Evolution of a Modified Binomial Random Graph by Agglomeration

Mihyun Kang1 · Angelica Pachon2 · Pablo M. Rodríguez3

Received: 16 March 2017 / Accepted: 8 December 2017 / Published online: 15 December 2017
© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract In the classical Erdős–Rényi random graph $G(n, p)$ there are $n$ vertices and each of the possible edges is independently present with probability $p$. The random graph $G(n, p)$ is homogeneous in the sense that all vertices have the same characteristics. On the other hand, numerous real-world networks are inhomogeneous in this respect. Such an inhomogeneity of vertices may influence the connection probability between pairs of vertices. The purpose of this paper is to propose a new inhomogeneous random graph model which is obtained in a constructive way from the Erdős-Rényi random graph $G(n, p)$. Given a configuration of $n$ vertices arranged in $N$ subsets of vertices (we call each subset a super-vertex), we define a random graph with $N$ super-vertices by letting two super-vertices be connected if and only if there is at least one edge between them in $G(n, p)$. Our main result concerns the threshold for connectedness. We also analyze the phase transition for the emergence of the giant component and the degree distribution. Even though our model begins with $G(n, p)$, it assumes the existence of some community structure encoded in the configuration. Furthermore, under certain conditions it exhibits a power law degree distribution. Both properties are important for real-world applications.

Keywords Erdős–Rényi model · Random graph · Inhomogeneous random graph · Connectedness · Phase transition

Pablo M. Rodríguez
pablor@icmc.usp.br

Mihyun Kang
kang@math.tugraz.at

Angelica Pachon
apachonpinzon@unito.it

1 Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria

2 Dipartimento di Matematica, Università di Torino, Via Carlo Alberto 10, 10123 Turin, Italy

3 Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Av. Trabalhador São-Carlense 400 - Centro, CEP 13560-970 São Carlos, SP, Brazil
1 Introduction

The subject of random graphs began in 1959–1960 with the papers “On random graphs I” and “On the evolution of random graphs” by Erdős and Rényi [16, 17]. Since then, many properties of the Erdős–Rényi random graph have been analyzed in order to answer questions of mathematical and physical interest.

The original model studied by Erdős and Rényi is the uniform random graph $G(n, M)$, which is a graph chosen uniformly at random among all graphs with vertex set $[n] := \{1, 2, \ldots, n\}$ and exactly $M$ edges. It is not difficult to see that $G(n, M)$ is closely related to the binomial random graph $G(n, p)$, which is a graph with vertex set $[n]$, in which each pair of vertices is connected by an edge with probability $p$, independently of each other. The latter model was introduced by Gilbert [19] at about the same time. It is well known that the two random graph models $G(n, p)$ and $G(n, M)$ are essentially equivalent for the correct choice of $M$ and $p$. Due to this equivalence and the deep and notable results proved in [16, 17], the binomial random graph $G(n, p)$ is also known in the literature as the Erdős–Rényi random graph.

One of the most classical results on the Erdős–Rényi random graph is the threshold for connectedness, which is closely related to the non-existence of isolated vertices. Clearly, when there exists at least one isolated vertex the graph is disconnected, but the opposite implication is not generally true. Remarkably, it was shown that when there are no isolated vertices, the random graph $G(n, p)$ is connected with high probability (in short whp) which means with probability tending to one as $n$ goes to $\infty$ (see [4, 17]).

Many other properties were also studied in [17]. One of the most striking results was the discovery of a drastic change in the size of the largest component when the number of edges passes through $n/2$. This phenomenon is related to the phase transition in percolation, a model that is well studied in mathematical physics and is one of the main branches of contemporary probability, see [3, 20]. However, perhaps the first to talk about this kind of phenomena were Flory and Stockmayer, using a very different language coming from polymer physics, concerning gelation rather than percolation, and a gel rather than a giant component, see for instance [18].

During the last few decades, increasing attention in the field of random graphs has been devoted to find models that describe the complexity of real-world networks. It has recently been observed that many real-world networks are inhomogeneous, in the sense that they may contain distinct groups of nodes with distinct types of probabilistic behavior (see, for example, [24]). A general theoretical model of an inhomogeneous random graph is proposed in the seminal paper of Bollobás et al. [6], who considered a conditional independence between the edges, where the number of edges is linear in the number of vertices. This model includes many models previously studied in the literature, for instance Bollobás et al. [5] and Durrett [14]. A special case of this general inhomogeneous random graph model is the so-called rank-1 case, which is closely related to other models studied by [7, 9–12, 25]. In the inhomogeneous random graph model introduced in [6], it is shown that under a weak (convergence) assumption on the expected number of edges, many interesting properties can be determined, in particular the critical point of the phase transition and the size of the giant component. More recently, van der Hofstad [29] analyzed the size of the largest connected component in the critical regime of the inhomogeneous random graph studied in [25], where...
weights are associated with the vertices of the graph, and edges are present between vertices with a probability that is approximately proportional to the product of the respective weights. These results are also extended to graphs studied in [7,10].

In this paper we propose a new inhomogeneous random graph model that is obtained in a constructive way from the classical Erdős-Rényi model. By “a constructive way”, we mean an explicit scheme for constructing the graph from a given realization of $G(n, p)$. Given a partition of the vertex set $[n]$ of $G(n, p)$, in which each partition class represents an agglomeration of nodes, we call each partition class a super-vertex and the partition a configuration of super-vertices. We define an inhomogeneous random graph model by letting two super-vertices be connected if and only if there is at least one edge between them in $G(n, p)$. Note that our model assumes the existence of a kind of community structure by the agglomeration of the nodes, which is encoded in the super-vertices. However, we are not assuming that the vertices inside each super-vertex should be all connected. In other words, each super-vertex is not necessarily a clique in $G(n, p)$ and can be any kind of subgraph of $G(n, p)$. Related random graph models are analyzed, for instance, by Janson and Spencer [22] and by Seshadhri et al. [26].

The main contributions of this paper are fourfold. We determine (i) the threshold for the connectedness of our inhomogeneous random graph model (Theorem 1); (ii) the threshold for the existence of the giant component formed by super-vertices (Proposition 1); (iii) the degree distribution of a super-vertex (Proposition 2). Finally we show that under certain conditions our model exhibits a power-law degree distribution (Example 1), which is an important property for real applications.

In order to determine the asymptotic probability of our model being connected, we analyze the distribution of the number of isolated super-vertices, using the second moment method (in Lemma 1), as well as Stein’s method (in Lemma 3). As for the threshold for the existence of the giant component and the degree distribution of super-vertices, we show that our model can be viewed as a special case of the inhomogeneous random graph (IRG) studied in [6], by identifying a graphical sequence of kernels and applying the corresponding results for IRG in [6]. We also discuss how these results for the threshold for the existence of the giant component and the degree distribution of super-vertices can be obtained, when our model is compared to the rank-1 random graph (in particular, the Poisson random graph as proposed by Norros and Reittu [25]).

The rest of the paper is organized as follows. In Sect. 2 we introduce our inhomogeneous random graph model, which is followed by our main results and related work. The proof of the location of the threshold for connectedness (Theorem 1) is provided in Sect. 3. The proofs for the emergence of the giant component and the degree distribution are given in Sect. 4. Finally, in Sect. 5 we add some concluding remarks.

2 Our Model and Main Results

2.1 The Model

The motivation of our model comes from real-world applications that exhibit a community structure. Think for instance communities like cities, families, criminal groups or criminal organizations. In order to reflect a possible community structure we shall define a random graph with a given number of super-vertices (also known as agglomerates) of given sizes, in which we assume that the underlying graph follows the $G(n, p)$-law.
More precisely speaking, for each $N \in \mathbb{N} := \{1, 2, \ldots\}$ we let $r \in \mathbb{N} \cup \{\infty\}$ be either a constant independent of $N$ or a function in $N$ such that $r = r(N)$ tends to a constant or $\infty$ as $N \to \infty$. We let $K^r := \{(k_1, \ldots, k_r) \in \mathbb{N}^r : \sum_{i=1}^{r} k_i = N\}$ and $p = p(N, K^r) \in [0, 1]$ be given. Note that $K^r$ and $k_1, \ldots, k_r$ depend on $N$, but for the sake of simplicity, we suppress this dependence in our notation. We may use the notation $K^\infty$ when there exists a super-vertex of size $r = r(N)$, where $r(N) \to \infty$ as $N \to \infty$.

Given a partition of the vertex set $[n]$ of $G(n, p)$, we call a partition class of size $i$ a super-vertex of size $i$. Note that the vertices in each super-vertex are not necessarily connected in $G(n, p)$. We define $G(N, K^r, p)$ to be a random graph with $N$ super-vertices with configuration $K^r$, in which for each $i, j = 1, 2, \ldots, r$ there are $k_i$ super-vertices of size $i$ and an edge between a pair of two distinct super-vertices of sizes $i$ and $j$ is present with probability

$$p_{ij} := 1 - (1 - p)^{ij},$$

independently of each other. In words, $p_{ij}$ is the probability that there is at least one edge between the corresponding partition classes of the vertex set $[n]$ of $G(n, p)$, see Fig. 1. Note that the number of super-vertices and the number of vertices are given by $N = \sum_{i=1}^{r} k_i$ and $n = \sum_{i=1}^{r} ik_i = \sum_{v \in [N]} \text{size}(v)$, respectively. Thus $n = n(N)$ and $N \to \infty$ implies $n \to \infty$.

Our goal is to study properties of the random graph $G(N, K^r, p)$ by considering different values of $p$ as a function of both $n$ and $N$. It seems difficult to obtain substantial results for $G(N, K^r, p)$ without further restrictions. Throughout the paper we therefore assume that for each $i \in \mathbb{Z}^+$ the following limit exists

$$\mu_i := \lim_{N \to \infty} \frac{k_i}{N},$$

and that $\mu_i > 0$ for some $i \in \mathbb{Z}^+$, which means that for $N$ sufficiently large, there are linearly many super-vertices of size $i$ for some $i \in \mathbb{Z}^+$. Note that for $i > r$, $k_i = 0$ so $\mu_i = 0$. Furthermore, we define

$$i_\ast = i_\ast(N) := \arg \max_i k_i^{1/(i(n-i))},$$

![Fig. 1](image-url)  
(a)

construction of $G(N, K^r, p)$. (a) Begin with a fixed configuration $K^r$ of $N$ super-vertices, that is, subsets of vertices (there are $n$ vertices in total). (b) Connect every pair of vertices independently with probability $p$, in other words, take a realization of $G(n, p)$. (c) A pair of super-vertices in $G(N, K^r, p)$ is connected if and only if there is at least one edge between the corresponding subsets of vertices in $G(n, p)$.
i.e. \( i_* = \left\{ i \mid k_i^{1/(n-r)} = \max_{1 \leq j \leq r} k_j^{1/(n-r)} \right\} \). The quantity \( i_* \) will turn out to be crucial for the existence of isolated super-vertices and therefore for the threshold for connectedness.

Observe that if \( r \) is either a constant independent of \( N \) or a function in \( N \) such that \( r = r(N) \) tends to a constant as \( N \to \infty \), we may assume that \( 1 \leq r \leq \tilde{r} \) for some constant \( \tilde{r} \in \mathbb{N} \), and for values of \( N \) sufficiently large. This guarantee that \( i_* \) is bounded from above by a constant independent of \( N \), which implies by (2) that \( \lim_{N \to \infty} n/N \) exists and is equal to \( u := \sum_{i \in \mathbb{Z}^+} i \mu_i \).

### 2.2 Main Results

The main result of this paper concerns the exact location of the threshold for connectedness in \( G(N, K^r, p) \). The property of connectedness has not been addressed in the study of the general inhomogeneous random graph model (IRG) by Bollobás et al. [6]. For a special class of IRG, the connectedness has been studied by Devroye and Fraiman [13]. The model presented in this paper is related to a case that has not been covered in [13] (see Sect. 5.2 for details).

**Theorem 1** Let \( r \in \mathbb{N} \) be either a constant independent of \( N \) or a function in \( N \) such that \( r = r(N) \) tends to a constant as \( N \to \infty \). Let \( i_* \) be defined by (3), let \( c(N) \) be a function in \( N \) satisfying

\[
-\ln k_{i_*} < c(N) < i_*(n - i_*) - \ln k_{i_*}
\]

and let

\[
\tilde{c} := \lim_{N \to \infty} c(N) \in [-\infty, \infty].
\]

Consider the random graph \( G(N, K^r, p) \) with

\[
p := \frac{\ln k_{i_*} + c(N)}{i_*(n - i_*)},
\]

and assume that (2) holds.

1. If \( \tilde{c} = -\infty \), then

\[
\lim_{N \to \infty} \mathbb{P}[ G(N, K^r, p) \text{ is connected } ] = 0.
\]

2. If \( \tilde{c} = c \in \mathbb{R} \) is a constant, then

\[
\lim_{N \to \infty} \mathbb{P}[ G(N, K^r, p) \text{ is connected } ] = e^{-e^{-c} - \gamma},
\]

where \( \gamma \) is defined as

\[
\gamma := \lim_{N \to \infty} \sum_{i \neq i_*} k_i (k_i e^c)^{i/(n-i_* - i)}. \tag{7}
\]

3. If \( \tilde{c} = +\infty \), then

\[
\lim_{N \to \infty} \mathbb{P}[ G(N, K^r, p) \text{ is connected } ] = 1.
\]

In order to state our next results we need some more notation. We use the standard notation \( \mu \text{-a.e. on some metric space } (S, \mu) \), which means that the set of elements for which a property does not hold is a set of \( \mu \)-measure zero. We shall suppress the measure \( \mu \) in our notation if
it is clear from context. We use also the following standard asymptotic notation, in which the limits are taken as \( N \to \infty \). For functions \( f = f(N) \) and \( g = g(N) \), we write \( f = O(g) \) if the limit of \( f/g \) is bounded; \( f = \Theta(g) \) if \( f = O(g) \) and \( g = O(f) \); \( f = o(g) \) if \( f/g \to 0 \); \( f \sim g \) if \( f/g \to 1 \).

Furthermore we denote the sizes of components of a graph \( G \) by \( L_1(G) \geq L_2(G) \geq \ldots \), with \( L_j(G) = 0 \) if \( G \) has fewer than \( j \) components. The following result deals with the threshold for the existence of the giant component (composed of super-vertices) in \( (G, K^r, p) \).

**Proposition 1** Let \( c \in \mathbb{R}^+ \) be a positive constant and let \( r \in \mathbb{N} \cup \{ \infty \} \) be either a constant independent of \( N \) or a function in \( N \) such that \( r = r(N) \) tends to a constant or \( \infty \) as \( N \to \infty \). Assume that \( \lim_{N \to \infty} n/N \) exists and is equal to \( u := \sum_{i \in \mathbb{Z}^+} i \mu_i \). Define

\[
\tilde{s}_2 := \frac{1}{u} \sum_{i \in \mathbb{Z}^+} i^2 \mu_i.
\]

Then the random graph \( (G(N, K^r, p)) \) with \( p = c/n \) satisfies the following properties.

1. If \( c \tilde{s}_2 \leq 1 \), then

\[
\lim_{N \to \infty} \frac{L_1(G(N, K^r, p))}{N} = 0 \quad \text{in probability},
\]

while if \( c \tilde{s}_2 > 1 \), then

\[
\lim_{N \to \infty} P[ L_1(G(N, K^r, p)) = \Theta(N) ] = 1.
\]

2. Let \( \mu = (\mu_i)_{i \geq 1} \) be the probability measure on \( \mathbb{Z}^+ \) defined by (2). Then,

\[
\lim_{N \to \infty} \frac{L_1(G(N, K^r, p))}{N} = \rho \quad \text{in probability},
\]

where \( \rho := \sum_{i \in \mathbb{Z}^+} h(i) \mu_i \) and the function \( h \) is \( \mu \)-a.e. equal to the maximum solution of the non-linear equation \( h = 1 - e^{-Th} \), where \( T \) is an integral operator defined by

\[
Th(i) := \frac{ci}{u} \sum_{j \in \mathbb{Z}^+} j h(j) \mu_j.
\]

Furthermore, \( \rho > 0 \) if and only if \( c \tilde{s}_2 > 1 \).

In the next result we characterize the degree distribution in \( (G(N, K^r, p)) \). We define the degree of a super-vertex \( v \in [N] \) as the number of edges which touch \( v \). Thus, \( deg(v) := \sum_{u \in [N], u \neq v} 1_{\{u,v\}} \), where \( 1_{\{u,v\}} = 1 \) if \( u \) is connected with \( v \), and zero otherwise.

**Proposition 2** Let \( c \in \mathbb{R}^+ \) be a positive constant and let \( r \in \mathbb{N} \cup \{ \infty \} \) be either a constant independent of \( N \) or a function in \( N \) such that \( r = r(N) \) tends to a constant or \( \infty \) as \( N \to \infty \). Assume that \( \lim_{N \to \infty} n/N \) exists and is equal to \( u := \sum_{i \in \mathbb{Z}^+} i \mu_i \). Consider the random graph \( (G(N, K^r, p)) \) with \( p = c/n \). For each integer \( k \geq 0 \), we let \( Z_k \) denote the number of super-vertices of degree \( k \) in \( (G(N, K^r, p)) \). Then we have

\[
\lim_{N \to \infty} \frac{Z_k}{N} = \mathbb{P}(\Xi = k) \quad \text{in probability},
\]

where

\[
\mathbb{P}(\Xi = k) := \sum_{i \in \mathbb{Z}^+} \mu_i \mathbb{P}(Po(ci) = k)
\]
and Po(ci) denotes a random variable with Poisson distribution with mean ci.

The example below shows how and under which conditions the upper tail of the degree distribution of a random super-vertex in $G(N, K^\infty, p)$ exhibits a power-law behavior. In recent years, it has been conjectured that power laws characterize the behavior of the upper tails of the degree distribution in many real-world networks (see e.g. [24] for a review of the empirical evidence of this property).

Example 1 Consider the random graph $G(N, K^\infty, p)$ with $p = 1/n$ and a configuration of super-vertices $K^\infty$ such that for some positive constants $C$ and $\alpha$,

$$\sum_{i=k}^\infty \mu_i \sim \frac{C}{k^\alpha}$$

for $k$ sufficiently large. For each integer $k \geq 0$, let $Z \geq k$ denote the number of super-vertices of degree of at least $k$ in $G(N, K^\infty, p)$. It follows as a corollary of Proposition 2 that

$$\frac{Z \geq k}{N} \to \mathbb{P}(Z \geq k) \sim \frac{C}{k^\alpha},$$

(9)

where the first convergence is in probability when $N \to \infty$ (for $k$ fixed) and the latter approximation holds for $k$ large enough. To see it, note that (8) says that $Z$ follows a mixed Poisson distribution with mixing distribution. More precisely, $Z$ has a Poisson distribution with a random mean $Y$ such that $\mathbb{P}(Y = i) = \mu_i$ for $i \geq 1$. We take $\varepsilon > 0$ arbitrarily small and obtain from (8) that

$$\mathbb{P}(Z \geq k) \sim \mathbb{P}(Z \geq k | Y > (1 - \varepsilon)k) \frac{C}{k^\alpha} + \mathbb{P}(Z \geq k | Y < (1 - \varepsilon)k) \left(1 - \frac{C}{k^\alpha}\right).$$

Then (9) follows from

$$\mathbb{P}(Z \geq k | Y > (1 - \varepsilon)k) = 1 - o(1),$$

$$\mathbb{P}(Z \geq k | Y < (1 - \varepsilon)k) = o(k^{-\alpha}),$$

which may be obtained by Chernoff estimates (e.g. Corollary 13.1 in [6]).

This example shows that our model provides a mechanism starting from the critical Erdős–Rényi random graph $G(n, 1/n)$ and leading to scale-free networks. Therefore it may be seen as an instance of the good-get-richer mechanism proposed by Caldarelli et al. in [8], which has been introduced as an alternative to the well-known scale-free networks obtained by dynamical properties or preferential attachment. In [8], the authors start with a random graph with a large number $N$ of vertices. They associate each vertex with the so-called fitness, which is a random number taken from a given probability distribution and measures the importance of the vertex. The probability of any two vertices being connected is given by a function of their respective fitness values. If we assume that the fitness of a super-vertex in our model is its size, and it is distributed according to $(\mu_i)_{i \geq 1}$ defined by (2), then $G(N, K^\alpha, p)$ provides a constructive example of the good-get-richer mechanism.

2.3 Related Work

2.3.1 Inhomogeneous Random Graphs (IRG)

The random graph $G(N, K^\alpha, p)$ can be seen as a special case of inhomogeneous random graphs studied by Bollobás et al. [6]. To this end, we briefly recall some definitions and
notations from [6]. Consider a graph with vertex set \([N] := \{1, 2, \ldots, N\}\). A vertex space \(V\) is a triple \((S, \mu, (x_N)_{N \geq 1})\) where \(S\) is a separable metric space, \(\mu\) is a Borel probability measure on \(S\) and for each \(N \in \mathbb{N}\), \(x_N\) is a random sequence \((x_1, x_2, \ldots, x_N)\) of points of \(S\) such that
\[
\frac{[\{k : x_k \in A\}]}{N} \rightarrow \mu(A) \quad \text{in probability} \quad (10)
\]
for every \(\mu\)-continuous set \(A \subset S\), where \(|C|\) denotes the cardinality of the set \(C\). A kernel \(\kappa_N\) on vertex space \(V\) is a symmetric non-negative Borel measurable function on \(S \times S\).

Let \(G^{V}(N, \kappa_N)\) be the inhomogeneous random graph with vertex set \([N]\), in which two vertices \(k\) and \(l\) are connected by an edge with probability
\[
p_{kl} := \min\left\{1, \frac{\kappa_N(x_k, x_l)}{N}\right\}.
\]
This model is an extension of the one defined by Söderberg [27] and its various properties are studied by [6] under specific restrictions. The main results regarding the existence and uniqueness of the giant component are proved by using an appropriate multi-type branching process and an integral operator to which the component structure is related.

We shall show that our random graph model \(G(N, \kappa', p)\) is a particular case of \(G^{V}(N, \kappa_N)\). To this end, we consider the set \(S = \mathbb{Z}^+\), the probability measure on \(S\) defined by \(\mu([i]) = \mu_i\) given by (2), and the sequence \(x_N = (x_1, x_2, \ldots, x_N)\) of points of \(S\) such that \(x_k\) represents the size of the \(k\)-th super-vertex, for \(k = 1, 2, \ldots, N\). Then the triple \(V := (S, \mu, (x_N)_{N \geq 1})\) is a vertex space. Observe that for all \(i = 1, 2, \ldots, r\),
\[
k_i = \sum_{k=1}^{N} I_{\{x_k = i\}},
\]
where \(I_A\) denotes the indicator random variable of the event \(A\). Then by our construction the \(x_k\)'s are deterministic and therefore (2) implies (10). Now we define the kernel \(\kappa_N\) on the vertex space \(V\) by
\[
\kappa_N(x_k, x_l) := N \left(1 - (1 - p)^{x_kx_l}\right) \quad (11)
\]
and let the connection probabilities between two super-vertices of sizes \(k\) and \(l\) in \(G(N, \kappa', p)\) be defined as
\[
p_{kl} := \frac{\kappa_N(x_k, x_l)}{N}.
\]
Then our model \(G(N, \kappa, p)\) corresponds to \(G^{V}(N, \kappa_N)\).

As far as we know, the threshold for connectedness on an IRG-like model has been studied only recently by Devroye and Fraiman [13], for a continuous version of the model defined by Söderberg [27]. The Devroye–Fraiman model is defined following the notation from [6] with some minor changes. Thus, it considers a random graph model on a set of \(N\) vertices in which each pair of vertices, say \(k\) and \(l\), are connected independently with probability
\[
p_{kl} := \min\{1, \kappa(x_k, x_l) p_N\},
\]
where \(\kappa\) is a kernel on the respective vertex space \(V\), \(x_k, x_l\) are \(\mu\)-distributed independent random variables on \(S\), which are associated with the vertices \(k\) and \(l\), respectively, and \(p_N := (\log N)/N\). For a discussion on a possible connection between \(G(N, \kappa', p)\) and the Devroye–Fraiman model see Sect. 5.2.
2.3.2 Norros–Reittu Model (NRN(w))

Another model related to ours is the one introduced by Norros and Reittu [25], which is a particular case of IRG model studied by Bollobás et al. [6]. Furthermore, a variant of this model is also studied by Aldous and Limic [1] in the construction of the standard multiplicative coalescent. The Norros–Reittu model is defined on the vertex set \([N]\) as follows.

Each vertex \(v \in [N]\) has an associated weight \(w_v \geq 0\). Write \(w = (w_v)_{v \in [N]}\) for the vector of weights and let \(\ell_N := \sum_{v \in [N]} w_v\) be the sum of these weights. Define the probabilities \(p_{u,v} := 1 - \exp(-w_u w_v/\ell_N), \ u, v \in [N]\), and construct the random graph \(G_{NRN}(w)\) by putting an edge \({u, v}\) between the vertices \(u, v \in [N]\) with probability \(p_{u,v}\), independent across edges.

This model has been extended to settings where weights are deterministic and the weights of a uniform vertex converges in distribution, see e.g. [28], while Janson [21] investigated when this random graph is asymptotically equivalent to two others famous models of inhomogeneous random graphs, Chung–Lu model [10] and Britton–Deijfen–Löf model [7].

Relating Norros–Reittu model to the model of super-vertices \(G(N, K^r, p)\), let \(u, v \in [N]\) be labels of two super-vertices of sizes \(i\) and \(j\), respectively. In this case we know that there is an edge between them with probability \(p_{ij} = 1 - (1 - p)^{ij}\). We can rewrite \(p_{ij}\) as

\[
p_{ij} = 1 - e^{-w_u w_v/\ell_N},
\]

where

\[
w_v = -\text{size}(v) \ln(1 - p) \sum_{u \in [N]} \text{size}(u) = -\text{size}(v)n \ln(1 - p),
\]

and

\[
\ell_N := \sum_{u \in [N]} w_u = -n^2 \ln(1 - p).
\]

Thus, \(G(N, K^r, p)\) coincides with the model proposed by Norros and Reittu [25].

3 Proof of Theorem 1

Let \(r \in \mathbb{N}\) be either a constant independent of \(N\) or a function in \(N\) such that \(r = r(N)\) tends to a constant as \(N \to \infty\). Throughout the section we will assume that \(N\) is sufficiently large so that we can consider \(1 \leq r \leq \bar{r}\) for some constant \(\bar{r} \in \mathbb{N}\).

First, we recollect some important conditions in Theorem 1. We have

\[
p := \frac{\ln k_{i_*} + c(N)}{i_*(n - i_*)},
\]

where \(c(N)\) satisfies (4) which ensures \(p \in (0, 1)\), and \(i_*\) is defined by (3), i.e. \(i_* = \arg \max_i k_i^{1/(i(n-i))}\).

Because the properties of being connected or of having no isolated super-vertices are monotone decreasing properties, we may assume that \(c(N) = o(\ln N)\).

The connectedness of \(G(N, K^r, p)\) is closely related to the distribution of the number of isolated super-vertices and “small” components, which will be the topics of the next three sections and will be used in the proof of Theorem 1 (1)–(3).
Let $X$ denote the number of isolated super-vertices in $G(N, K^r, p)$. Observe that the random variable $X$ depends on $N$. For simplicity we suppress this dependence in our notation. First we shall derive asymptotic expressions of the first moment and second moment of $X$, which will be used in the proofs of Lemma 1 (when $\lim_{N \to \infty} c(N) = \pm \infty$) and Lemma 3 (when $\lim_{N \to \infty} c(N) \to c$ a constant).

Take any arbitrary order of the super-vertices and let us write $X$ as a sum of indicator random variables

$$X = \sum_{i=1}^{r} \sum_{k=1}^{k_i} I_{ik},$$

where $I_{ik} = 1$ if the $k$-th super-vertex of size $i$ is isolated and 0 otherwise, for $k = 1, 2, \ldots , k_i$ and $i = 1, 2, \ldots , r$.

Note that $I_{ik} = 1$ if the $k$-th super-vertex of size $i$ is not connected to any other super-vertex of size $i$ and connected with any other super-vertex of size $j \neq i$ neither. Since any pair of super-vertices is connected independently of each other, we have

$$\mathbb{E}[I_{ik}] = (1 - p_{ii})^{k_i - 1} \prod_{j \neq i} (1 - p_{ij})^{k_j} = (1 - p)^{k_i - 1} \prod_{j \neq i} (1 - p)^{i j k_j} = (1 - p)^{i(n-i)},$$

and hence

$$\mathbb{E}[X] = \sum_{i=1}^{r} k_i (1 - p)^{i(n-i)}. \quad (15)$$

In order to estimate (16), observe that $N \leq n \leq \tilde{r} N$, $k_{i*} < N$, $i* \leq \tilde{r}$ and recall $c(N) = o(\ln N)$. Thus, $p = o(1)$ and $\frac{(\ln k_{i*} + c(N))^2}{n} = o(1)$. Furthermore, using Taylor series for $\ln(1 - x)$ and $e^x$ we get

$$(1 - p)^{i(n-i)} = \exp\left(-i(n-i)(p + O(p^2))\right) = \left(1 + O\left(\frac{(\ln k_{i*} + c(N))^2}{n}\right)\right) \exp\left(-\frac{i(n-i)}{i* (n-i*)} (\ln k_{i*} + c(N))\right) = (1 + o(1)) \frac{i(n-i)}{i* (n-i*)} \exp\left(-\frac{i(n-i)}{i* (n-i*)}\right). \quad (16)$$

To ease notation, we let

$$f(i) := f(i, c(N)) = \frac{i(n-i)}{i* (n-i*)} \exp\left(c(N) \frac{i(n-i)}{i* (n-i*)}\right), \quad (17)$$

and we obtain from (16) that

$$(1 - p)^{i(n-i)} = (1 + o(1)) f(i)^{-1} \quad \text{for any} \quad 1 \leq i \leq r. \quad (18)$$

By (15)–(18), the first moment of the number $X$ of isolated super-vertices in $G(N, K^r, p)$ satisfies

$$\mathbb{E}[X] = (1 + o(1)) \sum_{i=1}^{r} \frac{k_i}{f(i)}. \quad (19)$$
As for the second moment of $X$ we observe that

$$\mathbb{E}[X^2] = \mathbb{E} \left[ \left( \sum_{i=1}^{r} \sum_{k=1}^{k_i} I_k^i \right)^2 \right]$$

$$= \sum_{i=1}^{r} k_i \mathbb{E}(I_i^i) + \sum_{i=1}^{r} k_i (k_i - 1) \mathbb{E}(I_i^i I_j^j) + \sum_{i,j,i \neq j} k_i k_j \mathbb{E}(I_i^i I_j^j)$$

$$= \mathbb{E}[X] - \sum_{i=1}^{r} k_i \mathbb{E}[I_i^i I_j^j] + \sum_{i,j=1}^{r} k_i k_j \mathbb{E}[I_i^i I_j^j]. \quad (20)$$

Furthermore, we obtain by (18) that for any $1 \leq i, j \leq r$

$$\mathbb{E}[I_i^i I_j^j] = \mathbb{P}(I_i^i = 1 \mid I_j^j = 1) \mathbb{P}(I_j^j = 1)$$

$$= \left( 1 - p_{ii} \right)^{k_i-1} \left( 1 - p_{jj} \right)^{k_j-1} \prod_{l=1,l \neq i,j}^{r} (1 - p_{ll})^{k_l} \mathbb{P}(I_j^j = 1)$$

$$= \left( 1 - p \right)^{i(n-i)+j(n-j)-ij} \quad (21)$$

Recall that here we consider $1 \leq r \leq \tilde{r}$ for some constant $\tilde{r} \in \mathbb{N}$. Therefore, $(1 - p)^{ij} = 1 + o(1)$ for any $1 \leq i, j \leq r$. Thus (20)–(21) and (18)–(19) imply

$$\mathbb{E}[X^2] = \mathbb{E}[X] - (1 + o(1)) \sum_{i=1}^{r} \frac{k_i}{f(i)^2} + (1 + o(1)) \sum_{i,j=1}^{r} \frac{k_i k_j}{f(i) f(j)}$$

$$= \mathbb{E}[X] - (1 + o(1)) \sum_{i=1}^{r} \frac{k_i}{f(i)^2} + (1 + o(1)) \left( \mathbb{E}[X] \right)^2. \quad (22)$$

Using (19) and (22) we shall derive the existence of isolated super-vertices (in Lemma 1) and the exact distribution of the number of isolated super-vertices (in Lemma 3).

### 3.1 Existence of Isolated Super-vertices

The following result deals with the existence of isolated super-vertices when $\tilde{c} = \pm \infty$.

**Lemma 1** Let $r \in \mathbb{N}$ be either a constant independent of $N$ or a function in $N$ such that $r = r(N)$ tends to a constant as $N \to \infty$. Let $X$ denote the number of isolated super-vertices in $G(N, K^r, p)$ and let $\tilde{c} := \lim_{N \to \infty} c(N)$ as in Theorem 1.

1. If $\tilde{c} = +\infty$, then

$$\lim_{N \to \infty} \mathbb{P}[X \geq 1] = 0.$$

2. If $\tilde{c} = -\infty$, then

$$\lim_{N \to \infty} \mathbb{P}[X \geq 1] = 1.$$

**Proof** To prove (1) we will assume that $c(N) > 0$. Note that

$$\frac{i(n-i)}{i_n(n-i_n)} \geq \frac{(n-i)}{n(n-i_n)} = \frac{1+o(1)}{i_n}$$

and thus...
\[ f(i)^{-1} = k_{i}^{\frac{1}{\sum_{i=1}^{N} i^{-1}}} \exp \left( -c(N) \frac{i(n-i)}{\sum_{i=1}^{N} i^{-1}} \right) \]

\[ \leq k_{i}^{\frac{1}{\sum_{i=1}^{N} i^{-1}}} \exp \left( - (1 + o(1)) \frac{1}{i} c(N) \right). \] (23)

Furthermore, by definition of \( i_{n} \) we have that

\[ k_{i} \leq k_{i}^{\frac{1}{\sum_{i=1}^{N} i^{-1}}} , \] (24)

which together with (23) implies

\[ \frac{k_{i}}{f(i)} \leq \exp \left( - (1 + o(1)) \frac{1}{i} c(N) \right) , \]

for each \( 1 \leq i \leq r \). Then, we obtain by (19)

\[ \mathbb{E}[X] = (1 + o(1)) \sum_{i=1}^{r} \frac{k_{i}}{f(i)} \]

\[ \leq (1 + o(1)) r \exp \left( - (1 + o(1)) \frac{1}{i} c(N) \right) \]

\[ \leq (1 + o(1)) \bar{r} \exp \left( - (1 + o(1)) \frac{1}{\bar{r}} c(N) \right). \]

The last inequality follows, because \( r \leq \bar{r} \) and \( i_{n} \leq \bar{r} \). Thus by Markov’s inequality, we have

\[ \mathbb{P}(X \geq 1) \leq \mathbb{E}[X] \to 0 \text{ as } N \to \infty, \]

when \( \lim_{N \to \infty} c(N) = +\infty \).

To prove (2), we observe that

\[ \frac{k_{i}}{f(i)} \geq (1 + o(1)) \frac{k_{i}}{f(i)} \]

and since \( \sum_{i=1}^{r} \frac{k_{i}}{f(i)} \geq (1 + o(1)) \frac{k_{i}}{f(i)} \)

\[ \mathbb{E}[X] = (1 + o(1)) \sum_{i=1}^{r} \frac{k_{i}}{f(i)} \geq (1 + o(1)) \exp \left( - (1 + o(1)) \frac{1}{\bar{r}} c(N) \right) \to +\infty , \] (25)

as \( N \to \infty \). We shall prove that \( \mathbb{P}(X = 0) \to 0 \) as \( N \to \infty \), by applying Chebyshev’s inequality

\[ \mathbb{P}(X = 0) \leq \frac{\mathbb{E}[X^2]}{(\mathbb{E}[X])^2} - 1. \]

It suffices to show that \( \lim_{N \to \infty} \frac{\mathbb{E}[X^2]}{(\mathbb{E}[X])^2} = 1 \) as \( N \to \infty \). To this end, we observe that (22) implies

\[ \frac{\mathbb{E}[X^2]}{(\mathbb{E}[X])^2} = \mathbb{E}[X]^{-1} - (1 + o(1)) \sum_{i=1}^{r} \frac{k_{i}}{f(i)} (\mathbb{E}[X])^{-2} + 1 + o(1). \] (26)

Using that \( f(i) > 1 \) for each \( 1 \leq i \leq r \) (because \( c(N) > -\ln k_{i} \)), we have

\[ \sum_{i=1}^{r} \frac{k_{i}}{f(i)^2} \leq \sum_{i=1}^{r} \frac{k_{i}}{f(i)} = (1 + o(1)) \mathbb{E}[X]. \]
Since $\mathbb{E}[X] \to +\infty$ as $N \to \infty$ by (25), we obtain

$$
\left( \sum_{i=1}^{r} \frac{k_i}{f(i)^2} \right) (\mathbb{E}[X])^{-2} < (1 + o(1)) (\mathbb{E}[X])^{-1} \to 0 \quad \text{as} \quad N \to \infty. \quad (27)
$$

Putting (26) and (27) together, we have $\frac{\mathbb{E}[X^2]}{(\mathbb{E}[X])^2} \to 1$ as $N \to \infty$ as desired. \hfill \Box

### 3.2 Distribution of the Number of Isolated Super-vertices

In this section we deal with the exact asymptotic distribution of the number of isolated super-vertices when $\bar{c} = c \in \mathbb{R}$ is a constant, by applying Stein’s method, in particular Theorem 6.24 [23]. To this end we need a few more definitions and notations. The random variables $\{I_k\}_k$ are said to be positively related if they satisfy the following two conditions:

1. For each $k$ there exists a family of random variables $\{J_k^\ell\}_{\ell \neq k}$, such that the joint distribution of $\{J_k^\ell\}_{\ell \neq k}$ is the same as the conditional distribution of $\{I_\ell\}_\ell$ given $I_k = 1$; and
2. $J_k^\ell \geq I_\ell$ for every $\ell \neq k$.

Formally, $\mathcal{L}(\{J_k^\ell\}_\ell) = \mathcal{L}(\{I_\ell\}_\ell \mid I_k = 1)$, and $J_k^\ell \geq I_\ell$ for every $\ell \neq k$, where $\mathcal{L}(\{Y_\ell\}_\ell)$ denotes the joint distribution of the random variables $\{Y_\ell\}_\ell$.

**Lemma 2** (Theorem 6.24 in [23]) Given random variables $\{I_k\}_k$, let $X := \sum_k I_k$. If $\{I_k\}_k$ are positively related, then

$$
d_{TV}(X, \text{Poi}(\lambda)) \leq \frac{\sqrt{V(X)}}{\mathbb{E}[X]} + 1 + 2 \max_{1 \leq k \leq N} \left\{ \mathbb{E}(I_k) \right\}, \quad (28)
$$

where $\lambda := \mathbb{E}[X]$ and $d_{TV}(\cdot, \cdot)$ is the total variation distance.

**Lemma 3** Let $r \in \mathbb{N}$ be either a constant independent of $N$ or a function in $N$ such that $r = r(N)$ tends to a constant as $N \to \infty$. Let $X$ denote the number of isolated super-vertices in $G(N, K^r, p)$. If $\bar{c} = c \in \mathbb{R}$ is a constant, then $X$ has asymptotically Poisson distribution with mean $e^{-c} + \gamma$, where

$$
\gamma := \lim_{N \to \infty} \sum_{i \neq i_o} k_i (k_i e^c)^{-i(n-i)} \frac{i(n-i)}{i_o (n-i_o)}. \quad (29)
$$

**Proof** Without loss of generality we may assume that $c(N) = c \in \mathbb{R}$ is a constant.

By the definition of $i_o$, we have that $k_i \leq k_i \frac{i(n-i)}{i_o (n-i_o)}$ for each $1 \leq i \leq r$ and $1 \leq i_o \leq r$, $r \leq \bar{r}$. Note further that $(1 + o(1)) \frac{1}{r_o} \leq \frac{i(n-i)}{i_o (n-i_o)} \leq (1 + o(1)) \frac{\bar{r}}{r_o}$ for each $1 \leq i \leq r$. Thus, letting $M(0) := 0$, $M(c) := -((1 + o(1)) c \bar{r})$ if $c > 0$ and $M(c) := -(1 + o(1)) c \bar{r})$ if $c < 0$, we have that for each $1 \leq i \leq r$

$$
c \frac{i(n-i)}{i_o (n-i_o)} \geq -M(c)
$$

and hence

$$
\frac{k_i}{f(i)} = k_i k_i \frac{i(n-i)}{i_o (n-i_o)} \exp \left( -c \frac{i(n-i)}{i_o (n-i_o)} \right) \leq \exp(M(c)).
$$
Therefore,
\[
\gamma := \lim_{N \to \infty} \sum_{i \neq i_0} k_i (k_i e^c) \frac{j(i-n-i)}{i_0 (n-i_0)} = \lim_{N \to \infty} \sum_{i \neq i_0} \frac{k_i}{f(i)} \leq \bar{\gamma} \exp(M(c)) < \infty. \tag{29}
\]
Because \( \frac{k_i}{f(i)} = e^{-c} \), (19) and (29) imply that
\[
\mathbb{E}[X] = (1 + o(1)) \left( \frac{k_i}{f(i)} + \sum_{i \neq i_0} \frac{k_i}{f(i)} \right) \to e^{-c} + \gamma \text{ as } N \to \infty. \tag{30}
\]

In order to apply (28), we take an arbitrary order of the super-vertices and rewrite the number of isolated super-vertices in \( G(N, K^r, p) \) as
\[
X = \sum_{k=1}^{N} I_k, \tag{31}
\]
where, for each \( k = 1, 2, \ldots, N \), \( I_k = 1 \) if the \( k \)-th super-vertex is isolated and \( I_k = 0 \) otherwise. Let \( G_k(N, K^r, p) \) be the random graph \( G(N, K^r, p) \) with all edges from the \( k \)-th super-vertex removed, and let \( J_k^\ell = 1 \) if the \( \ell \)-th super-vertex is isolated in \( G_k(N, K^r, p) \), and \( J_k^\ell = 0 \) otherwise. Observe that
\[
J_k^\ell = (I_\ell | I_k = 1),
\]
and thus
\[
\mathcal{L}([J_k^\ell]) = \mathcal{L}([I_\ell] | I_k = 1).
\]

Moreover, if the \( \ell \)-th super-vertex in \( G(N, K^r, p) \) is isolated (i.e. \( I_\ell = 1 \)), then \( J_k^\ell = 1 \) as the \( \ell \)-th super-vertex is still isolated even if all the edges from the \( k \)-th super-vertex are removed. Otherwise, i.e. \( I_\ell = 0 \), then the \( \ell \)-th super-vertex is either connected with the \( k \)-th super-vertex or with any other super-vertex. So, if all edges from the \( k \)-th super-vertex are removed, the resulting \( \ell \)-th super-vertex could become either isolated or not, this means that \( J_k^\ell \in \{0, 1\} \). Thus \( J_k^\ell \succeq I_\ell \) for every \( \ell \neq k \), and thus the random variables \( \{I_k\} \) are positively related. So we can apply Lemma 2.

Next we shall show that \( d_{TV}(X, Po(\lambda)) \to 0 \) as \( N \to \infty \), by proving \( \frac{\mathbb{V}(X)}{\mathbb{E}[X]} \to 1 \) and \( \max_{1 \leq k \leq N} \{\mathbb{E}(I_k)\} \to 0 \) as \( N \to \infty \). By (19) and (22) as well as (30), we get
\[
\frac{\mathbb{V}(X)}{\mathbb{E}[X]} = \frac{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}{\mathbb{E}[X]}
\]
\[
= 1 - (1 + o(1)) \sum_{i=1}^{r} \frac{k_i}{f(i)^2} (\mathbb{E}[X])^{-1} + o(1)\mathbb{E}[X]
\]
\[
= 1 + o(1) - (1 + o(1)) \sum_{i=1}^{r} \frac{k_i}{f(i)} (\mathbb{E}[X])^{-1}. \tag{32}
\]

Let \( f_* := \min\{f(i) : i \in \{1, 2, \ldots, r\} \} \). By (19) we have
\[
\sum_{i=1}^{r} \frac{k_i}{f(i)^2} \leq \sum_{i=1}^{r} \frac{k_i}{f(i)} f_* \leq (1 + o(1)) \frac{\mathbb{E}[X]}{f_*}. \tag{33}
\]
By (32), in order to prove
\[
\frac{\mathbb{V}(X)}{\mathbb{E}[X]} \to 1 \quad \text{as} \quad N \to \infty,
\] (34)

it suffices to show that
\[
f_\ast \to \infty \quad \text{as} \quad N \to \infty,
\] (35)
because this together with (33) implies that
\[
\sum_{i=1}^{r} \frac{k_i}{f(i)(\mathbb{E}[X])^{-1}} \leq (1 + o(1)) \frac{1}{f_\ast} \to 0 \quad \text{as} \quad N \to \infty.
\]

In order to prove (35), we observe from the definition of \( M(c) \) that for each \( 1 \leq i \leq r \)
\[
c_i \frac{i(n-i)}{i_\ast(n-i_\ast)} \geq -M(c),
\]
and therefore we get
\[
f(i) = k_i \frac{i(n-i)}{i_\ast(n-i_\ast)} \exp \left( c \frac{i(n-i)}{i_\ast(n-i_\ast)} \right)
\geq k_i \frac{i(n-i)}{i_\ast(n-i_\ast)} \exp (-M(c))
\geq k_{i_\ast} \frac{(1+o(1))}{i_\ast} \exp (-M(c)).
\] (36)

Now we define
\[
\bar{\ell} := \min \{ \ell \in \{1, 2, \ldots, r\} : \mu_\ell > 0 \}.
\] (37)

As \( i_\ast = \arg \max_j k_j^{1/(i(n-i))} \), we have \( i_\ast \leq \bar{\ell} \) and thus
\[
k_{i_\ast} \geq k_{i_\ast} \frac{i_\ast(n-i_\ast)}{(n-i_\ast)} = k_{\bar{\ell}} \frac{(1+o(1))}{\bar{\ell}} \exp (-M(c)).
\]

Putting this in (36) we have that for each \( 1 \leq i \leq r \)
\[
f(i) \geq k_{i_\ast} \frac{(1+o(1))}{i_\ast} \exp (-M(c)) \geq k_{\bar{\ell}} \frac{(1+o(1))}{\bar{\ell}} \exp (-M(c)).
\]

Because \( k_{\bar{\ell}} = (1 + o(1))\mu_{\bar{\ell}}N \) with \( \mu_{\bar{\ell}} > 0 \), we get
\[
f_\ast := \min_{1 \leq i \leq r} f(i)
\geq \left( (1 + o(1))\mu_{\bar{\ell}}N \right)^{\frac{1}{1+o(1)}} \exp (-M(c)) \to \infty \quad \text{as} \quad N \to \infty.
\]

Finally we shall show that \( \max_{1 \leq k \leq N} \{ \mathbb{E}(I_k) \} = o(1) \). To this end, we consider the sequence of indicator random variables \( (I_k)_k \) used in (31). For each \( k = 1, 2, \ldots, N \) we obtain
\[
\mathbb{E}(I_k) = \sum_{i=1}^{r} (1 - p)^i(n-i) \frac{k_i}{N} = \frac{\mathbb{E}[X]}{N} \to 0 \quad \text{as} \quad N \to \infty,
\]
where the last equality and the limit behavior follow from (15) and (30) respectively. Therefore we have
\[
\max_{1 \leq k \leq N} \{ \mathbb{E}(I_k) \} = o(1).
\] (38)
Thus, we can conclude from (28), (34) and (38) that $X$ has asymptotic Poisson distribution with mean $e^{-c} + \gamma$. □

### 3.3 Small Components

Before we proceed to the proof of Theorem 1, we consider components of size $m$ for each $2 \leq m \leq N/2$.

**Lemma 4** Let $r \in \mathbb{N}$ be either a constant independent of $N$ or a function in $N$ such that $r = r(N)$ tends to a constant as $N \to \infty$. If $\bar{c} = +\infty$ or $\bar{c} = c \in \mathbb{R}$ is a constant, then whp $G(N, K^r, p)$ does not have any component of size $m$ for any $2 \leq m \leq N/2$.

**Proof** For each $m$ such that $2 \leq m \leq N/2$, we let $S^m$ be the set of all subsets of $m$ super-vertices, and for each $S \in S^m$, let $m_s(S)$ be the number of super-vertices of size $i$ in $S$. Note that $m = m_1(S) + \cdots + m_r(S)$. Observe that if the super-vertices in $S$ form a component in $G(N, K^r, p)$, then the following two events should hold:

(i) $A^S_i := \{\text{the super-vertices in } S \text{ are connected}\}$ and

(ii) $A^S_2 := \{\text{no super-vertex in } S \text{ is connected with a super-vertex in } S^c\}$,

where $S^c$ denotes the complement of $S$. Note that the events $A^S_i$ and $A^S_2$ are independent, because every pair of vertices are connected independently of each other with probability $p$.

Thus, we have

$$\mathbb{P}(\exists \text{ a component of size } m) \leq \sum_{S \in S^m} \mathbb{P}(A^S_1)\mathbb{P}(A^S_2). \tag{39}$$

Since a component of size $m$ contains a tree of size $m$ and the number of Cayley trees on $m$ vertices is $m^{m-2}$, moreover $p_{ij} = 1 - (1 - p)^{ij} \leq ij p \leq r^2 p$, we have

$$\mathbb{P}(A^S_1) \leq \mathbb{P}(S \text{ contains a tree}) \leq m^{m-2}(r^2 p)^{m-1}. \tag{40}$$

On the other hand, for each $S \in S^m$ letting $M_m(S) := \sum_{i=1}^{r} im_i(S)$, we have

$$\mathbb{P}(A^S_2) = (1 - p)^{M_m(S)(n - M_m(S))}.$$

Observe that $M_m(S) \geq m$. Since the function $f(x) = x(n - x)$ is increasing in $x \in (2, N/2)$, we have

$$\mathbb{P}(A^S_2) \leq (1 - p)^{m(n-m)}. \tag{41}$$

By (39)–(41), we obtain

$$\mathbb{P}(\exists \text{ a component of size } m) \leq {N \choose m} m^{m-2}(r^2 p)^{m-1}(1 - p)^{m(n-m)} \leq (eN)^m m^{-5/2}(r^2 p)^{m-1} e^{-sp(m-1)(n-m)}, \tag{42}$$

where the last inequality is because $\binom{N}{m} < (eN)^m / m^{m+1/2}$ and $(1 - p) < e^{-p}$. Summing up, we have

$$\sum_{m=2}^{N/2} \mathbb{P}(\exists \text{ a component of size } m) \leq \sum_{m=2}^{N/2} (eN)^m m^{-5/2}(r^2 p)^{m-1} e^{-p(m-1)(n-m)}. \tag{43}$$

We shall show that the right hand side of (43) tends to 0 as $N \to \infty$, making a case distinction depending on whether $2\bar{c} + 2 \leq m \leq N/2$ or $1 \leq m \leq 2\bar{c} + 1$, where $\bar{c} := \min\{\ell \in \{1, 2, \ldots, r\} : \mu_\ell > 0\}$ as defined in (37).
**Case 1** Let $2\bar{\ell} + 2 \leq m \leq N/2$. We first rewrite (42) as

$$eN m^{-5/2} (er^2 N)^{m-1} \exp \left\{ (m-1) \left[ \ln p - p (n-m) \right] \right\}.$$  \hspace{1cm} (44)

Noting that $p = \frac{\ln k_{i*} + c(N)}{i_*(n-i_*)}$, we have

$$\ln p - p (n-m) = \ln \left[ \ln k_{i*} + c(N) \right] - \ln \left[ i_*(n-i_*) \right] - \left( \frac{\ln k_{i*} + c(N)}{i_*} \right) \frac{(n-m)}{(n-i_*)},$$

so considering that $m \leq N/2 < n/2$ (and thus $\frac{n-m}{n-i_*} \geq \frac{1}{2}$), we obtain

$$\ln p - p (n-m) \leq \ln \left[ \ln k_{i*} + c(N) \right] - \frac{1}{2i_*} \left( \ln k_{i*} + c(N) \right) - \ln \left[ i_*(n-i_*) \right].$$

Observe now that for any $K > 0$ there exists $\tilde{x} > 0$ such that $\ln x < x/K$ for any $x > \tilde{x}$. Using this with $x = \ln k_{i*} + c(N)$, we have $\ln \left[ \ln k_{i*} + c(N) \right] < \frac{1}{K} (\ln k_{i*} + c(N))$ and so we have

$$\ln p - p (n-m) \leq \left( \frac{1}{K} - \frac{1}{2i_*} \right) (\ln k_{i*} + c(N)) - \ln \left[ i_*(n-i_*) \right].$$  \hspace{1cm} (45)

provided $N$ is sufficiently large. Using (45), we can bound (44) from above by

$$eN m^{-5/2} (er^2 N)^{m-1} \exp \left\{ (m-1) \left[ \left( \frac{1}{K} - \frac{1}{2i_*} \right) (\ln k_{i*} + c(N)) - \ln \left[ i_*(n-i_*) \right] \right] \right\}$$

$$= eN m^{-5/2} \left( \frac{er^2 N}{i_*(n-i_*)} \right)^{m-1} \exp \left\{ (m-1) \left[ \left( \frac{1}{K} - \frac{1}{2i_*} \right) (\ln k_{i*} + c(N)) \right] \right\}$$

$$\leq eN m^{-5/2} (er^2)^{m-1} \exp \left\{ (m-1) \left[ \left( \frac{1}{K} - \frac{1}{2i_*} \right) (\ln k_{i*} + c(N)) + \ln(e r^2) \right] \right\}.$$  

The last inequality follows because $N/i_*(n-i_*) \leq 1$ for $N$ sufficiently large. Therefore, we have

$$\mathbb{P}(\exists \text{ a component of size } m)$$

$$\leq eN m^{-5/2} \exp \left\{ (m-1) \left[ \left( \frac{1}{K} - \frac{1}{2i_*} \right) (\ln k_{i*} + c(N)) + \ln(e r^2) \right] \right\}.  \hspace{1cm} (46)$$

Now consider a constant $\alpha > 1$. Taking $K$ large enough such that $K > 2i_*(2\bar{\ell} + 1)a$ we have

$$(m-1) \left( \frac{1}{K} - \frac{1}{2i_*} \right) \leq (2\bar{\ell} + 1) \left( \frac{1}{K} - \frac{1}{2i_*} \right) \quad \text{(because } m \geq 2\bar{\ell} + 2)$$

$$\leq \left( \frac{1 - (2\bar{\ell} + 1)\alpha}{2i_*a} \right) \quad \text{(because } \frac{1}{K} < \frac{1}{2i_*(2\bar{\ell} + 1)a}).$$
Thus we have
\[
(m - 1) \left( \frac{1}{K} - \frac{1}{2i_a} \right) \ln k_{i_\bar{\ell}} \leq \left( \frac{1 - (2\bar{\ell} + 1)a}{2i_a} \right) \ln k_{i_\bar{\ell}} = \left( \frac{(2\bar{\ell} + 1)a - 1}{2i_a} \right) \ln k_{i_\bar{\ell}}^{-1} \leq -(1 + o(1)) \left( 1 + \frac{a - 1}{2a\bar{\ell}} \right) \ln k_{i_\bar{\ell}} \quad \text{(because } k_{i_\bar{\ell}} \geq k_{i_\bar{\ell}}^{i_a(n-i_a)/\ell(n-I)}) \right) .
\]
(47)

Furthermore, because \( m \geq 2\bar{\ell} + 2 \), we have
\[
(m - 1) \left[ \left( \frac{1}{K} - \frac{1}{2i_a} \right) c(N) + \ln (er^2) \right] \leq -(2\bar{\ell} + 1) \left[ \left( \frac{1}{2i_a} - \frac{1}{K} \right) c(N) - \ln (er^2) \right],
\]
and because \( c(N) \to +\infty \) or \( c \in \mathbb{R} \), we have
\[
\exp \left\{ 1 - (2\bar{\ell} + 1) \left[ \left( \frac{1}{2i_a} - \frac{1}{K} \right) c(N) - \ln (er^2) \right] \right\} \leq C_1,
\]
for \( K \) and \( N \) large enough. Using (46)–(48), we have that for \( 2\bar{\ell} + 2 \leq m \leq N/2 \)
\[
\mathbb{P}(\exists \text{ a component of size } m) \leq C_1 \left( Nk_{i_{\bar{\ell}}}^{-(1+o(1))(1+\frac{a-1}{2a\bar{\ell}})} \right) m^{-5/2}
\]
(49)

where \( \varphi_1(N) \to 0 \) as \( N \to \infty \), because \( k_{\bar{\ell}} = (1 + o(1))\mu_{\bar{\ell}}N, \mu_{\bar{\ell}} > 0 \) and \( a > 1 \).

Summing up, we have
\[
\sum_{m=2\bar{\ell}+2}^{N/2} \mathbb{P}(\exists \text{ a component of size } m) \leq C_1 \varphi_1(N) \sum_{m=2\bar{\ell}+2}^{N/2} m^{-5/2},
\]
(50)

which tends to 0 as \( N \to \infty \).

**Case 2** Let \( 2 \leq m \leq 2\bar{\ell} + 1 \). In this case, it suffices to consider components containing only super-vertices of size at least \( \tilde{\ell} := \min\{\ell \in \{1, 2, \ldots, r\} : \mu_\ell > 0\} \).

To see this let \( E \) be the event that there exists a component of size \( m \) in which all its super-vertices have size at least \( \tilde{\ell} \), and let \( Y \) be the number of super-vertices of size strictly less than \( \tilde{\ell} \) in a random sample of \( m \) super-vertices. Hence, \( Y \) is a random variable that follows the hypergeometric distribution with parameters \((N, v_{\tilde{\ell}}, m)\), where \( v_{\tilde{\ell}} \) denotes the number of super-vertices of size strictly less than \( \tilde{\ell} \), i.e., \( v_{\tilde{\ell}} := \sum_{i=1}^{\tilde{\ell}-1} k_i \). Observe that, as \( v_{\tilde{\ell}} = o(N) \) and \( m \leq 2\bar{\ell} + 1 \),
\[
\mathbb{P}(Y = 0) = \binom{N-v_{\tilde{\ell}}}{m-N} = \binom{N-v_{\tilde{\ell}}}{N} \cdots \binom{N-v_{\tilde{\ell}}-m+1}{N-m+1} = 1 - o(1),
\]
and therefore
\[
\mathbb{P}(E^c) \leq \mathbb{P}(Y > 0) = o(1).
\]

Springer
More precisely, we obtain
\[ \mathbb{P}(\exists \text{ a component of size } m) = \mathbb{P}(E) + o(1). \] (51)

Thus, applying arguments analogous to (39)-(42) to \( \mathbb{P}(E) \), and noting that in such case the component \( S \) has only super-vertices of size at least \( \ell \), we obtain that \( M_{m}(S) \geq m\ell \) and therefore, instead of (42), we obtain
\[ \mathbb{P}(E) \leq (en)^m m^{-5/2} (r^2 p)^{m-1} e^{-p(m\ell - 1)(n-m\ell)} = O(N) (enp)^{m-1} e^{-pm\ell(n-m\ell)+m\ln r^2} \quad (\text{because } m \geq 2 \text{ and } N \leq n). \] (52)

Because \( 2 \leq m \leq 2\ell + 1 \), we have \( m\ell(n-m\ell) \geq 2\ell[n - \ell(2\ell + 1)] \). Using \( p = \frac{\ln k_{i*} + c(N)}{i*(n-i*)} \) and \( k_{i*} \geq k_{\ell}^{i*(n-i*)/(n-\ell)} \), we obtain
\[
p\ell m(n-\ell m) \geq (\ln k_{i*} + c(N))\frac{2\ell [n - \ell(2\ell + 1)]}{i*(n-i*)} \\
\geq (\ln k_{\ell})\frac{2[n - \ell(2\ell + 1)]}{(n-\ell)} + c(N)\frac{2\ell [n - \ell(2\ell + 1)]}{i*(n-i*)} \\
= (2 + o(1))(\ln k_{\ell}) + (2 + o(1))\frac{\ell}{i*}c(N),
\]
and since \( N \leq n \),
\[
(enp)^{m-1} \leq \left(\frac{en[\ln k_{i*} + c(N)]}{i*(n-i*)}\right)^{2\ell} = O(1)[\ln k_{i*} + c(N)]^{2\ell}.
\]

Using the previous bounds in (52), and considering that \( m \leq N/2 \), we obtain
\[ \mathbb{P}(E) \leq O(N) [\ln k_{i*} + c(N)]^{2\ell} [k_{\ell}]^{-2+o(1)} e^{-(2+o(1))k_{\ell}c(N)+(2\ell+1)\ln r^2} =: \varphi_2(N) \rightarrow 0 \quad \text{as } N \rightarrow \infty, \]
because \( c(N) \rightarrow +\infty \) or \( c \in \mathbb{R} \) implies \( e^{-(2+o(1))k_{\ell}c(N)+(2\ell+1)\ln r^2} = O(1) \) and the conditions \( k_{\ell} = (1 + o(1))k_{\ell}N \) with \( \mu_{\ell} > 0 \) and \( k_{i*} \leq O(N) (\ln k_{i*} + c(N))^{2\ell} [k_{\ell}]^{-2+o(1)} \rightarrow 0 \) as \( N \rightarrow \infty \). Summing up, we have that
\[
\sum_{m=2}^{2\ell+1} \mathbb{P}(\exists \text{ a component of size } m) \leq (2\ell + 1)\varphi_2(N) + o(1) \rightarrow 0 \quad \text{as } N \rightarrow \infty \] (53)
as desired. \( \square \)

### 3.4 Proof of Theorem 1 (1) – (3)

We denote, as before, by \( X \) the number of isolated super-vertices in \( G(N, Kr, p) \). First, as
\[ \mathbb{P}[ G(N, Kr, p) \text{ is connected } ] \leq \mathbb{P}[X = 0], \]
we conclude Theorem 1 (1) from Lemma 1 (2). On the other hand, let \( B \) be the event that \( G(N, Kr, p) \) has components of size between 2 and \( N/2 \), and note that we can write
\[ \mathbb{P}[ G(N, Kr, p) \text{ is connected } ] = \mathbb{P}[X = 0] - \mathbb{P}(B) + \mathbb{P}[\{X \geq 1\} \cap B]. \]
Lemma 4 implies $P(B) = o(1)$, and as a consequence $P(X \geq 1 \cap B) < P(B) = o(1)$. Therefore,

$$P[ G(N, K', p) \text{ is connected } ] = P[X = 0] + o(1),$$

and thus, Theorem 1 (2) follows from Lemma 3 and Lemma 4 while Theorem 1 (3) follows from Lemma 1 (1) and Lemma 4.

4 Proof of Propositions 1 and 2

Consider the random graph $G(N, K', p)$ with $p = c/n$, where $c$ is a constant. Also assume that the limit $u := \lim_{N \to \infty} (n/N)$ exists and is equal to $u := \sum_{i \in \mathbb{Z}^+} i \mu_i$. As discussed in Sect. 2.3, the random graph model $G(N, K', p)$ belongs to the class of IRG studied in [6], and also it can be seen as the Poisson random graph NR$_N(w)$ proposed by Norros and Reittu in [25]. In this section we will analyze the emergence of the giant component and the degree distribution in the random graph $G(N, K', p)$ based on the relationship with the IRG model. In addition we will discuss how these results can be obtained through the connection with the NR$_N(w)$ model.

In order to obtain the results of Proposition 1 and Proposition 2 from the results in [6], we need to analyze some conditions on the sequence of kernels $\{\kappa_N\}_{N \geq 1}$ given by (11). Let us begin with some definitions.

Definition 1 Consider a set of $N$ vertices. Let $\mathcal{V} = (\mathcal{S}, \mu, (x_N)_{N \geq 1})$ be a vertex space and let $\kappa$ be a kernel on $\mathcal{V}$. A sequence $\{\kappa_N\}_{N \geq 1}$ of kernels is graphical on $(\mathcal{S}, \mu)$ with limit $\kappa$ if the following holds.

1. For a.e. $(x, y) \in \mathcal{S}^2$, $x_N \to x$ and $y_N \to y$ imply $\kappa_N(x_N, y_N) \to \kappa(x, y)$.
2. $\kappa$ is continuous a.e. on $\mathcal{S}^2$.
3. $\kappa \in L^1(\mathcal{S}^2, \mu \times \mu)$.
4. If $e(G)$ is the number of edges of $G^V(N, \kappa_N)$, then

$$\lim_{N \to \infty} \frac{E(e(G))}{N} = \frac{1}{2} \int \int \kappa(x, y) d\mu(x) d\mu(y).$$

Note that whether $\{\kappa_N\}_{N \geq 1}$ is graphical depends on the sequences $(x_N)_{N \geq 1}$. The next lemma says that the only condition in our random graph model $G(N, K', p)$ for $\{\kappa_N\}_{N \geq 1}$ to be graphical is that the limit $u$ exists.

Lemma 5 Consider the random graph $G(N, K', p)$ with $p = c/n$, where $c$ is a constant. If the limit $u$ exists and is equal to $\sum_{i=1}^\infty i \mu_i$, then the sequence of kernels $\{\kappa_N\}_{N \geq 1}$ given by (11) is graphical on $\mathcal{V}$ with limit given by the kernel $\kappa$ such that $\kappa(i, j) = (c/u) i j$, for $i, j \in \mathcal{S}$.

Proof It is not difficult to see that $\kappa$ is in fact a kernel on $\mathcal{V}$, and that the conditions (2) and (3) in Definition 1 are satisfied. To check condition (1) in Definition 1, we fix a point $(i, j)$ of $\mathbb{Z}^+ \times \mathbb{Z}^+$ and consider two sequences, $(i_N)_{N \geq 1}$ and $(j_N)_{N \geq 1}$, such that $i_N \to i$ and $j_N \to j$. Since

$$\kappa_N(i_N, j_N) = \frac{N}{n} c i_N j_N + o(1),$$

and $u = \lim_{N \to \infty} (n/N)$ exists, we conclude that

$$\lim_{N \to \infty} \kappa(i_N, j_N) = \left( \frac{c}{u} \right) i j.$$
To show that condition (4) is satisfied, note first that in \( G(N, K', \rho) \),

\[
\frac{1}{2} \int S^2 \kappa(x, y)d\mu(x)d\mu(y) = \frac{1}{2} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \kappa(i, j)\mu_i \mu_j = \frac{c}{2u} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} i \mu_i = \frac{cu}{2}.
\]

Furthermore, we have

\[
\mathbb{E}(e(G)) = \frac{1}{N} \sum_{1 \leq k < l \leq N} p_{x_k, x_l} = \frac{1}{N} \sum_{1 \leq k < l \leq N} \frac{\kappa_N(x_k, x_l)}{N} = \frac{1}{N^2} \sum_{1 \leq k < l \leq N} \frac{N}{n} c x_k x_l + o(1)
\]

\[
\leq \frac{N c}{n} \left( \frac{1}{N} \sum_{k=1}^{N} x_k \right)^2 + o(1)
\]

\[
= \frac{N c}{n} \left( \frac{n}{N} \right)^2 + o(1).
\]

Therefore, we get

\[
\lim_{N \to \infty} \frac{\mathbb{E}(e(G))}{N} \leq \frac{cu}{2}.
\]

On the other hand, we know by Lemma 8.1 ([6]) that if \( \kappa \) is a continuous kernel on a vertex space \( \mathcal{V} \), then

\[
\liminf_{N \to \infty} \frac{\mathbb{E}(e(G))}{N} \geq \frac{1}{2} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \kappa(i, j)\mu_i \mu_j.
\]

Thus, we have

\[
\lim_{N \to \infty} \frac{\mathbb{E}(e(G))}{N} = \frac{1}{2} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \kappa(i, j)\mu_i \mu_j
\]

and the sequence of kernels \( \{\kappa_N\}_{N \geq 1} \) is graphical on \( \mathcal{V} \).

In order to obtain results concerning the size of the giant component, one additional definition is required.

**Definition 2** (Definition 2.10-11 [6]) A kernel \( \kappa \) on \((S, \mu)\) is irreducible if for all \( A \subset S \) and \( \kappa = 0 \) a.e. on \( A \times (S \setminus A) \) implies \( \mu(A) = 0 \) or \( \mu(S \setminus A) = 0 \).

In fact, for technical reasons, a slight weakening of irreducibility is considered.

**Definition 3** (Definition 2.11 [6]) A kernel \( \kappa \) on \((S, \mu)\) is quasi-irreducible if there is a \( \mu \)-continuity set \( S' \subseteq S \) with \( \mu(S') > 0 \) such that the restriction of \( \kappa \) to \( S' \times S' \) is irreducible, and \( \kappa(x, y) = 0 \) if \( x \notin S' \) or \( y \notin S' \).

Now, Propositions 1 and 2 can be obtained as corollaries of Theorems 3.1 and 3.13 in ([6]), respectively, which are included here for the sake of completeness.
Theorem 2 (Theorem 3.1 [6]) Let \( \{\kappa_N\}_{N \geq 1} \) be a graphical sequence of kernels on a vertex space \( \mathcal{V} \) with limit \( \kappa \), and let \( G_N := G^\mathcal{V}(N, \kappa_N) \). Let \( T_\kappa \) be an integral operator defined by

\[
(T_\kappa f)(x) = \int_S \kappa(x, y) f(y) d\mu(y),
\]

for any (measurable) function \( f \), such that its integral is defined (finite or \(+\infty\)) for a.e. \( x \). Let \( ||T_\kappa|| := \sup\{||T_\kappa f||_2 : f \geq 0 \text{ and } ||f||_2 \leq 1\} \).

1. If \( ||T_\kappa|| \leq 1 \), then
   \[
   \lim_{N \to \infty} \frac{L_1(G_N)}{N} = 0
   \]
   in probability, while if \( ||T_\kappa|| > 1 \), then whp \( L_1(G_N) = \Theta(N) \).

2. For any \( \epsilon > 0 \), whp
   \[
   \frac{L_1(G_N)}{N} \leq \rho(\kappa) + \epsilon.
   \]

3. If \( \kappa \) is quasi-irreducible, then
   \[
   \lim_{N \to \infty} \frac{L_1(G_N)}{N} = \rho(\kappa),
   \]
   in probability, where \( \rho(\kappa) := \int_S \rho(\kappa, x) d\mu(x) \), and the function \( x \mapsto \rho(\kappa, x) \) is the maximal fixed point of the non-linear operator \( \Phi_\kappa \) defined by \( \Phi_\kappa f := 1 - e^{-T_\kappa f} \). In addition, \( \rho(\kappa) < 1 \), and \( \rho(\kappa) > 0 \) if and only if \( ||T_\kappa|| > 1 \).

More details regarding the definition and behavior of \( \rho(\kappa) \) can be found in Theorem 6.2, [6].

Theorem 3 (Theorem 3.13 [6]) Let \( \{\kappa_N\}_{N \geq 1} \) be a graphical sequence of kernels on a vertex space \( \mathcal{V} \) with limit \( \kappa \), and let \( G_N := G^\mathcal{V}(N, \kappa_N) \). Let \( Z_k \) the number of vertices of \( G_N \) with degree \( k \), for \( k \geq 0 \). Then, for any fixed \( k \geq 0 \),

\[
\lim_{N \to \infty} \frac{Z_k}{N} = \mathbb{P}(Z = k),
\]

in probability, where \( Z \) has the mixed Poisson distribution \( \int_S P_o(\lambda(x)) d\mu(x) \), and \( \lambda(x) := \int_S \kappa(x, y) d\mu(y) \).

4.1 Proof of Proposition 1

Observe that the kernel \( \kappa \) has the form \( \kappa(i, j) = \varphi(i) \varphi(j) \), with \( \varphi(i) := (c/u)^{1/2} i \), which is the rank-1 case studied in [6]. In this case, we have

\[
||T_\kappa|| = \int_S \varphi^2 d\mu = \sum_{i=1}^{\infty} \frac{c}{u} i^2 \mu_i = c \bar{s}_2.
\]

On the other hand, note that \( \kappa(i, j) = (c/u) ij \) defined on \( (\mathbb{Z}^+, \mu) \), where \( \mu([i]) = \mu_i \) given by (2), equals 0 only if \( i = 0 \) or \( j = 0 \). Hence, \( \kappa \) is irreducible and quasi-irreducible. Therefore, Proposition 1 follows as a consequence of Theorem 2 and Lemma 5.

4.2 Proof of Proposition 2

Proposition 2 follows as a consequence of Theorem 3 and Lemma 5.
4.3 A Second Approach Based on the Relation with the NR\(_N(w)\) Model

Through the relation with the NR\(_N(w)\) model, in this subsection we check certain conditions of the weights in order to obtain the results on phase transition and degree distribution in \(G(N, K', p)\). To do that, take any order of the super-vertices such that \(v_i^\ell\) denotes the \(\ell\)th vertex of size \(i\). Let \(W_N\) be the random weight of a uniformly chosen vertex in \([N]\), where weights are given by (12). Thus, for all \(x \geq 0\), the distribution function of \(W_N\) is given by

\[
F_N(x) = \frac{1}{N} \sum_{v \in [N]} 1_{[w_v \leq x]} = \frac{1}{N} \sum_{i=1}^{r} \sum_{\ell=1}^{k_i} 1_{[w_{v_i^\ell} \leq x]},
\]

\[
= \sum_{i=1}^{r} \frac{k_i}{N} 1_{[i \leq x / c]} 
\]

By Lebesgue’s dominated convergence theorem with respect to the counting measure on \(N\), we have

\[
\lim_{n \to \infty} F_N(x) = \sum_{i=1}^{\infty} \mu_i 1_{[i \leq x / c]} := F(x).
\]

Therefore, letting \(W = ic\) with probability \(\mu_i\), \(i = 1, 2, \ldots\), what we have is

\[
W_N \to W,
\]

in distribution. Furthermore, using Taylor series for \(\ln(1 - x)\), we have

\[
\mathbb{E}[W_n] = \frac{1}{N} \sum_{v \in [N]} w_v = \frac{n^2}{N} \ln \left(\frac{1}{1 - p}\right) = \frac{n^2}{N} \left[\frac{c}{n} + O\left(\frac{1}{n^2}\right)\right].
\]

Since \(u := \lim_{N \to \infty} n/N = \sum_i i \mu_i\), we obtain

\[
\lim_{N \to \infty} \mathbb{E}[W_N] = \mathbb{E}[W] = cu.
\]

These two conditions, (54) and (55), are essential to get the results in Propositions 1 and 2. Specifically, since (54) and (55) hold, an analogous result of Proposition 1 follows by Theorem 3.2 in [31], which uses a reformulation of the NR\(_N(w)\) model in terms of a two-stage branching process.

To see Proposition 2 we need a bit more work. Here we only sketch the proof. Let \(D_i\) denotes the degree of a super-vertex \(u \in [N]\) of size \(i\). By Theorem 3.6.1 in [15] we can prove that \(D_i \to Y_i\) in distribution, where \(Y_i\) is a Poisson random variable with mean \(ic\). Therefore, by (54) and Lebesgue’s dominated convergence theorem, the distribution of the degree of a uniformly chosen super-vertex, \(D_U\), converges to a Mixed Poisson distribution, i.e.,

\[
\mathbb{P}(D_U = k) \to \sum_{i=1}^{\infty} e^{-ci} (ci)^k k! \mu_i := p_k,
\]

as \(N \to \infty\). Now we extend the result to the convergence of the empirical degree sequence. Let

\[
p_k^N := \frac{\Xi_k}{N} = \frac{1}{N} \sum_{i=1}^{r} \sum_{\ell=1}^{k_i} 1_{[\text{deg}(v_i^\ell) = k]},
\]
where $\deg(v'_i)$ is the degree of the $\ell$-th super-vertex of size $i$. By (56) and (57), $\mathbb{E}[p_k^N] = (1/N) \sum_{i=1}^r k_i \mathbb{P}(D_i = k) = \mathbb{P}(D_U = k) \to p_k$, as $N \to \infty$. Thus, by Chebyshev’s inequality, for all $\epsilon > 0$

$$\mathbb{P}[(p_k^N - \mathbb{E}(p_k^N)) \geq \epsilon] \leq \frac{\mathbb{V}(p_k^N)}{\epsilon^2}.$$  

Finally, one may follow the same reasoning used in the proof of Theorem 6.10 in [30] to get that $\mathbb{V}(p_k^N) \to 0$ as $N \to \infty$.

## 5 Discussions

In this section we compare our results with related work.

### 5.1 Comparison with Connectedness of $G(n, p)$

Consider the Erdős–Rényi random graph $G(n, p)$, with $p = \frac{\ln n + c(n)}{n}$. It is well known (see for example [2]) that if $\lim_{n \to \infty} c(n) = -\infty$, then whp $G(n, p)$ is disconnected, but if $\lim_{n \to \infty} c(n) = +\infty$, then whp $G(n, p)$ is connected. Furthermore, if $\lim_{n \to \infty} c(n) = c$ is a constant,

$$\lim_{n \to \infty} \mathbb{P}[G(n, p) \text{ is connected}] = e^{-e^{-c}}.$$  

In order to compare Theorem 1 with the threshold for connectedness of $G(n, p)$, consider the model $G(N, K^r, p)$, with $p = \frac{\ln k_i + c(N)}{\ell(n-i)}$, and assume that $\lim_{N \to \infty}(n/N) = 1$. Since $N := \sum_{i=1}^r k_i$ and $n := \sum_{i=1}^r i k_i$, we have

$$\frac{n}{N} = 1 + \frac{1}{N} \sum_{i=2}^r (i-1) k_i,$$

and therefore it should be $\lim_{N \to \infty} k_i/N = 0$ for each $i = 2, 3, \ldots, r$. This in turns imply by (2) that $\mu_1 = 1$, $\mu_i = 0$ for each $i \in \{2, 3, \ldots, r\}$, and $\ell = 1$, where $\ell := \min\{\ell \in \{1, 2, \ldots, r\} : \mu_\ell > 0\}$. Note that $\mu_1 = 1$ if and only if $k_1 = (1 + o(1)) N$. On the other hand, by definition of $i_*$, i.e. $i_* = \arg \max_i k_i^{1/(n-i)}$, we have $k_{i_*} \geq k_1 \frac{\ln(n-i)}{n-i-1} = k_1 \frac{\ell}{\ell(n-i)}$, and therefore it should be $i_* = \ell = 1$. In brief, we have shown that when $\lim_{N \to \infty}(n/N) = 1$, then

$$\frac{\ln k_{i_*} + c(N)}{i_{*}(n-i_{*})} \sim \frac{n + c(n)}{n}.$$  

Furthermore, we have $\gamma = \lim_{N \to \infty} \sum_{i \neq 1} k_i (k_1 e^\gamma)^{-\frac{(n-i)}{(n-1)}} = 0$, and Theorem 1 (2) yields

$$\lim_{N \to \infty} \mathbb{P}[G(N, K^r, p) \text{ is connected}] = e^{-e^{-c}}.$$  

In words, the asymptotic probability of $G(N, K^r, p)$ being connected is the same as that of $G(n, p)$ being connected, provided $\lim_{N \to \infty}(n/N) = 1$. This is not surprising because in this case both models have the same asymptotic number of vertices.
5.2 Comparison with Connectedness of the Devroye—Fraiman Model

The property of an inhomogeneous random graph (IRG) being connected has recently been studied for a special case of IRG by Devroye and Fraiman in [13] (see Sect. 2.3 for details.) In [13], a connectivity threshold is obtained in terms of an isolation parameter \( \lambda_* := \text{ess inf} \lambda(x) \), where

\[
\lambda(x) := \int_{\Omega} \kappa(x, y) d\mu(y),
\]

and \( \text{ess inf} \lambda(x) := \sup \{ \alpha \in \mathbb{R} : \mu \{ x : \lambda(x) < \alpha \} = 0 \} \). More precisely, it is proved that when \( \lambda_* > 1 \) the graph is connected \( \text{whp} \), while when \( \lambda_* < 1 \) the graph is disconnected \( \text{whp} \).

In order to compare our model \( G(N, K', p) \) to that of Devroye and Fraiman [13], we consider \( G(N, K', p) \) with \( r \in \mathbb{N} \) constant, \( p = \frac{\ln k_\ell + c(N)}{i_\ell (n - i_\ell)} \), and \( K' \) is such that \( \lambda_* = \ell \), i.e., \( \mu_j > 0 \). Thus \( k_\ell = k_j = \mu_j N + o(N) \) and so without loss of generality we may assume \( c(N) = o(N) \). By (1), we have, for \( i, j = 1, \ldots, r \),

\[
pij := 1 - (1 - p)^{ij} = ij p + O(p^2)
\]

\[
= \left( \frac{i}{\ell} \right) \left( \frac{j}{n - \ell} \right) \ln N + o(1) + O(p^2) \sim \left( \frac{i}{u \ell} \right) \frac{\ln N}{N},
\]

where \( u = \lim_{N \to \infty} (n/N) \). So, if we consider the kernel

\[
\kappa(i, j) = \frac{i j}{u \ell},
\]

then the connection probability between pairs of super-vertices in our model is approximately the same as the connection probability between pairs of vertices in the model introduced by [13], provided \( N \) is sufficiently large. Moreover, this choice of kernel corresponds to the case \( \lambda_* = 1 \), because

\[
\lambda(i) = \sum_{j=1}^{r} \kappa(i, j) \mu_j = \sum_{j=1}^{r} \frac{i j}{u \ell} \lim_{N \to \infty} \frac{k_j}{N}
\]

\[
= \frac{i}{\ell} \sum_{j=1}^{r} \lim_{N \to \infty} \frac{j}{(n/N)} \lim_{N \to \infty} \frac{k_j}{N}
\]

\[
= \frac{i}{\ell} \sum_{j=1}^{r} \frac{j k_j}{n} = \frac{i}{\ell}
\]

and \( \sum_{i < a \ell} \mu_i = 0 \) if and only if \( a < 1 \). The case \( \lambda_* = 1 \) was not covered by [13].

5.3 Unbounded Sizes

The proof of Theorem 1 relies mainly on the analysis of the asymptotic number of isolated super-vertices in \( G(N, K', p) \). Note that if the number \( r \) of sizes of super-vertices is either a constant independent of \( N \) or \( r = r(N) \) is tending to a constant as \( N \to \infty \), the value \( \ell = \min \{ i \in \{ 1, \ldots, r \} : \mu_i > 0 \} \), which has been important in the proof of Lemma 7 and Lemma 8, is well defined. Moreover \( \lim_{N \to \infty} (n/N) \) exists. However, if \( r = r(N) \to \infty \) as \( N \to \infty \), we can not guarantee neither the existence of an integer \( i \in \{ 1, 2, 3, \ldots \} \) with \( \mu_i > 0 \) nor that of \( \lim_{N \to \infty} (n/N) \). In this case, the exact distribution of sizes of super-vertices is related to the existence of isolated super-vertices in a more complex manner. We
believe that an analysis should be extended to $r = r(N) \to \infty$ as $N \to \infty$. That analysis will need to deal with additional conditions regarding the velocity for $r(N)$ and appears to be an interesting subject for further work.

5.4 Susceptibility

Consider the model $G(N, K^r, p)$ and let $s_2 := \sum_{i=1}^{r} i^2 k_i / n$. Thus, $s_2$ is the average size of a super-vertex containing a uniformly chosen vertex. If the vertices in each super-vertex were all connected in $G(n, p)$, that is, if the super-vertices were connected components in $G(n, p)$, then $s_2$ would be the susceptibility which is defined as the average component size of $G(n, p)$ (see e.g. [22]).

In Proposition 1 we assume that the limit $\lim_{N \to \infty} (n/N)$ exists and is equal to $\sum_{i=1}^{\mu_i} i \mu_i$. This assumption guarantees that the sequence of kernels which will be given by (11) is graphical in the sense of [6] (see Sects. 2.3 and 4 for details). It is natural to ask what happens if the limit $\lim_{N \to \infty} (n/N)$ does not exist. More specifically, how does it affect the emergence and the size of the giant component in $G(N, K^r, p)$?

Acknowledgements We thank Serguei Popov for suggesting us the construction of the inhomogeneous random graph model studied in the paper. We also thank Luiz Renato Fontes for fruitful discussions during the early stages of this work. The first two authors were financially supported by DFG KA 2748/3-1 and the Austrian Science Fund (FWF): P26826, and the last one by FAPESP 2013/03898-8, 2015/03868-7 and CNPq 479313/2012-1. The second and the third author also thank, respectively, ICMC - Universidade de São Paulo and Università di Torino, for their hospitality. The authors are grateful to the anonymous reviewers for their interesting comments and suggestions.

References

1. Aldous, A., Limic, V.: The entrance boundary of the multiplicative coalescent. Electron. J. Probab. 3, 1–59 (1998)
2. Bollobás, B.: Random Graphs. Cambridge University Press, Cambridge (2001)
3. Bollobás, B., Riordan, O.: Percolation. Cambridge University Press, Cambridge (2006)
4. Bollobás, B., Thomason, A.: Random graphs of small order. Ann. Discret. Math. 118, 47–97 (1985)
5. Bollobás, B., Janson, S., Riordan, O.: The phase transition in a uniformly grown random graph has infinite order. Random Struct. Alg. 26, 1–36 (2005)
6. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. Random Struct. Alg. 31, 3–122 (2007)
7. Britton, T., Deijfen, M., Martin-Löf, A.: Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. 124(6), 1377–1397 (2006)
8. Caldarelli, G., Capocci, A., De Los Rios, P., Muñoz, M.A.: Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. 89, 258702 (2002)
9. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Ann. Comb. 6(2), 125–145 (2002)
10. Chung, F., Lu, L.: The average distance in a random graph with given expected degrees. Internet Math. 1(1), 91–113 (2003)
11. Chung, F., Lu, L.: Complex Graphs and Networks. CBMS Regional Conference Series in Mathematics, vol. 107. Published for the Conference Board of the Mathematical Sciences, Washington, DC (2006)
12. Chung, F., Lu, L.: The volume of the giant component of a random graph with given expected degrees. SIAM J. Discret. Math. 20, 395–411 (2006)
13. Devroye, L., Fraiman, N.: Connectivity of inhomogeneous random graphs. Random Struct. Alg. 45, 408–420 (2014)
14. Durrett, R.: Rigorous result for the CHKNS random graph model. In: Banderier, C., Krattenthaler, C. (eds.) Discrete Random Walks, DRW’03. Discrete Mathematics and Theoretical Computer Science Proceedings AC, pp. 95–104 (2003)
15. Durrett, R.: Probability: Theory and Examples, 3rd edn. Duxbury Advanced Series. Duxbury Press, Pacific Grove (2005)
16. Erdős, P., Rényi, A.: On random graphs I. Math. Debrecen 6, 290–297 (1959)
17. Erdős, P., Rényi, A.: On the Evolution of Random Graphs, vol. 5, pp. 17–61. Institute of Mathematics Hungarian Academy of Sciences, Hungary (1960)
18. Flory, P.J.: Principles of Polymer Chemistry. Cornell University Press, New York (1953)
19. Gilbert, E.N.: Random graphs. Ann. Math. Stat. 30, 1141–1144 (1959)
20. Grimmett, G.: Percolation, 2nd edn. Springer, Berlin (1999)
21. Janson, S.: Asymptotic equivalence and contiguity of some random graphs. Random Struct. Alg. 36(1), 26–45 (2010)
22. Janson, S., Spencer, J.: Phase transition for modified Erdős–Rényi processes. Ark. Mat. 50, 305–329 (2012)
23. Janson, S., Łuczak, T., Ruciński, A.: Random Graphs. Wiley, New York (2000)
24. Newman, M.E.J.: Power laws, Pareto distributions and Zipf’s law. Contemp. Phys. 46, 323–351 (2005)
25. Norros, I., Reittu, H.: On a conditionally Poissonian graph process. Adv. Appl. Probab. 38(1), 59–75 (2008)
26. Seshadhri, C., Kolda, T.G., Pinar, A.: Community structure and scale-free collections of Erdős–Rényi graphs. Phys. Rev. E 85, 056109 (2012)
27. Söderberg, B.: General formalism for inhomogeneous random graphs. Phys. Rev. E 66, 066121 (2002)
28. van den Esker, H., van der Hofstad, R., Hooghiemstra, G.: Universality for the distance in finite variance random graphs. J. Stat. Phys. 124(6), 169–202 (2008)
29. van der Hofstad, R.: Critical behavior in inhomogeneous random graphs. Random Struct. Alg. 42, 480–508 (2012)
30. van der Hofstad, R.: Random Graphs and Complex Networks, vol. 1. Cambridge University Press, Cambridge (2017)
31. van der Hofstad, R.: Random Graphs and Complex Networks, vol. 2. http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf (2018+)