Received: 3 May 2021 | Revised: 30 April 2022 | Accepted: 3 August 2022 | Published on: 17 February 2023

DOI: 10.1002/rsa.21140

RESEARCH ARTICLE

WILEY

Clustering and percolation on superpositions of Bernoulli random graphs

Mindaugas Bloznelis[1] | **Lasse Leskelä**[2]

[1]Vilnius University, Faculty of Mathematics and Informatics, Vilnius, Lithuania

[2]School of Science, Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

Correspondence

Lasse Leskelä, School of Science, Department of Mathematics and Systems Analysis, Aalto University, Otakaari 1, Espoo 02015, Finland.
Email: lasse.leskela@aalto.fi

Abstract

A simple but powerful network model with n nodes and m partly overlapping layers is generated as an overlay of independent random graphs $G_1, \dots, G_m$ with variable sizes and densities. The model is parameterized by a joint distribution P_n of layer sizes and densities. When m grows linearly and $P_n \to P$ as $n \to \infty$, the model generates sparse random graphs with a rich statistical structure, admitting a nonvanishing clustering coefficient together with a limiting degree distribution and clustering spectrum with tunable power-law exponents. Remarkably, the model admits parameter regimes in which bond percolation exhibits two phase transitions: the first related to the emergence of a giant connected component, and the second to the appearance of gigantic single-layer components.

KEYWORDS

Overlapping communities, power law, clustering coefficient, random graph, intersection graph, complex network, bond percolation, site percolation, giant component

1 | INTRODUCTION

Applications in natural sciences, social sciences, and technology often deal with large networks of nodes linked by pairwise interactions which involve uncertainty due to noisy observations and missing data. Such uncertainties have been investigated using statistical models ranging from classical Bernoulli random graphs and uniform random graphs with given degree distributions to stochastic block models and more complex generative models involving various preferential attachment and rewiring mechanisms [1, 25, 30, 38, 46]. While succeeding to obtain a good fit for degree distributions and tractable percolation analysis, most earlier models fail to capture second-order effects related to clustering and transitivity. Random intersection graphs [5, 11, 18, 33, 41], spatial preferential attachment models [27–29], and hyperbolic random geometric graphs [13, 24, 34, 35] have been introduced to conduct percolation analysis on networks with nonvanishing transitivity and clustering properties.

Despite remarkable methodological advances, most sparse network models still appear somewhat rigid in what comes to modeling finer clustering properties, such as the *clustering spectrum* (degree-dependent local clustering coefficient) [3, 43, 47], which may significantly impact the percolation properties of the network [4, 19]. A decreasing clustering spectrum manifests the fact that *high-degree nodes tend to have sparser local neighborhoods than low-degree nodes*. Motivated by analyzing this phenomenon in a tractable quantitative framework, this article discusses a statistical network model generated as an overlay of mutually independent Bernoulli random graphs $G_1, \dots, G_m$ which can be interpreted as *layers* or *communities*. The layers have a variable size (number of nodes) and strength (link probability), and they may overlap each other. A key feature of the model is that the layer sizes and layer strengths are assumed to be correlated, which allows to model and analyze a rich class of networks with a tunable frequency of strong small communities and weak large communities.

1.1 | Main contributions

This article presents a rigorous mathematical analysis of clustering and percolation of the overlay graph model in the natural sparse limiting regime where the number of nodes n tends to infinity, the number of layers m is linear in the number of nodes, and the joint distribution P_n of layer sizes and layer strengths converges to a limiting distribution P. We derive exact formulas for the limiting degree distribution, clustering coefficient, clustering spectrum, and the largest component size in terms of cross-factorial moments and functional transforms of P. We also investigate the model under bond and site percolation, and characterize critical parameter values of the associated phase transitions.

The descriptive power of the model is illustrated by a detailed investigation of an instance where the layer size follows a power law, and the layer strength is a deterministic function of the layer size following another power law. This setting leads to a power-law degree distribution and a power-law clustering spectrum with tunable exponents in ranges $(1, \infty)$ and $[0,2]$, respectively. A special case in which layer strengths are inversely proportional to their sizes corresponds to layers of bounded average degree. In this natural parameter regime we discover a remarkable *double phase transition* phenomenon with two critical values: the first characterizing the emergence of a giant component in the overlay graph, and the second characterizing the emergence of gigantic components in layers covering a typical node.

Finally, we highlight that the modelling framework in this article covers *both deterministic and random layer types*. Our approach of characterizing the regularity of layer types using averaged empirical distributions allows both cases to be treated in a uniform manner.

1.2 | Related work

The overlay network model discussed in this article is naturally motivated and implicitly described by classical works in social networks [17, 22]. The explanatory power and wide applicability of the model in the context of social, collaboration, and information networks has been demonstrated in [48, 49] by experimental studies of a *community-affiliation graph*, which represents an instance of the present model where the node sets of layers are nonrandom or otherwise known to the observer. The superposition of Bernoulli random graphs considered here serves as a null model for sparse community-affiliation graphs.

The mathematical analysis in this article builds on earlier works on component evolution and clustering in inhomogeneous random graphs [14] and random intersection graphs [8, 9]. The special model instance with unit layer strengths reduces to the so-called *passive random intersection graph* [26], and as a byproduct, the present article also provides the first rigorous analysis of giant components in general passive random intersection graphs, extending [16, 37]. When layer strengths are constant but not necessarily one, clustering properties and subgraph densities of the model have been analyzed in [31, 32, 40], and the recovery of the layers in [21]. Another related work [45] (also part of [44]) on percolation in overlapping community networks assumes that layers are sampled from an arbitrary distribution on the space of finite connected graphs, and the layers are assigned to nodes via a bipartite configuration model. The restriction to connected layers and the use of a configuration model makes the model in [45] and its analysis fundamentally different from the present one, and limits its applicability by ruling out networks composed of weak communities.

Clustering spectra with power-law exponent 1 have been shown for random intersection graph models [7, 9] and spatial preferential attachment models [27, 36], and with a tunable power-law exponent in $[0, 1]$ for random intersection graphs [10, 12] and recently also for a hyperbolic random geometric graph model [24]. Furthermore, [43] discusses an inhomogeneous Bernoulli graph model where the clustering spectrum vanishes, but its normalized version displays evidence of a power-law behavior with exponent in range (0,2).

To the best of our knowledge, the present work is the first of its kind where a degree dependent clustering with a tunable power-law exponent in the extended range [0,2] is rigorously analyzed in terms of a simple statistical network model. This model admits a clear explanation of the values of power-law exponents, and introduces a new analytical framework for studying ordinary and double phase transitions in bond and site percolation on sparse networks of overlapping communities of variable size and strength.

1.3 | Outline

In the rest of the article, Section 2 presents model details and notations, and Section 3 the main results. Section 4 illustrates the main results in a power-law setting, and confirms the existence of double phase transition. The remaining Sections 5–10 are devoted to proofs, with technical details postponed to Appendix 11.

2 | MODEL DESCRIPTION

2.1 | Multilayer network

A multilayer network model with n nodes and m layers is defined by a list $\big((G_1, X_1, Q_1), \dots, (G_m, X_m, Q_m)\big)$ of mutually independent random variables with values in $\mathcal{G}_n \times \{0, \dots, n\} \times [0, 1]$, where $\mathcal{G}_n$ is the set of undirected graphs with node set contained in $\{1, \dots, n\}$. We assume that conditionally on (X_k, Q_k), the probability distribution of the vertex set $V(G_k)$ of G_k is uniform on the subsets of $\{1, \dots, n\}$ of size X_k, and conditionally on $(V(G_k), X_k, Q_k)$, each node pair of $V(G_k)$ is linked with probability Q_k, independently of other node pairs. Thus, G_k is a Bernoulli random graph on node set $V(G_k)$, with edge set denoted $E(G_k)$. The variables X_k, Q_k, and (X_k, Q_k) are called the *size*, *strength*, and *type* of layer k, respectively. Aggregation of layers produces an overlay random graph G defined by

$$V(G) = \{1, \dots, n\} \qquad \text{and} \qquad E(G) = \cup_{k=1}^{m} E(G_k). \tag{1}$$

This setting includes as special cases: (i) models with deterministic layer types, and (ii) models where the layer types are independent and identically distributed random variables.

2.2 | Large networks

A large network is analyzed by considering a sequence of network models $\big((G_{n,1}, X_{n,1}, Q_{n,1}), \ldots, (G_{n,m}, X_{n,m}, Q_{n,m})\big)$ indexed by the number of nodes $n = 1, 2, \ldots$ so that the number of layers $m = m_n$ tends to infinity as $n \to \infty$. The sequence of corresponding overlay random graphs is denoted $(G_{(n)})$. We shall focus on a sparse parameter regime where there exists a probability measure P on (Borel's σ-algebra $\mathcal{B}$ of) $\{0, 1, \ldots\} \times [0, 1]$ which approximates in sufficiently strong sense the averaged layer type distribution

$$P_n(B) = \frac{1}{m} \sum_{k=1}^{m} \mathbb{P}\left((X_{n,k}, Q_{n,k}) \in B\right), \qquad B \in \mathcal{B}. \tag{2}$$

In this fundamental regime, the network features are described by limiting formulas with rich expressive power captured by cross moments and tail characteristics of P.

2.3 | Notations

We denote $\mathbb{Z}_+ = \{0, 1, \ldots\}$, $(a)_+ = \max\{0, a\}$, and $(x)_s = x(x-1)\cdots(x-s+1)$. The indicator function of a condition A is denoted by $\mathbb{I}(A)$ or $\mathbb{I}_A$, whichever is more convenient. Sets of size x are called *x-sets*. Unordered pairs and triples are abbreviated as $ij = \{i, j\}$ and $ijk = \{i, j, k\}$. We write $\sum'_{ij}$ and $\sum'_{ijk}$ to indicate sums over ordered pairs and ordered triples with distinct elements. We write $a_n \ll b_n$ and $a_n = o(b_n)$ when $a_n/b_n \to 0$, $a_n \lesssim b_n$ and $a_n = O(b_n)$ when $limsup|a_n/b_n| < \infty$, and $a_n \sim b_n$ when $a_n/b_n \to 1$. For a sequence of bivariate random variables (α_n, β_n), we write $\alpha_n = o_P(\beta_n)$ whenever $\lim_{n\to\infty} \mathbb{P}(|\alpha_n| < \varepsilon|\beta_n|) = 1$ for any $\varepsilon > 0$; and $\alpha_n = O_P(\beta_n)$ if for every $\varepsilon > 0$ there exists a constant $c_\varepsilon > 0$ such that $\lim_{n\to\infty} \mathbb{P}(|\alpha_n| < c_\varepsilon|\beta_n|) > 1 - \varepsilon$. Notation $\alpha_n = o_P(\beta_n)$, $\alpha_n = O_P(\beta_n)$ extends to the case where the sequence β_n is deterministic (nonrandom).

A graph is a pair $G = (V, E)$ where E is a set of unordered pairs of elements of V. The degree and component of node i in graph G are denoted by $\deg_G(i)$ and $C_G(i)$, respectively. The transitive closure of graph G is defined as the graph $\overline{G}$ with $V(\overline{G}) = V(G)$ and $E(\overline{G}) = \{ij : i \in C_G(j), j \in V(G)\}$ consisting of unordered node pairs connected by a path in G.

The probability distribution of a random variable X is denoted by $\mathcal{L}(X)$. For probability measures, $d_{\mathrm{tv}}(f, g)$ denotes the total variation distance, $f * g$ the convolution, and $f_n \xrightarrow{w} f$ refers to weak convergence. Convergence in probability is denoted $\xrightarrow{\mathbb{P}}$. On countable spaces, the same letter is used for both a probability measure $f(A)$ and its density $f(t)$ with respect to the counting measure. The Dirac measure at x is denoted by δ_x. The densities of the binomial distribution $\mathrm{Bin}(x, q)$, with $x \in \mathbb{Z}_+$ and $q \in [0, 1]$, and the Poisson distribution $\mathrm{Poi}(\lambda)$ with $\lambda \geq 0$, are denoted by

$$\mathrm{Bin}(x, q)(t) = \binom{x}{t} (1-q)^{x-t} q^t, \qquad \mathrm{Poi}(\lambda)(t) = e^{-\lambda} \frac{\lambda^t}{t!},$$

with the convention that the densities are zero for t outside $\{0, \ldots, x\}$ and $\mathbb{Z}_+$, respectively. The Bernoulli distribution is denoted $\mathrm{Ber}(q)(t) = \mathrm{Bin}(1, q)(t)$. We also denote by

$$\mathrm{Bin}^+(x, q)(t) = \mathbb{P}\left(\deg_{\overline{H}_{x+1,q}}(1) = t\right) \tag{3}$$

the degree distribution of any particular node in the transitive closure $\overline{H}_{x+1,q}$ of a Bernoulli random graph $H_{x+1,q}$ on node set $\{1, \dots, x+1\}$, where each node pair is linked with probability q, independently of other node pairs. Alternatively, $\mathrm{Bin}^+(x,q)(t)$ equals the probability that the connected component of any particular node in $H_{x+1,q}$ has size $t+1$. Both distributions have the same support $\{0, \dots, x\}$, and $\mathrm{Bin}(x,q) \le_{\mathrm{st}} \mathrm{Bin}^+(x,q)$ in the strong stochastic order. No simple closed form expression is known for $\mathrm{Bin}^+(x,q)(t)$, but its values can be efficiently computed with the help of Gontcharoff polynomials [2, 5]. The compound Poisson distribution with rate parameter λ and increment distribution g is denoted $\mathrm{CPoi}(\lambda, g)$; recall that this is the law of a random variable $\sum_{k=1}^{\Lambda} X_k$ where $\Lambda, X_1, X_2, \dots$ are mutually independent and such that $\mathcal{L}(\Lambda) = \mathrm{Poi}(\lambda)$ and $\mathcal{L}(X_k) = g$.

For any probability measure P on $\mathbb{Z}_+ \times [0,1]$, any P-distributed random variable (X,Q), and integers $r, s \ge 0$, we denote

$$(P)_{rs} = \mathbb{E}(X)_r Q^s = \int (x)_r q^s \, P(dx, dq), \tag{4}$$

and when this quantity is finite and nonzero, we define mixed probability distributions $\mathrm{Bin}_{rs}(P)$ and $\mathrm{Bin}^+_{rs}(P)$ on $\mathbb{Z}_+$ with probability mass functions

$$\mathrm{Bin}_{rs}(P)(t) = \mathbb{E}\left(\mathrm{Bin}(X-r, Q)(t) \, \frac{(X)_r Q^s}{(P)_{rs}} \right), \tag{5}$$

$$\mathrm{Bin}^+_{rs}(P)(t) = \mathbb{E}\left(\mathrm{Bin}^+(X-r, Q)(t) \, \frac{(X)_r Q^s}{(P)_{rs}} \right), \tag{6}$$

where $\mathrm{Bin}(0,q)(t) = \mathrm{Bin}^+(0,q)(t) = \mathbb{I}_{\{t=0\}}$ and $\mathrm{Bin}(x,q) \equiv 0$, $\mathrm{Bin}^+(x,q) \equiv 0$ for $x < 0$.

3 | MAIN RESULTS

3.1 | Degree distribution

The model degree distribution is defined by

$$f^{(n)}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}\left(\deg_{G_{(n)}}(i) = t \right), \tag{7}$$

and represents the probability distribution of the number of neighbors of a randomly chosen node. Because $G_{(n)}$ is an exchangeable random graph, we see that $f^{(n)} = \mathcal{L}\big(\deg_{G_{(n)}}(1) \big)$.

Theorem 3.1. *Assume that $\frac{m}{n} \to \mu \in (0,\infty)$ and $P_n \to P$ weakly together with $(P_n)_{10} \to (P)_{10} \in (0,\infty)$ for some probability measure P on $\mathbb{Z}_+ \times [0,1]$. Then the model degree distribution $f^{(n)}$ converges weakly to the compound Poisson distribution $f = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(P))$.*

The limiting degree distribution f in Theorem 3.1 can be represented as the law of $\zeta = \sum_{k=1}^{\Lambda} \zeta_k$ where Λ is Poisson distributed with mean $\mu(P)_{10}$, $\zeta_1, \zeta_2, \dots$ follow a mixed binomial distribution $\mathrm{Bin}_{10}(P)$, and the random variables in the sum are mutually independent. Here, Λ represents the number of layers covering a particular node, and ζ_k the number of neighbors in a typical layer covering the node. The mean equals $\mathbb{E}\,\zeta = \mu(P)_{21} \le \infty$, and the variance equals $\mathrm{Var}(\zeta) = \mu\left((P)_{21} + (P)_{32}\right)$ for $(P_{21}), (P_{32}) < \infty$. Moreover, $\mathbb{E}(\zeta^r) < \infty$ if and only if $(P)_{r+1,r} < \infty$. The generating function is given by $\mathbb{E}(z^{\zeta}) = e^{\lambda(\hat{g}_{10}(z)-1)}$, where $\hat{g}_{10}(z) = \int (1-q+qz)^{x-1}\frac{xP(dx,dq)}{(P)_{10}}$. The structure of P determines whether or not the limiting degree distribution is light-tailed or heavy-tailed. Section 4 illustrates both cases and provides examples of power laws with a tunable exponent. In Theorem 3.1 we have assumed that $(P)_{10} > 0$. One can show that for $(P)_{10} = 0$ the asymptotic degree distribution is degenerate at zero.

3.2 | Clustering

Given a finite (nonrandom) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the global clustering coefficient $\tau_{\mathcal{G}}$ and the degree dependent local clustering coefficient (also called clustering spectrum) $\sigma_{\mathcal{G}}(k)$ are defined as follows. Let $N_{\triangle}$ and $N_{\vee}$ denote the number of triangles and cherries (paths of length 2) of $\mathcal{G}$, respectively. Let $N_{\triangle}(v)$ be the number of triangles containing vertex v. Then

$$\tau_{\mathcal{G}} = \frac{3N_{\triangle}}{N_{\vee}}, \qquad \sigma_{\mathcal{G}}(k) = \frac{\sum_{v\in\mathcal{V}} N_{\triangle}(v)\mathbb{I}_{\{\deg_{\mathcal{G}}(v)=k\}}}{\sum_{v\in\mathcal{V}} \binom{k}{2}\mathbb{I}_{\{\deg_{\mathcal{G}}(v)=k\}}}. \tag{8}$$

These network characteristics represent conditional probabilities of a link between two neighbors of a randomly selected vertex. Let (v_1^*, v_2^*, v_3^*) be an ordered triple of vertices sampled uniformly at random. Let $\triangle_{v_1^*,v_2^*,v_3^*}$ denote the event that v_1^*, v_2^*, v_3^* induce the triangle. Similarly, let $\vee_{v_1^*,v_2^*,v_3^*}$ denote the event that v_3^* is adjacent to v_1^* and v_2^*. A straightforward calculation shows that the ratios (8) can be written in the form

$$\tau_{\mathcal{G}} = \mathbb{P}^*\left(\triangle_{v_1^*,v_2^*,v_3^*} | \vee_{v_1^*,v_2^*,v_3^*}\right), \qquad \sigma_{\mathcal{G}}(k) = \mathbb{P}^*\left(\triangle_{v_1^*,v_2^*,v_3^*} | \vee_{v_1^*,v_2^*,v_3^*}, \deg_{\mathcal{G}}(v_3^*) = k\right).$$

Here, the probability $\mathbb{P}^*$ refers to random sampling of vertices v_1^*, v_2^*, v_3^*. Below we consider similar conditional probabilities, but defined for the random graph G (instead of $\mathcal{G}$)

$$\tau(G) = \mathbb{P}\left(\triangle_{v_1^*,v_2^*,v_3^*} | \vee_{v_1^*,v_2^*,v_3^*}\right), \qquad \sigma(G)(k) = \mathbb{P}\left(\triangle_{v_1^*,v_2^*,v_3^*} | \vee_{v_1^*,v_2^*,v_3^*}, \deg_{G}(v_3^*) = k\right). \tag{9}$$

We call $\tau(G)$ the model (global) clustering coefficient and $\sigma(G)(k)$ the model clustering spectrum of the random overlay graph G. Note that conditional probabilities (9) refer to two sources of randomness, namely, the random graph G and the random sampling of vertices v_1^*, v_2^*, v_3^* (which is assumed to be independent of G). It is interesting to compare the probabilities (9) with the respective ratios τ_G and $\sigma_G(k)$ of (8), where $\mathcal{G}$ is replaced by an instance of the random graph G. An argument bearing on the law of large numbers (applied to the sums of random variables in the numerators and denominators of ratios (8)) suggests that $\tau_G - \tau(G) = o_P(1)$ and $\sigma_G(k) - \sigma(G)(k) = o_P(1)$ as the number of vertices $n \to +\infty$. Therefore, the model characteristics $\tau(G)$ and $\sigma(G)(k)$ can be viewed as approximations to the clustering coefficients τ_G and $\sigma_G(k)$ and our asymptotic results for $\tau(G)$ and $\sigma(G)(k)$ shown below can likely be extended to τ_G and $\sigma_G(k)$.

Now we focus on the model clustering characteristics $\tau(G)$ and $\sigma(G)(k)$. We observe that since the distribution of G is invariant under permutation of its vertices, we have that

$$\tau(G) = \frac{\sum'_{ijk}\mathbb{P}(G(ij), G(ik), G(jk))}{\sum'_{ijk}\mathbb{P}(G(ij), G(ik))}, \quad \sigma(G)(k) = \frac{\sum'_{ij\ell}\mathbb{P}(\deg_G(i) = k, G(ij), G(i\ell), G(j\ell))}{\sum'_{ij\ell}\mathbb{P}(\deg_G(i) = k, G(ij), G(i\ell))},$$

where $G(ij)$ represents the event that nodes i and j are adjacent, and the sums are taken over ordered triples of distinct nodes. We denote $\tau^{(n)} = \tau(G_{(n)})$ and $\sigma^{(n)}(k) = \sigma(G_{(n)})(k)$.

Theorem 3.2. *Assume that* $(P_n)_{rs} \to (P)_{rs} < \infty$ *for* $rs = 21, 32, 33$*, and* $(P)_{21} > 0$*. Then* $\lim_{n\to\infty} \tau^{(n)} = \tau$*, where*

$$\tau = \begin{cases} \dfrac{(P)_{33}}{(P)_{32}} & \text{when } m \ll n \text{ and } (P)_{32} > 0, \\ \dfrac{(P)_{33}}{(P)_{32} + \mu(P)_{21}^2} & \text{when } \frac{m}{n} \to \mu \in (0, \infty), \\ 0 & \text{when } n \ll m \ll n^2. \end{cases}$$

Remark (constant layer strengths)**.** When $Q_k = q$ is constant for all k, we see that $(P)_{rs} = (p)_r q^s$ where $(p)_r$ equals the rth factorial moment of the limiting layer size distribution. In this case the limiting model clustering equals $\frac{q(p)_3}{(p)_3+\mu(p)_2^2}$ and agrees with [9, 32].

Theorem 3.3. *Assume that* $\frac{m}{n} \to \mu \in (0, \infty)$*, and* $P_n \to P$ *weakly together with* $(P_n)_{rs} \to (P)_{rs} \in (0, \infty)$ *for* $rs = 10, 21, 32$*. Then* $\sigma^{(n)} \to \sigma$ *pointwise to the limit*

$$\sigma(t) = \frac{(P)_{33}\ (f * g_{33})(t-2)}{(P)_{32}(f * g_{32})(t-2) + \mu(P)_{21}^2(f * g_{21} * g_{21})(t-2)}, \tag{10}$$

where $f = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(P))$ *is the limiting degree distribution in Theorem 3.1, and the distributions* $g_{rs} = \mathrm{Bin}_{rs}(P)$ *are defined by (5).*

Section 4 illustrates examples where the limiting clustering spectrum $\sigma(t)$ follows a power law.

3.3 | Connected components

We denote by $N_1(G_{(n)}) \geq N_2(G_{(n)})$ the two largest component sizes in $G_{(n)}$. For a probability distribution h on $\mathbb{Z}_+$, we denote by

$$\rho_{(h)} = 1 - \min\left\{ s \geq 0 : \sum_{x\geq 0} s^x h(x) = s \right\}$$

the probability of eternal survival of a Galton–Watson branching process with offspring distribution h.

Theorem 3.4. *Assume that* $\frac{m}{n} \to \mu \in (0, \infty)$ *and* $P_n \to P$ *weakly together with* $(P_n)_{10} \to (P)_{10} \in (0, \infty)$*. Then*

$$\frac{N_1(G_{(n)})}{n} \xrightarrow{\mathbb{P}} \rho_{(f^+)} \quad \textit{and} \quad \frac{N_2(G_{(n)})}{n} \xrightarrow{\mathbb{P}} 0,$$

where $f^+ = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}^+(P))$ *is a compound Poisson distribution with rate parameter* $\mu(P)_{10}$ *and increment distribution* $\mathrm{Bin}_{10}^+(P)$ *defined by (6).*

In Theorem 3.4 we have assumed that $(P)_{10} > 0$. One can show that for $(P)_{10} = 0$ we have $\frac{N_1(G_{(n)})}{n} \xrightarrow{\mathbb{P}} 0$.

3.4 | Site percolation

We may analyze how a subset of nodes $S \subset \{1, \dots, n\}$ is connected by considering a *site-percolated graph* defined as the subgraph

$$\check{G} = G[S] \tag{11}$$

of G induced by S. The site-percolated overlay graph is an instance of the overlay graph model (1) on the vertex set S with layers $(\check{G}_1, \check{X}_1, \check{Q}_1), \dots, (\check{G}_m, \check{X}_m, \check{Q}_m)$ such that $\check{G}_k$ has vertex set $V(\check{G}_k) = S \cap V(G_k)$ of size $\check{X}_k := |V(\check{G}_k)|$ and $\check{Q}_k = Q_k$. Note that the conditional distribution of $\check{X}_k$ given $X_k = |V(G_k)|$ is hypergeometric. An approximation of the hypergeometric distribution by a binomial distribution $\mathrm{Bin}(X_k, \theta)$ with $\frac{|S|}{n} \approx \theta$ suggests replacing the limiting layer type distribution P by

$$\check{P}(A) = \int \left(\mathrm{Bin}(x, \theta) \times \delta_q\right)(A)\ P(dx, dq).$$

The following result confirms that this modification is well justified, and summarizes the results of Theorems 3.1–3.4 adjusted to site percolation.

Theorem 3.5. *Assume that* $\frac{m}{n} \to \mu \in (0, \infty)$, $P_n \to P$ *weakly together with* $(P_n)_{10} \to (P)_{10} \in (0, \infty)$, *and* $S_n \subset \{1, \dots, n\}$ *satisfies* $\frac{|S_n|}{n} \to \theta \in (0, 1]$. *Then the following approximations are valid for the site-percolated graph* $\check{G}_{(n)} = \check{G}_{(n)}[S_n]$:

(i) The degree distribution converges weakly to $\check{f} = \mathrm{CPoi}\left(\mu(\check{P})_{10}, \mathrm{Bin}_{10}(\check{P})\right)$.

(ii) $|S_n|^{-1} N_1\left(\check{G}_{(n)}\right) \xrightarrow{\mathbb{P}} \rho_{(\check{f}^+)}$ *and* $|S_n|^{-1} N_2\left(\check{G}_{(n)}\right) \xrightarrow{\mathbb{P}} 0$ *with* $\check{f}^+ = \mathrm{CPoi}\left(\mu(\check{P})_{10}, \mathrm{Bin}^+_{10}(\check{P})\right)$.

If we also assume that $(P_n)_{rs} \to (P)_{rs} \in (0, \infty)$ *for* $rs = 21, 32, 33$*, then*

(iii) The model clustering coefficient $\tau(\check{G}_{(n)})$ *converges to* τ *where* τ *is the corresponding limit of the nonpercolated graph* $G_{(n)}$.

(iv) The model clustering spectrum $\sigma(\check{G}_{(n)})$ *converges pointwise to* $\check{\sigma}$ *defined by replacing* f *and* g_{rs} *in (10) by* $\check{f}$ *and* $\check{g}_{rs} = \mathrm{Bin}_{rs}(\check{P})$.

3.5 | Bond percolation

Bond percolation studies how well the nodes of a graph are connected along a subset of links obtained by random sampling. In a multilayer network, we may either sample (i) a subset of links of the overlay graph, or (ii) independent subsets of links for each layer separately. To analyze these cases for the overlay graph model $G = G_{(n)}$ in (1), we define an *overlay bond-percolated graph* by

$$\hat{G} = G \cap H, \tag{12}$$

and a *layerwise bond-percolated graph* $\tilde{G}$ by

$$V(\tilde{G}) = \{1, \dots, n\} \qquad \text{and} \qquad E(\tilde{G}) = \cup_{k=1}^{m} E\left(G_k \cap H_k\right), \tag{13}$$

where $H, H_1, \dots, H_m$ are mutually independent random graphs on $\{1, \dots, n\}$ in which each node pair is linked with probability θ, independently of other node pairs, and independently of the layers (G_k, X_k, Q_k). To emphasize the dependence on θ we sometimes write $\hat{G} = \hat{G}(\theta)$ and $\tilde{G} = \tilde{G}(\theta)$.

In an epidemic modeling context, the standard SIR epidemic model is used to model individuals who infect their neighbors with probability θ, independently of each other [2]. The links of a graph G represent social contacts, and the bond-percolated component of node i corresponds to the set of eventually infected individuals in a population where node i is initially infectious and the other nodes susceptible. Bond percolation on the overlay graph can be used to develop finer models to model contacts of individuals generated by social communities (households, workplaces, schools) of variable size and strength. Layerwise percolation $\tilde{G}$ then models the case where infections occur independently inside the communities, and the overlay bond-percolation $\hat{G}$ models the case where infections occur between individuals regardless of the underlying community structure.

The layerwise bond-percolated graph is an instance of the overlay model (1) with layer types $(X_k, \theta Q_k)$. This suggests considering a modified limiting layer type distribution

$$\hat{P}(A) = \int (\delta_x \times \delta_{\theta q})(A)\, P(dx, dq).$$

We expect the overlay bond-percolated model to behave similarly to the layerwise bond-percolated model in sparse regimes where the layers do not overlap much. The following result confirms this, and summarizes the results of Theorems 3.1–3.4 adjusted to bond percolation.

Theorem 3.6. *Assume that* $\frac{m}{n} \to \mu \in (0, \infty)$, *and* $P_n \xrightarrow{w} P$ *together with* $(P_n)_{10} \to (P)_{10} \in (0, \infty)$, *and* $\theta \in (0, 1]$. *Then the following approximations are valid for both the overlay bond-percolated graph* $\hat{G}_{(n)} = \hat{G}_{(n)}(\theta)$ *and layerwise bond-percolated graph* $\tilde{G}_{(n)} = \tilde{G}_{(n)}(\theta)$:

(i) The degree distribution converges weakly to $\hat{f} = \mathrm{CPoi}\big(\mu(\hat{P})_{10}, \mathrm{Bin}_{10}(\hat{P})\big)$.

(ii) For N_1 *and* N_2 *denoting the largest and the second largest component sizes we have* $n^{-1}N_1 \xrightarrow{\mathbb{P}} \rho_{(\hat{f}^+)}$ *and* $n^{-1}N_2 \xrightarrow{\mathbb{P}} 0$ *with* $\hat{f}^+ = \mathrm{CPoi}\big(\mu(\hat{P})_{10}, \mathrm{Bin}^+_{10}(\hat{P})\big)$.

If we also assume that $(P_n)_{rs} \to (P)_{rs} \in (0, \infty)$ *for* $rs = 21, 32, 33$, *then*:

(iii) The model clustering coefficient converges to $\theta\tau$ *where* τ *is the corresponding limit of the nonpercolated graph* $G_{(n)}$.

(iv) The model clustering spectrum converges pointwise to $\hat{\sigma}$ *defined by replacing* P, f, *and* g_{rs} *in (10) by* $\hat{P}, \hat{f}$, *and* $\hat{g}_{rs} = \mathrm{Bin}_{rs}(\hat{P})$.

3.6 | Double phase transition

Theorem 3.6 shows that the largest relative component size in the bond-percolated graph is approximated by the survival probability $\rho_{(\hat{f}^+)}$ of a Galton–Watson process with compound Poisson offspring distribution $\hat{f}^+ = \mathrm{CPoi}\big(\mu(\hat{P})_{10}, \mathrm{Bin}^+_{10}(\hat{P})\big)$. The mean of the offspring distribution can be written as[1]

$$R_0(\theta) = \mu \int R(x-1, \theta q)\, xP(dx, dq), \tag{14}$$

where $R(x, q) = \sum_{t \ge 0} t\, \mathrm{Bin}^+(x, q)(t)$ defined using (3) represents the expected degree of a node in the transitive closure of Bernoulli random graph with $x + 1$ nodes and link probability q. Classical branching process theory tells that $\rho_{(\hat{f}^+)} > 0$ if and only if $R_0(\theta) > 1$. Hence the largest component in

[1] $R_0(\theta)$ can be interpreted as the basic reproduction number "R naught" in the epidemiological context.

the bond-percolated graph is sublinear for $\theta < \theta_1$, and linear for $\theta > \theta_1$, where the critical threshold is defined by

$$\theta_1 = \sup\{\theta \in [0,1] : R_0(\theta) < 1\}.$$

The overlay graph model studied in this article involves another nontrivial phase transition associated with a critical threshold value

$$\theta_2 = \sup\{\theta \in [0,1] : R_0(\theta) < \infty\}.$$

Section 4 describes an example where $0 < \theta_1 < \theta_2 < 1$.

The first phase transition at θ_1 characterizes the emergence of a giant component in a bond-percolated overlay graph. To understand the second phase transition, note that $R_0(\theta)$ is proportional to the expected number of nodes which can be reached by paths within a typical bond-percolated layer covering a particular node. The second phase transition at θ_2 hence amounts to the emergence of gigantic components inside bond-percolated layers covering a typical node.

In the epidemic context discussed in Section 3.5, we note that the critical quantity $R_0(\theta)$ does *not* refer to the number of individuals directly infected by a reference individual in an otherwise susceptible population, unlike in classical SIR models. Rather, $R_0(\theta)$ also counts the number of individuals indirectly infected by the reference individual via single-layer infection paths.

4 | POWER-LAW MODELS

This section illustrates the rich statistical features of the overlay model by discussing the results of Section 3 in a setting where the limiting layer strength is a deterministic function of layer size according to $Q = g(X)$ for some $g : \mathbb{Z}_+ \to [0,1]$, and the limiting layer type distribution factorizes according to

$$P(dx, dq) = p(dx)\delta_{g(x)}(dq), \tag{15}$$

where the layer size distribution p is a probability on $\mathbb{Z}_+$. We are especially interested in the case where the probability mass function $p(x) = \mathbb{P}(X = x)$ of the layer size distribution and $g(x)$ follow power laws

$$p(x) = L(x)x^{-\alpha} \quad \text{and} \quad g(x) = bx^{-\beta} \tag{16}$$

with exponents $\alpha \geq 2$ and $\beta \geq 0$. Here $L(x)$ is a slowly varying function at $+\infty$, $b > 0$ is a constant, and we choose $b \leq 1$ for $\beta = 0$. We will assume that (16) holds for large x. Note that for r, s satisfying $\alpha + s\beta > r + 1$ the cross moment

$$(P)_{rs} = \sum_{x \geq 0} (x)_r g(x)^s p(x)$$

is finite. It is also finite in the case where $\alpha + s\beta = r + 1$ and $x^{-1}L(x)$ is integrable at $+\infty$.

4.1 | Degree distribution and clustering spectrum

Theorems 4.1 and 4.2 below establish power laws for the limiting degree distribution and clustering spectrum. Figures 1 and 2 illustrate how the associated power-law exponents relate to the

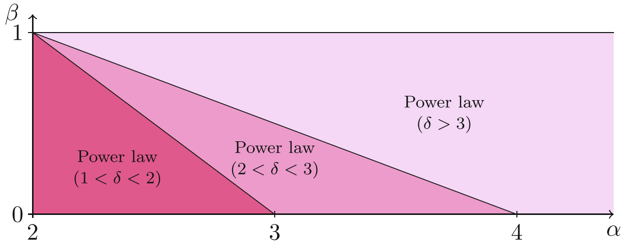

FIGURE 1 Power-law exponent of degree distribution as a function of layer size exponent α and layer strength exponent β.

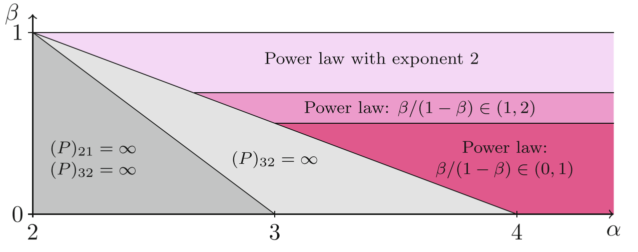

FIGURE 2 Power-law exponent of clustering spectrum as a function of layer size exponent α and layer strength exponent β. The assumptions of Theorem 4.2 do not hold in the grey areas where $(P)_{32} = \infty$.

corresponding exponents of layer sizes and layer strengths. Remarkably, the power law of the clustering spectrum admits a tunable exponent in $[0, 2]$. A similar power law with exponent 1 has earlier been established for a random intersection graph [9] and for a spatial preferential attachment random graph [27], and with exponent restricted to $[0, 1]$ for inhomogeneous random intersection graphs [7, 10, 12] and a hyperbolic random geometric graph model [24].

Theorem 4.1. *Let $\alpha \geq 2$, $\beta \geq 0$, and $b > 0$. Assume (16) and $(P)_{10} < \infty$.*

(i) If $\beta \in (0, 1)$, then the limiting degree distribution f satisfies as $t \to +\infty$

$$f(t) \sim cL\left(t^{1/(1-\beta)}\right) t^{-\delta} \tag{17}$$

for $\delta = 1 + \frac{\alpha-2}{1-\beta}$ and $c = \mu(1-\beta)^{-1} b^{\delta-1}$.

(ii) Relation (17) holds also for $\beta = 0$. Note that in this case we have $b \leq 1$.

(iii) If $\beta \geq 1$, then the limiting degree distribution is light-tailed with the moment generating function bounded by $\sum_{t \geq 0} e^{st} f(t) \leq \exp\left\{(e^{B(e^s-1)} - 1)\mu(P)_{10}\right\}$ $\forall s$. Here $B = \max_x xg(x)$.

Theorem 4.2. *Let $\alpha > 2$. Let $\beta \in [0, 1)$ be such that $\alpha + 2\beta > 4$. Assume that for some $a, b > 0$ (16) holds with $L(x) = a + o(1)$ as $x \to +\infty$. For $\beta = 0$ we assume in addition that $b < 1$. Then the model clustering spectrum defined by (10) follows a power law according to*

$$\sigma(t) \sim \begin{cases} c_1 t^{-\beta/(1-\beta)}, & \beta < 2/3, \\ c_2 t^{-2}, & \beta = 2/3, \\ c_3 t^{-2}, & \beta > 2/3, \end{cases}$$

where $c_1 = b^{1/(1-\beta)}$, $c_3 = \mu(P)_{33}$, and $c_2 = c_1 + c_3$. In particular, $\sigma(t) \sim b$ for $\beta = 0$.

We remark that the inequality $\alpha + 2\beta > 4$ implies $(P)_{10}, (P)_{21}, (P)_{32} < \infty$; in particular the asymptotic degree distribution has a finite second moment/variance.

Networks with $\sigma(t) \ll t^{-1}$ are sometimes called weakly clustered, and those with $\sigma(t) \gg t^{-1}$ strongly clustered [4]. According to Theorem 4.2, the overlay graph model produces weakly clustered networks for $\beta > \frac{1}{2}$, and strongly clustered networks for $\beta < \frac{1}{2}$. We believe Theorem 4.2 can be extended to more general subexponential distributions $p(x)$. We do not pursue this line here to avoid unnecessary technicalities.

4.2 | Existence of double phase transition

For the power-law model (16), the function in (14) can be computed as $R_0(\theta) = \mu \sum_x R(x-1, \theta g(x)) x p(x)$. By applying a classical giant component result for Bernoulli random graphs [30, Theorem 5.4], one may verify that[2]

$$\begin{aligned} \limsup_{x\to\infty} \theta x g(x) \le 1-\epsilon \quad &\Rightarrow \quad \limsup_{x\to\infty} R(x-1, \theta g(x)) \le 2\epsilon^{-2}, \\ \liminf_{x\to\infty} \theta x g(x) \ge 1+\epsilon \quad &\Rightarrow \quad \liminf_{x\to\infty} x^{-1} R(x-1, \theta g(x)) > 0. \end{aligned} \tag{18}$$

If $\alpha > 3$, then the limiting layer size distribution p has a finite second moment and $R(x-1, q) \le x-1$ implies that $R_0(1) < \infty$. Hence $\theta_2 = 1$, and the second phase transition cannot occur. On the other hand, when $\alpha \in (2, 3]$, the limiting layer size distribution has infinite second moment. In this case (18) yields the following conclusions:

(1) $\beta = 1$ with $b > 1$. Then $R_0(\theta) < \infty$ for $\theta < b^{-1}$, and $R_0(\theta) = \infty$ for $\theta > b^{-1}$. Hence $\theta_2 = b^{-1} \in (0, 1)$. Assume in addition that the constant a in (16) is large enough so that $\mu\theta(P)_{21} \ge 1$ for $\theta = \frac{1}{2}\theta_2$. Then $\hat{f}^+ \ge_{\rm st} \hat{f}$ implies that $R_0(\theta) = \sum_t t\hat{f}^+(t) \ge \sum_t t\hat{f}(t) = \mu\theta(P)_{21} \ge 1$ for $\theta = \frac{1}{2}\theta_2$, and the continuity of $R_0(\theta)$ on $[0, \theta_2)$ implies that $\theta_1 \in (0, \frac{1}{2}\theta_2)$. There are hence two critical values $0 < \theta_1 < \theta_2 < 1$ in which the model displays two distinct phase transitions.

(2) $\beta \in (1, \infty)$, or $\beta = 1$ with $b < 1$. Then $R_0(\theta) < \infty$ for all $\theta \in [0, 1]$, so that $\theta_2 = 1$, and the second-type phase transition cannot occur.

(3) $\beta \in [0, 1)$. Then one can show that $R_0(\theta) = \infty$ for all $\theta \in (0, 1]$, and hence $\theta_1 = \theta_2 = 0$, and there are no phase transitions of either type.

The above observations confirm the existence of a double phase transition in bond percolation, as postulated in [19], for a natural network model admitting tunable power-law exponents for both the degree distribution and the clustering spectrum. Together with Theorems 4.1 and 4.2, this opens up a flexible framework for studying the significance and interrelations of these power laws to bond and site percolation properties in clustered complex networks. The investigation of how these phase transitions are reflected in the core-periphery organization of the network [4, 19] remains an important topic for future research.

5 | NOTATION USED IN PROOFS

Let (X, Q) and $(X_{n,\pi}, Q_{n,\pi})$ be bivariate random variables with the distributions P and P_n respectively (one may interpret π as a integer selected uniformly at random from $\{1, \dots, m\}$). For a random variable

[2]The first implication in (18) follows by noting that if $xq < 1$, then the proof of Theorem 5.4 [30] shows that $\mathbb{E}|C_{H_{x,q}}(1)| = \sum_{t\ge1} \mathbb{P}(|C_{H_{x,q}}(1)| \ge t) \le \sum_{t\ge1} e^{-\frac{1}{2}(1-xq)^2 t} \le \int_0^\infty e^{-\frac{1}{2}(1-xq)^2 t} dt \le 2(1-xq)^{-2}$, so that $R(x-1, q) = \mathbb{E}|C_{H_{x,q}}(1)| - 1 \le 2(1-xq)^{-2}$.

ξ we denote by P_ξ the probability distribution of ξ and write $\xi \sim P_\xi$ (equivalently, $P_\xi = \mathcal{L}(\xi)$). For a bivariate random variable (ξ, η) we denote by P_ξ^η the conditional probability distribution of ξ given η. In particular, $P_\xi^\eta(B) = \mathbb{E}(\mathbb{I}_{\{\xi \in B\}} | \eta)$, for a Borel set B. In the proof of Theorem 3.1 below we use the inequality

$$d_{\rm tv}(P_\xi, P_\zeta) \le \mathbb{E}\, d_{\rm tv}\left(P_\xi^\eta, P_\zeta^\eta\right). \tag{19}$$

For vectors $\bar{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$ and $\bar{q} = (q_1, \dots, q_m) \in [0,1]^m$ we denote $(\bar{x}, \bar{q}) = ((x_1, q_1), \dots, (x_m, q_m))$. With a little abuse of notation we write $(\bar{x}_n, \bar{q}_n) = ((x_{n,1}, q_{n,1}), \dots, (x_{n,m}, q_{n,m}))$ and $(\overline{X}_n, \overline{Q}_n) = ((X_{n,1}, Q_{n,1}), \dots, (X_{n,m}, Q_{n,m}))$. To stress the dependence of the overlay graph on the sequence $(\overline{X}_n, \overline{Q}_n)$ we write $G_{(n)} = G_{(\overline{X}_n, \overline{Q}_n)}$. By $V = \{v_1, \dots, v_n\}$ we denote the vertex set of $G_{(n)}$. The vertex sets of the layers $G_{n,j}$ are denoted D_j, $1 \le j \le m$.

Next we introduce an ε-discretization of the space $[0,1]$ of admissible layer strengths (edge densities) which helps to reduce the analysis of the general model with potentially uncountably many layer types into a one with finitely many layer types. For any $\varepsilon \in (0,1)$ we fix numbers $0 = s_0 < s_1 < \cdots < s_r = 1$ (with $r \le 2/\varepsilon$) such that $\mathbb{P}(Q = s_i) = 0$ for $0 < s_i < 1$ and $|s_i - s_{i-1}| < \varepsilon$ for $1 \le i \le r$. Relative to this mesh, we define down-rounding and up-rounding operations $q \to q^-$ and $q \to q^+$ on $[0,1]$ using the formulas

$$q^+ = s_1 \mathbb{I}_{\{q=0\}} + \sum_{i=1}^r s_i \mathbb{I}_{\{s_{i-1} < q \le s_i\}}, \qquad q^- = \sum_{i=1}^r s_{i-1} \mathbb{I}_{\{s_{i-1} < q \le s_i\}}. \tag{20}$$

For $\bar{q} \in [0,1]^m$ we denote $\bar{q}^{\,-} = (q_1^-, \dots, q_m^-)$, $\bar{q}^+ = (q_1^+, \dots, q_m^+)$ and write

$$(\bar{x}, \bar{q}^{\,-}) = \left((x_1, q_1^-), \dots, (x_m, q_m^-)\right), \qquad (\bar{x}, \bar{q}^+) = \left((x_1, q_1^+), \dots, (x_m, q_m^+)\right). \tag{21}$$

Furthermore, we denote for short $G^+ = G_{(\overline{X}_n, \overline{Q}_n^+)}$ and $G^- = G_{(\overline{X}_n, \overline{Q}_n^-)}$. In particular, G^+ is the superposition of the layers $(G_{n,1}^+, X_{n,1}, Q_{n,1}^+), \dots, (G_{n,m}^+, X_{n,m}, Q_{n,m}^+)$. In view of the coordinate-wise inequalities $Q_{n,i}^- \le Q_{n,i} \le Q_{n,i}^+$, $1 \le i \le m$, there is a natural coupling of random graphs G^-, G^+, and $G_{(n)}$ such that $\mathbb{P}(G^- \subset G_{(n)} \subset G^+) = 1$. By d, d^+, and d^- we denote the degree of vertex v_1 in $G_{(n)}$, G^+ and G^- respectively. The discretization is used in the proofs of Theorems 3.1 and 3.4. Given $\varepsilon \in (0,1)$ we first establish respective results for G^- and G^+ and then letting $\varepsilon \downarrow 0$ we carry them over to $G_{(n)}$ using the coupling $G^- \subset G_{(n)} \subset G^+$.

By $\mathbb{E}_{(\overline{X}, \overline{Q})}$ and $\mathbb{P}_{(\overline{X}, \overline{Q})}$ we denote the conditional expectation and probability given the random vector $(\overline{X}_n, \overline{Q}_n)$.

6 | ANALYSIS OF DEGREE DISTRIBUTIONS

Here we prove Theorem 3.1. Before the proof we collect some useful facts about the compound Poisson distribution and the total variation distance. Let $\lambda_1, \lambda_2 > 0$. Let ξ, η be random variables and let $Z_\xi \sim \mathrm{CPoi}(\lambda_1, \mathcal{L}(\xi))$ and $Z_\eta \sim \mathrm{CPoi}(\lambda_2, \mathcal{L}(\eta))$ be independent compound Poisson random variables. Let I be Bernoulli random variable independent of (ξ, η) having the success probability $\mathbb{P}(I = 1) = \lambda_1 / (\lambda_1 + \lambda_2)$. Then

$$Z_\xi + Z_\eta \sim \mathrm{CPoi}(\lambda_1 + \lambda_2, \mathcal{L}(\zeta)), \qquad \text{where} \qquad \zeta = I\xi + (1 - I)\eta. \tag{22}$$

In the particular case where ξ and η take values in $\mathbb{Z}_+$ we have

$$\mathbb{P}(\zeta = k) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\mathbb{P}(\xi = k) + \frac{\lambda_2}{\lambda_1 + \lambda_2}\mathbb{P}(\eta = k), \qquad k \in \mathbb{Z}_+. \tag{23}$$

If, in addition, $\mathcal{L}(\xi) = \mathcal{L}(\eta)$ then we have $\mathcal{L}(\zeta) = \mathcal{L}(\xi) = \mathcal{L}(\eta)$.

Lemma 6.1. *Let γ, η be non-negative integer valued random variables. Let $\{\xi_t,\ t \geq 0\}$ and $\{\zeta_t,\ t \geq 0\}$ be sequences of independent random variables. Assume that $\{\xi_t,\ t \geq 0\}$ is independent of γ, and $\{\zeta_t,\ t \geq 0\}$ is independent of η. Then*

$$d_{\text{tv}}\left(\mathcal{L}(\xi_\gamma), \mathcal{L}(\zeta_\eta)\right) \leq 2d_{\text{tv}}\left(\mathcal{L}(\gamma), \mathcal{L}(\eta)\right) + \sum_{t\geq 0} d_{\text{tv}}\left(\mathcal{L}(\xi_t), \mathcal{L}(\zeta_t)\right)\mathbb{P}(\eta = t). \tag{24}$$

Proof. We can assume that $\{\xi_t,\ t \geq 0\}$ and η are independent. By the triangle inequality

$$d_{\text{tv}}\left(\mathcal{L}(\xi_\gamma), \mathcal{L}(\zeta_\eta)\right) \leq d_{\text{tv}}\left(\mathcal{L}(\xi_\gamma), \mathcal{L}(\xi_\eta)\right) + d_{\text{tv}}\left(\mathcal{L}(\xi_\eta), \mathcal{L}(\zeta_\eta)\right).$$

Furthermore, the identities that hold for any Borel set $B \subset \mathbb{R}$

$$\mathbb{P}(\xi_\gamma \in B) - \mathbb{P}(\xi_\eta \in B) = \sum_{t\geq 0}\left(\mathbb{P}(\gamma = t) - \mathbb{P}(\eta = t)\right)\mathbb{P}(\xi_t \in B)$$
$$\mathbb{P}(\xi_\eta \in B) - \mathbb{P}(\zeta_\eta \in B) = \sum_{t\geq 0}\left(\mathbb{P}(\xi_t \in B) - \mathbb{P}(\zeta_t \in B)\right)\mathbb{P}(\eta = t)$$

imply

$$d_{\text{tv}}\left(\mathcal{L}(\xi_\gamma), \mathcal{L}(\xi_\eta)\right) \leq 2d_{\text{tv}}(\mathcal{L}(\gamma), \mathcal{L}(\eta)),$$
$$d_{\text{tv}}\left(\mathcal{L}(\xi_\eta), \mathcal{L}(\zeta_\eta)\right) \leq \sum_{t\geq 0} d_{\text{tv}}(\mathcal{L}(\xi_t), \mathcal{L}(\zeta_t))\mathbb{P}(\eta = t).$$

■

We apply Lemma 6.1 to compound Poisson random variables. Let $a, b > 0$. Let ξ, ζ be random variables. It follows from the lemma that

$$d_{\text{tv}}\left(\text{CPoi}(a, \mathcal{L}(\xi)), \text{CPoi}(b, \mathcal{L}(\zeta))\right) \leq 2|a - b| + b d_{\text{tv}}\left(\mathcal{L}(\xi), \mathcal{L}(\zeta)\right) \tag{25}$$

To see how (25) follows from (24) put $\gamma \sim \text{Poi}(a)$, $\eta \sim \text{Poi}(b)$ and $\xi_t := \xi_1^{(1)} + \cdots + \xi^{(t)}$, $\zeta_t := \zeta^{(1)} + \cdots + \zeta^{(t)}$, where $\xi^{(i)}$, $i \geq 1$, and $\zeta^{(j)}$, $j \geq 1$, are independent copies of ξ and ζ respectively. Then use the triangle inequality $d_{\text{tv}}((\mathcal{L}(\xi_t), \mathcal{L}(\zeta_t)) \leq d_{\text{tv}}(\mathcal{L}(\xi), \mathcal{L}(\zeta))$ and the fact that $\mathbb{E}\eta = b$.

Proof of Theorem 3.1. In the proof we drop the subscript n when it does not cause an ambiguity. Thus we write $G = G_{(n)}$, $G_j = G_{n,j}$, $G_{n,j}^+ = G_j^+$ and $X_j = X_{n,j}$, $Q_j = Q_{n,j}$, $(\overline{X}, \overline{Q}) = (\overline{X}_n, \overline{Q}_n)$, $X_\pi = X_{n,\pi}$. We begin by outlining the idea of the proof. The degree d of vertex v_1 in the overlay random graph G is approximated by the sum of degrees of v_1 in the layers G_j containing this vertex. We denote this sum

$$L_A = \sum_{1\leq j\leq m} \mathbb{I}_{\{v_1 \in D_j\}} H(j). \tag{26}$$

Here, $\mathbb{I}_{\{v_1 \in D_j\}}$ is the indicator of the event $\{v_1 \in D_j\}$ and $H(j)$ stands for the degree of v_1 in G_j. In the sparse regime considered it is rather unlikely that two layers intersect in more than one point. Hence we approximate

$$d = L_A + o_P(1). \tag{27}$$

Note that only the layers of size at least 2 may contribute to the sum L_A. We denote the respective number of layers $S_A = \sum_{1\le j\le m} \mathbb{I}_{\{v_1 \in D_j, X_j \ge 2\}}$. In order to analyze the distributions of L_A and S_A it is convenient to condition on (X_j, Q_j), $1 \le j \le m$. Then S_A becomes the sum of independent Bernoulli random variables and its distribution is approximately $\mathrm{Poi}\left(\frac{m}{n}\tilde{x}_*\right)$, where $\tilde{x}_* = \frac{1}{m}\sum_{1\le j\le m} X_j \mathbb{I}_{\{X_j\ge 2\}}$. Furthermore, each $H(j)$ has binomial distribution $\mathrm{Bin}(X_j - 1, Q_j)$ and the probability that it will contribute to the sum L_A is proportional to X_j. This in turn yields that the typical contribution to the sum (26) by a layer of size at least 2 has the size biased mixed binomial distribution

$$h(t) = h(\overline{X}, \overline{Q})(t) = \frac{1}{\tilde{x}_*}\frac{1}{m}\sum_{1\le j\le m} \mathrm{Bin}(X_j - 1, Q_j)(t) X_j \mathbb{I}_{\{X_j \ge 2\}}, \qquad t = 0, 1, \ldots, \tag{28}$$

while the conditional distribution of L_A is approximately $\mathrm{CPoi}\left(\frac{m}{n}\tilde{x}_*, h\right)$.

Finally, letting $n, m \to +\infty$ we approximate $\frac{m}{n}\tilde{x}_* \to \mu x_*$, where $x_* = \mathbb{E} X \mathbb{I}_{\{X\ge 2\}}$. Furthermore, we approximate h by the distribution h_*, where

$$h_*(t) = \frac{1}{x_*}\mathbb{E}\mathrm{Bin}(X-1, Q)(t) X \mathbb{I}_{\{X\ge 2\}}, \qquad t = 0, 1, 2 \ldots . \tag{29}$$

In this way we establish the approximation $\mathrm{CPoi}\left(\frac{m}{n}\tilde{x}_*, h\right) \xrightarrow{w} \mathrm{CPoi}(\mu x_*, h_*)$. We conclude the outline with the observation that

$$\mathrm{CPoi}(\mu x_*, h_*) = \mathrm{CPoi}\left(\mu (P)_{10}, \mathrm{Bin}_{10}(P)\right). \tag{30}$$

Indeed, a simple calculation shows that

$$\begin{aligned}
&(P)_{10} = \mathbb{E}X = \mathbb{P}(X = 1) + x_*,\\
&(P)_{10}\mathrm{Bin}_{10}(P)(0) = x_* \left(\mathbb{P}(X = 1) + h_*(0)\right),\\
&(P)_{10}\mathrm{Bin}_{10}(P)(t) = x_* h_*(t), \qquad t \ge 1.
\end{aligned}$$

These identities imply the relation $(P)_{10}(\hat{g} - 1) = x_*(\hat{h}_* - 1)$ between the Fourier transforms $\hat{g}$ and $\hat{h}_*$ of the increment distributions $\mathrm{Bin}_{10}(P)$ and h_*. The latter relation establishes the correspondence between the Fourier transforms of respective compound distributions (30).

Finally, we mention that the rigorous proof of Theorem 3.1 is a bit more complex and the proof idea is somewhat hidden by technicalities including the truncation and discretization.

Now we give a rigorous proof. We first consider the special case where there exists $M > 0$ such that $X_{n,j} \le M$ almost surely for each n and $1 \le j \le m$.

Given $0 < \varepsilon < 1$, consider ε-discretized random variables Q^+ and Q^- and the overlay graphs G^+ and G^- defined by the vectors $(\overline{X}_n, \overline{Q}_n^+)$ and $(\overline{X}_n, \overline{Q}_n^-)$, see (20), (21). Let h_*^+ and h_*^- denote the distributions defined by (29), where Q is replaced by Q^+ and Q^- respectively. Let $d_*^+ = d_*^+(\varepsilon)$ and $d_*^- = d_*^-(\varepsilon)$

be compound Poisson random variables with the distributions $\mathrm{CPoi}(\mu x_*, h_*^+)$ and $\mathrm{CPoi}(\mu x_*, h_*^-)$. Note that as $\varepsilon \to 0$

$$\mathcal{L}(d_*^+) \xrightarrow{w} \mathrm{CPoi}(\mu x_*, h_*) \qquad \text{and} \qquad \mathcal{L}(d_*^-) \xrightarrow{w} \mathrm{CPoi}(\mu x_*, h_*). \tag{31}$$

Now we turn to the analysis of the degrees d, d^+ and d^-. We will show below that

$$\mathcal{L}(d^-) \xrightarrow{w} \mathcal{L}(d_*^-) \qquad \text{and} \qquad \mathcal{L}(d^+) \xrightarrow{w} \mathcal{L}(d_*^+). \tag{32}$$

Since the coupling $G^- \subset G \subset G^+$ implies the coupling $d^- \le d \le d^+$ we have

$$\mathbb{P}(d^+ \le t) \le \mathbb{P}(d \le t) \le \mathbb{P}(d^- \le t) \qquad \forall\, t \ge 0. \tag{33}$$

Combining (32) and (33) we obtain

$$\begin{aligned} \mathbb{P}(d_*^+ \le t) &= \liminf_n \mathbb{P}(d^+ \le t) \le \liminf_n \mathbb{P}(d \le t) \\ &\le \limsup_n \mathbb{P}(d \le t) \le \limsup_n \mathbb{P}(d^- \le t) = \mathbb{P}(d_*^- \le t). \end{aligned} \tag{34}$$

Finally, letting $\varepsilon \downarrow 0$ we obtain from (31), (34) that $\mathcal{L}(d) \xrightarrow{w} \mathrm{CPoi}(\mu x_*, h_*)$.

It remains to prove (32). We only show the second relation. The proof of the first one is the same. At this point we need some more notation. Recall ε-discretization (20). Let $\eta_{k,i}^{(j)}$, $\xi_{k,i}^{(j)}$, $H_{k,i}$, $H_{k,i}^{(j)}$, $H_{k,i}^{(j,l)}$ for $k \ge 2$ and $1 \le i \le r$ and $j, l \ge 1$ be independent random variables. We assume that $\eta_{k,i}^{(j)}$ and $\xi_{k,i}^{(j)}$ have Poisson and Bernoulli distributions with mean values $\mathbb{E}\eta_{k,i}^{(j)} = \mathbb{E}\xi_{k,i}^{(j)} = k/n$. Furthermore, $H_{k,i}$, $H_{k,i}^{(j)}$, $H_{k,i}^{(j,l)}$ have binomial distribution $\mathrm{Bin}(k-1, s_i)$, where s_i, $1 \le i \le r$ are the same as in (20). Let

$$L_1 = \sum_{1 \le j \le m} \mathbb{I}_j(v_1) H^+(j) = \sum_{1 \le j \le m} \mathbb{I}_j(v_1) H^+(j) \mathbb{I}_{\{X_j \ge 2\}}, \tag{35}$$

where $H^+(j)$ stands for the number of neighbors of v_1 created by the layer G_j^+. In particular, given $(\overline{X}, \overline{Q})$, the random variable $H^+(j)$ has binomial distribution $\mathrm{Bin}(X_j - 1, Q_j^+)$. Introduce random sets $\mathcal{M}_{k,i} = \{j : (X_j, Q_j^+) = (k, s_i)\}$ and denote their sizes $m_{k,i} = |\mathcal{M}_{k,i}|$. Furthermore, put

$$S = \sum_{1 \le j \le m} X_j \mathbb{I}_{\{X_j \ge 2\}}, \qquad \hat{\lambda} = \frac{S}{n} = \frac{m}{n} \tilde{x}_*, \qquad \hat{p}_{k,i} = \frac{k m_{k,i}}{S} \mathbb{I}_{\{S > 0\}}, \qquad 1 \le i \le r,\ k \ge 2.$$

Next we condition on $(\overline{X}, \overline{Q})$. Given $(\overline{X}, \overline{Q})$ such that $S > 0$ define random variables

$$L_2 = L_2(\overline{X}, \overline{Q}) = \sum_{2 \le k \le M} \sum_{i=1}^{r} \sum_{j=1}^{m_{k,i}} \xi_{k,i}^{(j)} H_{k,i}^{(j)}, \quad L_3 = L_3(\overline{X}, \overline{Q}) = \sum_{2 \le k \le M} \sum_{i=1}^{r} \sum_{j=1}^{m_{k,i}} \sum_{l=1}^{\eta_{k,i}^{(j)}} H_{k,i}^{(j,l)}.$$

For $S = 0$ we put $L_2 \equiv 0$ and $L_3 \equiv 0$.

We observe that the conditional distributions $P_{L_2}^{(\overline{X},\overline{Q})}$ and $P_{L_1}^{(\overline{X},\overline{Q})}$ do coincide. To see this we partition the index set $\{j : X_j \ge 2\} = \cup_{(k,i)} \mathcal{M}_{k,i}$ of the second sum of (35) and note that the conditional distributions of $\sum_{j=1}^{m_{k,i}} \xi_{k,i}^{(j)} H_{k,i}^{(j)}$ and $\sum_{j \in \mathcal{M}_{k,i}} \mathbb{I}_j(v_1) H^+(j)$ do coincide. Another useful fact is that for $S > 0$

the conditional distribution $P_{L_3}^{(\overline{X},\overline{Q})}$ is a compound Poisson distribution. Moreover, using the property (22), (23) one can easily show that $P_{L_3}^{(\overline{X},\overline{Q})} = \mathrm{CPoi}(\hat{\lambda}, h^+)$, where

$$h^+(t) = h^+(\overline{X},\overline{Q})(t) = \sum_{2\le k\le M}\sum_{1\le i\le r} \mathbb{P}(H_{k,i}=t)\hat{p}_{k,i}, \qquad t = 0,1,2,\ \dots .$$

Now we are ready for the proof of (32). We observe that $d^+ \neq L_1$ implies that some $v_l \neq v_1$ is a neighbor of v_1 in at least two different layers. We have, by the union bound and symmetry, that

$$\begin{aligned}\mathbb{P}(d^+ \neq L_1) &\le (n-1)\sum_{i<j}\mathbb{P}(v_1,v_2\in D_i\cap D_j) = (n-1)\sum_{i<j}\left(\mathbb{E}\frac{(X_i)_2}{(n)_2}\right)\left(\mathbb{E}\frac{(X_j)_2}{(n)_2}\right)\\ &\le (n-1)\frac{m^2}{n^4}\left(\mathbb{E}X_\pi^2\right)^2.\end{aligned}$$

Invoking $\mathbb{E}X_\pi^2 \le M^2$ and using the fact that $\mathcal{L}(L_2) = \mathcal{L}(L_1)$ we obtain

$$d_{\mathrm{tv}}(\mathcal{L}(d^+),\mathcal{L}(L_2)) = d_{\mathrm{tv}}(\mathcal{L}(d^+),\mathcal{L}(L_1)) \le \mathbb{P}(d^+\neq L_1) = O(n^{-1}). \tag{36}$$

Next we evaluate $d_{\mathrm{tv}}(\mathcal{L}(L_2),\mathcal{L}(L_3))$. To this aim we consider an array of random variables starting with L_2 and ending at L_3 where each subsequent element of the sequence is obtained from the previous one by replacing $\xi_{k,i}^{(j)}H_{k,i}^{(j)} = \sum_{1\le l\le \xi_{k,i}^{(j)}} H_{k,i}^{(j,l)}$ by $\sum_{1\le l\le \eta_{k,i}^{(j)}} H_{k,i}^{(j,l)}$. We proceed until all the products $\xi_{k,i}^{(j)}H_{k,i}^{(j)}$ are replaced so that at the array ends with L_3. By the triangle inequality, the total variation distance between the conditional distributions

$$d_{\mathrm{tv}}\left(P_{L_2}^{(\overline{X},\overline{Q})}, P_{L_3}^{(\overline{X},\overline{Q})}\right) \le \sum_{2\le k\le M}\ \sum_{1\le i\le r}\ \sum_{1\le j\le m_{k,i}} d_{\mathrm{tv}}\left(\mathcal{L}\left(\xi_{k,i}^{(j)}\right),\mathcal{L}\left(\eta_{k,i}^{(j)}\right)\right).$$

Invoking the bound $d_{\mathrm{tv}}\left(\mathcal{L}\left(\xi_{k,i}^{(j)}\right),\mathcal{L}\left(\eta_{k,i}^{(j)}\right)\right) \le 2k^2/n^2$, which follows by Le Cam's inequality [42], we obtain

$$d_{\mathrm{tv}}\left(P_{L_2}^{(\overline{X},\overline{Q})}, P_{L_3}^{(\overline{X},\overline{Q})}\right) \le 2\sum_{1\le j\le m} X_j^2 n^{-2} \le \frac{2m}{n^2}\mathbb{E}_{(\overline{X},\overline{Q})}X_\pi^2.$$

Now an application of (19) yields

$$d_{\mathrm{tv}}\left(\mathcal{L}(L_2),\mathcal{L}(L_3)\right) \le \frac{2m}{n^2}\mathbb{E}X_\pi^2 \le \frac{2m}{n^2}M^2 = O(n^{-1}). \tag{37}$$

Finally, we evaluate the distance $d_{\mathrm{tv}}(\mathcal{L}(L_3),\mathcal{L}(d^+))$, where $\mathcal{L}(d^+) = \mathrm{CPoi}(\mu x_*, h_*^+)$. For this purpose it is convenient to write h_*^+ in the form

$$h_*^+(t) = \sum_{2\le k\le M}\sum_{1\le i\le r}\mathbb{P}(H_{k,i}=t)p_{k,i}^+, \qquad p_{k,i}^+ := \frac{k}{x_*}\mathbb{P}(X=k, Q^+=s_i).$$

Recall that given $(\overline{X},\overline{Q})$ the conditional distribution of L_3 is $\mathrm{CPoi}(\hat{\lambda},h^+)$. It follows from (25) that

$$d_{\mathrm{tv}}\left(\mathrm{CPoi}(\hat{\lambda},h^+),\mathrm{CPoi}(\mu x_*,h_*^+)\right) \le 2|\hat{\lambda}-\mu x_*| + \mu x_*\ d_{\mathrm{tv}}(h^+,h_*^+). \tag{38}$$

Furthermore, Lemma 6.1 implies

$$d_{\text{tv}}(h^+, h_*^+) \le 2 \sum_{2\le k\le 1} \sum_{\le i \le r} |\hat{p}_{k,i} - p_{k,i}^+|. \tag{39}$$

(Here we apply Lemma 6.1 to $\xi_{k,i} = \zeta_{k,i} = H_{k,i}$ in the case where γ and η are bivariate random variables with the distributions $\mathbb{P}(\gamma = (k,i)) = \hat{p}_{k,i}$ and $\mathbb{P}(\eta = (k,i)) = p_{k,i}^+$.) We observe that both terms on the right of (38) vanish in probability. Indeed, by the weak law of large numbers (we use Chebyshev's inequality), our assumption $P_n \xrightarrow{w} P$ implies $\hat{\lambda} - \lambda \xrightarrow{\mathbb{P}} 0$ and $\hat{p}_{k,i} - p_{k,i}^+ \xrightarrow{\mathbb{P}} 0$ for each k and i. Since the number of different pairs (k,i) is finite, we conclude that $\tau := d_{\text{tv}}(\text{CPoi}(\hat{\lambda}, h^+), \text{CPoi}(\mu x_*, h_*^+) \xrightarrow{\mathbb{P}} 0$. In view of the obvious inequality $\tau \le 1$ we have also that $\mathbb{E}\tau = o(1)$. Now (19) implies $d_{\text{tv}}(\mathcal{L}(L_3), \mathcal{L}(d_*^+)) = o(1)$.

Finally, from the latter bound combined with (36), (37) we derive (32), by the triangle inequality:

$$d_{\text{tv}}\left(\mathcal{L}(d^+), \mathcal{L}(d_*^+)\right) \le d_{\text{tv}}\left(\mathcal{L}(d^+), \mathcal{L}(L_2)\right) + d_{\text{tv}}\left(\mathcal{L}(L_2), \mathcal{L}(L_3)\right) + d_{\text{tv}}\left(\mathcal{L}(L_3), \mathcal{L}(d_*^+)\right) = o(1).$$

Now we revoke the extra condition that $X_{n,j} \le M$ almost surely for each n and $1 \le j \le m$. From now on let $\{(X_{n,j}, Q_{n,j}),\ n \ge 1, 1 \le j \le m_n\}$ be arbitrary bivariate random variables satisfying conditions of the theorem. Given $M > 0$, let $G_{(n)}^{[M]}$ be the random overlay graph defined by the sequence $\left((X_{n,1}^{[M]}, Q_{n,1}), \ldots, (X_{n,m}^{[M]}, Q_m)\right)$, where $X_{n,j}^{[M]} = X_{n,j}\mathbb{I}_{\{X_{n,j} \le M\}}$. In the proof above we have shown that the degree $d^{[M]}$ of vertex v_1 in $G_{(n)}^{[M]}$ has asymptotic compound Poisson distribution $\text{CPoi}\left(\mu x_*^{[M]}, h_*^{[M]}\right)$. Here $x_*^{[M]}$ and $h_*^{[M]}$ are defined in the same way as x_* and h_* above, but with X replaced by $X^{[M]} = X\mathbb{I}_{\{X\le M\}}$.

Now we let $M \to \infty$ and observe that $\text{CPoi}\left(\mu x_*^{[M]}, h_*^{[M]}\right) \xrightarrow{w} \text{CPoi}\left(\mu x_*, h_*\right)$. Furthermore, the natural coupling $G_{(n)}^{[M]} \subset G_{(n)}$ implies $d^{[M]} \le d$, where d stands for the degree of v_1 in $G_{(n)}$. Moreover, $d \ne d^{[M]}$ implies that v_1 belongs to a layer of size greater than M. Hence, by the union bound,

$$\begin{aligned}\mathbb{P}\left(d \ne d_{[M]}\right) &\le \sum_{1\le j\le m} \mathbb{P}(v_1 \in D_j,\ X_{n,j} > M)\\ &= \sum_{1\le j\le m} \mathbb{E}\frac{X_{n,j}}{n}\mathbb{I}_{\{X_{n,j}>M\}} = \frac{m}{n}\mathbb{E}\left(X_{n,\pi}\mathbb{I}_{\{X_{n,\pi}>M\}}\right).\end{aligned}$$

Note that the quantity on the right is $o(1)$ uniformly in n as $M \to \infty$. Indeed our assumptions $P_n \xrightarrow{w} P$ and $(P)_{10} \to (P)_{10}$ imply $\lim_{M\to\infty} \sup_n \mathbb{E}X_{n,\pi}\mathbb{I}_{\{X_{n,\pi}>M\}} = 0$. Therefore, $\mathcal{L}(d) \xrightarrow{w} \text{CPoi}\left(\mu x_*, h_*\right)$.

7 | ANALYSIS OF CLUSTERING

Here we prove Theorems 3.2 and 3.3. In the proof we drop the subscript n when it does not cause an ambiguity. Thus we write $G = G_{(n)}$, $V = V(G_{(n)}) = \{1, 2, \ldots, n\}$ and $G_k = G_{n,k}$. Let K_{12} be the two-star on the vertex set $\{1,2,3\}$ with links $\{12, 13\}$. Let K_3 be the triangle on the vertex set $\{1,2,3\}$. Denote events $\mathcal{K}_3 = \{G \supset K_3\}$ and $\mathcal{K}_{12} = \{G \supset K_{12}\}$. We also denote $d = \deg_G(1)$. Theorems 3.2 and 3.3 are derived from Theorems 7.3 and 7.4 below, where we evaluate the probabilities $\mathbb{P}(\mathcal{K}_3)$, $\mathbb{P}(\mathcal{K}_{12})$,

$\mathbb{P}(d=t,\ \mathcal{K}_3)$ and $\mathbb{P}(d=t,\ \mathcal{K}_{12})$ defining the ratios $\tau^{(n)} = \frac{\mathbb{P}(\mathcal{K}_3)}{\mathbb{P}(\mathcal{K}_{12})}$ and $\sigma^{(n)}(t) = \frac{\mathbb{P}(d=t,\ \mathcal{K}_3)}{\mathbb{P}(d=t,\ \mathcal{K}_{12})}$. In the proof of Theorem 7.3 we approximate $\mathbb{P}(\mathcal{K}_3)$ by the probability that K_3 is produced by a single layer (with m different layers available). Similarly, in the proof of Theorem 7.4 $\mathbb{P}(\mathcal{K}_{12})$ is approximated by the sum of two probabilities: the first one being the probability that K_{12} is produced by a single layer (with m different layers available) and the second one being the probability that the edges of K_{12} are produced by different layers (with $m(m-1)$ layer pairs available). We note the corresponding sum in the denominator of (10).

Before the proof we introduce some notation and collect auxiliary results. Subgraph frequencies in the overlay graph will be characterized using cross moments

$$(P_n)_{rs} = \int (x)_r y^s \, dP_n, \qquad (P_n)_{rs,tu} = \int (x)_r y^s \ (x)_t y^u \, dP_n \tag{40}$$

of the averaged layer type distribution P_n defined by (2), and normalized cross moments defined by

$$\mu_{rs} = \sum_{k=1}^{m} p_{rs}(k), \qquad \mu_{rs,tu} = \sum_{k=1}^{m} p_{rs}(k)\ p_{tu}(k), \tag{41}$$

where $p_{rs}(k) = (n)_r^{-1}\mathbb{E}(X_{n,k})_r Q_{n,k}^s$. Note that $\mu_{rs} = m(n)_r^{-1}(P_n)_{rs}$.

Lemma 7.1. *Recall the averaged empirical distribution P_n defined by (2). If $P_n \xrightarrow{w} P$ and $(P_n)_{rs} \to (P)_{rs} < \infty$, then the cross moments defined in (40)–(41) satisfy $\mu_{10,rs} \ll m(n)_r^{-1}$ and $(P_n)_{10,rs} \ll n$.*

Proof. Denote $A_k = X_k$ and $B_k = (X_k)_r Q_k^s$. Observe that $A_k \le a + A_k \mathbb{I}(A_k > a)$ and $B_k \le b + B_k \mathbb{I}(B_k > b)$ for any $a, b > 0$. Because $A_k \le n$, we find that

$$\begin{aligned} A_k\ \mathbb{E}B_k &\le (a + A_k\mathbb{I}(A_k > a))\mathbb{E}B_k \\ &\le a\mathbb{E}B_k + bn\mathbb{I}(A_k > a) + n\mathbb{E}B_k\mathbb{I}(B_k > b). \end{aligned} \tag{42}$$

By taking expectations and averaging with respect to k, we find that

$$\frac{1}{m}\sum_{k=1}^{m}\mathbb{E}A_k\ \mathbb{E}B_k \le a\mathbb{E}B_\pi + bn\mathbb{P}(A_\pi > a) + n\mathbb{E}B_\pi\mathbb{I}(B_\pi > b), \tag{43}$$

where $A_\pi = X_\pi$, $B_\pi = (X_\pi)_r Q_\pi^s$, and (X_π, Y_π) is a generic P_n-distributed random variable. Because the left side above equals $m^{-1}n(n)_r\mu_{10,rs}$, we conclude that

$$m^{-1}(n)_r\mu_{10,rs} \le \frac{a}{n}c + b\phi(a) + \psi(b),$$

where $c = \sup_n (P_n)_{rs}$, $\phi(t) = \sup_n \int \mathbb{I}(x > t)dP_n$, and $\psi(t) = \sup_n \int (x)_r y^s \mathbb{I}((x)_r y^s > t)dP_n$. Then the tightness of P_n implies that $\phi(a_n) \to 0$ for $a_n = n^{1/2}$. Hence also $b_n\phi(a_n) \to 0$ for $b_n = \phi(a_n)^{-1/2} \to \infty$. The uniform $(x)_r y^s$-integrability of P_n further implies that $\psi(b_n) \to 0$. Hence the right side above vanishes and first claim follows.

For the second claim, we may repeat the above reasoning to verify that (42) holds also with the $\mathbb{E}$-symbol removed. Therefore, (43) also holds when the left side is replaced by $(P_n)_{10,rs} = \frac{1}{m}\sum_{k=1}^{m}\mathbb{E}A_k B_k$. Hence the second claim follows by the same argument. ∎

Let k^* be a random number uniformly distributed in $\{1, \dots, m\}$ and independent of $G = G_{(n)}$. In the following result $G_{k^*} = G_{n,k^*}$ represents a randomly chosen layer.

Lemma 7.2. *Let F_{rs} be a graph with node set in $\{1, \dots, n\}$ such that $|V(F_{rs})| = r$ and $|E(F_{rs})| = s$, and let i be a node in $V(F_{rs})$ with $\deg_{F_{rs}}(i) = r - 1$. Select $k^* \in \{1, \dots, m\}$ uniformly at random and independently of the layers. Then*:

(i) $\mathbb{P}(G_{k^*} \supset F_{rs}) = m^{-1}\mu_{rs}$,
(ii) $\mathbb{P}(\deg_{G_{k^*}}(i) = t \,|\, G_{k^*} \supset F_{rs}) = \mathrm{Bin}_{rs}(P_n)(t - r + 1)$ *for all* t.

Proof. (i) Because $\mathbb{P}(V(G_k) \supset V(F_{rs}) \,|\, X_k, Q_k) = \frac{(X_k)_r}{(n)_r}$ for any k, we see that $\mathbb{P}(G_k \supset F_{rs}) = \mathbb{E}\frac{(X_k)_r}{(n)_r}Q_k^s = p_{rs}(k)$. The corresponding probability for a randomly selected k^* equals $\mathbb{P}(G_{k^*} \supset F_{rs}) = \frac{1}{m}\sum_{k=1}^m p_{rs}(k) = (n)_r^{-1}(P_n)_{rs}$. Finally, recall that $(n)_r^{-1}(P_n)_{rs} = m^{-1}\mu_{rs}$.

(ii) Denote $d_k = \deg_{G_k}(i)$. On the event that $G_k \supset F_{rs}$, we see that $d_k = \ell + d'_k$ where $d'_k = |N_{G_k}(i) \setminus V(F_{rs})|$ and $\ell = r - 1$. Conditionally on $(X_k, Q_k) = (x, q)$ and $G_k \supset F_{rs}$, the random integer d'_k is $\mathrm{Bin}(x - r, q)$-distributed. Hence

$$\mathbb{P}\left(d_k = t,\ G_k \supset F_{rs}\right) = \mathbb{E}\left(\mathrm{Bin}(X_k - r, Q_k)(t - \ell)\frac{(X_k)_r}{(n)_r}Q_k^s\right).$$

The corresponding probability for a randomly chosen k^* is

$$\mathbb{P}(d_{k^*} = t,\ G_{k^*} \supset F_{rs}) = \int \left(\mathrm{Bin}(x - r, q)(t - \ell)\frac{(x)_r}{(n)_r}q^s\right) P_n(dx, dq),$$

so the claim follows by dividing both sides by $\mathbb{P}(G_{k^*} \supset F_{rs}) = (n)_r^{-1}(P_n)_{rs}$. ∎

Now we are ready to state and prove Theorems 7.3 and 7.4. We use the short hand notation $g_{rs}^{(n)} = \mathrm{Bin}_{rs}(P_n)$, where the mixed binomial distribution $\mathrm{Bin}_{rs}(P_n)$ is defined in (5).

Theorem 7.3. *We have*

(i) $|\mathbb{P}(G \supset K_3) - \mu_{33}| \le 4\mu_{21}\mu_{32} + \mu_{21}^3$.
(ii) $\mathbb{P}(\deg_G(1) = t,\ G \supset K_3) = \mu_{33}\, f^{(n)} * g_{33}^{(n)}(t - 2) + \varepsilon(t)$, *where* $f^{(n)}$ *is the model degree distribution defined by (7) and the approximation error is bounded by*

$$|\varepsilon(t)| \le (4 + t)\mu_{21}\mu_{32} + \mu_{21}^3 + 2\mu_{10,33}. \tag{44}$$

Proof. Denote by $\mathcal{A}_k = \{G_k \supset K_3\}$ the event that all node pairs of the triangle are linked by the layer k. We also denote $d_k = \deg_{G_k}(1)$, and $d_{-k} = \deg_{G_{-k}}(1)$ with $G_{-k} = \cup_{k' \neq k} G_{k'}$.

Proof of (i). Denote

$$\varepsilon_1(t) = \mathbb{P}(d = t, \mathcal{K}_3) - \mathbb{P}(d = t, \cup_k \mathcal{A}_k),$$

and observe that $0 \le \varepsilon_1(t) \le \mathbb{P}(d = t, \mathcal{E}_{12}) + \mathbb{P}(d = t, \mathcal{E}_{111})$, where $\mathcal{E}_{12}$ is the event that there exists one layer covering one link and a different layer covering the remaining two links of K_3, and $\mathcal{E}_{111}$ is the event that three distinct layers cover distinct links of K_3. We write $p(abc) = \mathbb{P}\big(\mathcal{G}_a^{12}, \mathcal{G}_b^{13}, \mathcal{G}_c^{23}\big)$, where

$\mathcal{G}_a^{ij}$ denotes the event that node pair ij is linked in layer a. We note that $p(abc) = p_{21}(a)p_{21}(b)p_{21}(c)$, $p(aab) = p_{32}(a)p_{21}(b)$, and $p(aaa) = p_{33}(a)$ for distinct layers a, b, c. Hence

$$\mathbb{P}(\mathcal{E}_{12}) \le \sum_{a,b}{}' (p(aab) + p(aba) + p(baa)) \le 3\mu_{21}\mu_{32},$$

and $\mathbb{P}(\mathcal{E}_{111}) \le \sum'_{a,b,c} p(abc) \le \mu_{21}^3$. Thus, $\sum_{t\ge 0} |\epsilon_1(t)| \le 3\mu_{21}\mu_{32} + \mu_{21}^3$.

Then denote

$$\epsilon_2(t) = \mathbb{P}(d = t, \cup_k \mathcal{A}_k) - \sum_k \mathbb{P}(d = t, \mathcal{A}_k).$$

Bonferroni's inequalities imply that $0 \le -\epsilon_2(t) \le \sum'_{k,k'} \mathbb{P}(d = t, \mathcal{A}_k, \mathcal{A}_{k'})$, and hence, noting that $\mu_{33} \le \mu_{32} \le \mu_{21}$,

$$\sum_{t\ge 0} |\epsilon_2(t)| \le \sum{}'_{k,k'} \mathbb{P}(\mathcal{A}_k, \mathcal{A}_{k'}) = \sum{}'_{k,k'} p_{33}(k)p_{33}(k') \le \mu_{33}^2 \le \mu_{21}\mu_{32}.$$

By combining this with the bound for $\epsilon_1(t)$, we conclude that

$$\mathbb{P}(d = t, \mathcal{K}_3) = \sum_k \mathbb{P}(d = t, \mathcal{A}_k) + \epsilon_1(t) + \epsilon_2(t), \tag{45}$$

where $\sum_{t\ge 0}(|\epsilon_1(t)| + |\epsilon_2(t)|) \le 4\mu_{21}\mu_{32} + \mu_{21}^3$. Hence claim (i) follows by summing the above equality over t, and noting that $\sum_k \mathbb{P}(\mathcal{A}_k) = \mu_{33}$.

Proof of (ii). We start with (45) and approximate

$$\begin{aligned}
\sum_k \mathbb{P}(d = t, \mathcal{A}_k) &\approx \sum_k \mathbb{P}(d_{-k} + d_k = t, \mathcal{A}_k) && (46)\\
&= \sum_k \sum_{r+s=t} \mathbb{P}(d_{-k} = r)\, \mathbb{P}(d_k = s, \mathcal{A}_k)\\
&\approx \sum_k \sum_{r+s=t} \mathbb{P}(d = r)\, \mathbb{P}(d_k = s, \mathcal{A}_k). && (47)
\end{aligned}$$

It follows from Lemma 7.2 that

$$\begin{aligned}
\sum_k \mathbb{P}(d_k = s, \mathcal{A}_k) &= m\mathbb{P}(d_{k*} = s, G_{k^*} \supset K_3) && (48)\\
&= m\mathbb{P}(d_{k^*} = s | G_{k^*} \supset K_3)\mathbb{P}(G_{k^*} \supset K_3) = \mu_{33}\mathrm{Bin}_{33}(P_n)(s-2).
\end{aligned}$$

Hence the last term above equals $\mu_{33}\, f^{(n)} * g_{33}^{(n)}(t-2)$, and to prove the claim it suffices to analyze the approximation errors in (46)–(47).

The approximation error in (46) equals $\epsilon_3(t) = \sum_k \epsilon_{3k}(t)$, where

$$\epsilon_{3k}(t) = \mathbb{P}(d = t, \mathcal{A}_k) - \mathbb{P}(d_{-k} + d_k = t, \mathcal{A}_k).$$

By applying Lemma 11.2 with $A = \{k\}$, $B = [n] \setminus \{k\}$, $\mathcal{E}_A = \mathcal{A}_k$, and $\mathcal{E}_B$ being the sure event, we see that $|\epsilon_{3k}(t)| \le c_B t\mathbb{P}(d_k \le t, \mathcal{A}_k) \le c_B t\mathbb{P}(\mathcal{A}_k)$, where $c_B = \mathbb{P}(G_{-k} \ni 12) \le \sum_{\ell\ne k} p_{21}(\ell) \le \mu_{21}$. Hence

$$|\epsilon_3(t)| \le t\mu_{21} \sum_k p_{33}(k) = t\mu_{21}\mu_{33} \le t\mu_{21}\mu_{32}.$$

The approximation error in (47) equals $\varepsilon_4(t) = \sum_k \varepsilon_{4k}(t)$ where

$$\varepsilon_{4k}(t) = \sum_{r+s=t} (\mathbb{P}(d=r) - \mathbb{P}(d_{-k}=r))\,\mathbb{P}(d_k = s, \mathcal{A}_k).$$

By Lemma 11.1, $\sum_{t\geq 0} |\varepsilon_{4k}(t)| \leq 2\mathbb{P}(d_k > 0)\mathbb{P}(\mathcal{A}_k)$. Because $\mathbb{P}(d_k > 0) \leq p_{10}(k)$ and $\mathbb{P}(\mathcal{A}_k) = p_{33}(k)$, it follows that $\sum_{t\geq 0} |\varepsilon_4(t)| \leq 2\mu_{10,33}$. Claim (ii) follows by combining the above estimates for the total approximation error $\varepsilon(t) = \varepsilon_1(t) + \varepsilon_2(t) + \varepsilon_3(t) + \varepsilon_4(t)$. ∎

Theorem 7.4. *We have*

(i) $|\mathbb{P}(G \supset K_{12}) - (\mu_{32} + \mu_{21}^2)| \leq 6\mu_{21}\mu_{32} + 6\mu_{21}^3 + \mu_{21}^4 + \mu_{21,21}$.
(ii) $\mathbb{P}(\deg_G(1) = t, G \supset K_{12}) = \mu_{32}\, f^{(n)} * g_{32}^{(n)}(t-2) + \mu_{21}^2 f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2) + \varepsilon(t)$, *where* $f^{(n)}$ *is the degree distribution of* G, *and the approximation error is bounded by*

$$|\varepsilon(t)| \leq (6+3t)(\mu_{21}\mu_{32} + \mu_{21}^3) + \mu_{21}^4 + 4\mu_{10,32} + 4\mu_{21}\mu_{10,21} + \mu_{21,21}. \tag{49}$$

Proof. Recall that K_{12} is the two-star with node set $\{1,2,3\}$ and link set $\{e_1, e_2\} := \{12, 13\}$. We denote by $\mathcal{G}_k^{ij}$ the event that $ij \in E(G_k)$ and we set $\mathcal{A}_{k\ell} = \mathcal{G}_k^{12} \cap \mathcal{G}_\ell^{13}$. We denote $G_{k\ell} = G_k \cup G_\ell$ and $G_{-k\ell} = \cup_{q \notin \{k,\ell\}} G_q$, and we set $d = \deg_G(1)$, $d_{k\ell} = \deg_{G_{k\ell}}(1)$ and $d_{-k\ell} = \deg_{G_{-k\ell}}(1)$. We also denote $h_{k\ell}(s) = \mathbb{P}(d_{k\ell} = s, \mathcal{A}_{k\ell})$.

We start with an outline of the proof. First we approximate

$$\begin{aligned}
\mathbb{P}(d=t, \mathcal{K}_{12}) &\approx \sum_{k,\ell} \mathbb{P}(d=t, \mathcal{A}_{k\ell}) && (50)\\
&\approx \sum_{k,\ell} \mathbb{P}(d_{k\ell} + d_{-k\ell} = t, \mathcal{A}_{k\ell}) && (51)\\
&= \sum_{k,\ell} \sum_{r+s=t} \mathbb{P}(d_{-k\ell} = r)\, h_{k\ell}(s) &&\\
&\approx \sum_{k,\ell} \sum_{r+s=t} \mathbb{P}(d=r)\, h_{k\ell}(s), && (52)
\end{aligned}$$

so that

$$\mathbb{P}(d=t, \mathcal{K}_{12}) \approx \sum_{r+s=t} f^{(n)}(r) \sum_k h_{kk}(s) + \sum_{r+s=t} f^{(n)}(r) \sideset{}{'}\sum_{k,\ell} h_{k\ell}(s). \tag{53}$$

We note that

$$\begin{aligned}
\sum_k h_{kk}(s) &= m\mathbb{P}\left(\deg_{G_{k^*}}(1) = s, \mathcal{K}_{12} \subset G_{k^*}\right) && (54)\\
&= m\mathbb{P}\left(\deg_{G_{k^*}}(1) = s | \mathcal{K}_{12} \subset G_{k^*}\right)\mathbb{P}(\mathcal{K}_{12} \subset G_{k^*}) = \mu_{32}\, g_{32}^{(n)}(s-2), &&
\end{aligned}$$

where the conditional probability is evaluated using Lemma 7.2. Hence the first term on the right of (53) equals

$$\sum_{r+s=t} f^{(n)}(r) \sum_k h_{kk}(s) = \mu_{32} f^{(n)} * g_{32}^{(n)}(t-2). \tag{55}$$

Next we approximate, denoting $h_k(s) = \mathbb{P}(d_k = s, \mathcal{G}_k^{12})$,

$$\begin{aligned}\sum_{k,\ell}{}' h_{k\ell}(s) &= \sum_{k,\ell}{}' \mathbb{P}\left(d_{k\ell} = s, \mathcal{G}_k^{12}, \mathcal{G}_\ell^{13}\right) \\ &\approx \sum_{k,\ell}{}' \mathbb{P}(d_k + d_\ell = s, \mathcal{G}_k^{12}, \mathcal{G}_\ell^{13}) \qquad (56)\\ &= \sum_{k,\ell}{}' \sum_{s_1+s_2=s} h_k(s_1)h_\ell(s_2) \\ &\approx \sum_{k,\ell} \sum_{s_1+s_2=s} h_k(s_1)h_\ell(s_2). \qquad (57)\end{aligned}$$

After noting (see Lemma 7.2) that $\sum_k h_k(s) = \mu_{21} g_{21}^{(n)}(s-1)$, we conclude that

$$\sum_{k,\ell} \sum_{s_1+s_2=s} h_k(s_1)h_\ell(s_2) = \mu_{21}^2 g_{21}^{(n)} * g_{21}^{(n)}(s-2),$$

and hence the second term on the right side of (53) is approximately

$$\sum_{r+s=t} f^{(n)}(r) \sum_{k,\ell}{}' h_{k\ell}(s) \approx \mu_{21}^2 f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2). \qquad (58)$$

By combining (53), (55) and (58), we conclude that

$$\mathbb{P}(d = t, \mathcal{K}_{12}) \approx \mu_{32} f^{(n)} * g_{32}^{(n)}(t-2) + \mu_{21}^2 f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2). \qquad (59)$$

The total approximation error in (59) can be written as $\varepsilon(t) = \varepsilon_1(t) + \varepsilon_2(t) + \varepsilon_3(t) + \varepsilon_4(t)$, where $\varepsilon_1(t), \varepsilon_2(t), \varepsilon_3(t)$ are the approximation errors in (50), (51), (52), respectively, and the approximation error in (58) equals

$$\varepsilon_4(t) = \sum_{r+s=t} f^{(n)}(r) \left(\varepsilon_{41}(s) + \varepsilon_{42}(s)\right), \qquad (60)$$

where $\varepsilon_{41}(s), \varepsilon_{42}(s)$ denote the errors made in (56), (57), respectively.

Now we give a rigorous proof of (i) and (ii), where we analyze the individual approximation errors one by one.

Proof of claim (i). The union bound shows that the approximation error $\varepsilon_1(t)$ in (50) is nonpositive for all t, and hence $\sum_{t\geq 0} |\varepsilon_1(t)| = \sum_{k,\ell} \mathbb{P}(\mathcal{A}_{k\ell}) - \mathbb{P}(\bigcup_{k,\ell} \mathcal{A}_{k\ell})$. Bonferroni's inequalities imply that

$$\sum_{t\geq 0} |\varepsilon_1(t)| \leq \sum_{(k_1,k_2),(\ell_1,\ell_2)}{}' \mathbb{P}(\mathcal{A}_{k_1k_2}, \mathcal{A}_{\ell_1\ell_2}) =: \Delta.$$

We split the right side above by $\Delta = \Delta_2 + \Delta_3 + \Delta_4$, where Δ_i, $i = 2, 3, 4$, is the sum over layer pairs $(k_1, k_2) \neq (\ell_1, \ell_2)$ such that the list $(k_1, k_2, \ell_1, \ell_2)$ contains precisely i distinct elements. Denote

$$p(k_1k_2\ell_1\ell_2) = \mathbb{P}(G_{k_1} \ni e_1, G_{k_2} \ni e_2, G_{\ell_1} \ni e_1, G_{\ell_2} \ni e_2).$$

Then

$$\Delta_2 = \sum_{a,b}{}' (p(aabb) + p(abba) + p(aaab) + p(aaba) + p(abaa) + p(baaa)),$$

$$\Delta_3 = \sum_{a,b,c}{}' (p(aabc) + p(abac) + p(abca) + p(baac) + p(baca) + p(bcaa)).$$

In the sum of Δ_2, the terms $p(aabb)$ and $p(abba)$ equal $p_{32}(a)p_{32}(b)$ and the other terms equal $p_{32}(a)p_{21}(b)$. Because $p_{32}(b) \leq p_{21}(b)$, it follows that $\Delta_2 \leq 6\sum'_{a,b} p_{21}(a)p_{32}(b) \leq 6\mu_{21}\mu_{32}$. In the sum of Δ_3, the terms $p(abac)$ and $p(baca)$ equal $p_{21}(a)p_{21}(b)p_{21}(c)$ and the other terms equal $p_{32}(a)p_{21}(b)p_{21}(c)$. Because $p_{32}(a) \leq p_{21}(a)$, it follows that $\Delta_3 \leq 6\mu_{21}^3$. Furthermore, $\Delta_4 = \sum'_{a,b,c,d} p(abcd) \leq \mu_{21}^4$. As a conclusion, it follows that

$$\sum_{t\geq 0} |\varepsilon_1(t)| \leq 6\mu_{21}\mu_{32} + 6\mu_{21}^3 + \mu_{21}^4. \tag{61}$$

Claim (i) now follows by combining the above bound with the equality

$$\sum_{k,\ell} \mathbb{P}(\mathcal{A}_{k\ell}) = \sum_k p_{32}(k) + \sum_{k,\ell}{}' p_{21}(k)p_{21}(\ell) = \mu_{32} + \mu_{21}^2 - \mu_{21,21}.$$

Proof of claim (ii). The approximation error in (51) equals $\varepsilon_2(t) = \sum_{k,\ell} \varepsilon_{2k\ell}(t)$ where

$$\varepsilon_{2k\ell}(t) = \mathbb{P}(d = t, \mathcal{A}_{k\ell}) - \mathbb{P}(d_{k\ell} + d_{-k\ell} = t, \mathcal{A}_{k\ell}).$$

By applying Lemma 11.2 with $A = \{k, \ell\}$, $B = [m] \setminus \{k, \ell\}$, $\mathcal{E}_A = \mathcal{A}_{k\ell}$, and $\mathcal{E}_B$ being the sure event, we see that

$$|\varepsilon_{2k\ell}(t)| \leq tc_B\mathbb{P}(d_{k\ell} \leq t, \mathcal{A}_{k\ell}) \leq tc_B\mathbb{P}(\mathcal{A}_{k\ell}),$$

where $c_B \leq \mathbb{P}(G_{-k\ell} \ni 12) \leq \mathbb{P}(G \ni 12) \leq \mu_{21}$. Hence

$$|\varepsilon_2(t)| \leq t\mu_{21} \sum_{k,\ell} \mathbb{P}(\mathcal{A}_{k\ell}) \leq t(\mu_{21}\mu_{32} + \mu_{21}^3). \tag{62}$$

The approximation error in (52) equals $\varepsilon_3(t) = \sum_{k,\ell} \varepsilon_{3k\ell}(t)$ where

$$\varepsilon_{3k\ell}(t) = \sum_{r+s=t} (\mathbb{P}(d = r) - \mathbb{P}(d_{-k\ell} = r))\, h_{k\ell}(s).$$

By applying Lemma 11.1 with $g(s) = \frac{h_{k\ell}(s)}{\mathbb{P}(\mathcal{A}_{k\ell})}$, it follows that $\sum_{t\geq 0} |\varepsilon_{3k\ell}(t)| \leq 2\mathbb{P}(\mathcal{A}_{k\ell})\mathbb{P}(d_{k\ell} > 0)$. Observe now that $\mathbb{P}(d_{k\ell} > 0) \leq p_{10}(k) + p_{10}(\ell)$. Hence,

$$\begin{aligned}
\sum_{t\geq 0} |\varepsilon_3(t)| &\leq 2\sum_{k,\ell} (p_{10}(k) + p_{10}(\ell))\, \mathbb{P}(\mathcal{A}_{k\ell}) \\
&\leq 4\sum_k p_{10}(k)p_{32}(k) + 4\sum_{k,\ell}{}' p_{10}(k)p_{21}(k)p_{21}(\ell) \\
&\leq 4\mu_{10,32} + 4\mu_{21}\mu_{10,21}.
\end{aligned} \tag{63}$$

The approximation error in (56) equals $\varepsilon_{41}(s) = \sum'_{k,\ell}\varepsilon_{4k\ell}(s)$ where

$$\varepsilon_{4k\ell}(s) \ = \mathbb{P}(d_{k\ell} = s, \mathcal{A}_{k\ell}) - \mathbb{P}(d_k + d_\ell = s, \mathcal{A}_{k\ell}).$$

By applying Lemma 11.2 with $A = \{k\}$ and $B = \{\ell\}$, together with $\mathcal{E}_A = \{12 \in G_k\}$ and $\mathcal{E}_B = \{13 \in G_\ell\}$, it follows that $|\varepsilon_{4k\ell}(s)| \le sp_{21}(k)p_{32}(\ell) + sp_{21}(\ell)p_{32}(k)$. By summing the above inequality with respect to k, ℓ, it follows that $|\varepsilon_{41}(s)| \le 2s\mu_{21}\mu_{32}$.

The approximation error in (57) equals

$$|\varepsilon_{42}(s)| = \sum_k \sum_{s_1+s_2=s} \mathbb{P}(d_k = s_1, \mathcal{G}_k^{12})\mathbb{P}(d_k = s_2, \mathcal{G}_k^{12}),$$

where $\sum_{s\ge0} |\varepsilon_{42}(s)| = \sum_k p_{21}(k)^2 = \mu_{21,21}$. Hence, by (60),

$$\begin{aligned} |\varepsilon_4(t)| \ &\le \sum_{r+s=t} f^{(n)}(r)\,(|\varepsilon_{41}(s)| + |\varepsilon_{42}(s)|) \le \max_{s\le t}|\varepsilon_{41}(s)| + \max_{s\le t}|\varepsilon_{42}(s)| \\ &\le 2t\mu_{21}\mu_{32} + \mu_{21,21}. \end{aligned} \tag{64}$$

Claim (ii) follows by collecting all the bounds in (61–64) together. ∎

Proof of Theorem 3.2. We evaluate the ratio $\tau^{(n)} = \frac{\mathbb{P}(\mathcal{K}_3)}{\mathbb{P}(\mathcal{K}_{12})}$. By Theorems 7.3 and 7.4,

$$\begin{aligned} \mathbb{P}(\mathcal{K}_3) \ &= \mu_{33} + O\left(\mu_{21}\mu_{32} + \mu_{21}^3\right), \\ \mathbb{P}(\mathcal{K}_{12}) \ &= \mu_{32} + \mu_{21}^2 + O\left(\mu_{21}\mu_{32} + \mu_{21}^3 + \mu_{21}^4 + \mu_{21,21}\right), \end{aligned}$$

where $\mu_{rs} = m(n)_r^{-1}(P_n)_{rs}$. Now relations $(P_n)_{21} \lesssim 1$ and $(P_n)_{32} \lesssim 1$ imply $\mu_{21} \lesssim m/n^2$ and $\mu_{32} \lesssim m/n^3$. Hence $\mu_{21}\mu_{32} \lesssim m^2n^{-5}$, $\mu_{21}^3 \lesssim m^3n^{-6}$, and $\mu_{21}^4 \lesssim m^4n^{-8}$. Next, we note that $\mu_{21,21} \le m(n)_2^{-2}(P_n)_{21,21}$ by Jensen's inequality. Note also that $((x)_2q)^2 \le 2x(x)_3q^2$ for $x \ge 3$. Hence, $((x)_2q)^2 \le 4 + 2x(x)_3q^2$, and $(P_n)_{21,21} \le 4 + 2(P_n)_{10,32}$. Furthermore, Lemma 7.1 implies that $(P_n)_{10,32} \ll n$. Hence $\mu_{21,21} \ll mn^{-3}$.

(i) Consider the case $\frac{m}{n} \to \mu \in [0, \infty)$. Then $\mu_{32} = ((P)_{32} + o(1))\,mn^{-3}$ and $\mu_{21}^2 = \left(\mu(P)_{21}^2 + o(1)\right)mn^{-3}$ imply that

$$\mathbb{P}(\mathcal{K}_{12}) \ = (P)_{32}mn^{-3} + \mu(P)_{21}^2mn^{-3} + o\left(mn^{-3}\right).$$

Similarly, $\mu_{33} = ((P)_{33} + o(1))\,n^{-3}m$ implies

$$\mathbb{P}(\mathcal{K}_3) = (P)_{33}mn^{-3} + o(mn^{-3}),$$

and hence the first two claims of Theorem 3.2 follow (the first claim corresponds to $\mu = 0$).

(ii) Assume now that $n \ll m \ll n^2$. Then $mn^{-3}, m^2n^{-5}, m^3n^{-6} \ll m^2n^{-4}$. Hence, $\mathbb{P}(\mathcal{K}_3) \ll m^2n^{-4}$. Furthermore, $m^4n^{-8} \ll m^2n^{-4}$, and we conclude that $\mathbb{P}(\mathcal{K}_{12}) = (P)_{21}^2m^2n^{-4} + o(m^2n^{-4})$. Hence, $\frac{\mathbb{P}(\mathcal{K}_3)}{\mathbb{P}(\mathcal{K}_{12})} \to 0$ implies the third claim of Theorem 3.2. ∎

Proof of Theorem 3.3. We evaluate the ratio $\sigma^{(n)}(t) = \frac{\mathbb{P}(d=t,\ \mathcal{K}_3)}{\mathbb{P}(d=t,\ \mathcal{K}_{12})}$. By Theorems 7.3 and 7.4,

$$\begin{aligned} \mathbb{P}(d = t,\ \mathcal{K}_3) \ &= \mu_{33}\, f^{(n)} * g_{33}^{(n)}(t-2) + \varepsilon_A(t), \\ \mathbb{P}(d = t,\ \mathcal{K}_{12}) \ &= \mu_{32}\, f^{(n)} * g_{32}^{(n)}(t-2) + (\mu_{21})^2 f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2) + \varepsilon_B(t), \end{aligned}$$

where the remainder terms $\varepsilon_A(t)$ and $\varepsilon_B(t)$ are upper bounded in (44) and (49) respectively. The conditions $m/n \to \mu$ and $(P_n)_{rs} \to (P)_{rs}$ imply $\mu_{rs} = (\mu + o(1))(P)_{rs} n^{1-r}$ for $rs = 21, 32, 33$. Invoking these relations in (44), (49) and using the bounds $\mu_{10,33} \ll n^{-2}$, $\mu_{10,32} \ll n^{-2}$ and $\mu_{21,21} \ll n^{-2}$ (the latter bound is shown in the proof of Theorem 3.2 above) we upper bound the remainders $\varepsilon_A(t) \ll n^{-2}$ and $\varepsilon_B(t) \ll n^{-2}$. We obtain

$$\begin{aligned}
\mu_{33} f^{(n)} * g_{33}^{(n)}(t-2) &= (P)_{33} \mu n^{-2} f^{(n)} * g_{33}^{(n)}(t-2) + o(n^{-2}), \\
\mu_{32} f^{(n)} * g_{32}^{(n)}(t-2) &= (P)_{32} \mu n^{-2} f^{(n)} * g_{32}^{(n)}(t-2) + o(n^{-2}), \\
\mu_{21}^2 f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2) &= (P)_{21}^2 \mu^2 n^{-2} f^{(n)} * g_{21}^{(n)} * g_{21}^{(n)}(t-2) + o(n^{-2}).
\end{aligned}$$

Now the claim follows by the fact that $f^{(n)} \xrightarrow{w} f = \mathrm{CPoi}(\mu(P)_{10}, g_{10})$ (see Theorem 3.1) and $g_{rs}^{(n)} \xrightarrow{w} g_{rs}$ for $rs = 21, 32, 33$. ■

8 | CLUSTERING AND DEGREE IN POWER-LAW MODELS

Here we prove Theorems 4.1 and 4.2. In the proof we use the fact that a compound Poisson distribution $\mathrm{CPoi}(\lambda, h)$ is heavy-tailed whenever the increment distribution h is [23, Theorem 4.30]. Namely, for a subexponential distribution h we have as $t \to +\infty$

$$\mathrm{CPoi}(\lambda, h)(t) \sim \lambda h(t). \tag{65}$$

Furthermore, we show below that for P satisfying (15), (16) the mixed binomial distribution (5) follows a power law

$$\mathrm{Bin}_{rs}(P)(t) \;\sim\; d_{rs} L(t^{1/(1-\beta)}) t^{-\delta_{rs}}, \tag{66}$$

with parameters

$$\delta_{rs} = 1 + \frac{\alpha + s\beta - r - 1}{1-\beta} \qquad \text{and} \qquad d_{rs} = \frac{b^s}{(P)_{rs}} \frac{b^{\delta_{rs}-1}}{1-\beta}. \tag{67}$$

It is an immediate consequence of (65) and (66) that the limiting degree distribution $f = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(P))$ obeys a power law (17). We similarly establish respective power law asymptotics for the distributions $f * g_{33}, f * g_{32}$ and $f * g_{21} * g_{21}$ that appear in (10). Now a simple analysis of the ratio (10) as $t \to +\infty$ shows the asymptotics of Theorem 4.2.

Proof of Theorem 4.1. Statements (i) and (ii) are immediate consequences of (65), (66) as described above. To prove (iii) we note that $\beta \geq 1$ implies $XQ = Xg(X) \leq B$ almost surely. Denote for short $\lambda = \mu(P)_{10}$ and let H be a random variable with the distribution $\mathrm{Bin}_{10}(P)$ and Λ be a Poisson random variable with parameter λ. We have

$$\sum_{t \geq 0} e^{st} f(t) = \mathbb{E}\left(\mathbb{E} e^{sH}\right)^{\Lambda} = e^{(\mathbb{E}e^{sH}-1)\lambda},$$

where

$$\mathbb{E}e^{sH} = \sum_{h \geq 0} e^{sh} \mathbb{E}\left(\binom{X-1}{h} Q^h (1-Q)^{X-1-h} \frac{X}{(P)_{10}}\right)$$

$$= \mathbb{E}\left((1+Q(e^s-1))^{X-1} \frac{X}{(P)_{10}} \right)$$
$$\leq \mathbb{E}\left(e^{Q(X-1)(e^s-1)} \frac{X}{(P)_{10}} \right).$$

The almost sure inequality $QX \leq B$ together with the identity $(P)_{10} = \mathbb{E}X$ complete the proof. ∎

Proof of Theorem 4.2. Theory of discrete subexponential densities [23, Lemmas 4.9 and 4.14], implies that $(f_1 * f_2)(t) \sim f_1(t) + f_2(t)$ for all probability densities on the positive integers such that $f_i(t) \sim a_i t^{-\alpha_i}$ with $a_i > 0$ and $\alpha_i > 1$. By Theorem 4.1, we know that $f(t) \sim \mu(P)_{10} d_{10} t^{-\delta_{10}}$, and by (66), we find that $g_{rs}(t) = \mathrm{Bin}_{rs}(P)(t) \sim d_{rs} t^{-\delta_{rs}}$ with parameters given by (67). Because $\delta_{32} < \delta_{21} < \delta_{10}$, it follows that

$$(f * g_{32})(t) \sim f(t) + g_{32}(t) \sim g_{32}(t)$$

and

$$(f * g_{21} * g_{21})(t) \sim f(t) + g_{21}(t) + g_{21}(t) \ll g_{32}(t).$$

Hence by formula (10),

$$\sigma(t) \sim \frac{(P)_{33}}{(P)_{32}} \frac{f(t) + g_{33}(t)}{g_{32}(t)} \sim \frac{(P)_{33}}{(P)_{32}} \frac{\mu(P)_{10} d_{10} t^{-\delta_{10}} + d_{33} t^{-\delta_{33}}}{d_{32} t^{-\delta_{32}}}.$$

Because $\delta_{33} - \delta_{10} = \frac{3\beta-2}{1-\beta}$, we see that $\sigma(t)$ follows a power law with density exponent $\delta_{33} - \delta_{32} = \frac{\beta}{1-\beta}$ for $\beta \leq \frac{2}{3}$, and density exponent $\delta_{10} - \delta_{32} = 2$ for $\beta \geq \frac{2}{3}$. The constant term of the power law is determined by (67). ∎

Let us prove (66). Given $r \in \mathbb{Z}_+$, let $H_{(r)}$ be a mixed binomial random variable with the distribution

$$\mathbb{P}(H_{(r)} = t) = \mathbb{E}\left(\mathrm{Bin}(X - r, g(X))(t)\right), \qquad t = 0, 1, 2, \ldots , \tag{68}$$

where the distribution $p(x) = \mathbb{P}(X = x)$ of X and function g are given by (16). In Lemma 8.1 below we show that the distribution of $H_{(r)}$ follows a power law

$$\mathbb{P}(H_{(r)} = t) \sim \frac{b^{\frac{\alpha-1}{1-\beta}}}{1-\beta} L\left(t^{1/(1-\beta)}\right) t^{-1-\frac{\alpha-1}{1-\beta}} \quad \text{as} \quad t \to +\infty. \tag{69}$$

Next we observe that the distribution $\mathrm{Bin}_{rs}(P)$ can be written in the form

$$\mathrm{Bin}_{rs}(P)(t) = \mathbb{E}\left(\mathrm{Bin}(\tilde{X} - r, g(\tilde{X})(t)\right) = \sum_{x=0}^{\infty} \mathrm{Bin}(x - r, g(x))(t)\, \tilde{p}_{rs}(x),$$

where the random variable $\tilde{X}$ has the power law distribution

$$\mathbb{P}(\tilde{X} = x) = \tilde{p}_{rs}(x) = \frac{(x)_r g(x)^s p(x)}{(P)_{rs}} \sim \frac{b^s}{(P)_{rs}} L(x) x^{-(\alpha+s\beta-r)}.$$

Hence (69) yields (66). It remains to prove (69). In the proof we use the fact that binomial distribution is highly concentrated around its mean. We apply this fact to mixed binomial random variable $H_{(r)}$ conditionally, given the mixing random variable X.

Lemma 8.1. *Let $\alpha \geq 1$, $0 \leq \beta < 1$ and $b > 0$. Assume that (15), (16) hold. For $\beta = 0$ we assume, in addition, that $b \leq 1$. Then for each $r \in \mathbb{Z}_+$ (69) holds.*

Proof of Lemma 8.1. For $\beta = 0$, $b = 1$ relation (69) follows from the identity $\mathbb{P}(H_{(r)} = t) = \mathbb{P}(X = t + r)$, which holds for large t. Indeed, for large t the second relation of (16) implies $q(x) \equiv 1$ for $x \geq t$. Hence $\text{Bin}(x - r, g(x))(t) = \mathbb{I}_{\{x-r=t\}}$, and thus $\mathbb{P}(H_{(r)} = t) = \mathbb{P}(X = t + r)$.

In what follows we assume that either $0 < \beta < 1$ or $\beta = 0$ and $b < 1$. We only consider the case where $r = 0$. The proof for $r \geq 1$ is much the same.

In the proof limits are taken as $t \to +\infty$. Let $H_k \sim Bin(k, g(k))$ be a binomial random variable. We use the shorthand notation

$$\mu_k = \mathbb{E}H_k = kg(k), \qquad \sigma_k^2 = \text{Var}H_k = kg(k)(1 - g(k)), \qquad \tilde{p}_k = \mathbb{P}(X = k).$$

Given t, let $\delta_t = t^{1/2}\ln^4 t$. We split the probability

$$\mathbb{P}(H_{(0)} = t) = \sum_{k \geq t} \mathbb{P}(H_k = t)\tilde{p}_k = I_1 + I_2 + I_3, \qquad I_j := \sum_{k \in A_j} \mathbb{P}(H_k = t)\tilde{p}_k,$$

where

$$A_1 = \{k \geq t : \ \mu_k < t - \delta_t\}, \quad A_2 = \{k : \ |t - \mu_k| \leq \delta_t\}, \quad A_3 = \{k : \ \mu_k > t + \delta_t\}.$$

We assume that t is large enough so that $g(k) = bk^{-\beta}$ for $k > t$. Then $\mu_k = bk^{1-\beta}$ and $A_1 \neq \emptyset$. To prove (69) we show that

$$I_2 = \frac{1}{1 - \beta} b^{\frac{\alpha-1}{1-\beta}} L\left(t^{1/(1-\beta)}\right) \ t^{-1-\frac{\alpha-1}{1-\beta}}(1 + o(1)) \qquad \text{and} \qquad I_1, \ I_3 = O(e^{-0.5\ln^8 t}). \tag{70}$$

Let us evaluate I_2. By the local limit theorem [39], [50] we approximate uniformly in $k \in A_2$

$$\mathbb{P}(H_k = t) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(t-\mu_k)^2}{2\sigma_k^2}} (1 + o(1)). \tag{71}$$

Furthermore, we have $\tilde{p}_k = L\left(t^{1/(1-\beta)}\right)(b/t)^{\alpha/(1-\beta)}(1 + o(1))$ uniformly in $k \in A_2$. In what follows we consider the cases $0 < \beta < 1$ and $\beta = 0$ separately.

For $0 < \beta < 1$ we have $\sigma_k^2 = \mu_k(1 - g(k)) = t(1 + O(\delta_t/t) + O(t^{-\beta/(1-\beta)}))$ uniformly in $k \in A_2$. Now (71) implies

$$I_2 = (1 + o(1)) \sum_{k \in A_2} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(t-\mu_k)^2}{2\sigma_k^2}} \tilde{p}_k = (1 + o(1)) \frac{S}{\sqrt{2\pi t}} L\left(t^{1/(1-\beta)}\right)(b/t)^{\alpha/(1-\beta)},$$

where $S = \sum_{k \in A_2} e^{-\frac{(t-\mu_k)^2}{2t}}$. Next we approximate

$$S = I + O(1), \qquad I = \int_{|t-by^{1-\beta}|\le \delta_t} e^{-(t-by^{1-\beta})^2/(2t)} dy$$

and use the substitution $x = (by^{1-\beta} - t)/\sqrt{t}$ to write the integral in the form

$$I = \frac{1}{1-\beta} b^{-\frac{1}{1-\beta}} t^{\frac{1}{1-\beta}-\frac{1}{2}} I_1, \qquad I_1 := \int_{|x|<\ln^4 t} e^{-x^2/2}\left(1 + xt^{-1/2}\right)^{\frac{\beta}{1-\beta}} dx.$$

Now it is easily seen that the integral I_1 converges to $\sqrt{2\pi}$ as $t \to +\infty$. Hence,

$$S = (1 + o(1))\frac{1}{1-\beta} b^{-\frac{1}{1-\beta}} t^{\frac{1}{1-\beta}-\frac{1}{2}} \sqrt{2\pi} \qquad \text{as} \quad t \to +\infty.$$

We have arrived to the first relation of (70).

For $\beta = 0$ and $b < 1$ we have $\sigma_k^2 = t(1-b)(1 + O(\delta_t/t))$. Now (71) implies

$$I_2 = \frac{S'}{\sqrt{2\pi t(1-b)}} L(t)(b/t)^\alpha (1 + o(1)), \qquad \text{where} \qquad S' = \sum_{k \in A_2} e^{-\frac{(t-\mu_k)^2}{2t(1-b)}}.$$

Invoking the approximation $S' = I' + O(1)$, where

$$I' = \int_{|t-by|\le \delta_t} e^{-(t-by)^2/(2t(1-b))} dy = \sqrt{2\pi t(1-b)}\, b^{-1}(1 + o(1)),$$

we obtain the first relation of (70).

We derive the bounds on the right of (70) from the upper bounds

$$\mathbb{P}(H_k = t) \le e^{-0.5\ln^8 t}(1 + o(1)). \tag{72}$$

that hold uniformly in $k \in A_1 \cup A_3$. Indeed, (72) follows from the well-known exponential inequalities for Binomial probabilities see for example, Theorem 2.1 of [30],

$$\mathbb{P}(H_k \ge \mu_k + s) \le e^{-\frac{s^2}{2(\mu_k + s/3)}}, \qquad \mathbb{P}(H_k \le \mu_k - s) \le e^{-\frac{s^2}{2\mu_k}}, \qquad s > 0. \tag{73}$$

We only show (72) for $k \in A_3$. Let $\overline{\mu} = \min\{\mu_k : k \in A_3\}$. The function $h(x) = (x-t)^2 x^{-1}$ is increasing for $x > t$. Hence $h(\mu_k) \ge h(\overline{\mu})$ for $k \in A_3$. The second inequality of (73) implies

$$\mathbb{P}(H_k = t) \le \mathbb{P}(H_k \le \mu_k - (\mu_k - t)) \le e^{-0.5h(\mu_k)} \le e^{-0.5h(\overline{\mu})} = e^{-0.5\ln^8 t}(1 + o(1)).$$

■

9 | GIANT COMPONENT

Here we prove Theorem 3.4. We start with an outline of the proof. In the proof we apply the approach developed in [15]: we approximate the number $N_1(G_{(n)})$ of vertices in the largest connected component of $G_{(n)}$ by the number of vertices u with the property that the breadth first search (BFS) tree rooted at u contains at least ω vertices, where $\omega = \omega(n) \to +\infty$ as $n \to +\infty$. Then the BFS exploration process rooted at u is approximated by respective branching process and the fraction of vertices u having large BFS trees (of size at least ω) is approximated by the survival probability $\rho = \rho_{(f^+)}$ of the branching process. Hence $N_1(G_{(n)})/n \xrightarrow{\mathbb{P}} \rho$. We briefly comment on the branching process, which is much different from that of [15]. Let us consider the BFS exploration process at the moment when it enters a new layer, say, $G_{n,i}$. The first vertex of $G_{n,i}$ detected by BFS will be included into the BFS tree together with the component of $G_{n,i}$ the vertex belongs to. The other members of the component are called children of the vertex. Since the vertex may belong to several layers, it can have children from the other layers as well. Consequently, the total number of children is approximated by the sum of degrees of the vertex in the transitive closures $\overline{G}_{n,i}$ of layers $G_{n,i}$ that cover this vertex. Accordingly, the offspring distribution is approximated by f^+. We note that layers do not need to be connected. Each layer $G_{n,i}$ may split into components, but for moderately growing $\omega = \omega(n)$ the BFS exploration process will not visit the same layer twice within the first $O(\omega)$ steps with high probability. Therefore the offspring numbers are approximately independent.

The proof is organized as follows. We start with Lemma 9.1 which establishes the result in the special case where the layer types are deterministic, that is, $\forall n\ \mathbb{P}((\overline{X}_n, \overline{Q}_n) = (\overline{x}_n, \overline{q}_n)) = 1$ for some $\overline{x}_n = (x_{n,1}, \dots, x_{n,m})$ and $\overline{q}_n = (q_{n,1}, \dots, q_{n,m})$. Furthermore, we assume that the distribution of (X, Q) has a finite support, say, $A \subset \{0, 1, 2, \dots, M\} \times [0, 1]$, where $M \geq 2$ is an integer. Hence, $\mathbb{P}(X = t, Q = q) > 0$ whenever $(t, q) \in A$. Moreover, we assume in Lemma 9.1 that the set

$$A^0 = \{(t, q) : \ t \geq 2,\ q > 0,\ (t, q) \in A\}$$

is nonempty and

$$\forall\, n \geq 1 \qquad \forall\, 1 \leq i \leq m_n \qquad (x_{n,i}, q_{n,i}) \in A. \tag{74}$$

Note that for each $(x_{n,i}, q_{n,i}) \in A \setminus A^0$ respective layer $G_{n,i}$ has no edges. Next, in Lemma 9.2 we relax condition (74) by allowing a negligible fraction of layer types $(x_{n,i}, q_{n,i})$ to take their values outside A. In the last step of the proof we reduce the general case to that considered in Lemma 9.2. To this aim we truncate the layer sizes $X_{n,i}$ (at level M) and ϵ-discretize the edge densities $Q_{n,i}$ as in (20). Then we apply Lemma 9.2 conditionally given the truncated and discretized layers $\left(\overline{X}_n^{[M]}, \overline{Q}_n^{\pm}\right)$. Here we use notation (21) and for $\overline{x}_n = (x_{n,1}, \dots, x_{n,m})$ we denote $\overline{x}_n^{[M]} = (x_{n,1}^{[M]}, \dots, x_{n,m}^{[M]})$, where $x^{[M]} = x\mathbb{I}_{\{x \leq M\}}$. Finally, letting $\epsilon \downarrow 0$ and $M \uparrow \infty$ we approximate the distribution of (X, Q) and survival probability ρ by respective characteristics of $(X^{[M]}, Q^{\pm})$ thus completing the proof of Theorem 3.4.

Notation. Before the proof we introduce some notation. Given a Galton-Watson (G-W) branching process $\mathcal{X}$ we denote by $|\mathcal{X}|$ the total progeny of $\mathcal{X}$, $\rho^{(k)}(\mathcal{X}) = \mathbb{P}(|\mathcal{X}| \geq k)$ and $\rho(\mathcal{X}) = \mathbb{P}(|\mathcal{X}| = \infty)$. Let $T(t, q) = \deg_{\overline{H}_{t,q}}(\ell)$ be the degree of a randomly selected vertex ℓ in the transitive closure $\overline{H}_{t,q}$ of Bernoulli random graph $H_{t,q}$ on t vertices and with edge density q. Let $\mathcal{T} = \{T_s(t, q), T_s^{(j)}(t, q) : \ q \in [0, 1], j, s, t \in \mathbb{N}\}$ be a collection of independent random variables such that $T_s(t, q)$ and $T_s^{(j)}(t, q)$ have the same distribution as $T(t, q)$.

Let $C_v \subset V$ be the vertex set of the connected component of $G = G_{(n)} = G_{(\overline{X}_n, \overline{Q}_n)}$ containing vertex $v \in V$. Let $\omega : \mathbb{N} \to \mathbb{N}$ be a function satisfying $\omega(n) \to \infty$, $\omega(n) = o(n)$ as $n \to +\infty$. Let $B^k = \{v : \ |C_v| \geq k\} \subset V$ be the set of vertices belonging to connected components of G of size at least k. We write $B^\omega = B^{\omega(n)}$. Let $C \subset V$ denote the vertex set of the largest connected component of G. Note that for each integer $k \geq 1$

$$|C| \leq \max\{k, |B^k|\}. \tag{75}$$

In Lemma 9.1 below we assume that the distribution of (X, Q) has a finite support A and denote

$$q_0 = \min\{q : (t,q) \in A^0\} \qquad \text{and} \qquad h_{t,q} = \mathbb{P}(X = t, Q = q), \quad (t,q) \in A.$$

Given $0 < \delta < 1/4$, let $\mathcal{Y}_\delta^+, \mathcal{Y}_\delta^-, \mathcal{Y}$ be G-W processes with the offspring numbers

$$Y_\delta^+ = \sum_{(t,q)\in A^0} \sum_{s=1}^{\Lambda_{t,q}^+} T_s(t,q), \quad Y_\delta^- = \sum_{(t,q)\in A^0} \sum_{s=1}^{\Lambda_{t,q}^-} T_s(t,q), \quad Y = \sum_{(t,q)\in A^0} \sum_{s=1}^{\Lambda_{t,q}} T_s(t,q). \tag{76}$$

Here $\Lambda_{t,q} \sim \text{Poi}(\lambda_{t,q})$ and $\Lambda_{t,q}^\pm \sim \text{Poi}(\lambda_{t,q}^\pm)$ with $\lambda_{t,q} = th_{t,q}\mu$ and $\lambda_{t,q}^\pm = \lambda_{t,q}(1 \pm \delta)$. Furthermore, we assume that the collection of random variables $\{\Lambda_{t,q}, \Lambda_{t,q}^+, \Lambda_{t,q}^-, (t,q) \in A^0\}$ and $\mathcal{T}$ are independent. Note that offspring numbers (76) have compound Poisson distributions. In particular, Y has the probability distribution $\text{CPoi}(\lambda, \mathcal{L}(T_*))$, where $\lambda = \mu x_*$ with $x_* = \mathbb{E}(X\mathbb{I}_{\{X\geq 2\}})$ and the random variable T_* has the probability distribution

$$\begin{aligned}\mathbb{P}(T_* = \ell) &= x_*^{-1}\mathbb{E}\left(\mathbb{P}(T(X,Q) = \ell | X, Q) X\mathbb{I}_{\{X\geq 2\}}\right) \\ &= x_*^{-1}\mathbb{E}\left(\text{Bin}^+(X-1, Q)(\ell) X\mathbb{I}_{\{X\geq 2\}}\right), \qquad \ell = 0, 1, \dots .\end{aligned} \tag{77}$$

By the same reasoning as in (30) above, we obtain the equality of distributions

$$\text{CPoi}(\lambda, \mathcal{L}(T_*)) = \text{CPoi}(\mu(P)_{10}, \text{Bin}_{10}^+(P)), \tag{78}$$

where the increment distribution $\text{Bin}_{10}^+(P)$ is defined by (6).

We recall that D_i denotes the vertex set of layer $G_{n,i}$, $1 \leq i \leq m$. With a little abuse of notation we also refer to D_i as the layer $G_{n,i}$. Furthermore, we assign label D_i to the edges of $G_{n,i}$. In particular, an edge of G may receive several labels if it is present in several layers. Given n and $(t,q) \in A$, let $\mathbb{D}_{t,q}$ be the collection of layers D_i having size $x_{n,i} = t$ and edge density $q_{n,i} = q$. Put $\mathbb{D}_0 = \cup_{(t,q)\in A^0} \mathbb{D}_{t,q}$. For every $D_i \in \mathbb{D}_0$ the probability that a randomly chosen vertex of D_i has a neighbor in D_i connected by an edge labeled D_i is

$$1 - (1 - q_{n,i})^{x_{n,i}-1} \geq \min\left\{1 - (1-q)^{t-1} : (t,q) \in A^0\right\} =: \hat{q}.$$

Note that $\hat{q} \geq q_0$.

Lemma 9.1. *Let $\mu > 0$. Let $M \geq 2$ be an integer. Let $n, m \to +\infty$. Assume that $m/n \to \mu$. Let $\varepsilon_n \downarrow 0$ be a positive sequence. Assume that $\mathbb{P}(X \leq M) = 1$ and (X, Q) has a finite support denoted by A. Assume that $A^0 = \{(t,q) \in A : t \geq 2, q > 0\}$ is nonempty. Assume that (74) holds and the numbers*

$$m_{t,q} := \#\left\{i \in \{1, \dots, m\} : (x_{n,i}, q_{n,i}) = (t,q)\right\}$$

satisfy

$$\forall\, n \qquad \max_{(t,q)\in A^0} \left| h_{t,q} - m_{t,q}/m \right| \leq \varepsilon_n. \tag{79}$$

Assume that $\omega(n) \leq n \ln^{-2} n$ and

$$\forall n \qquad |m/n - \mu| < \varepsilon_n. \tag{80}$$

There exists sequences $\varepsilon_n' \downarrow 0, \varepsilon_n'' \downarrow 0$ (depending on $\{\varepsilon_n\}, \omega$ and μ, A^0, $\{h_{t,q}, (t,q) \in A^0\}$) such that for each n we have

$$\max_{v\in V}\left|\mathbb{P}(|C_v| \ge \omega(n)) - \rho(\mathcal{Y})\right| \le \varepsilon_n', \tag{81}$$

$$\mathbb{P}\left(\left||B^{\omega}| - n\rho(\mathcal{Y})\right| > \varepsilon_n'' n\right) \le \varepsilon_n'. \tag{82}$$

There exists sequences $\varepsilon_n' \downarrow 0, \varepsilon_n'' \downarrow 0$ (depending on $\{\varepsilon_n\}$, μ, A^0, $\{h_{t,q}, (t,q) \in A^0\}$) such that for each n we have

$$\mathbb{P}\left(\left||C| - n\rho(\mathcal{Y})\right| > \varepsilon_n'' n\right) \le \varepsilon_n'. \tag{83}$$

We note that $A^0 \neq \emptyset$ implies $q_0 > 0$ and $\hat{q} > 0$.

Proof of Lemma 9.1. We recall that the idea of the proof is outlined in the first paragraph of the section above. Proofs of (81), (82), (83) are given in separate steps.

Proof of (81). Note that the distribution of $|C_v|$ is the same for each $v \in V$. Hence, it suffices to approximate $\mathbb{P}(|C_v| \ge \omega(n))$ for $v = v_1$.

Before the proof we introduce some notation. Given $0 < \delta < 4^{-1}$ we denote

$$m_{t,q}(\delta) = m h_{t,q}(1-\delta) \quad \text{and} \quad p_{t,\delta}^- = t(1-\delta)n^{-1}. \tag{84}$$

We assume that n, m are large enough so that $m_{t,q} > m_{t,q}(\delta)$, for $(t,q) \in A_0$. Let $\omega' : \mathbb{N} \to \mathbb{N}$ be such that $\omega'(n) = o(\sqrt{n})$, $\omega'(n) \to +\infty$, $\omega' \le \omega$. We write, for short, $\rho^{(\omega)} = \rho^{(\omega(n))}(\mathcal{Y})$, $\rho^{(\omega\prime)} = \rho^{(\omega\prime(n))}(\mathcal{Y})$. Let

$$\mathcal{N}^+ = \{N_{t,q}^+,\ 2 \le t \le M\}, \qquad \mathcal{N}^- = \{N_{t,q}^-,\ 2 \le t \le M\},$$
$$\tilde{\mathcal{N}}^+ = \{\tilde{N}_{t,q}^+,\ 2 \le t \le M\}, \qquad \tilde{\mathcal{N}}^- = \{\tilde{N}_{t,q}^-,\ 2 \le t \le M\}$$

be collections of independent random variables having binomial and Poisson distributions:

$$N_{t,q}^+ \sim \mathrm{Bin}\left(m_{t,q}, t/(n-\omega'(n))\right), \qquad N_{t,q}^- \sim \mathrm{Bin}(m_{t,q}(2\delta), p_{t,\delta}^-), \tag{85}$$

$$\tilde{N}_{t,q}^+ \sim \mathrm{Poi}\left(t m_{t,q}/(n-\omega'(n))\right), \qquad \tilde{N}_{t,q}^- \sim \mathrm{Poi}\left(m_{t,q}(2\delta)p_{t,\delta}^-\right). \tag{86}$$

Note that $\mathbb{E}N_{t,q}^+ = \mathbb{E}\tilde{N}_{t,q}^+$ and $\mathbb{E}N_{t,q}^- = \mathbb{E}\tilde{N}_{t,q}^-$. We assume that $\mathcal{T}$ is independent of $\mathcal{N}^-, \mathcal{N}^+, \tilde{\mathcal{N}}^-, \tilde{\mathcal{N}}^+$. Let $Z^{\pm}$ (respectively $\tilde{Z}^{\pm}$) be defined as $Y_{\delta}^{\pm}$ in (76), but with $\Lambda_{t,q}^{\pm}$ replaced by $N_{t,q}^{\pm}$ (respectively $\tilde{N}_{t,q}^{\pm}$). Let $\mathcal{X}^{\pm}$ and $\tilde{\mathcal{X}}^{\pm}$ be Galton-Watson processes with the offspring numbers $Z^{\pm}$ and $\tilde{Z}^{\pm}$ respectively.

Using the total variation distance bound $d_{tv}(\mathrm{Bin}(n,p), \mathrm{Poi}(np)) \le p$, see (1.23) in [6], we show by coupling the offspring numbers of $\mathcal{X}^{\pm}$ and $\tilde{\mathcal{X}}^{\pm}$ that

$$\left|\rho^{(k)}(\mathcal{X}^-) - \rho^{(k)}\left(\tilde{\mathcal{X}}^-\right)\right| \le Mkn^{-1}, \tag{87}$$
$$\left|\rho^{(k)}\left(\mathcal{X}^+\right) - \rho^{(k)}\left(\tilde{\mathcal{X}}^+\right)\right| \le Mk(n-\omega(n))^{-1}.$$

From (79), (80), (87) we obtain for $k = k(n) = o(n)$ as $m, n \to +\infty$

$$\rho^{(k)}\left(\mathcal{Y}_{3\delta}^-\right) \le \rho^{(k)}\left(\tilde{\mathcal{X}}^-\right) \le \rho^{(k)}(\mathcal{X}^-) + Mkn^{-1} = \rho^{(k)}(\mathcal{X}^-) + o(1), \tag{88}$$

$$\rho^{(k)}\left(\mathcal{Y}_{2\delta}^{+}\right) \geq \rho^{(k)}\left(\tilde{\mathcal{X}}^{+}\right) \geq \rho^{(k)}\left(\mathcal{X}^{+}\right) - Mk(n-\omega(n))^{-1} \geq \rho^{(k)}\left(\mathcal{X}^{+}\right) - o(1). \tag{89}$$

To show the first inequality of (88) we couple the offspring numbers $Y_{3\delta}^{-} \leq \tilde{Z}^{-}$. The coupling is feasible whenever $\lambda_{t,q}(1-3\delta) \leq m_{t,q}(2\delta)p_{t,q}^{-}$ $\forall (t,q) \in A^0$ or, equivalently, $\mu(1-3\delta) \leq (m/n)(1-3\delta+2\delta^2)$. In view of (80) the latter inequality holds true for sufficiently large n, say, $n > n_*$, where n_* depends on δ and $\{\varepsilon_n\}$. Similarly, the first inequality of (89) holds whenever $\lambda_{t,q}(1+2\delta) \geq tm_{t,q}/(n-\omega'(n))$ $\forall (t,q) \in A^0$. In view of (79), (80) the latter inequality holds for sufficiently large n, say, $n > n_{**}$, where n_{**} depends on δ and $\{\varepsilon_n\}$.

Furthermore, we show in (101), (109) below that

$$\rho^{(\omega)}(\mathcal{X}^{-}) - o(1) \leq \mathbb{P}(|C_v| \geq \omega(n)) \leq \mathbb{P}(|C_v| \geq \omega'(n)) \leq \rho^{(\omega')}(\mathcal{X}^{+}) + o(1) \tag{90}$$

(the second inequality follows by $\omega \geq \omega'$). We note that $o(1)$ on the left of (90) depends on δ, ω, M, μ, $|A^0|$, and $\{\varepsilon_n\}$, while $o(1)$ on the right depends on ω', M and $\hat{q}$. (88), (89), (90) imply

$$\rho^{(\omega)}\left(\mathcal{Y}_{3\delta}^{-}\right) - o(1) \leq \mathbb{P}(|C_v| \geq \omega(n)) \leq \rho^{(\omega')}\left(\mathcal{Y}_{2\delta}^{+}\right) + o(1). \tag{91}$$

Finally, letting $\delta \downarrow 0$ we obtain as $n \to +\infty$

$$\mathbb{P}(|C_v| \geq \omega(n)) = \rho(\mathcal{Y}) + o(1), \tag{92}$$

where the remainder $o(1)$ depends on M, A^0, $\{h_{t,q}, (t,q) \in A^0\}$, μ, $\{\varepsilon_n\}$ and function ω. Indeed the lower bound of (92) follows from $\rho^{(\omega)}(\mathcal{Y}_{3\delta}^{-}) \geq \rho(\mathcal{Y}_{3\delta}^{-}) \to \rho(\mathcal{Y})$ as $\delta \downarrow 0$. More precisely, given $\tau > 0$ we choose $\delta_\tau > 0$ such that $\delta \leq \delta_\tau$ implies $\rho(\mathcal{Y}_{3\delta}^{-}) \geq \rho(\mathcal{Y}) - \tau$ and then letting $n \to \infty$ we apply the left inequality of (91) to get

$$\mathbb{P}(|C_v| \geq \omega(n)) \geq \rho^{(\omega)}(\mathcal{Y}_{3\delta}^{-}) - \tau \geq \rho(\mathcal{Y}) - 2\tau.$$

For the upper bound we push $\rho^{(k)}(\mathcal{Y}_{2\delta}^{+})$ arbitrarily close to $\rho(\mathcal{Y})$ choosing large k and small δ. Indeed, given $\tau > 0$ we find large $k_\tau > 0$ such that $\rho^{(k_\tau)}(\mathcal{Y}) \leq \rho(\mathcal{Y}) + \tau$. Next, we find small δ_τ' such that $\delta \leq \delta_\tau'$ implies $\rho^{(k_\tau)}(\mathcal{Y}_{2\delta}^{+}) \leq \rho^{(k_\tau)}(\mathcal{Y}) + \tau$. For $\delta \leq \delta_\tau'$ and $k > k_\tau$ we have

$$\rho^{(k)}\left(\mathcal{Y}_{2\delta}^{+}\right) \leq \rho^{(k_\tau)}\left(\mathcal{Y}_{2\delta}^{+}\right) \leq \rho^{(k_\tau)}(\mathcal{Y}) + \tau \leq \rho(\mathcal{Y}) + 2\tau. \tag{93}$$

Now letting $n \to \infty$ we apply the right inequality of (91) to get

$$\mathbb{P}\left(|C_v| \geq \omega(n)\right) \leq \rho^{(\omega\prime)}\left(\mathcal{Y}_{2\delta}^{+}\right) + \tau \leq \rho(\mathcal{Y}) + 2\tau.$$

Finally, we note that (92) implies (81). It remains to prove the first and the last inequality of (90).

Proof of (90). We fix an order $v_1 < v_2 < \cdots < v_n$ of elements of V. Let $m' = \sum_{(t,q)\in A^0} m_{t,q}$ be the number of sets in the collection $\mathbb{D}_0$. We can assume without loss of generality that $\mathbb{D}_0 = \{D_1, \ldots, D_{m'}\}$ and $|D_1| \leq |D_2| \leq \cdots \leq |D_{m'}|$.

Upper bound (*the last inequality of* (90)). Given $v \in V$, define the list L_v of vertices using a BFS type exploration procedure. In the beginning all vertices are uncolored, all sets $D_i \in \mathbb{D}_0$ are unmarked, and $L_v = \emptyset$. After a vertex is added to L_v the vertex is colored white. We add v to the list. Next we proceed recursively. We choose the oldest (with respect to inclusion to L_v) white vertex, say u, from L_v.

For $i = 1, 2, \dots, m'$ such that $u \in D_i$ and D_i is not marked, we mark D_i (we say that D_i is marked by u) and add to L_v (in increasing order) all uncolored vertices of D_i that are connected to u by paths of edges labelled D_i. We say that D_i brings these vertices to the list and attach label D_i to each of them. Afterwards we color u black. Vertices added to L_v in this step are called children of u. We then chose the oldest white vertex from L_v, add to L_v its children and color this vertex black etc. We stop when there are no more white vertices in L_v or there are no more unmarked sets D_i left. We denote $L_v = \{u_1, u_2, \dots\}$, where elements are listed in the order of their inclusion to the list (u_i is older than u_j for $i < j$ and $u_1 = v$). We denote $L_{v,k} = \{u_1, \dots, u_k\}$ the set of k oldest vertices of L_v. Note that L_v is a subset of C_v. For any $u_i \in L_v$ with $i \geq 2$ there is unique $i^* \in [1, i)$ such that u_i is a child of u_{i^*} (equivalently, u_{i^*} is the parent of u_i). While constructing the list L_v we keep track of the sets $D_{i_1}, D_{i_2}, \dots$ that have been marked one after another (D_{i_s} was marked before D_{i_t} for $s < t$). For $u_j \in L_v$ the number $r = r(j)$ tells us that u_j was brought to the list by D_{i_r}, the rth member of the sequence $\{D_{i_1}, D_{i_2}, \dots\} =: \mathbb{D}_v$. A set D_{i_s} marked by $u \in L_v$ is called void if u has no neighbors in D_{i_s} linked to u by edges labeled D_{i_s} (in this case D_{i_s} brings no children to u). Note that any D_{i_j} is void with probability at most $1 - \hat{q}$. A set $D_{i_s} \in \mathbb{D}_v$ is called regular if $\cup_{j=1}^{s-1} D_{i_j}$ and D_{i_s} intersect in a single point. Vertex v is called k-regular if $|L_v| \geq k$ and each D_{i_j} is regular for $j = 2, 3, \dots, r(k)$. The set of k-regular vertices of G is denoted $V_k = \{v \in V : v \text{ is } k-\text{regular}\}$. Note that the events $\{|C_v| \geq k, v \in V_k\}$ and $\{|L_v| \geq k, v \in V_k\}$ are equivalent.

We observe that the number of vertices brought to the list L_v by a regular set $D_{i_s} \in \mathbb{D}_{t,q}$ has the same distribution as $T(t, q)$. For a nonregular set this number may be smaller, since white vertices of a nonregular set D_{i_s} that have been colored in previous steps of the exploration cannot be brought to L_v by D_{i_s}. Therefore as long as $k \leq \omega'(n)$ a coupling of the exploration process with the branching process $\mathcal{X}^+$ shows that

$$\mathbb{P}(|L_v| \geq k) \leq \mathbb{P}(|\mathcal{X}^+| \geq k). \tag{94}$$

Next we show that

$$\mathbb{P}(|L_v| \geq k, v \notin V_k) \leq \hat{k}^2 M^2 n^{-1} + 2k^{-1}, \tag{95}$$

where $\hat{k} := 2k/\hat{q}$. For v with $|L_v| \geq k$ the event $\{v \notin V_k\}$ implies that one or more nonregular sets have been marked during the exploration. Then either the first marked nonregular set D_{i_s} has index s satisfying $s \leq \hat{k}$ (we denote this event $\mathcal{A}_k$) or we have $s > \hat{k}$. In the latter case there are at least $\hat{k} - k + 2$ void sets D_{i_l} with $l \leq \hat{k}$ (this event we denote $\mathcal{B}_k$). Indeed, on the event $\overline{\mathcal{A}}_k \cap \{v \notin V_k\} \cap \{|L_v| \geq k\}$ we have that the index s of the first observed nonregular set D_{i_s} satisfies $\hat{k} < s \leq r(k)$. But the inequality $\hat{k} < r(k)$ implies that among the first $\hat{k}$ sets from $\mathbb{D}_v$ there are less than $k - 1$ nonvoid ones as each nonvoid set contributes at least one new vertex to the list. Now (95) follows from the inequalities

$$\mathbb{P}(\mathcal{B}_k) \leq \mathbb{P}(Y < k - 1) \leq 2k^{-1}, \tag{96}$$

$$\mathbb{P}(\mathcal{A}_k) \leq \sum_{2 \leq s \leq \hat{k}} \mathbb{P}(D_{i_s} \text{ is non regular}) \leq (\hat{k} - 1)\frac{(\hat{k} - 1)M^2}{n}. \tag{97}$$

Here, $Y \sim \text{Bin}(\hat{k}, \hat{q})$ and (96) follows by Chebyshev's inequality. In (97) we estimated $\mathbb{P}(D_{i_s}$ is non regular$) \leq (\hat{k} - 1)M^2/(n - (\hat{k} - 1))$. Indeed, given $H_{s-1} = \cup_{1 \leq \ell \leq s-1} D_{i_\ell}$, the size $|D_{i_s}| = t$ and the event that D_{i_s} is marked by u_j, the probability that D_{i_s} is non regular is the conditional probability

$p^* = \mathbb{P}\left(|H_{s-1} \cap D^*| \geq 2 \ |u_j \in D^*\right)$, where D^* is a random subset of size t of the set $V \setminus \{u_1, \dots, u_{j-1}\}$. For $|H_{s-1}| = h$ we have

$$p^* = \frac{\mathbb{P}(|H_{s-1} \cap D^*| \geq 2, u_j \in D^*)}{\mathbb{P}(u_j \in D^*)} \leq \frac{(h-j)(t-1)}{n-j}. \tag{98}$$

The last fraction upper bounds the probability that $D^* \setminus \{u_j\}$ of size $t-1$ intersects with $H_{s-1} \setminus \{u_1, \dots, u_j\}$ of size $h-j$. Note that $(h-j)/(n-j) \leq h/n$ and $h \leq (\hat{k}-1)M$ and $t \leq M$. Therefore the right side of (98) is at most $(\hat{k}-1)M^2/n$. This shows (97) and we arrive to (95). It follows from (95) that for $k = k_n \to +\infty$

$$k = o(\sqrt{n}) \ \Rightarrow \ \mathbb{P}(|L_v| \geq k, v \notin V_k) = o(1). \tag{99}$$

We similarly show that

$$k = o(\sqrt{n}) \ \Rightarrow \ \mathbb{P}(|C_v| \geq k, v \notin V_k) = o(1). \tag{100}$$

Namely, given v with $|C_v| \geq k$, the event $v \notin V_k$ implies that either $|L_v| < k$ or $|L_v| \geq k$ and D_{i_j} is nonregular for some $2 \leq j \leq r(k)$. The probability of the latter event is bounded by (99). Now we show that the remaining event $\mathcal{C}_k := \{|C_v| \geq k, |L_v| < k\}$ has probability $\mathbb{P}(\mathcal{C}_k) = o(1)$. We observe[3] that event $\mathcal{C}_k$ implies that a nonregular set D_{i_s} has been marked by some $u_r \in L_v$, where $r < k$. Next we consider two alternatives: either the index s of the first marked nonregular set D_{i_s} satisfies $s \leq \hat{k}$ (the probability of such event is upperbounded in (97) and is $o(1)$) or $s > \hat{k}$. But the inequality $s > \hat{k}$ implies that at most $k-1$ elements of the list L_w have marked at least $\hat{k} - k + 2$ void sets before a nonregular set was marked. The probability of such event is upperbounded by (96) and is $o(1)$.

Finally, we observe that the events $\{|C_v| \geq k, v \in V_k\}$ and $\{|L_v| \geq k, v \in V_k\}$ are equal. Now (99), (100) combined with (94) imply

$$\mathbb{P}(|C_v| \geq k) = \mathbb{P}(|L_v| \geq k) + o(1) \leq \mathbb{P}(|\mathcal{X}^+| \geq k) + o(1). \tag{101}$$

Lower bound (the first inequality of (90)). We modify a bit our exploration procedure. Given $v \in V$, we construct the list $L_v^* = \{u_1, u_2, \dots\}$ similarly as L_v above, but now each $u_j \in L_v^*$ only accepts children brought by regular sets. Moreover, not every regular set is allowed to contribute to the list L_v^*. Permission to contribute is granted at random. The construction of L_v^* is described in the algorithm $\mathbb{A}$, which uses a slightly modified definition of regular set.

In the algorithm $\mathbb{A}$ we use the following notation. $D_1^*, D_2^*, \dots$ denote the regular marked sets that were allowed to contribute to the list L_v^* one after another during the exploration; $H_s^* = \{u_1\} \cup \left(\cup_{1 \leq l \leq s} D_l^*\right)$, $s \geq 1$. We set $H_0^* = \{u_1\}$. Furthermore, $M^{(1)}, M^{(2)}, \dots$ denote the numbers of sets marked by $u_1, u_2, \dots \in L_v^*$ respectively; $M_{t,q}^{(j)}$ denotes the number of sets from $\mathbb{D}_{t,q}$ marked by u_j (so that $M^{(j)} = \sum_{(t,q) \in A^0} M_{t,q}^{(j)}$). For each $(t,q) \in A^0$ we define the integer sequence $m_{t,q}^{(j)} = m_{t,q}(\delta) - (j-1)\lfloor 3 \ln m \rfloor$, $j \geq 1$. For integers h, t we denote, see (84),

$$p^*(h,t,j) = \frac{p_{t,\delta}^-}{p_1^*(h,t,j)}, \qquad p_1^*(h,t,j) = \binom{n-h}{t-1}\binom{n-j+1}{t}^{-1}. \tag{102}$$

[3]Indeed, we always have $L_v \subset C_v$. The opposite, $L_v \neq C_v$, may happen if, for example, some $u_r \in L_v$ marks a nonregular set D_{i_s} containing a white vertex u_{r+j} (which is already on the list L_v) and vertices u_{r+j} and u_r belong to distinct components of the layer corresponding to D_{i_s}. The set D_{i_s} once marked by u_r will not be allowed to bring children to u_{r+j}.

Algorithm $\mathbb{A}$

1. for $(t,q) \in A^0$ set $\tilde{\mathbb{D}}_{t,q}^{(0)} := \mathbb{D}_{t,q}$;
2. $L_v^* \leftarrow v$, color$[v]$ = white, $j \leftarrow 0$;
3. while L_v^* contains a white vertex:
4. $j \leftarrow j+1$, $u_j \leftarrow$ the oldest white vertex of L_v^*, color$[u_j]$ = black,
5. for $(t,q) \in A^0$:
6. select a subset $\mathbb{D}_{t,q}^{(j)} \subset \tilde{\mathbb{D}}_{t,q}^{(j-1)}$ of size $|\mathbb{D}_{t,q}^{(j)}| = m_{t,q}^{(j)}$,
7. examine each $D \in \mathbb{D}_{t,q}^{(j)}$: mark D whenever $u_j \in D$, and if D is regular[4] then accept D with probability $p^*(h,t,j)$, where $t = |D|$ and where $h = |H_s^*|$ refers to the union H_s^* of regular sets accepted so far. Color white the children of u_j brought by D and insert them to the list L_v^*. Set $\tilde{\mathbb{D}}_{t,q}^{(j)}$ to be the family of unmarked sets from the collection $\mathbb{D}_{t,q}^{(j)}$. Stop if either (a) $|L_v^*| = k$ or (b) $M_{t,q}^{(j)} > \lfloor 3\ln m \rfloor$.

Let us show that for any sequence $k = k_n \to +\infty$, $k_n \le n\ln^{-2} n$ we have

$$\mathbb{P}(|L_v^*| \ge k) \ge \mathbb{P}(|\mathcal{X}^-| \ge k) + o(1). \tag{103}$$

More precisely, we show that there exist integer n'_* depending on δ and $\{h_{t,q}, (t,q) \in A^0\}$ such that for $n > n'_*$ we have $\mathbb{P}(|L_v^*| \ge k) \ge \mathbb{P}(|\mathcal{X}^-| \ge k) + r_{*n}$, where the sequence $r_{*n} = o(1)$ may depend on M, μ, $\{\varepsilon_n\}$ and $|A^0|$.

We firstly show that the acceptance probability $p^*(h,t,j)$ of step 7 is well defined, that is, for sufficiently large n, m it is always less than 1. Indeed, while deciding whether to accept a regular set D in step 7 we know that $|L_v^*| < k$. Hence, the number of sets marked so far is at most $\sum_{u_j \in L_v^*} M^{(j)} < k|A^0| 3\ln m$ as each $M^{(j)}$ contribute at most $|A^0| 3 \ln m$. The number of marked regular accepted sets is even less. Hence $h \le Mk|A^0|3\ln m = O(n\ln^{-1} n)$. Here we used the fact that the layer sizes $|D_l^*| \le M$, the set A^0 is finite and $m/n \to \mu \in (0,+\infty)$. Now for $j = O(n\ln^{-2} n)$ and $2 \le t \le M$ we have

$$p_1^*(h,t,j) = t\frac{\left(n - O(n\ln^{-1} n)\right)_{t-1}}{\left(n - O(n\ln^{-2} n)\right)_t} = (1+o(1))\frac{t}{n} > (1-\delta)\frac{t}{n} = p_{t,\delta}^-.$$

In the inequality above we use $n > n'_*$. A similar reasoning shows that each time we perform step 6 the collection $\tilde{\mathbb{D}}_{t,q}^{(j-1)}$ has more than $m_{t,q}(2\delta)$ members. Indeed, we have for $j < n\ln^{-2} n$

$$|\tilde{\mathbb{D}}_{t,q}^{(j)}| = |\mathbb{D}_{t,q}^{(j)}| - M_{t,q}^{(j)} \ge m_{t,q}^{(j)} - \lfloor 3\ln m \rfloor \ge m_{t,q}(\delta) - j\lfloor 3\ln m \rfloor \ge m_{t,q}(2\delta).$$

In the last inequality above we use $n > n'_*$.

We secondly show (see (105), (106) below) that the probability that algorithm $\mathbb{A}$ stops for the reason (b) is negligibly small. Introduce the event

$$\mathcal{E}_r = \{M_{t,q}^{(j)} > 3\ln m, \text{ for some } 1 \le j \le r \text{ and some } (t,q) \in A^0\}.$$

If the algorithm did not stop before u_j started exploring its children in the layers of size t and strength q, then $M_{t,q}^{(j)}$ has binomial distribution Bin $\left(m_{t,q}^{(j)}, t/(n-j+1)\right)$. In this case we have for some constant c depending on μ and M

$$\mathbb{P}(M_{t,q}^{(j)} > \lfloor 3\ln m \rfloor) \le cm^{-2}. \tag{104}$$

[4] A set D marked by u_j is called regular if u_j is the only common vertex shared by D and the union of previously marked and accepted regular sets.

To show (104) we couple $M_{t,q}^{(j)}$ with binomial random variable $M_j^* \sim \text{Bin}(m, t/(n-j+1))$ so that $\mathbb{P}\left(M_{t,q}^{(j)} \leq M_j^*\right) = 1$. Then we apply exponential Markov inequality $\mathbb{P}(M_j^* > x) \leq e^{-x}\mathbb{E}e^{M_j^*}$ with $x = \lfloor 3\ln m \rfloor$ and use the bound[5] $\mathbb{E}e^{M_j^*} \leq c$, where c depends on μ, M and the sequence $\{\epsilon_n\}$. If the algorithm stops before u_j starts exploring its children in the layers of size t and strength q, then we set $M_{t,q}^{(j)} \equiv 0$. In the latter case (104) is obvious. For $r \leq k$ we obtain from (104) by the union bound that

$$\mathbb{P}(\mathcal{E}_r) \leq \mathbb{P}(\mathcal{E}_k) \leq c|A^0|m^{-1} = O(m^{-1}). \tag{105}$$

By $i^*(k)$ and $r^*(k)$ we denote the positive integers such that $u_k \in L_v^*$ is a child of $u_{i^*(k)} \in L_v^*$ and u_k is brought to the list by the set $D^*_{r^*(k)}$. Observe that $r^*(k) \leq M_1 + \cdots + M_{i^*(k)}$ and $i^*(k) < k$. Therefore (105) implies

$$\mathbb{P}(\overline{\mathcal{E}}_{i^*(k)}) = 1 - O(m^{-1}), \qquad \mathbb{P}\left(|L_v^*| \geq k \mid \overline{\mathcal{E}}_{i^*(k)}\right) = \mathbb{P}(|L_v^*| \geq k) + O(m^{-1}). \tag{106}$$

Here, the second relation follows from the first one. The conditioning on $\overline{\mathcal{E}}_{i^*(k)}$ means that the algorithm has not stopped for the reason (b).

We now are ready to prove (103). We consider the probability $\mathbb{P}\left(|L_v^*| \geq k \mid \overline{\mathcal{E}}_{i^*(k)}\right)$. We claim that as long as the algorithm does not stop for the reason (b), we have for each $1 \leq j \leq i^*(k)$, each $(t,q) \in A^0$ and each $D \in \mathbb{D}_{t,q}^{(j)}$ that the probability that D is marked by u_j, D is regular and it is allowed to contribute to L_v^* is $p^-_{t,\delta}$. To show this we examine the probability $p^*(h,t,j)$ of step 7. Note that $p_1^*(h,t,j)$ is the probability that given $H \subset W$ of size $|H| = h$ and $u_1, \ldots, u_j \in H$, a random subset $D \subset V \setminus \{u_1, \ldots, u_{j-1}\}$ of size t intersect with H and the intersection $D \cap H = \{u_j\}$ (i.e., $p_1^*(h,t,j)$ is the probability that D is marked by u_j and it is regular). The random acceptance of D with probability $p^*(h,t,j)$ (in step 7) makes the final acceptance probability equal $p^-_{t,\delta}$. Now we can write the total number of children of u_j, $1 \leq j \leq i^*(k)$, in the form

$$\sum_{(t,q)\in A^0} \sum_{s=1}^{\eta_{t,q}^{(j)}} \mathbb{I}_s^{(j)}(t,q) T_s^{(j)}(t,q). \tag{107}$$

Here, $\mathbb{I}_s^{(j)}(t,q)$ is a Bernoulli random variable (independent of all the other random variables) with success probability

$$p'_{j,\delta} := \frac{p^-_{t,\delta}}{\mathbb{P}(D \text{ is marked by } u_j)} = \frac{p^-_{t,\delta}}{t/(n-j+1)} = (1-\delta)\frac{n-j+1}{n}$$

(note that $p'_{j,\delta}$ is the conditional probability that $D \in \mathbb{D}_{t,q}^{(j)}$ is allowed to contribute to L_w^* given that D is marked by u_j). Furthermore, by $\eta_{t,q}^{(j)}$ we denote the random variable $M_{t,q}^{(j)}$ conditioned on the event $M_{t,q}^{(j)} \leq \lfloor 3\ln m \rfloor$.

Let us compare the exploration process L_v^* (conditioned on the event $\overline{\mathcal{E}}_{i^*(k)}$) with the branching process $\mathcal{L}$, which produces an ordered list of particles $\{u_1, u_2, \ldots\}$ and where the offspring number

[5] We have $\mathbb{E}e^{M_j^*} = \left(1 + (e-1)\frac{t}{n-j-1}\right)^m \leq e^{m\frac{t(e-1)}{n-n\ln^{-2}n}} \leq e^{m\frac{M(e-1)}{n-n\ln^{-2}n}} \leq e^{3M\frac{m}{n}}$, for $n \geq 5$.

of u_j is defined by (107), but with $\eta_{t,q}^{(j)}$ replaced by $M_{t,q}^{(j)}$. Note that the total variation distance between their distributions $\mathcal{L}(\eta_{t,q}^{(j)})$ and $\mathcal{L}(M_{t,q}^{(j)})$

$$d_{\rm tv}\left(\mathcal{L}\left(\eta_{t,q}^{(j)}\right),\mathcal{L}\left(M_{t,q}^{(j)}\right)\right)\le \mathbb{P}\left(M_{t,q}^{(j)}>3\ln m\right)\le cm^{-2}.$$

Here, the second inequality is shown in (104) and the first one follows from the inequality[6] $|\mathbb{P}(A|B)-\mathbb{P}(A)|\le \mathbb{P}(\overline{B})$, which holds for any events A,B with $\mathbb{P}(B)>0$. Hence, we have

$$\mathbb{P}\left(|L_v^*|\ge k\mid \overline{\mathcal{E}}_{i^*(k)}\right)=\mathbb{P}(|\mathcal{L}|\ge k)+O(k/m^2). \tag{108}$$

Furthermore, we have $\mathbb{P}(|\mathcal{L}|\ge k)\ge \mathbb{P}(|\mathcal{X}^-|\ge k\}$. Indeed, we can represent the offspring number of $\mathcal{L}$ as

$$\sum_{(t,q)\in A^0}\sum_{s=1}^{M_{t,q}^{(j)}}\mathbb{I}_s^{(j)}(t,q)T_s^{(j)}(t,q)=\sum_{(t,q)\in A^0}\sum_{s=1}^{\overline{M}_{t,q}^{(j)}}T_s^{(j)}(t,q),$$

where $\overline{M}_{t,q}^{(j)}\sim \mathrm{Bin}(m_{t,q}^{(j)},p_{t,\delta}^-)$, and then couple $\overline{M}_{t,q}^{(j)}$ with $N_{t,q}^-$ so that $\mathbb{P}\left(\overline{M}_{t,q}^{(j)}\ge N_{t,q}^-\right)=1$. Now (106), (108) imply (103). Finally, (103) together with the simple inequality $\mathbb{P}(|C_v|\ge k)\ge \mathbb{P}(|L_v^*|\ge k)$ shows

$$\mathbb{P}(|C_v|\ge k)\ge \mathbb{P}(|\mathcal{X}^-|\ge k\}+o(1). \tag{109}$$

Proof of (82). We use the shorthand notation $\mathbb{I}_v:=\mathbb{I}_{\{|C_v|\ge\omega(n)\}}$. We have

$$|B^\omega|=\sum_{v\in V}\mathbb{I}_v,\qquad \binom{|B^\omega|}{2}=\sum_{\{u,v\}\subset V}\mathbb{I}_u\mathbb{I}_v. \tag{110}$$

The first identity combined with (81) yield

$$\mathbb{E}|B^\omega|=n\rho(\mathcal{Y})+o(n). \tag{111}$$

For $\rho(\mathcal{Y})=0$ this implies (82). For $\rho(\mathcal{Y})>0$ we establish (82) by showing that $|B^\omega|$ concentrates around its mean $\mathbb{E}|B^\omega|$.

We first consider the special case of $\omega=\overline{\omega}$, where $\overline{\omega}(n)=\ln n$. Let $\{x,y\}\subset V$ denote a pair of vertices selected uniformly at random. We show below that

$$\mathbb{E}(\mathbb{I}_x\mathbb{I}_y)\le \rho(\mathcal{Y})\times\rho(\mathcal{Y})+o(1), \tag{112}$$

where the remainder $o(1)$ only depends on M, A^0 and $\{\varepsilon_n\}$. (112) combined with (110), (111) imply $\mathbb{E}|B^{\overline{\omega}}|^2\le(\mathbb{E}|B^{\overline{\omega}}|)^2+o(n^2)$. From the latter inequality we conclude that $\mathrm{Var}|B^{\overline{\omega}}|=o(n^2)$. Now Chebyshev's inequality implies

$$\forall\gamma>0\quad \mathbb{P}\left\{\left||B^{\overline{\omega}}|-\mathbb{E}|B^{\overline{\omega}}|\right|>\gamma n\right\}\le(\gamma n)^{-2}\mathrm{Var}(|B^{\overline{\omega}}|)=o(1). \tag{113}$$

Letting $\gamma\downarrow 0$ we obtain (82). It remains to prove (112).

[6]For $\Delta:=\mathbb{P}(A|B)-P(A)>0$ we have $\Delta\le\mathbb{P}(A|B)-\mathbb{P}(A,B)=\mathbb{P}(A|B)\mathbb{P}(\overline{B})\le\mathbb{P}(\overline{B})$. For $\Delta\le 0$ we have $|\Delta|=\mathbb{P}(A)-\mathbb{P}(A|B)\le \mathbb{P}(A)-\mathbb{P}(A,B)\le\mathbb{P}(\overline{B})$.

Proof of (112). We start with an outline. Denote for short $\bar{k} = \overline{\omega}(n)$. We select y and perform exploration L_y (as in the proof of the upper bound of (90)). We stop the exploration after the first $\bar{k}$ elements of the list L_y are discovered. Let $\mathbb{D}_y$ denote the collection of sets marked during this exploration and $H(y) = \cup_{D\in\mathbb{D}_y} D$. We will show that $|H(y)| = O_P(\bar{k})$ (see (115) below). Next we select x. The event $\mathcal{L}_{x|y} := \{x \notin H(y)\}$ has probability $1 - O(\bar{k})/n = 1 - o(1)$ (see (116) below). We then consider the exploration L_x (until the first $\bar{k}$ elements of the list L_x are discovered) conditionally on the event $\mathcal{L}_{x|y}$. Using the fact that $|H(y)| = O_P(\bar{k})$ we show that the sets marked during the exploration L_x do not intersect with $H(y)$ with high probability (see (118)). Hence the event $\mathcal{L}_x := \{|L_x| \geq \bar{k}\}$ is asymptotically independent of $\mathcal{L}_y := \{|L_y| \geq \bar{k}\}$. Finally, we establish (112) by approximating $\mathbb{P}(\mathcal{L}_y)$ and $\mathbb{P}(\mathcal{L}_x|\mathcal{L}_y) = \mathbb{P}(\mathcal{L}_x)(1 + o(1))$ by the survival probabilities of related branching processes. The rigorous argument below adds the details.

Recall that V_k denotes the set of k-regular vertices. Denote the events

$$\mathcal{H}_y = \{|H(y)| \leq M\hat{\hat{k}},\ |\mathbb{D}_y| \leq \hat{\hat{k}}\}, \qquad \mathcal{L}_z^+ = \mathcal{L}_z \cap \{z \in V_{\bar{k}}\}, \quad z = x, y,$$

where $\hat{\hat{k}} = 2\bar{k}/\hat{q}$. (100) and the fact that events $\{|C_v| \geq \bar{k}, v \in V_{\bar{k}}\}$ and $\mathcal{L}_v^+$ are equal imply

$$\mathbb{E}(\mathbb{I}_x\mathbb{I}_y) = \mathbb{E}\left(\mathbb{I}_x\mathbb{I}_y\mathbb{I}_{\{x\in V_{\bar{k}}\}}\mathbb{I}_{\{y\in V_{\bar{k}}\}}\right) + o(1), \qquad \mathbb{E}\left(\mathbb{I}_x\mathbb{I}_y\mathbb{I}_{\{x\in V_{\bar{k}}\}}\mathbb{I}_{\{y\in V_{\bar{k}}\}}\right) = \mathbb{P}\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+\}. \tag{114}$$

Note that event $\mathcal{L}_y^+ \cap \{|\mathbb{D}_y| > \hat{\hat{k}}\}$ implies that among the first $\hat{\hat{k}}$ sets marked by L_y less than $\bar{k} - 1$ are nonvoid. The probability of such event is $o(1)$, see (96). Since $|\mathbb{D}_y| \leq \hat{\hat{k}}$ implies $\mathcal{H}_y$ we conclude that

$$\mathbb{P}(\mathcal{L}_y^+ \cap \overline{\mathcal{H}}_y) = o(1). \tag{115}$$

Combining the latter bound with the obvious bound

$$\mathbb{P}(\overline{\mathcal{L}}_{x|y}|\mathcal{H}_y) \leq M\hat{\hat{k}}/n = o(1) \tag{116}$$

we obtain

$$\mathbb{P}\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+\} = \mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y\right\} + o(1) = \mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y}\right\} + o(1). \tag{117}$$

In the last step we used the inequalities

$$\mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \overline{\mathcal{L}}_{x|y}\right\} \leq \mathbb{P}\left\{\mathcal{H}_y \cap \overline{\mathcal{L}}_{x|y}\right\} \leq \mathbb{P}(\overline{\mathcal{L}}_{x|y}|\mathcal{H}_y) = o(1).$$

Let $\mathbb{D}_x = \{D_{i_1}, D_{i_2}, \dots\}$ denote the sets marked during the exploration L_x (D_{i_s} is marked before $D_{i_{s+1}}$). We call D_{i_s} healthy whenever $D_{i_s} \cap H(y) = \emptyset$. Exploration L_x is called healthy if all marked sets D_{i_s} are healthy (recall that we stop marking the sets after L_x collects $\bar{k}$ elements). Introduce events

$$\mathcal{S}_x = \{L_x \text{ is healthy}\}, \qquad \mathcal{S}_x^* = \{\text{there is no nonhealthy } D_{i_s} \text{ with } s \leq \hat{\hat{k}}\}.$$

Next we show that

$$\mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y}\right\} = \mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y} \cap \mathcal{S}_x\right\} + o(1). \tag{118}$$

Given integer $0 < t \le M$, let $D \subset V$ be a random set of size $|D| = t$. Assuming that D and $H(y)$ are independent we estimate the conditional probability

$$\mathbb{P}\left(D \cap H(y) \neq \emptyset | H(y), x \notin H(y), x \in D\right) \le (t-1)|H(y)|(n-1)^{-1}. \tag{119}$$

Now we consider the exploration L_x conditionally, given the event $\mathcal{H}_y \cap \mathcal{L}_{x|y}$. The conditional probability that D_{i_1} marked by x is not healthy is at most $(M-1)M\hat{\hat{k}}(n-1)^{-1}$. Here, we applied (119) and used the fact that $x \notin H(y)$ implies $D_{i_1} \notin \mathbb{D}_y$ and therefore D_{i_1} and $H(y)$ are (conditionally) independent. Furthermore, for $s = 1, 2, \ldots$, given the event that $D_{i_1}, \ldots, D_{i_s}$ are all healthy and that $D_{i_{s+1}}$ was marked by the jth element (where $j < \overline{k}$) of the list L_x, the probability that $D_{i_{s+1}}$ is not healthy is at most $(M-1)M\hat{\hat{k}}(n-j)^{-1} \le (M-1)M\hat{\hat{k}}(n-\overline{k})^{-1}$. Here we used the fact that $u_j \notin H(y)$ implies $D_{i_{s+1}} \notin \mathbb{D}_y$. By the union bound applied to $\overline{\mathcal{S}_x^*} = \cup_{1 \le s \le \hat{\hat{k}}} \{D_{i_1}, \ldots, D_{i_{s-1}}$ are healthy and D_{i_s} is not healthy$\}$, we have

$$\mathbb{P}\left\{\overline{\mathcal{S}_x^*} | \mathcal{H}_y, x \notin H(y)\right\} \le \hat{\hat{k}} \cdot (M-1)M\hat{\hat{k}}(n-\overline{k})^{-1} = o(1).$$

This bound implies

$$\mathbb{P}(\mathcal{S}_x^* \cap \mathcal{H}_y \cap \mathcal{L}_{x|y}) = \mathbb{P}(\mathcal{H}_y \cap \mathcal{L}_{x|y}) - o(1). \tag{120}$$

Furthermore, on the event $\mathcal{S}_x^*$ the exploration L_x does not encounter $H(y)$ and therefore L_x is determined solely by the sets $\mathbb{D}_x = \{D_{i_1}, D_{i_2}, \ldots\}$ (which are subsets of $V \setminus H(y)$). The same argument as that of (96), (97) above yields that the event $\mathcal{L}_x^+ \cap \{|\mathbb{D}_x| > \hat{\hat{k}}\}$ has probability $o(1)$, that is, we have $\mathbb{P}\left\{\mathcal{L}_x^+ \cap \{|\mathbb{D}_x| > \hat{\hat{k}}\} | \mathcal{S}_x^*, \mathcal{L}_{x|y}, \mathcal{H}_y\right\} = o(1)$. Consequently,

$$\mathbb{P}\left\{\mathcal{L}_x^+ \cap \{|\mathbb{D}_x| \le \hat{\hat{k}}\} \cap \mathcal{S}_x^* \cap \mathcal{L}_{x|y} \cap \mathcal{H}_y\right\} = \mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{S}_x^* \cap \mathcal{L}_{x|y}, \mathcal{H}_y\right\} - o(1). \tag{121}$$

In view of the fact that the event $\{|\mathbb{D}_x| \le \hat{\hat{k}}\} \cap \mathcal{S}_x^*$ implies $\mathcal{S}_x$ and event $\mathcal{S}_x$ implies $\mathcal{S}_x^*$ we obtain from (121) that

$$\mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{S}_x \cap \mathcal{L}_{x|y} \cap \mathcal{H}_y\right\} = \mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{S}_x^* \cap \mathcal{L}_{x|y}, \mathcal{H}_y\right\} - o(1).$$

This relation together with (117), (120) imply (118).

In the last step of the proof of (112) we estimate

$$\mathbb{P}\left\{\mathcal{L}_x^+ \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y} \cap \mathcal{S}_x\right\} \le \mathbb{P}\left\{\mathcal{L}_x^+ | \mathcal{S}_x \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y}\right\} \mathbb{P}(\mathcal{L}_y^+) \tag{122}$$

and

$$\mathbb{P}(\mathcal{L}_y^+) \le \rho(\mathcal{Y}) + o(1), \qquad \mathbb{P}\left\{\mathcal{L}_x^+ | \mathcal{S}_x \cap \mathcal{L}_y^+ \cap \mathcal{H}_y \cap \mathcal{L}_{x|y}\right\} \le \rho(\mathcal{Y}) + o(1). \tag{123}$$

The first bound of (123) follows from (89,93,94). The second one is obtained by a similar argument, but now we perform exploration L_x in the subgraph of G induced by the vertex set $V \setminus H(y)$. In particular, we set $N_{t,q}^+ \sim \mathrm{Bin}(m_{t,q}, p_t^*)$ and $\tilde{N}_{t,q}^+ \sim \mathrm{Poi}(m_{t,q} p_t^*)$ in (85), (86). Here $p_t^* = t/(n - \overline{k} - M\hat{\hat{k}})$ upperbounds the probability $\mathbb{P}(u_j \in D)$ that $u_j \in L_x = \{u_1, u_2, \ldots\}$ with $j < \overline{k}$ marks a "healthy" random set $D \subset V \setminus (\{u_1, \ldots, u_{j-1}\} \cup H(y))$ of size $|D| = t$. Note that $2 < t \le M$ implies $p_t^* = tn^{-1}(1 + o(1))$.

Relation (112) follows from (114), (117), (118), (122), and (123). We have shown (82) in the special case of $\omega = \overline{\omega}$, where $\overline{\omega}(n) = \ln n$.

Next we prove (82) for general ω. To this aim we show that $\left||B^\omega| - |B^{\overline{\omega}}|\right| = o_p(n)$. Let $\omega_1 = \omega \vee \overline{\omega}$ and $\omega_2 = \omega \wedge \overline{\omega}$ so that $\left||B^\omega| - |B^{\overline{\omega}}|\right| = |B^{\omega_2}| - |B^{\omega_1}| \geq 0$. Now (111) implies

$$\mathbb{E}\left||B^\omega| - |B^{\overline{\omega}}|\right| = \mathbb{E}\left(|B^{\omega_2}| - |B^{\omega_1}|\right) = o(n).$$

Hence $\left||B^\omega| - |B^{\overline{\omega}}|\right| = o_p(n)$.

Proof of (83). The upper bound $\mathbb{P}\left(|C| \leq n\rho(\mathcal{Y}) + \varepsilon_n'' n\right) \geq 1 - o(1)$ follows from (82). Indeed, we can assume without loss of generality that ε_n'' from (82) satisfies $\varepsilon_n'' n \geq \ln^2 n$. For $\omega(n) = \ln n$ we obtain from (75) for large n, m that

$$\begin{aligned}\mathbb{P}\left(|C| > n\rho(\mathcal{Y}) + \varepsilon_n'' n\right) &\leq \mathbb{P}\left(\max\{\omega(n), |B^\omega|\} > n\rho(\mathcal{Y}) + \varepsilon_n'' n\right) \\ &= \mathbb{P}\left(|B^\omega| > n\rho(\mathcal{Y}) + \varepsilon_n'' n\right) = o(1). \end{aligned} \tag{124}$$

For $\rho(\mathcal{Y}) = 0$ relation (83) follows from (124). It remains to prove (83) for $\rho(\mathcal{Y}) > 0$. To this aim we show the matching lower bound $|C| \geq \rho(\mathcal{Y})n + o_P(n)$. Fix $(t,q) \in A^0$. Choose $\delta = \delta_n > 0$ such that $\rho(\mathcal{Y}_\delta^-) > 0$. We select a subset $\mathbb{D}_{t,q}^\delta \subset \mathbb{D}_{t,q}$ of size $|\mathbb{D}_{t,q}^\delta| = \lfloor \delta m_{t,q} \rfloor$ and color sets from $\mathbb{D}_{t,q}^\delta$ blue. The collection $\mathbb{D}_0^* = \mathbb{D}_0 \setminus \mathbb{D}_{t,q}^\delta$ is obtained from $\mathbb{D}_0$ by removal of the blue sets. Let G^δ (respectively G^*) be the overlay graph on the vertex set V defined by the collection of layers $\mathbb{D}_{t,q}^\delta$ (respectively $\mathbb{D}_0^*$). We color edges of G^δ blue. We couple G, G^δ and G^* so that $G = G^\delta \cup G^*$. Let $\omega(n) = n^{2/3}$ and let $B^* \subset V$ be the set of vertices belonging to connected components of G^* having at least $\omega(n)$ vertices. Clearly, there are at most $n^{1/3}$ such components. Given a pair of such components $C', C'' \subset V$, for any $D \in \mathbb{D}_{t,q}^\delta$, the probability that C', C'' are connected by a blue edge labeled D is at least

$$p^\star := \mathbb{P}\left(D \cap C', D \cap C'' \mid C', C''\right) \cdot q \geq 2n^{4/3}n^{-2}q.$$

Indeed, $\mathbb{P}\left(D \cap C', D \cap C'' \mid C', C''\right)$, the probability that a randomly selected pair of elements of D intersects with C' and C'' simultaneously, is at least $n^{2/3} \times n^{2/3} / \binom{n}{2}$. Furthermore, by Chernoff's bound (see formula (2.6) of [30]), for any $\tilde{c} > 0$ the probability that there are less than $\tilde{c} \ln n$ blue edges between C' and C'' is at most

$$\mathbb{P}\left\{X^\star < \tilde{c} \ln n\right\} \leq e^{-c\delta n^{1/3}}, \tag{125}$$

where the random variable $X^\star$ has binomial distribution $\mathrm{Bin}(\lfloor \delta m_{t,q} \rfloor, p^\star)$. Furthermore, the constant $c > 0$ in (125) depends on $h_{t,q}$, constant $\tilde{c} > 0$ and the sequence $m_n/n \to \mu$. Next, by the union bound, the probability that there exists a pair of components connected by less than $\tilde{c} \ln n$ blue edges is at most

$$\binom{\lceil n^{1/3} \rceil}{2} e^{-c\delta n^{1/3}} = o(1), \tag{126}$$

provided that $\delta n^{1/3} \geq \ln^2 n$.

We let $\delta = n^{-1/6}$ and apply (82) to the set B^* of vertices of G^*. In view of (126) these vertices belong to the same connected component of $G = G^\delta \cup G^*$ with high probability. Hence, $|C| \geq |B^*| \geq n\rho(\mathcal{Y}) + o_P(n)$. ∎

In the next Lemma, we relax condition (74) by allowing a negligible fraction of layer types $(x_{n,i}, q_{n,i})$ to take their values outside A.

Lemma 9.2. *Statements (82), (83) of Lemma 9.1 remain true if we replace condition (74) by the condition*

$$\forall\, n \geq 1 \qquad \forall\, 1 \leq i \leq m_n \qquad x_{n,i} \leq M.$$

and condition (79) by the condition

$$\forall\, n \qquad \max_{(t,q)\in A} \left| h_{t,q} - m_{t,q}/m \right| \leq \epsilon_n. \tag{127}$$

Proof. Given n and $(\bar{x}_n, \bar{q}_n)$ we color a pair $(x_{n,i}, q_{n,i})$ red whenever $(x_{n,i}, q_{n,i}) \in A$. Otherwise we color $(x_{n,i}, q_{n,i})$ blue. The layers defined by red (blue) pairs are colored red (blue) as well. By (127), the number m_B of blue layers satisfies $m_B = o(m)$. The overlay graphs defined by the families of red (blue) pairs are denoted by G_R (G_B). Then $G = G_{(\bar{x}_n, \bar{q}_n)}$ is the union $G = G_B \cup G_R$. Let C_R and C be the vertex sets of the largest components of G_R and G.

We first show that (83) holds (under conditions of Lemma 9.2). We observe that results (81), (82), (83) of Lemma 9.1 apply to G_R because $m_B = o(m)$. In particular, (83) remains true with C replaced by C_R. This observation together with the simple inequality $|C| \geq |C_R|$ yields the lower bound $|C| \geq n\rho(\mathcal{Y}) + o_P(n)$. To prove the matching upper bound we apply the inequality

$$|B^k| \leq |B^k_R| + m_B M k, \tag{128}$$

where B^k_R is the set of vertices that belong to components of G_R of sizes at least k. Let us show inequality (128). We observe that each $v \in B^k \setminus B^k_R$ belongs to a component of G_R of size less than k and this component intersects with some blue layer. Furthermore, each blue layer (being of size at most M) may intersect with at most M distinct components. Hence each blue layer may contribute at most kM vertices to $B^k \setminus B^k_R$. Consequently, blue layers altogether contribute at most $m_B M k$ vertices. From (75) and (128) we obtain

$$|C| \leq \max\left\{k, |B^k|\right\} \leq |B^k_R| + m_B M k + k. \tag{129}$$

Choosing $k = k_n \to +\infty$ so that

$$m_B k_n = o(n) \tag{130}$$

we obtain $|C| \leq |B^{k_n}_R| + o(n)$. Finally we apply (82) to $|B^{k_n}_R|$ and obtain $|C| \leq n\rho(\mathcal{Y}) + o_P(n)$.

Next we show that (82) holds (under conditions of Lemma 9.2). For k_n satisfying (130) the upper bound $|B^{k_n}| \leq n\rho(\mathcal{Y}) + o_P(n)$ follows from (128) and relation (82) applied to $B^{k_n}_R$. We extend this upper bound to arbitrary ω. Fix some $\{k_n\}$ satisfying (130). Then $\omega_2(n) = \omega(n) \wedge k_n$ satisfies (130) and the upper bound applies to $|B^{\omega_2}| \geq |B^{\omega}|$.

The lower bound $|B^{\omega}| \geq n\rho(\mathcal{Y}) + o_P(n)$ makes sense when $\rho(\mathcal{Y}) > 0$. For $\rho(\mathcal{Y}) > 0$ the lower bound follows from the lower bound $|C| \geq |C_R| \geq n\rho(\mathcal{Y}) + o_P(n)$ (see (83)) and the fact that $|B^{\omega}| \geq |C|$ provided that $\omega(n) = o(n)$ and $|C| \geq 0.5n\rho(\mathcal{Y})$. ∎

Proof of Theorem 3.4. The proof of Theorem 3.4 rests on Lemma 9.1. With the aid of truncation and discretization we reduce the general case to the case of a finite set of layer types considered in Lemma 9.1.

We start with some notation. Let $M \geq 2$ be a positive integer. Recall the notation $x^{[M]}$ and $\bar{x}^{[M]}_n$: given $\bar{x}_n = (x_{n,1}, \ldots, x_{n,m})$ we denote $\bar{x}^{[M]}_n = \left(x^{[M]}_{n,1}, \ldots, x^{[M]}_{n,m}\right)$, where $x^{[M]} = x\mathbb{I}_{\{x\leq M\}}$. Furthermore, given $0 < \epsilon < 1$ and $\overline{Q}_n = (Q_{n,1}, \ldots, Q_{n,m})$ we denote $\overline{Q}^{\pm}_n = (Q^{\pm}_{n,1}, \ldots, Q^{\pm}_{n,m})$ the ϵ-discretization as in (20). Let $A^{\pm}_{\varepsilon}$ denote the support of $(\overline{X}^{[M]}, \overline{Q}^{\pm})$. That is, $A^{\pm}_{\varepsilon}$ is a subset of $\{0, 1, \dots, M\} \times \{s_0, \dots, s_r\}$, where $s_0, \dots, s_r$ are possible values of $Q^{\pm}$, and $(t, q) \in A^{\pm}_{\varepsilon}$ whenever $h^{\pm}_{t,q,\varepsilon} > 0$. Here,

$$h^{\pm}_{t,q,\varepsilon} := \mathbb{P}\left((X^{[M]}, Q^{\pm}) = (t, q)\right), \qquad (t, q) \in \{0, 1, \dots, M\} \times \{s_0, \dots, s_r\}.$$

We consider the overlay random graphs $G^{+}_{[M]} = G_{(\overline{X}^{[M]}_n, \overline{Q}^{+}_n)}$ and $G^{-}_{[M]} = G_{(\overline{X}^{[M]}_n, \overline{Q}^{-}_n)}$ defined by the sequences of layer types $\left(\overline{X}^{[M]}_n, \overline{Q}^{+}_n\right)$ and $\left(\overline{X}^{[M]}_n, \overline{Q}^{-}_n\right)$ respectively. We can couple $G^{\pm}_{[M]}$ with $G_{[M]} = G_{(\overline{X}^{[M]}_n, \overline{Q}_n)}$ so that $\mathbb{P}\left(G^{-}_{[M]} \subset G_{[M]} \subset G^{+}_{[M]}\right) = 1$. Furthermore, we couple $G_{[M]}$ and $G = G_{(\overline{X}_n, \overline{Q}_n)}$ so that $\mathbb{P}(G_{[M]} \subset G) = 1$. Let $C^{\pm}_{[M]}$ and $C_{[M]}$ denote the vertex sets of the largest components of $G^{\pm}_{[M]}$ and $G_{[M]}$ respectively. The couplings above imply the couplings

$$\mathbb{P}\left(|C^{-}_{[M]}| \le |C_{[M]}| \le |C^{+}_{[M]}|\right) = 1 \qquad \text{and} \qquad \mathbb{P}\left(|C_{[M]}| \le |C|\right) = 1. \tag{131}$$

Denote $\lambda_{[M]} = \mu x_{*[M]}$, where $x_{*[M]} = \mathbb{E}(X\mathbb{I}_{\{2 \le X \le M\}})$. Let $T_{*[M]}$ be a random variable with the distribution

$$\mathbb{P}\left(T_{*[M]} = \ell\right) = x^{-1}_{*[M]}\mathbb{E}\left(\mathrm{Bin}^{+}(X - 1, Q)(\ell)X\mathbb{I}_{\{2 \le X \le M\}}\right), \qquad \ell = 0, 1, \dots .$$

Note that $T_{*[M]}$ has the same distribution as (77) above, but with (X, Q) replaced by $(X^{[M]}, Q)$. Let $\mathcal{Y}_M$ be a G-W process with the offspring number $Y_M \sim \mathrm{CPoi}(\lambda_{[M]}, \mathcal{L}(T_{*[M]}))$.

Now we are ready to prove the theorem. We will assume that $\mathbb{P}(X \ge 2, Q > 0) > 0$. (The case where $\mathbb{P}(X \ge 2, Q > 0) = 0$ is treated at the very end of the proof.) Let M be large enough so that $\mathbb{P}(2 \le X \le M, Q > 0) > 0$ and $\varepsilon > 0$ is small enough so that A^{+}_{ε} and A^{-}_{ε} are both nonempty (this will ensure the condition $A^0 \neq \emptyset$ of Lemma 9.1).

In the first step of the proof we show that as $n \to +\infty$

$$\left|C_{[M]}\right| = n\rho(\mathcal{Y}_M) + o_P(n). \tag{132}$$

We note that the remainder term $o_P(n)$ depends on M. To prove (132) we apply Lemma 9.2 to overlay graphs $G^{+}_{[M]}$ and $G^{-}_{[M]}$ conditionally, given their layer types $(\overline{X}^{[M]}_n, \overline{Q}^{+}_n)$ and $(\overline{X}^{[M]}_n, \overline{Q}^{-}_n)$ respectively. We will show below that as $n \to +\infty$

$$\left|C^{\pm}_{[M]}\right| = n\rho\left(\mathcal{Y}^{\pm}_M\right) + o_P(n), \tag{133}$$

where $o_P(n)$ depends on M, ε, $A^{\pm}_{\varepsilon}$, $\{h_{t,q,\varepsilon}, (t, q) \in A^{\pm}_{\varepsilon}\}$, the sequence $m/n \to \mu$, and sequences $\delta^{\pm}_n \downarrow 0$ constructed below. Here $\mathcal{Y}^{\pm}_M$ is a G-W processes with offspring numbers $Y^{\pm}_M \sim \mathrm{CPoi}\left(\lambda_{[M]}, \mathcal{L}\left(T^{\pm}_{*[M]}\right)\right)$, where $T^{\pm}_{*[M]}$ is a random variable with the probability distribution

$$\mathbb{P}\left(T^{\pm}_{*[M]} = \ell\right) = x^{-1}_{*[M]}\mathbb{E}\left(\mathrm{Bin}^{+}(X - 1, Q^{\pm})(\ell)X\mathbb{I}_{\{2 \le X \le M\}}\right), \qquad \ell = 0, 1, \dots .$$

Letting $\varepsilon \downarrow 0$ we obtain $\rho\left(\mathcal{Y}^{\pm}_M\right) \to \rho(\mathcal{Y}_M)$. Now (133) together with the first identity of (131) yield (132). We explain the implication (131), (133) $\Rightarrow$ (132) in more detail. Given $\tau > 0$ we choose small ε so that $|\rho(\mathcal{Y}^{\pm}_M) - \rho(\mathcal{Y}_M)| < \tau$. Then we build sets $A^{\pm}_{\varepsilon}$, $\{h_{t,q,\varepsilon}, (t, q) \in A^{\pm}_{\varepsilon}\}$ and sequences $\delta^{\pm}_n$ (depending on ε and M, see the proof of (133) below) and apply (133). We obtain

$$\left||C^{\pm}_{[M]}| - n\rho(\mathcal{Y}_M)\right| \le \tau n + \left||C^{\pm}_{[M]}| - n\rho\left(\mathcal{Y}^{\pm}_M\right)\right| \le \tau n + o_P(n).$$

This yields (132): $\forall\, \tau > 0$ we have $\mathbb{P}\left(\left||C^{\pm}_{[M]}| - n\rho(\mathcal{Y}_M)\right| > 2\tau n\right) = o(1)$ as $n \to +\infty$.

Let us show (133). We only consider $C^+_{[M]}$. For $(t,q) \in A^+_\varepsilon$ denote

$$m_{t,q} = \#\left\{ i\ :\ (X^{[M]}_{n,i}, Q^+_{n,i}) = (t,q) \right\}.$$

Let $\mathcal{A}_n(\delta)$ denote the event that $\max_{(t,q)\in A^+_\varepsilon} |h^+_{t,q,\varepsilon} - m_{t,q}/m| \le \delta$. We claim that there exists a positive sequence $\delta^+_n \downarrow 0$ such that

$$\mathbb{P}(\mathcal{A}_n(\delta^+_n)) = 1 - o(1) \qquad \text{as} \qquad n, m \to +\infty. \tag{134}$$

Indeed, for each $(t,q) \in A^+_\varepsilon$ we have

$$\mathbb{E}(m_{t,q}/m) - h^+_{t,q,\varepsilon} = \mathbb{P}\left((X^{[M]}_{n,\pi}, Q^+_{n,\pi}) = (t,q) \right) - \mathbb{P}\left((X^{[M]}, Q^+) = (t,q) \right) = o(1). \tag{135}$$

In the very last step we use the fact that our assumption $P_n \xrightarrow{w} P$ implies $\mathcal{L}\left((X^{[M]}_n, Q^+_n) \right) \xrightarrow{w} \mathcal{L}\left((X^{[M]}, Q^+) \right)$. Noting that $m_{t,q}$ is a sum of independent Bernoulli random variables (some of them may be degenerate) we have, by Chebyshev's inequality, for any $\delta > 0$

$$\mathbb{P}\left((m_{t,q} - \mathbb{E}m_{t,q}) > m\delta \right) \le \frac{\mathrm{Var}(m_{t,q})}{m^2\delta^2} \le \frac{1}{m\delta^2}. \tag{136}$$

Combining (135) and (136) we obtain (134). Now we are ready to derive (133) from Lemma 9.2. We apply (83) of Lemma 9.2 to $G^+_{[M]}$ conditionally, given the event $\mathcal{A}_n(\delta^+_n)$. By (83) there exists a sequence $\varepsilon''_n \downarrow 0$ (depending on M, A^+_ε, $\{h^+_{t,q,\varepsilon}, (t,q) \in A^+_\varepsilon\}$ and the sequences δ^+_n and $m/n \to \mu$) such that

$$\mathbb{P}\left(\left| |C^+_{[M]}| - n\rho(\mathcal{Y}^+_{[M]}) \right| > \varepsilon''_n n \middle| \mathcal{A}_n(\delta^+_n) \right) = o(1).$$

Combining this bound with (134) we obtain (133).

In the second step of the proof we let $M \to +\infty$. Denote $\lambda = \mu x_*$, where $x_* = \mathbb{E}(X\mathbb{I}_{\{X\ge 2\}})$, and let T_* be a random variable with the distribution

$$\mathbb{P}(T_* = \ell) = x_*^{-1}\mathbb{E}\left(\mathrm{Bin}^+(X-1, Q)(\ell) X \mathbb{I}_{\{2\le X\}} \right), \qquad \ell = 0, 1, \ldots .$$

Note that T_* has the same distribution as (77) above, but now we drop the restriction on (X, Q) of having a finite support. Let $\mathcal{Y}$ be a G-W process with the offspring number $Y \sim \mathrm{CPoi}(\lambda, \mathcal{L}(T_*))$. We mention that the equality of distributions (78) extends to the general setup (where we drop the restriction on (X, Q) of having a finite support). Consequently, the respective survival probabilities $\rho(\mathcal{Y})$ and $\rho_{(f^+)}$ are the same.

To prove the first relation of Theorem 3.4 we show that

$$|C| = n\rho(\mathcal{Y}) + o_P(n).$$

We have, by continuity, that $\rho(\mathcal{Y}_M) \to \rho(\mathcal{Y})$ as $M \to +\infty$. Now the second identity of (131) together with (132) yield the lower bound $|C| \ge n\rho(\mathcal{Y}) + o_P(n)$ as $n \to +\infty$. Indeed, given $\tau > 0$ we choose large M such that $|\rho(\mathcal{Y}_M) - \rho(\mathcal{Y})| < \tau$. Then we apply (131), (132) to get $|C| \ge |C_{[M]}| \ge \big(\rho(\mathcal{Y}) - \tau\big)n + o_P(n)$.

To show the matching upper bound $|C| \leq n\rho(\mathcal{Y}) + o_P(n)$ we use the inequalities

$$|C| \leq \max\{k, |B^k|\} \leq \left|B^k_{[M]}\right| + k \sum_{1\leq i\leq m} X_{n,i}\mathbb{I}_{\{X_{n,i}>M\}} + k. \tag{137}$$

Here, B^k (respectively $B^k_{[M]}$) is the set of vertices of G (respectively $G_{[M]}$) that belong to components of size at least k. The first inequality of (137) is obvious. The second one is obtained in the same way as (128), (129) above.

Now we upper bound $\left|B^k_{[M]}\right|$. An inspection of the proof of (82) in Lemmas 9.1 and 9.2 shows that (82) remains true if we replace B^ω and $\rho(\mathcal{Y})$ by B^k and $\rho^{(k)}(\mathcal{Y})$, respectively, where integer $k \geq 1$ is fixed. (Note that sequences $\varepsilon'_n \downarrow 0$, $\varepsilon''_n \downarrow 0$ of (82) now depends on k (instead of ω). Using this fact and proceeding as in the proof of (132) above we show that as $n \to +\infty$

$$\left|B^k_{[M]}\right| = n\rho^{(k)}(\mathcal{Y}_M) + o_P(n). \tag{138}$$

Note that the remainder term $o_P(n)$ depends on k and M.

Next we upper bound the sum $\sum(\cdots)$ on the right of (137). We observe that as $M \to \infty$

$$Z_{M,n} := m^{-1} \sum_{1\leq i\leq m} X_{n,i}\mathbb{I}_{\{X_{n,i}>M\}} \xrightarrow{\mathbb{P}} 0 \tag{139}$$

uniformly in n. Indeed, our conditions $P_n \xrightarrow{w} P$ and $(P_n)_{10} \to (P)_{10}$ imply the uniform integrability of the sequence $\{X_{n,\pi},\ n \geq 1\}$. That is, we have

$$\varphi_M := \sup_n \mathbb{E}\left(X_{n,\pi}\mathbb{I}_{\{X_{n,\pi}>M\}}\right) \to 0 \qquad \text{as} \qquad M \to +\infty.$$

Then for any $\tau > 0$ and n we have, by Markov's inequality, that

$$\mathbb{P}\left(Z_{M,n} > \tau\right) \leq \tau^{-1}\mathbb{E}Z_{M,n} = \tau^{-1}\mathbb{E}\left(X_{n,\pi}\mathbb{I}_{\{X_{n,\pi}>M\}}\right) \leq \tau^{-1}\varphi_M.$$

Hence (139) holds uniformly in n.

Finally, combining (137), (138) and invoking the simple inequality $\rho^{(k)}(\mathcal{Y}_M) \leq \rho^{(k)}(\mathcal{Y})$ we obtain for any k, M as $n, m \to +\infty$

$$|C| \leq n\rho^{(k)}(\mathcal{Y}) + kmZ_{M,n} + k + o_P(n),$$

where $o_P(n)$ depends on k and M. Now, by choosing large k, we can make $\rho^{(k)}(\mathcal{Y})$ arbitrarily close to $\rho(\mathcal{Y})$. Then, by choosing large M, we can make $kZ_{M,n}$ arbitrarily close to zero whp. In this way we obtain the desired upper bound $|C| \leq n\rho(\mathcal{Y}) + o_p(n)$. For readers convenience we explain these steps in more detail. From the fact that $m/n \to \mu$ we conclude that the sequence $\{m/n\}$ is bounded, that is, $m/n < C_\star$ for some $C_\star > 0$ and all n (recall that $m = m_n$). Given $\tau > 0$ we choose (sufficiently large) k such that $\rho^{(k)}(\mathcal{Y}) \leq \rho(\mathcal{Y}) + \tau$. Then we choose (sufficiently large) M such that $\varphi_M < \tau^2/(C_\star k)$. Now (137), (138) yields

$$|C| \leq n\rho(\mathcal{Y}) + \tau n + kmZ_{M,n} + k + nr_{k,M,n},$$

where $nr_{k,M,n}$ stands for the remainder $o_P(n)$ in (138). In particular, $r_{k,M,n} = o_P(1)$ as $n \to +\infty$. Consequently,

$$\mathbb{P}(|C| > n\rho(\mathcal{Y}) + 4\tau n) \le \mathbb{P}(kmZ_{M,n} > \tau n) + \mathbb{P}(k > \tau n) + \mathbb{P}(r_{k,M,n} > \tau).$$

The first probability on the right

$$\mathbb{P}(kmZ_{M,n} > \tau n) \le (km/\tau n)\mathbb{E}Z_{M,n} \le (km/\tau n)\varphi_M \le \tau,$$

by Markov's inequality. Furthermore, $\mathbb{P}(k > \tau n) + \mathbb{P}(r_{k,M,n} > \tau) = o(1)$ as $n \to +\infty$. Hence $\mathbb{P}(|C| > n\rho(\mathcal{Y}) + 4\tau n) \le \tau + o(1)$. The latter inequality yields the bound $|C| \le n\rho(\mathcal{Y}) + o_p(n)$.

It remains to show that $N_2(G_{(n)}) = o_P(n)$. We write for short $N_i = N_i(G_{(n)})$. For $\rho = 0$ we have $N_2 \le N_1 = o_P(n)$. For $\rho > 0$ we use the simple inequality $N_1 + N_2 \le |B^k| + 2k$ $\forall k = 2, 3, \dots$. From (138) and the second inequality of (137) we obtain $N_1 + N_2 \le n\rho^{(k)}(\mathcal{Y}) + kmZ_{M,n} + 2k + o_P(n)$. Hence $N_1 + N_2 \le n\rho + o_p(n)$. Now $N_2 = o_P(n)$ follows from the relation $N_1 \ge n\rho + o_P(n)$ shown above (recall that $N_1 = |C|$).

Now we consider the case where $\mathbb{P}(X \ge 2, Q > 0) = 0$. In this case the distribution $f^+ = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}^+(P))$ is degenerate, $f^+(0) = 1$, and the theorem claims that $N_i(G_{(n)}) = o_P(n)$, for $i = 1, 2$. To prove this claim we show that $\mathbb{E}|N_1(G_{(n)})| = o(n)$. Note that the asymptotic degree distribution $f = \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(P))$ defined by Theorem 3.1 is degenerate as well. In particular, we have $f^{(n)}(0) \to f(0) = 1$. Consequently, we have

$$\mathbb{E}|N_1(G_{(n)})| \le \mathbb{E}\left(\sum_{i \in V} \mathbb{I}_{\{\deg_{G_{(n)}}(i) \ge 1\}}\right) = \sum_{i \in V} \mathbb{P}(\deg_{G_{(n)}}(i) \ge 1) = n(1 - f^{(n)}(0)) = o(n).$$

■

10 | PERCOLATION MODELS

Proof of Theorem 3.5. The site-percolated graph $\check{G}$ is an instance of the overlay graph model (1) with $\check{n} = |S|$ nodes and m layers $\check{G}_1, \dots, \check{G}_m$ where $\check{G}_k$ is the subgraph of G_k induced by vertices belonging to S, and $G_1, \dots, G_m$ are the original layers generating the overlay graph G. The layer types $(\check{X}_k, Q_k)$ in the site-percolated model are mutually independent, and $\mathcal{L}(\check{X}_k \mid X_k = x_k)$ is hypergeometric with probability mass function

$$\mathrm{Hyp}(n, \check{n}, x_k)(t) = \frac{\binom{\check{n}}{t}\binom{n-\check{n}}{x_k - t}}{\binom{n}{x_k}}.$$

We consider the sequence of site-percolated graphs $(\check{G}_{(n)})$ defined by a sequence of overlay graphs $(G_{(n)})$ and sets (S_n). For each n, the site-percolated graph $\check{G}_{(n)}$ is an instance of the overlay model with $|S_n| = \check{n} = \check{n}_n$ nodes, $m = m_n$ layers, and averaged layer type distribution

$$\check{P}_n(A) = \int (\mathrm{Hyp}(n, \check{n}, x) \times \delta_q)(A)\, P_n(dx, dq).$$

We claim that $\check{P}_n \xrightarrow{w} \check{P}$ as $n \to +\infty$. In the proof of the claim we use the following bounds

$$d_{\rm tv}\left(\mathrm{Hyp}(n,\check{n},x),\mathrm{Bin}\left(x,\frac{\check{n}}{n}\right)\right) \le 4\frac{x}{n}, \quad d_{\rm tv}\left(\mathrm{Bin}\left(x,\frac{\check{n}}{n}\right),\mathrm{Bin}(x,\theta)\right) \le \left|\frac{\check{n}}{n}-\theta\right| x. \tag{140}$$

The first bound is shown in [20, Theorem 4]. The second one is obtained by coupling of coin flips. Recall that $(X_{n,\pi}, Q_{n,\pi})$ is a bivariate random variable with the distribution P_n. Let S_n^* be a random subset of $\{1,\dots,n\} = V$ of size $\check{n}$ independent of $(X_{n,\pi}, Q_{n,\pi})$. Then the random variable $\check{X}_{n,\pi} = |\{1,\dots,X_{n,\pi}\} \cap S_n^*|$ has the mixed hypergeometric distribution $\mathbb{P}(\check{X}_{n,\pi} = t) = \mathbb{E}\left(\mathrm{Hyp}(n,\check{n},X_{n,\pi})(t)\right)$ and the random vector $(\check{X}_{n,\pi}, Q_{n,\pi})$ has the distribution $\check{P}_n$. Let $(\check{X}, Q)$ be a random variable with the distribution $\check{P}$. To prove the weak convergence $\check{P}_n \xrightarrow{w} \check{P}$ we show for every $(t,s) \in \mathbb{Z}_+ \times [0,1]$ that $\Delta := \check{P}_n(t \times [0,s]) - \check{P}(t\times[0,s]) = o(1)$. We split

$$\Delta = \mathbb{E}\left(\mathrm{Hyp}(n,\check{n},X_{n,\pi})(t)\mathbb{I}_{\{Q_{n,\pi}\le s\}}\right) - \mathbb{E}\left(\mathrm{Bin}(X,\theta)(t)\mathbb{I}_{\{Q\le s\}}\right) = \Delta_1 + \Delta_2,$$

where

$$\begin{aligned}\Delta_1 &= \mathbb{E}\left(\mathrm{Hyp}(n,\check{n},X_{n,\pi})(t)\mathbb{I}_{\{Q_{n,\pi}\le s\}}\right) - \mathbb{E}\left(\mathrm{Bin}(X_{n,\pi},\theta)(t)\mathbb{I}_{\{Q_{n,\pi}\le s\}}\right)\\ &= \mathbb{E}\left(\left(\mathrm{Hyp}(n,\check{n},X_{n,\pi})(t) - \mathrm{Bin}(X_{n,\pi},\theta)(t)\right)\mathbb{I}_{\{Q_{n,\pi}\le s\}}\right)\end{aligned}$$

and

$$\Delta_2 = \mathbb{E}\left(\mathrm{Bin}(X_{n,\pi},\theta)(t)\mathbb{I}_{\{Q_{n,\pi}\le s\}}\right) - \mathbb{E}\left(\mathrm{Bin}(X,\theta)(t)\mathbb{I}_{\{Q\le s\}}\right).$$

By (140), we have $|\Delta_1| \le \left(|\frac{\check{n}}{n}-\theta| + \frac{4}{n}\right)\mathbb{E}X_{n,\pi}$. Our assumption $(P_n)_{10} \to (P)_{10}$ implies $limsup_n \mathbb{E}X_{n,\pi} < \infty$. Hence $\Delta_1 = o(1)$. Furthermore, the assumption $P_n \xrightarrow{w} P$ implies $\Delta_2 = o(1)$. We obtain $\Delta = o(1)$ thus proving the claim.

Next we observe that $(\check{P}_n)_{10} = \frac{\check{n}}{n}(P_n)_{10} \to \theta(P)_{10} = (\check{P})_{10}$, and $\frac{m}{\check{n}} \to \check{\mu} = \theta^{-1}\mu$. Theorem 3.5(i)–(ii) now follow by applying Theorems 3.1 and 3.4 to $\check{G}_{(n)}$ and noting that $\check{\mu}(\check{P})_{10} = \mu(P)_{10}$.

Now assume that $(P_n)_{rs} \to (P)_{rs} \in (0,\infty)$ for $rs = 21, 32, 33$. A direct computation shows that $(\check{P})_{rs} = \theta^r(P)_{rs}$. Theorem 3.5(iii) now follows by applying Theorem 3.2 to conclude that the clustering coefficient $\tau(\check{G}_{(n)})$ converges to $\frac{(\check{P})_{33}}{(\check{P})_{32}+\check{\mu}(\check{P})_{21}^2} = \frac{(P)_{33}}{(P)_{32}+\mu(P)_{21}^2} = \tau$. Theorem 3.5(iv) follows similarly from Theorem 3.3. ■

Proof of Theorem 3.6 for the layerwise bond-percolated graph $\tilde{G}_{(n)}$. The graph $\tilde{G}_{(n)}$ is an instance of the overlay model with n nodes and m layers $\tilde{G}_{n,1},\dots,\tilde{G}_{n,m}$ where $\tilde{G}_{n,k}$ has size $X_{n,k}$ and strength $\theta Q_{n,k}$. The layers $(\tilde{G}_{n,k}, X_{n,k}, \theta Q_{n,k})$ are mutually independent, with averaged layer type distribution

$$\tilde{P}_n(A) = \int (\delta_x \times \delta_{\theta q})(A)\, P_n(dx,dq) \tag{141}$$

converging according to $\tilde{P}_n \xrightarrow{w} \hat{P}$ and $(\tilde{P})_{10} \to (\hat{P})_{10}$. Furthermore, a direct computation shows that $(\hat{P})_{rs} = \theta^s(P)_{rs}$. Statements (i)–(ii) of Theorem 3.6 now follow by Theorems 3.1 and 3.4, and noting that $(\hat{P})_{10} = (P)_{10}$. Statements (iii)–(iv) follow analogously by Theorems 3.2 and 3.3.

Proof of Theorem 3.6 for overlay bond percolated graph $\hat{G}_{(n)}$. In the proof we use a coupling argument. We will utilize the fact that the overlay bond-percolated graph does not differ much from the layerwise bond-percolated graph $\tilde{G}_{(n)}$, for which the theorem has already been proved. The conditional distribution of $\hat{G}_{(n)}$ given the layers $(G_{n,k}, X_{n,k}, Q_{n,k})$ is an inhomogeneous Bernoulli graph on $\{1, \dots, n\}$ where each node pair ij is linked with probability $\hat{p}_{ij} = \theta(M_{ij} \wedge 1)$ where $M_{ij} = \sum_k \mathbb{I}(E(G_{n,k}) \ni ij)$ is the number of layers linking a node pair ij. The corresponding conditional distribution of $\tilde{G}_{(n)}$ is a similar inhomogeneous Bernoulli graph with link probabilities $\tilde{p}_{ij} = 1 - (1-\theta)^{M_{ij}}$. Because $\hat{p}_{ij} \le \tilde{p}_{ij}$, this suggest the following coupling construction:

(i) Sample the layers $(G_{n,k}, X_{n,k}, Q_{n,k})$, $k = 1, \dots, m$.
(ii) Sample independent inhomogeneous Bernoulli graphs $\tilde{H}$ and H^* with link probabilities $\tilde{p}_{ij}$ and $p^*_{ij} = \frac{\hat{p}_{ij}}{\tilde{p}_{ij}}$ with the convention $\frac{0}{0} = 1$. Note that $p^*_{ij} \ge \theta$.
(iii) Define $\hat{G}_{(n)} = G_{(n)} \cap \hat{H}$ and $\tilde{G}_{(n)} = G_{(n)} \cap \tilde{H}$ with $G_{(n)}$ defined by (1) and $\hat{H} = \tilde{H} \cap H^*$. Note the identity $\hat{G}_{(n)} = \tilde{G}_{(n)} \cap H^*$.

Then $(\hat{G}_{(n)}, \tilde{G}_{(n)}, G_{(n)})$ constitutes a coupling of the overlay bond-percolated, layerwise bond-percolated, and nonpercolated graphs such that

$$\hat{G}_{(n)} \subset \tilde{G}_{(n)} \subset G_{(n)} \qquad \text{almost surely.} \tag{142}$$

Proof of Theorem 3.6(i). Let us denote by $\hat{d}_n = \deg_{\hat{G}_{(n)}}(i)$ and $\tilde{d}_n = \deg_{\tilde{G}_{(n)}}(i)$ the degrees of node i in the overlay bond-percolated and layerwise bond-percolated graph, respectively. By the coupling (142), we have $\hat{d}_n = \tilde{d}_n$ on the event $M_i \le 1$, where $M_i = \max_{j \ne i} M_{ij}$. Hence $d_{\mathrm{tv}}(\mathcal{L}(\hat{d}_n), \mathcal{L}(\tilde{d}_n)) \le \mathbb{P}(M_i > 1)$. The union bound implies that

$$\mathbb{P}(M_{ij} > 1) \le \sum_{k,\ell}{}' \mathbb{P}(E(G_{n,k}) \ni ij)\, \mathbb{P}(E(G_{n,\ell}) \ni ij) \le \left(\sum_k \mathbb{P}(E(G_{n,k}) \ni ij) \right)^2.$$

By noting that $\mathbb{P}(E(G_{n,k}) \ni ij) = \mathbb{E}\frac{(X_{n,k})_2}{(n)_2} Q_{n,k}$, we conclude that

$$\mathbb{P}(M_{ij} > 1) \le \left(m(n)_2^{-1} (P_n)_{21} \right)^2, \tag{143}$$

Another union bound shows that $\mathbb{P}(M_i > 1) \le \sum_{j \ne i} \mathbb{P}(M_{ij} > 1)$ and hence

$$d_{\mathrm{tv}}(\mathcal{L}(\hat{d}_n), \mathcal{L}(\tilde{d}_n)) \le (m/n)^2 (P_n)_{21}^2 (n-1)^{-1}. \tag{144}$$

In the particular case, where there the layer sizes are bounded, that is, there exists constant $M > 0$ such that

$$\mathbb{P}(X_{n,k} \le M) = 1 \quad \forall \ n = 1, 2, \dots \quad \text{and} \quad k = 1, \dots, m_n, \tag{145}$$

the right side of (144) is $o(1)$ as $n \to +\infty$. Now (144) together with the weak convergence (shown above) $\mathcal{L}(\tilde{d}_n) \xrightarrow{w} \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(\hat{P}))$ yields the weak convergence $\mathcal{L}(\hat{d}_n) \xrightarrow{w} \mathrm{CPoi}(\mu(P)_{10}, \mathrm{Bin}_{10}(\hat{P}))$.

Finally, to treated the general case we revoke the boundedness condition (145) in the same way as in the proof of Theorem 3.1. The proof of Theorem 3.6(i) is complete.

Proof of Theorem 3.6(iii). By the fact that random overlay graph $G_{(n)}$ is independent of the percolating graph H, see (12), for any distinct nodes i, j, k, we have

$$\mathbb{P}\left(\hat{G}_{(n)}(ij), \hat{G}_{(n)}(ik), \hat{G}_{(n)}(jk)\right) = \theta^3 \mathbb{P}\left(G_{(n)}(ij), G_{(n)}(ik), G_{(n)}(jk)\right),$$
$$\mathbb{P}\left(\hat{G}_{(n)}(ij), \hat{G}_{(n)}(ik)\right) = \theta^2 \mathbb{P}\left(G_{(n)}(ij), G_{(n)}(ik)\right).$$

Hence $\tau(\hat{G}_{(n)}) = \theta\tau(G_{(n)})$ and the claim follows by applying Theorem 3.2 to the nonpercolated model.

Proof of Theorem 3.6(iv). The proof is similar to that of Theorem 3.3, but it is a bit more technical. For reader's convenience we give it in the Appendix (Section A) below.

Proof of Theorem 3.6(ii). Let $\tilde{N}_1$ and $\tilde{N}_2$ (respectively, $\hat{N}_1$ and $\hat{N}_2$) denote the number of vertices of the largest and second largest component of $\tilde{G}_{(n)}$ (respectively $\hat{G}_{(n)}$). Let $\tilde{\rho} = \rho_{(\hat{f}^+)}$ denote the survival probability of the Galton-Watson branching process with the offspring distribution $\hat{f}^+$.

We observe that coupling (142) yields a coupling of $\hat{N}_1$ and $\tilde{N}_1$ such that $\Pr(\hat{N}_1 \le \tilde{N}_1) = 1$. Furthermore, an application of Theorem 3.4 to the sequence of overlay graphs $(\tilde{G}_{(n)})$ yields the approximation $\tilde{N}_1 = \tilde{\rho}n + o_P(n)$. These two facts taken together imply the upper bound $\hat{N}_1 \le \tilde{\rho}n + o_P(n)$. To show the matching lower bound $\hat{N}_1 \ge n\tilde{\rho} + o_P(n)$ and the bound $\hat{N}_2 = o_P(n)$ we use the same argument as that of the proof of Theorem 3.4. The only place where we need a minor modification of the argument is the proof of Lemma 9.1. In what follows we review the proof of Lemma 9.1 and pinpoint the changes needed to be made.

At this point we need some notation. Let $\hat{C}$, $\hat{C}_v$, $\hat{B}^k$ and $\tilde{C}$, $\tilde{C}_v$, $\tilde{B}^k$ denote the largest component, the component containing vertex v and the set of vertices belonging to components of size at least k in $\hat{G}_{(n)}$ and $\tilde{G}_{(n)}$ respectively. Recall the notation $\lambda = \mu x_*$ and $x_* = \mathbb{E}(X\mathbb{I}_{\{X\ge 2\}})$, and let $\tilde{T}_*$ be a random variable with the distribution

$$\mathbb{P}(\tilde{T}_* = \ell) = x_*^{-1}\mathbb{E}\left(\mathrm{Bin}^+(X-1, \theta Q)(\ell) X\mathbb{I}_{\{X\ge 2\}}\right), \qquad \ell = 0, 1, \dots,$$

which is obtained from (77) by attaching the factor θ to Q. We note the equality of distributions $\hat{f}^+ = \mathrm{CPoi}(\lambda, \mathcal{L}(\tilde{T}_*))$, which is shown by the same argument as (30) above. In particular, $\tilde{\rho}$ is the survival probability of the Galton-Watson branching process with the offspring distribution $\mathrm{CPoi}(\lambda, \mathcal{L}(\tilde{T}_*))$.

We claim that the results (81), (82), (83) of Lemma 9.1 hold true if we replace C, C_v, B^ω, $\rho(\mathcal{Y})$ by $\hat{C}$, $\hat{C}_v$, $\hat{B}^\omega$, $\tilde{\rho}$, respectively. In the proof we use the fact that the coupling $\hat{G}_{(n)} \subset \tilde{G}_{(n)}$ implies couplings $\hat{C} \subset \tilde{C}$, $\hat{C}_v \subset \tilde{C}_v$, $\hat{B}^\omega \subset \tilde{B}^\omega$, and that the results of Lemma 9.1 hold true for $\tilde{C}$, $\tilde{C}_v$, $\tilde{B}^\omega$, $\tilde{\rho}$.

Proof of (81). We apply Lemma 9.1 to $\tilde{G}_{(n)}$ and obtain the upper bound of (81) via the coupling $\hat{C}_v \subset \tilde{C}_v$:

$$\mathbb{P}\left(|\hat{C}_v| \ge \omega(n)\right) \le \tilde{\rho} + \varepsilon'_n.$$

The corresponding lower bound

$$\mathbb{P}\left(|\hat{C}_v| \ge \omega(n)\right) \ge \tilde{\rho} - \varepsilon'_n$$

is obtained by the same argument as in the proof of respective result of Lemma 9.1 (note that regular exploration will not detect any difference between the coupled graphs $\hat{G}_{(n)} \subset \tilde{G}_{(n)}$).

Proof of (82). In the proof we take a shortcut (compared to the original argument of Lemma 9.1) while establishing the analog of the main intermediate inequality (112). For the overlay graph $\tilde{G}_{(n)}$ the inequality (112) holds and it reads as follows

$$\mathbb{E}\left(\mathbb{I}_{\{|\tilde{G}_x|\ge\omega(n)\}}\mathbb{I}_{\{|\tilde{G}_y|\ge\omega(n)\}}\right) \le \tilde{\rho}\times\tilde{\rho} + o(1).$$

The coupling $\hat{C}_x \subset \tilde{G}_x$, $\hat{C}_y \subset \tilde{G}_y$ implies

$$\mathbb{E}\left(\mathbb{I}_{\{|\hat{C}_x|\geq\omega(n)\}}\mathbb{I}_{\{|\hat{C}_y|\geq\omega(n)\}}\right) \leq \mathbb{E}\left(\mathbb{I}_{\{|\tilde{G}_x|\geq\omega(n)\}}\mathbb{I}_{\{|\tilde{G}_y|\geq\omega(n)\}}\right).$$

Hence

$$\mathbb{E}\left(\mathbb{I}_{\{|\hat{C}_x|\geq\omega(n)\}}\mathbb{I}_{\{|\hat{C}_y|\geq\omega(n)\}}\right) \leq \tilde{\rho}\times\tilde{\rho}+o(1).$$

This is the analogue of (112) that yields (82) for $\hat{G}_{(n)}$. The rest of the proof of (82) goes without changes.

Proof of (83). The upper bound

$$\mathbb{P}(|\hat{C}| \leq n\tilde{\rho}+\varepsilon_n'' n) \geq 1-o(1)$$

follows from the respective upper bound for $|\tilde{C}|$ combined with the coupling $\hat{C} \subset \tilde{C}$. To show the matching lower bound

$$\mathbb{P}(|\hat{C}| \geq n\tilde{\rho}-\varepsilon_n'' n) \geq 1-o(1)$$

we use the coupling $\hat{G}_{(n)} \subset \tilde{G}_{(n)}$ and slightly modify the corresponding argument of the proof of Lemma 9.1. Recall that $\hat{G}_{(n)}$ is obtained from $\tilde{G}_{(n)}$ by deleting certain edges at random and the probability of deletion is at most $1-p$. We call this process thinning. Let $\hat{C}'$, $\hat{C}''$ be connected components of $\hat{G}_{(n)}$ that have at least $n^{2/3}$ vertices each. We show that for any pair of such components $\hat{C}'$, $\hat{C}''$ there is at least one blue edge of $\tilde{G}_{(n)}$ (blue edges are defined in the proof of Lemma 9.1 when applied to $\tilde{G}_{(n)}$) connecting $\hat{C}'$ and $\hat{C}''$ that has not been removed when we intersect H^* with $\tilde{G}_{(n)}$ to get $\hat{G}_{(n)} = \tilde{G}_{(n)} \cap H^*$ in our coupling construction (142). For this purpose we choose $\tilde{c} = 10\theta^{-1}$ in (125). Now the probability that all blue edges of $\tilde{G}_{(n)}$ connecting a given pair $\hat{C}'$, $\hat{C}''$ have been removed by the intersection of $\tilde{G}_{(n)}$ with H^* is at most $(1-\theta)^{\tilde{c}\ln n} \leq n^{-10}$. As the number of pairs does not exceed $\binom{\lceil n^{1/3}\rceil}{2}$ the probability that at least one pair $\hat{C}'$, $\hat{C}''$ is not connected by a blue edge is at most $\binom{\lceil n^{1/3}\rceil}{2} n^{-10} = o(1)$, by the union bound. The rest of the proof of the lower bound of (83) goes without changes. Hence (83) holds. □

11 | SUPPLEMENTARY RESULTS

Recall the overlay graph G defined by (1). Given $A \subset [m]$, we consider the subgraph $G_A \subset G$ on the vertex set $V(G_A) = V(G) = \{1, \dots, n\}$ defined by the layers (G_a, X_a, Q_a), $a \in A$. Thus, the edge set of G_A is $E(G_A) = \cup_{a\in A} E(G_a)$. In the following two results, we denote by N_A the set of neighbors of node i in G_A, and we set $D_A = |N_A|$ to denote the degree of i in G_A.

Lemma 11.1. *Let g be a probability density on $\mathbb{Z}_+$. Let*

$$\varepsilon(t) = \sum_{r+s=t} (\mathbb{P}(D_{A\cup B}=r)-\mathbb{P}(D_A=r))\, g(s).$$

Then $\sum_{t\geq 0} |\varepsilon(t)| \leq 2\mathbb{P}(D_B > 0)$.

Proof. Denote the densities of the degrees by $f_{A\cup B} = \mathcal{L}(D_{A\cup B})$ and $f_A = \mathcal{L}(D_A)$. Then

$$\sum_{t\geq 0} |\varepsilon(t)| = \|f_{A\cup B} * g - f_A * g\|_1 = 2d_{\rm tv}(f_{A\cup B} * g, f_A * g) \leq 2d_{\rm tv}(f_{A\cup B}, f_A).$$

Further, $d_{\rm tv}(f_{A\cup B}, f_A) \leq \mathbb{P}(D_{A\cup B} \neq D_A) \leq \mathbb{P}(D_B > 0)$. ∎

Lemma 11.2. *Assume that $G_1, \ldots, G_m$ are mutually independent, let $A, B \subset [m]$ be disjoint, and let $\mathcal{E}_A, \mathcal{E}_B$ be events determined by $(G_a)_{a\in A}$ and $(G_b)_{b\in B}$, respectively. Then*

$$\mathbb{P}(D_{A\cup B} = t, \mathcal{E}_A, \mathcal{E}_B) = \mathbb{P}(D_A + D_B = t, \mathcal{E}_A, \mathcal{E}_B) + \varepsilon(t),$$

where the error term is bounded by $|\varepsilon(t)| \leq c_B t \mathbb{P}(D_A \leq t, \mathcal{E}_A)$, and where $c_B = \max_{j\neq i} \mathbb{P}(ij \in E(G_B), \mathcal{E}_B)$. In the particular case where $i = 1, A = \{k\}, B = \{\ell\}, \mathcal{E}_A = \{12 \in G_k\}, \mathcal{E}_B = \{13 \in G_\ell\}$ we have $|\varepsilon(t)| \leq t\mathbb{P}(12, 13 \in G_k)\mathbb{P}(13 \in G_\ell) + t\mathbb{P}(12, 13 \in G_\ell)\mathbb{P}(12 \in G_k)$.

Proof. Because $D_{A\cup B} = D_A + D_B$ outside the event $\mathcal{F} = \{|N_A \cap N_B| > 0\}$, we see that

$$\varepsilon(t) = \mathbb{P}(D_{A\cup B} = t, \mathcal{E}_A, \mathcal{E}_B, \mathcal{F}) - \mathbb{P}(D_A + D_B = t, \mathcal{E}_A, \mathcal{E}_B, \mathcal{F}).$$

Hence, it follows that $|\varepsilon(t)| \leq \mathbb{P}(D_A \leq t, \mathcal{E}_A, \mathcal{E}_B, \mathcal{F})$, where the upper bound can be expressed as

$$\mathbb{P}(D_A \leq t, \mathcal{E}_A, \mathcal{E}_B, \mathcal{F}) = \sum_{U : |U|\leq t, i\notin U} \mathbb{P}(N_A = U, \mathcal{E}_A)\, \mathbb{P}\left(|U \cap N_B| > 0, \mathcal{E}_B\right).$$

Because $\mathbb{P}(|U \cap N_B| > 0, \ \mathcal{E}_B) \leq \sum_{j\in U} \mathbb{P}(ij \in E(G_B), \ \mathcal{E}_B) \leq c_B t$ whenever $|U| \leq t$, the inequality $|\varepsilon(t)| \leq c_B t\mathbb{P}(D_A \leq t, \mathcal{E}_A)$ follows.

To show the remaining bound we note that $N_A \cap N_B \neq \emptyset$ implies that at least one of the events $\{3 \in N_A\}$ and $\{\exists j \in N_A \cap N_B, j \neq 3\}$ occurs. Hence

$$\begin{aligned}\mathbb{P}(D_A \leq t, \mathcal{E}_A, \mathcal{E}_B, \mathcal{F}) &= \mathbb{P}(\mathcal{E}_A, \mathcal{E}_B, 13 \in G_k) \\ &\quad + \sum_{U : |U|\leq t, i\notin U} \mathbb{P}(N_A = U, \mathcal{E}_A) \sum_{j\in U, j\neq 3} \mathbb{P}(\mathcal{E}_B, 1j \in G_\ell).\end{aligned}$$

By symmetry, the inner sum is at most $t\mathbb{P}(12, 13 \in G_\ell)$. Consequently, the second term on the right is at most $\mathbb{P}(\mathcal{E}_A)\mathbb{P}(12, 13 \in G_\ell)$. The first term equals $\mathbb{P}(12, 13 \in G_k)(\mathbb{P}(13 \in G_\ell)$. ∎

ACKNOWLEDGMENT

We thank an anonymous referee for a careful reading of the article and insightful comments.

REFERENCES

1. E. Abbé, *Community detection and stochastic block models: Recent developments*, J. Mach. Learn. Res. **18** (2018), 1–86.
2. H. Andersson and T. Britton, Stochastic epidemic models and their statistical analysis, Springer, New York, 2000.
3. M. Ángeles Serrano and M. Boguñá, *Clustering in complex networks. I. General formalism*, Phys. Rev. E **74** (2006), 056114.

4. M. Ángeles Serrano and M. Boguñá, *Clustering in complex networks. II. Percolation properties*, Phys. Rev. E **74** (2006), 056115.
5. F. G. Ball, D. J. Sirl, and P. Trapman, *Epidemics on random intersection graphs*, Ann. Appl. Probab. **24** (2014), no. 3, 1081–1128.
6. A. D. Barbour, L. Holst, and S. Janson, Poisson approximation, Oxford University Press, Oxford, 1992.
7. A. R. Benson, P. Liu, and H. Yin, A simple bipartite graph projection model for clustering in networks, 2020. arXiv preprint, arXiv:2007.00761.
8. M. Bloznelis, *The largest component in an inhomogeneous random intersection graph with clustering*, Electron. J. Combinator. **17** (2010), no. 1.
9. M. Bloznelis, *Degree and clustering coefficient in sparse random intersection graphs*, Ann. Appl. Probab. **23** (2013), no. 3, 1254–1289.
10. M. Bloznelis, *Local probabilities of randomly stopped sums of power law lattice random variables*, Lith. Math. J. **59** (2019), 437–468.
11. M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk, "*Recent progress in complex network analysis: Properties of random intersection graphs*," Data science, learning by latent structures, and knowledge discovery, Springer, Berlin, Heidelberg, 2015, pp. 79–88.
12. M. Bloznelis and J. Petuchovas, "*Correlation between clustering and degree in affiliation networks*," International workshop on algorithms and models for the web graph (WAW), Springer, Cham, 2017.
13. M. Bode, N. Fountoulakis, and T. Müller, *On the largest component of a hyperbolic model of complex networks*, Electron. J. Combinator. **22** (2015), no. 3.
14. B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Struct. Algoritm. **31** (2007), no. 1, 3–122.
15. B. Bollobás, S. Janson, and O. Riordan, *Sparse random graphs with clustering*, Random Struct. Algoritm. **38** (2011), no. 3, 269–323.
16. M. Bradonjić, A. Hagberg, N. W. Hengartner, and A. G. Percus, "*Component evolution in general random intersection graphs*," International workshop on algorithms and models for the web graph (WAW), Springer, Berlin, Heidelberg, 2010.
17. R. L. Breiger, *The duality of persons and groups*, Soc. Forces **53** (1974), no. 2, 181–190.
18. T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm, *Epidemics on random graphs with tunable clustering*, J. Appl. Probab. **45** (2008), no. 3, 743–756.
19. P. Colomer-de-Simón and M. Boguñá, *Double percolation phase transition in clustered complex networks*, Phys. Rev. X **4** (2014), 041020.
20. P. Diaconis and D. Freedman, *Finite exchangeable sequences*, Ann. Probab. **8** (1980), no. 4, 745–764.
21. A. Epasto, S. Lattanzi, and R. Paes Leme, *Ego-splitting framework: From non-overlapping to overlapping clusters*, ACM SIGKDD conference on knowledge discovery and data mining (KDD), Association for Computing Machinery, New York, (2017).
22. S. L. Feld, *The focused organization of social ties*, Am. J. Sociol. **86** (1981), no. 5, 1015–1035.
23. S. Foss, D. Korshunov, and S. Zachary, An introduction to heavy-tailed and subexponential distributions, Springer, New York, 2013.
24. N. Fountoulakis, P. van der Hoorn, T. Müller, and M. Schepers, *Clustering in a hyperbolic model of complex networks*, Electron. J. Probab. **26** (2021), 1–132.
25. A. Frieze and M. Karoński, Introduction to random graphs, Cambridge University Press, Cambridge, 2016.
26. E. Godehardt and J. Jaworski, *Two models of random intersection graphs and their applications*, Electron. Notes Discrete Math. **10** (2001), 129–132.
27. L. Iskhakov, B. Kamiński, M. Mironov, P. Prałat, L. Prokhorenkova, and D. Higham, *Local clustering coefficient of spatial preferential attachment model*, J. Complex Networks **8** (2020), no. 1, 1–32.
28. E. Jacob and P. Mörters, *Spatial preferential attachment networks: Power laws and clustering coefficients*, Ann. Appl. Probab. **25** (2015), no. 2, 632–662.
29. E. Jacob and P. Mörters, *Robustness of scale-free spatial networks*, Ann. Probab. **45** (2017), no. 3, 1680–1722.
30. S. Janson, T. Łuczak, and A. Ruciński, Random graphs, Wiley, 2000.
31. J. Karjalainen and L. Leskelä, "*Moment-based parameter estimation in binomial random intersection graph models*," International workshop on algorithms and models for the web graph (WAW), Springer, Cham, 2017.
32. J. Karjalainen, J. S. H. van Leeuwaarden, and L. Leskelä, "*Parameter estimators of sparse random intersection graphs with thinned communities*," International workshop on algorithms and models for the web graph (WAW), Springer, Cham, 2018.
33. M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen, *On random intersection graphs: The subgraph problem*, Comb. Probab. Comput. **8** (1999), no. 1–2, 131–159.
34. M. Kiwi and D. Mitsche, *On the second largest component of random hyperbolic graphs*, SIAM J. Discret. Math. **33** (2019), no. 4, 2200–2217.

35. D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, *Hyperbolic geometry of complex networks*, Phys. Rev. E **82** (2010), no. 3, 036106.
36. A. Krot and L. Ostroumova Prokhorenkova, "*Local clustering coefficient in generalized preferential attachment models*," International workshop on algorithms and models for the web graph (WAW), Springer, Cham, 2015.
37. A. N. Lagerås and M. Lindholm, *A note on the component structure in random intersection graphs with tunable clustering*, Electron. J. Combinator. **15** (2008), no. 1.
38. M. E. J. Newman, *The structure and function of complex networks*, SIAM Rev. **45** (2003), no. 2, 167–256.
39. V. V. Petrov, Sums of independent random variables, Springer, Berlin, Heidelberg, 1975. (Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82).
40. S. Petti and S. Vempala, *Approximating sparse graphs: The random overlapping communities model*, Random Struct. Algoritm. **61** (2022), 844–908.
41. P. G. Spirakis, S. Nikoletseas, and C. Raptopoulos, "*A guided tour in random intersection graphs*," International colloquium on automata, languages, and programming (ICALP), Springer, Berlin, Heidelberg, 2013.
42. J. M. Steele, *Le Cam's inequality and Poisson approximations*, Am. Math. Mon. **101** (1994), no. 1, 48–54.
43. C. Stegehuis, R. van der Hofstad, A. J. E. M. Janssen, and J. S. H. van Leeuwaarden, *Clustering spectrum of scale-free networks*, Phys. Rev. E **96** (2017), 042309.
44. V. Vadon, Local and global structure of networks with communities. PhD thesis, Technische Universiteit Eindhoven, 2020.
45. V. Vadon, J. Komjáthy, and R. van der Hofstad, *A new model for overlapping communities with arbitrary internal structure*, Appl. Network Sci. **4** (2019), no. 1, 42.
46. R. van der Hofstad, Random graphs and complex networks, Vol **I**, Cambridge University Press, Cambridge, 2017.
47. A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Large-scale topological and dynamical properties of the internet*, Phys. Rev. E **65** (2002), 066130.
48. J. Yang and J. Leskovec, "*Community-affiliation graph model for overlapping network community detection*," IEEE international conference on data mining (ICDM), IEEE, 2012.
49. J. Yang and J. Leskovec, *Structure and overlaps of ground-truth communities in networks*, ACM Trans. Intell. Syst. Technol. **5** (2014), no. 2, 1–35.
50. A. Zolotukhin, S. Nagaev, and V. Chebotarev, *On a bound of the absolute constant in the berry-Esseen inequality for i.i.d Bernoulli random variables*, Mod. Stoch.: Theory Appl. **5** (2018), no. 3, 385–410.

How to cite this article: M. Bloznelis and L. Leskelä, *Clustering and percolation on superpositions of Bernoulli random graphs*, Random Struct. Alg. **63** (2023), 283–342. https://doi.org/10.1002/rsa.21140

APPENDIX

Here, we prove Theorem 3.6(iv). The proof is similar to that of Theorem 3.3 above. In particular, we use notation of section 7. Furthermore, for a graph $F \subset G_{(n)}$ we denote by $\hat{F} = F \cap H$ the percolated graph, where each edge of F is retained independently with probability θ, see (12). We write $\hat{G} = \hat{G}_{(n)}$, $\hat{d} = \deg_{\hat{G}}(1)$. We also denote $\tilde{g}_{rs}^{(n)} = \mathrm{Bin}_{rs}(\tilde{P}_n)$, where $\tilde{P}_n$ is defined in (141) and where the mixed binomial distribution Bin_{rs} is defined in (5). The degree distribution $\hat{f}^{(n)}$ of $\hat{G}$ is defined as in (7), but with $\deg_{G_{(n)}}(i)$ replaced by $\deg_{\hat{G}_{(n)}}(i)$. Introduce events $\hat{\mathcal{K}}_3 = \{\hat{G} \supset K_3\}$ and $\hat{\mathcal{K}}_{12} = \{\hat{G} \supset K_{12}\}$ and denote

$$\overline{\Delta}_{rs} = \frac{m^3}{((n)_2)^3}(P_n)_{rs}(P_n)_{21}(P_n)_{21}, \qquad \hat{\Delta}_{rs} = \frac{m^2}{(n)_3(n)_2}(P_n)_{rs}(P_n)_{21}. \tag{A1}$$

We derive Theorem 3.6(iv) from the relations

(i) $\mathbb{P}(\hat{d}=t,\ \hat{\mathcal{K}}_3)=\theta^3\mu_{33}\hat{f}^{(n)}*\tilde{g}_{33}^{(n)}(t-2)+\hat{\varepsilon}(t)$, where

$$|\hat{\varepsilon}(t)|\le 4\mu_{21}\mu_{32}+\mu_{21}^3+\mu_{10,33}+2\hat{\Delta}_{44}+4\hat{\Delta}_{33};$$

(ii) $\mathbb{P}(\hat{d}=t,\ \hat{\mathcal{K}}_{12})=\theta^2\mu_{32}\hat{f}^{(n)}*\tilde{g}_{32}^{(n)}(t-2)+\theta^2\mu_{21}^2\hat{f}^{(n)}*\tilde{g}_{21}^{(n)}*\tilde{g}_{21}^{(n)}(t-2)+\overline{\varepsilon}(t)$, where

$$\begin{aligned}|\overline{\varepsilon}(t)|\le\ & 10\mu_{21}\mu_{32}+6\mu_{21}^3+\mu_{21}^4+4\mu_{10,32}+4\mu_{21}\mu_{10,21}+\mu_{21,21}+2(n-3)\mu_{32}^2\\ &+2\hat{\Delta}_{43}+4(\hat{\Delta}_{32}+\overline{\Delta}_{32}+\overline{\Delta}_{21}).\end{aligned}$$

We first show (i) and (ii). Afterwards we prove Theorem 3.6 (iv). Introduce events $\hat{\mathcal{A}}_k=\{\hat{G}_k\supset K_3\}$ and $\hat{\mathcal{A}}_{k\ell}=\hat{\mathcal{G}}_k^{12}\cap\hat{\mathcal{G}}_\ell^{13}$, where $\hat{\mathcal{G}}_k^{ij}$ is the event that $ij\in E(\hat{G}_k)$. Recall the overlay graphs $G_{-k}=\cup_{k'\neq k}G_{k'}$, $G_{k\ell}=G_k\cup G_\ell$ and $G_{-k\ell}=\cup_{q\notin\{k,\ell\}}G_q$ and denote

$$\begin{aligned}&\hat{d}_k=\deg_{\hat{G}_k}(1),\quad \hat{d}_{-k}=\deg_{\hat{G}_{-k}}(1),\quad \hat{d}_{k\ell}=\deg_{\hat{G}_{k\ell}}(1),\quad \hat{d}_{-k\ell}=\deg_{\hat{G}_{-k\ell}}(1),\\ &\delta_k=\deg_{G_k\cap G_{-k}}(1),\quad \delta_{k\ell}=\deg_{G_{k\ell}\cap G_{-k\ell}}(1),\quad \overline{\delta}_{k\ell}=\deg_{G_k\cap G_\ell}(1).\end{aligned}$$

We also denote $\hat{h}_{k\ell}(s)=\mathbb{P}(\hat{d}_{k\ell}=s,\hat{\mathcal{A}}_{k\ell})$, and $\overline{h}_{ki}(s)=\mathbb{P}(\hat{d}_k=s,\hat{\mathcal{G}}_k^{1i})$.

Proof of (i). The proof is similar to that of Theorem 7.3 (ii): we denote

$$\hat{\varepsilon}_1(t)=\mathbb{P}(\hat{d}=t,\hat{\mathcal{K}}_3)-\mathbb{P}(\hat{d}=t,\cup_k\hat{\mathcal{A}}_k),\quad \hat{\varepsilon}_2(t)=\mathbb{P}(\hat{d}=t,\cup_k\hat{\mathcal{A}}_k)-\sum_k\mathbb{P}(\hat{d}=t,\hat{\mathcal{A}}_k)$$

and estimate (cf. proof of Theorem 7.3 (ii))

$$\begin{aligned}&0\le\hat{\varepsilon}_1(t)\le\mathbb{P}(\mathcal{E}_{12})+\mathbb{P}(\mathcal{E}_{111})\le 3\mu_{21}\mu_{32}+\mu_{21}^3,\\ &0\le-\hat{\varepsilon}_2(t)\le\sum_{k,k'}{}'\mathbb{P}(\hat{d}=t,\mathcal{A}_k,\mathcal{A}_{k'})\le\sum_{k,k'}{}'\mathbb{P}(\mathcal{A}_k,\mathcal{A}_{k'})\le\mu_{21}\mu_{32}.\end{aligned}$$

Hence $\left|\mathbb{P}(\hat{d}=t,\hat{\mathcal{K}}_3)-\sum_k\mathbb{P}(\hat{d}=t,\hat{\mathcal{A}}_k)\right|\le 4\mu_{21}\mu_{32}+\mu_{21}^3$. We next approximate

$$\sum_k\mathbb{P}(\hat{d}=t,\hat{\mathcal{A}}_k)\ \approx\ \sum_k\sum_{r+s=t}\mathbb{P}(\hat{d}_{-k}=r)\,\mathbb{P}(\hat{d}_k=s,\hat{\mathcal{A}}_k)\tag{A2}$$

$$\approx\ \sum_k\sum_{r+s=t}\mathbb{P}(\hat{d}=r)\,\mathbb{P}(\hat{d}_k=s,\hat{\mathcal{A}}_k)\tag{A3}$$

and observe that $\sum_k\mathbb{P}(\hat{d}_k=s,\hat{\mathcal{A}}_k)=\theta^3\mu_{33}\tilde{g}_{33}^{(n)}(s-2)$ (the identity follows from Lemma 7.2 applied to $\hat{G}_{k^*}=G_{k^*}\cap H$, cf. (48)). Hence the right side of (A3) equals $\theta^3\mu_{33}\hat{f}^{(n)}*\tilde{g}_{33}^{(n)}(t-2)$, and to prove (i) it suffices to analyze the approximation errors in (A2), (A3).

To upper bound the approximation error in (A2), denoted $\hat{\varepsilon}_3(t)$, we apply the first inequality of Lemma A.1 with $F=K_3, A=\{k\}, B=[m]\setminus\{k\}\ \forall k$. We obtain

$$|\hat{\varepsilon}_3(t)|\le\sum_k 2\mathbb{P}(\mathcal{A}_k,\delta_k>0)\le 2\hat{\Delta}_{44}+4\hat{\Delta}_{33},$$

where in the last step we apply inequality (A12) of Lemma A.2.

The approximation error in (A3) equals $\hat{\varepsilon}_4(t) = \sum_k \hat{\varepsilon}_{4k}(t)$ where

$$\hat{\varepsilon}_{4k}(t) = \sum_{r+s=t} \left(\mathbb{P}(\hat{d} = r) - \mathbb{P}(\hat{d}_{-k} = r)\right) \mathbb{P}(\hat{d}_k = s, \hat{\mathcal{A}}_k).$$

Noting that $\deg_{G_k}(1) = 0$ implies $\hat{d} = \hat{d}_{-k}$ we estimate

$$\left|\mathbb{P}(\hat{d} = r) - \mathbb{P}(\hat{d}_{-k} = r)\right| \le \mathbb{P}(\deg_{G_k}(1) > 0) \le \mathbb{P}(1 \in V(G_k)) = p_{10}(k) \tag{A4}$$

and subsequantly

$$|\hat{\varepsilon}_{4k}(t)| \le p_{10}(k) \sum_{s=2}^{t} \mathbb{P}(\hat{d}_k = s, \mathcal{A}_k) \le p_{10}(k)\mathbb{P}(\mathcal{A}_k) = p_{10}(k)p_{33}(k).$$

Hence, $|\hat{\varepsilon}_4(t)| \le \sum_k p_{10}(k)p_{33}(k) = \mu_{10,33}$. Claim (i) follows by combining the above estimates for the total approximation error $\hat{\varepsilon}(t) = \sum_{i=1}^4 \hat{\varepsilon}_i(t)$.

Proof of (ii). The proof is similar to that of Theorem 7.4 (ii): we approximate

$$\mathbb{P}(\hat{d} = t, \hat{\mathcal{K}}_{12}) \approx \sum_{k,\ell} \mathbb{P}(\hat{d} = t, \hat{\mathcal{A}}_{k\ell}), \tag{A5}$$

$$\approx \sum_{k,\ell} \sum_{r+s=t} \mathbb{P}(\hat{d}_{-k\ell} = r)\, \hat{h}_{k\ell}(s), \tag{A6}$$

$$\approx \sum_{k,\ell} \sum_{r+s=t} \mathbb{P}(\hat{d} = r)\, \hat{h}_{k\ell}(s), \tag{A7}$$

so that

$$\mathbb{P}(\hat{d} = t, \hat{\mathcal{K}}_{12}) \approx \sum_{r+s=t} \hat{f}^{(n)}(r) \sum_{k} \hat{h}_{kk}(s) + \sum_{r+s=t} \hat{f}^{(n)}(r) {\sum_{k,\ell}}' \hat{h}_{k\ell}(s). \tag{A8}$$

Invoking identity $\sum_k \hat{h}_{kk}(s) = \theta^2 \mu_{32}\, \tilde{g}_{32}^{(n)}(s-2)$ (which follows by Lemma 7.2 applied to $\hat{G}_{k^*}$, cf. (54)) we write the first term on the right of (A8) in the form $\theta^2 \mu_{32} \hat{f}^{(n)} * \tilde{g}_{32}^{(n)}(t-2)$.

Next, we approximate the second inner sum on the right of (A8)

$${\sum_{k,\ell}}' \hat{h}_{k\ell}(s) \approx {\sum_{k,\ell}}' \sum_{s_1+s_2=s} \overline{h}_{k2}(s_1)\overline{h}_{\ell 3}(s_2), \tag{A9}$$

$$\approx \sum_{k,\ell} \sum_{s_1+s_2=s} \overline{h}_{k2}(s_1)\overline{h}_{\ell 3}(s_2). \tag{A10}$$

Invoking identities $\sum_k \overline{h}_{k2}(s) = \sum_\ell \overline{h}_{\ell 3}(s) = \theta \mu_{21} \tilde{g}_{21}^{(n)}(s-1)$ (the first one follows by symmetry; the second one follows by Lemma 7.2 applied to $\hat{G}_{k^*}$, cf. (54)) we write the right side of (A10) in the form $\theta^2 \mu_{21}^2 \tilde{g}_{21}^{(n)} * \tilde{g}_{21}^{(n)}(s-2)$. Hence, the second term on the right side of (A8) is approximately $\theta^2 \mu_{21}^2 \hat{f}^{(n)} * \tilde{g}_{21}^{(n)} * \tilde{g}_{21}^{(n)}(t-2)$. Finally, we conclude that

$$\mathbb{P}(\hat{d} = t, \hat{\mathcal{K}}_{12}) \approx \theta^2 \mu_{32} \hat{f}^{(n)} * \tilde{g}_{32}^{(n)}(t-2) + \theta^2 \mu_{21}^2 \hat{f}^{(n)} * \tilde{g}_{21}^{(n)} * \tilde{g}_{21}^{(n)}(t-2),$$

where the total approximation error can be written as $\overline{\varepsilon}(t) = \sum_{i=1}^{4} \overline{\varepsilon}_i(t)$. Here $\overline{\varepsilon}_1(t), \overline{\varepsilon}_2(t), \overline{\varepsilon}_3(t)$ are the approximation errors in (A5), (A6), (A7), respectively, and

$$\overline{\varepsilon}_4(t) = \sum_{r+s=t} \hat{f}^{(n)}(r) \left(\overline{\varepsilon}_{41}(s) + \overline{\varepsilon}_{42}(s)\right),$$

where $\overline{\varepsilon}_{41}(s)$ and $\overline{\varepsilon}_{42}(s)$ denote the errors made in (A9) and (A10).

Now, we analyze individual approximation errors $\overline{\varepsilon}_i(t)$. The cases $i = 1, 3$ are treated similarly as in the corresponding proof of Theorem 7.4 (ii) Bonferroni's inequalities imply

$$|\overline{\varepsilon}_1(t)| \leq \sum_{(k_1,k_2),(\ell_1,\ell_2)}^{\prime} \mathbb{P}(\hat{\mathcal{A}}_{k_1k_2}, \hat{\mathcal{A}}_{\ell_1\ell_2}) \leq \sum_{(k_1,k_2),(\ell_1,\ell_2)}^{\prime} \mathbb{P}(\mathcal{A}_{k_1k_2}, \mathcal{A}_{\ell_1\ell_2})$$

and therefore (61) extends to $|\overline{\varepsilon}_1(t)|$. Furthermore, invoking inequalities (cf. (A4))

$$\left|\mathbb{P}(\hat{d} = r) - \mathbb{P}(\hat{d}_{-k\ell} = r)\right| \leq \mathbb{P}(1 \in V(G_k)) + \mathbb{P}(1 \in V(G_\ell)) = p_{10}(k) + p_{10}(\ell)$$

and $\sum_{r+s=t} \hat{h}_{k\ell}(s) \leq \mathbb{P}(\hat{A}_{k\ell}) \leq \mathbb{P}(\mathcal{A}_{k\ell})$ we obtain

$$|\overline{\varepsilon}_3(t)| \leq \sum_{k,\ell} (p_{10}(k) + p_{10}(\ell))\, \mathbb{P}(\hat{\mathcal{A}}_{k\ell}) \leq \sum_{k,\ell} (p_{10}(k) + p_{10}(\ell))\, \mathbb{P}(\mathcal{A}_{k\ell}).$$

Therefore (63) extends to $|\overline{\varepsilon}_3(t)|$.

To upper bound $\overline{\varepsilon}_2(t)$ we apply the first inequality of Lemma A.1 with $F = K_{12}$, where we set $A = \{k, \ell\}$, $B = [m] \setminus \{k, \ell\}$ for $k \neq \ell$, and we set $A = \{k\}$, $B = [m] \setminus \{k\}$ for $k = \ell$. We obtain

$$|\overline{\varepsilon}_2(t)| \leq 2 \sum_{k,\ell} \mathbb{P}(\mathcal{A}_{k\ell}, \delta_{k\ell} > 0) \leq 2\hat{\Delta}_{43} + 4(\hat{\Delta}_{32} + \overline{\Delta}_{32} + \overline{\Delta}_{21}),$$

where the last inequality follows by inequality (A13) of Lemma A.2.

To upper bound $\overline{\varepsilon}_{41}(s)$ we apply the second inequality of Lemma A.1 with $A = \{k\}$, $B = \{\ell\}$ and F, F' being the edges 12, 13. We obtain

$$|\overline{\varepsilon}_{41}(s)| \leq \sum_{k\ell}^{\prime} 2\mathbb{P}\left(\mathcal{A}_{k\ell}, \overline{\delta}_{k\ell} > 0\right) \leq 2(n-3)\mu_{32}^2 + 2\mu_{32}\mu_{21},$$

where the last inequality follows by inequality (A11) of Lemma A.2.

Finally the approximation error in (A10)

$$|\overline{\varepsilon}_{42}(s)| = \sum_{k} \sum_{s_1+s_2=s} \mathbb{P}\left(\hat{d}_k = s_1, \hat{\mathcal{G}}_{12}^k\right) \mathbb{P}\left(\hat{d}_k = s_2, \hat{\mathcal{G}}_{12}^k\right)$$

satisfies $\sum_{s\geq 0} |\overline{\varepsilon}_{42}(s)| = \sum_k \left(\mathbb{P}\left(\hat{\mathcal{G}}_k^{12}\right)\right)^2 = \theta^2 \sum_k \left(p_{21}(k)\right)^2 = \theta^2 \mu_{21,21}$. We conclude that $|\overline{\varepsilon}_4(t)| \leq 2(n-3)\mu_{32}^2 + 2\mu_{32}\mu_{21} + \mu_{21,21}$.

Claim (ii) follows by collecting the bounds for $\overline{\varepsilon}_i$, $i = 1, 2, 3, 4$ altogether.

Proof of Theorem 3.6(iv). We evaluate the ratio $\sigma(\hat{G})(k) = \frac{\mathbb{P}(\hat{d}=t,\ \hat{\mathcal{K}}_3)}{\mathbb{P}(\hat{d}=t,\ \hat{\mathcal{K}}_{12})}$. To this aim we invoke approximations (i) and (ii) above, where respective leading terms

$$\theta^3\mu_{33}\,\hat{f}^{(n)} * \tilde{g}_{33}^{(n)}(t-2) = (\hat{P})_{33}\mu n^{-2}\hat{f} * \hat{g}_{33}(t-2) + o(n^{-2})$$

and

$$\begin{aligned}&\theta^2\mu_{32}\,\hat{f}^{(n)} * \tilde{g}_{32}^{(n)}(t-2) + (\theta\mu_{21})^2\hat{f}^{(n)} * \tilde{g}_{21}^{(n)} * \tilde{g}_{21}^{(n)}(t-2)\\ &= (\hat{P})_{32}\mu n^{-2}\hat{f} * \hat{g}_{32}(t-2) + (\hat{P})_{21}^2\mu^2 n^{-2}\hat{f} * \hat{g}_{21} * \hat{g}_{21}(t-2) + o(n^{-2})\end{aligned}$$

are of order $\Theta(n^{-2})$ as $n \to +\infty$. It remains to show that the remainders $\hat{\varepsilon}(t)$ and $\overline{\varepsilon}(t)$ of (i) and (ii) are of order $o(n^{-2})$. In the proof of Theorem 3.3 all the terms contributing to $\hat{\varepsilon}(t)$ and $\overline{\varepsilon}(t)$ are shown to be $o(n^{-2})$, but $\hat{\Delta}_{43}$, $\hat{\Delta}_{44}$, $\hat{\Delta}_{33}$, $\hat{\Delta}_{32}$, $\overline{\Delta}_{32}$, $\overline{\Delta}_{21}$, and $(n-3)\mu_{32}^2$.

Here, we only show that $\hat{\Delta}_{43} = o(n^{-2})$. The proof of $o(n^{-2})$ bound for the remaining terms is easy. We have $\hat{\Delta}_{4,3} \lesssim n^{-3}(P_n)_{43} = n^{-3}\mathbb{E}(X_{n,\pi}^4 Q_{n,\pi}^3)$, where $(X_{n,\pi}, Q_{n,\pi})$ is a bivariate random variable with the distribution P_n. Our conditions $(P_n)_{3,3} \to (P)_{33}$ and $P_n \xrightarrow{w} P$ implies that the sequence $\{X_{n,\pi}^3 Q_{n,\pi}^3, n \geq 1\}$ is uniformly integrable. Note also that $\mathbb{P}(X_{n\pi} \leq n) = 1\ \forall n$. Now fix $0 < \tau < 1$ and split

$$\begin{aligned}(P_n)_{43} = \mathbb{E}(X_{n,\pi}^4 Q_{n,\pi}^3) &= \mathbb{E}\left(X_{n,\pi}^4 Q_{n,\pi}^3)\mathbb{I}_{\{X_{n,\pi}<\tau n\}}\right) + \mathbb{E}\left(X_{n,\pi}^4 Q_{n,\pi}^3\mathbb{I}_{\{X_{n,\pi}\geq\tau n\}}\right)\\ &\leq \tau n\mathbb{E}\left(X_{n,\pi}^3 Q_{n,\pi}^3\right) + n\mathbb{E}\left(X_{n,\pi}^3 Q_{n,\pi}^3\mathbb{I}_{\{X_{n,\pi}\geq\tau n\}}\right).\end{aligned}$$

The last term above is $o(n)$, by the uniform integrability property. Hence $(P_n)_{43} = o(n)$. Consequently, we have $\hat{\Delta}_{43} = o(n^{-2})$.

In the next lemma we use notation $G_A = (V(G_A), E(G_A))$, $A \subset [m]$, introduced in section 11. We also denote N_A the set of neighbors of vertex 1 in G_A and write $D_A = |N_A|$. For $A, B \subset [m]$ we denote $\delta_{A,B} = |N_A \cap N_B|$. We denote by $\hat{D}_A$ the degree of vertex 1 in the percolated graph $\hat{G}_A = G_A \cap H$. Given a graph F with the vertex set $V(F) \subset V = \{1, \ldots, n\}$ we denote by $\mathcal{A}_{F,A}$ and $\hat{\mathcal{A}}_{F,A}$ the events that $F \subset G_A$ and $F \subset \hat{G}_A$ respectively. Lemma A.1 is used in the proof of claims (i), (ii) above where we take $F = K_3, K_{12}$.

Lemma A.1. *Let $A, B \subset [m]$. Assume that $A \cap B = \emptyset$. Assume that $1 \in V(F)$. Then*

$$\left|\mathbb{P}(\hat{D}_{A\cup B} = t, \hat{\mathcal{A}}_{F,A}) - \sum_{r+s=t}\mathbb{P}(\hat{D}_B = r)\,\mathbb{P}(\hat{D}_A = s, \hat{\mathcal{A}}_{F,A})\right| \leq 2\mathbb{P}(\mathcal{A}_{F,A}, \delta_{A,B} > 0).$$

Assume, in addition, that graph F' has vertex set $V(F') \subset V$ and $1 = V(F) \cap V(F')$. Then

$$\begin{aligned}&\left|\mathbb{P}(\hat{D}_{A\cup B} = t, \hat{\mathcal{A}}_{F\cup F',A\cup B}, \mathcal{A}_{F,A}, \mathcal{A}_{F',B}) - \sum_{r+s=t}\mathbb{P}(\hat{D}_B = r, \hat{\mathcal{A}}_{F',B})\,\mathbb{P}(\hat{D}_A = s, \hat{\mathcal{A}}_{F,A})\right|\\ &\leq 2\mathbb{P}(\mathcal{A}_{F,A}, \mathcal{A}_{F',B}, \delta_{A,B} > 0).\end{aligned}$$

Proof. We only prove the first inequality. The proof of the second one is much the same.

In the proof we write, for short, $\mathcal{A} = \mathcal{A}_{F,A}$ and $\hat{\mathcal{A}} = \hat{\mathcal{A}}_{F,A}$. We note that random variables D_A and D_B are independent and $\delta_{A,B} = 0$ implies $D_{A\cup B} = D_A + D_B$ and, consequently, $\hat{D}_{A\cup B} = \hat{D}_A + \hat{D}_B$.

Introduce Bernoulli random graph $\tilde{H}$ with $V(H) = V$ and edge probability θ, which is independent of $G_1, \dots, G_m$ and H. Let $\tilde{D}_B$ denote the degree of vertex 1 in the percolated graph $\tilde{G}_B = G_B \cap \tilde{H}$. Clearly, random variables $\hat{D}_B$ and $\tilde{D}_B$ have the same probability distributions. Moreover, on the event $\delta_{A,B} = 0$ the pairs $\big((\hat{D}_A, \hat{\mathcal{A}}); (\hat{D}_B)\big)$ and $\big((\hat{D}_A, \hat{\mathcal{A}}); (\tilde{D}_B)\big)$ have the same distributions. Hence

$$\mathbb{P}\big(\hat{D}_{A\cup B} = t, \hat{\mathcal{A}}, \delta_{A,B} = 0\big) = \mathbb{P}\big(\hat{D}_A + \hat{D}_B = t, \hat{\mathcal{A}}, \delta_{A,B} = 0\big) = \mathbb{P}\big(\hat{D}_A + \tilde{D}_B = t, \hat{\mathcal{A}}, \delta_{A,B} = 0\big).$$

Consequently, we obtain

$$\left|\mathbb{P}(\hat{D}_{A\cup B} = t, \hat{\mathcal{A}}) - \mathbb{P}(\hat{D}_A + \tilde{D}_B = t, \hat{\mathcal{A}})\right| \le 2\mathbb{P}(\mathcal{A}, \delta_{A,B} > 0).$$

Next, by the independence of $\tilde{D}_B$ and $(\hat{D}_A, \hat{\mathcal{A}})$ we can factorize the probability $\mathbb{P}(\hat{D}_A + \tilde{D}_B = t, \hat{\mathcal{A}}) = \sum_{r+s=t} \mathbb{P}(\hat{D}_A = s, \hat{\mathcal{A}})\mathbb{P}(\tilde{D}_B = r)$. Now the identity $\mathcal{L}(\tilde{D}_B) = \mathcal{L}(\hat{D}_B)$ completes the proof.

To prove the second inequality of the lemma we note that $V(F) \cap V(F') = 1$ implies that F and F' have no common edges. Therefore, given the event $F \subset G_A$, $F' \subset G_B$ and $\delta_{A,B} = 0$, the percolation process acts on F and F' independently. The rest of the proof is much the same as above. ∎

Lemma A.2. *For $\overline{\Delta}_{rs}$ and $\hat{\Delta}_{rs}$ defined in (A1), we have*

$$\sum_{k\ell}{}' \mathbb{P}(\mathcal{A}_{k\ell}, \overline{\delta}_{k\ell} > 0) \le (n-3)\mu_{32}^2 + 2\mu_{21}\mu_{32}, \tag{A11}$$

$$\sum_{k\in[m]} \mathbb{P}(\mathcal{A}_k, \delta_k > 0) \le \hat{\Delta}_{44} + 2\hat{\Delta}_{33}, \tag{A12}$$

$$\sum_{k,\ell} \mathbb{P}(\mathcal{A}_{k\ell}, \delta_{k\ell} > 0) \le \hat{\Delta}_{43} + 2(\hat{\Delta}_{32} + \overline{\Delta}_{32} + \overline{\Delta}_{21}). \tag{A13}$$

Proof. By $\overline{\mathbb{P}} = \mathbb{P}_{(\overline{X}_n, \overline{Q}_n)}$ we denote the conditional probability given $(\overline{X}_n, \overline{Q}_n)$.

Proof of (A11). We write event $\{\mathcal{A}_{k\ell}, \overline{\delta}_{k\ell} > 0\}$ in the form

$$\mathcal{G}_k^{12} \cap \mathcal{G}_\ell^{13} \cap \left(\mathcal{G}_k^{13} \cup \mathcal{G}_\ell^{12} \cup \bigcup_{s=4}^{n} \left(\mathcal{G}_k^{1s} \cap \mathcal{G}_\ell^{1s}\right)\right).$$

Invoking identities

$$\overline{\mathbb{P}}(\mathcal{G}_k^{12}, \mathcal{G}_\ell^{13}, \mathcal{G}_k^{13}) = \frac{(X_k)_3}{(n)_3} Q_k^2 \frac{(X_\ell)_2}{(n)_2} Q_\ell, \qquad \overline{\mathbb{P}}(\mathcal{G}_k^{12}, \mathcal{G}_\ell^{13}, \mathcal{G}_\ell^{12}) = \frac{(X_k)_2}{(n)_2} Q_k \frac{(X_\ell)_3}{(n)_3} Q_\ell^2,$$
$$\overline{\mathbb{P}}(\mathcal{G}_k^{12}, \mathcal{G}_\ell^{13}, \mathcal{G}_k^{1s}, \mathcal{G}_\ell^{1s}) = \frac{(X_k)_3}{(n)_3} Q_k^2 \frac{(X_\ell)_3}{(n)_3} Q_\ell^2$$

we obtain, by the union bound and symmetry,

$$\overline{\mathbb{P}}(\mathcal{A}_{k\ell}, \overline{\delta}_{k\ell} > 0) \le \frac{(X_k)_3}{(n)_3} Q_k^2 \frac{(X_\ell)_2}{(n)_2} Q_\ell + \frac{(X_k)_2}{(n)_2} Q_k \frac{(X_\ell)_3}{(n)_3} Q_\ell^2 + (n-3)\frac{(X_k)_3}{(n)_3} Q_k^2 \frac{(X_\ell)_3}{(n)_3} Q_\ell^2.$$

Consequently,

$$\mathbb{P}\big(\mathcal{A}_{k\ell},\bar{\delta}_{k\ell}>0\big)=\mathbb{E}\overline{\mathbb{P}}\big(\mathcal{A}_{k\ell},\bar{\delta}_{k\ell}>0\big)\\ \le p_{32}(k)p_{21}(\ell)+p_{21}(k)p_{32}(\ell)+(n-3)p_{32}(k)p_{32}(\ell).$$

Now, summing over $k\neq\ell$ we obtain (A11).

Proof of (A12). We write $\delta_k=\delta'_k+\delta''_k$, where δ'_k (respectively δ''_k) is the number of vertices $s\in V(G)\setminus\{1,2,3\}$ (respectively $s\in\{2,3\}$) such that $1s\in E(G_k)\cap E(G_{-k})$. Then $\mathbb{P}(\mathcal{A}_k,\delta_k>0)\le\mathbb{P}(\mathcal{A}_k,\delta'_k>0)+\mathbb{P}(\mathcal{A}_k,\delta''_k>0)$ and we estimate the probabilities on the right separately.

To evaluate $\mathbb{P}(\mathcal{A}_k,\delta'_k>0)$ we write event $\{\delta'_k>0\}$ in the form

$$\bigcup_{s\in V(G_k)\setminus\{1,2,3\}}\left(\mathcal{G}_k^{1s}\cap\left(\bigcup_{j\in[m]\setminus\{k\}}\mathcal{G}_j^{1s}\right)\right). \tag{A14}$$

Then, using the union bound and symmetry, we upper bound the conditional probability

$$\overline{\mathbb{P}}(\mathcal{A}_k,\delta'_k>0)\le(X_k-3)\frac{(X_k)_3}{(n)_3}Q_k^4\sum_{j\in[m]\setminus\{k\}}\frac{(X_j)_2}{(n)_2}Q_j. \tag{A15}$$

Here, X_k-3 is the number of vertices $s\in V(G_k)\setminus\{1,2,3\}$. $\frac{(X_k)_3}{(n)_3}$ is the probability that $\{1,2,3\}\subset V(G_k)$. Similarly, $\frac{(X_j)_2}{(n)_2}$ is the probability that $\{1,s\}\subset V(G_j)$. Next, we take the expectation $\mathbb{E}\overline{\mathbb{P}}(\mathcal{A}_k,\delta'_k>0)=\mathbb{P}(\mathcal{A}_k,\delta'_k>0)$ and sum over k to obtain

$$\sum_k\mathbb{P}(\mathcal{A}_k,\delta'_k>0)\le{\sum_{kj}}'\frac{\mathbb{E}((X_k)_4Q_k^4)}{(n)^3}\frac{\mathbb{E}((X_j)_2Q_j)}{(n)_2}. \tag{A16}$$

Note that, the right side is upper bounded by

$$\sum_k\frac{\mathbb{E}((X_k)_4Q_k^4)}{(n)^3}\times\sum_j\frac{\mathbb{E}((X_j)_2Q_j)}{(n)_2}=\hat{\Delta}_{44}. \tag{A17}$$

Now, we evaluate $\mathbb{P}(\mathcal{A}_k,\delta''_k>0)$. We write event $\{\mathcal{A}_k,\delta''_k>0\}$ in the form

$$\mathcal{A}_k\cap\bigcup_{s\in\{2,3\}}\left(\bigcup_{j\in[m]\setminus\{k\}}\mathcal{G}_j^{1s}\right)$$

and proceeding similarly as in (A15) above we obtain that

$$\mathbb{P}(\mathcal{A}_k,\delta''_k>0)\le2\sum_{j\in[m]\setminus k}\frac{\mathbb{E}((X_k)_3Q_k^3)}{(n)^3}\frac{\mathbb{E}((X_j)_2Q_j)}{(n)_2}\le2\frac{\mathbb{E}((X_k)_3Q_k^3)}{(n)_3}m\frac{(P_n)_{21}}{(n)_2}. \tag{A18}$$

Consequently, the sum $\sum_k\mathbb{P}(\mathcal{A}_k,\delta''_k>0)\le2\hat{\Delta}_{33}$. The latter bound together with (A17) implies (A12).

Proof of (A13). We write $\delta_{k\ell}=\delta'_{k\ell}+\delta''_{k\ell}$, where $\delta'_{k\ell}$ (respectively $\delta''_{k\ell}$) is the number of vertices $s\in V(G)\setminus\{1,2,3\}$ (respectively $s\in\{2,3\}$) such that $1s\in E(G_{k\ell})\cap E(G_{-k\ell})$.

We split

$$\sum_{k,\ell}\mathbb{P}\big(\mathcal{A}_{k\ell},\delta_{k\ell}>0\big)=\sum_{k\ell}{}^{'}\mathbb{P}\big(\mathcal{A}_{k\ell},\delta_{k\ell}>0\big)+\sum_{k}\mathbb{P}\big(\mathcal{A}_{kk},\delta_{kk}>0\big)=:S_1+S_2$$

and estimate

$$S_1\le\sum_{k\ell}{}^{'}\mathbb{P}\big(\mathcal{A}_{k\ell},\delta'_{k\ell}>0\big)+\sum_{k\ell}{}^{'}\mathbb{P}\big(\mathcal{A}_{k\ell},\delta''_{k\ell}>0\big)=:S'_1+S''_1.$$

In what follows we upper bound S'_1, S''_1 and S_2. To estimate S'_1 we write event $\{\delta'_{k\ell}>0\}$ in the form

$$\bigcup_{s\in V(G_{k\ell})\setminus\{1,2,3\}}\left(\mathcal{G}_k^{1s}\cup\mathcal{G}_\ell^{1s}\cap\left(\bigcup_{j\in[m]\setminus\{k,\ell\}}\mathcal{G}_j^{1s}\right)\right)$$

and upper bound the conditional probability $\overline{\mathbb{P}}(\mathcal{A}_{k\ell},\delta'_{k\ell}>0)$ using the union bound and symmetry,

$$\begin{aligned}\overline{\mathbb{P}}(\mathcal{A}_{k\ell},\delta'_{k\ell}>0)\le\frac{(X_k)_2Q_k}{(n)_2}\frac{(X_\ell)_2Q_\ell}{(n)_2}\left((X_k-2)Q_k+(X_\ell-2)Q_\ell\right)\\ \times\sum_{j\in[m]\setminus\{k,\ell\}}\frac{(X_j)_2Q_j}{(n)_2}.\end{aligned}\tag{A19}$$

Here, $\frac{(X_k)_2Q_k}{(n)_2}\frac{(X_\ell)_2Q_\ell}{(n)_2}$ is the (conditional) probability of event $\mathcal{A}_{k\ell}$; $(X_k-2)Q_k$ counts $s\in V(G_k)\setminus\{1,2\}$ linked to 1 by the layer G_k; $(X_\ell-2)Q_\ell$ counts $s\in V(G_\ell)\setminus\{1,3\}$ linked to 1 by the layer G_ℓ. Furthermore, each such s is linked to 1 by $G_{-k\ell}$ with probability $\overline{\mathbb{P}}(\cup_{j\in[m]\setminus\{k,\ell\}}\mathcal{G}_j^{1s})\le\sum_{j\in[m]\setminus\{k,\ell\}}\frac{(X_j)_2Q_j}{(n)_2}$. It follows from (A19) that

$$S'_1=\sum_{k\ell}{}^{'}\mathbb{E}\overline{\mathbb{P}}(\mathcal{A}_{k\ell},\delta'_{k\ell}>0)\le2\frac{m^3}{(n)_2^3}(P_n)_{32}(P_n)_{21}^2=2\overline{\Delta}_{32}.$$

We similarly setimate S''_1. The event $\delta''_{k\ell}>0$ means that at least one link 12 or/and 13 is present in $G_{-k\ell}$. Hence

$$\overline{\mathbb{P}}(\mathcal{A}_{k\ell},\delta''_{k\ell}>0)\le2\frac{(X_k)_2Q_k}{(n)_2}\frac{(X_\ell)_2Q_\ell}{(n)_2}\times\sum_{j\in[m]\setminus\{k,\ell\}}\frac{(X_j)_2Q_j}{(n)_2}.$$

Consequently, $S''_1\le\sum'_{k\ell}\mathbb{E}\overline{\mathbb{P}}(\mathcal{A}_{k\ell},\delta_{k\ell}-\delta'_{k\ell}>0)\le2\frac{m^3}{(n)_2^3}(P_n)_{21}^3=2\overline{\Delta}_{21}$.

Finally, we show $S_2\le\hat{\Delta}_{43}+2\hat{\Delta}_{32}$ in much the same way as (A12) above. Collecting the bounds for S'_1, S''_1, and S_2 we obtain (A13). ■