The probability of nonexistence of a subgraph in a moderately sparse random graph

Dudley Stark*
School of Mathematical Sciences
Queen Mary College
University of London

Nick Wormald†
School of Mathematical Sciences
Monash University

Abstract

We develop a general procedure that finds recursions for statistics counting isomorphic copies of a graph G_0 in the common random graph models $\mathcal{G}(n,m)$ and $\mathcal{G}(n,p)$. Our results apply when the average degrees of the random graphs are below the threshold at which each edge is included in a copy of G_0. This extends an argument given earlier by the second author for $G_0 = K_3$ with a more restricted range of average degree. For all strictly balanced subgraphs G_0, our results give much information on the distribution of the number of copies of G_0 that are not in large "clusters" of copies. The probability that a random graph in $\mathcal{G}(n,p)$ has no copies of G_0 is shown to be given asymptotically by the exponential of a power series in n and p, over a fairly wide range of p. A corresponding result is also given for $\mathcal{G}(n,m)$, which gives an asymptotic formula for the number of graphs with n vertices, m edges and no copies of G_0, for the applicable range of m. An example is given, computing the asymptotic probability that a random graph has no triangles for $p = o(n^{-7/11})$ in $\mathcal{G}(n,p)$ and for $m = o(n^{15/11})$ in $\mathcal{G}(n,m)$, extending results of the second author.

*Research initially undertaken while this author was a member of the Department of Mathematics and Statistics, University of Melbourne

†Research supported by the Australian Laureate Fellowships grant FL120100125. Also supported by the Australian Research Council while this author was a member of the Department of Mathematics and Statistics, University of Melbourne, and by the Canada Research Chairs program and NSERC while in the Department of Combinatorics and Optimization, University of Waterloo

1 Introduction

Our topic is the number of subgraphs of a random graph that are isomorphic to some given graph G_0. The perturbation method of [11] is used to derive recursions of ratios of random graph statistics describing the occurence of different types of clusters formed as edge-overlapping groups of copies of G_0. These recursions are used to investigate the probability of no occurrences of G_0, as well as other aspects of the distribution of clusters. For certain graphs G_0 and restrictions on p, we show that the probability that there are no copies of the graph in $\mathcal{G}(n,p)$ is the exponential of an appropriate truncation of a power series in n and p, with error factor $(1+o(1))$. (As is usual, $\mathcal{G}(n,p)$ denotes the random graph on n vertices obtained by choosing each edge in the graph to be present independently with probability p and $\mathcal{G}(n,m)$ denotes the random graph on n vertices obtained by choosing uniformly at random from the $\binom{\binom{n}{2}}{m}$ graphs having m edges.) By considering recursions involving both G_0 and isolated edges, we build on this result to show that the probability that there are no copies of G_0 in $\mathcal{G}(n,m)$ is given in the same way but by a different power series in n and d, where

$$d = \frac{m}{\binom{n}{2}}, \tag{1.1}$$

under corresponding restrictions on d.

Let $\nu(G)$ and $\mu(G)$ denote the number of vertices and number of edges of a graph G. A graph G_0 is *strictly balanced* if all its subgraphs are strictly less dense than G_0; that is,

$$\frac{\mu(G_0)}{\nu(G_0)} > \frac{\mu(G_1)}{\nu(G_1)}$$

for all nontrivial proper subgraphs G_1 of G_0. For example, the graph K_n is strictly balanced for all $n \geq 2$, as is every cycle. Let G_0 be strictly balanced, and let X be the number of copies of G_0 in the random graph $\mathcal{G}(n,p)$. Denote the set of proper subgraphs of G_0 which contain at least one edge by $\mathcal{E}$ and let $\chi > 0$ be defined by

$$\chi = \chi(G_0) = \max_{G_1 \in \mathcal{E}} \frac{\nu(G_0) - \nu(G_1)}{\mu(G_0) - \mu(G_1)}. \tag{1.2}$$

We will restrict the growth of p to $p = O(n^{-\chi-\epsilon})$ for some $\epsilon > 0$. The reason for this restriction is that when p is a little larger than $n^{-\chi}$ (sometimes called the *2-threshold*), each edge of $\mathcal{G}(n,p)$ will expect to be contained in many copies of G_0. Thus, there will be subgraphs consisting of arbitrarily large numbers of copies of G_0 "chained" together by shared edges. In this case our analysis will not apply, since it relies on a copy of G_0 being unlikely to overlap with any others, as happens when restricting to $p = O(n^{-\chi-\epsilon})$.

Here is our main result. Note that χ should not be confused with the chromatic number, which does not appear in this paper.

Theorem 1.1 *Let G_0 be strictly balanced and put $\chi = \chi(G_0)$. Let X be the number of copies of G_0 in $\mathcal{G}(n,p)$, or let X be the number of copies of G_0 in $\mathcal{G}(n,m)$ and set $p = m/\binom{n}{2}$. In each case, there is a formal power series $F = F(G_0) = \sum_{\ell \geq 0} c_\ell n^{i_\ell} p^{j_\ell}$, with i_ℓ and j_ℓ strictly* *positive for all ℓ, depending only on G_0, such that the following holds. For any $\epsilon > 0$, if $p = O(n^{-\chi-\epsilon})$, then*

$$\mathbb{P}(X = 0) = \exp\left(\sum_{\ell=0}^{M_\epsilon} c_\ell n^{i_\ell} p^{j_\ell} + o(1)\right), \tag{1.3}$$

where the bound implicit in $o(1)$ is uniform over all such p (but depends on ϵ), and M_ϵ is a constant depending only on ϵ and G_0. Moreover, $\ell > M_\epsilon$ if and only if $i_\ell < j_\ell(\chi + \epsilon)$.

Remarks

1. The theorem immediately gives an asymptotic formula for the number of G_0-free graphs on n vertices and m edges, for the values of m covered, by multiplying the $\mathcal{G}(n, m)$ case of (1.3) by $\binom{n(n-1)/2}{m}$.

2. Note that $i_\ell < j_\ell(\chi + \epsilon)$ if and only if $n^{i_\ell} p^{j_\ell} = o(1)$ when $p = n^{-\chi-\epsilon}$, so each term with $\ell > M_\epsilon$ is $o(1)$. We also note that the issue of non-convergence of the power series $F(G_0)$ for a given fixed n and p is not relevant in the present context.

3. The proof of the theorem contains a definition of the coefficients c_ℓ in Theorem 1.1 in terms of an algorithm by which they may be computed. It involves summing over a set of graphs whose size is bounded for fixed $\epsilon > 0$, but not as $\epsilon \to 0$.

We next give two specific examples of the main result, by restricting to that case that G_0 is a triangle, or K_3, and computing only the first few terms of the power series explicitly.

Theorem 1.2 *If $p = p(n) = o(n^{-7/11})$, the probability that the random graph $\mathcal{G}(n, p)$ is triangle-free is asymptotic to*

$$\exp\left(-\frac{1}{6}n^3p^3 + \frac{1}{4}n^4p^5 - \frac{7}{12}n^5p^7 + \frac{1}{2}n^2p^3 - \frac{3}{8}n^4p^6 + \frac{27}{16}n^6p^9\right).$$

Similarly, we determine the coefficients c_ℓ in the case of $\mathcal{G}(n, m)$ where $G_0 = K_3$ and $d = o(n^{-7/11})$, or equivalently $m = o(n^{15/11})$, in the next theorem.

Theorem 1.3 *If $m = m(n) = o(n^{15/11})$, the probability that the random graph $\mathcal{G}(n, m)$ is triangle-free is asymptotic to*

$$\exp\left(-\frac{1}{6}n^3d^3 - \frac{1}{8}n^4d^6 + \frac{1}{2}n^2d^3\right),$$

where $d = m/\binom{n}{2}$.

These two results on triangles agree with and extend those of the second author in [11], which applied for $p = o(n^{-2/3})$. The result for $\mathcal{G}(n, m)$ extended an earlier one of Frieze [3] which applied for even smaller p.

For $\mathcal{G}(n, m)$, the expected value of X is easily found to be

$$\begin{aligned} \lambda(G_0) &:= \binom{n}{\nu}\binom{m}{\mu}\binom{\binom{n}{2}}{\mu}^{-1} \nu! |\mathrm{aut}(G_0)|^{-1} \\ &\sim \hat{\lambda}(G_0) := \frac{(2m)^\mu}{n^{2\mu-\nu}|\mathrm{aut}(G_0)|} \end{aligned}$$

where $\nu = \nu(G_0)$, $\mu = \mu(G_0)$, and $|\mathrm{aut}(G_0)|$ denotes the number of automorphisms of G_0. Ruciński [10] showed that the distibution of X is asymptotically Poisson essentially for d up to $n^{-\chi}$. Frieze [3, Remark 2, P.69] raised the possibility that, for the same range of d, the number of graphs with k copies of G_0 in $\mathcal{G}(n,m)$ is asymptotic to the probability that the Posson random variable with mean $\hat{\lambda}(G_0)$ is equal to k, for all "small" k. Theorem 1.3 shows (for the first time!) that this is false in particular for $k=0$ and $G_0 = K_3$, since in this case, $\chi = 1/2$ but already for $m = n^{4/3}$, other terms are entering the asymptotic formula in a significant way. Moreover, the situation is not remedied by using (the more natural) Poisson with mean $\lambda(G_0)$, since $\lambda(G_0) = \frac{1}{6}n^3d^3 - \frac{1}{2}n^2d^3 + o(1)$ (using $nd^2 = O(m^2/n^3) = o(1)$ for the range of m under consideration).

We note that it may be possible to modify our approach to cater also for subgraphs that are not strictly balanced. In some cases, for instance where G_0 has a unique densest subgraph, the desired result can be deduced immediately from our results. However, other cases are more delicate, with different subgraphs of G_0 'competing'. One would need to incorporate considerations similar to those in the determination the threshold of appearance of G_0, as was done by Bollobás [1].

Our concern here is to obtain an asymptotic formula for the probability that a random graph in $\mathcal{G}(n,p)$ or $\mathcal{G}(n,m)$ is G_0-free, for a fixed graph G_0, where the density of the random graph is small enough that there are no large clusters of copies of G_0. Our methods will not work for the denser case, but some results are already known there, and for arbitrary densities. Recall, as in Remark 2 above, that for $\mathcal{G}(n,m)$ our problem is equivalent to enumerating m-edged graphs with a forbidden subgraph. The classic paper of Erdős, Kleitman and Rothschild [2] gives the number of triangle-free graphs with n vertices, in total, asymptotically (and asymptotics of the logarithm of the number when $G_0 = K_t$). These results also demonstrate the connection between enumeration and the extremal numbers of edges for G_0-free graphs. There are many other similar results, which we refrain from mentioning as they do not take into account the edge density of the host graph. More related to the problem at hand, Prömel and Steger [9] found an asymptotic formula for the number of triangle-free graphs with n vertices and m edges when $m > cn^{7/4}\log n$, by showing that they are almost all bipartite. This was extended by Osthus, Prömel and Taraz [7] to cover all m that are at least slightly above $n^{3/2}$. Before this, Łuczak [6] had found asymptotics of the logarithm of the number.

For more general subgraphs than the triangle, and general p, asymptotic formulae for the actual numbers (or probabilities) are elusive. The *logarithm* of the probability that $\mathcal{G}(n,p)$ is G_0-free was estimated within a constant factor by Janson, Łuczak and Ruciński [4]. This was extended by Prömel and Steger [8] to similar bounds on $\mathbb{P}\big(\mathcal{G}(n,m)$ is G_0-free$\big)$.

Many results are known on the distribution of the number of copies of a fixed subgraph in $\mathcal{G}(n,p)$ and $\mathcal{G}(n,m)$; see for example [5, Chapter 6], but this is not our concern in this paper.

Our basic approach, and its background, are discussed in [11]. The proof for $\mathcal{G}(n,p)$ estimates ratios of numbers of graphs using induction on the numbers of edge-overlapping clusters of copies of G_0 up to a given size; for $\mathcal{G}(n,m)$ the number of edges not in copies of G_0 is also used, and the base step of this induction is essentially given by the n-vertex graph with no edges. There are two major extensions to the argument in [11]. One is that the graph G_0 is no longer restricted to K_3. This extension requires mainly graph theoretic arguments related to the ways that multiple copies of a graph can overlap. The other is that the range of p permits edge-overlapping clusters containing arbitrarily many copies of G_0 to appear in the typical random graph under consideration. Thus our asymptotic estimates involve polynomials of unbounded size, and this poses significant problems in characterising and managing those estimates (see Corollary 2.7 for example).

The working assumption on $p = p(n)$ we will make in our proofs is $p = n^{-\kappa+o(1)}$ where $\kappa \geq \chi + \epsilon$ is *fixed*. This assumption can be weakened to obtain asymptotic results that hold uniformly over more general $p = p(n) = O(n^{-\chi-\epsilon})$ by using the following lemma. Here a and b are finite but the same result holds (with appropriate interpretation) without this assumption.

Lemma 1.4 *For a closed interval $[a, b]$, suppose that $f(n, p)$ is a function such that $f(n, p) \to 0$ as $n \to \infty$ for all p of the form $p = n^{-\kappa+o(1)}$ when $\kappa \in [a, b]$ is fixed. Then $f(n, p) \to 0$ uniformly for all $p(n)$ satisfying $p(n) = n^{-\kappa(n)}$ with $\kappa(n) \in [a, b]$ for all n.*

Proof. If $p(n)$ satisfies $-\log_n p \in [a, b]$ for all n, then any subsequence of $\big(p(n)\big)_{n\geq 1}$ has a subsubsequence for which $-\log_n p \to \kappa'$ for some fixed $\kappa' \in [a, b]$. On this subsubsequence, $f(n, p) \to 0$ by assumption. So the lemma follows from the subsubsequence principle (see [5, p.12]) applied to the sequence $\big(f(n, p(n))\big)_{n\geq 1}$. ∎

Our results will give information on the distribution of the number of copies of a strictly balanced subgraph, not just the probability that the number is 0, but we postpone this investigation to another paper. We believe that it should be possible to modify our approach so as to obtain accuracy in the formulae to any desired power of n^{-1}. Specifically, the power series in Theorem 1.1 should give valid lower order correction terms to the asymptotic formulae. However, we have avoided attempting this and there are some steps in the present argument that would have to be replaced in order to carry it out.

Some basic definitions are made and results are proved in Section 2; the $\mathcal{G}(n, p)$ case of Theorem 1.1 is proved in in Section 3; the $\mathcal{G}(n, m)$ case is proved in in Section 4; Theorems 1.2 and 1.3 are proved in Appendix A.

2 Clusters and recursions for counting maximal clusters

We assume for a general framework that Ω is any finite set. A family $\mathcal{K}$ of subsets of Ω is called a *clustering* if $C_1 \in \mathcal{K}$, $C_2 \in \mathcal{K}$ and $C_1 \cap C_2 \neq \emptyset$ imply that $C_1 \cup C_2 \in \mathcal{K}$. The elements of $\mathcal{K}$ are called *clusters*.

We will consider here only the case that $\Omega = \Omega_n$ is the set of edges of the complete graph K_n on n vertices, although the same principles can also be applied to clusterings in general. As a further restriction, to focus on small subgraph counts, we only consider very special clusterings, for which simplification occurs by taking advantage of the symmetries of K_n. We take a fixed graph G_0 throughout this paper, and will investigate the distribution of the number of subgraphs of a random graph isomorphic to G_0. The edge set of any subgraph of K_n isomorphic to G_0 is called an *elementary G_0-cluster*. Mostly, we deal with the minimal clustering which has every elementary G_0-cluster as a member. We call this the *G_0-clustering* of Ω. Equivalently, $J \subseteq \Omega$ is in the G_0-clustering if and only if there is a sequence $J_1, \dots, J_i$ of subsets of Ω such that each J_j is an elementary G_0-cluster, $\bigcup_{j=1}^{i} J_j = J$, and $J_k \cap \bigcup_{i=1}^{k-1} J_j \neq \emptyset$ for $2 \leq k \leq i$. (This definition of clusters corrects an error in the definition in [11]. The usage of it in [11] is consistent with the present definition.)

More generally, suppose $\mathcal{R}$ is any fixed set of nonempty graphs, and information is desired on the joint distribution of the subgraph counts for the graphs in $\mathcal{R}$. Then the appropriate clustering to consider is the minimal clustering containing every elementary G-cluster for every $G \in \mathcal{R}$. We call this the *clustering generated by* $\mathcal{R}$. Of course, if $\mathcal{R} = \{G_0\}$, this is simply the G_0-clustering.

Henceforth in this paper we consider the clustering generated by a fixed set of graphs $\mathcal{R}$, and assume that each graph in $\mathcal{R}$ has no isolated vertices. Our first proposition considers a general set $\mathcal{R}$, and after that we restrict to only two kinds of clustering: the G_0-clustering, and the one generated by $\mathcal{R} = \{G_0, K_2\}$, which we call the G_0^**-clustering*. Note that a 1-element subset of Ω cannot have a nontrivial proper intersection with any other cluster. It follows that the G_0^*-clustering consists of the clusters of the G_0-clustering, together with all the 1-element subsets of Ω. We assume in all cases that $|E(G_0)| \geq 2$.

For $H \subseteq \Omega$, a *cluster of* H is any cluster in $\mathcal{K}$ contained in H. A *maximal cluster* Q of H is cluster of H which is contained in no larger cluster of H. Equivalently, Q is a subset of H such that $Q \in \mathcal{K}$ and such that for every $J \in \mathcal{K}$ with $J \subseteq H$, either $J \subseteq Q$ or $J \cap Q = \emptyset$. (The case of nonempty intersection is excluded by the definition of a clustering.) For example, if $\mathcal{K}$ is the G_0-clustering and H is an arbitrary subset of Ω, a maximal cluster of H whose cardinality is $|E(G_0)|$ must be an elementary G_0-cluster contained in H having empty intersection with every other elementary G_0-cluster in H.

Being a subset of Ω, a cluster induces a subgraph of K_n. The isomorphism class of the subgraph is called the *type* of the cluster and also of the subgraph. The set of types will be denoted $\mathcal{T}$, and we use τ to denote the function which maps a cluster or the corresponding graph to its type. Given $t \in \mathcal{T}$, we use the notation $|t| := |\{S \subseteq \Omega : \tau(S) = t\}|$. Note that this depends on n, whereas t is fixed.

We will define a special nonempty finite set $\mathcal{S}$ of types which is closed under taking subsets, i.e. which satisfies

$$\text{if } S,\, S' \in \mathcal{K},\, \tau(S) \in \mathcal{S} \text{ and } S' \subseteq S \text{ then } \tau(S') \in \mathcal{S}.$$

Let $s = |\mathcal{S}|$ be the number of types in $\mathcal{S}$.

The types in $\mathcal{S}$ will be called *small*, and any cluster Q with $\tau(Q) \in \mathcal{S}$ is also called small. Any type or cluster which is not small is called *large*. An *unavoidable* cluster is any large cluster which is a union of a small cluster Q and a set of small clusters all pairwise disjoint and all having nonempty intersection with Q. The set of types of unavoidable clusters is denoted by $\mathcal{U}$. (The term "unavoidable" refers to the fact that large clusters created in a certain way, to be specified later, cannot avoid being in $\mathcal{U}$.)

We will need to record how many subgraphs of every small type are present in a given graph. So we consider the set $\mathcal{F}$ of all non-negative integer functions defined on $\mathcal{S}$. For any $H \subseteq \Omega$, define s_H to be the function in $\mathcal{F}$ such that, for all $t \in \mathcal{S}$, $s_H(t)$ is the number of maximal clusters of H of type t. The function $\delta_t \in \mathcal{F}$ has value 1 at t and 0 elsewhere.

All our basic work is in $\mathcal{G}(n,p)$, the standard edge-independent (binomial) model for random graphs, and $\mathbb{P}$ and $\mathbb{E}$ denote probability and expectation in this space. G denotes a random graph in $\mathcal{G}(n,p)$ and q always denotes $1-p$. For $H \subseteq \Omega$, the event $H \subseteq E(G)$ is denoted by A_H, so that $\mathbb{P}(A_H) = p^{|H|}$. The main objects we work with are, for each $f \in \mathcal{F}$, the set $\mathcal{C}_f$ consisting of graphs G on n vertices containing no large clusters and such that $s_{E(G)} = f$. For $f \notin \mathcal{F}$, for example if f has a negative value on $\mathcal{S}$, we define $\mathcal{C}_f = \emptyset$. We write $\mathbb{P}(\mathcal{C}_f)$ for $\mathbb{P}(\mathcal{G}(n,p) \in \mathcal{C}_f)$.

For types $u, t \in \mathcal{S}$ and for $h \in \mathcal{F}$, define, for any fixed cluster J of type u,

$$c(u,t,h) = \sum_{\substack{Q \in \mathcal{K} \\ Q \subseteq J \\ \tau(Q)=t}} \sum_{\substack{H \subseteq J \\ H \cup Q = J \\ s_H = h}} p^{|Q\cap H|} q^{|J \setminus H|}. \tag{2.1}$$

Since the clustering generated by any set $\mathcal{R}$ is symmetrical, $c(u,t,h)$ is clearly independent of the choice of J with $\tau(J) = u$. Note that in the special case $u = t$,

$$c(t,t,h) = \sum_{\substack{H \subseteq J \\ s_H = h}} p^{|H|} q^{|J \setminus H|}, \tag{2.2}$$

and in particular

$$c(t,t,\mathbf{0}) = 1 + O(p). \tag{2.3}$$

We use $\nu(G)$ and $\mu(G)$ for the numbers of vertices and edges of a graph G respectively, and extend the notation to arbitrary subsets H of Ω, so that $\nu(H)$ is the number of vertices of the graph induced by H and $\mu(H)$ is the number of edges. In particular, this applies to clusters H. We also use $\nu(t)$ for the number of vertices in each cluster of type t and $\mu(t)$ for the number of edges.

Let $[n]_k$ denote $n(n-1)\cdots(n-k+1)$. For $t \in \mathcal{T}$, let Q be any cluster of type t and $|\mathrm{aut}(Q)|$ the number of automorphisms of the graph induced by Q. Then

$$|t| = \frac{[n]_{\nu(Q)}}{|\mathrm{aut}(Q)|}, \tag{2.4}$$

and

$$\lambda_t := |t| p^{\mu(Q)} = \Theta(n^{\nu(Q)} p^{\mu(Q)}) \tag{2.5}$$

is the expected number of different copies, in $G \in \mathcal{G}(n,p)$, of the subgraph induced by Q.

Our first result is obtained by simple counting.

Proposition 2.1 *For $f \in \mathcal{F}$ and $t \in \mathcal{S}$,*

$$\frac{\mathbb{P}(\mathcal{C}_{f+\delta_t})}{\mathbb{P}(\mathcal{C}_f)} = \frac{\lambda_t}{(f(t)+1)c(t,t,\mathbf{0})} \left(1 - \Sigma - \frac{\theta(f,\delta_t)}{|t|\mathbb{P}(\mathcal{C}_f)}\right)$$

where

$$\Sigma = \sum_{\substack{u \in \mathcal{S} \\ h \in \mathcal{F} \\ (u,h) \neq (t,\mathbf{0})}} \frac{(f(u)-h(u)+1)c(u,t,h)\mathbb{P}(\mathcal{C}_{f-h+\delta_u})}{\lambda_t \mathbb{P}(\mathcal{C}_f)} \tag{2.6}$$

and

$$0 \le \theta(f, \delta_t) \le \sum_{L\,:\,\tau(L)\in\mathcal{U}} \sum_{\substack{Q,H\subseteq L\\ \tau(Q)=t\\ L\setminus Q\subseteq H}} \mathbb{P}(\mathcal{C}_{f-s_H}) \left(\frac{p}{q}\right)^{|H|}. \tag{2.7}$$

Proof. Note that

$$c(u,t,h)p^{-\mu(t)} = \sum_{\substack{Q\in\mathcal{K}\\ Q\subseteq J\\ \tau(Q)=t}} \sum_{\substack{H\subseteq J\\ H\cup Q=J\\ s_H=h}} \frac{q^{|J\setminus H|}}{p^{|Q\setminus H|}} = \sum_{\substack{Q\in\mathcal{K}\\ Q\subseteq J\\ \tau(Q)=t}} \sum_{\substack{H\subseteq J\\ H\cup Q=J\\ s_H=h}} \left(\frac{q}{p}\right)^{|J\setminus H|}, \tag{2.8}$$

where we have used the fact that $J \setminus H = Q \setminus H$ follows from $H \cup Q = J$. Consider a pair (E, Q) where E is the edge set of a graph G in $\mathcal{C}_f$ and Q is a cluster of type t. Let J be the maximal cluster of $E \cup Q$ containing Q. If G' is the graph with edge set $E \cup Q$, then the expression $(q/p)^{|J\setminus H|}$ in (2.8) is $\mathbb{P}(G)/\mathbb{P}(G')$. Classifying $E \cup Q$ according to $u = \tau(J)$, and, in the case that $u \in \mathcal{S}$, subclassifying according to $h = s_H$ where $H = E \cap J$, gives

$$|t|\mathbb{P}(\mathcal{C}_f) = \Big(\sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}}} (f(u) - h(u) + 1)c(u,t,h)p^{-\mu(t)}\mathbb{P}(\mathcal{C}_{f-h+\delta_u})\Big) + \theta(f, \delta_t), \tag{2.9}$$

where the θ term is bounded as in the statement of the proposition. This term comes from observing that if J is a large cluster L, then it must be unavoidable since E has no large clusters, and from considering the subset of Ω obtained by removing the set H of all edges of E in L. Multiplying (2.9) by $p^{\mu(t)}$ gives

$$\lambda_t\mathbb{P}(\mathcal{C}_f) = \Big(\sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}}} (f(u) - h(u) + 1)c(u,t,h)\mathbb{P}(\mathcal{C}_{f-h+\delta_u})\Big) + \theta(f, \delta_t)p^{\mu(t)}$$

and rearranging the terms, isolating the one with $(u, h) = (t, \mathbf{0})$, finishes the proof. ∎

We now lay the groundwork for asymptotic results. Henceforth, we consider only the G_0- and G_0^*-clusterings for some fixed graph G_0 with at least two edges. Recalling that $|E(G_0)| \ge 2$, we define the *extension value* of G_0 to be

$$x = x(G_0, p, n) = \max_{G_1\in\mathcal{E}} n^{\nu(G_0)-\nu(G_1)}p^{\mu(G_0)-\mu(G_1)}. \tag{2.10}$$

For example, if G_0 is a triangle,

$$x = \max(np^2, p, p^2) = \max(np^2, p). \tag{2.11}$$

The significance of the extension value lies in the fact that $n^{\nu(G_0)-\nu(G_1)}p^{\mu(G_0)-\mu(G_1)}$ is the asymptotically important part of

$$\binom{n-\nu(G_1)}{\nu(G_0)-\nu(G_1)}p^{\mu(G_0)-\mu(G_1)}.$$

To interpret this quantity, first distinguish one of the subgraphs of G_0 isomorphic to G_1. For G_2 isomorphic to G_1, conditional upon $G_2 \subseteq \mathcal{G}(n,p)$, the quantity above is the expected number of isomorphisms from G_0 to a subgraph of $\mathcal{G}(n,p)$ that map the distinguished copy of G_1 onto G_2.

For $H \subseteq \Omega$ define $\Phi(H, G_0)$ to be the expected number of subgraphs of $G \in \mathcal{G}(n,p)$ that are isomorphic to G_0 and whose edge set contains H, conditional on $H \subseteq E(G)$. If H is moreover a nonempty proper subset of the edge set of a copy of G_0, it follows from the remarks above that $\Phi(H, G_0)$ is $O(x)$, since there is a bounded number of ways to distinguish one of the subgraphs of G_0 isomorphic to G_1.

Put a partial ordering on the set of types by defining t to be strictly less than u in the poset, denoted by $t \prec u$, if, and only if, any cluster of type u properly contains a cluster of type t. If $t \prec u$, then a cluster of type u can be obtained from a cluster Q of type t by a finite sequence of non-disjoint unions with clusters $Q_0, \ldots, Q_k$ such that each Q_i is the edge set of a graph isomorphic to some $G_i \in \mathcal{R}$ and $Q_i \not\subseteq Q \cup (\bigcup_{j=0}^{i-1} Q_j)$. (Note that, in the G_0^*-clustering, it must be that $G_i = G_0$ for all i.) Thus, for $G \in \mathcal{G}(n,p)$ the expected number of clusters of type u in $E(G)$ can be bounded above by a finite sum whose terms are all of the form $\lambda_t \prod_{i=0}^{k} \Phi(H_i, G_i)$ where H_i corresponds to the intersection of Q_i with $Q \cup (\bigcup_{j=0}^{i-1} Q_j)$. Hence, from the conclusion of the previous paragraph, provided $x = o(1)$ we have

$$\text{if } t \prec u \text{ then } \frac{\lambda_u}{\lambda_t} = O(x). \tag{2.12}$$

Henceforth in this paper, we assume that G_0 is strictly balanced, with at least two edges. Let X be the number of copies of G_0 in the random graph $\mathcal{G}(n,p)$. It follows easily from the definition (2.10) of x that the constant χ defined in (1.2) is the smallest number such that $p = o(n^{-\chi})$ implies $x = o(1)$. Hence, there are functions $p = p(n)$ such that $\lambda_{\tau(G_0)} \to \infty$ while $x(G_0, p, n) = o(1)$. We also assume henceforth that $p = p(n)$ is restricted so that for some fixed $\kappa > \chi$,

$$p = n^{-\kappa + o(1)}. \tag{2.13}$$

This will be enough for our purposes in view of Lemma 1.4.

Fix $\epsilon > 0$ and let $\kappa \geq \chi + \epsilon$. Since $\mu(G_1) < \mu(G_0)$ for all $G_1 \in \mathcal{E}$, the expression maximised in (2.10) is at most $(n^\chi p)^{\mu(G_0) - \mu(G_1)} \leq n^\chi p$. Thus,

$$x(G_0, p, n) = O(n^{-\epsilon + o(1)}). \tag{2.14}$$

See [5] for a general introduction to the considerations relevant here. Note that

$$p \leq x \tag{2.15}$$

by definition, as shown by setting the graph G_1 in (2.10) equal to G_0 minus an edge.

For our asymptotic results, we work with a particular set of small cluster types defined as follows:

$$\mathcal{S} = \{t : \nu(t)/\mu(t) \geq \kappa\}. \tag{2.16}$$

Then for $t \in \mathcal{S}$, the expected number λ_t of subgraphs of type t is bounded below by $\lambda_t \geq n^{-o(1)}$ (here the negative sign is not necessary, just indicative, since $o()$ bounds the absolute value), since by (2.4), (2.5) and (2.13),

$$\lambda_t = \Theta(n^{\nu(t)-\kappa\mu(t)+o(1)}). \tag{2.17}$$

The set $\mathcal{S}$ is finite by (2.14) and (2.12). Hence, defining

$$\lambda_{\mathcal{L}} := \sup_{t \notin \mathcal{S}} \lambda_t \tag{2.18}$$

we obtain

$$\lambda_{\mathcal{L}} = O(n^{-\epsilon'}) \tag{2.19}$$

for some $\epsilon' > 0$ by our definition of $\mathcal{S}$. While we are at it, due to a technicality we assume $\kappa < 2$, so that p satisfies the very weak growth condition

$$n^2 p > n^{\epsilon''} \tag{2.20}$$

for some $\epsilon'' > 0$. This ensures that the number of edges in the random graph tends to infinity at a reasonable rate. Imposing this condition is without loss of generality, since the omitted case follows from the case considered. For example, the p such that $p \sim n^{-c\nu(G_0)/\mu(G_0)}$ are covered for all $1 < c < 2\mu(G_0)/\nu(G_0)$, and this is well below the threshold of appearance of copies of G_0. Hence, each term in the power series must tend to zero for such c, and must also tend to 0 when $\kappa \geq 2$. The assumption $\kappa < 2$ also ensures that, in the case of the G_0^*-clustering, the single edge cluster is in $\mathcal{S}$. Note that if $n^2 p = o(\sqrt{n})$, the random graph is in any case not interesting, as it is asymptotically almost surely a matching.

Define

$$\mathcal{S}_0 = \{t \,:\, \nu(t)/\mu(t) = \kappa\}, \quad \mathcal{S}_1 = \mathcal{S} \setminus \mathcal{S}_0,$$
$$m_t = \begin{cases} 3\lambda_t & \text{if } t \in \mathcal{S}_1 \\ \lambda_t \log n & \text{if } t \in \mathcal{S}_0. \end{cases} \tag{2.21}$$

Note that $\mathcal{S}_0$ will often be empty, but if it is nonempty, the types in $\mathcal{S}_0$ are the rarest types of small clusters in the random graph, and for $t \in \mathcal{S}_0$, we have $\lambda_t = n^{o(1)}$ and hence $m_t = n^{o(1)}$. Any type in $\mathcal{S}_0$ is maximal in $\mathcal{S}$ by (2.12). Thus, for later reference we may note that, for some positive ϵ''',

$$\lambda_t > n^{\epsilon'''} \text{ for } t \in \mathcal{S}_1, \quad \lambda_t = n^{o(1)} \text{ for } t \in \mathcal{S}_0. \tag{2.22}$$

Let $\mathcal{F}_{\mathcal{S}} = \mathcal{F}_{\mathcal{S}}(n)$ be the set containing those functions $f \in \mathcal{F}$ such that for all $t \in \mathcal{S}$,

$$f(t) \leq m_t. \tag{2.23}$$

For integer-valued h with $f, f+h \in \mathcal{F}$, we define

$$\rho(f,h) = \frac{\mathbb{P}(\mathcal{C}_{f+h})}{\mathbb{P}(\mathcal{C}_f)} \tag{2.24}$$

and for $t \in \mathcal{T}$, $f \in \mathcal{F}$ define

$$\gamma(f,t) = \frac{\rho(f,\delta_t)(f(t)+1)}{\lambda_t}. \tag{2.25}$$

The motivation for focussing on γ is that if the numbers of clusters of the various small types were independent Poisson variables, then all the γ's would be exactly 1. Proving that they are close to 1 shows that the variables are approximately Poisson. We will be measuring the difference between the Poisson probability and the true probability of $\mathcal{C}_f$ very accurately for some values of f.

Ultimately, we wish to estimate $\gamma(f,t)$, and will achieve this in Corollary 2.7. The proof is complicated, so is broken up into several parts, obtaining progressively simpler approximations. The downside of breaking it up like this is that it requires repeating the same kinds of inductive arguments several times. We first obtain a more useful bound on the function $\theta(f,\delta_t)$ appearing in Proposition 2.1. Let t^* denote the type of the single edge cluster, which of course only appears in the G_0^*-clustering.

Proposition 2.2 *Uniformly for every $f \in \mathcal{F}_\mathcal{S}$ and every $t \in \mathcal{S}$,*

$$\frac{\theta(f,\delta_t)}{|t|\mathbb{P}(\mathcal{C}_f)} = O\left(\frac{\phi_t \lambda_\mathcal{L}}{\lambda_t}\right),$$

where $\phi_t = n^{o(1)}$ for $t = t^*$ and $\phi_t = 1$ otherwise. Moreover, for all $f \in \mathcal{F}_\mathcal{S}$ and $t \in \mathcal{S}$, uniformly,

$$\gamma(f,t) = 1 + O(xn^{o(1)}).$$

Note. The proof will reveal that the factor $n^{o(1)}$ can be replaced by the maximum of $f(t')/\lambda_{t'}$ for $t' \in \mathcal{S}$, which is always at most $\log n$. However, $n^{o(1)}$ is tight enough for our purposes here. Also, $\lambda_\mathcal{L}$ can be replaced by the maximum value of λ_u over all $u \in \mathcal{L}$ such that $t \prec u$.

Proof. In this proof, as in the proposition's statement, the constants implicit in the $O()$ terms depend only on the choice of clustering and κ, as do the bounds implicit in the notation $\sim$ and $o(1)$. We will use induction on $f \in \mathcal{F}_\mathcal{S}$. Order $\mathcal{F}_\mathcal{S}$ lexicographically; that is $g < f$ if, and only if, $g \neq f$ and g has a smaller value than f in the first component at which they differ. This induction is crucial to the whole approach of this paper, and is rather unusually complex, since for the G_0^*-clustering, the induction actually begins with the graph on n vertices and no edges. So we formulate a statement that pays explicit attention to the implicit constants in $O()$: what we claim is that there exists constants C and C', a number N_0 and a function $1 \le \phi^* = \phi^*(n) = n^{o(1)}$ (all depending only on the clustering and κ) such that, for $n \ge N_0$ and all relevant f and t,

$$\frac{\theta(f,\delta_t)}{|t|\mathbb{P}(\mathcal{C}_f)} \le C\phi_t \frac{\lambda_\mathcal{L}}{\lambda_t}, \tag{2.26}$$

where $\phi_t = \phi^*$ for $t = t^*$ and $\phi_t = 1$ otherwise, and furthermore

$$|\gamma(f,t) - 1| \le C'\phi_t x \le 1/2. \tag{2.27}$$

To prove this, we can assume that for this particular C, and n large enough, these inequalities hold when f is replaced by any $g < f$ (in the lexicographic ordering).

We first discuss the bound involving θ. Here, by (2.18), it is enough to show the bound $C\phi_t\lambda_{\tau(L)}/\lambda_t$ where $\tau(L) \notin \mathcal{S}$ (which then justifies the second part of the note after the statement of the proposition). Moreover, of (2.27) we will only use the inequality

$$|\gamma(f,t) - 1| \le 1/2. \tag{2.28}$$

Since the number of clusters of the complete graph K_n which are isomorphic to a given L is $O(n^{\nu(L)})$, and since the number of types of unvoidable clusters is by definition bounded, we may use (2.7) and $q \sim 1$ to obtain the bound

$$\frac{\theta(f,t)}{|t|\mathbb{P}(\mathcal{C}_f)} = O(1) \max_{\substack{\tau(L)\in\mathcal{U} \\ \tau(Q)=t, Q\subseteq L \\ L\setminus Q\subseteq H\subseteq L}} n^{\nu(L)-\nu(Q)} p^{|H|} \frac{\mathbb{P}(\mathcal{C}_{f-s_H})}{\mathbb{P}(\mathcal{C}_f)} \tag{2.29}$$

for n sufficiently large (which in particular ensures that $\mathbb{P}(\mathcal{C}_{f-s_H}) \neq 0$). Here, recalling (2.5) we see that

$$|H| \geq |L| - |Q|, \quad \lambda_t = O(n^{\nu(Q)} p^{|Q|}), \quad n^{\nu(L)} p^{|L|} = O(\lambda_{\tau(L)}). \tag{2.30}$$

In the case $f = \mathbf{0}$, we may assume $s_H = \mathbf{0}$ in (2.29), since otherwise, $\mathcal{C}_{f-s_H}$ is empty. Thus, by (2.30), we have the bound $O(\lambda_{\tau(L)}/\lambda_t)$ on each term in (2.29). Since $\tau(L) \notin \mathcal{S}$, we are done in this case.

In the case $\mathbf{0} \neq f \in \mathcal{F}_{\mathcal{S}}$, suppose the claim has been shown when f is replaced by any $g < f$. We need to show that, when C is large enough, the very same C applies in the statement for f. Denoting a general term in the maximum in (2.29) by M, since $\tau(L) \in \mathcal{U} \subseteq \mathcal{L}$, it suffices to show that $M = O(\lambda_{\tau(L)}/\lambda_t)$, or $M = O(n^{o(1)}\lambda_{\tau(L)}/\lambda_t)$ in the case of the G_0^*-clustering (and then choosing ϕ^* appropriately). We may write

$$\frac{\mathbb{P}(\mathcal{C}_{f-s_H})}{\mathbb{P}(\mathcal{C}_f)} = \prod_{i=1}^{k} \rho(f_i\,, -\delta_{u_i}) \tag{2.31}$$

for some sequence $u_1, u_2, \ldots, u_k$ in $\mathcal{S}$ such that $\sum_{i=1}^{k} \delta_{u_i} = s_H$ and where $f_i = f - \sum_{j=1}^{i-1} \delta_{u_j}$. By definition, an unavoidable cluster has size at most $r(r-1)$ where r is the size of the largest small cluster. Hence, the upper index k in the above product is at most $r(r-1)$. Note that each f_i occurs before f in the lexicographic order, and (2.28) inductively implies $1/2 \leq \gamma(f_j - \delta_{u_j}, \delta_{u_j}) \leq 3/2$ for all $j \geq 1$. Note also that

$$\rho(f_j, -\delta_{u_j}) = \frac{1}{\rho(f_j - \delta_{u_j}, \delta_{u_j})} = \frac{f_j(u_j)}{\lambda_{u_j}\gamma(f_j - \delta_{u_j}, u_j)}.$$

Suppose firstly that, in (2.31), $u_i \in \mathcal{S}_1$ for all i. Then by (2.23), $f_j(u_i)/\lambda_{u_i} \leq 3$ for all i, and by (2.28) inductively $\gamma(f_i - \delta_{u_i}, u_i)^{-1} \leq 2$, so we deduce that the product in (2.31) is $O(1)$. Now (2.30) implies that $M = O\left(\lambda_{\tau(L)}/\lambda_t\right)$, as required.

Suppose on the other hand that, for $i = j'$ in (2.31), we have $u_{j'} \in \mathcal{S}_0$. Recall that $\lambda_{u_{j'}} = n^{o(1)}$ by (2.22), and hence

$$\frac{\mathbb{P}(\mathcal{C}_{f-s_H})}{\mathbb{P}(\mathcal{C}_f)} = \rho(f, -s_H) = O(n^{o(1)}) \tag{2.32}$$

using the same argument as for analysing (2.31) above. Also note that

$$n^{\nu(L)-\nu(Q)} p^{|H|} = n^{\nu(L)-\nu(Q)} p^{|L\setminus Q|} p^{|H\cap Q|} = O(\lambda_{\tau(L)}/\lambda_t) p^{|H\cap Q|}. \tag{2.33}$$

There are two subcases to consider. Firstly, if $|H \cap Q| \geq 1$, then $p^{|H\cap Q|}n^{o(1)} \leq pn^{o(1)} = o(1)$ and hence $M = O(\lambda_{\tau(L)}/\lambda_t)$ as required. The second subcase is $|H \cap Q| = 0$. Then a cluster Q' of type $u_{j'}$ that H contains must be disjoint from Q. It follows that there is a sequence $Q_1, \dots, Q_\ell$ of elementary clusters, each nontrivially intersecting the next, with $Q_1 \cap Q' \neq \emptyset$ and $Q_\ell \cap Q \neq \emptyset$. We also suppose that ℓ is minimal, so that $Q_i \cap Q = \emptyset$ for all $i < \ell$, and in particular this implies $Q_\ell \neq Q$. We will consider two subsubcases of this second case.

Suppose firstly that $Q \not\subseteq Q_\ell$, and so $Q'' := Q' \cup \bigcup_{i=1}^{\ell} Q_i$ is a cluster satisfying $Q' \subset Q'' \subset L$, where the inclusions are proper and $\tau(Q') = u_{j'}$. It follows by (2.12) and (2.22) that $\lambda_{\tau(Q'')} = O(\lambda_{u_{j'}} x) = O(n^{o(1)}x)$ since $u_{j'} \in \mathcal{S}_0$. Thus $\tau(Q'') \in \mathcal{L}$, and hence by the definition (2.18) of $\lambda_{\mathcal{L}}$, we have $\lambda_{\tau(Q'')} \leq \lambda_{\mathcal{L}}$. Similarly, $\lambda_{\tau(L)} = O(x\lambda_{\tau(Q'')}) = O(x\lambda_{\mathcal{L}})$, and now using (2.32) and (2.33) in (2.29) gives $M = O\left(x\lambda_{\tau(L)}n^{o(1)}/\lambda_t\right) = O\left(\lambda_{\tau(L)}/\lambda_t\right)$ as required.

For the other subsubcase $Q \subseteq Q_\ell$, recall that $Q_\ell \neq Q$. As Q_ℓ is elementary, it follows that this can only occur for the G_0^*-clustering, and Q must be a single edge (and its type t equals t^*). Using (2.32) and (2.33) in (2.29) gives $M = O\left(\lambda_{\tau(L)}n^{o(1)}/\lambda_t\right)$ in this case, as required. We note that in fact the bound can be strengthened to $O\left(\lambda_{\tau(L)}/\lambda_t\right)$ unless $Q_\ell = L$, $\ell = 1$ and $j = 1$, and looking back at the above argument, we may use $f_j(u_j)/\lambda_{u_j}$ in place of $n^{o(1)}$, as noted after the proposition's statement.

We turn now to proving the bounds

$$|\gamma(f, t) - 1| \leq C'\phi_t x$$

for all $t \in \mathcal{S}$, and here we may assume by induction that (2.28) holds with f replaced by any $g < f$, and that, as we have just shown, (2.26) holds. We also know that $c(t, t, \mathbf{0}) = 1 + O(p)$ from (2.3). So it suffices to show that Σ in the statement of the Proposition 2.1 is $O(\phi_t x)$. Since $\mathcal{S}$ is fixed, there is a bounded number of terms in the sum, and each may be written as

$$\gamma(f - h, u)\,\frac{\lambda_u}{\lambda_t}\,c(u, t, h)\rho(f, -h). \tag{2.34}$$

Note that the argument that produced (2.32) gives, in this case, $\rho(f, -h) = O(n^{o(1)})$. So (again by appropriate choice of ϕ^*) we only need to show that the product of the remaining factors in (2.34) is $O(xn^{o(1)})$.

Let $\mathcal{F}_1$ denote the set of $h \in \mathcal{F}_{\mathcal{S}}$ for which there are $t, u \in \mathcal{S}$ such that $c(u, t, h) \neq 0$. Note that the cardinality of $\mathcal{F}_1$ is bounded.

Inside the present main inductive step, we use a second level of induction on t, going from greatest to smallest in the relation '$\prec$'. Assume first that t is maximal. Since $u \in \mathcal{S}$, it is necessary that $u = t$ and $h \neq \mathbf{0}$ for such a term to be included in Σ. Then $\gamma(f - h, t) \leq 3/2$ by (2.28) inductively. Furthermore, since the graphs in $\mathcal{R}$ are nonempty and $H \neq \emptyset$ in (2.2), we have $c(t, t, h) = O(p) = O(x)$, which gives the desired result.

Suppose next that t is not maximal. A term (2.34) with $u = t$ and $h \neq \mathbf{0}$ is $O(xn^{o(1)})$ for reasons as in the previous paragraph. On the other hand, for $u \neq t$ and $h \in \mathcal{F}_1$, clearly $c(u, t, h) = O(1)$. If $c(u, t, h) \neq 0$, then by the definition (2.1), $t \prec u$, and then $\gamma(f - h, u) \leq 3/2$ by (2.28) inductively, and $\lambda_u/\lambda_t = O(x)$ by (2.12). Once again, (2.34) is $O(xn^{o(1)})$. For appropriate choice of ϕ^* and C', we now have $|\gamma(f, t) - 1| \leq C'\phi_t x$. Thus, in view of the bound (2.14) on x, for appropriate choice of N_0, we have (2.27) in full. This completes the inductive step, and (2.26) and (2.27) imply the lemma. ∎

It is useful to rewrite Proposition 2.1 in terms of the γ's. It says that for $f \in \mathcal{F}$ and $t \in \mathcal{S}$,

$$\gamma(f,t) = \frac{1}{c(t,t,\mathbf{0})}\left(1 - \Sigma - \frac{\theta(f,\delta_t)}{|t|\mathbb{P}(\mathcal{C}_f)}\right), \tag{2.35}$$

where Σ is defined by (2.6). Writing

$$\frac{\mathbb{P}(\mathcal{C}_{f-h+\delta_u})}{\mathbb{P}(\mathcal{C}_f)} = \frac{\mathbb{P}(\mathcal{C}_{f-h})}{\mathbb{P}(\mathcal{C}_f)} \cdot \frac{\mathbb{P}(\mathcal{C}_{f-h+\delta_u})}{\mathbb{P}(\mathcal{C}_{f-h})}$$

and using (2.31) for the first factor gives

$$\Sigma = \sum_{\substack{u\in\mathcal{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} \frac{\lambda_u}{\lambda_t} c(u,t,h)\gamma(f-h,u) \prod_{i=1}^{k} \frac{f_i(t_i)+1}{\lambda_{t_i}\gamma(f_i,t_i)}, \tag{2.36}$$

which is a function of f and t, where, for each h, t_i, $i = 1,\ldots,k$ is a sequence in $\mathcal{S}$ such that $h = \sum_{i=1}^{k} \delta_{t_i}$ and $f_i = f - \sum_{j=1}^{i} \delta_{t_j}$. Here and henceforth, we may choose a canonical sequence $t_1,\ldots,t_k$ for each h such that $c(u,t,h) \neq 0$ for some $u,t \in \mathcal{S}$. Note that k is bounded because $\mathcal{S}$ is finite.

Approximations to the γ's may be defined recursively by ignoring the term containing $\theta(f,\delta_t)$ in (2.35). Thus, we define:

$$\hat{\gamma}(f,t) = \frac{1}{c(t,t,\mathbf{0})}\left(1 - \hat{\Sigma}\right)$$

where

$$\hat{\Sigma} = \sum_{\substack{u\in\mathcal{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} \frac{\lambda_u}{\lambda_t} c(u,t,h)\hat{\gamma}(f-h,u) \prod_{i=1}^{k} \frac{f_i(t_i)+1}{\lambda_{t_i}\hat{\gamma}(f_i,t_i)} \tag{2.37}$$

is a function of f and t.

Proposition 2.3 *Uniformly for all $f \in \mathcal{F}_{\mathcal{S}}$ and $t \in \mathcal{S}$*

$$|\hat{\gamma}(f,t) - \gamma(f,t)| = O\left(\frac{x + \phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right)$$

where $\phi_t = n^{o(1)}$ for $t = t^*$ and $\phi_t = 1$ otherwise.

Proof. We use an inductive scheme as we did for Proposition 2.2. The initial step of the outer induction is $f = \mathbf{0}$, and the initial step of the inner induction has t maximal in $\mathcal{S}$. The initial steps are considered below.

We aim to show inductively that

$$\gamma(f,t) = \hat{\gamma}(f,t) + O_t\left(\frac{x + \phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{2.38}$$

where $O_t()$ denotes $O()$ with the implicit constant depending on t. (Although this implies the same statement for a uniformly defined implicit constant, the induction argument requires different constants for each t, larger constants for "smaller" t. Constraints on the sizes of these constants are implicitly determined in the proof below.) By (2.3), the definition (2.10) of x, and Proposition 2.2, it suffices to show

$$\Sigma = \hat{\Sigma} + O_t\left(\frac{x + \phi_t \lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{2.39}$$

Instead of proceeding step by step through the induction, the argument is made by focussing on the relevant considerations for an arbitrary step, whether it be an initial step (for f or for t) or an arbitrary inductive step.

First, notice that if some $t_i = u$ in (2.37) when $c(u,t,h) \neq 0$, then it must be that $k = 1$, $h = \delta_u$, $f_1 = f - h$ and the $\hat{\gamma}$'s cancel. This means that the corresponding terms in Σ and $\hat{\Sigma}$ are equal, so henceforth whenever $k \geq 1$, we may assume that $t_i \prec u$ for all i.

If $h = \mathbf{0}$ in a term in Σ, or $\hat{\Sigma}$, then the value of k in that term is 0, and the product in that term is empty, and equal to 1. On the other hand, suppose that $h \neq \mathbf{0}$. When evaluating the factor $\hat{\gamma}(f_i, t_i)$ in (2.37), the definition of $\hat{\Sigma}$ invokes (2.37) a second time (recursively); for the second level of invocation we will use j in place of i, as the index of the product. As shown above, we may assume that each $t_j \prec u$. Since $u \in \mathcal{S}$, we have $\lambda_u \geq n^{-o(1)}$. Thus, for all j in (2.37), $\lambda_{t_j} = \Omega(\lambda_u/x) > n^{\epsilon - o(1)}$ by (2.14). The ratios $(f_i(t_i)+1)/\lambda_{t_i}$ in (2.36) and (2.37) are therefore $O(1)$ by (2.23). We have from Proposition 2.2 that $\gamma(f,t) \sim 1$ uniformly, and it is also immediate that $c(u,t,h) = O(1)$, and $1/c(t,t,\mathbf{0}) = O(1)$ by (2.3). The combination of these facts shows that each $\hat{\gamma}(f_i, t_i)$ in (2.37) is $1 + o(1)$, with the convergence uniform over all f_i and t_i. This implies in particular that the product in (2.37) is in all cases $O(1)$.

We will estimate the difference between the summands in (2.36) and (2.37) using

$$(A + \delta_A)(B + \delta_B) - AB = O(|\delta_A B| + |A \delta_B|), \tag{2.40}$$

which holds provided that $\delta_A = O(A)$ or $\delta_B = O(B)$. We will show that for (u,h) as in the scope of the summation in (2.36),

$$|\hat{\gamma}(f-h,u) - \gamma(f-h,u)| \frac{\lambda_u}{\lambda_t} c(u,t,h) = \begin{cases} O_u\left(\dfrac{x + \lambda_{\mathcal{L}}}{\lambda_t}\right) & \text{if } t \prec u \\[2ex] O_t\left(x \dfrac{x + \phi_t \lambda_{\mathcal{L}}}{\lambda_t}\right) & \text{if } u = t, \end{cases} \tag{2.41}$$

and, for factors appearing in the product in (2.37) with $t_i \prec u$,

$$|\hat{\gamma}(f_i,t_i) - \gamma(f_i,t_i)| \frac{\lambda_u}{\lambda_t} = O_{t_i}\left(x \cdot \frac{x + \phi_{t_i} \lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{2.42}$$

In view of the above observations, these imply

$$\Sigma = \hat{\Sigma} + \sum_{u \in \mathcal{S} \colon t \prec u} \frac{x + \phi_t \lambda_{\mathcal{L}}}{\lambda_t} O_u(1) + \sum_{v \in \mathcal{S}} \frac{x + \phi_{t^*} \lambda_{\mathcal{L}}}{\lambda_t} O_v(x).$$

Equation (2.39) will then follow, since the summations contain a bounded number of terms, and in the first summation the constant implicit in $O_u()$ may be used in defining the constant implicit in $O_t()$, whilst in the second summation the bound is $o\big((x+\phi_t\lambda_{\mathcal{L}})/\lambda_t\big)$ by induction using $x\phi_{t^*}=o(1)$. Note that for the initial step of the inner induction, when t is maximal in $\mathcal{S}$, it must be that $u=t$.

For each term in (2.36) and (2.37) we have $(u,h)\neq(t,\mathbf{0})$, so the inductive statement (2.38) implies

$$|\gamma(f-h,u)-\hat{\gamma}(f-h,u)|\,\frac{\lambda_u}{\lambda_t}=O_u\left(\frac{x+\phi_u\lambda_{\mathcal{L}}}{\lambda_t}\right).$$

Note that $t\prec u$ implies $u\neq t^*$ and hence $\phi_u=1$. Recalling $c(u,t,h)=O(1)$, and noting that in particular $c(t,t,h)=O(x)$ when $t=u$ (as $h\neq\mathbf{0}$ in that case), we have (2.41). By the outer induction (which is on f) using (2.38), the left side of (2.42) is of order

$$O_{t_i}\left(\frac{x+\phi_{t_i}\lambda_{\mathcal{L}}}{\lambda_{t_i}}\,\frac{\lambda_u}{\lambda_t}\right)=O_{t_i}\left(\frac{x+\phi_{t_i}\lambda_{\mathcal{L}}}{\lambda_t}\,\frac{\lambda_u}{\lambda_{t_i}}\right) \tag{2.43}$$

and by (2.12) and (2.14) (noting that $t_i\prec u$ as discussed above), $\lambda_u/\lambda_{t_i}=O(x)$, which completes the proof. ∎

A recursive calculation of $\hat{\gamma}$ using its definition, including (2.37), would need to keep track of $\hat{\gamma}(f,t)$ for each $f\in\mathcal{F}_{\mathcal{S}}$ and $t\in\mathcal{S}$. By making further approximations, we may obtain a simpler recursion for functions which are explicitly defined in a compact form, and not depending on f. Recalling that $|\mathcal{S}|=s$, without loss of generality we denote $\mathcal{S}$ by $[s]=\{1,\dots,s\}$. (Thus $t\in\mathcal{S}$ is represented by an integer. We apologise to the reader for the possible confusion resulting; in particular the definition (2.4) of the function $|t|$, where t is a type, overrides the notation for absolute value of the integer. It only appears once or twice more.) The simpler recursion will define $\overline{\gamma}_t\in\mathbb{R}[[n,p,g_1,\dots,g_s]]$, i.e. a formal power series in n, p and $g_1,\dots,g_s$ with real coefficients. Occasionally it will be useful to regard $\overline{\gamma}_t$ also as an element of $\mathbb{R}[[n,p]][[\mathbf{g}]]$ where $\mathbf{g}=(g_1,\dots,g_s)$, meaning a formal power series with indeterminates $g_1,\dots,g_s$ and coefficients in $\mathbb{R}[[n,p]]$. Later, we will calculate the new estimates of $\gamma(f,t)$ by setting $g_i=f(i)/\lambda_i$ in $\overline{\gamma}_t$ for each i.

Note that $c(u,t,h)$ is a polynomial in p, and $1/c(t,t,\mathbf{0})=1+O(p)$ and can be expanded as power series in p. Also, by (2.5), for $t\prec u$, λ_u/λ_t is a polynomial in n and p with terms of the form $p^{\mu(u)-\mu(t)}n^i$, and, since $\mu(u)>\mu(t)$, λ_u/λ_t has zero constant term. With these interpretations, we will define $\overline{\gamma}_t=\overline{\gamma}_t(n,p,\mathbf{g})\in\mathbb{R}[[n,p,g_1,\dots,g_s]]$ using

$$\overline{\gamma}_t=\frac{1}{c(t,t,\mathbf{0})}\left(1-\sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}}\frac{\lambda_u}{\lambda_t}c(u,t,h)\overline{\gamma}_u\prod_{i=1}^{k}\frac{g_{t_i}}{\overline{\gamma}_{t_i}}\right),\quad \overline{\gamma}_t(0,0,\mathbf{0})=1 \tag{2.44}$$

simultaneously for all $t\in\mathcal{S}$, where the t_i are defined as in (2.36). Since $c(t,t,h)=O(p)$ for $h\neq\mathbf{0}$ and $(\lambda_u/\lambda_t)c(u,t,h)$ has zero constant term for $u\neq t$, there is a unique set of formal power series $\overline{\gamma}_t(n,p,\mathbf{g})$, $t\in\mathcal{S}$, defined by (2.44), and they all have constant term 1. It will also be useful to rewrite (2.44) as

$$\overline{\gamma}_t = 1 + w_0(t) - \sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} w(u,t,h)\overline{\gamma}_u \prod_{i=1}^{k} \frac{1}{\overline{\gamma}_{t_i}}, \tag{2.45}$$

$$w_0(t) = \frac{1}{c(t,t,\mathbf{0})} - 1, \quad w(u,t,h) = \frac{\lambda_u c(u,t,h)}{\lambda_t c(t,t,\mathbf{0})} \prod_{i=1}^{k} g_{t_i}. \tag{2.46}$$

Here (2.45) defines $\overline{\gamma}_t$ as a power series in the w's, which, if substituted appropriately as power series in n, p and $\mathbf{g}$ using (2.46), results in the same series as defined in (2.44).

Given a function $f \in \mathcal{F}$, with a slight abuse of notation, define

$$\overline{\gamma}_t(f) = \overline{\gamma}_t(n, p, \tilde{\mathbf{g}}) \tag{2.47}$$

where

$$\tilde{\mathbf{g}} = (f(1)/\lambda_1, \dots, f(s)/\lambda_s).$$

Thus, given n and p, $\overline{\gamma}_t(\cdot)$ maps functions $f \in \mathcal{F}$ to numbers, whereas $\overline{\gamma}_t$ is a power series.

Returning to our original setting, $f \in \mathcal{F}_{\mathcal{S}}$ (as defined at (2.23)), and p is a function of n such that $x = x(n,p) = O(n^{-\epsilon})$ by (2.14). It might help to observe at this point that, for given n, p and f satisfying these constraints, there is a unique value of $\overline{\gamma}_t(f)$ determined from the equations (2.44) and (2.47), as long as n is large enough. One way to prove this is to consider an initial approximation for each $\overline{\gamma}_t(f)$, and then, iterating the approximations using (2.44), with g_t set equal to $f(t)/\lambda_t$, the current values of $\overline{\gamma}_t$ on the right side giving rise to updated values on the left side. This determines a contractive mapping on the vector whose entries are $\overline{\gamma}_t(f)$ ($t \in \mathcal{S}$) which has a fixed point near the initial approximate solution determined by $\overline{\gamma}_t(f) = 1$ for all t. To flesh this out, we first examine the definition of $\overline{\gamma}_t$ in order to bound the error of approximations. Recalling (2.13) and (2.14), we have the following lemma.

First, given particular values of n, p and f, we define

$$\tilde{g}_t = f(t)/\lambda_t,$$

so that $\tilde{\mathbf{g}} = (\tilde{g}_1, \dots, \tilde{g}_s)$, and let $\tilde{w}(u,t,h)$ denote the value of $w(u,t,h)$ obtained if we replace g_{t_i} by $\tilde{g}_{t_i}$ in (2.46). For convenience, similarly set $\tilde{w}_0(t) = w_0(t)$. Recall that p has been assigned a function of n satisfying (2.13), which is significant when considering issues of uniformity.

Lemma 2.4 *Suppose that* $0 \le \tilde{g}_t = \tilde{g}_t(n) = O(n^{o(1)})$, *with* $\tilde{g}_t(n) = O(1)$ *if* $t \in \mathcal{S}_1$. *Then* $\tilde{w}_0(t) = O(p)$ *and* $\tilde{w}(u,t,h) = O(x)$ *for each term in (2.45), where the bounds in the* $O()$ *terms are uniform.*

Proof. From (2.3), $\tilde{w}_0(t) = O(p)$ and, recalling that k is bounded in (2.46) and that $c(t,t,\mathbf{0}) \sim 1$,

$$\tilde{w}(u,t,h) = O\left(\frac{\lambda_u c(u,t,h)}{\lambda_t} (\max_i \tilde{g}_{t_i})^k\right). \tag{2.48}$$

Firstly, if $h = \mathbf{0}$, then $k = 0$, and $u \succ t$ by the condition in the summation. So $\tilde{w}(u,t,h) = O(x)$ by (2.12).

Secondly, suppose that $h \neq \mathbf{0}$ and $u = t$. If $h = \delta_{t^*}$ (recall that t^* is the type of the single-edge cluster), then $c(u,t,h) = O(p)$ in view of (2.2). By (2.20), we have $t^* \in \mathcal{S}_1$. So, using the hypothesis of this lemma, the maximum in (2.48) is $O(1)$, and thus $\tilde{w}(u,t,h) = O(p) = O(x)$. In all other cases, if $c(u,t,h) \neq 0$ then (2.2) gives $c(t,t,h) = O(p^2)$ since $s_H = h$ implies $|H| \geq 2$. By (2.48), again $\tilde{w}(u,t,h) = O(x)$.

Lastly, suppose that $h \neq \mathbf{0}$ and $u \succ t$. Here $\lambda_u/\lambda_t = O(x)$ by (2.12), and so we are done if the maximum in (2.48) is $O(1)$. But this must happen unless $t_i \in \mathcal{S}_0$ for some i. Since H contains only subclusters of a cluster of type $u \in \mathcal{S}$, (2.12) shows that this requires $t_i = u$. Then we have $h = \delta_u$, and hence in (2.1), $Q \subseteq J$ and $|Q \cap H| \geq 1$, and so $c(u,t,h) = O(p) = O(x)$. Since the maximum in (2.48) is $O(n^{o(1)})$, the bound obtained is $O(x^2 n^{o(1)})$, and the result follows in this case also. ∎

Recall that $\overline{\gamma}_t(f)$ is a function of n, p and f.

Lemma 2.5 *For $f \in \mathcal{F}_\mathcal{S}$ and p satisfying (2.13), the series definition of $\overline{\gamma}_t(f)$ in (2.47) converges absolutely for n sufficiently large, and $\overline{\gamma}_t(f) = 1 + O(x)$, where the bound in the $O()$ notation is uniform.*

Proof. For any $t \in \mathcal{S}_0$, it follows from the definition of $\tilde{g}_t$, the upper bounds (2.21) and (2.23) on $f(t)$, and the asymptotics (2.13) of p, that $\tilde{g}_t = O(n^{o(1)})$. On the other hand, if $t \in \mathcal{S}_1$ then $\tilde{g}_t \in [0,3]$ for similar reasons. Thus the conditions of Lemma 2.4 are satisfied.

For polynomials or formal power series P and $\hat{P}$, denote by P^+ the formal power series obtained by replacing all coefficients of P by their absolute values, and write $P \leq \hat{P}$ if the coefficient of any monomial in P is no greater than the corresponding coefficient in $\hat{P}$. We will use the obvious fact that if P^+ is absolutely convergent (for a particular assignment of the indeterminates) then so is P.

With (2.45) in mind, and with the aim of obtaining the useful inequality (2.50) below, define the power series γ_t^* for each $t \in \mathcal{S}$ by

$$\gamma_t^* = 1 + w_0^+ + \sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} w(u,t,h)^+ \gamma_u^* \prod_{i=1}^{k} \frac{1}{2-\gamma_{t_i}^*}, \tag{2.49}$$

which by induction has a unique solution in formal power series with constant terms all 1. Then

$$\frac{1}{2-\gamma_{t_i}^*} = \sum_{j\geq 0} (\gamma_{t_i}^* - 1)^j$$

and so by induction, all coefficients of γ_t^* are nonnegative for each $t \in \mathcal{S}$. Thus

$$\frac{1}{2-\gamma_{t_i}^*} \geq \sum_{j\geq 0} (1-\gamma_{t_i}^*)^j = \frac{1}{\gamma_{t_i}^*}$$

and, again by induction, comparing (2.45) with (2.49) gives

$$\overline{\gamma}_t^+ \leq \gamma_t^* \tag{2.50}$$

for each $t \in \mathcal{S}$.

Now consider summing the terms of $\gamma_t^*(n, p, \tilde{\mathbf{g}})$ for p and f as in the lemma, when n is sufficiently large. Since all coefficients of γ_t^* are nonnegative, we are at liberty to sum the terms in any convenient order. It is immediate from the proof of Lemma 2.4 that $w(u,t,h)^+ = O(x)$ and $w_0^+ = O(p) = O(x)$. It is now straightforward to verify from (2.49), by a sequence of successive approximations beginning with $\gamma^* \approx 1$ for all t, that

$$\gamma_t^*(n, p, \tilde{\mathbf{g}}) = 1 + O(x). \tag{2.51}$$

The lemma now follows since from (2.50), and the fact that the constant terms in all $\overline{\gamma}$'s and γ^*'s are all 1, $(\overline{\gamma}_t - 1)^+ \le \gamma_t^* - 1$. ∎

If p and f satisfy the conditions of Lemma 2.5, we may treat $\overline{\gamma}_t(f)$ as a number, being the sum of the series, for n sufficiently large. Since we may ignore small values of n, and since p is a function of n, this makes $\overline{\gamma}_t(f)$ a real-valued function of f and n, and henceforth in this section we treat it as such.

Proposition 2.6 *Uniformly for all $f \in \mathcal{F}_\mathcal{S}$ and $t \in \mathcal{S}$,*

$$|\overline{\gamma}_t(f) - \gamma(f,t)| = O\left(\frac{x + \phi_t \lambda_\mathcal{L}}{\lambda_t}\right)$$

where $\phi_t = n^{o(1)}$ for $t = t^*$ and $\phi_t = 1$ otherwise.

Proof. An induction like the one proving Proposition 2.3 is used. The inductive hypothesis is

$$|\overline{\gamma}_t(f) - \hat{\gamma}(f,t)| = O_t\left(\frac{x + \phi_t \lambda_\mathcal{L}}{\lambda_t}\right),$$

where O_t denotes a bound depending only on t. The result then follows by Proposition 2.3.

Suppose that $f = \mathbf{0}$. Then $h = \mathbf{0}$ in (2.37) and the terms in (2.44) with $h \ne \mathbf{0}$ are 0 because $\tilde{g}_{t_i} = 0$ for all i by (2.47). Hence, the products in (2.37) and (2.44) are empty, and by simple (downwards) induction on t, $\hat{\gamma}(f,t) = \overline{\gamma}_t(f)$ for all $t \in \mathcal{S}$.

It remains to prove the lemma when $f \ne \mathbf{0}$, which we assume henceforth.

Note that (2.44) contains terms such that, for some values of f, the corresponding terms are excluded (2.37) because $f - h \notin \mathcal{F}$. For the inductive step, we bound these terms first. After this, we consider the error caused by replacing $\hat{\gamma}(f-h,u)$ by $\overline{\gamma}_u(f)$ in (2.37), as well as $\hat{\gamma}(f_i,t_i)$ by $\overline{\gamma}_{t_i}(f)$, and $f_i(t_i)+1$ by $f_i(t_i)$.

Since $\overline{\gamma}_t(f) = 1 + O(x)$ by Lemma 2.5, and $\tilde{w}_0 = O(p)$ and $\tilde{w}(u,t,h) = O(x)$ from Lemma 2.4, all terms in the summation in (2.45) are $O(x)$. If $f - h \notin \mathcal{F}$ in (2.45), so that $f(t_{i'}) - h(t_{i'}) < 0$ for some $t_{i'}$, then $f(t_{i'}) = O(1)$ and so $g_{t_{i'}} = O(1/\lambda_{t_{i'}})$. The contribution of such a term in (2.44) is $O\left(\lambda_u/\lambda_t\lambda_{t_{i'}}\right)$, which in the case $t_{i'} \prec u$ is $O(x/\lambda_t)$. On the other hand, if $t_{i'} = u$, we have the same situation as in the second paragraph after (2.39), so the γ's cancel, $c(u,t,h) = O(x)$, and again the term is $O(x/\lambda_t)$.

For those h satisfying $f - h \in \mathcal{F}$, first recall, as observed in the middle of the proof of Proposition 2.3, the product in (2.37), which we will denote by Π, is $O(1)$. Analogous to (2.41) and (2.42) in the proof of Proposition 2.3, we will show that, for the same values of (u, h) as in that Proposition,

$$|\hat{\gamma}(f-h,u)-\overline{\gamma}_u(f)|\,\frac{\lambda_u}{\lambda_t}c(u,t,h)\Pi = \begin{cases} O_u\left(\frac{x+\lambda_{\mathcal{L}}}{\lambda_t}\right)+O\left(\frac{x}{\lambda_t}\right) & \text{if } t\prec u \\ O_t\left(x\frac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right)+O\left(\frac{x}{\lambda_t}\right) & \text{if } u=t, \end{cases} \tag{2.52}$$

and

$$\left|\hat{\gamma}(f_i,t_i)-\overline{\gamma}_{t_i}(f)\right|\frac{\lambda_u}{\lambda_t}c(u,t,h)\Pi = O_{t_i}\left(x\frac{x+\phi_{t_i}\lambda_{\mathcal{L}}}{\lambda_t}\right)+O\left(\frac{x}{\lambda_t}\right), \quad t_i\prec u \tag{2.53}$$

and, for the replacement of $f_i(t_i)+1$ by $f_i(t_i)$ when evaluating g_{t_i},

$$\frac{\lambda_u c(u,t,h)}{\lambda_t\lambda_{t_i}} = O\left(\frac{x}{\lambda_t}\right). \tag{2.54}$$

The lemma follows from these claims, using (2.40) along the lines of the proof of Proposition 2.3, combined with the observation that, by the inductive hypothesis combinded with Lemma 2.5, we may assume that $\hat{\gamma}(f-h,u)=\Theta(1)$ uniformly whenever $h>\mathbf{0}$, or $h=\mathbf{0}$ and $t\prec u$.

The treatment of the $O_t()$ terms in this proof is rather delicate and is explained in detail in the proof of Proposition 2.3. In this case, there are extra terms $O(x/\lambda_t)$ in (2.52–2.54), which we write separately to make the recursive argument clearer. Note that the $O_u()$ and $O_t()$ terms contain the same implicit constants as in the inductive hypothesis.

It is convenient to treat (2.54) first. If $t_i\prec u$, then we are done by (2.12) applied with t replaced by t_i, and the fact that $c(u,t,h)=O(1)$. On the other hand, if $t_i=u$ then $k=1$ and $h=\delta_u$, and, as in the last part of the proof of Lemma 2.4, $c(u,t,h)=O(x)$, as required.

Now consider (2.52). Since either $f-h<f$ or $t\prec u$, the inductive hypothesis may be applied, with Π referring to $f-h$ rather than f, yielding

$$\begin{aligned} |\hat{\gamma}(f-h,u)-\overline{\gamma}_u(f-h)|\frac{\lambda_u}{\lambda_t}c(u,t,h)\Pi &= O_u(1)\frac{x+\phi_u\lambda_{\mathcal{L}}}{\lambda_u}\frac{\lambda_u}{\lambda_t}c(u,t,h)\Pi \\ &= O_u\left(\frac{x+\phi_u\lambda_{\mathcal{L}}}{\lambda_t}c(u,t,h)\right). \end{aligned} \tag{2.55}$$

Recalling also from the proof of Lemma 2.4 that $c(t,t,h)=O(x)$ (and $c(u,t,h)=O(1)$ always), and that $\phi_u=1$ when $t\prec u$, now shows that this expression is bounded by $O_t\big(x(x+\phi_t\lambda_{\mathcal{L}})/\lambda_t\big)$ (respectively $O_u\big((x+\lambda_{\mathcal{L}})/\lambda_t\big)$) as required for the cases $u=t$ and $t\prec u$ in the right hand side of (2.52). Next we bound

$$|\overline{\gamma}_u(f)-\overline{\gamma}_u(f-h_0)| \tag{2.56}$$

for any fixed h_0 with bounded entries. We can assume $h_0\neq\mathbf{0}$. By Lemma 2.5, equation (2.45) can be expanded in increasing powers of the w's, which are $O(x)$ under the substitution $g_v=f(v)/\lambda_v$ by Lemma 2.4. By (2.14), we may ignore terms whose total degree in w's is larger than some fixed value. Into the truncated expression, substitute $f(t_i)/\lambda_{t_i}$ and $(f(t_i) - h(t_i))/\lambda_{t_i}$ for g_{t_i} in the definition of $w(u,t,h)$ at (2.46) and subtract the two resulting expressions term by term. Since the entries of h_0 are bounded, the dominating terms are exactly of the type estimated in (2.54), and hence are bounded by $O(x/\lambda_t)$. Equation (2.52) now follows (with room to spare) in view of the fact that, by Lemma 2.4,

$$\frac{\lambda_u}{\lambda_t} c(u,t,h)\Pi = O(x).$$

The proof of (2.53) involves firstly consideration of $\left|\hat{\gamma}(f_i,t_i) - \overline{\gamma}_{t_i}(f_i)\right|$ (multiplied by the other factors). This yields an expression as in the right hand side of (2.55), but with O_u replaced by O_{t_i}, $\lambda_{\mathcal{L}}/\lambda_u$ replaced with $\lambda_{\mathcal{L}}/\lambda_{t_i}$ and $f-h$ becoming f_i. The error term is bounded similarly to the bound (2.43) for the analogous term in the proof of Proposition 2.3, and also using $\lambda_u/\lambda_{t_i} = O(x)$ (as $t_i \prec u$), giving the first error term in (2.53). Then, $\left|\overline{\gamma}_{t_i}(f_i) - \overline{\gamma}_{t_i}(f)\right|$ is bounded by the expression in (2.54), by the same argument as for (2.56). ∎

From Lemmas 2.4 and 2.5, we may use (2.45) to expand all the functions $\overline{\gamma}_t$ ($t \in \mathcal{S}$) recursively in power series in n, p and the variables g_i. Iterating r times determines $\overline{\gamma}_t$ to arbitrarily small error $O(x^r)$ when the appropriate values are assigned to p and the g_i. However, instead of pursuing arbitrary accuracy in this paper, we desire a final formula which is shown to exhibit a uniformity over all relevant κ, and for this we need the following. We use $\mathbf{g}^{\mathbf{i}}$ to denote $g_1^{i_1} g_2^{i_2} \cdots g_s^{i_s}$; if $\mathbf{i} = \mathbf{0}$, this is the multiplicative identity of the ring $\mathbb{R}[[n,p]][[\mathbf{g}]]$ of formal power series over $\mathbf{g}$ whose coefficients are in $\mathbb{R}[[n,p]]$.

Corollary 2.7 *There are power series ξ_t, $t \in \mathcal{T}$, in n, p, $\mathbf{g}$, independent of κ, and, for all $\epsilon > 0$, truncations $\xi_{t,\epsilon}$ of the series ξ_t, to a finite number of terms, such that for all $t \in \mathcal{S}$*

- *(a) For $\mathbf{i} \neq \mathbf{0}$, we have $[\mathbf{g}^{\mathbf{i}}]\xi_t = O(x)$, for p satisfying (2.13) with $\kappa \geq \chi + \epsilon$, as $n \to \infty$;*
- *(b) for each $\mathbf{i}$, the coefficient $[\mathbf{g}^{\mathbf{i}}]\xi_t$ is a multiple of $\prod_{u \in \mathcal{S}} p^{\mu(u) i_u}$;*
- *(c) With p satisfying (2.13), and $\xi_{t,\epsilon}(f)$ defined from $\xi_{t,\epsilon}$ analogously to $\overline{\gamma}_t(f)$ in (2.47), there exists $\overline{\epsilon} > 0$ such that uniformly for all $f \in \mathcal{F}_{\mathcal{S}}$, and all $\kappa \geq \chi + \epsilon$,*

$$\xi_{t,\epsilon}(f) = \overline{\gamma}_t(f) + O\left(\frac{n^{-\overline{\epsilon}}}{\lambda_t}\right). \tag{2.57}$$

Proof. Instead of (c) we show the obviously stronger

$$\xi_{t,\epsilon}(f) = \overline{\gamma}_t(f) + O\left(\frac{x + n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{2.58}$$

We start by essentially focusing on this, but with one eye fixed on (a). Define the function $F_t = F_t(n,p,\mathbf{g},\overline{\gamma}_1,\ldots,\overline{\gamma}_s)$ by

$$F_t(n,p,\mathbf{g},\overline{\gamma}_1,\ldots,\overline{\gamma}_s) = \frac{1}{c(t,t,\mathbf{0})}\left(1 - \sum_{\substack{u\in\mathcal{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} \frac{\lambda_u}{\lambda_t} c(u,t,h)\overline{\gamma}_u \prod_{i=1}^{k} \frac{g_{t_i}}{\overline{\gamma}_{t_i}}\right) - 1. \tag{2.59}$$

We obtain successive power series approximations $F_t^{(j)}$ and $\overline{\gamma}_t^{(j)}$ for all the F_t and $\overline{\gamma}_t$ ($j = 0, 1, \ldots$). Initially, set $F_t^{(0)} = 0$ and $\overline{\gamma}_t^{(0)} = 1$ for all t. For $j \geq 0$, substituting $\overline{\gamma}_t^{(j)}$ for $\overline{\gamma}_t$ in (2.59) simultaneously for all $t \in \mathcal{S}$ defines $F_t^{(j+1)}$ as a power series (recalling the observations made before (2.44) that λ_u/λ_t is a polynomial in n and p, and so on). Next, define $\overline{\gamma}_t^{(j+1)} = 1 + F_t^{(j+1)}$ to complete the iterative definition. Define $\overline{\gamma}_t^{(i)}(f)$ from $\overline{\gamma}_t^{(i)}$ analogously to $\overline{\gamma}_t(f)$ in (2.47), and similarly $F_t^{(i)}(f)$. By Lemma 2.5, $\overline{\gamma}_t^{(0)}(f) = \overline{\gamma}_t(f)(1+O(x))$ for all relevant f and p. Thus

$$F_t^{(1)}(f) = F_t(n, p, \tilde{\mathbf{g}}, \overline{\gamma}_1, \ldots, \overline{\gamma}_s)(1 + O(x)).$$

By Lemma 2.5, this is $O(x)$, and so by (2.44), $\overline{\gamma}_t^{(1)}(f) = \overline{\gamma}_t(f) + O(x^2)$. Repeating the same argument r times shows that

$$\overline{\gamma}_t^{(r)}(f) = \overline{\gamma}_t(f) + O(x^{r+1}). \tag{2.60}$$

As with Lemma 2.5, the argument to this point is for fixed $\kappa > \chi$. The definition of $\mathcal{S}$ by (2.16), and hence the formula (2.44), depends on κ. However, for all $\kappa \geq \chi + \epsilon$, $\mathcal{S}$ is a subset of $\hat{\mathcal{S}} = \{t \in \mathcal{T} : \nu(t)/\mu(t) \geq \chi + \epsilon\}$, which is the value of $\mathcal{S}$ when $\kappa = \kappa_0 = \chi + \epsilon$. So define r_t to be such that $x^{r_t} = O(1/\lambda_t)$ when $\kappa = \kappa_0$. Then set $\xi_{t,\epsilon}$ equal to the truncation of $\overline{\gamma}_t^{(r_t)}$ to those terms whose value, with $\mathbf{g}$ set equal to 1, is not $o(x/\lambda_t)$ (when $\kappa = \kappa_0$). By (2.60), (2.58) holds for $\kappa = \kappa_0$.

Also note for later use that, in view of (2.60), using $\overline{\gamma}_t^{(r)}$ for any $r > r_t$ would define the same $\xi_{t,\epsilon}$. From (2.50) and (2.51), the coefficients of any non-constant monomial $\mathbf{g}^{\mathbf{i}}$ in $\xi_{t,\epsilon}$, as it arises recursively from (2.59), are $O(x)$, which proves part (a) with ξ_t interpreted as $\xi_{t,\epsilon}$.

We next claim that (2.58) is also valid when $\kappa > \kappa_0$. In this case, the recursive definition of $\overline{\gamma}_t^{(r)}$ is the same as for κ_0 except that the definition of $\mathcal{S}$ is different. Any terms in the summation in (2.44) corresponding to types t that are in $\mathcal{S}$ for κ_0, and not in $\mathcal{S}$ for κ, are now missing. These terms are of the form λ_u/λ_t times a finite product of g_i, for some $u \notin \mathcal{S}$. Since all g_i are substituted with values $n^{o(1)}$, the claim holds.

The remaining portion of the claim in part (c) of the corollary relates to uniformity. This follows from the above observations once we show that these functions $\xi_{t,\epsilon}$ are all common truncations of the power series ξ_t. Now of course (a) is justified in its original form, for ξ_t.

If $\epsilon' < \epsilon$ is considered, then new types enter $\mathcal{S}$, but the terms in $\xi_{t,\epsilon'}$ due to these are of smaller order (as with consideration of $\kappa > \kappa_0$ above) and cannot be included in $\xi_{t,\epsilon}$. Also, the appropriate value of r_t may be larger for ϵ' than for ϵ, but as noted above, truncating with the larger value of r gives the same function $\xi_{t,\epsilon}$, so the extra terms generated cannot include any of the same monomials as appearing in $\xi_{t,\epsilon}$. The power series ξ_t is now well-defined to be the termwise limit of $\xi_{t,\epsilon}$ as $\epsilon \to 0$.

Finally, to verify part (b), note that in the recursive use of (2.59), every new product $\prod_{i=1}^{k} g_{t_i}$ that is introduced is accompanied by the factor $\frac{\lambda_u}{\lambda_t} c(u, t, h)$. By its definition (2.1), each term of $c(u, t, h)$ is associated with a cluster of J of type u, a cluster Q of type t, and pairwise edge-disjoint clusters $J_1, \ldots, J_k$ of types $t_1, \ldots, t_k$, with $c(u, t, h)$ divisible by p^a where $a = |Q \cap (\bigcup J_i)|$. Since λ_u/λ_t is divisible by p^b where $b = \mu(u) - \mu(t) = \mu(u) - |Q|$, the term itself must be divisible by $p^{\sum |J_i|}$, as required for part (b). Of course, the expansions of $1/c(t, t, \mathbf{0})$ and $1/\overline{\gamma}_{t_i}$ do not affect this as their terms have nonnegative exponents. ∎

3 Graphs with forbidden subgraphs in $\mathcal{G}(n,p)$

In this section we prove our main result for subgraphs of the random graph $\mathcal{G}(n,p)$. Let G_0 be a strictly balanced graph and recall that χ is defined by (1.2). Let X be the number of copies of G_0 in $\mathcal{G}(n,p)$.

Proof of the $\mathcal{G}(n,p)$ case of Theorem 1.1

The proof works roughly as follows. We estimate the ratios of 'adjacent' probabilities $\mathbb{P}(\mathcal{C}_f)$ by estimating $\gamma(f,t)$ defined in (2.25). This is approximated by $\overline{\gamma}_t(f)$, as shown in Proposition 2.6, which in turn is approximated by $\xi_{t,\epsilon}$ as found in Corollary 2.7. Fix $\epsilon > 0$. We assume at first that $p = n^{-\kappa+o(1)}$ for fixed $\kappa \geq \chi + \epsilon$, in accordance with (2.13), so that (2.14), Proposition 2.6 and Corollary 2.7 can be applied. The theorem will then be shown in full generality, with assistance from Lemma 1.4. In this section, we work only with the G_0-clustering. As a consequence of this, the parts of the theorems in the previous section relating to t^* are not needed. The set $\mathcal{S}$ is defined, as before, to contain just those types t in this clustering for which $\nu(t)/\mu(t) \geq \kappa$. Recall by the discussion after (2.13) that $\mathcal{S}$ is finite.

The expected number of sets of j disjoint clusters of type $t \in \mathcal{S}$ is, recalling (2.4) and (2.5), at most

$$\binom{|t|}{j} p^{\mu(t)j} \leq \left(\frac{e|t|p^{\mu(t)}}{j}\right)^j = \left(\frac{e\lambda_t}{j}\right)^j.$$

Taking $j = \lfloor m_t \rfloor + 1$ for each $t \in \mathcal{S}$ shows by (2.21) (using $e < 3$) that $\sum_{f \notin \mathcal{F}_\mathcal{S}} \mathbb{P}(\mathcal{C}_f) = o(1)$. (This reveals the relevance of the constant 3 in the definition of m_t.) Furthermore, every large cluster contains an unavoidable cluster, of which there are a finite number. Applying (2.19) to all such clusters, we see that $\sum_{f \in \mathcal{F}} \mathbb{P}(\mathcal{C}_f) \sim 1$. Hence

$$\mathbb{P}(X=0)^{-1} = \frac{1}{\mathbb{P}(\mathcal{C}_\mathbf{0})} \sim \sum_{f \in \mathcal{F}_\mathcal{S}} \frac{\mathbb{P}(\mathcal{C}_f)}{\mathbb{P}(\mathcal{C}_\mathbf{0})}. \tag{3.1}$$

By renaming the cluster types in $\mathcal{S}$ if necessary, extend the poset on $\mathcal{S}$ to a unique linear ordering on $\mathcal{S} = [s] := \{1, 2, \ldots, s\}$ denoted by $<$, in decreasing order of $\nu(t) - \kappa\mu(t)$, breaking ties in a canonical way independent of the choice of κ (i.e. depending only on the graph structure of the types). This is possible in view of (2.4), (2.5), and (2.12). Although the values of p can "wobble" around $p^{-\kappa}$, so that λ_{t+1} and λ_t are not always in the same order when a tie occurred, we do have

$$\lambda_{t+1} < n^{o(1)}\lambda_t \quad \text{for all } t < s. \tag{3.2}$$

(That observation is in fact the main motivation behind the restriction of p in (2.13).)

Fix $(j_1, j_2, \ldots, j_s)$ with $j_u \in [0, m_u]$ for all $u \in \mathcal{S}$ and define f so that $f(t) = j_t$ for each $t \in \mathcal{S}$. Then for each t and j define the function $f_{t,j}$ on $\mathcal{S}$ by $f_{t,j}(t') = j_{t'}$ for $t' < t$; $f_{t,j}(t) = j$; $f_{t,j}(t') = 0$ for $t' > t$. Then $f_{s,j_s} = f$.

By Proposition 2.6, we have $\overline{\gamma}_t(f) = \gamma(f,t) + O((x+\lambda_\mathcal{L})/\lambda_t)$ uniformly for all $f \in \mathcal{F}_\mathcal{S}$ and $t \in \mathcal{S}$. Moreover, by Proposition 2.2, $\gamma(f,t) \sim 1$ uniformly, so that $\overline{\gamma}_t(f) = \gamma(f,t)\big(1+O((x+\lambda_\mathcal{L})/\lambda_t)\big)$. Note that $\big(1+O((x+\lambda_\mathcal{L})/\lambda_t)\big) = \big(1+O(n^{-\overline{\epsilon}}/\lambda_t)\big)$ by (2.14) and (2.19). Using these estimates, then Corollary 2.7, and finally the fact that $\big(1+O(n^{-\bar\epsilon}/\lambda_t)\big)^{m_t} = 1+o(1)$ by the definition of m_t in (2.21), we have

$$\begin{aligned}
\frac{\mathbb{P}(\mathcal{C}_f)}{\mathbb{P}(\mathcal{C}_{\mathbf{0}})} &= \prod_{t=1}^{s}\prod_{j=0}^{j_t-1}\rho(f_{t,j},\delta_t) \\
&= \prod_{t=1}^{s}\frac{\lambda_t^{j_t}}{j_t!}\prod_{j=0}^{j_t-1}\gamma\left(f_{t,j},t\right) \\
&= \prod_{t=1}^{s}\frac{\lambda_t^{j_t}}{j_t!}\prod_{j=0}^{j_t-1}\overline{\gamma}_t\left(f_{t,j}\right)\big(1+O((x+\lambda_{\mathcal{L}})/\lambda_t)\big) \qquad (3.3)\\
&= \prod_{t=1}^{s}\frac{\lambda_t^{j_t}}{j_t!}\prod_{j=0}^{j_t-1}\xi_{t,\epsilon}\left(f_{t,j}\right)\big(1+O(n^{-\bar\epsilon}/\lambda_t)\big) \\
&= \big(1+o(1)\big)\prod_{t=1}^{s}\frac{\lambda_t^{j_t}}{j_t!}\prod_{j=0}^{j_t-1}\xi_{t,\epsilon}\left(f_{t,j}\right). \qquad (3.4)
\end{aligned}$$

Our basic method is to sum the above expression over all f for which $\mathbb{P}(\mathcal{C}_f)$ is significant, thereby obtaining an estimate for the reciprocal of $\mathbb{P}(\mathcal{C}_{\mathbf{0}})$. To facilitate analysis of the summation, we employ various partial sums defined as follows. For $t\in\mathcal{S}\cup\{0\}$, define the functions S_t by $S_s(j_1,j_2,\dots,j_s)=1$, and recursively for t decreasing from $s-1$ to 0, by

$$S_t(j_1,j_2,\dots,j_t) = \sum_{j=0}^{\lfloor m_{t+1}\rfloor} S_{t+1}(j_1,j_2,\dots,j_t,j)\left(\prod_{i=0}^{j-1}\overline{\gamma}_{t+1}\left(f_{t+1,i}\right)\right)\frac{\lambda_{t+1}^{j}}{j!}. \qquad (3.5)$$

We next show (see (3.6)) that this quantity approximates the reciprocal of the conditional probability of having no small clusters of type $u>t$, given j_u clusters of type u for all $u\le t$. Recalling the error bounds involved in (3.3), and then the bound on m_t used in deriving (3.4), we have, uniformly,

$$\begin{aligned}
S_t(j_1,j_2,\dots,j_t) &\sim \sum_{j=0}^{\lfloor m_{t+1}\rfloor} S_{t+1}(j_1,j_2,\dots,j_t,j)\left(\prod_{i=0}^{j-1}\gamma\left(f_{t+1,i},t+1\right)\right)\frac{\lambda_{t+1}^{j}}{j!} \\
&= \sum_{j_{t+1}=0}^{\lfloor m_{t+1}\rfloor} S_{t+1}(j_1,j_2,\dots,j_t,j_{t+1})\frac{\mathbb{P}\left(\mathcal{C}_{f_{t+1,j_{t+1}}}\right)}{\mathbb{P}\left(\mathcal{C}_{f_{t+1,0}}\right)}.
\end{aligned}$$

An inductive argument immediately shows that for all $t<s$ and $j_i\in[0,m_t]$, $i\in[1,t]$,

$$S_t(j_1,j_2,\dots,j_t)\sim\sum_{j_{t+1}=0}^{\lfloor m_{t+1}\rfloor}\cdots\sum_{j_s=0}^{\lfloor m_s\rfloor}\frac{\mathbb{P}(\mathcal{C}_{f_{t+1,0}+j_{t+1}\delta_{t+1}+\cdots+j_s\delta_s})}{\mathbb{P}(\mathcal{C}_{f_{t+1,0}})}. \qquad (3.6)$$

Therefore, by (3.1), noting that S_0 has no arguments,

$$S_0\sim\mathbb{P}(X=0)^{-1}. \qquad (3.7)$$

Thus, we have reduced the problem to that of estimating S_0.

For use in the following, we define

$$\zeta(j) = (j_1/\lambda_1, \dots, j_{t-1}/\lambda_{t-1}, j/\lambda_t) \tag{3.8}$$

(with the dependence on $j_1, \dots, j_{t-1}$ suppressed for compactness of notation), and we say that $\zeta(j)$ is *appropriate* if $j_i \in [0, m_i]$ for all $i \in [1, t-1]$ and $j \in [0, m_t]$.

It is useful to define $\mathcal{P}^+(\mathbf{g_t})$ to be the ring of polynomials in $\mathbf{g_t} = (g_1, \dots, g_t)$ whose coefficients are polynomials in n, p and n^{-1} (as formal indeterminates), and $\mathcal{P}(\mathbf{g_t})$ to be the subring of $\mathcal{P}^+(\mathbf{g_t})$ consisting of those polynomials whose coefficient of $g_1^{i_1} \cdots g_t^{i_t}$ is divisible by $p^{\sum_j \mu(j) i_j}$. By permitting p, n and n^{-1} to commute, and setting $n \cdot n^{-1} = 1$, we can regard these coefficients as a ring consisting simply of the union of the set of polynomials in n and p with the set of polynomials in n^{-1} and p.

We will use the definition of S_t, together with Lemma 2.5 and Corollary 2.7 and an induction argument to prove that

$$S_0 = \exp\big(P_{0,\kappa} + o(1)\big), \qquad S_t(j_1, \dots, j_t) = \exp\big(P_{t,\kappa}\big(\zeta(j_t)\big) + o(1)\big) \quad (1 \le t \le s) \tag{3.9}$$

for all $j_1, \dots, j_t$ such that $\zeta(j_t)$ is appropriate, for some polynomials $P_{t,\kappa}$ such that

(i) $P_{t,\kappa} \in \mathcal{P}(\mathbf{g_t})$ (and so in particular for $t = 0$, $P_{0,\kappa}$ is a polynomial in n, p and n^{-1});

(ii) the constant coefficient of $P_{t,\kappa}$ (i.e. $P_{t,\kappa}(0, 0, \dots, 0)$) is equal to $(1 + O(xn^{o(1)})) \sum_{t'=t+1}^{s} \lambda_{t'}$ and the other coefficients are $O(xn^{o(1)} P_{t,\kappa}(0, 0, \dots, 0))$, where the implicit bounds in $O(\cdot)$ are independent of ϵ. Note that by (3.2), it follows that the constant coefficient is $O(n^{o(1)}\lambda_t)$;

(iii) the convergence expressed by $o(1)$ in (3.9) is uniform over all appropriate $j_1, j_2, \dots, j_t$.

The induction begins with $t = s$ and then proceeds through decreasing values of t. It finishes with the case $t = 0$ of (3.9), which is used to show that the polynomials $P_{t,\kappa}(\mathbf{g_t})$ are of such a form that the theorem follows using (3.7).

The initial step of the induction argument, $t = s$, is trivial, since S_s is identically equal to 1 and we may set $P_{s,\kappa} = 0$. So now suppose that (3.9) holds for some particular value of t. We must prove that it also holds when $t - 1$ is substituted for t. Define T_j by

$$T_j = \exp\big(P_{t,\kappa}\big(\zeta(j)\big)\big) \left(\prod_{i=0}^{j-1} \xi_{t,\epsilon}\left(f_{t,i}\right) \right) \frac{\lambda_t^j}{j!}. \tag{3.10}$$

We now use (3.5), (3.9) and Corollary 2.7 to replace $\overline{\gamma}$ in (3.5) by ξ, the fact that $\zeta(j)$ is appropriate and $m_t = O(\lambda_t \log n)$, together with (2.19), to obtain

$$S_{t-1}(j_1, j_2, \dots, j_{t-1}) \sim \sum_{j=0}^{\lfloor m_t \rfloor} T_j. \tag{3.11}$$

First assume that $t \in \mathcal{S}_0$. Note that (with square brackets for extraction of coefficients)

$$\frac{\exp P_{t,\kappa}\big(\zeta(j)\big)}{\exp P_{t,\kappa}\big(\zeta(0)\big)} = \exp \sum_{\mathbf{i}} \left([\mathbf{g}^{\mathbf{i}}]P_{t,\kappa}\right) \left(\prod_{\ell=1}^{t-1} (j_\ell/\lambda_\ell)^{i_\ell}\right) \left((j/\lambda_t)^{i_t} - 0^{i_t}\right)$$

where $0^0 = 1$ as usual, and the summation is over the set of $\mathbf{i}$ for which the coefficient is nonzero. The number of such $\mathbf{i}$ is bounded, given $P_{t,\kappa}$. The only terms contributing have $i_t > 0$, and in particular the constant term does not contribute. Let $v_{\max}$ be the total degree of $P_{t,\kappa}(\mathbf{g_t})$. Each factor j_ℓ/λ_ℓ is by (2.22) at most $\log n = n^{o(1)}$, and the same goes for j/λ_t. By the inductive hypothesis (ii) and (2.21), we now obtain

$$\frac{\exp P_{t,\kappa}\big(\zeta(j)\big)}{\exp P_{t,\kappa}\big(\zeta(0)\big)} = \exp\left(xn^{o(1)}(n^{o(1)})^{v_{\max}}\right) = 1 + O(xn^{o(1)}).$$

By Lemma 2.5, $\overline{\gamma}(t) = 1 + O(x)$, and so Corollary 2.7 gives that each factor $\xi_{t,\epsilon}\,(f_{t,i})$ in (3.10) is $1 + O\big(x + n^{-\overline{\epsilon}+o(1)}\big)$ Hence for $j \le m_t = n^{o(1)}$, the product of j factors in (3.10) is

$$\Big(1 + O\big(x + n^{o(1)-\epsilon'}\big)\Big)^{m_t} \sim 1$$

using (2.14). Thus

$$T_j \sim \exp P_{t,\kappa}\big(\zeta(0)\big)\lambda_t^j/j! \sim S_t(j_1, j_2, \ldots, j_{t-1}, 0)\lambda_t^j/j!$$

by the inductive hypothesis (3.9). Since in this case $m_t = \lambda_t \log n$, it follows that

$$S_{t-1}(j_1, j_2, \ldots, j_{t-1}) = S_t(j_1, j_2, \ldots, j_{t-1}, 0) \exp\left(\lambda_t + o(1)\right).$$

Here we used the uniformity of the convergence in the estimates, including that asserted in part (iii) of the induction hypothesis. To establish the inductive hypothesis in this case, we thus set $P_{t-1,\kappa}$ equal to $P_{t,\kappa} + \lambda_t$, which is a polynomial in n and p (thus a constant in $\mathcal{P}(\mathbf{g})$) by (2.4) and (2.5). This clearly gives the inductive hypotheses (i) and (ii), whilst the uniformity in (iii) implies that (iii) holds with t replaced by $t-1$.

We next suppose that $t \in \mathcal{S}_1$, so that in particular $\lambda_t \to \infty$ by (2.22). We need to estimate the ratio of consecutive terms T_j quite accurately. We have

$$\frac{T_j}{T_{j-1}} = \exp\Big(P_{t,\kappa}\big(\zeta(j)\big) - P_{t,\kappa}\big(\zeta(j-1)\big)\Big)\xi_{t,\epsilon}(f_{t,j-1})\frac{\lambda_t}{j}. \tag{3.12}$$

Let $R_v = [g_t^v]P_{t,\kappa}(\mathbf{g_t})$, so that $R_v \in \mathcal{P}(\mathbf{g}_{t-1})$. Put

$$\hat{\zeta} = (j_1/\lambda_1, \ldots, j_{t-1}/\lambda_{t-1})$$

and

$$\eta = n^{-\epsilon/2}/\lambda_t. \tag{3.13}$$

Then

$$
\begin{aligned}
P_{t,\kappa}\big(\zeta(j)\big) - P_{t,\kappa}\big(\zeta(j-1)\big) &= \sum_{v=1}^{v_{\max}} R_v(\hat{\zeta})\left(\left(\frac{j}{\lambda_t}\right)^v - \left(\frac{j-1}{\lambda_t}\right)^v\right) \\
&= \sum_{v=1}^{v_{\max}} R_v(\hat{\zeta})\left(\frac{vj^{v-1} + O(j^{v-2})}{\lambda_t^v}\right) \\
&= \sum_{v=1}^{v_{\max}} \frac{vR_v(\hat{\zeta})}{\lambda_t} \cdot \frac{j^{v-1}}{\lambda_t^{v-1}} + O(R_v(\hat{\zeta})/\lambda_t^2) \\
&= O(\eta) + \sum_{v=1}^{v_{\max}} \frac{vR_v(\hat{\zeta})}{\lambda_t} \cdot \frac{j^{v-1}}{\lambda_t^{v-1}}
\end{aligned}
$$

since $j = O(\lambda_t)$ by (2.21) (and $j_i = O(\lambda_i)$ for $i < t$), and using the inductive hypothesis (ii), which implies that the coefficients of R_v for $v \geq 1$ are all $O(n^{o(1)}x\lambda_t) = O(\eta\lambda_t^2)$ by (2.14). For the same reason, the terms in this summation are all $O(n^{o(1)}x)$.

We call a polynomial $\tilde{P}$ *acceptable* if $\tilde{P} = 1 + P$ for some polynomial $P \in \mathcal{P}(\mathbf{g})$ whose coefficients are all $O(n^{o(1)}x)$ for the range of p under consideration, i.e. satisfying (2.13). Note that $n^{o(1)}x = o(n^{-\epsilon/2}) = o(1)$ by (2.14). A polynomial $\tilde{P}$ is *t-acceptable* if $\tilde{P} = 1 + P$ for some polynomial $P \in \mathcal{P}^+(\mathbf{g_t})$ whose coefficient of $g_1^{i_1} \cdots g_t^{i_t}$ is divisible by $p^{\sum_{j<t}\mu(j)i_j}$, and whose coefficients are all $O(n^{o(1)}x)$ for p satisfying (2.13). That is, $\tilde{P}$ satisfies the definition of an acceptable polynomial in $\mathcal{P}(\mathbf{g_t})$ except that the powers of p in the terms in P are only required to pay their respect to the variables $g_1, \ldots, g_{t-1}$.

By (2.5), λ_t^{-1} can be expanded as $p^{-\mu(t)}$ times a power series in n^{-1}. So by the inductive assumption that $P_{t,\kappa} \in \mathcal{P}(\mathbf{g_t})$, it follows that there exists $\tilde{R}_v \in \mathcal{P}(\mathbf{g}_{t-1})$ such that $R_v(\hat{\zeta})/\lambda_t = \tilde{R}_v(\hat{\zeta}) + O(\eta)$. To verify this, we note that g_t does not appear in R_v and hence the lower bound on the exponent of p required for $P_{t,\kappa}$'s membership in $\mathcal{P}(\mathbf{g_t})$ is enough to compensate for $p^{-\mu(t)}$; the power series in n^{-1} can be truncated at an appropriate point to obtain a polynomial in n^{-1}, producing the error term $O(\eta)$.

We conclude that

$$\exp\Big(P_{t,\kappa}\big(\zeta(j)\big) - P_{t,\kappa}\big(\zeta(j-1)\big)\Big) = A_{t,\kappa}^{(1)}(\zeta(j))(1 + O(\eta))$$

for a t-acceptable polynomial $A_{t,\kappa}^{(1)}$ (with constant term precisely 1 in this case).

By Lemma 2.5 and Corollary 2.7(b), we see that $\xi_{t,\epsilon}$ is acceptable and consequently t-acceptable. Consequently, $\xi_{t,\epsilon}(f_{t,j-1})$ that occurs in (3.12) is equal to $\xi_{t,\epsilon}(f_{t,j})(1 + O(\eta))$. Moreover, the product of two t-acceptable polynomials is t-acceptable. Thus (3.12) gives

$$\frac{T_j}{T_{j-1}} = \frac{A_{t,\kappa}^{(2)}(\zeta(j))\lambda_t(1 + O(\eta))}{j} \tag{3.14}$$

for the t-acceptable polynomial

$$A_{t,\kappa}^{(2)}(\mathbf{g_t}) := A_{t,\kappa}^{(1)} \cdot \tilde{\xi}_{t,\epsilon}, \tag{3.15}$$

where $\tilde{\xi}_{t,\epsilon}$ is obtained from $\xi_{t,\epsilon}$ by setting $g_{t+1} = \cdots = g_s = 0$.

To identify (approximately) the maximum term of the summation in (3.11), we note that since $A^{(2)}_{t,\kappa}$ is t-acceptable, $A^{(2)}_{t,\kappa}(\zeta(j)) \sim 1$ and so (3.14) shows that we are interested in $j \sim \lambda_t$. Furthermore, again using t-acceptability, the derivative of $A^{(2)}_{t,\kappa}(g_1, \dots, g_{t-1}, y)$ with respect to y is $o(n^{-\epsilon/2})$ when $\mathbf{g}_{t-1}$ is set equal to $\hat{\zeta}$. So, at least for large n, this function has a fixed point y that is $1 + o(1)$. In other words, there must exist $j^* \sim \lambda_t$ satisfying

$$j^* = \lambda_t A^{(2)}_{t,\kappa}(\zeta(j^*)). \tag{3.16}$$

Since $A^{(2)}_{t,\kappa}$ is t-acceptable, we can use repeated substitutions in

$$Q_\ell = A^{(2)}_{t,\kappa}(g_1, \dots, g_{t-1}, Q_{\ell-1})$$

beginning with $Q_0 = 1$ to obtain a polynomial $Q_\ell \in \mathcal{P}(\mathbf{g}_{t-1})$ such that $Q_\ell(\hat{\zeta})$ is an approximation to j^*/λ_t. Clearly, replacing the variable g_t of a t-acceptable polynomial by another t-acceptable polynomial produces yet another t-acceptable polynomial. So each Q_ℓ is t-acceptable. For each iteration, the error in the approximation is multiplied by $o(n^{-\epsilon/2})$. Hence, for ℓ sufficiently large, Q_ℓ is an acceptable polynomial $A^{(3)}_{t,\kappa} \in \mathcal{P}(\mathbf{g}_{t-1})$ satisfying

$$j^* = \lambda_t A^{(3)}_{t,\kappa}(\hat{\zeta}) + o(1) \tag{3.17}$$

uniformly for all $\hat{\zeta}$ under consideration.

For the product in (3.10) we will use the following. Recall that $\xi_{t,\epsilon}$ is a polynomial, whereas $\xi_{t,\epsilon}(f_{t,i})$ is a number given n and p (and in the present context n determines p). Since $\xi_{t,\epsilon}$ is acceptable, we may expand its logarithm and hence obtain

$$\log \xi_{t,\epsilon}(f_{t,i}) = \sum_{v=0}^{v^{(1)}_{\max}} R^{(1)}_v(\hat{\zeta}) \left(\frac{i}{\lambda_t}\right)^v + o(\lambda_t^{-1}) \tag{3.18}$$

for some $v^{(1)}_{\max}$, with $R^{(1)}_v \in \mathcal{P}(\mathbf{g}_{t-1})$ having all coefficients $O(n^{o(1)}x)$ for all $v \le v_{\max}$. (That is, $1 + R^{(1)}_v$ is acceptable.) Then

$$\begin{aligned}
\sum_{i=0}^{j-1} \log \xi_{t,\epsilon}(f_{t,i}) &= \sum_{i=0}^{j-1} \sum_{v=0}^{v^{(1)}_{\max}} R^{(1)}_v(\hat{\zeta}) \left(\frac{i}{\lambda_t}\right)^v + o(j/\lambda_t) \\
&= \sum_{v=0}^{v^{(1)}_{\max}} \frac{R^{(1)}_v(\hat{\zeta})}{v+1} \cdot \frac{j^{v+1}}{\lambda_t^v} + \sum_{v=0}^{v^{(1)}_{\max}} \frac{O(j^v) R^{(1)}_v(\hat{\zeta})}{\lambda_t^v} + o(j/\lambda_t) \\
&= o(1) + \lambda_t \sum_{v=0}^{v^{(1)}_{\max}} \frac{R^{(1)}_v(\hat{\zeta})}{v+1} \left(\frac{j}{\lambda_t}\right)^{v+1}.
\end{aligned} \tag{3.19}$$

We wish to approximate the terms in (3.11) by expanding the formula for T_j given in (3.10) about $j = j^*$, beginning with (3.14) written as

$$\log(T_j/T_{j-1}) = q(j) + O(\eta) \tag{3.20}$$

where

$$q(j) = \log A^{(2)}_{t,\kappa}(\zeta(j)) + \log\lambda_t - \log j. \tag{3.21}$$

Note that this equation also defines $q(y)$ for an arbitrary non-integer real y, so we can consider its derivative $q'(y)$. Since $A^{(2)}_{t,\kappa}$ is t-acceptable, we have for some $v^{(2)}_{\max}$ and $R^{(2)}_v \in \mathcal{P}(g_1,\dots,g_{t-1})$ with all coefficients of size $O(n^{o(1)}x)$ that

$$\begin{aligned} q'(y) &= \frac{d}{dy}\left(\sum_{v=0}^{v_{\max}{}^{(2)}} R^{(2)}_v(\hat\zeta)\left(\frac{y}{\lambda_t}\right)^v\right) - \frac{1}{y} \\ &= -\frac{1}{y} + O\left(\frac{n^{o(1)}x}{\lambda_t}\right) \\ &= -\frac{1}{j^*} + O\left(\frac{n^{o(1)}x}{\lambda_t} + \frac{|y-j^*|}{(j^*)^2}\right) \end{aligned} \tag{3.22}$$

for $|y-j^*| = o(j^*)$, and on the other hand, from the definition of j^*, $q(j^*)=0$. It follows that for $k = j^* + O(\sqrt{j^*}\log j^*)$, we have (again noting $j^* \sim \lambda_t$)

$$q(k) = \int_{j^*}^{k} q'(y)\,dy = -\frac{k-j^*}{j^*} + o\left((j^*)^{-1/2}xn^{o(1)}\right).$$

Thus, for the same range of k, summing (3.20) over j between $\tilde{j} := \lfloor j^*\rfloor$ and k gives

$$\log(T_k/T_{\tilde{j}}) = \frac{-(k-\tilde{j})^2}{2j^*} + o(1) \tag{3.23}$$

(and this argument applies whether k is smaller or larger than $\tilde{j}$). Hence, the sum of T_k for $k = j^* + O(\sqrt{j}^*\log j^*)$ is asymptotic to $T_{\tilde{j}}$ times the sum of $e^{-(k-\tilde{j})^2/2j^*}$ over the same range, and is hence

$$(2\pi j^*)^{1/2}T_{\tilde{j}}(1+o(1)).$$

Also, (3.23) is valid at the extreme ends of the range, i.e. $k = j^* + \Theta(\sqrt{j}^*\log j^*)$. Thus, recalling (3.14), all the terms in (3.11) outside the range $k = j^* + O(\sqrt{j}^*\log j^*)$ are negligible and

$$\sum_{j=0}^{\lfloor m_t\rfloor} T_j \sim (2\pi\tilde{j})^{1/2}T_{\tilde{j}}. \tag{3.24}$$

To estimate $T_{\tilde{j}}$, we use Stirling's formula and then $j^* \sim \lambda_t$ and $|\tilde{j} - j^*| < 1$ to write

$$\frac{\lambda_t^{\tilde{j}}}{\tilde{j}!} \sim \frac{(e\lambda_t/\tilde{j})^{\tilde{j}}}{\sqrt{2\pi\tilde{j}}} \sim \frac{(e\lambda_t/j^*)^{j^*}}{\sqrt{2\pi\tilde{j}}}. \tag{3.25}$$

Using (3.17) we may expand the logarithm of $1/A^{(3)}_{t,\kappa}$ to obtain, for some acceptable polynomials $A^{(4)}_{t,\kappa}$ and $A^{(5)}_{t,\kappa}$ in $\mathcal{P}(\mathbf{g}_{t-1})$,

$$\log(\lambda_t/j^*) = A^{(4)}_{t,\kappa}(\hat\zeta) - 1 + o(1/\lambda_t)$$

and then

$$(e\lambda_t/j^*)^{j^*} = \exp(j^* \log(e\lambda_t/j^*)) = \exp(\lambda_t A^{(5)}_{t,\kappa}(\hat{\zeta}) + o(1)). \tag{3.26}$$

(Here $A^{(5)}_{t,\kappa}$ just contains the significant terms of $A^{(3)}_{t,\kappa} \cdot A^{(4)}_{t,\kappa}$.) Next, from (3.19) with $j = \tilde{j}$ we have, for some t-acceptable polynomial $A^{(6)}_{t,\kappa}$,

$$\sum_{i=0}^{\tilde{j}-1} \log \xi_{t,\epsilon}(f_{t,i}) = \lambda_t\big(A^{(6)}_{t,\kappa}(\zeta(\tilde{j})) - 1\big) + o(1). \tag{3.27}$$

For example, if $\xi_{t,\epsilon}$ happens not to contain g_t, then $A^{(6)}_{t,\kappa}$ is equal to $1 + g_t \widehat{\log}\, \xi_{t,\epsilon}$, where $\widehat{\log}$ denotes the logarithm truncated to significant terms. Since $|j^* - \tilde{j}| < 1$ and $A^{(6)}_{t,\kappa}$ is t-acceptable, we may replace $\tilde{j}$ in the right hand side of (3.27) by j^*, with no other change to the equation. Using this, together with (3.25) and (3.26), in (3.10) with $j = \tilde{j}$, we may transform (3.24) into

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_j \sim \exp\Big(P_{t,\kappa}\big(\zeta(\tilde{j})\big) + \lambda_t A^{(6)}_{t,\kappa}\big(\zeta(j^*)\big) - \lambda_t + \lambda_t A^{(5)}_{t,\kappa}\big(\hat{\zeta}\big)\Big). \tag{3.28}$$

Note that $A^{(6)}_{t,\kappa} - 1 + A^{(5)}_{t,\kappa}$ is t-acceptable. Then the expansion (3.17) calls for replacing g_t in $A^{(6)}_{t,\kappa}$ by $A^{(3)}_{t,\kappa}(\hat{\zeta})$:

$$A^{(6)}_{t,\kappa}\big(\zeta(j^*)\big) - 1 + A^{(5)}_{t,\kappa}\big(\hat{\zeta}\big) = A^{(6)}_{t,\kappa}\big(\zeta\big(\lambda_t A^{(3)}_{t,\kappa}(\hat{\zeta})\big)\big) - 1 + A^{(5)}_{t,\kappa}\big(\hat{\zeta}\big) = A^{(7)}_{t,\kappa}\big(\hat{\zeta}\big) + o(1/\lambda_t)$$

for some acceptable polynomial $A^{(7)}_{t,\kappa} \in \mathcal{P}(\mathbf{g}_{t-1})$. Also, by hypothesis (ii) and the fact that $|j^* - \tilde{j}| < 1$, we have $P_{t,\kappa}\big(\zeta(\tilde{j})\big) = P_{t,\kappa}\big(\zeta(j^*)\big) + o(1)$. Again replacing g_t by $A^{(3)}_{t,\kappa}(\hat{\zeta})$, using (3.17) we obtain

$$P_{t,\kappa}\big(\zeta(\tilde{j})\big) = \tilde{P}_{t,\kappa}(\hat{\zeta}) + o(1)$$

for a polynomial $\tilde{P}_{t,\kappa} \in \mathcal{P}(\mathbf{g}_{t-1})$ that has exactly the properties described in (ii) for $P_{t,\kappa}$.

Note that there are multiple valid choices for $\tilde{P}_{t,\kappa}$ at this point, due to the possible inclusion of negligible terms. To avoid ambiguity, we specify that the terms that are retained are exactly those that are significant in this argument when p is precisely $n^{-\kappa}$, that is, terms of order $n^a p^b$ for which $a/b \geq k$.

Now from (3.11) and (3.28) we have

$$S_{t-1}(j_1, j_2, \ldots, j_{t-1}) \sim \exp\Big(\tilde{P}_{t,\kappa}(\hat{\zeta}) + \lambda_t A^{(7)}_{t,\kappa}\big(\hat{\zeta}\big)\Big). \tag{3.29}$$

We may now set

$$P_{t-1,\kappa} = \tilde{P}_{t,\kappa} + \lambda_t A^{(7)}_{t,\kappa}$$

to obtain parts (i) and (ii) of the inductive hypothesis. Indeed, by this recursive definition we obtain that

$$P_{t,\kappa} = \sum_{t'=t+1}^{s} \lambda_{t'} A_{t',\kappa}$$

for some acceptable polynomials $A_{t',\kappa}$. Verifying part (iii) of the inductive hypothesis requires simply noticing that the estimates in the above derivation are, inductively, uniform over all appropriate $\hat{\zeta}$. This uses the uniformity of the estimates in Lemma 2.5 and Corollary 2.7.

The inductive step is now fully established, and we have (3.9) for all t. Taking $t = 0$, (3.7) shows that

$$\mathbb{P}(X = 0) \sim \exp(-P_{0,\kappa}). \tag{3.30}$$

By part (ii) of the inductive hypothesis, $P_{0,\kappa} = (1 + O(n^{o(1)}x)) \sum_{t \in \mathcal{S}} \lambda_t$.

We now show that

$$\text{the polynomial } P_{0,\kappa} \text{ is a truncation of } P_{0,\chi+\epsilon} \text{ for all } \chi + \epsilon \leq \kappa < 2 - \epsilon'', \tag{3.31}$$

(where the upper bound $2 - \epsilon''$ arises from (2.20)). This statement immediately requires some qualification. In the definition of $P_{t,\kappa}$, it is important to note that any expansions during the proof above must be taken in the formal sense. For instance, if $\chi + \epsilon$ happens to take certain rational values, then some terms in an expansion of the form $n^a p^b$ might happen to be equal to other terms $n^c p^d$, but these terms should be kept separate when comparing polynomials.

We begin by showing that there is no ambiguity in the definition of $P_{t,\kappa}$ due to the arbitrariness of ordering of the types in $\mathcal{S}$. That is, we show that the various orderings of types that are valid all lead to the same terms in $P_{t,\kappa}$. Consider two possible orderings of types π and $\tilde{\pi}$. For each choice of ordering there corresponds a polynomial $P_{t,\kappa}$ in (3.30). Let us refer to the function $o(1)$ in (2.13) as $g(n)$. Since $g(n)$ may be taken so that $n^{g(n)}$ is any positive constant function, and for all such functions the two polynomials must have equal values to within $o(1)$, all terms in the polynomials that are bounded below when $n^{g(n)}$ is constant must be equal. Terms that tend to 0 when $n^{g(n)}$ is constant must be $n^{-\epsilon'}$ for some $\epsilon' > 0$ and hence cannot occur in these polynomials.

We continue with the main part of the proof of (3.31). Note first, as an easy argument shows, that as κ increases smoothly from $\chi + \epsilon$ to $2 - \epsilon''$, there is a finite number of values of κ at which the ordering of the types can change, or a type changes from small to large. (Recall that, as κ increases, p decreases, and hence every λ_i decreases, and hence a type can move from $\mathcal{S}_1$ to $\mathcal{S}_0$, and at essentially the same κ from $\mathcal{S}_0$ to large, but not in the reverse direction.) These are special values for our argument, since the ordering of types determines the order of expansions in the inductive arguments concerning S_t. We designate the minimum value, $\chi + \epsilon$, also as one of these special values, κ_0, and let the others be $\kappa_1, \kappa_2, \ldots$, with $\kappa_0 < \kappa_1 < \cdots$.

Let us first fix two of these distinct values of κ, $\kappa_i < \kappa_{i+1}$, and consider κ in the open interval (κ_i, κ_{i+1}). First, we will show that in the inductive argument given above, for such κ, we may use P_{t,κ_i} in the argument in place of $P_{t,\kappa}$ (subject to some near-trivial modification we will describe). We show moreover that $P_{t,\kappa}$ is a truncation of P_{t,κ_i}. To be precise, we claim that all the expansions in the argument for κ can be replaced by the corresponding ones from the argument for κ_i. The difference between the corresponding expansions lies only in the terms that are absorbed by the error terms in the argument for κ. To see this inductively, we need only to modify the argument for κ slightly. We describe various aspects of the two arguments as being "for κ" or "for κ_i" to distinguish between the two versions.

The inductive argument for κ begins with a maximal $t \in \mathcal{S}$. Since κ is not a special value, it cannot be true that $t \in \mathcal{S}_0$. However, it may happen that a type t' is large for κ but small (and hence in $\mathcal{S}_0$) for κ_i. By what has been shown about ordering types arbitrarily, we may assume that types that are small for κ have the same ordering for κ as they do for κ_i. For any type like the above-mentioned t', we may extend the definitions in the argument for κ by putting $S_{t'} = 1$, and it is easy to verify that $P_{t',\kappa_i} = o(1)$ when evaluated at the value of p occurring in the argument for κ, i.e. $p = n^{-\kappa+o(1)}$. As the remaining types have identical order, it remains to be shown that if $t \in \mathcal{S}$ for κ, then $P_{t,\kappa}$ equals P_{t,κ_i} except for those terms of P_{t,κ_i} which are $o(1)$ for κ.

At every point in the argument above for arbitrary κ that an expansion is called for, beginning with the use of $\overline{\gamma}$ in (3.12), we may add the extra terms called for in the κ_i argument, and note that they fall into the error terms in the equation concerned. In particular, for (3.12) this is true because of the assertion about the truncations in Corollary 2.7. Then, since this equation (and those following it) is true with these extra terms, the argument works as before, with expansions being carried out and with truncations determined by the argument for κ_i rather than κ. Every step of the argument then preserves the expansions obtained in the argument for κ_i, but all other aspects of the argument are as for κ. This is immediately obvious in places where products of series, and logarithms, are expanded, but it is a little more subtle in the part involving j^*, so we examine this in more detail.

We need to show that $A^{(3)}_{t,\kappa}$ equals $A^{(3)}_{t,\kappa_i}$ up to insignificant terms. Let $\tilde{\lambda}_t$ be λ_t with $p = n^{-\kappa_i+o(1)}$. Let $\tilde{A}^{(3)}_{t,\kappa_i}$ be the polynomial derived with $p = n^{-\kappa_i+o(1)}$ but evaluated at $p = n^{-\kappa+o(1)}$. We can write $A^{(3)}_{t,\kappa_i}$ as $A^{(3)}_{t,\kappa_i} = B^{(3)}_{t,\kappa_i} + C^{(3)}_{t,\kappa_i} + D^{(3)}_{t,\kappa_i}$, where $D^{(3)}_{t,\kappa_i} = o(1/\tilde{\lambda}_t)$ and where $C^{(3)}_{t,\kappa_i}$ is significant for $p = n^{-\kappa_i+o(1)}$ but such that $\tilde{C}^{(3)}_{t,\kappa_i} = o(1/\lambda_t)$. We constructed $D^{(3)}_{t,\kappa_i}$ from a given number of contractions and the contraction constant is smaller for κ than it is for κ_i (for the contractions obtained when the coefficients in $A^{(2)}_{t,\kappa_i}$ and $A^{(2)}_{t,\kappa}$ are replaced by their absolute values), hence $\tilde{D}^{(3)}_{t,\kappa_i} \leq D^{(3)}_{t,\kappa}$ and $\tilde{D}^{(3)}_{t,\kappa_i} = o(1/\lambda_t)$. All of the remaining steps in the argument for $p = n^{-\kappa+o(1)}$ involve sums, products, expansions of logarithms or substitutions into polynomials and so everything arising from $\tilde{C}^{(3)}_{t,\kappa_i}$ is of the order $o(1/\lambda_t)$. Thus, ignoring $o(1/\lambda_t)$ terms, $\tilde{A}^{(3)}_{t,\kappa_i} = A^{(3)}_{t,\kappa}$.

Next, we will show that the inductive argument given above, for κ_{i+1}, remains valid if we use $P_{t,\kappa}$ in the argument in place of $P_{\kappa_{i+1}}$, and that $P_{\kappa_{i+1}}$ is a truncation of $P_{t,\kappa}$. In this case, no type can move from being small for the κ argument to being large for the κ_{i+1} argument (since $\kappa_{i+1} > \kappa$), but possibly a type t is in $\mathcal{S}_1$ for the case of κ and in $\mathcal{S}_0$ for the case of κ_{i+1}. By part (ii) of the inductive hypothesis, the contribution from the type t to $P_{t,\kappa}$ is $\lambda_t + o(1)$ when p is taken in the appropriate range for κ_{i+1} because then $\lambda_t = n^{o(1)}$, and moreover this is also the contribution to $P_{t,\kappa_{i+1}}$. The rest of the argument for this case only involves considering the expansions, so is similar to the argument above.

Statement (3.31) now follows by induction from the statements that $P_{t,\kappa}$ is a truncation of P_{t,κ_i} and that $P_{\kappa_{i+1}}$ is a truncation of $P_{t,\kappa}$. In view of the argument above that decreasing κ simply adds more terms to $P_{0,\kappa}$, we see that decreasing ϵ does the same thing to $P_{0,\chi+\epsilon}$. Hence, this is the truncation to a finite number of terms of a power series $F(G_0)$ in n and p. Since there is a bounded number of terms in (1.3) that are $o(1)$ for a given κ, we have now established (1.3) for this power series $F(G_0)$ and for $p = n^{-\kappa+o(1)}$ (whenever $\kappa \geq \chi + \epsilon$). In particular, with the terms $c_\ell n^{i_\ell} p^{j_\ell}$ arranged in decreasing order of i_ℓ/j_ℓ, the claimed characterisation of M_ϵ

follows. Note that the function represented by $o(1)$ in (1.3) is given explicitly by

$$f(n,p) = \log\left(\mathbb{P}(X=0)\right) - \sum_{\ell=0}^{M_\epsilon} c_\ell n^{i_\ell} p^{j_\ell}.$$

We may now apply Lemma 1.4 with $a = \chi + \epsilon$ and $b = 2 - \epsilon''$ to deduce that the convergence in (1.3) is uniform over all $\kappa \in [\chi + \epsilon, 2 - \epsilon'']$.

All that remains is to show the strict positivity of the exponents i_ℓ and j_ℓ in $F(G_0)$. Note that a term $n^{i_\ell} p^{j_\ell}$ with $i_\ell \le 0$ must have $j_\ell < 0$, otherwise it is always $o(1)$ and can simply be omitted. However, such a term is decreasing in p, so, if it is ever significant, must be so when $p \le n^{-2+\epsilon''}$. However, at that point we know $\mathbb{P}(X=0) \sim 1$, and hence the term must be insignificant here too. Thus, such terms can be dropped. It follows that we may assume $i_\ell > 0$. Given (by the same argument) that the term must be insignificant for small p, we deduce that $j_\ell > 0$ also. The $\mathcal{G}(n,p)$ case of the theorem follows. ∎

4 Graphs with forbidden subgraphs in $\mathcal{G}(n,m)$

We will show that the $\mathcal{G}(n,p)$ case of Theorem 1.1 can be extended to give a similar result in $\mathcal{G}(n,m)$ without much difficulty. Specifically, we provide asymptotics for the probability of $\mathcal{G}(n,m)$ not containing a fixed subgraph isomorphic to G_0. The asymptotics could be expressed in terms of n and m, but it is more convenient to use n and the parameter $d = m/\binom{n}{2}$ defined in (1.1). We employ the $\mathcal{G}(n,p)$ case inside the proof, for a value of p that is close, but not quite equal, to d, though for the statement of the theorem we have renamed d as p for convenience.

Proof of the $\mathcal{G}(n,m)$ case of Theorem 1.1.

Let Y denote the number of edges of a graph. The probability that $X=0$ in $\mathcal{G}(n,m)$ is precisely $\mathbb{P}(X=0 \mid Y=m)$ in $\mathcal{G}(n,p)$. In the rest of the proof we estimate this quantity, with all probabilities referring to $\mathcal{G}(n,p)$. By Bayes' Theorem, what we desire is

$$\mathbb{P}(X=0 \mid Y=m) = \mathbb{P}(Y=m \mid X=0)\frac{\mathbb{P}(X=0)}{\mathbb{P}(Y=m)} \tag{4.1}$$

This formula is valid for all $0<p<1$. The value of p we will use, which is specified below, is asymptotic to d and hence lies in the range required for the $\mathcal{G}(n,p)$ case of Theorem 1.1, given by (2.13) with the same restrictions on κ, which determines $\mathcal{S}$ via (2.16). Thus, Theorem 1.1 gives us $\mathbb{P}(X=0)$ in $\mathcal{G}(n,p)$.

The main difficulty is computing $\mathbb{P}(Y=m \mid X=0)$. For this, we will first alter the analysis in Section 3 to consider the G_0^*-clustering in $\mathcal{G}(n,p)$. Recall that this is obtained by adding to $\mathcal{S}$ the type t^* of maximal cluster corresponding to a single edge. For convenience, we henceforth denote the cluster type t^* by 0, and modify the definition of $\tilde{g}$ for (2.47) accordingly with g_0 substituted by $f(0)/\lambda_0$.

Considering the polynomial $\xi_{0,\epsilon}(n,p,\mathbf{g})$ in Corollary 2.7, for $j/\lambda_0 \le 3$ (in accordance with (2.23)), by part (c) of that Corollary

$$\overline{\gamma}_0(j\delta_0) = \xi_{0,\epsilon}(n,p,\hat{\mathbf{g}}(j)) + o(\lambda_0^{-1}), \tag{4.2}$$

where $\hat{g}_0(j) = j/\lambda_0$, $\hat{g}_i(j) = 0$ for $i \geq 1$, provided that $p = p(n) = O(n^{-\chi-\epsilon})$ and satisfies (2.20).

Also define $\tilde{\mathbf{g}}$ by $\tilde{g}_0 = m/\lambda_0$ and $\tilde{g}_i = 0$ for $i \geq 1$, and let ξ denote $\xi_{0,\epsilon}(n, p, \tilde{\mathbf{g}})$ (noting that ξ is a function of n, p and m. As $\hat{g}_0(j) \leq 3 = O(1)$ we have

$$\xi_{0,\epsilon}\big(n, p, \hat{\mathbf{g}}(j)\big) = \xi + O\big(x(m-j)/\lambda_0\big) \tag{4.3}$$

by Corollary 2.7(a).

By the definitions of $\rho(f, h)$ and $\gamma(f, t)$ before Proposition 2.2, one would expect that the probability that $\mathcal{G}(n, p)$ has no copies of G_0 and m' edges will be maximised, given p, at $m' \approx m$ provided that $\rho(m\delta_0, \delta_0) \approx 1$, or $\gamma(m\delta_0, 0) \approx m/\lambda_0$. On the other hand, in $\mathcal{G}(n, p)$ the ratio of the probabilities of having a given number of edges, when increasing m to $m+1$, is approximately d/p. Consequently, we define p by

$$p = d/\xi \tag{4.4}$$

(recalling that (1.1) gives d as a function of n and m). Then (4.2) and Lemma 2.5 imply that

$$p = d\big(1 + O(x + \lambda_0^{-1})\big), \tag{4.5}$$

and hence our assumptions on d imply the necessary properties of p such as (2.20), perhaps with different values of the unimportant constants.

From the $\mathcal{G}(n, p)$ case of Theorem 1.1, $\mathbb{P}(X = 0)$ in $\mathcal{G}(n, p)$ is $e^{-\Theta(\lambda_t)}$, and $\lambda_t = o(\lambda_0)$ by (2.12). On the other hand, The number Y of edges in $\mathcal{G}(n, p)$ is distributed as $\mathrm{Bin}(N, p)$ where $N = \binom{n}{2}$, with mean $\lambda_0 = Np \sim m$. Hence, $\mathbb{P}(Y > 2m) < e^{-cm} < e^{-\Omega(\lambda_0)}$ (for instance by Chernoff's bound). It follows that

$$\mathbb{P}(X = 0) \sim \sum_{j \leq 2m} \mathbb{P}(\mathcal{C}_{j\delta_0}). \tag{4.6}$$

Using the definition of ρ, Proposition 2.6 and Lemma 2.5, and then (4.2) and (4.3), we have

$$\begin{aligned}
\rho(j\delta_0, \delta_0) &= \frac{\lambda_0}{j+1}\gamma(j\delta_0, 0) \\
&= \frac{\lambda_0 \overline{\gamma}_0(j\delta_0)}{j+1}\big(1 + o(\lambda_0^{-1})\big) \\
&= \frac{\lambda_0 \xi}{j+1}\big(1 + o(\lambda_0^{-1}) + O(x(m-j)/\lambda_0)\big) \\
&= \frac{m}{j+1}\exp\big(o(\lambda_0^{-1}) + O(x(m-j)/\lambda_0)\big)
\end{aligned}$$

by (4.4) and (1.1). Hence (4.6) gives

$$
\begin{aligned}
\frac{\mathbb{P}(X=0)}{\mathbb{P}(Y=m, X=0)} &= \frac{\mathbb{P}(X=0)}{\mathbb{P}(\mathcal{C}_{m\delta_0})} \\
&\sim \sum_{j\le 2m} \frac{\mathbb{P}(\mathcal{C}_{j\delta_0})}{\mathbb{P}(\mathcal{C}_{m\delta_0})} \\
&= \sum_{j\le 2m} \rho(m\delta_0, (j-m)\delta_0) \\
&= \sum_{j=m}^{2m} \prod_{i=m}^{j-1} \rho(i\delta_0, \delta_0) + \sum_{j=0}^{m} \prod_{i=j}^{m-1} \rho(i\delta_0, \delta_0)^{-1} \\
&= \frac{m!}{m^m} \sum_{j=0}^{2m} \frac{m^j}{j!} \exp\big(o((m-j)/\lambda_0) + O(x(m-j)^2/\lambda_0)\big) \\
&\sim \frac{m!}{m^m} \sum_{j=0}^{2m} \frac{m^j}{j!} \\
&\sim \frac{m! e^m}{m^m} \sim \sqrt{2\pi m}.
\end{aligned}
$$

In the third-last line, the main terms of the summation have $|m-j| \approx \sqrt{m} \sim \sqrt{\lambda_0}$, for which the error terms are $o(1)$ as $x \to 0$. The remaining terms are insignificant since the absolute value of the jth term in the sum is $\frac{m^m}{m!}\exp(-\Omega(m-j)^2/\lambda_0)$, which dominates the error term. The last line uses Stirling's formula.

Taking the multiplicative inverse of the previous asymptotic formula produces

$$\mathbb{P}(Y=m \mid X=0) \sim \frac{1}{\sqrt{2\pi m}}. \tag{4.7}$$

For the other factors in (4.1), first recall that ξ comes ultimately as a truncation of the power series ξ_0, in n and p (here $t=0$) in Corollary 2.7. Thus, we can use (4.4) and (4.5) to expand p as a power series in n and d. Specifically, we obtain $p = d\tilde{J}_1\big(1 + o(\lambda_0^{-1})\big)$ where $\tilde{J}_1$ is the truncation of a power series J_1 in n and d to significant terms. Here J_1 is independent of κ, being the termwise limit of the power series obtained for κ as $\kappa \downarrow \chi$ (which represents increasing p). This can be substituted into the polynomial obtained by truncating the power series for $\log \mathbb{P}(X=0)$ obtained from the $\mathcal{G}(n,p)$ case of Theorem 1.1, at an appropriate level, to express $\log \mathbb{P}(X=0)$ as $\tilde{J}_2 + o(1)$ where $\tilde{J}_2$ is a truncation of a power series J_2 in n and d, with J_2 independent of κ. Similarly, $\mathbb{P}(Y=m)$ is simply the binomial probability which can be estimated as follows. For $N = \binom{n}{2}$ and rewriting $p = d(1+\epsilon_0) = m(1+\epsilon_0)/N$ where $\epsilon_0 = O(x + \lambda_0^{-1})$ by (4.5), we have (using Stirling's formula)

$$
\begin{aligned}
\mathbb{P}(Y=m) &= \mathbb{P}(Y=dN) = \binom{N}{m} p^{dN} (1-p)^{N(1-d)} \\
&\sim \frac{1}{\sqrt{2\pi m}} \left((1+\epsilon_0)^d \left(\frac{1-d(1+\epsilon_0)}{1-d} \right)^{1-d} \right)^N
\end{aligned}
$$

using $m = o(N)$ (which follows since $p = o(n^{-\epsilon})$. Then from (4.1) and (4.7), we have

$$\mathbb{P}(X = 0 \mid Y = m) \sim \exp(\tilde{J}_2) \left((1+\epsilon_0)^d \left(1 - \frac{d\epsilon_0}{1-d} \right)^{1-d} \right)^{-N}. \tag{4.8}$$

The obvious expansion gives a power series in n and d, and due to the way we unified the theorem statements, we must replace each "d" in this expression by "p" to obtain the power series $F(G_0)$ for the $\mathcal{G}(n,m)$ case of the theorem. To verify that F has the required properties, we note that $1 + \epsilon_0 = \tilde{J}_1 + o(\lambda_0^{-1})$, where $\tilde{J}_1$ was derived from the κ-free J_1 as above.

The positivity of the exponents i_ℓ and j_ℓ follows by arguing as in the proof of the $\mathcal{G}(n,p)$ case. ∎

Acknowledgment. The authors are thankful for the careful checking of a referee who provided a number of corrections. One of these *almost* altered the final result of Theorem 1.3. The authors also thank Frank Mousset, Kosta Panagiotou and Wojciech Somotij who pointed out a term that was mistakenly dropped in the final calculations for Theorem 1.3, in an earlier version of this paper.

References

[1] B. Bollobás, Random graphs, In *Combinatorics*, Proceedings (Swansea, 1981), pp. 80–102, London Math. Soc. Lecture Note Ser. 52, Cambridge Univ. Press, Cambridge, 1981.

[2] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, pp. 1927. Atti dei Convegni Lincei, No. 17, Accad. Naz. Lincei, Rome, 1976.

[3] A. Frieze, On small subgraphs of random graphs. In *Random Graphs, Volume 2*, Wiley, New Tork, (1992), 67–90.

[4] S. Janson, T. Łuczak and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, in: M. Karoński, J. Jaworski, A. Ruciński, eds., Random Graphs '87 (Wiley), pp. 73–87, 1990.

[5] S. Janson, T. Łuczak and A. Ruciński, *Random graphs.* Wiley, New York, 2000.

[6] T. Łuczak, On triangle-free random graphs, *Random Structures & Algorithms* **16** (2000), 260–276.

[7] D. Osthus, H.J. Prömel and A. Taraz, For which densities are random triangle-free graphs almost surely bipartite? *Paul Erdős and his mathematics (Budapest, 1999). Combinatorica* **23**, 105–150.

[8] H.J. Prömel and A. Steger [8] Counting H-free graphs. *Discrete Math.*, **154** (1996), 311–315.

[9] H.J. Prömel and A. Steger, On the asymptotic structure of sparse triangle free graphs, *J. Graph Theory*, **21** (1996), 137–151.

[10] A. Ruciński, When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** (1988), 1–10.

[11] N.C. Wormald, The perturbation method and triangle-free random graphs, *Random Structures & Algorithms*, **9** (1996), 253–270.

Appendix A Calculations for triangle-free graphs

Proof of Theorem 1.2

Section 3 shows that an asymptotic formula for the probability a subgraph G_0 is not present in $\mathcal{G}(n,p)$ exists, but it does not state the formula explicitly. Nevertheless, the proof fully prescribes a method of calculating the formula for any particular case. At its heart, the proof uses Corollary 2.7, in which the power series $\xi_{t,\epsilon}$ are not stated explicitly. To obtain a formula in practice, these must be determined to a required accuracy, along with the quantities $c(u,t,h)$ defined in (2.1). In this section we demonstrate how the necessary calculations are performed in the case when G_0 is a triangle.

We let $G_0 = K_3$, the complete graph on 3 vertices, and proceed to estimate the probability that $\mathcal{G}(n,p)$ contains no triangles in the case that $p < n^{-7/11-\epsilon}$. (This constraint will be relaxed to $p = o(n^{-7/11})$ at the end.) It is easy to check that that (1.2) determines $\chi = \frac{1}{2}$ when $G_0 = K_3$. If we make the restriction $p = n^{-\kappa+o(1)}$ with $\kappa > 7/11$, then there are then 10 possible cluster types possible in $\mathcal{S}$ according to (2.16). We thus have $\mathcal{S} = \{1, 2, \ldots, 10\}$ as depicted in Figure 1.

All these types are present in $\mathcal{S}$ when κ is at most $2/3$. All other cluster types have expected number tending to 0 as $\kappa > 7/11$, and are therefore are not in $\mathcal{S}$. Recall that the poset ordering $\prec$ on $\mathcal{S}$ is not necessarily a linear ordering; for example, the types $\{5,6,7,8,9,10\}$ are all maximal, and therefore not comparable. The ordering $\prec$ is extended to the usual real linear ordering on $\mathcal{S}$ denoted by $<$.

The first step is to calculate λ_t for $t \in \mathcal{S}$. In accordance with (2.4) and (2.5), we obtain the λ_t as in Table 1.

$\lambda_1 = \frac{1}{6}[n]_3 p^3$	$\lambda_2 = \frac{1}{4}[n]_4 p^5$	$\lambda_3 = \frac{1}{2}[n]_5 p^7$	$\lambda_4 = \frac{1}{12}[n]_5 p^7$	$\lambda_5 = \frac{1}{2}[n]_6 p^9$
$\lambda_6 = \frac{1}{6}[n]_6 p^9$	$\lambda_7 = \frac{1}{2}[n]_6 p^9$	$\lambda_8 = \frac{1}{2}[n]_6 p^9$	$\lambda_9 = \frac{1}{48}[n]_6 p^9$	$\lambda_{10} = \frac{1}{24}[n]_4 p^6$

Table 1: Expected numbers of small clusters.

Our next task is to find the polynomial $\xi_{t,\epsilon}$ of Corollary 2.7 for all $t \in \mathcal{S}$. For this, the proof of the corollary describes an iterative scheme to compute the $F_t^{(r)}$ and hence $\overline{\gamma}_t^{(r)}$.

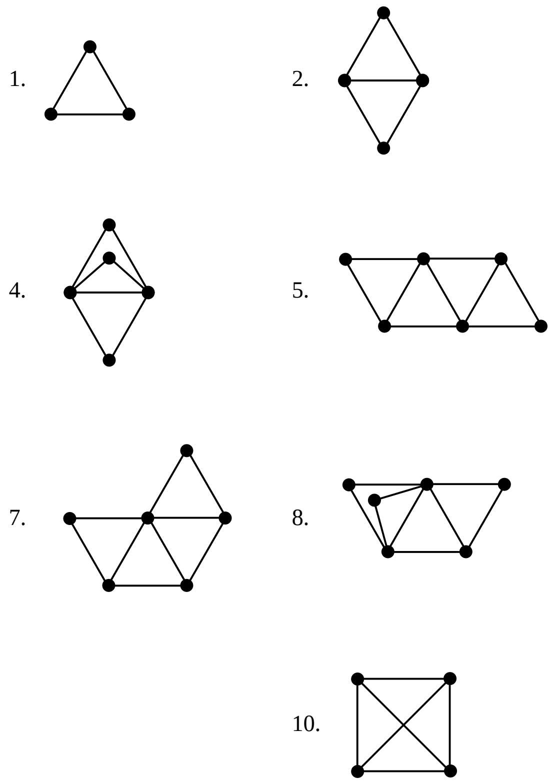

Figure 1: *Ten types of cluster*

We can drop all terms that would yield coefficients of variables g_{t_i} that are $O(n^{-\overline{\epsilon}}/\lambda_t)$ for some $\overline{\epsilon} > 0$. This is because in $\xi_{t,\epsilon}(f)$, each g_{t_i} is assigned a value that is $n^{o(1)}$, and hence the dropped terms are subsumed into the error term in (2.57) when $\overline{\epsilon}$ is sufficiently small (recalling that $\overline{\gamma}_t(f) \sim 1$ by Lemma 2.5). For similar reasons, we can drop any $O(p^2)$ term in the expansion of $c(t, t, \mathbf{0})$ at the front of (2.59), as $p^2\lambda_t = O(p^2\lambda_1) = O(p^5n^3) = o(n^{-2/11})$. Note that $c(1, 1, \mathbf{0}) = 1 - p^3$ since it is simply the probability that three vertices do not form a triangle. Hence, we can treat $c(1, 1, \mathbf{0})$ as 1. A similar argument applies to $c(t, t, \mathbf{0})$ for all other $t \in \mathcal{S}$.

Moving on to the quantities $c(u, t, h)$ inside the summation in (2.59), for any non-zero $h \in \mathcal{F}$, clearly $c(t, t, h) = O(p^3)$, and so these terms can be ignored completely for the same reason, for all t.

For the other terms in the summation, we only need to compute $c(u, t, h)$ to $O(n^{-\overline{\epsilon}}/\lambda_u)$. For $u = 2$, note that $n^{-\overline{\epsilon}}/n^4p^5 = \Omega(p^{10/7})$ for sufficiently small $\overline{\epsilon}$ since $n < p^{-11/7}$. Thus, we may drop p^2 terms in $c(2, t, h)$. First consider $c(2, 1, \mathbf{0})$. In (2.1), J is a cluster of type 2, i.e. (the edge set of) two triangles with a common edge. Q corresponds to one of the two triangles of J (so there are two choices for Q). There are four cases for H, as it must contain $J \setminus Q$

but no triangles. Letting $q = 1 - p$, we get

$$c(2, 1, \mathbf{0}) = 2(q^3 + 2pq^2 + p^2q) = 2(1 - p).$$

The other cases of $c(2, t, h)$ can be computed similarly, and only $h = \delta_1$ is significant (i.e. not $O(p^2)$). Similarly, for $u = 3$, $1/n^5p^7 = \Omega(p^{6/7})$ and we may drop the $O(p)$ terms. The same clearly holds for all $u > 3$ as well. In this way, we obtain all significant terms of $c(u, t, h)$ for $u > t$, as shown in Table 2. In computing these, note that h is quite restrictive. For instance, for $c(3, 1, \mathbf{0})$, the deletion of Q from J must leave no triangles, and there is only one such choice for Q.

u	t	h	$c(u,t,h)$	cofactor
2	1	$\mathbf{0}$	$2(1-p)$	$\frac{3}{2}np^2 \cdot \overline{\gamma}_2$
2	1	δ_1	$2p$	$\frac{3}{2}np^2 \cdot \overline{\gamma}_2 g_1 \overline{\gamma}_1^{-1}$
3	1	$\mathbf{0}$	1	$3n^2p^4 \cdot \overline{\gamma}_3$
3	1	δ_1	2	$3n^2p^4 \cdot \overline{\gamma}_3 g_1 \overline{\gamma}_1^{-1}$
4	1	$\mathbf{0}$	3	$\frac{1}{2}n^2p^4 \cdot \overline{\gamma}_4$
5	1	δ_1	2	$3n^3p^6 \cdot g_1$
5	1	δ_2	2	$3n^3p^6 \cdot g_2$
6	1	$\mathbf{0}$	1	n^3p^6
6	1	$2\delta_1$	3	$n^3p^6 \cdot g_1^2$
7	1	δ_1	2	$3n^3p^6 \cdot g_1$
7	1	δ_2	2	$3n^3p^6 \cdot g_2$
8	1	$\mathbf{0}$	1	$3n^3p^6$
8	1	δ_1	2	$3n^3p^6 \cdot g_1$
8	1	δ_2	1	$3n^3p^6 \cdot g_2$
9	1	$\mathbf{0}$	4	$\frac{1}{8}n^3p^6$
10	1	$\mathbf{0}$	4	$\frac{1}{4}np^3$

3	2	$\mathbf{0}$	2	$2np^2 \cdot \overline{\gamma}_3$
4	2	$\mathbf{0}$	3	$\frac{1}{3}np^2 \cdot \overline{\gamma}_4$
5	2	$\mathbf{0}$	1	$2n^2p^4$
5	2	δ_1	2	$2n^2p^4 \cdot g_1$
6	2	$\mathbf{0}$	3	$\frac{2}{3}n^2p^4$
7	2	$\mathbf{0}$	1	$2n^2p^4$
7	2	δ_1	2	$2n^2p^4 \cdot g_1$
8	2	$\mathbf{0}$	3	$2n^2p^4$
8	2	δ_1	1	$2n^2p^4 \cdot g_1$
9	2	$\mathbf{0}$	6	$\frac{1}{12}n^2p^4$
10	2	$\mathbf{0}$	6	$\frac{1}{6}p$
5	3	$\mathbf{0}$	2	np^2
6	3	$\mathbf{0}$	3	$\frac{1}{3}np^2$
7	3	$\mathbf{0}$	2	np^2
8	3	$\mathbf{0}$	2	np^2
8	4	$\mathbf{0}$	1	$6np^2$
9	4	$\mathbf{0}$	4	$\frac{1}{4}np^2$

Table 2: Significant contributions to (2.59)

The "cofactor" column of Table 2 shows the significant contribution to those terms in F_t from

$$\frac{\lambda_u}{\lambda_t}\overline{\gamma}_u \prod_{i=1}^{k} \frac{g_{t_i}}{\overline{\gamma}_{t_i}}.$$

Here, and in the rest of the calculation, we assume $\overline{\epsilon} > 0$ is as small as we like, and any terms that are $O(n^{-\overline{\epsilon}}/\lambda_t)$ are dropped. In each case, λ_u/λ_t is the first item in the column, with any others (that are not equal to 1) appearing after "$\cdot$". In each case only the leading term of λ_u/λ_t turns out to be significant, since the correction terms are $O(1/n)$ and $\lambda_u/n = O(n^3p^5)$ for $u \geq 2$. Any other factors which appear to be missing have simply been replaced by 1, with the following justification. In the initial iteration, for computing $F_t^{(1)}$ we have all $\overline{\gamma}_v$ equal to 1, and by induction, thereafter they are $1+O(np^2)$ (if we treat each g_{t_i} as 1). In the end each g_{t_i} is substituted by something that is $n^{o(1)}$. Hence we may set any $\overline{\gamma}_u$ or $\overline{\gamma}_{t_i}$ equal to 1 in all iterations for all $u \geq 5$, since then $\lambda_u np^2 = O(n^7p^{11} + n^5p^8) = O(n^{-\overline{\epsilon}}/\lambda_t)$. Of course there are no contributions from $t \geq 5$ since all such t are maximal in $\mathcal{S}$, and $c(u,t,h) = 0$ unless $t \prec u$ (and we have already dealt with the case $u = t$).

The significant terms of (2.59) are now deduced to be

$$\begin{aligned} F_1 &= -np^2\left(3(1-p)\overline{\gamma}_2 + 3p\overline{\gamma}_2 g_1/\overline{\gamma}_1\right) - n^2p^4\left(3\overline{\gamma}_3 + 6\overline{\gamma}_3 g_1/\overline{\gamma}_1 + \frac{3}{2}\overline{\gamma}_4\right) \\ &\quad -n^3p^6\left(18g_1 + 15g_2 + 3g_1^2 + \frac{9}{2}\right) - np^3, \\ F_2 &= -np^2\left(4\overline{\gamma}_3 + \overline{\gamma}_4\right) - n^2p^4\left(10g_1 + 25/2\right) - p, \\ F_3 &= -7np^2, \\ F_4 &= -7np^2, \\ F_t &= 0 \quad (t \geq 5). \end{aligned}$$

Write $y = np^2$ and solve (2.59) iteratively as described after that equation. It may help to note that any terms of order yp^2, y^2p or y^4 can be dropped. After three iterations (actually the expressions don't change after the second update), the error is of order $x^4 = \max\{y^4, p^4\}$ by (2.11), which is neglibible for each t. This gives $\xi_{t,\epsilon} = 1 + F_t^{(4)}$ given as follows.

$$\xi_{1,\epsilon} = 1 - 3y + 5py - 3g_1py + \frac{21}{2}y^2 - 6g_1y^2 - \frac{81}{2}y^3 + 36g_1y^3 - 3g_1^2y^3 - 15g_2y^3, \quad \text{(A.1)}$$

$$\xi_{2,\epsilon} = 1 - p - 5y + \frac{45}{2}y^2 - 10g_1y^2, \quad \text{(A.2)}$$

$$\xi_{3,\epsilon} = 1 - 7y, \quad \text{(A.3)}$$

$$\xi_{4,\epsilon} = 1 - 7y. \quad \text{(A.4)}$$

$$\xi_{t,\epsilon} = 1 \quad (t \geq 5). \quad \text{(A.5)}$$

We will evaluate the expressions given in Section 3 with $7/11 < \kappa < 2/3$, so that $\mathcal{S}_1 = [10]$ and $\mathcal{S}_0 = \emptyset$ (and actually $y = x$ as per (2.10)). The ultimate result will then be valid for all values of $\kappa > 7/11$ by (3.31). We also fix ϵ in the range $0 < \epsilon < 7/11 - \chi = 3/22$. With κ and ϵ in these ranges, the $\xi_{t,\epsilon}$ are given by the expressions (A.1)–(A.5).

The recursive definition (3.5) of S_t for $t \leq 10$ starts with $S_{10} = 1$ and hence, in (3.9), $P_{10,\kappa} = 0$. Hence (just before (3.14)) $A_{10,\kappa}^{(1)} = 1$. Of course there are options in choosing A's since they are only determined up to an error term; we use the natural choices.

The next step is to determine S_9 and $P_{9,\kappa}$. From (3.15), $A_{10,\kappa}^{(2)} = 1$. Now (3.16) implies $j^* = \lambda_{10}$ and hence from (3.17) $A_{10,\kappa}^{(3)} = 1$. It is now easy to check that $A_{10,\kappa}^{(4)} = A_{10,\kappa}^{(5)} = 1$ at (3.26), and then similarly $A_{10,\kappa}^{(6)} = A_{10,\kappa}^{(7)} = 1$. (Much more detail in the steps here is provided in the less trivial case when $t = 1$ below.) Finally, we conclude that, at (3.29), $S_9(j_1, j_2, \ldots, j_9) \sim e^{\lambda_{10}}$ and then $P_{9,\kappa} = \lambda_{10}$. In the same way one can show that $S_t \sim \exp\left(\sum_{u=t+1}^{10} \lambda_u\right)$ for $t = 8, 7, 6, 5, 4$. In particular we have $S_4 \sim \exp\left(\sum_{u=5}^{10} \lambda_u\right)$ and $P_{4,\kappa} = \sum_{u=5}^{10} \lambda_u$.

Next consider S_3 and $P_{3,\kappa}$. We have that $P_{4,\kappa}(\zeta)$ is independent of ζ, and so $A^{(1)}_{4,\kappa} = 1$. The ratio in (3.12) is $T_j/T_{j-1} = (1-7y)\lambda_4(1+O(\eta))/j$, so $A^{(2)}_{4,\kappa} = 1-7y$, $j^* = (1-7y)\lambda_4$ and $A^{(3)}_{4,\kappa} = 1-7y$. Moreover, $\log(\lambda_4/j^*) = -\log(1-7y) = 7y+O(y^2) = 7y+O(\lambda_4^{-1})$, so $A^{(4)}_{4,\kappa} = 1+7y$ and

$$\left(\frac{e\lambda_4}{j^*}\right)^{j*} = \left(\frac{e}{1-7y}\right)^{\lambda_4(1-7y)} = e^{\lambda_4+o(1)}$$

from which we deduce $A^{(5)}_{4,\kappa} = 1$. Now, $\sum_{i=0}^{\tilde{j}-1}\log\xi_{4,\epsilon} = \tilde{j}\log(1-7y) = -7y(1-7y)\lambda_4 + o(1) = -7y\lambda_4 + o(1)$ implies that $A^{(6)}_{4,\kappa} = 1-7yg_4$ and $A^{(7)}_{4,\kappa} = 1-7y$. Because $P_{4,\kappa}$ does not depend on g_4, $\tilde{P}_{4,\kappa} = P_{4,\kappa} = \sum_{u=5}^{10}\lambda_u$. Finally, we have

$$S_3(j_1,j_2,j_3) \sim \exp\left(\sum_{u=5}^{10}\lambda_u + (1-7y)\lambda_4\right)$$

and $P_{3,\kappa} = \tilde{P}_{4,\kappa} + \lambda_4 A^{(7)}_{4,\kappa} = \sum_{u=5}^{10}\lambda_u + (1-7y)\lambda_4$. Similar analyses which we omit show that

$$S_2(j_1,j_2) \sim \exp\left((1-7y)(\lambda_3+\lambda_4) + \sum_{u=5}^{10}\lambda_u\right).$$

Next, note that $A^{(1)}_{2,\kappa} = 1$, $A^{(2)}_{2,\kappa} = A^{(3)}_{2,\kappa} = \xi_{2,\epsilon}$, $j^* = \lambda_2\xi_{2,\epsilon}$, $A^{(4)}_{2,\kappa}$ is a truncation of the expansion of $1-\log\xi_{2,\epsilon}$, $A^{(5)}_{2,\kappa} = 1-\frac{25}{2}y^2$, $A^{(6)}_{2,\kappa}$ is the truncation of $1+g_2\log\xi_{2,\epsilon}$, which is $1+g_2(-p-5y+(10-10g_1)y^2)$, and $A^{(7)}_{2,\kappa} = 1-p-5y+\frac{45}{2}y^2-10g_1y^2$. Eventually $S_1(j_1) \sim \exp\left(P_{1,\kappa}(\zeta(j_1))\right)$ where

$$P_{1,\kappa} = \left(1-p-5y+\frac{45}{2}y^2-10g_1y^2\right)\lambda_2 + (1-7y)(\lambda_3+\lambda_4) + \sum_{u=5}^{10}\lambda_u. \qquad \text{(A.6)}$$

The final step of the induction is a little more involved. We have

$$\begin{aligned}\exp\left(P_{1,\kappa}(\zeta(j)) - P_{1,\kappa}(\zeta(j-1))\right) &= \exp\left(\frac{-10y^2\lambda_2}{\lambda_1}\right)\\ &= \exp\left(-15y^3 + O(y^3/n)\right)\end{aligned}$$

and hence $A^{(1)}_{1,\kappa} = 1-15y^3$. For (3.15) we set $g_2 = 0$ to get $\tilde{\xi}_{1,\epsilon}$ and obtain

$$A^{(2)}_{1,\kappa} = 1 + c_1 + c_2g_1 + c_3g_1^2.$$

where

$$\begin{aligned}c_1 &= -3y + 5py + \frac{21}{2}y^2 - \frac{81}{2}y^3 - 15y^3\\ &= -3y + 5py + \frac{21}{2}y^2 - \frac{111}{2}y^3,\\ c_2 &= -3py - 6y^2 + 36y^3\\ c_3 &= -3y^3,\end{aligned}$$

The equation (3.16) for j^* becomes

$$\frac{j^*}{\lambda_1} = 1 + c_1 + c_2\left(\frac{j^*}{\lambda_1}\right) + c_3\left(\frac{j^*}{\lambda_1}\right)^2. \tag{A.7}$$

Since the c_i are $O(y^i)$, $j^* \sim \lambda_1$ and $\lambda_1 y^4 = o(1)$, we find iteratively that $j^* = (1 + c_1 + c_2 + c_1c_2 + c_3)\lambda_1 + o(1)$, and so

$$A^{(3)}_{1,\kappa} = 1 + c_1 + c_2 + c_1c_2 + c_3. \tag{A.8}$$

Expanding $1 - \log A^{(3)}_{1,\kappa}$ gives

$$A^{(4)}_{1,\kappa} = 1 - c_1 - c_2 - c_3 + c_1^2/2 - c_1^3/3$$

and then truncating $A^{(3)}_{1,\kappa} \cdot A^{(4)}_{1,\kappa}$ gives

$$A^{(5)}_{1,\kappa} = 1 - c_1c_2 - c_1^2/2 + c_1^3/6.$$

Next, referring to (A.9) and writing $\tilde{c}_1 = c_1 + 15y^3$,

$$\begin{aligned}
\sum_{i=0}^{\tilde{j}-1} \log \xi_{1,\epsilon}(f_{1,i}) &= \sum_{i=0}^{\tilde{j}-1} \log\left(1 + \tilde{c}_1 + \frac{c_2 i}{\lambda_1} + \frac{c_3 i^2}{\lambda_1^2}\right) \\
&= o(1) + \sum_{i=0}^{\tilde{j}-1}\left(\tilde{c}_1 - \frac{\tilde{c}_1^2}{2} + \frac{\tilde{c}_1^3}{3} + \frac{(c_2 - \tilde{c}_1c_2)i}{\lambda_1} + \frac{c_3 i^2}{\lambda_1^2}\right) \\
&= \left[\left(\tilde{c}_1 - \frac{1}{2}\tilde{c}_1^2 + \frac{1}{3}\tilde{c}_1^3\right)\frac{\tilde{j}}{\lambda_1} + \left(\frac{1}{2}c_2 - \frac{1}{2}\tilde{c}_1c_2\right)\left(\frac{\tilde{j}}{\lambda_1}\right)^2 + \frac{1}{3}c_3\left(\frac{\tilde{j}}{\lambda_1}\right)^3\right]\lambda_1 + o(1),
\end{aligned}$$

so

$$A^{(6)}_{1,\kappa} = 1 + \left(\tilde{c}_1 - \frac{1}{2}\tilde{c}_1^2 + \frac{1}{3}\tilde{c}_1^3\right)g_1 + \left(\frac{1}{2}c_2 - \frac{1}{2}\tilde{c}_1c_2\right)g_1^2 + \frac{1}{3}c_3g_1^3. \tag{A.9}$$

Substituting (A.8) for g_1 in (A.9), dropping insignificant terms, and adding $-1 + A^{(5)}_{t,\kappa}$, we obtain after some algebra

$$\begin{aligned}
A^{(7)}_{1,\kappa} &= 1 + \tilde{c}_1 + \tilde{c}_1c_1 - \frac{1}{2}\tilde{c}_1^2 + \frac{1}{2}c_2 + \frac{1}{2}\tilde{c}_1c_2 + \frac{1}{3}\tilde{c}_1^3 - \frac{1}{2}\tilde{c}_1^2c_1 + \frac{1}{3}c_3 - \frac{1}{2}c_1^2 + \frac{1}{6}c_1^3 + o(\lambda_1^{-1}) \\
&= 1 - 3y + \frac{7}{2}py + \frac{15}{2}y^2 - \frac{29}{2}y^3.
\end{aligned}$$

Since $j^* = \lambda_1(1 + O(y))$, changing from $g_1 = \tilde{j}/\lambda_1$ or $g_1 = j^*/\lambda_1$ to $g_1 = 1$ in (A.6) induces a change to $P_{1,\kappa}$ of order $O((j^* - \lambda_1)y^2\lambda_2/\lambda_1) = O(y^3\lambda_2) = o(1)$ and therefore $\tilde{P}_{1,\kappa} = P_{1,\kappa}|_{g_1=1}$. Finally,

$$\begin{aligned}
P_{0,\kappa} &= \tilde{P}_{1,\kappa}(1) + \lambda_1 A^{(7)}_{1,\kappa} \\
&= \sum_{u=5}^{10}\lambda_u + (1 - 7y)\lambda_4 + (1 - 7y)\lambda_3 + \left[1 - p - 5y + \frac{25}{2}y^2\right]\lambda_2 \\
&\quad + \left[1 - 3y + \frac{7}{2}py + \frac{15}{2}y^2 - \frac{29}{2}y^3\right]\lambda_1 + o(1).
\end{aligned}$$

The remaining task is to plug in the expansions for the λ_t's given in Table 1, simplify, and apply (3.30). Since $p = o(n^{-7/11})$ we approximate λ_1 by $\frac{1}{6}n^3p^3 - \frac{1}{2}n^2p^3$, whilst for λ_t, $t \geq 2$ only the first order term is important: $\lambda_2 \sim \frac{1}{4}n^4p^5$ etc. This determines $P_{0,\kappa}$ and hence the coefficients in the statement of Theorem 1.1, resulting in the statement of Theorem 1.2 for $p < n^{-7/11-\epsilon}$. To relax this to $p = o(n^{-7/11})$, we only need to note that, from this conclusion, all other terms in the series F in Theorem 1.1 must have $i_\ell/j_\ell \leq 7/11$. Such terms tend to 0 for $p = o(n^{-7/11})$, and the theorem follows. ∎

Proof of Theorem 1.3

Here we extend the previous proof to obtain the probablity that $\mathcal{G}(n,m)$ contains no copies of K_3. We just need to find $\tilde{J}_2$ and ϵ_0 in (4.8). We require additive error $o(1)$ for the former. For the latter, we note that expanding the logarithm of the large factor in (4.8) gives $-\frac{1}{2}d\epsilon_0^2 + \cdots$. Since this factor is raised to the power $-N$ and $\epsilon_0 = O(x)$, this indicates that the absolute error required for ϵ_0 is simply $o(1/dNx) = o(1/n^3d^3)$. (This assumes $x = np^2$. In the other case, when $x = p$, we have $np \leq 1$ and the main result follows from known results easily.) Since $p = d(1+\epsilon_0) = d/\xi$ as per (4.4), we will need to find the asymptotic expansion of $\xi = \xi_{0,\epsilon}(n,p,\tilde{\mathbf{g}})$ to relative error $o(1/n^3d^3)$.

The existing terms in Table 2 have the accuracy desired for the present computation. Table 3 is essentially an extension of Table 2, showing significant contributions to F_t from (2.59) as needed to calculate $\overline{\gamma}_0(n,p,\tilde{\mathbf{g}})$, under the same assumption that $p = O(n^{-7/11-\epsilon})$. Note that F_1 and F_2 need to be recomputed in this new clustering as the expression for $\overline{\gamma}_0$ contains $\overline{\gamma}_1$ and $\overline{\gamma}_2$.

Since

$$\tilde{g}_0 = m/\lambda_0 = mp^{-1}\binom{n}{2}^{-1} = d/p = \xi = \xi_{0,\epsilon}(n,p,\tilde{\mathbf{g}}) = \overline{\gamma}_0(m\delta_0) + o(\lambda_0^{-1}) \qquad \text{(A.10)}$$

by (4.2), it is straightforward to see that the factors $\tilde{g}_0/\overline{\gamma}_0$ can at this point be replaced by 1. Strictly this needs to be justified in the context of the recursive computation of ξ in Corollary 2.7, and this can be seen in a straightforward way by going back to the original equations in Proposition 2.1 with the altered equations and observe that the same argument as in Section 2 applies to these altered equations, resulting in the modified definition of F_t in (2.59); alternatively, one could include the factors explicitly and watch them turn naturally into 1. Note that terms like $c(1,0,\delta_1)$ cannot affect this computation since they contain a factor g_t for $t > 0$, and to evaluate ξ we must set such g_t equal to 0.

The denominator of (2.59) is $c(0,0,\mathbf{0}) = 1-p$ in the case of $t = 0$. As with the $\mathcal{G}(n,p)$ calculation, we can ignore certain terms in the product of $c(u,t,h)$ with its cofactor. In the case of $t = 0$, as explained above for calculating ϵ_0 or ξ we can ignore any terms that are $O(n^{-\bar{\epsilon}}/n^3p^3)$. Note that $\lambda_0 = n(n-1)p/2$. Since $\overline{\gamma}_1$ only arises in terms with a cofactor that is $O(np^2)$, we ignore terms in its expression that are $O(n^{-\bar{\epsilon}}/n^4p^5)$ such as p^2. For similar reasons, terms in $\overline{\gamma}_2$ of order $O(n^{-\bar{\epsilon}}/n^5p^7)$ are ignored.

Note that we now have $c(1,1,\mathbf{0}) = (1-p)^3 \approx 1-3p$ since the only relevant possibility for H in (2.1) in this case is the empty set. Plugging the values in Table 3 into (2.44) or (2.59)

u	t	h	$c(u,t,h)$	cofactor
0	0	δ_0	p	$\overline{\gamma}_0$
1	0	$2\delta_0$	$3(1-p)$	$\frac{1}{3}np^2 \cdot \overline{\gamma}_1$
1	1	δ_0	$3p$	$\overline{\gamma}_1$
2	0	$4\delta_0$	1	$\frac{1}{2}n^2p^4 \cdot \overline{\gamma}_2$
4	0	$6\delta_0$	1	$\frac{1}{6}n^3p^6$
2	1	$2\delta_0$	2	$\frac{3}{2}np^2 \cdot \overline{\gamma}_2$
3	1	$4\delta_0$	1	$3n^2p^4$
4	1	$4\delta_0$	3	$\frac{1}{2}n^2p^4$
3	2	$2\delta_0$	2	$2np^2$
4	2	$2\delta_0$	3	$\frac{1}{3}np^2$

Table 3: Significant contributions to (2.59)

gives the following partly truncated expressions for $\overline{\gamma}_0$, $\overline{\gamma}_1$ and $\overline{\gamma}_2$

$$\overline{\gamma}_0 = \frac{1}{1-p}\left(1 - p\overline{\gamma}_0 - (1-p)np^2\overline{\gamma}_1 - \frac{1}{2}n^2p^4\overline{\gamma}_2 - \frac{1}{6}n^3p^6\right),$$

$$\overline{\gamma}_1 = (1+3p)\left(1 - 3p\overline{\gamma}_1 - 3np^2\overline{\gamma}_2 - \frac{9}{2}n^2p^4\right),$$

and

$$\overline{\gamma}_2 = 1 - 5np^2.$$

Solving for $\overline{\gamma}_0$ gives

$$\overline{\gamma}_0 = 1 - np^2 + \frac{5}{2}n^2p^4 - \frac{49}{6}n^3p^6 + np^3 + O(n^2p^5)$$

to additive error $o(1/\lambda_0) = o(1/n^2p)$. By (A.10) we can use this expression for ξ and find $1/\xi \approx 1 + y - \frac{3}{2}y^2 + \frac{25}{6}y^3 - yp$ where $y = np^2$ and the terms of order y^4, y^2p and p^2 are omitted. Substituting $p = d/\xi$ into itself three times gives

$$p = (d + nd^3 - nd^4 + \frac{1}{2}n^2d^5 + \frac{1}{6}n^3d^7)(1 + o(1/n^3d^3)). \tag{A.11}$$

This lets us eliminate p from the formula for $\mathbb{P}(X=0)$ obtained from Theorem 1.2, which gives $\exp(\tilde{J}_2)$ asymptotically as required for (4.8). The other ingredient for that estimate is the value of ϵ_0, to additive error $o(1/n^3d^3)$, which is determined directly from (A.11). Then (4.8) gives the probability that $X = 0$ in $G \in \mathcal{G}(m,n)$, and simple computations give

$$\mathbb{P}(X = 0|Y = m) \sim \exp\left(-\frac{1}{6}n^3d^3 - \frac{1}{8}n^4d^6 + \frac{1}{2}n^2d^3\right).$$

For the same reasons as in the $\mathcal{G}(n,p)$ case, the validity extends to all $d = o(n^{-7/11})$. ∎