Merging percolation and classical random graphs: Phase transition in dimension 1

TATYANA S. TUROVA[1] and THOMAS VALLIER

Mathematical Center, University of Lund, Box 118, Lund S-221 00, Sweden.

Abstract

We study a random graph model which combines properties of the edge percolation model on Z^d and a classical random graph $G(n, c/n)$. We show that this model, being a *homogeneous* random graph, has a natural relation to the so-called "rank 1 case" of *inhomogeneous* random graphs. This allows us to use the newly developed theory of inhomogeneous random graphs to describe completely the phase diagram in the case $d = 1$. The phase transition is similar to the classical random graph, it is of the second order. We also find the scaled size of the largest connected component above the phase transition.

1 Introduction.

We consider a graph on the set of vertices $V_N^d := \{1, \dots, N\}^d$ in Z^d, where the edges between any two different vertices i and j are presented independently with probabilities

$$p_{ij} = \begin{cases} q, & \text{if } |i-j| = 1, \\ c/N^d, & \text{if } |i-j| > 1, \end{cases}$$

where $0 \le q \le 1$ and $0 < c < N$ are constants. This graph, call it $G_N^d(q, c)$ is a mixture of percolation model, where each pair of neighbours in Z^d is connected with probability q, and a random graph model, where each vertex is connected to any other vertex with probability $c/|V_N^d|$.

The introduced model is a simplification of the most common graphs designed to study natural phenomena, in particular, biological neural networks [8]. Observe the difference between this and the so-called "small-world" models intensively studied after [9]. In the "small-world" models where edges from the grid may be kept or removed, only finite number (often at most $2d$) of the long-range edges may come out of each vertex, and the probability of those is a fixed number.

We are interested in the limiting behaviour of the introduced graph $G_N^d(q, c)$ as $N \to \infty$. One can consider this model as a graph on Z^d or on a torus, in the limit the results are the

[1]Research was supported by the Swedish Natural Science Research Council.

same. The one-dimensional case which we study here, is exactly solvable. We shall write $G^1_N(q,c) = G_N(q,c)$.

Let X be a random variable with $Fs(1-q)$-distribution, i.e.,

$$\mathbf{P}\{X = k\} = (1-q)q^{k-1}, \qquad k = 1, 2, \dots\ , \tag{1.1}$$

with

$$\mathbf{E}X = \frac{1}{1-q}\ .$$

Let further $C_1\Big(G\Big)$ denote the size of the largest connected component in a graph G.

Theorem 1.1. *For any* $0 \le q < 1$ *define*

$$c^{cr}(q) = \frac{\mathbf{E}X}{\mathbf{E}X^2} = \frac{1-q}{1+q}\ . \tag{1.2}$$

i) If $c < c^{cr}(q)$ *then there exists a constant* $\alpha = \alpha(q,c) > \frac{1}{|\log q|}$ *such that*

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) > \alpha\, \log N\right\} \to 0, \tag{1.3}$$

and for any $\alpha_1 < \frac{1}{|\log q|}$

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) < \alpha_1 \log N\right\} \to 0, \tag{1.4}$$

as $N \to \infty$.

ii) If $c \ge c^{cr}(q)$ *then*

$$\frac{C_1\Big(G_N(q,c)\Big)}{N} \xrightarrow{P} \beta \tag{1.5}$$

as $N \to \infty$, *with* $\beta = \beta(q,c)$ *defined as the maximal solution to*

$$\beta = 1 - \frac{1}{\mathbf{E}X}\ \mathbf{E}\left\{Xe^{-cX\beta}\right\}. \tag{1.6}$$

Observe the following duality of this result. For any $c < 1$ we know that the subgraph induced in our model by the long-range edges may have at most $O(\log N)$ vertices in a connected component. According to Theorem 1.1, for any $c < 1$ there is

$$q^{cr}(c) = \frac{1-c}{1+c}$$

such that for all $q^{cr}(c) < q < 1$ our model will have a giant component with a size of order N, while any $q < q^{cr}(c)$ is insufficient to produce a giant component in $G_N(q,c)$. Hence, Theorem 1.1 may also tell us something about the "distances" between the components of a random graph when it is considered on the vertices of $\mathbf{Z}$.

Remark 1.1. *In the proof of (1.3) we will show how to obtain $\alpha(q,c)$, and will discuss how optimal this value is. Statement (1.4) is rather trivial (and far from being optimal): it follows from a simple observation (see the details below) that*

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) < \frac{1}{|\log q|}\log N\right\} \leq \mathbf{P}\left\{C_1\Big(G_N(q,0)\Big) < \frac{1}{|\log q|}\log N\right\} \to 0.$$

Remark 1.2. *For any fixed c function $\beta(q,c)$ is continuous at $q=0$: if $q=0$, i.e., when our graph is merely a classical $G_{n,c/n}$ random graph, then $X \equiv 1$ and (1.6) becomes a well-known relation. Equation (1.6) can be written in an exact form:*

$$\beta = 1 - \frac{e^{c\beta}}{\left(e^{c\beta}-q\right)^2}(1-q)^2 \, .$$

It is easy to check that if $c \leq c^{cr}$ then the equation (1.6) does not have a strictly positive solution, while $\beta = 0$ is always a solution to (1.6). Therefore one can derive

$$\beta_c' \mid_{c\downarrow c^{cr}} = 2\frac{\big(\mathbf{E}(X^2)\big)^3}{\big(\mathbf{E}X\big)^2\mathbf{E}(X^3)} = 2\frac{(1-q^2)^2}{q^2+4q+1}. \tag{1.7}$$

This shows that the emergence of the giant component at critical parameter $c = c^{cr}$ becomes slower as q increases, but the phase transition remains of the second order (exponent 1) for any $q < 1$.

We conjecture that similar results hold in the higher dimensions if $q < Q^{cr}(d)$, where $Q^{cr}(d)$ is the percolation threshold in the dimension d. More exactly, Theorem 1.1 (as well as the first equality in (1.7)) should hold with X replaced by another random variable, which is stochastically not larger than the size of the open cluster at the origin in the edge percolation model with a probability of edge q. It is known from the percolation theory (see, e.g., [3]) that the tail of the distribution of the size of an open cluster in the subcritical phase decays exponentially. This should make possible to extend our arguments (where we use essentially the distribution of X) to the general case.

Our result in the supercritical case, namely equation (1.6) looks somewhat similar to the equation obtained in [2] for the "volume" (the sum of degrees of the involved vertices) of the giant component in the graph with a given sequence of the expected degrees. Note, however, that the model in [2] (as well as the derivations of the results) differs essentially from the one studied here. In particular, in our model the critical mean degree when $c = c^{cr}$ and $N \to \infty$ is given by

$$2q + c^{cr} = 2q + \frac{1-q}{1+q} = 1 + \frac{2q^2}{1+q} \tag{1.8}$$

which is strictly greater than 1 for all positive q. This is in a contrast with the model studied in [2], where the critical expected average degree is still 1 as in the classical random graph.

Although our model (when considered on the ring or torus in higher dimensions) is a perfectly *homogeneous* random graph, in the sense that the degree distribution is the same for any vertex, we study it via *inhomogeneous* random graphs, making use of the recently developed theory from [1]. The idea is the following. First, we consider the subgraph induced by the short-range edges, i.e., the edges which connect two neighbouring nodes with probability q. It is composed of the consecutive connected paths (which may consist just of one single vertex) on $V_N = \{1, \dots, N\}$. Call a *macro-vertex* each of the component of this subgraph. We say that a macro-vertex is of type k, if k is the number of vertices in it. Conditionally on the set of macro-vertices, we consider a graph on these macro-vertices induced by the long-range connections. Two macro-vertices are said to be connected if there is at least one (long-range type) edge between two vertices belonging to different macro-vertices. Thus the probability of an edge between two macro-vertices v_i and v_j of types x and y correspondingly, is

$$\widetilde{p}_{xy}(N) := 1 - \left(1 - \frac{c}{N}\right)^{xy}. \tag{1.9}$$

Below we argue that this model fits the conditions of a general inhomogeneous graph model defined in [1], find the critical parameters and characteristics for the graph on macro-vertices, and then we turn back to the original model. We use essentially the results from [1] to derive (1.6), while in the subcritical case our approach somewhat differs from the one in [1]; we discuss this in the end of Section 2.4. We shall also note that our graph on macro-vertices is similar to the model studied in [5], and our results on the critical value agree with those in [5].

Finally we comment that our result should help to study more general model for the propagation of the neuronal activity introduced in [8]. Here we show that a giant component in the graph can emerge from two sources, none of which can be neglected, but each of which may be in the subcritical phase, i.e., even when both $q < 1$ and $c < 1$. In particular, for any $0 < c < 1$ we can find $q < 1$ which allows with a positive probability the propagation of impulses through the large part of the network due to the local activity.

2 Proof

2.1 Random graph on macro-vertices.

Denote X a random number of the vertices connected through short-range edges to the vertex 1 on $V_\infty^1 = \{1, 2, \ldots\}$. Clearly, X has the First success distribution defined in (1.1). Let $X_1, X_2, \ldots$, be independent copies of X, and define for any $N > 1$

$$T(N) := \min\{n \geq 0 : \sum_{i=1}^{n} X_i \leq N, \quad \sum_{i=1}^{n+1} X_i > N\},$$

where we assume that a sum over an empty set equals zero.

Consider now the subgraph on $V_N = \{1, \ldots, N\}$ induced by the short-range edges. This means that any two vertices i and $i+1$ from V_N are connected with probability q independent of the rest. By the construction this subgraph, call it $G_N^{(s)}(q)$, is composed of a random number of connected paths of random sizes. We call here the size of a path the number of its vertices. Clearly, there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where the number of paths in $G_N^{(s)}(q)$ equals $T(N)$ if $\sum_{i=1}^{T(N)} X_i = N$, or $T(N) + 1$ if $\sum_{i=1}^{T(N)} X_i < N$. Correspondingly, the sizes of the paths follow the distribution of

$$\mathbf{X} = (X_1, X_2, \ldots, X_{T(N)}, N - \sum_{i=1}^{T(N)} X_i) \tag{2.1}$$

(where the last entry may take zero value).

On the other hand, the number of the connected components of $G_N^{(s)}(q)$ exceeds exactly by one the number of "missed" short edges on V_N. This means that on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ there is a random variable Y_N distributed as $Bin(N-1, 1-q)$, such that either $T(N) = Y_N + 1$ or $T(N) + 1 = Y_N + 1$, and in any case

$$0 \leq T(N) - Y_N \leq 1. \tag{2.2}$$

This together with the Strong Law of Large Numbers implies

Proposition 2.1.

$$\frac{T(N)}{N} \stackrel{a.s.}{\to} 1 - q = \frac{1}{\mathbf{E}X}$$

as $N \to \infty$. □

Also the relation (2.2) allows us to use the large deviation inequality from [4] (formula (2.9), p.27 in [4]) for the binomial random variables in order to obtain the following rate of convergence

$$\mathbf{P}\left\{\left|\frac{T(N)}{N}-\frac{1}{\mathbf{E}X}\right|>\delta\right\}\leq 2\exp\left(-\frac{\delta^2}{12(1-q)}N\right) \tag{2.3}$$

for all $\delta>0$ and $N>2/\delta$.

Define for any $k\geq 1$ an indicator function

$$I_k(x)=\begin{cases}1, & \text{if } x=k,\\ 0, & \text{otherwise.}\end{cases}$$

As an immediate corollary of Proposition 2.1 and the Law of Large Numbers we also get the following result.

Proposition 2.2. *For any fixed* $k\geq 1$

$$\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i) \stackrel{P}{\to} \mathbf{P}\{X=k\}=(1-q)q^{k-1}=:\mu(k) \tag{2.4}$$

as $N\to\infty$. □

Given a vector of paths $\mathbf{X}$ defined in (2.1), we introduce another graph $\widetilde{G}_N(\mathbf{X},q,c)$ as follows. The set of vertices of $\widetilde{G}_N(\mathbf{X},q,c)$ we denote $\{v_1,\ldots,v_{T(N)}\}$. Each vertex v_i is said to be of type X_i, which means that v_i corresponds to the set of X_i connected vertices on V_N. We shall also call any vertex v_i of $\widetilde{G}_N(\mathbf{X},q,c)$ a *macro-vertex*, and write

$$v_i=\begin{cases}\{1,\ldots,X_1\}, & \text{if } i=1;\\ \\ \{\sum_{j=1}^{i-1}X_j+1,\ldots,\sum_{j=1}^{i-1}X_j+X_i\}, & \text{if } i>1.\end{cases} \tag{2.5}$$

With this notation the type of a vertex v_i is simply the cardinality of set v_i. The space of the types of macro-vertices is $S=\{1,2,\ldots\}$. According to (2.4) the distribution of type of a (macro-)vertex in graph $\widetilde{G}_N(\mathbf{X},q,c)$ converges to measure μ on S. The edges between the vertices of $\widetilde{G}_N(\mathbf{X},q,c)$ are presented independently with probabilities induced by the original graph $G_N(q,c)$. More precisely, the probability of an edge between any two vertices v_i and v_j of types x and y correspondingly, is $\widetilde{p}_{xy}(N)$ introduced in (1.9). Clearly, this construction provides a one-to-one correspondence between the connected components in the graphs $\widetilde{G}_N(\mathbf{X},q,c)$ and $G_N(q,c)$: the number of the connected components is the same for both graphs, as well as the number of the involved vertices from V_N in two corresponding components. In fact, considering conditionally on $\mathbf{X}$ graph $\widetilde{G}_N(\mathbf{X},q,c)$ we neglect only those long-range edges from $G_N(q,c)$, which connect vertices within each v_i, i.e., the vertices which are already connected through the short-range edges.

Consider now

$$\widetilde{p}_{xy}(N) = 1 - \left(1 - \frac{c}{N}\right)^{xy} =: \frac{\kappa'_N(x,y)}{N}. \tag{2.6}$$

Observe that if $x(N) \to x$ and $y(N) \to y$ then

$$\kappa'_N(x(N),y(N)) \to cxy \tag{2.7}$$

for all $x, y \in S$. In order to place our model into the framework of the inhomogeneous random graphs from [1] let us introduce another (random) kernel

$$\kappa_{T(N)}(x,y) = \frac{T(N)}{N}\kappa'_N(x,y),$$

so that we can rewrite the probability $\widetilde{p}_{xy}(N)$ in a graph $\widetilde{G}_N(\mathbf{X},q,c)$ taking into account the size of the graph:

$$\widetilde{p}_{xy}(N) = \frac{\kappa_{T(N)}(x,y)}{T(N)}. \tag{2.8}$$

(We use notations from [1] whenever it is appropriate.) According to Proposition 2.1 and (2.7), if $x(N) \to x$ and $y(N) \to y$ then

$$\kappa_{T(N)}(x(N),y(N)) \to \kappa(x,y) := \frac{c}{\mathbf{E}X}xy \quad a.s. \tag{2.9}$$

as $N \to \infty$ for all $x, y \in S$.

Hence, in view of Proposition 2.2 we conclude that conditionally on $T(N) = t(N)$, where $t(N)/N \to 1/\mathbf{E}X$, our model falls into the so-called "rank 1 case" of the general inhomogeneous random graph model $G^{\mathcal{V}}(t(N),\kappa_{t(N)})$ with a vertex space $\mathcal{V} = (S,\mu,(X_1,\dots,X_{t(N)})_{N\geq 1})$ from [1] (Chapter 16.4). Furthermore, it is not difficult to verify with a help of the Propositions 2.1 and 2.2 that

$$\kappa \in L^1(S\times S, \mu\times\mu), \tag{2.10}$$

since

$$\sum_{y=1}^{\infty}\sum_{x=1}^{\infty}(1-q)xq^{x-1}(1-q)yq^{y-1} = \left(\frac{1}{1-q}\right)^2,$$

and for any $t(N)$ such that $t(N)/N \to 1/\mathbf{E}X$

$$\frac{1}{t(N)}\mathbf{E}\{e(\widetilde{G}_N(\mathbf{X},q,c))|T(N)=t(N)\} \to \frac{1}{2}\sum_{y=1}^{\infty}\sum_{x=1}^{\infty}\kappa(x,y)\mu(x)\mu(y), \tag{2.11}$$

where $e(G)$ denotes the number of edges in a graph G. According to Definition 2.7 from [1], under the conditions (2.11), (2.10) and (2.9) the sequence of kernels $\kappa_{t(N)}$ (on the countable space $S \times S$) is called *graphical* on $\mathcal{V}$ with limit κ.

2.2 A branching process related to $\widetilde{G}_N(\mathbf{X}, q, c)$.

Here we closely follow the approach from [1]. We shall use a well-known technique of branching processes to reveal the connected component in graph $\widetilde{G}_N(\mathbf{X}, q, c)$. Recall first the usual algorithm of finding a connected component. Conditionally on the set of macro-vertices, take any vertex v_i to be the root. Find all the vertices $\{v^1_{i_1}, v^1_{i_2}, ..., v^1_{i_n}\}$ connected to this vertex v_i in the graph $\widetilde{G}_N(\mathbf{X}, q, c)$, call them the first generation of v_i, and then mark v_i as "saturated". Then for each non-saturated but already revealed vertex, we find all the vertices connected to them but which have not been used previously. We continue this process until we end up with a tree of saturated vertices.

Denote $\tau_N(x)$ the set of the macro-vertices in the tree constructed according to the above algorithm with the root at a vertex of type x.

It is plausible to think (and in our case it is correct, as will be seen below) that this algorithm with a high probability as $N \to \infty$ reveals a tree of the offspring of the following multi-type Galton-Watson process with type space $S = \{1, 2, \ldots\}$: at any step, a particle of type $x \in S$ is replaced in the next generation by a set of particles where the number of particles of type y has a Poisson distribution $Po(\kappa(x, y)\mu(y))$. Let $\rho(\kappa; x)$ denote the probability that a particle of type x produces an infinite population.

Proposition 2.3. *The function $\rho(\kappa; x)$, $x \in S$, is the maximum solution to*

$$\rho(\kappa; x) = 1 - e^{-\sum_{y=1}^{\infty} \kappa(x,y)\mu(y)\rho(\kappa;y)}. \tag{2.12}$$

Proof. We have

$$\sum_{y=1}^{\infty} \kappa(x, y)\mu(y) = \frac{c}{\mathbf{E}X} \frac{x}{1-q} < \infty \text{ for any } x,$$

which together with (2.10) verifies that the conditions of Theorem 6.1 from [1] are satisfied, and the result (2.12) follows by this theorem. □

Notice that it also follows by the same Theorem 6.1 from [1] that $\rho(\kappa; x) > 0$ for all $x \in S$ if and only if

$$\frac{c}{\mathbf{E}X} \sum_{y=1}^{\infty} y^2 \mu(y) = c\,\frac{\mathbf{E}X^2}{\mathbf{E}X} = c\,\frac{1+q}{1-q} > 1; \tag{2.13}$$

otherwise, $\rho(\kappa; x) = 0$ for all $x \in S$. Hence, the formula (1.2) for the critical value follows from (2.13).

As we showed above, conditionally on $T(N)$ (so that $T(N)/N \to 1/\mathbf{E}X$) the sequence $\kappa_{T(N)}$ is graphical on $\mathcal{V}$. Hence, the conditions of Theorem 3.1 from [1] are satisfied and we derive (first, conditionally on $T(N)$, and therefore unconditionally) that

$$\frac{C_1(\widetilde{G}_N(\mathbf{X}, q, c))}{T(N)} \xrightarrow{P} \rho(\kappa),$$

where $\rho(\kappa) = \sum_{x=1}^{\infty} \rho(\kappa; x)\mu(x)$. This together with Proposition 2.1 on the *a.s.* convergence of $T(N)$ implies

$$\frac{C_1(\widetilde{G}_N(\mathbf{X}, q, c))}{N} \xrightarrow{P} (1-q)\rho(\kappa). \tag{2.14}$$

Notice that here $C_1(\widetilde{G}_N(\mathbf{X}, q, c))$ is the number of macro-vertices in $\widetilde{G}_N(\mathbf{X}, q, c)$.

2.3 On the distribution of types of vertices in $\widetilde{G}_N(\mathbf{X}, q, c)$.

Given a vector of paths $\mathbf{X}$ (see (2.1)) we define a random sequence

$$\mathcal{N} = \{\mathcal{N}_1, \dots \mathcal{N}_N\},$$

where

$$\mathcal{N}_k = \mathcal{N}_k(\mathbf{X}) = \sum_{i=1}^{T(N)} I_k(X_i).$$

In words, $\mathcal{N}_k$ is the number of (macro-)vertices of type k in the set of vertices of graph $\widetilde{G}_N(\mathbf{X}, q, c)$. We shall prove here a useful result on the distribution of $\mathcal{N}$ (which is stronger than Proposition 2.2).

Lemma 2.1. *For any fixed $\varepsilon > 0$*

$$\mathbf{P}\{|\mathcal{N}_k/T(N) - \mu(k)| > \varepsilon\, k\mu(k) \quad \textit{for some } 1 \le k \le N\} = o(1) \tag{2.15}$$

as $N \to \infty$.

Proof. Let us fix $\varepsilon > 0$ arbitrarily. Observe that for any $K > 1/\varepsilon$

$$\mathbf{P}\{|\mathcal{N}_k/T(N) - \mu(k)| > \varepsilon\, k\mu(k) \quad \text{for some } 1 \le k \le N\} \tag{2.16}$$

$$\le \mathbf{P}\{\max_{1 \le i \le T(N)} X_i > K\}$$

$$+\mathbf{P}\{|\mathcal{N}_k/T(N)-\mu(k)|>\varepsilon\, k\mu(k) \quad \text{for some } 1\le k\le K\}.$$

Next we shall choose an appropriate $K=K(N)$ so that we will be able to bound from above by $o(1)$ (as $N\to\infty$) each of the summands on the right in (2.16).

Let us fix $\delta>0$ arbitrarily, and define an event

$$\mathcal{A}_{\delta,N}=\left\{\left|\frac{T(N)}{N}-\frac{1}{\mathbf{E}X}\right|\le\delta\right\}. \tag{2.17}$$

Recall that according to (2.3)

$$\mathbf{P}(\mathcal{A}_{\delta,N})\ge 1-2\exp\left(-\frac{\delta^2}{12(1-q)}\,N\right)=1-o(1) \tag{2.18}$$

as $N\to\infty$. Now we derive

$$\mathbf{P}\{\max_{1\le i\le T(N)}X_i>K\}\le\mathbf{P}\{\max_{1\le i\le T(N)}X_i>K\mid\mathcal{A}_{\delta,N}\}\mathbf{P}(\mathcal{A}_{\delta,N})+\mathbf{P}\{\overline{\mathcal{A}_{\delta,N}}\} \tag{2.19}$$

$$\le\left(\frac{1}{\mathbf{E}X}+\delta\right)N\,\mathbf{P}\{X_1>K\mid\mathcal{A}_{\delta,N}\}\mathbf{P}(\mathcal{A}_{\delta,N})+\mathbf{P}\{\overline{\mathcal{A}_{\delta,N}}\}$$

$$\le\left(\frac{1}{\mathbf{E}X}+\delta\right)N\,\mathbf{P}\{X_1>K\}+\mathbf{P}\{\overline{\mathcal{A}_{\delta,N}}\}$$

as $N\to\infty$. Making use of the formula (1.1) for the distribution of X_1 we obtain from (2.19) and (2.18)

$$\mathbf{P}\{\max_{1\le i\le T(N)}\ge K\}\le CNq^K+2\exp\left(-\frac{\delta^2}{12(1-q)}\,N\right) \tag{2.20}$$

as $N\to\infty$, where $C=C(\delta,q)$ is some finite positive constant. Let now $\omega_1(N)<N$ be any function tending to infinity with N, and set

$$K(N)=\frac{1}{|\log q|}\log N+\omega_1(N). \tag{2.21}$$

Clearly, bound (2.20) with K replaced by $K(N)$ implies

$$\mathbf{P}\{\max_{1\le i\le T(N)}X_i\ge K(N)\}=o(1) \tag{2.22}$$

as $N\to\infty$.

Now we consider the last term in (2.16). Let us define

$$k_0:=\max\left\{\left[\frac{1}{\varepsilon}\right],\left[\frac{1}{|\log q|}\right]\right\}+2. \tag{2.23}$$

Then we obtain making use of (2.18)

$$\mathbf{P}\{|\mathcal{N}_k/T(N)-\mu(k)|>\varepsilon\, k\mu(k) \quad \text{for some } 1\le k\le K(N)\} \tag{2.24}$$

$$\le \sum_{k=1}^{k_0}\mathbf{P}\{|\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i)-\mu(k)|>\varepsilon\, k\mu(k)\}$$

$$+\sum_{k=k_0+1}^{K(N)}\mathbf{P}\{|\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i)-\mu(k)|>\varepsilon\, k\mu(k)\mid \mathcal{A}_{\delta,N}\}\mathbf{P}(\mathcal{A}_{\delta,N})+o(1)$$

$$=\sum_{k=k_0+1}^{K(N)}\mathbf{P}\{|\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i)-\mu(k)|>\varepsilon\, k\mu(k)\mid \mathcal{A}_{\delta,N}\}\mathbf{P}(\mathcal{A}_{\delta,N})+o(1)$$

as $N\to\infty$, where the last equality is due to Proposition 2.2. Notice that for each $k>k_0$ we have $\varepsilon\, k>1$ and therefore

$$\mathbf{P}\left\{\left|\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i)-\mu(k)\right|>\varepsilon\, k\mu(k)\mid \mathcal{A}_{\delta,N}\right\} \tag{2.25}$$

$$=\mathbf{P}\left\{\frac{1}{T(N)}\sum_{i=1}^{T(N)} I_k(X_i)-\mu(k)>\varepsilon\, k\mu(k)\mid \mathcal{A}_{\delta,N}\right\}$$

$$\le \mathbf{P}\left\{\frac{1}{\left(\frac{1}{\mathbf{E}X}-\delta\right)N}\sum_{i=1}^{\left[\left(\frac{1}{\mathbf{E}X}+\delta\right)N\right]+1} I_k(X_i)>\mu(k)+\varepsilon\, k\mu(k)\mid \mathcal{A}_{\delta,N}\right\}=:\mathbf{P}(k).$$

Set $t(N)=\left[\left(\frac{1}{\mathbf{E}X}+\delta\right)N\right]+1$. Then using the bound

$$\frac{\left(\frac{1}{\mathbf{E}X}-\delta\right)N}{t(N)}>1-\frac{5}{2}\mathbf{E}X\,\delta$$

for all $N>2/\delta$, we derive

$$\mathbf{P}(k)\le \mathbf{P}\left\{\frac{1}{t(N)}\sum_{i=1}^{t(N)} I_k(X_i)>\mu(k)(1+\varepsilon\, k)(1-\frac{5\mathbf{E}X}{2}\delta)\mid \mathcal{A}_{\delta,N}\right\} \tag{2.26}$$

for all $N > 2/\delta$. Now for all $k > k_0$ and $0 < \delta < \frac{1}{10\mathbf{E}X}$ we have $(1+\varepsilon\, k)(1-\frac{5\mathbf{E}X}{2}\delta) \geq 1 + \frac{\varepsilon}{2}\, k$, and therefore

$$\mathbf{P}(k) \leq \mathbf{P}\left\{ \frac{1}{t(N)} \sum_{i=1}^{t(N)} I_k(X_i) > \mu(k)(1+\frac{\varepsilon}{2}\, k) \mid \mathcal{A}_{\delta,N} \right\} \tag{2.27}$$

$$\leq \mathbf{P}\left\{ \frac{1}{t(N)} \sum_{i=1}^{t(N)} I_k(X_i) > \mu(k)(1+\frac{\varepsilon}{2}\, k) \right\} / \, \mathbf{P}(\mathcal{A}_{\delta,N}).$$

Note that $\sum_{i=1}^{t(N)} I_k(X_i)$ follows the binomial distribution $Bin(t(N), \mu(k))$. This allows us to use the large deviation inequality from [4] (see (2.5), p.26 in [4]) and derive

$$\mathbf{P}\left\{ \frac{1}{t(N)} \sum_{i=1}^{t(N)} I_k(X_i) > \mu(k)(1+\frac{\varepsilon}{2}\, k) \right\} \tag{2.28}$$

$$\leq \exp\left(- \frac{(\frac{\varepsilon}{2} k\mu(k)t(N))^2}{\frac{1}{3}\varepsilon k\mu(k)t(N) + 2\mu(k)t(N)} \right) \leq \exp\big(- \frac{1}{10}\varepsilon k\mu(k)t(N)\big)$$

for all $k > k_0$. Substituting this into (2.27) we obtain

$$\mathbf{P}(k) \leq \exp\big(- \frac{1}{10}\varepsilon k\mu(k)t(N)\big) / \, \mathbf{P}(\mathcal{A}_{\delta,N}) \tag{2.29}$$

for all $k > k_0$. The last bound combined with (2.25) and (2.24) leads to

$$\mathbf{P}\{|\mathcal{N}_k/T(N) - \mu(k)| > \varepsilon\, k\mu(k) \quad \text{for some } 1 \leq k \leq K(N)\}$$

$$\leq \sum_{k=k_0+1}^{K(N)} \exp\big(- \frac{1}{10}\varepsilon k\mu(k)t(N)\big) + o(1),$$

as $N \to \infty$. Taking into account that function $k\mu(k)$ is decreasing for $k > k_0$ we derive from the last bound:

$$\mathbf{P}\{|\mathcal{N}_k/T(N) - \mu(k)| > \varepsilon\, k\mu(k) \quad \text{for some } 1 \leq k \leq K(N)\} \tag{2.30}$$

$$\leq \exp\big(- \frac{1}{10}\varepsilon K(N)\mu(K(N))t(N) + \log K(N)\big) + o(1),$$

as $N \to \infty$.

Setting now $\omega_1(N) = \log\log\log N$ in (2.21), it is easy to check that for

$$K(N) = \frac{1}{|\log q|} \log N + \log\log\log N$$

the entire right-hand side of the inequality (2.30) is $o(1)$ as $N \to \infty$. This together with the previous bound (2.22) and inequality (2.16) finishes the proof of lemma. □

2.4 Proof of Theorem 1.1 in the subcritical case $c < c^{cr}(q)$.

Let us fix $0 \le q < 1$ and then $c < c^{cr}(q)$ arbitrarily. Given $\mathbf{X}$ let again v_i denote the macro-vertices with types X_i, $i = 1, 2, \ldots$, respectively, and let $\widetilde{L}$ denote a connected component in $\widetilde{G}_N(\mathbf{X}, q, c)$. Firstly, for any $K > 0$ and $0 < \delta < 1/\mathbf{E}X$ we derive with help of (2.18)

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) < K\right\} \le \mathbf{P}\left\{C_1\Big(G_N(q,0)\Big) < K\right\} = \mathbf{P}\left\{\max_{1\le i\le T(N)} X_i < K\right\} \tag{2.31}$$

$$\le \mathbf{P}\left\{\max_{1\le i\le T(N)} X_i < K \mid \mathcal{A}_{\delta,N}\right\} + o(1) \le \left(1 - \mathbf{P}\left\{X \ge K\right\}\right)^{N\left(\frac{1}{\mathbf{E}X}-\delta\right)} + o(1),$$

as $N \to \infty$, where X has the $Fs(1-q)$-distribution. Since

$$\mathbf{P}\left\{X \ge K\right\} = q^{K-1},$$

we derive from (2.31) for any $a_1 < \frac{1}{|\log q|}$ and $K = a_1 \log N$

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) < a_1 \log N\right\} = o(1),$$

which proves statement (1.4).

Consider now for any positive constant a and a function $w = w(N) \ge \log N$

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) > aw\right\} = \mathbf{P}\left\{\max_{\widetilde{L}} \sum_{v_i \in \widetilde{L}} X_i > aw\right\}. \tag{2.32}$$

We know already from (2.14) that in the subcritical case the size (the number of macro-vertices) of any $\widetilde{L}$ is **whp** $o(N)$. Note that when the kernel $\kappa(x, y)$ is not bounded uniformly in both arguments, which is our case, it is not granted that the largest component in the subcritical case is at most of order $\log N$ (see, e.g., discussion of Theorem 3.1 in [1]). Therefore first we shall prove the following intermediate result.

Lemma 2.2. *If $c < c^{cr}(q)$ then*

$$\mathbf{P}\left\{C_1\Big(\widetilde{G}_N(\mathbf{X}, q, c)\Big) > N^{1/2}\right\} = o(1). \tag{2.33}$$

Proof. Let us fix $\varepsilon > 0$ and $\delta > 0$ arbitrarily and introduce the following event

$$\mathcal{B}_N := \mathcal{A}_{\delta,N} \cap \left(\max_{1\le i\le T(N)} X_i \le \frac{2}{|\log q|}\log N\right) \cap \left(\cap_{k=1}^{N}\left\{\left|\frac{\mathcal{N}_k}{T(N)} - \mu(k)\right| \le \varepsilon k \mu(k)\right\}\right). \tag{2.34}$$

According to (2.18), (2.22) and (2.15) we have

$$\mathbf{P}\left\{\mathcal{B}_N\right\} = 1 - o(1) \tag{2.35}$$

as $N \to \infty$.

Recall that $\tau_N(x)$ denotes the set of the macro-vertices in the tree constructed according to the algorithm of revealing of connected component described above. Let $|\tau_N(x)|$ denotes the number of macro-vertices in $\tau_N(x)$. Then we easily derive

$$\mathbf{P}\left\{C_1\left(\widetilde{G}_N(\mathbf{X},q,c)\right) > N^{1/2}\right\} \le \mathbf{P}\left\{\max_{1\le i\le T(N)} |\tau_N(X_i)| > N^{1/2} \mid \mathcal{B}_N\right\} + o(1) \tag{2.36}$$

$$\le N\sum_{k=1}^{N}(1+\varepsilon k)\mu(k)\left(\delta + 1/\mathbf{E}X\right)\mathbf{P}\left\{|\tau_N(k)| > N^{1/2} \mid \mathcal{B}_N\right\} + o(1)$$

as $N \to \infty$. We shall use the multi-type branching process introduced above (Section 2.2) to approximate the distribution of $|\tau_N(k)|$. Let further $\mathcal{X}^{c,q}(k)$ denote the number of the particles (including the initial one) in the branching process starting with a single particle of type k. Observe that at each step of the exploration algorithm, the number of new neighbours of x of type y has a binomial distribution $Bin(N'_y, \widetilde{p}_{xy}(N))$ where N'_y is the number of remaining vertices of type y, so that $N'_y \le \mathcal{N}_y$.

We shall explore the following obvious relation between the Poisson and the binomial distributions. Let $Y_{n,p} \in Bin(n,p)$ and $Z_a \in Po(a)$, where $0 < p < 1/4$ and $a > 0$. Then for all $k \ge 0$

$$\mathbf{P}\{Y_{n,p} = k\} \le (1 + Cp^2)^n\,\mathbf{P}\{Z_{n\frac{p}{1-p}} = k\}, \tag{2.37}$$

where C is some positive constant (independent of n and p). Notice that for all $x, y \le \frac{2}{|\log q|}\log N$

$$\widetilde{p}_{xy}(N) = 1 - \left(1 - \frac{c}{N}\right)^{xy} = \frac{c}{N}\,xy\,(1 + o(1)), \tag{2.38}$$

and clearly, $\widetilde{p}_{xy}(N) \le 1/4$ for all large N. Therefore for any fixed positive ε_1 we can choose small ε and δ in (2.34) so that conditionally on $\mathcal{B}_N$ we have

$$N'_y \frac{\widetilde{p}_{xy}(N)}{1 - \widetilde{p}_{xy}(N)} \le (1 + y\varepsilon_1)\mu(y)\kappa(x,y) \tag{2.39}$$

for all large N. Let us write further

$$\mu(y) = \mu_q(y), \qquad \mu_q = \sum_{y\ge 1} y\mu_q(y)\ (= \mathbf{E}X), \qquad \kappa(x,y) = \kappa_{c,q}(x,y)$$

emphasizing dependence on q and c. Then for any $\varepsilon_2 > 0$ and any $q' > q$ such that $\frac{1-q}{q}\frac{q'}{1-q'} < 1+\varepsilon_2$ we can choose $\varepsilon_1 < \log(q'/q)$, and derive from (2.39)

$$N'_y \frac{\widetilde{p}_{xy}(N)}{1-\widetilde{p}_{xy}(N)} \leq (1+\varepsilon_2)\mu_{q'}(y)\kappa_{c,q}(x,y) = \mu_{q'}(y)\frac{(1+\varepsilon_2)\,c}{\mu_q}xy. \tag{2.40}$$

Setting now $c' := (1+\varepsilon_2)\frac{\mu_{q'}}{\mu_q}\,c$ we rewrite (2.40) as follows

$$N'_y \frac{\widetilde{p}_{xy}(N)}{1-\widetilde{p}_{xy}(N)} \leq \mu_{q'}(y)\kappa_{c',q'}(x,y). \tag{2.41}$$

Recall that above we fixed q and $c < c^{cr}(q)$, where $c^{cr}(q)$ is decreasing and continuous in q. Hence, we can choose $q' > q$ and $c' := (1+\varepsilon_2)\frac{\mu_{q'}}{\mu_q}\,c$ so that

$$c < c' < c^{cr}(q') < c^{cr}(q), \tag{2.42}$$

and moreover c' and q' can be chosen arbitrarily close to c and q, respectively.

Now conditionally on $\mathcal{B}_N$ we can replace according to (2.37) at each (of at most N) step of the exploration algorithm the $Bin(N'_y, \widetilde{p}_{xy}(N))$ variable with $Po(N'_y\frac{\widetilde{p}_{xy}(N)}{1-\widetilde{p}_{xy}(N)})$, and further replace the last variables with the stochastically larger ones $Po(\mu_{q'}(y)\kappa_{c',q'}(x,y))$ (recall (2.41)). As a result we get the following bound using branching process:

$$\mathbf{P}\left\{|\tau_N(k)| > N^{1/2} \mid \mathcal{B}_N\right\} \tag{2.43}$$

$$\leq \left(1 + C\left(\max_{x,y\leq 2\log N/|\log q|}\widetilde{p}_{xy}(N)\right)^2\right)^{N^2} \mathbf{P}\left\{\mathcal{X}^{c',q'}(k) > N^{1/2}\right\}.$$

This together with (2.38) implies

$$\mathbf{P}\left\{|\tau_N(k)| > N^{1/2} \mid \mathcal{B}_N\right\} \leq e^{b(\log N)^4}\,\mathbf{P}\left\{\mathcal{X}^{c',q'}(k) > N^{1/2}\right\}, \tag{2.44}$$

where b is some positive constant. Substituting the last bound into (2.36) we derive

$$\mathbf{P}\left\{C_1\Big(\widetilde{G}_N(\mathbf{X},q,c)\Big) > N^{1/2}\right\} \tag{2.45}$$

$$\leq b_2 N e^{b(\log N)^4}\sum_{k=1}^{N} k\mu_{q'}(k)\mathbf{P}\left\{\mathcal{X}^{c',q'}(k) > N^{1/2}\right\} + o(1)$$

as $N \to \infty$, where b_2 is some positive constant. By the Markov's inequality

$$\mathbf{P}\big\{\mathcal{X}^{c',q'}(k) > N^{1/2}\big\} \le z^{-N^{1/2}} \mathbf{E} z^{\mathcal{X}^{c',q'}(k)} \tag{2.46}$$

for all $z \ge 1$. Denote $h_z(k) = \mathbf{E} z^{\mathcal{X}^{c',q'}(k)}$; then with a help of (2.46) we get from (2.45)

$$\mathbf{P}\left\{C_1\Big(\widetilde{G}_N(\mathbf{X},q,c)\Big) > N^{1/2}\right\} \le b_1 N e^{b(\log N)^2} z^{-N^{1/2}} \sum_{k=1}^{N} k\mu_{q'}(k) h_z(k) + o(1). \tag{2.47}$$

Now we will show that there exists $z > 1$ such that the series

$$B_z(c', q') = \sum_{k=1}^{\infty} k\mu_{q'}(k) h_z(k)$$

converge. This together with (2.47) will clearly imply the statement of the lemma.

Note that function $h_z(k)$ (as a generating function for a branching process) satisfies the following equation

$$\begin{aligned} h_z(k) &= z \exp\left\{\textstyle\sum_{x=1}^{\infty} \kappa_{c',q'}(k,x)\mu_{q'}(x)(h_z(x)-1)\right\} \\ &= z \exp\left\{\tfrac{c'}{\mu_{q'}} k \left(\textstyle\sum_{x=1}^{\infty} x\mu_{q'}(x)h_z(x) - \mu_{q'}\right)\right\} \\ &= z \exp\left\{\tfrac{c'}{\mu_{q'}} k (B_z(c',q') - \mu_{q'})\right\}. \end{aligned}$$

Multiplying both sides by $k\mu_{q'}(k)$ and summing up over k we find

$$B_z(c', q') = \sum_{k=1}^{\infty} k\mu_{q'}(k) z \exp\left\{\frac{c'}{\mu_{q'}} k (B_z(c',q') - \mu_{q'})\right\}.$$

Let us write for simplicity $B_z = B_z(c, q)$. Hence, as long as B_z is finite, it should satisfy equation

$$B_z = z\mathbf{E}X\, e^{\frac{c}{\mathbf{E}X} X (B_z - \mathbf{E}X)}, \tag{2.48}$$

which implies in turn that B_z is finite for some $z > 1$ if and only if (2.48) has at least one solution (for the same value of z). Notice that

$$B_z \ge B_1 = \mathbf{E}X \tag{2.49}$$

for $z \geq 1$. Let us fix $z > 1$ and consider equation

$$y/z = \mathbf{E}X\, e^{\frac{c}{\mathbf{E}X}X(y-\mathbf{E}X)} =: F(y) \tag{2.50}$$

for $y \geq \mathbf{E}X$. Using the properties of the distribution of X it is easy to derive that function

$$F(y) = \frac{1}{\mathbf{E}X} \frac{e^{c(\frac{y}{\mathbf{E}X}-1)}}{\left(1 - qe^{c(\frac{y}{\mathbf{E}X}-1)}\right)^2}$$

is increasing and has positive second derivative if $\mathbf{E}X \leq y \leq y_0$, where y_0 is the root of $1 = q\exp\left\{c(\frac{y_0}{\mathbf{E}X} - 1)\right\}$. Compute now

$$\frac{\partial}{\partial y}F(y)|_{y=\mathbf{E}X} = \frac{c}{\mathbf{E}X}\mathbf{E}X^2 = \frac{c}{c^{cr}}. \tag{2.51}$$

Hence, if $c < c^{cr}$ then there exists $z > 1$ such that there is a finite solution y to (2.50). Taking into account condition (2.42), we find that $B_z(c', q')$ is also finite for some $z > 1$, which finishes the proof of the lemma. □

Now we are ready to complete the proof of (1.3), following almost the same arguments as in the proof of the previous lemma. Let $S_N(x) = \sum_{v_i \in \tau_N(x)} X_i$ denote the number of vertices from V_N which compose the macro-vertices of $\tau_N(x)$. Denote

$$\mathcal{B}'_N := \mathcal{B}_N \cap \left(\max_{1 \leq i \leq T(N)} |\tau_N(X_i)| < N^{1/2} \right).$$

According to (2.35) and Lemma 2.2 we have

$$\mathbf{P}\left\{\mathcal{B}'_N\right\} = 1 - o(1).$$

This allows us to derive from (2.32)

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) > aw\right\} \leq \mathbf{P}\left\{\max_{1 \leq i \leq T(N)} S_N(X_i) > aw \mid \mathcal{B}'_N\right\} + o(1) \tag{2.52}$$

$$\leq N \sum_{k=1}^{N} (1 + \varepsilon k)\mu(k)\Big(\delta + 1/\mathbf{E}X\Big)\mathbf{P}\left\{S_N(k) > aw \mid \mathcal{B}'_N\right\} + o(1).$$

Let now $S^{c,q}(y)$ denote the sum of types (including the one of the initial particle) in the total progeny of the introduced above branching process starting with initial particle of type

y. Repeating the same argument which led to (2.43), we get the following bound using the introduced branching process:

$$\mathbf{P}\left\{S_N(k) > aw \mid \mathcal{B}'_N\right\} \le \left(1 + C\left(\max_{x,y\le 2\log N/|\log q|} \widetilde{p}_{xy}(N)\right)^2\right)^{b_1 N\sqrt{N}} \mathbf{P}\left\{S^{c',q'}(k) > aw\right\}$$

as $N \to \infty$, where we take into account that we can perform at most $\sqrt{N}$ steps of exploration (the maximal possible number of macro-vertices in any $\widetilde{L}$). This together with (2.38) implies

$$\mathbf{P}\left\{\tau_N(k) > aw \mid \mathcal{B}'_N\right\} \le (1+o(1))\mathbf{P}\left\{S^{c',q'}(k) > aw\right\} \tag{2.53}$$

as $N \to \infty$. Substituting the last bound into (2.52) we derive

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) > aw\right\} \le Nb\sum_{k=1}^{N} k\mu_{q'}(k)\mathbf{P}\left\{S^{c',q'}(k) > aw\right\} + o(1) \tag{2.54}$$

as $N \to \infty$, where b is some positive constant. Denote $g_z(k) = \mathbf{E}z^{S^{c',q'}(k)}$; then similar to (2.47) we derive from (2.54)

$$\mathbf{P}\left\{C_1\Big(G_N(q,c)\Big) > aw(N)\right\} \le bN\sum_{k=1}^{N} k\mu_{q'}(k)g_z(k)z^{-aw(N)} + o(1). \tag{2.55}$$

We shall search for all $z \ge 1$ for which the series

$$A_z(c',q') = \sum_{k=1}^{\infty} k\mu_{q'}(k)g_z(k)$$

converge. Function $g_z(k)$ (as a generating function for a certain branching process) satisfies the following equation

$$\begin{aligned} g_z(k) &= z^k \exp\left\{\textstyle\sum_{x=1}^{\infty} \kappa_{c',q'}(k,x)\mu_{q'}(x)(g_z(x)-1)\right\} \\ &= z^k \exp\left\{\tfrac{c'}{\mu_{q'}}k\left(\textstyle\sum_{x=1}^{\infty} x\mu_{q'}(x)g_z(x) - \mu_{q'}\right)\right\} \\ &= z^k \exp\left\{\tfrac{c'}{\mu_{q'}}k(A_z(c',q') - \mu_{q'})\right\}. \end{aligned}$$

Multiplying both sides by $k\mu_{q'}(k)$ and summing up over k we find

$$A_z(c',q') = \sum_{k=1}^{\infty} k\mu_{q'}(k)z^k \exp\left\{\frac{c'}{\mu_{q'}}k(A_z(c',q') - \mu_{q'})\right\}. \tag{2.56}$$

It follows from here (and the fact that $A_z(c', q') \geq \mu_{q'}$ for all $z \geq 1$) that if there exists $z > 1$ for which the series $A_z(c', q')$ converge, it should satisfy

$$z < \frac{1}{q'}. \tag{2.57}$$

According to (2.56), as long as $A_z = A_z(c, q)$ is finite it satisfies the equation

$$A_z = \mathbf{E} X z^X \, e^{\frac{c}{\mathbf{E}X} X (A_z - \mathbf{E}X)},$$

which implies that A_z is finite for some $z > 1$ if and only if the last equation has at least one solution

$$A_z \geq A_1 = \mathbf{E}X. \tag{2.58}$$

Let us fix $z > 1$ and consider equation

$$y = \mathbf{E} X z^X \, e^{\frac{c}{\mathbf{E}X} X (y - \mathbf{E}X)} =: f(y, z). \tag{2.59}$$

Using the properties of the distribution of X it is easy to derive that

$$f(y, z) = \frac{1}{\mathbf{E}X} \frac{z e^{c(\frac{y}{\mathbf{E}X} - 1)}}{\left(1 - q z e^{c(\frac{y}{\mathbf{E}X} - 1)}\right)^2}.$$

We shall consider $f(y, z)$ for $y \geq \mathbf{E}X$ and $1 \leq z < \frac{1}{q} e^{-c(\frac{y}{\mathbf{E}X} - 1)}$. It is easy to check that in this area function $f(y, z)$ is increasing, it has all the derivatives of the second order, and $\frac{\partial^2}{\partial y^2} f(y, z) > 0$. Compute now

$$\frac{\partial}{\partial y} f(y, z)|_{y=1, z=1} = \frac{c}{\mathbf{E}X} \mathbf{E}X^2 = \frac{c}{c^{cr}}. \tag{2.60}$$

Hence, if $c > c^{cr}$ there is no solution $y \geq \mathbf{E}X$ to (2.59) for any $z > 1$. On the other hand, if $c < c^{cr}$ then there exists $1 < z_0 < 1/q$ such that for all $1 \leq z < z_0$ there is a finite solution $y \geq \mathbf{E}X$ to (2.59). One could find z_0 for example, as the (unique!) value for which function y is tangent to $f(y, z_0)$ if $y \geq \mathbf{E}X$.

Now taking into account that $c' > c$ and $q' > q$ can be chosen arbitrarily close to c and q, respectively, we derive from (2.55) that for all $1 < z < z_0$

$$\mathbf{P}\left\{ C_1\Big(G_N(q, c)\Big) > a w(N) \right\} \leq b(z) N z^{-a w(N)} + o(1) \tag{2.61}$$

as $N \to \infty$, where $b(z) < \infty$. This implies that for any $a > 1/\log z_0 > 1/|\log q|$

$$\mathbf{P}\left\{ C_1\Big(G_N(q, c)\Big) > a \log N \right\} = o(1) \tag{2.62}$$

as $N \to \infty$, which proves (1.3). □

To conclude this section we comment on the methods used here. It is shown in [6] that in the subcritical case of classical random graphs the same method of generating functions combined with the Markov inequality leads to a constant which is known to be the principal term for the asymptotics of the size of the largest component (scaled to $\log N$). This gives us hope that a constant a chosen here to satisfy $a > 1/\log z_0$ is close to the minimal constant for which statement (2.62) still holds.

Similar methods were used in [7] for some class of inhomogeneous random graphs, and in [1] for a general class of models. Note, however, some difference with the approach in [1]. It is assumed in [1], Section 12, that the generating function for the corresponding branching process with the initial state k (e.g., our function $g_z(k)$, $k \geq 1$) is bounded uniformly in k. As we prove here this condition is not always necessary: we need only convergence of the series A_z, while $g_z(k)$ is unbounded in k in our case. Furthermore, our approach allows one to construct constant $\alpha(q,c)$ as a function of the parameters of the model.

2.5 Proof of Theorem 1.1 in the supercritical case.

Let $\mathcal{C}_k$ denote the set of vertices in the k-th largest component in graph $G_N(q,c)$, and conditionally on $\mathbf{X}$ let $\widetilde{\mathcal{C}}_k$ denote the set of macro-vertices in the k-th largest component in graph $\widetilde{G}_N(\mathbf{X},q,c)$ (ordered in any way if there are ties). Let also C_k and $\widetilde{C}_k$ denote correspondingly, their sizes. According to our construction for any connected component $\widetilde{L}$ in $\widetilde{G}_N(\mathbf{X},q,c)$ there is a unique component L in $G_N(q,c)$ such that they are composed of the same vertices from V_N, i.e., in the notations (2.5)

$$L = \cup_{v\in\widetilde{L}} \cup_{k\in v} \{k\} =: V(\widetilde{L}).$$

Next we prove that with a high probability the largest components in both graphs consist of the same vertices.

Lemma 2.3. *For any $0 \leq q < 1$ if $c > c^{cr}(q)$ then*

$$\mathbf{P}\{\mathcal{C}_1 = V(\widetilde{\mathcal{C}}_1)\} = 1 - o(1) \tag{2.63}$$

as $N \to \infty$.

Proof. In a view of the argument preceeding this lemma we have

$$\mathbf{P}\{\mathcal{C}_1 \neq V(\widetilde{\mathcal{C}}_1)\} = \mathbf{P}\{\mathcal{C}_1 = V(\widetilde{\mathcal{C}}_k) \text{ for some } k \geq 2\}.$$

According to Theorem 12.6 from [1], conditions of which are satisfied here, in the supercritical case conditionally on $T(N)$ such that $T(N)/N \to 1/\mathbf{E}X$, we have **whp** $\widetilde{C}_2 = O(\log(T(N)))$, which by Proposition 2.1 implies $\widetilde{C}_2 = O(\log N)$ **whp** . Also we know already from (2.14) that in the supercritical case $\widetilde{C}_1 = O(N)$ **whp**, and therefore $C_1 = O(N)$ **whp**. Hence, for some positive constants a and b

$$\mathbf{P}\{\mathcal{C}_1 \neq V(\widetilde{\mathcal{C}}_1)\} = \mathbf{P}\{\mathcal{C}_1 = V(\widetilde{\mathcal{C}}_k) \text{ for some } k \geq 2\} \tag{2.64}$$

$$\leq \mathbf{P}\left\{\left(\max_{k\geq 2}|V(\widetilde{\mathcal{C}}_k)| > bN\right) \cap \left(\max_{k\geq 2}\widetilde{C}_k < a\log N\right)\right\} + o(1).$$

Define now for any $K = K(N)$ a set

$$B_N := \{\exists X_i \geq K \text{ for some } 1 \leq i \leq T(N)\}.$$

According to (2.20)

$$\mathbf{P}(B_N) \leq CNq^K + o(1)$$

as $N \to \infty$ for some constant C independent of K and N. Setting from now on $K = \sqrt{N}$ we have $\mathbf{P}(B_N) = o(1)$ as $N \to \infty$. Then we derive

$$\mathbf{P}\left\{\left(\max_{k\geq 2}|V(\widetilde{\mathcal{C}}_k)| > bN\right) \cap \left(\max_{k\geq 2}\widetilde{C}_k < a\log N\right)\right\} \tag{2.65}$$

$$\leq \mathbf{P}\left\{\left(\max_{k\geq 2}|V(\widetilde{\mathcal{C}}_k)| > bN\right) \cap \left(\max_{k\geq 2}\widetilde{C}_k < a\log N\right) \cap \left(\max_{1\leq i\leq T(N)} X_i < \sqrt{N}\right)\right\} + o(1)$$

$$\leq \mathbf{P}\left\{\sqrt{N}\,a\log N > bN\right\} + o(1) = o(1).$$

Substituting this bound into (2.64) we immediately get (2.63). □

Conditionally on $\mathcal{C}_1 = V(\widetilde{\mathcal{C}}_1)$ we have

$$\begin{aligned}\tfrac{C_1}{N} &= \tfrac{1}{N}\textstyle\sum_{i=1}^{T(N)} X_i\mathbf{1}\{v_i \in \widetilde{\mathcal{C}}_1\}\\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{T(N)}\sum_{k=1}^{N} k\mathbf{1}\{X_i = k\}\mathbf{1}\{v_i \in \widetilde{\mathcal{C}}_1\}\\ &= \tfrac{T(N)}{N}\textstyle\sum_{k=1}^{N} k\tfrac{1}{T(N)}\#\{v_i \in \widetilde{\mathcal{C}}_1 : X_i = k\}.\end{aligned} \tag{2.66}$$

Note that Theorem 9.10 from [1] (together with Proposition 2.1 in our case) implies that

$$\nu_N(k) := \frac{1}{T(N)}\#\{v_i \in \widetilde{\mathcal{C}}_1(N) : X_i = k\} \xrightarrow{P} \rho(\kappa; k)\mu(k) \tag{2.67}$$

for each $k \ge 1$.

We shall prove below that also

$$W_N := \sum_{k=1}^{N} k\nu_N(k) \xrightarrow{P} \sum_{k=1}^{\infty} k\rho(\kappa;k)\mu(k) =: \beta\,\mathbf{E}X. \tag{2.68}$$

Observe that $\rho(\kappa;k)$ is the maximal solution to (2.12), therefore β is the maximal solution to

$$\beta = \frac{1}{\mathbf{E}X}\sum_{k=1}^{\infty} k\rho(\kappa;k)\mu(k) = \frac{1}{\mathbf{E}X}\sum_{k=1}^{\infty} k\left(1-e^{-\sum_{y=1}^{\infty}\kappa(k,y)\mu(y)\rho(\kappa;y)}\right)\mu(k)$$
$$= 1-\frac{1}{\mathbf{E}X}\mathbf{E}\Big(Xe^{-cX\beta}\Big).$$

This proves that β is the maximal root of (1.6). Then (2.68) together with Proposition 2.1, which states that $T(N)/N \stackrel{a.s.}{\to} 1/\mathbf{E}X$, will allow us to derive from (2.66) that for any positive ε

$$\mathbf{P}\Big\{\Big|\frac{C_1\Big(G_N(c,q)\Big)}{N}-\beta\Big| > \varepsilon \mid \mathcal{C}_1 = V(\widetilde{\mathcal{C}}_1)\Big\} \to 0$$

as $N \to \infty$. This combined with Lemma 2.3 would immediately imply

$$\frac{C_1\Big(G_N(c,q)\Big)}{N} \xrightarrow{P} \beta, \tag{2.69}$$

and hence the statement of the theorem follows.

Now we are left with proving (2.68). For any $1 \le R < N$ write $W_N := W_N^R + w_N^R$, where

$$W_N^R := \sum_{k=1}^{R} k\nu_N(k), \quad w_N^R := \sum_{k=R+1}^{N} k\nu_N(k).$$

By (2.67) we have for any fixed $R \ge 1$

$$W_N^R \xrightarrow{P} \sum_{k=1}^{R} k\rho(\kappa;k)\mu(k) \tag{2.70}$$

as $N \to \infty$. Consider w_N^R. Note that for any $k \ge 1$

$$\mathbf{E}\nu_N(k) = \mathbf{E}\mathbf{E}\Big\{\nu_N(k) \mid T(N)\Big\} \le \mathbf{E}\frac{1}{T(N)}\sum_{i=1}^{T(N)} \mathbf{P}\{X_i = k \mid T(N)\}. \tag{2.71}$$

Using events $\mathcal{A}_{\delta,N}$ with bound (2.18) and Proposition 2.1 we obtain from (2.71) for any fixed $0 < \delta < 1/(2\mathbf{E}X)$

$$\mathbf{E}\nu_N(k) \le \mathbf{E}\frac{1}{T(N)} \sum_{i=1}^{T(N)} \mathbf{P}\{X_i = k \mid T(N)\}\mathbf{1}\{\mathcal{A}_{\delta,N}\} + \mathbf{P}\{\overline{\mathcal{A}_{\delta,N}}\}$$

$$\le \frac{(1+\delta\mathbf{E}X)}{(1-\delta\mathbf{E}X)}\, \mathbf{P}\{X_1 = k\}(1+o(1)) + \mathbf{P}\{\overline{\mathcal{A}_{\delta,N}}\}.$$

Bound (2.18) allows us to derive from here that

$$\mathbf{E}\nu_N(k) \le A_1(\mu(k) + e^{-a_1 N}) \tag{2.72}$$

for some positive constants A_1 and a_1 independent of k and N. This yields

$$\mathbf{E}w_N^R = \sum_{k=R+1}^{N} k\mathbf{E}\nu_N(k) \le A_2 e^{-a_2 R} \tag{2.73}$$

for some positive constants A_2 and a_2.

Now for any $\varepsilon > 0$ we can choose R so that

$$\sum_{k=R+1}^{\infty} k\rho(\kappa;k)\mu(k) < \varepsilon/3,$$

and then we have

$$\mathbf{P}\{|W_N - \sum_{k=1}^{\infty} k\rho(\kappa;k)\mu(k)| > \varepsilon\} \tag{2.74}$$

$$= \mathbf{P}\{|(W_N^R - \sum_{k=1}^{R} k\rho(\kappa;k)\mu(k)) + w_N^R - \sum_{k=R+1}^{\infty} k\rho(\kappa;k)\mu(k)| > \varepsilon\}$$

$$\le \mathbf{P}\{|W_N^R - \sum_{k=1}^{R} k\rho(\kappa;k)\mu(k)| > \varepsilon/3\} + \mathbf{P}\{w_N^R > \varepsilon/3\}.$$

Markov's inequality together with bound (2.73) gives us

$$\mathbf{P}\{w_N^R > \varepsilon/3\} \le \frac{3\mathbf{E}w_N^R}{\varepsilon} \le \frac{3A_2 e^{-a_2 R}}{\varepsilon}. \tag{2.75}$$

Making use of (2.75) and (2.70) we immediately derive from (2.74)

$$\mathbf{P}\{|W_N - \sum_{k=1}^{\infty} k\rho(\kappa;k)\mu(k)| > \varepsilon\} \leq o(1) + \frac{3A_2 e^{-a_2 R}}{\varepsilon} \tag{2.76}$$

as $N \to \infty$. Hence, for any given positive ε and ε_0 we can choose finite R so large that

$$\lim_{N\to\infty} \mathbf{P}\{|W_N - \sum_{k=1}^{\infty} k\rho(\kappa;k)\mu(k)| > \varepsilon\} < \varepsilon_0. \tag{2.77}$$

This clearly proves statement (2.68), and therefore finishes the proof of the theorem. □

References

[1] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, to appear. (arXiv:math.PR/0504589)

[2] F. Chung and L. Lu , The volume of the giant component of a random graph with given expected degree. SIAM *J. Discrete Math.* **20** (2006), 395-411

[3] G. Grimmett, Percolation. Springer-Verlag, Berlin, 1999.

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

[5] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence. Combin. Probab. Comput. 7 (1998), no. 3, 295–305.

[6] T.S. Turova, Note on the random graphs in the subcritical case. Dynamical systems from number theory to probability - 2, (ed. A.Yu. Khrennikov), Växjö University Press (2003), 187–192.

[7] T.S. Turova, Phase Transitions in Dynamical Random Graphs, Journal of Statistical Physics, 123 (2006), no. 5, 1007-1032.

[8] T. Turova and A. Villa, On a phase diagram for random neural networks with embedded spike timing dependent plasticity. To appear in *BioSystems*

[9] D.J. Watts and S.H. Strogatz, Collective dynamics of "small-world" networks. Nature, 393 (1998), 440–442.