w. Furthermore, given any pair $(u,v)\in\mathscr{W}$ it is easy to construct a bipartite forest F such that the Prüfer code for $T(F)$ consists of u and v interleaved in some manner followed by $s+t$ many 0s. Hence, the number of such forest is $|\mathscr{W}|$, which can clearly be seen to be $si^{j-t-1}j^{i-s}+tj^{i-s-1}i^{j-t}$ as claimed. □

Using Lemma 5.1, we can prove Theorem 3.3.

Proof of Theorem 3.3. We note that every unicyclic bipartite graph with i vertices in one partition class and j in the other contains a unique cycle, which has length $2r$ for some $r\le\min\{i,j\}$, and if we delete the edges of this cycle, then what remains is a forest with $2r$ components, each meeting one vertex of the cycle (see Figure 5).

Hence, we can count $C(i,j,0)$ by first choosing a cycle of length $2r$, of which there are $\frac{(i)_r(j)_r}{2r}$ many possibilities, and then choosing from the $F(i,j,r,r)$ many possibilities for the forest left by the deletion of this cycle.

Hence, it follows from (32) that

$$C(i,j,0)=\sum_{r=2}^{\min\{i,j\}}\frac{(i)_r(j)_r}{2r}F(i,j,r,r)=\frac{1}{2}\left(\frac{1}{i}+\frac{1}{j}\right)i^j j^i\sum_{r=2}^{\min\{i,j\}}\frac{(i)_r(j)_r}{i^r j^r},\tag{33}$$

proving the first part of Theorem 3.3.

So let us suppose further that $i,j=\omega(1)$ and $\frac{1}{2}\le\frac{i}{j}\le 2$. By (3) we can conclude that

$$\frac{(i)_r(j)_r}{i^r j^r}\le\exp\left(-\frac{r^2}{2i}-\frac{r^2}{2j}\right),$$

and furthermore by (2) it follows that if $r=o\left(i^{\frac{2}{3}}\right)=o\left(j^{\frac{2}{3}}\right)$, then

$$\frac{(i)_r(j)_r}{i^r j^r}\approx\exp\left(-\frac{r^2}{2i}-\frac{r^2}{2j}\right).$$

We split (33) into two parts. Firstly, when $r\le i^{\frac{5}{9}}$ we note that $r=o\left(i^{\frac{2}{3}}\right)=o\left(j^{\frac{2}{3}}\right)$, and hence

$$\sum_{r=2}^{i^{\frac{5}{9}}}\frac{(i)_r(j)_r}{i^r j^r}\approx\sum_{r=2}^{i^{\frac{5}{9}}}\exp\left(-\frac{r^2}{2i}-\frac{r^2}{2j}\right)\approx\sqrt{\frac{\pi ij}{2(i+j)}},$$

where the final line follows from a standard estimate that

$$\sum_{r=1}^{\infty}e^{-\frac{r^2}{2n}}\approx\int_0^{\infty}e^{-\frac{x^2}{2n}}dx=\sqrt{\frac{\pi n}{2}}.$$

Conversely, when $r\ge i^{\frac{5}{9}}$ we can naively bound

$$\sum_{r=i^{\frac{5}{9}}}^{\min\{i,j\}}\frac{(i)_r(j)_r}{i^r j^r}\le i\exp\left(-\frac{i^{\frac{10}{9}}}{2i}-\frac{i^{\frac{10}{9}}}{2j}\right)\le i\exp\left(-\Omega\left(i^{\frac{1}{9}}\right)\right)=o(1).$$

It follows that

$$C(i,j,0)\approx\sqrt{\frac{\pi ij}{8(i+j)}}\left(\frac{1}{i}+\frac{1}{j}\right)i^j j^i=\sqrt{\frac{\pi}{8}}\sqrt{i+j}\,i^{j-\frac{1}{2}}j^{i-\frac{1}{2}}.$$

□

5.2. **Bipartite graphs with positive excess: proof of Theorem 3.5.** In order to count the number of balanced bipartite graphs with positive excess, we will split into two cases, when the excess is either small or large compared to the number of vertices. The latter case will turn out to be much simpler.

When the excess is small compared to the number of vertices, we will count the number of such graphs using an ingenious probabilistic method due to Łuczak [18], which turns the observation that $\mathbb{E}\left(X(i,j,\ell)\right)$ is intimately related to $C(i,j,\ell)$ on its head and uses rather the former to bound the latter.

Lemma 5.2. *There exists a constant $C>0$ such that for any $i,j,\ell\in\mathbb{N}$ satisfying $\frac{1}{2}\le\frac{i}{j}\le 2$ and $\ell\le i+j$ we have*

$$C(i,j,\ell)\le i^{j+\frac{1}{2}}j^{i+\frac{1}{2}}(i+j)^{\frac{3\ell+1}{2}}\left(\frac{i}{j}\right)^{\frac{i-j}{2}}\frac{1}{\sqrt{\ell}}\left(\frac{C}{\ell}\right)^{\frac{\ell}{2}}.$$

Proof. Let $Z(i,j,\ell)$ denote the number of components with i vertices in the partition class N_1, j vertices in the partition class N_2 and ℓ excess edges in the binomial random bipartite graph $G(ir,jr,q)$, where we will choose particular r and q later. Note, that it suffices to prove the statement for $i+j$ sufficiently large.

Using (2) together with the bound

$$1-q=\exp\left(-q-\frac{q^2}{2}+O\left(q^3\right)\right),$$

which holds for $q\ll 1$, and Stirling's approximation, and letting $k=i+j$, we see that, as long as $q\ll 1$,

$$\begin{aligned}\mathbb{E}(Z(i,j,\ell))&=\binom{ir}{i}\binom{jr}{j}C(i,j,\ell)q^{i+j+\ell}(1-q)^{i(jr-j)+j(ir-i)+ij-i-j-\ell}\\&=\frac{(ir)_i}{i!}\frac{(jr)_j}{j!}C(i,j,\ell)q^{k+\ell}(1-q)^{2ijr-ij-k-\ell}\\&=C(i,j,\ell)\frac{1}{i!j!}(ir)^i(jr)^jq^{k+\ell}\exp\left(-\frac{k}{2r}-\frac{k}{6r^2}-O\left(\frac{k}{r^3}\right)-\left(q+\frac{q^2}{2}+O\left(q^3\right)\right)(2ijr-ij-k-\ell)\right)\\&\ge C(i,j,\ell)\frac{i^ij^j}{i!j!}\frac{(qr)^{k+\ell}}{r^\ell}\exp\left(-\frac{k}{2r}-\frac{k}{6r^2}-O\left(\frac{k}{r^3}\right)-\left(q+\frac{q^2}{2}+O\left(q^3\right)\right)(2ijr-ij)\right).\end{aligned}\tag{34}$$

Since $|N_1|=ir$, there can clearly be at most r components containing i vertices in N_1, and so $\mathbb{E}(Z(i,j,\ell))\le r$. This together with (34) implies that

$$C(i,j,\ell)\le\frac{i!j!r^{\ell+1}}{i^ij^j}\frac{1}{(qr)^{k+\ell}}\exp\left(\frac{k}{2r}+\frac{k}{6r^2}+O\left(\frac{k}{r^3}\right)+\left(q+\frac{q^2}{2}+O\left(q^3\right)\right)(2ijr-ij)\right).$$

We set $q=\frac{k(1+\eta)}{2ijr}$, where we will choose $\eta\le 1$ later. Note that, in this case as long as i and j are sufficiently large, $q\ll 1$. Hence, using Stirling's inequality we have

$$C(i,j,\ell)\le\frac{\sqrt{ij}r^{\ell+1}}{e^k}\frac{(2ij)^{k+\ell}}{k^{k+\ell}}\frac{1}{(1+\eta)^{k+\ell}}\exp\left(\frac{k}{2r}+\frac{k}{6r^2}+O\left(\frac{k}{r^3}\right)+k(1+\eta)-\frac{k(1+\eta)}{2r}+\left(\frac{q^2}{2}+O\left(q^3\right)\right)2ijr\right).$$

Then, as long as $\eta\le 1$ we can conclude that $1+\eta\ge e^{\eta-\eta^2}$ and so

$$C(i,j,\ell)\le\sqrt{ij}r^{\ell+1}\left(\frac{2ij}{k}\right)^{k+\ell}\exp\left(-\eta\ell+\eta^2(k+\ell)-\frac{\eta k}{2r}+\frac{k}{6r^2}+O\left(\frac{k}{r^3}\right)+\left(\frac{q^2}{2}+O\left(q^3\right)\right)2ijr\right).$$

Let us take $r=\sqrt{\frac{k}{\ell}}$ and $\eta=\sqrt{\frac{\ell^3}{k^3}}$, so that $\eta\le 1$, and let us consider each of the terms inside the exponent of the above inequality. We shall show that they are all $O(\ell)$. Indeed, for the first two terms, we note that $\eta\le 1$ and so $\eta\ell=O(\ell)$, and furthermore, we see that

$$\eta^2(k+\ell)=\ell\frac{\ell^2(k+\ell)}{k^3}\le\ell\frac{k^2 2k}{k^3}=O(\ell).$$

For the next three terms, we note that

$$\frac{\eta k}{2r}=\frac{\ell^2}{2k}=O(\ell)\qquad\text{and}\qquad\frac{k}{r^3}\le\frac{k}{r^2}=\ell.$$

Finally, for the last two terms, we see that

$$q^2 ijr = \frac{k^2(1+\eta)^2}{4ijr} \le \frac{k^2}{ij} = O(1),$$

as long as $\frac{1}{2} \le \frac{i}{j} \le 2$, and similarly $q^3 ijr = \frac{k^3(1+\eta)^3}{8i^2j^2r^2} \le \frac{k^3}{i^2j^2} = o(1)$. This implies that

$$C(i,j,\ell) \le \sqrt{ij}\,r^{\ell+1}\left(\frac{2ij}{k}\right)^{k+\ell} \exp\left(O(\ell)\right) = \sqrt{ij}\sqrt{\frac{k}{\ell}}k^{\frac{\ell}{2}}\left(\frac{2ij}{k}\right)^{k+\ell}\left(\frac{C}{\ell}\right)^{\frac{\ell}{2}}.$$

Then, since by the AM-GM inequality we have that $k^{k+2\ell} \ge 2^{k+2\ell}(ij)^{\frac{k}{2}+\ell}$, we can conclude that

$$\begin{aligned} C(i,j,\ell) &\le \sqrt{ij}\sqrt{\frac{k}{\ell}}k^{\frac{\ell}{2}}\left(\frac{2ij}{k}\right)^{k+\ell}\left(\frac{C}{\ell}\right)^{\frac{\ell}{2}} = i^{j+\frac{1}{2}}j^{i+\frac{1}{2}}\frac{1}{\sqrt{\ell}}k^{\frac{3\ell+1}{2}}\left(\frac{C}{\ell}\right)^{\frac{\ell}{2}}\frac{2^{k+\ell}i^{i+\ell}j^{j+\ell}}{k^{k+2\ell}} \\ &\le i^{j+\frac{1}{2}}j^{i+\frac{1}{2}}k^{\frac{3\ell+1}{2}}\left(\frac{i}{j}\right)^{\frac{i-j}{2}}\frac{1}{\sqrt{\ell}}\left(\frac{C}{\ell}\right)^{\frac{\ell}{2}}. \end{aligned}$$

□

In the other case, when ℓ is larger than $i+j$, it is much simpler to bound $C(i,j,\ell)$.

Lemma 5.3. *For any $i,j,\ell \in \mathbb{N}$ satisfying $\ell \ge i+j$ and $\frac{1}{2} \le \frac{i}{j} \le 2$, we have*

$$C(i,j,\ell) \le i^{j-\frac{1}{2}}j^{i-\frac{1}{2}}(i+j)^{\frac{3\ell+1}{2}}\ell^{-\frac{\ell}{2}}.$$

Proof. Let $k = i+j$. We can naively bound $C(i,j,\ell)$ by looking at the total number of ways of choosing $k+\ell$ edges from the ij many possible edges. Note, this is an overcount, since we are also counting disconnected graphs.

Since $\binom{n}{r} \le \left(\frac{en}{r}\right)^r$, it follows that

$$C(i,j,\ell) \le \binom{ij}{k+\ell} \le \left(\frac{eij}{k+\ell}\right)^{k+\ell} \le i^{j-\frac{1}{2}}j^{i-\frac{1}{2}}k^{\frac{3\ell+1}{2}}\ell^{-\frac{\ell}{2}}\left(\frac{e^{k+\ell}i^{i+\ell+\frac{1}{2}}j^{j+\ell+\frac{1}{2}}\ell^{\frac{\ell}{2}}}{k^{\frac{3\ell}{2}+1}(k+\ell)^{k+\ell}}\right).$$

So, it is sufficient to show that this final term is at most one. However, by the AM-GM inequality we have that $k^2 \ge 4ij$, and hence

$$i^{i+\ell+\frac{1}{2}}j^{j+\ell+\frac{1}{2}} \le \frac{k^{k+2\ell+1}}{2^{k+2\ell+1}}\left(\frac{i}{j}\right)^{\frac{i-j}{2}} \le \frac{k^{k+2\ell+1}}{2^{k+\ell}},$$

since $\left(\frac{i}{j}\right)^{\frac{i-j}{2}} \le 2^{\ell}$ by our assumptions on i,j and ℓ.

Therefore, we obtain

$$\begin{aligned} \left(\frac{e^{k+\ell}i^{i+\ell+\frac{1}{2}}j^{j+\ell+\frac{1}{2}}\ell^{\frac{\ell}{2}}}{k^{\frac{3\ell}{2}+1}(k+\ell)^{k+\ell}}\right) &\le \left(\frac{e}{2}\right)^{k+\ell}\left(\frac{k}{k+\ell}\right)^{k+\frac{\ell}{2}}\left(\frac{\ell}{k+\ell}\right)^{\frac{\ell}{2}} \\ &\le \left(\frac{e}{2}\right)^{k+\ell}\exp\left(-\frac{\ell\left(k+\frac{\ell}{2}\right)}{k+\ell} - \frac{\ell k}{2(k+\ell)}\right) \le \left(\frac{e}{2}\right)^{k+\ell}\exp\left(-\frac{k+\ell}{2}\right) \le 1, \end{aligned}$$

where we used that

$$\frac{\ell\left(k+\frac{\ell}{2}\right)}{k+\ell} + \frac{\ell k}{2(k+\ell)} \ge \frac{2\ell k+\ell^2+k^2}{2(k+\ell)} = \frac{k+\ell}{2}$$

since $\ell \ge k$. □

Proof of Theorem 3.5. The theorem follows as a direct consequence of Lemmas 5.2 and 5.3. □

6. DISCUSSION

6.1. **Critical window.** The fact that some of our main results (Theorems 1.5, 1.6, and 1.9(ii)) only hold for $\epsilon^4 n \to \infty$ rather than in what might be the expected range of $\epsilon^3 n \to \infty$ seems to be an artefact of the proof: our proof relies on enumerative results for the number of bipartite graphs with positive excess (Theorem 3.5) which we suspect are not optimal. There is a natural bound to conjecture for the number of such bipartite graphs (see Section 6.2), generalising the known bounds for the number of graphs, and assuming this bound it is easy to adapt the proofs of our main results to cover the 'correct' range of ϵ (in fact, many of the calculations in the proofs become more natural with this parameterisation).

6.2. **Number of bipartite graphs with positive excess.** More explicitly, we conjecture that the following bound should hold.

Conjecture 6.1. *There is a constant $c > 0$ such that for all $i, j, \ell \in \mathbb{N}$ with $\frac{1}{2} \le \frac{i}{j} \le 2$,*

$$C(i, j, \ell) \le i^{j-\frac{1}{2}} j^{i-\frac{1}{2}} (i+j)^{\frac{3\ell+1}{2}} \left(\frac{i}{j}\right)^{\frac{i-j}{2}} \left(\frac{c}{\ell}\right)^{\frac{\ell}{2}}.$$

There are a few natural methods that one might use to try to prove such a bound. One would be via the so-called *core and kernel* method, used to prove similar results in the $G(n, p)$ model (See for example [4, 17]). Another would be to follow the methods of Bender, Canfield and McKay [3], who gave an asymptotic formula for the number of graphs with n vertices and k edges, which we denote as $c(n, k)$, which they derived by analysing the following recursive formula

$$kc(n,k) = \left(\binom{n}{2} - k + 1\right) c(n, k-1) + \frac{1}{2}\sum_{t=1}^{n-1}\sum_{s=-1}^{k-n}\binom{n}{t} t(n-t)c(t, t+s)c(n-t, k-t-s-1),$$

which can be seen by deleting an edge from a graph with n vertices and k edges and splitting into two cases as to whether this edge is a bridge or not. However, we were not able to implement either approach in the bipartite case.

6.3. **Open problems.** We have presented some initial results about the structure of $G(n, n, p)$ in the weakly supercritical regime, however many interesting questions still remain. For example, Łuczak [16] described in more detail the distribution of cycles in $G(n, p)$ in this regime. In particular, if we let the *girth* of a graph be the length of the shortest cycle and the *circumference* be the length of the longest cycle, then Łuczak determined asymptotically the girth and circumference of the giant component of $G(n, p)$ and the length of the longest cycle outside of the giant component.

Question 6.2. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. What is the girth and circumference of the giant component in $G(n, n, p)$? What is the length of the longest cycle outside of the giant component?*

Using some of the results of Łuczak [16] on the distribution of cycles in the weakly supercritical regime in $G(n, p)$, together with Euler's formula, Dowden, Kang and Krivelevich [9] were able to determine asymptotically the genus of $G(n, p)$ in this regime, in particular showing that whp the genus is asymptotically given by half of the excess of the giant component. It is natural to ask if a similar statement holds in the bipartite model.

Question 6.3. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^3 n \to \infty$ and $\epsilon = o(1)$, and let $p = \frac{1+\epsilon}{n}$. Is it true that whp the genus g of $G(n, n, p)$ is such that*

$$g \approx \frac{1}{2} excess\big(L_1\left(G(n, n, p)\right)\big) \approx \frac{2}{3}\epsilon^3 n?$$

Theorems 1.7-1.9 suggest an interesting relationship between the component structure of $G\left(n, n, \frac{1+\epsilon}{n}\right)$ and that of $G\left(n, n, \frac{1-\epsilon}{n}\right)$ in the weakly super- and subcritical regimes. In the case of the binomial random graph model, a much more precise relationship can be given. Given a graph G, let us write G^L for the graph obtained by deleting a component of G of maximum order, say L. Roughly speaking, it is known that $G^L\left(n, \frac{1+\epsilon}{n}\right)$ and $G\left(n - |L|, \frac{1-\epsilon}{n-|L|}\right)$ have approximately the same distribution. For a more detailed discussion of this phenomenon, known as the *symmetry rule*, see for example [12, Section 5.6]. Using similar techniques as in [17], which uses bounds on the excess of the giant component to prove a symmetry rule, we expect that Theorem 1.6 can be used to show a similar statement in the bipartite binomial random graph model.

Conjecture 6.4. *Let $\epsilon = \epsilon(n) > 0$ be such that $\epsilon^4 n \gg \omega \to \infty$ and $\epsilon \le \frac{1}{\omega}$, and let $p = \frac{1+\epsilon}{n}$. If we let*

$$n^{\pm} = (1 - 2\epsilon \pm o(\epsilon))n \qquad \text{and} \qquad p^{\pm} = \frac{1 - \epsilon \pm o(\epsilon)}{n^{\pm}},$$

then we can couple $G^L(n,n,p)$ with $G(n^-, n^-, p^-)$ and $G(n^+, n^+, p^+)$ such that whp

$$G(n^-, n^-, p^-) \subseteq G^L(n,n,p) \subseteq G(n^+, n^+, p^+).$$

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover Publications, Inc., New York, 1992. Reprint of the 1972 edition.

[2] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[3] E.A. Bender, E.R. Canfield, and B.D. McKay. The asymptotic number of labeled connected graphs with a given number of vertices and edges. *Random Structures & Algorithms*, 1(2):127–169, 1990.

[4] B. Bollobás. The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1):257–274, 1984.

[5] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, second edition, 2001.

[6] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[7] B. Bollobas, Y. Kohayakawa, and T. Łuczak. The evolution of random subgraphs of the cube. *Random Structures & Algorithms*, 3(1):55–90, 1992.

[8] T.A. Do, J. Erde, and M. Kang. Planarity and genus of sparse random bipartite graphs. *arxiv.org/abs/2005.03920*, 2020.

[9] C. Dowden, M. Kang, and M. Krivelevich. The genus of the Erdős-Rényi random graph and the fragile genus property. *Random Structures & Algorithms*, 56(1):97–121, 2020.

[10] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[11] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[12] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[13] Y. Jing and B. Mohar. The genus of a random bipartite graph. *Canad. J. Math.*, 72(6):1607–1623, 2020.

[14] T. Johansson. The giant component of the random bipartite graph. *Master thesis in Engineering and Computational Science*, 2012.

[15] M. Kang, C. Koch, and A. Pachón. The phase transition in multitype binomial random graphs. *SIAM J. Discrete Math.*, 29(2):1042–1064, 2015.

[16] T. Ł uczak. Cycles in a random graph near the critical point. *Random Structures & Algorithms*, 2(4):421–439, 1991.

[17] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310, 1990.

[18] T. Łuczak. On the number of sparse connected graphs. *Random Structures & Algorithms*, 1(2):171–173, 1990.

[19] H. I. Scoins. The number of trees with nodes of alternate parity. *Proc. Cambridge Philos. Soc.*, 58:12–16, 1962.

[20] J. Spencer. *Asymptopia*, volume 71 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2014. With Laura Florescu.

Appendix A. Proof of Lemma 2.2

Let us write

$$g(y) := \frac{1}{(k^2 - y^2)^m} \left(\frac{k-y}{k+y}\right)^{cy} \exp\left(-\frac{y^2}{2n}\right).$$

If we let $h(y) := \log\big(g(y)\big) = -m\log(k^2 - y^2) + cy\log\left(\frac{k-y}{k+y}\right) - \frac{y^2}{2n}$, then

$$h'(y) = \frac{2my}{k^2 - y^2} + c\log\left(\frac{k-y}{k+y}\right) - \frac{2cky}{k^2 - y^2} - \frac{y}{n},$$

$$h''(y) = \frac{2mk^2 - 4ck^3 + 2my^2}{(k^2 - y^2)^2} - \frac{1}{n},$$

$$h'''(y) = \frac{2y(6mk^4 - 4mk^2y^2 - 2my^4 - 8ck^5 + 8ck^3y^2)}{(k^2 - y^2)^4}.$$

Note that 0 is a solution of $h'(y) = 0$ and, since m is fixed and $h''(y) < 0$ on $[-L-1, L+1]$, 0 is the unique solution on $[-L-1, L+1]$. Hence, $h(y)$ is increasing on $[-L-1, 0]$ and deceasing on $[0, L+1]$, and this is also true for $g(y)$.

Therefore, by Lemma 2.1 we can bound the difference between

$$I := \int_{-L}^{L} g(y)dy,$$

and S as $|S - I| \le 12g(0)$. We will later show that $I = \omega\big(g(0)\big)$, and hence $S \approx I$.

In order to estimate I, we approximate g by a Gaussian function. By the mean value form of the remainder in Taylor's theorem, for any $y \in [-L, L]$ there is a real number z between 0 and y such that

$$h(y) = h(0) + \frac{h''(0)}{2}y^2 + \frac{h'''(z)}{6}y^3.$$

Note that, if $|z| = o(k)$, then $|h'''(z)| = o\left(\frac{1}{k^2}\right)$. Therefore, for any $|y| \le k^{\frac{3}{5}}$ we have

$$h(y) = h(0) + \frac{h''(0)}{2}y^2 + o\left(\frac{y^3}{k^2}\right) = h(0) + \frac{h''(0)}{2}y^2 + o(1).$$

Hence, if we let $R = \min\{k^{\frac{3}{5}}, L\}$ then

$$I = \int_{-R}^{R} \exp\left(h(0) + \frac{h''(0)}{2}y^2 + o(1)\right)dy + \int_{L \ge |y| \ge R} e^{h(y)}dy.$$

The first integral we can evaluate in a standard manner as

$$\int_{-R}^{R} \exp\left(h(0) + \frac{h''(0)}{2}y^2 + o(1)\right)dy \approx \int_{-R}^{R} \exp\left(h(0) + \frac{h''(0)}{2}y^2\right)dy \approx e^{h(0)}\int_{-\infty}^{\infty} \exp\left(\frac{h''(0)}{2}y^2\right)dy$$
$$\approx \sqrt{\frac{2\pi}{|h''(0)|}}e^{h(0)} = \sqrt{\frac{\pi}{2c}}k^{\frac{1}{2}-2m},$$

where we used that $h''(0) = \frac{2m}{k^2} - \frac{4c}{k} - \frac{1}{n} \approx -\frac{4c}{k}$ and also that $R = \omega(1)$.

If $R = L$, then the second integral is 0, and so we may assume that $R = k^{\frac{3}{5}}$. In order to bound the second integral we note that all the terms in $h(y)$ are negative, and in particular if $|y| \le L \le k$

$$\log\left(\frac{k-y}{k+y}\right) = \log\left(1 - \frac{2y}{k+y}\right) \le -\frac{2y}{k+y}.$$

Hence, if $L \ge |y| \ge R$, then

$$h(y) \le cy\log\left(\frac{k-y}{k+y}\right) \le -\frac{cy^2}{k+y} \le -\frac{ck^{\frac{1}{5}}}{2}.$$

It follows that

$$\int_{L \ge |y| \ge R} e^{h(y)}dy \le 2\int_{k^{\frac{3}{5}}}^{\infty} \exp\left(-\frac{cy^{\frac{1}{5}}}{2}\right)dy = O\left(e^{-\frac{c}{2}k^{\frac{3}{25}}}k^{\frac{12}{25}}\right) = o\left(k^{\frac{1}{2}-2m}\right).$$

Hence, $I \approx \sqrt{\frac{\pi}{2c}}k^{\frac{1}{2}-2m}$ and, noting that $g(0) = k^{-2m} = o(I)$, the result follows. □

Appendix B. Proof of Lemma 3.9

Recall that we write $X(i, j, -1)$ for the number of tree components with i vertices in N_1 and j vertices in N_2, and let

$$\Lambda_k = \left\{(i, j) \in \mathbb{N}^2 : i + j = k \text{ and } |i - j| < \epsilon^{\frac{1}{4}}\sqrt{n}\right\}.$$

Then,

$$Z_a = \sum_{k=1}^{\tilde{k}} k^a \sum_{(i,j)\in\Lambda_k} X(i, j, -1),$$

and so

$$\mathbb{E}\left(Z_1^2\right) = \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}} k_1k_2 \sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}} \mathbb{E}\big(X(i, j, -1)X(s, t, -1)\big).$$

Let us write $\mu_{i,j} = \mathbb{E}\big(X(i,j,-1)\big)$. Then, when $(i,j) \neq (s,t)$ we have, by comparison with (8),

$$\begin{aligned}\mathbb{E}\big(X(i,j,-1)X(s,t,-1)\big) &= \binom{n}{i}\binom{n}{j}\binom{n-i}{s}\binom{n-j}{t}C(i,j,-1)C(s,t,-1)p^{k_1+k_2-2}(1-p)^{n(k_1+k_2)-ij-st-sj-ij-k_1-k_2+2}\\ &= \mu_{i,j}\mu_{s,t}\frac{(n)_{i+s}}{(n)_i(n)_s}\frac{(n)_{j+t}}{(n)_j(n)_t}(1-p)^{-it-sj},\end{aligned}$$

and when $(i,j) = (s,t)$ we have

$$\mathbb{E}\big(X(i,j,-1)^2\big) = \mu_{i,j} + \mu_{i,j}^2\frac{(n)_{2i}}{(n)_i^2}\frac{(n)_{2j}}{(n)_j^2}(1-p)^{-2ij}.$$

Now, it can be seen that if $0 \le x \le y \le 1$, then

$$1-y \le (1-x)e^{x-y},$$

and so

$$\frac{(n)_{i+s}}{(n)_i(n)_s} = \prod_{m=0}^{i-1}\frac{1-\frac{s+m}{n}}{1-\frac{m}{n}} \le \exp\left(\sum_{m=0}^{i-1}\frac{m}{n}-\frac{s+m}{n}\right) = \exp\left(-\frac{is}{n}\right), \tag{35}$$

and a similar bound holds for $\frac{(n)_{j+t}}{(n)_j(n)_t}$. Hence, using (35) and the fact that $(1-p)^x \le e^{-px}$ for any positive p and x, we have

$$\begin{aligned}\mathbb{E}\big(Z_1^2\big) &= \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mathbb{E}\big(X(i,j,-1)X(s,t,-1)\big)\\ &= \mathbb{E}(Z_2) + \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mu_{i,j}\mu_{s,t}\frac{(n)_{i+s}}{(n)_i(n)_s}\frac{(n)_{j+t}}{(n)_j(n)_t}(1-p)^{-it-sj}\\ &\le \mathbb{E}(Z_2) + \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mu_{i,j}\mu_{s,t}\exp\left(-\frac{is}{n}-\frac{jt}{n}+(it+sj)\left(\frac{1+\epsilon}{n}\right)\right)\\ &= \mathbb{E}(Z_2) + \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mu_{i,j}\mu_{s,t}\exp\left(\frac{(i-j)(t-s)}{n}+(it+sj)\frac{\epsilon}{n}\right)\\ &\le \mathbb{E}(Z_2) + \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mu_{i,j}\mu_{s,t}\exp\left(\frac{(i-j)(t-s)}{n}+\frac{2\epsilon k_1k_2}{n}\right).\end{aligned} \tag{36}$$

Now, since we are only looking at ϵ-uniform components, if $(i,j)\in\Lambda_{k_1}$ and $(s,t)\in\Lambda_{k_2}$, then

$$\frac{(i-j)(t-s)}{n} \le \sqrt{\epsilon} = o(1).$$

Hence, since $0 \le \frac{2\epsilon k_1k_2}{n} \le \frac{2}{3}$ and $e^x \le 1+x+x^2$ for $|x|\le 1$ it follows that

$$\exp\left(\frac{(i-j)(t-s)}{n}+\frac{2\epsilon k_1k_2}{n}\right) = 1+\frac{(i-j)(t-s)}{n}\left(1+\frac{2\epsilon k_1k_2}{n}\right)+\frac{4\epsilon k_1k_2}{n}+\frac{(i-j)^2(t-s)^2}{n^2}. \tag{37}$$

So, from (36) and (37) we can conclude that

$$\begin{aligned}\mathbb{E}\big(Z_1^2\big) \le \mathbb{E}(Z_2) + \sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}}k_1k_2\sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\mu_{i,j}\mu_{s,t}\\ \cdot\left(1+\frac{(i-j)(t-s)}{n}\left(1+\frac{2\epsilon k_1k_2}{n}\right)+\frac{4\epsilon k_1k_2}{n}+\frac{(i-j)^2(t-s)^2}{n^2}\right).\end{aligned} \tag{38}$$

We split the sum in (38) into four terms and consider them separately. The first three terms are relatively easy to bound.

Firstly, we have that

$$\sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}} k_1k_2 \sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}} \mu_{i,j}\mu_{s,t} = \mathbb{E}(Z_1)^2. \tag{39}$$

Secondly, since $\mu_{i,j}$ is symmetric in i and j and $\mu_{s,t}$ is symmetric in s and t and $(i-j)$ and $(s-t)$ are antisymmetric, it follows that

$$\sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}} k_1k_2\left(1+\frac{2\epsilon k_1k_2}{n}\right) \sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}} \mu_{i,j}\mu_{s,t}\frac{(i-j)(t-s)}{n} = 0. \tag{40}$$

The third term can be seen to be

$$\frac{4\epsilon}{n}\sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}} k_1^2k_2^2 \sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}}\left(1+\frac{2\epsilon k_1k_2}{n}\right) = \frac{4\epsilon}{n}\mathbb{E}(Z_2)^2. \tag{41}$$

For the fourth term, we have to be a bit more careful. Let us consider

$$\begin{aligned}
&\sum_{k_1=1}^{\tilde{k}}\sum_{k_2=1}^{\tilde{k}} k_1k_2 \sum_{(i,j)\in\Lambda_{k_1}}\sum_{(s,t)\in\Lambda_{k_2}} \mu_{i,j}\mu_{s,t}\frac{(i-j)^2(t-s)^2}{n^2}\\
&=\left(\sum_{k_1=1}^{\tilde{k}} k_1\sum_{(i,j)\in\Lambda_{k_1}}\mu_{i,j}\frac{(i-j)^2}{n}\right)\left(\sum_{k_2=1}^{\tilde{k}} k_2\sum_{(s,t)\in\Lambda_{k_2}}\mu_{s,t}\frac{(t-s)^2}{n}\right)\\
&=S^2,
\end{aligned}$$

where

$$S:=\sum_{k=1}^{\tilde{k}} k\sum_{(i,j)\in\Lambda_k}\mu_{i,j}\frac{(i-j)^2}{n}.$$

Using (7), we see that, since $\tilde{k}\le n^{\frac{2}{3}}$, then

$$\begin{aligned}
S &= O\left(\sum_{k=1}^{\tilde{k}} ke^{-\frac{\delta k}{2}}\sum_{(i,j)\in\Lambda_k}\frac{(i-j)^2}{(ij)^{\frac{3}{2}}}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right)\right) = O\left(\sum_{k=1}^{\tilde{k}} ke^{-\frac{\delta k}{2}}\sum_{i+j=k}\frac{(i-j)^2}{(ij)^{\frac{3}{2}}}\left(\frac{i}{j}\right)^{j-i}\exp\left(-\frac{(i-j)^2}{2n}\right)\right)\\
&= O\left(\sum_{k=1}^{\tilde{k}} ke^{-\frac{\delta k}{2}}\sum_{d=-k}^{k}\frac{d^2}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right)\right).
\end{aligned} \tag{42}$$

Firstly, we note that for small k the sum is negligible. Indeed,

$$\sum_{k=1}^{\epsilon^{-\frac{2}{5}}} ke^{-\frac{\delta k}{2}}\sum_{d=-k}^{k}\frac{d^2}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right) \le \sum_{k=1}^{\epsilon^{-\frac{2}{5}}} k\sum_{d=-k}^{k} d^2 \le \epsilon^{-2} = O\left(\sqrt{\frac{n}{\epsilon}}\right). \tag{43}$$

For $k\ge\epsilon^{-\frac{2}{5}}$, we split the inner sum up further into two ranges

$$T_1 := \sum_{|d|\le k^{\frac{3}{5}}}\frac{d^2}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right) \qquad\text{and}\qquad T_2 := \sum_{k\ge|d|\ge k^{\frac{3}{5}}}\frac{d^2}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right).$$

By the same argument as in Lemma 2.2, we see that, since $k=\omega(1)$,

$$T_2 \lesssim \int_{k^{\frac{3}{5}}}^{\infty} y^2\exp\left(-\frac{y^{\frac{1}{5}}}{2}\right)dy = O\left(e^{-\frac{1}{2}k^{\frac{3}{25}}}k^{\frac{42}{25}}\right) = o\left(k^{-\frac{5}{4}}\right). \tag{44}$$

Furthermore, we can bound T_1 naively, using Hölder's inequality and Lemma 2.2, to obtain

$$\begin{aligned}
T_1 &= \sum_{|d|\le k^{\frac{3}{5}}}\frac{d^2}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right) \le \sqrt{\sum_{|d|\le k^{\frac{3}{5}}} d^4}\sqrt{\sum_{|d|\le k^{\frac{3}{5}}}\frac{1}{(k^2-d^2)^{\frac{3}{2}}}\left(\frac{k-d}{k+d}\right)^{d}\exp\left(-\frac{d^2}{2n}\right)}\\
&= O\left(\sqrt{k^3}\sqrt{k^{-\frac{11}{2}}}\right) = O\left(k^{-\frac{5}{4}}\right).
\end{aligned} \tag{45}$$

Therefore, by (44) and(45), we have

$$\sum_{k=\epsilon^{-\frac{2}{5}}}^{\tilde{k}} ke^{-\frac{\delta k}{2}} \sum_{d=-k}^{k} \frac{d^2}{(k^2-d^2)^{\frac{3}{2}}} \left(\frac{k-d}{k+d}\right)^d \exp\left(-\frac{d^2}{2n}\right) = O\left(\sum_{k=\epsilon^{-\frac{2}{5}}}^{\tilde{k}} k^{-\frac{1}{4}} e^{-\frac{\delta k}{2}}\right) = O\left(\int_{y=1}^{\infty} y^{-\frac{1}{4}} e^{-\frac{\delta y}{2}} dy\right)$$
$$= \left(\epsilon^{-\frac{3}{2}} \int_{x=\frac{\epsilon^2}{4}}^{\infty} x^{-\frac{1}{4}} e^{-x} dx\right) = O\left(\epsilon^{-\frac{3}{2}}\right). \tag{46}$$

Hence, by (42), (43), and (46) we see that

$$S = O\left(\sqrt{\frac{n}{\epsilon}} + \epsilon^{-\frac{3}{2}}\right) = O\left(\sqrt{\frac{n}{\epsilon}}\right),$$

and so

$$\sum_{k_1=1}^{\tilde{k}} \sum_{k_2=1}^{\tilde{k}} k_1 k_2 \sum_{(i,j)\in\Lambda_{k_1}} \sum_{(s,t)\in\Lambda_{k_2}} \mu_{i,j}\mu_{s,t} \frac{(i-j)^2(t-s)^2}{n^2} = S^2 = O\left(\frac{n}{\epsilon}\right). \tag{47}$$

Hence, by (38), (39), (40), (41) and (47) we can conclude that

$$\mathrm{Var}(Z_1) \le \mathbb{E}(Z_2) + \frac{4\epsilon}{n}\mathbb{E}(Z_2)^2 + O\left(\frac{n}{\epsilon}\right). \tag{48}$$

Using (7) and Lemma 2.2, we can bound

$$\mathbb{E}(Z_2) \le \sum_{k=1}^{n^{\frac{2}{3}}} k^2 \sum_{(i,j)\in\Lambda_k} \mu_{i,j} = O\left(n \sum_{k=1}^{n^{\frac{2}{3}}} k^2 e^{-\frac{\delta k}{2}} \sum_{(i,j)\in\Lambda_k} \frac{1}{(ij)^{\frac{3}{2}}} \left(\frac{i}{j}\right)^{j-i} \exp\left(-\frac{(i-j)^2}{2n}\right)\right)$$
$$= O\left(n \sum_{k=1}^{n^{\frac{2}{3}}} \frac{1}{\sqrt{k}} e^{-\frac{\delta k}{2}}\right) = O\left(n \int_{y=1}^{\infty} \frac{1}{\sqrt{y}} e^{-\frac{\delta y}{2}} dy\right) = O\left(\frac{n}{\sqrt{\delta}} \int_{x=\frac{\delta}{2}}^{\infty} \frac{1}{\sqrt{x}} e^{-x} dx\right) = O\left(\frac{n}{\epsilon}\right).$$

Finally, putting this together with (48), we can conclude that

$$\mathrm{Var}(Z_1) \le O\left(\frac{n}{\epsilon}\right) + \frac{4\epsilon}{n} O\left(\frac{n^2}{\epsilon^2}\right) + O\left(\frac{n}{\epsilon}\right) = O\left(\frac{n}{\epsilon}\right).$$

Appendix C. Proof of Theorem 4.1

A standard argument tells us that, for a fixed vertex v the order of the component in $G(n,n,p)$ containing v is stochastically dominated by the order of the component of the root in a random subgraph of T_n, the infinite $(n+1)$-regular rooted tree, where we include each edge independently with probability p.

It is shown in [7, Corollary 3], that if we let $t(k,n)$ be the number of subtrees of T_n that contain the root and have order k and $k = \omega(1)$, then

$$t(k,n) \approx \frac{1}{\sqrt{2\pi} k^{\frac{3}{2}}} n^{k-1} \left(\frac{n}{n-1}\right)^{k(n-1)+2}.$$

Hence, the probability that the component of the root in a random subgraph of T_n has order k is given by

$$P_k(n,p) = t(k,n) p^{k-1} (1-p)^{k(n-1)+2} \approx \frac{(pn)^{k-1}}{\sqrt{2\pi} k^{1.5}} \left(\frac{n(1-p)}{n-1}\right)^{k(n-1)+2}.$$

It follows that, if we let $p = \frac{1-\epsilon}{n}$, then

$$P_k(n,p) \approx \frac{1}{\sqrt{2\pi} k^{1.5}} (1-\epsilon)^{k-1} \left(1 + \frac{\epsilon}{n-1}\right)^{k(n-1)+2} \le \frac{1}{\sqrt{2\pi} k^{1.5}} (1-\epsilon)^{k-1} \exp\left(\epsilon k + O\left(\frac{\epsilon}{n}\right)\right) \lesssim \frac{1}{\sqrt{2\pi} k^{1.5}} \left((1-\epsilon)e^{\epsilon}\right)^{k-1}.$$

Furthermore, it is clear by comparison with a branching process that with probability 1 the component of the root is finite, and hence $\sum_k P_k(n,p) = 1$. It follows that the probability that a vertex in $G(n,n,p)$ belongs to a component of order larger than $k_0 \in \mathbb{N}$ is equal to

$$\sum_{k \ge k_0} P_k(n,p).$$

Hence, if we let $Y_{\geq k_0}$ be the number of vertices in $G(n,n,p)$ which belong to a component of order larger than k_0, then we have that

$$\mathbb{E}(Y_{\geq k_0}) = n \sum_{k \geq k_0} P_k(n,p) \lesssim n \sum_{k \geq k_0} \frac{1}{\sqrt{2\pi}k^{1.5}} \left((1-\epsilon)e^{\epsilon}\right)^{k-1} \lesssim n k_0^{-\frac{3}{2}} \frac{((1-\epsilon)e^{\epsilon})^{k_0}}{1-(1-\epsilon)e^{\epsilon}}.$$

However, $(1-\epsilon)e^{\epsilon} = 1 - \frac{\epsilon^2}{2} + O(\epsilon^3)$ and so

$$\mathbb{E}(Y_{\geq k_0}) \lesssim n k_0^{-\frac{3}{2}} \frac{4}{\epsilon^2}.$$

Taking $k_0 = \sqrt{\frac{n}{3\epsilon}}$, we see that

$$\mathbb{E}\left(Y_{\geq \sqrt{\frac{n}{3\epsilon}}}\right) = O\left(\frac{n^{\frac{1}{4}}}{\epsilon^{\frac{5}{4}}}\right) = o\left(\sqrt{\frac{n}{\epsilon}}\right).$$