

Critical Window for Connectivity in the Configuration Model

LORENZO FEDERICO and REMCO VAN DER HOFSTAD

Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513,
5600 MB Eindhoven, The Netherlands
(e-mail: l.federico@tue.nl, r.w.v.d.hofstad@tue.nl)

Received 18 July 2016; revised 10 March 2017; first published online 29 May 2017

We identify the asymptotic probability of a configuration model $\text{CM}_n(\mathbf{d})$ producing a *connected* graph within its critical window for connectivity that is identified by the number of vertices of degree 1 and 2, as well as the expected degree. In this window, the probability that the graph is connected converges to a non-trivial value, and the size of the complement of the giant component weakly converges to a finite random variable. Under a finite second moment condition we also derive the asymptotics of the connectivity probability conditioned on simplicity, from which follows the asymptotic number of simple connected graphs with a prescribed degree sequence.

2010 *Mathematics subject classification*: Primary 05C80, 05A16
Secondary 05C30, 05C40, 60C05

1. Introduction

In this paper we consider the configuration model $\text{CM}_n(\mathbf{d})$ on n vertices with a prescribed degree sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$. We identify the probability that $\text{CM}_n(\mathbf{d})$ is connected in terms of \mathbf{d} in the limit as $n \rightarrow \infty$. We further analyse the behaviour of the model in the critical window for connectivity (*i.e.* when the asymptotic probability of producing a connected graph is in the interval $(0, 1)$). Given a vertex $v \in [n] := \{1, 2, \dots, n\}$, we let d_v denote its degree. The configuration model is constructed by assigning d_v half-edges to each vertex v , after which the half-edges are paired randomly: first we pick two half-edges at random and create an edge out of them, then we pick two half-edges at random from the set of remaining half-edges and pair them into an edge, *etc.* We assume the total degree $\sum_{v \in [n]} d_v$ is even. The construction can give rise to self-loops and multiple edges between vertices, but these imperfections are relatively rare when n is large; see [3, 5, 6].

We define the random variable D_n as the degree of a vertex chosen uniformly at random from the vertex set $[n]$. We let \mathcal{N}_i denote the set of all vertices of degree i , and n_i its cardinality. The configuration model is known to have a phase transition for the existence

of a giant component with critical point at

$$v_n = \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} = 1$$

(see e.g. [10] or [7]). When $v_n \rightarrow v > 1$, there is a (unique) giant component \mathcal{C}_{\max} containing a positive proportion of the vertices, while when $v_n \rightarrow v \leq 1$, the maximal connected component contains a vanishing proportion of the vertices. Assuming that the second moment of D_n remains uniformly bounded, the subcritical behaviour was analysed by Janson in [4].

In this paper, we focus on the *connectivity transition* of the configuration model. Let us first describe the history of this problem. Wormald [12] showed that for $k \geq 3$ a random k -regular graph on n vertices is with high probability k -connected as $n \rightarrow \infty$ (see also [2]). Tomasz Łuczak [9] proved that also if the graph is not regular, but $d_v \geq k$ for every $v \in [n]$, then $CM_n(\mathbf{d})$ is with high probability k -connected, and found the asymptotic probability of having a connected graph when $d_v \geq 2$ and the graph is simple. Actually Łuczak’s model was defined in a different way from the configuration model and does not allow for vertices of degree 1, but the results could easily be adapted to the configuration model.

In this paper we extend Łuczak’s results to all possible kinds of weakly converging degree distributions, allowing vertices of degree 1 and removing any condition on maximum degree or finiteness of moments of the degree distribution. This implies that our results include heavy-tailed degree distribution that has recently received considerable attention in research on the configuration model. Moreover, using the multivariate method of moments, we identify the limiting distribution of the size of the *complement* of the maximal connected component $[n] \setminus \mathcal{C}_{\max}$ as a mixture of Poisson variables.

We start by introducing some notation.

Notation. All limits in this paper are taken as n tends to infinity unless stated otherwise. A sequence of events $(\mathcal{A}_n)_{n \geq 1}$ happens *with high probability* (w.h.p.) if $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$. For random variables $(X_n)_{n \geq 1}, X$, we write $X_n \xrightarrow{d} X$ and $X_n \xrightarrow{\mathbb{P}} X$ to denote convergence in distribution and in probability, respectively. For real-valued sequences $(a_n)_{n \geq 1}, (b_n)_{n \geq 1}$, we write $a_n = O(b_n)$ if the sequence $(a_n/b_n)_{n \geq 1}$ is bounded, $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$, $a_n = \Theta(b_n)$ if the sequences $(a_n/b_n)_{n \geq 1}$ and $(b_n/a_n)_{n \geq 1}$ are both bounded. Similarly, for sequences $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$ of random variables, we write $X_n = O_{\mathbb{P}}(Y_n)$ if the sequence $(X_n/Y_n)_{n \geq 1}$ is tight, and $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \xrightarrow{\mathbb{P}} 0$. Moreover, $\text{Poi}(\lambda)$ always denotes a Poisson-distributed random variable with mean λ and $\text{Bin}(n, p)$ denotes a random variable with binomial distribution with parameters n and p .

2. Main results

We start by defining the conditions for $CM_n(\mathbf{d})$ to be in the connectivity critical window. We define the random variable D_n as the degree of a vertex chosen uniformly at random in $[n]$. We state the conditions we assume to hold throughout this paper.

Condition 2.1 (critical window for connectivity). We define a sequence $\text{CM}_n(\mathbf{d})$ to be in the critical window for connectivity when the following conditions are satisfied.

- (1) There exists a limiting degree variable D such that $D_n \xrightarrow{d} D$.
- (2) $n_0 = 0$.
- (3) $\lim_{n \rightarrow \infty} n_1 / \sqrt{n} = \rho_1 \in [0, \infty)$.
- (4) $\lim_{n \rightarrow \infty} n_2 / n = p_2 \in [0, 1)$.
- (5) $\lim_{n \rightarrow \infty} \mathbb{E}[D_n] = d < \infty$.

Notice that we use different symbols for ρ_1 and p_2 to stress the fact that p_2 is actually the limit probability that a uniformly chosen vertex has degree 2 while ρ_1 is obtained through a rescaling (in the critical window for connectivity $p_1 = \lim_{n \rightarrow \infty} n_1 / n = 0$). Under these conditions, we prove our main theorem. In its statement, we write \mathcal{C}_{\max} for the maximal connected component in $\text{CM}_n(\mathbf{d})$.

Theorem 2.2 (connectivity threshold for the configuration model). Consider $\text{CM}_n(\mathbf{d})$ in the critical window for connectivity as described in Condition 2.1. Then

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) = \left(\frac{d - 2p_2}{d} \right)^{1/2} \exp \left\{ - \frac{\rho_1^2}{2(d - 2p_2)} \right\}. \tag{2.1}$$

Moreover,

$$n - |\mathcal{C}_{\max}| \xrightarrow{d} X, \tag{2.2}$$

where $X = \sum_k k(C_k + L_k)$, and $(C_k, L_k)_{k \geq 1}$ are independent random variables such that

$$L_k \stackrel{d}{=} \text{Poi} \left(\frac{\rho_1^2 (2p_2)^{k-2}}{2d^{k-1}} \right), \quad C_k \stackrel{d}{=} \text{Poi} \left(\frac{(2p_2)^k}{2kd^k} \right).$$

Finally,

$$\lim_{n \rightarrow \infty} \mathbb{E}[n - |\mathcal{C}_{\max}|] = \frac{\rho_1^2 (2d - p_2)}{2(d - p_2)^2} + \frac{p_2}{d - 2p_2}. \tag{2.3}$$

The convergence in distribution of $n - |\mathcal{C}_{\max}|$ to a proper random variable with finite mean is a stronger result than proved by Łuczak [9], who instead proved that

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} 1. \tag{2.4}$$

Our improvement is achieved by an application of the multivariate method of moments, as well as a careful estimate of the probability that there exists $v \in [n]$ that is not part of $|\mathcal{C}_{\max}|$. We next investigate the boundary cases.

Remark 1 (boundary cases). Our proof also applies to the boundary cases where $\rho_1 = \infty, p_2 = 0$ or $d = \infty$. When $d < \infty$, we obtain

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) \rightarrow \begin{cases} 0 & \text{when } \rho_1 = \infty, \\ 1 & \text{when } \rho_1, p_2 = 0. \end{cases} \tag{2.5}$$

When $d = \infty$, instead

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) = \lim_{n \rightarrow \infty} \exp\left\{-\frac{n_1^2}{2\ell_n}\right\}, \tag{2.6}$$

where $\ell_n = \sum_{i \in [n]} d_i$ denotes the total degree.

Assuming also that D has finite second moment, the configuration model is simple with non-vanishing probability. Under this extra assumption, the next theorem states how many connected graphs there are with prescribed degrees in the connectivity window defined in Condition 2.1.

Theorem 2.3 (connectivity conditioned on simplicity and connected simple graphs). *Consider $\text{CM}_n(\mathbf{d})$ in the connectivity critical window defined in Condition 2.1. If*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} = v < \infty,$$

then

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ &= \left(\frac{d - 2p_2}{d}\right)^{1/2} \exp\left(-\frac{\rho_1^2}{2(d - 2p_2)} + \frac{p_2^2 + dp_2}{d^2}\right). \end{aligned} \tag{2.7}$$

Let $\mathcal{N}_n^C(\mathbf{d})$ be the number of connected simple graphs with degree distribution \mathbf{d} . Then

$$\begin{aligned} \mathcal{N}_n^C(\mathbf{d}) &= \frac{(\ell_n - 1)!!}{\prod_{i \in [n]} d_i!} \left(\frac{d - 2p_2}{d}\right)^{1/2} \\ &\times \exp\left\{-\frac{v}{2} - \frac{v^2}{4} - \frac{\rho_1^2}{2(d - 2p_2)} + \frac{p_2^2 + dp_2}{d^2}\right\} (1 + o(1)). \end{aligned} \tag{2.8}$$

It is instructive to compare the asymptotics of the number of connected simple graphs with degree distribution \mathbf{d} to the number $\mathcal{N}_n(\mathbf{d})$ of simple graphs with a given degree sequence \mathbf{d} , as identified by Janson [5]:

$$\mathcal{N}_n(\mathbf{d}) = \frac{(\ell_n - 1)!!}{\prod_{i \in [n]} d_i!} \exp\left\{-\frac{v}{2} - \frac{v^2}{4}\right\} (1 + o(1)). \tag{2.9}$$

With these results, the connectivity critical window is fully explored. Indeed, we have determined the asymptotic probability that the configuration model will produce a connected graph for all possible choices of the limiting degree distribution. What remains is to find the asymptotic of the number of connected simple graphs with degree distribution \mathbf{d} when it is below the connectivity critical window (*i.e.* when $n_1 \gg n^{1/2}$). In this case we should analyse how fast the probability of producing a connected graph vanishes, which is a hard problem.

To address the last boundary case, not considered in Remark 1, it is also worth noticing that the size of the largest component is very sensitive to the precise way in which $n_2/n \rightarrow 1$ (recall that we assume that $p_2 < 1$ in Condition 2.1), as we describe now. We

define $\mathcal{C}(v)$ to be the connected component of a uniformly chosen vertex. When $n_2 = n$, it is not hard to see that

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{d} S, \quad \frac{|\mathcal{C}(v)|}{n} \xrightarrow{d} T, \tag{2.10}$$

where S, T are proper random variables that satisfy the relation $S \stackrel{d}{=} T \vee [(1 - T)S]$. Instead, $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ when $n_2 = n - n_1$, with $n_1 \rightarrow \infty$, while $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 1$ when $n_2 = n - n_4$, with $n_4 \rightarrow \infty$. The latter two statements can be proved by relating it to the case where $n_2 = n$. Indeed, for the case where $n_1 > 0$, we take $n'_2 = n_2 + n_1/2$, and produce $\text{CM}_n(\mathbf{d})$ from the configuration model with n_2 vertices of degree 2 by ‘splitting’ $n_1/2$ vertices of degree 2 into two vertices of degree 1. For the case where $n_4 > 0$, we take $n'_2 = n_2 + 2n_4$, and produce $\text{CM}_n(\mathbf{d})$ from the configuration model with n_2 vertices of degree 2 by ‘merging’ $2n_4$ vertices of degree 2 into a vertex of degree 4. This explains why it is important to us to assume that $p_2 < 1$ in Condition 2.1.

2.1. Outline of the proof

We first notice that in the connectivity critical window our configuration model is supercritical, that is, w.h.p. it has a unique component of linear size with respect to the whole graph. In more detail, for finite $\rho_1 < \infty$ and $p_2 < 1$,

$$\lim_{n \rightarrow \infty} v_n = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} \geq \frac{2p_2 + 6(1 - p_2)}{2p_2 + 3(1 - p_2)} > 1. \tag{2.11}$$

Thus the results from [7, 11] imply that $|\mathcal{C}_{\max}| = \Theta_{\mathbb{P}}(n)$, while the second largest connected component $\mathcal{C}_{(2)}$ satisfies $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n)$ and $|E(\mathcal{C}_{(2)})| = o_{\mathbb{P}}(n)$, where $|E(G)|$ indicates the number of edges of the graph G . The proof of our main theorem is now divided into two parts:

- (1) to identify the limit distribution of the number of lines and cycles that form $[n] \setminus \mathcal{C}_{\max}$, which we do in Section 3,
- (2) to prove that w.h.p. all vertices $v \in [n]$ with $d_v \geq 3$ are in the giant component \mathcal{C}_{\max} , which we do in Section 4.

The proofs of our main theorems are then completed in Section 5.

3. Poisson convergence of the number of lines and cycles

In this section we prove that the number of cycles (components made by k vertices of degree 2) and lines (components made by two vertices of degree 1 and $k - 2$ vertices of degree 2) jointly converge to independent Poisson random variables. In Section 4, we will show that $[n] \setminus \mathcal{C}_{\max}$ w.h.p. only contains vertices of degree 1 and 2, so that all the other components are either cycles or lines. We define the sequences of random variables $(\mathbf{C}_n, \mathbf{L}_n) = (C_k(n), L_k(n))_{k \geq 1}$ as

- ▷ $C_k(n) = \# \{ \text{cycles of length } k \text{ in } \text{CM}_n(\mathbf{d}) \},$
- ▷ $L_k(n) = \# \{ \text{lines of length } k \text{ in } \text{CM}_n(\mathbf{d}) \}.$

We consider a vertex of degree 2 with a self-loop as a cycle of length 1. By convention, $L_1(n) = 0$ for all n since a vertex of degree 1 cannot have a self-loop.

We define $\mathcal{C}_k = \{\{v_1, v_2, \dots, v_k\} \subseteq \mathcal{N}_2\}$ to be the set of all collections of k vertices that could form a cycle, and denote

$$C_k(n) = \sum_{c \in \mathcal{C}_k} \mathbb{1}_{\{c \text{ forms a cycle}\}}, \tag{3.1}$$

where $\mathbb{1}_A$ denotes the indicator of the event A . In a similar way we define

$$\mathcal{L}_k = \{\{v_1, v_2, \dots, v_k\} : v_1, v_k \in \mathcal{N}_1; v_2, \dots, v_{k-1} \in \mathcal{N}_2\}$$

to be the set of all collections of k vertices that could form a line, and denote

$$L_k(n) = \sum_{l \in \mathcal{L}_k} \mathbb{1}_{\{l \text{ forms a line}\}}. \tag{3.2}$$

We will use the multivariate method of moments to show that $((C_n(k), L_k(n)))_{k \geq 1}$ converges to a vector of independent Poisson random variables. For a random variable X , we define $(X)_r = X(X - 1) \cdots (X - r + 1)$. For the multivariate method of moments, we recall two useful lemmas, whose proofs are given in [3, Section 2.1].

Lemma 3.1 (multivariate moment method with Poisson limit). *A sequence of vectors of non-negative integer-valued random variables $(X_1^{(n)}, X_2^{(n)}, \dots, X_k^{(n)})_{n \geq 1}$ converges in distribution to a vector of independent Poisson random variables with parameters $(\lambda_1, \lambda_2, \dots, \lambda_k)$ when, for all possible choices of $(r_1, r_2, \dots, r_k) \in \mathbb{N}^k$,*

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_1^{(n)})_{r_1} (X_2^{(n)})_{r_2} \cdots (X_k^{(n)})_{r_k}] = \lambda_1^{r_1} \lambda_2^{r_2} \cdots \lambda_k^{r_k}. \tag{3.3}$$

Lemma 3.2 (factorial moments of sums of indicators). *When $X_j = \sum_{i \in \mathcal{J}_j} I_{i^{(j)}}$ for all $j = 1, \dots, k$,*

$$\mathbb{E}[(X_1^{(n)})_{r_1} (X_2^{(n)})_{r_2} \cdots (X_k^{(n)})_{r_k}] = \sum_{i_1^{(1)}, \dots, i_{r_1}^{(1)} \in \mathcal{J}_1}^* \cdots \sum_{i_1^{(k)}, \dots, i_{r_k}^{(k)} \in \mathcal{J}_k}^* \mathbb{E} \left[\prod_{j=1}^k \prod_{s=1}^{r_k} I_{i_s^{(j)}} \right], \tag{3.4}$$

where \sum^* denotes a sum over distinct indices.

See also [8, Chapter 6] for more general versions of the method of moments. We can now state the main result of this section.

Theorem 3.3 (Poisson convergence of number of lines and cycles). *Consider $\text{CM}_n(\mathbf{d})$ in the critical window for connectivity defined in Condition 2.1. Then*

$$(\mathbf{C}_n, \mathbf{L}_n) \xrightarrow{d} (\mathbf{C}, \mathbf{L}), \tag{3.5}$$

where $(\mathbf{C}, \mathbf{L}) = (C_k, L_k)_{k \geq 1}$ is a sequence of independent random variables with

$$L_k \stackrel{d}{=} \text{Poi} \left(\frac{\rho_1^2 (2p_2)^{k-2}}{2d^{k-1}} \right), \quad C_k \stackrel{d}{=} \text{Poi} \left(\frac{(2p_2)^k}{2kd^k} \right), \tag{3.6}$$

and the convergence in (3.5) is in the product topology on \mathbb{N}^∞ .

Proof. We want to find the combined factorial moments of $(L_j(n), C_j(n))_{j \leq k}$ and show that

$$\begin{aligned} &\mathbb{E}[(C_1(n))_{r_1} (L_2(n))_{s_2} \cdots (C_k(n))_{r_k} (L_k(n))_{s_k}] \\ &\rightarrow \prod_{j=2}^k \left(\frac{\rho_1^2 (2p_2)^{j-2}}{d^{j-1}} \right)^{s_j} \prod_{j=1}^k \left(\frac{(2p_2)^j}{2kd^j} \right)^{r_j}. \end{aligned} \tag{3.7}$$

We argue by induction on k . When $k = 0$, both sides in (3.7) are equal to 1, which initializes the induction hypothesis.

We next argue how to advance the induction hypothesis. We define

$$w_{k,j}(\mathbf{r}, \mathbf{s}) = \{c_i(1), \dots, c_i(r_i) \in \mathcal{C}_i, 1 \leq i \leq k; l_i(1), \dots, l_i(s_i) \in \mathcal{L}_i, 2 \leq i \leq j\},$$

where all $c_i(1), \dots, c_i(r_i)$ and $l_i(1), \dots, l_i(s_i)$ are ordered lists without repetitions. Further, we let $\mathcal{E}(w_{k,j}(\mathbf{r}, \mathbf{s}))$ denote the event that all $c_i(h) \in w_{k,j}(\mathbf{r}, \mathbf{s})$ form a cycle and all $l_i(h) \in w_{k,j}(\mathbf{r}, \mathbf{s})$ form a line. By Lemma 3.2,

$$\mathbb{E}[(C_1(n))_{r_1} (L_2(n))_{s_2} \cdots (C_k(n))_{r_k} (L_k(n))_{s_k}] = \sum_{w_{k,k}(\mathbf{r}, \mathbf{s})} \mathbb{P}(\mathcal{E}(w_{k,k}(\mathbf{r}, \mathbf{s}))). \tag{3.8}$$

We rewrite this as

$$\sum_{w_{k,k-1}(\mathbf{r}, \mathbf{s})} \mathbb{P}(\mathcal{E}(w_{k,k-1}(\mathbf{r}, \mathbf{s}))) \sum_{l_1, \dots, l_{s_k} \in \mathcal{L}_k}^* \mathbb{E}[I_{l_1} I_{l_2} \cdots I_{l_{s_k}} \mid \mathcal{E}(w_{k,k-1}(\mathbf{r}, \mathbf{s}))], \tag{3.9}$$

where I_{l_s} is the indicator that the vertices in c_{l_s} form a line.

We let a_1 and a_2 denote the number of vertices of degree 1 and 2 necessary to create the cycles and lines prescribed by $w_{k,k-1}(\mathbf{r}, \mathbf{s})$, and let $a_e = a_1 + 2a_2$ denote the number of half-edges they have. The values of a_1, a_2 are completely independent from the exact choice of $w_{k,k-1}(\mathbf{r}, \mathbf{s})$ as long as all sets are disjoint (otherwise the event $\mathcal{E}(w_{k,k-1}(\mathbf{r}, \mathbf{s}))$ is impossible). The number of possible choices of s_k different disjoint $l \in \mathcal{L}_k$ without using the vertices allocated for $w_{k,k-1}(\mathbf{r}, \mathbf{s})$ are

$$\begin{aligned} &\frac{(n_1 - a_1)!}{2^{s_k} (n_1 - a_1 - 2s_k)!} \frac{(n_2 - a_2)!}{(k-2)^{s_k} (n_2 - a_2 - (k-2)s_k)!} (1 + o(1)) \\ &= \frac{n_1^{2s_k}}{2^{s_k}} \frac{n_2^{(k-2)s_k}}{(k-2)^{s_k}} (1 + o(1)), \end{aligned} \tag{3.10}$$

provided that $n_1, n_2 \rightarrow \infty$. The probability that the first forms a line is

$$\begin{aligned} &\frac{2k-4}{\ell_n - a_e - 1} \frac{2k-6}{\ell_n - a_e - 3} \cdots \frac{2}{\ell_n - a_e - 2k + 5} \frac{1}{\ell_n - a_e - 2k + 3} \\ &= \frac{(2k-4)!!}{\ell_n^{k-1}} (1 + o(1)). \end{aligned} \tag{3.11}$$

For all the other lines we just have to subtract from $\ell_n - a_e$ the $2k - 2$ half-edges that we have used for each of the previous ones, so that

$$\mathbb{E}[I_{l_1} I_{l_2} \cdots I_{l_{s_k}} \mid \mathcal{E}(w_{k,k-1}(\mathbf{r}, \mathbf{s}))] = \frac{(2k-4)!!^{s_k}}{\ell_n^{s_k(k-1)}} (1 + o(1)). \tag{3.12}$$

Finally we obtain

$$\begin{aligned} & \sum_{I_1, \dots, I_{s_k} \in \mathcal{L}_k}^* \mathbb{E}[I_{i_1} I_{i_2} \cdots I_{i_{s_k}} | \mathcal{E}(w_{k,k-1}(\mathbf{r}, \mathbf{s}))] \\ &= \frac{(\rho_1^2 n)^{s_k}}{2^{s_k}} \frac{(p_2 n)^{(k-2)s_k}}{(k-2)!^{s_k}} \frac{(2k-4)!^{s_k}}{\ell_n^{s_k(k-1)}} (1 + o(1)) \\ &= \left(\frac{\rho_1^2 (2n_2)^{k-2}}{2d^{k-1}} \right)^{s_k} (1 + o(1)). \end{aligned} \tag{3.13}$$

We do the same for the cycles $C_k(n)$, writing

$$\sum_{w_{k-1,k-1}(\mathbf{r}, \mathbf{s})} \mathbb{P}(\mathcal{E}(w_{k-1,k-1}(\mathbf{r}, \mathbf{s}))) \sum_{c_1, \dots, c_{r_k} \in \mathcal{C}_k}^* \mathbb{E}[I_{i_1} \cdots I_{i_{r_k}} | \mathcal{E}(w_{k-1,k-1}(\mathbf{r}, \mathbf{s}))]. \tag{3.14}$$

The number of possible choices of r_k different disjoint $c \in \mathcal{C}_k$ without using the vertices allocated for $w_{k-1,k-1}(\mathbf{r}, \mathbf{s})$ are

$$\frac{(n_2 - a_2)!}{k!^{r_k} (n_2 - a_2 - kr_k)!} (1 + o(1)) = \frac{(n_2)^{kr_k}}{k!^{r_k}} (1 + o(1)), \tag{3.15}$$

provided that $n_2 \rightarrow \infty$.

The probability that the first forms a cycle is

$$\frac{2k-2}{\ell_n - a_e - 3} \frac{2k-4}{\ell_n - a_e - 5} \cdots \frac{2}{\ell_n - a_e - 2k + 3} \frac{1}{\ell_n - a_e - 2k + 1} (1 + o(1)). \tag{3.16}$$

Again, for all the other cycles we just have to subtract the $2k$ half-edges that we have used for the previous ones so that

$$\mathbb{E}[I_{i_1} I_{i_2} \cdots I_{i_{r_k}} | \mathcal{E}(w_{k-1,k-1}(\mathbf{r}, \mathbf{s}))] = \frac{(2k-2)!^{r_k}}{\ell_n^{r_k k}} (1 + o(1)). \tag{3.17}$$

Thus, we obtain

$$\begin{aligned} & \sum_{c_1, \dots, c_{r_k} \in \mathcal{C}_k}^* \mathbb{E}[I_{i_1} I_{i_2} \cdots I_{i_{r_k}} | \mathcal{E}(w_{k-1,k-1}(\mathbf{r}, \mathbf{s}))] \\ &= \frac{n_2^{kr_k}}{k!^{r_k}} \frac{(2k-2)!^{r_k}}{(\ell_n)^{r_k k}} (1 + o(1)) = \left(\frac{(2p_2)^k}{2kd^k} \right)^{r_k} (1 + o(1)). \end{aligned} \tag{3.18}$$

This advances the induction hypothesis. We now use induction to show that (3.7) holds for every $k \geq 0$, and consequently prove the claim through the method of moments in Lemma 3.1.

In (3.10) and (3.15), we have assumed that $n_1, n_2 \rightarrow \infty$. When this is not satisfied we can just use a first moment method to show that $L_k(n) \xrightarrow{\mathbb{P}} 0$ for all k if $n_1 = O(1)$, and $C_k(n) \xrightarrow{\mathbb{P}} 0$ for all k and $L_k(n) \xrightarrow{\mathbb{P}} 0$ for $k \geq 3$ if $n_2 = O(1)$. □

We next show that, for a finite second moment and under the condition

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} \rightarrow v < \infty,$$

the asymptotic distribution of the number of self-loops and multiple edges is independent of $(C_k)_{k \geq 3}$ and $(L_k)_{k \geq 2}$.

We first notice that connectivity and simplicity are *not* independent, since self-loops and multiple edges among vertices of degree 2 make the graph simultaneously disconnected and not simple, so for $CM_n(\mathbf{d})$ to be simple, we have to require $C_1(n) = C_2(n) = 0$.

We define the number of self-loops and multiple edges in $CM_n(\mathbf{d})$ by $S(n), M(n)$ and show the following joint convergence.

Theorem 3.4 (Poisson convergence of number self-loops and multiple edges). *Consider $CM_n(\mathbf{d})$ in the critical window for connectivity defined in Condition 2.1, and let*

$$v_n = \mathbb{E}[D_n(D_n - 1)] / \mathbb{E}[D_n] \rightarrow v \leq \infty.$$

Then

$$((L_k(n))_{k \geq 2}, (C_k(n))_{k \geq 3}, S(n), M(n)) \xrightarrow{d} ((L_k)_{k \geq 2}, (C_k)_{k \geq 3}, S, M), \tag{3.19}$$

with $((L_k)_{k \geq 2}, (C_k)_{k \geq 3}, S, M)$ independent Poisson random variables with

$$\begin{aligned} L_k &\stackrel{d}{=} \text{Poi}\left(\frac{\rho_1^2 (2p_2)^{k-2}}{2d^{k-1}}\right), & C_k &\stackrel{d}{=} \text{Poi}\left(\frac{(2p_2)^k}{2kd^k}\right), \\ S &\stackrel{d}{=} \text{Poi}\left(\frac{v}{2}\right), & M &\stackrel{d}{=} \text{Poi}\left(\frac{v^2}{4}\right). \end{aligned} \tag{3.20}$$

Proof. We again use the multivariate method of moments in Lemma 3.1. We aim to find the combined factorial moments of

$$((L_j(n))_{2 \leq j \leq k}, (C_j(n))_{3 \leq j \leq k}, S(n), M(n)),$$

and show that

$$\begin{aligned} &\mathbb{E}[(L_2(n))_{s_2} (C_3(n))_{r_3} \cdots (C_k(n))_{r_k} (L_k(n))_{s_k} (S(n))_t (M(n))_u] \\ &\rightarrow \left(\frac{v}{2}\right)^{t+2u} \prod_{j=2}^k \left(\frac{\rho_1^2 (2p_2)^{j-2}}{2d^{j-1}}\right)^{s_j} \prod_{j=1}^k \left(\frac{(2p_2)^j}{2kd^j}\right)^{r_j}. \end{aligned} \tag{3.21}$$

We now define

$$w'_{k,j}(\mathbf{r}, \mathbf{s}) = \{c_i(1), \dots, c_i(r_i) \in \mathcal{C}_i, 3 \leq i \leq k; l_i(1), \dots, l_i(s_i) \in \mathcal{L}_i, 2 \leq i \leq j\}$$

as the choice of subsets that can form such lines and cycles, and by Lemma 3.1,

$$\sum_{w'_{k,k}(\mathbf{r}, \mathbf{s})} \mathbb{P}(\mathcal{E}(w'_{k,k}(\mathbf{r}, \mathbf{s}))) \mathbb{E}[(S(n))_t (M(n))_u \mid \mathcal{E}(w'_{k,k}(\mathbf{r}, \mathbf{s}))]. \tag{3.22}$$

Conditionally on $\mathcal{E}(w'_{k,k}(\mathbf{r}, \mathbf{s}))$, the random vector $(S(n), M(n))$ has the same law as the number of self-loops and multiple edges in a configuration model with degree sequence \mathbf{d}' , which is obtained from \mathbf{d} by removing the vertices appearing in $w'_{k,k}(\mathbf{r}, \mathbf{s})$. We notice that \mathbf{d}' is independent from the exact choice of $w'_{k,k}(\mathbf{r}, \mathbf{s})$. Thus, when D'_n denotes the degree of

a uniform random vertex selected from \mathbf{d}' and

$$v' = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[D'_n(D'_n - 1)]}{\mathbb{E}[D'_n]},$$

we obtain

$$\mathbb{E}[(S(n))_t(M(n))_u] \rightarrow \left(\frac{v'}{2}\right)^{t+2u}$$

(see e.g. [5, 6]). Since we are removing only a finite number of vertices from \mathbf{d} , we have that $v' = v$, and we thus obtain

$$\left(\frac{v}{2}\right)^{t+2u} \sum_{w'_{k,k}(\mathbf{r}, \mathbf{s})} \mathbb{P}(\mathcal{E}(w'_{k,k}(\mathbf{r}, \mathbf{s}))). \tag{3.23}$$

We finally obtain (3.21) using the same induction argument used to prove (3.7), which completes the proof. □

4. Connectivity among vertices of degree at least three

In this section we show that in the connectivity critical window w.h.p. all vertices v with $d_v \geq 3$ are in the giant component. This result is already known when $\min_{i \in [n]} d_i \geq 2$; we show that it still holds even in the presence of a sufficiently small amount of vertices of degree 1, and state it in the following theorem.

Theorem 4.1 (connectivity among vertices with $d_v \geq 3$). *Consider $\text{CM}_n(\mathbf{d})$ in the connectivity critical window defined in Condition 2.1. Then*

$$\mathbb{E}[\#\{v \in [n] : d_v \geq 3, |\mathcal{C}(v)| < n/2\}] \rightarrow 0. \tag{4.1}$$

Consequently,

$$\mathbb{E}[\#\{v \in [n] \setminus \mathcal{C}_{\max} : d_v \geq 3\}] \rightarrow 0. \tag{4.2}$$

We will use the usual exploration process of the configuration model, as we describe now. At each time t , we define the sets of half-edges $\{\mathcal{A}_t, \mathcal{D}_t, \mathcal{N}_t\}$ (the active, dead and neutral sets), and explore them in the following way.

Initialize We pick a vertex $v \in [n]$ uniformly at random with $d_v \geq 3$ and we set all its half-edges as active. All other half-edges are set as neutral.

Step At each step t , we pick a half-edge $e_1(t)$ in \mathcal{A}_t uniformly at random, and we pair it with another half-edge $e_2(t)$ chosen uniformly at random in $\mathcal{A}_t \cup \mathcal{N}_t$. We set $e_1(t), e_2(t)$ as dead.

If $e_2(t) \in \mathcal{N}_t$, then we find the vertex $v(e_2(t))$ incident to $e_2(t)$ and activate all its other half-edges.

As usual, the above exploration forms the graph at the same time as it explores the neighbourhood of the vertex v . A convenient way to encode the randomness in the exploration algorithm is to first choose a permutation ξ of the half-edges, chosen uniformly

at random from the set of all permutations of the half-edges. Then we run the exploration choosing as $e_1(t)$ and $e_2(t)$ always the first feasible half-edges in the permutation according to the exploration rules. This means that we take the first available active half-edge as $e_1(t)$, pair it to the first available active or neutral half-edge as $e_2(t)$ to create an edge consisting of $e_1(t)$ and $e_2(t)$, and then to update the status of all the half-edges as above.

The above description, which we will rely on for the remainder of this document, offers the possibility to analyse some properties of the exploration before running it, and it will be useful in proving that w.h.p. we will not run out of high-degree vertices too early in the exploration.

We define the process $S_t^{(v)} = |\mathcal{A}_t|$. The update rules of $S_t^{(v)}$ are

$$S_0^{(v)} = d_v, \quad S_{t+1}^{(v)} - S_t^{(v)} = \begin{cases} d_{v(e_2(t))} - 2 & \text{if } e_2(t) \in \mathcal{N}_t, \\ -2 & \text{if } e_2(t) \in \mathcal{A}_t. \end{cases} \tag{4.3}$$

We define T_0 as the smallest t such that $X_t = 0$ and

$$T_{1/2} = \max\{t : |\mathcal{N}_t| > n/2\}. \tag{4.4}$$

By definition of the exploration process, if $T_0 \geq T_{1/2}$ then $|\mathcal{C}(v)| \geq n/2$ (and, in particular, $v \in \mathcal{C}_{\max}$), so that proving the following proposition is sufficient to prove Theorem 4.1.

Proposition 4.2 (no hit of zero of exploration). *Consider $\text{CM}_n(\mathbf{d})$ in the critical window for connectivity defined in Condition 2.1. Let v be such that $d_v \geq 3$. Then*

$$\mathbb{P}(\exists t \leq T_{1/2} : S_t^{(v)} = 0) = o(n^{-1}). \tag{4.5}$$

Since there are n vertices in the graph, Proposition 4.2 indeed proves (4.1) in Theorem 4.1. We start the proof with a bound on the depletion of high-degree vertices.

Lemma 4.3 (bound on the depletion of high-degree vertices). *Consider $\text{CM}_n(\mathbf{d})$ in the connectivity critical window defined in Condition 2.1, and perform the exploration up to time*

$$T_{1/2} = \max\{t : |\mathcal{N}_t| > n/2\}.$$

Then there exists $\varepsilon > 0$ such that

$$\mathbb{P}(\#\{v \in \mathcal{N}_{T_{1/2}} : d_v \geq 3\} < \varepsilon n) = o(n^{-1}). \tag{4.6}$$

Proof. Let us consider the exploration from a permutation ξ of the set of the half-edges chosen uniformly at random, as described above (4.3). We let $\mathcal{T}_{n/2}(\xi)$ denote the set of vertices such that all their half-edges are among the last $n/2$ of the permutation ξ . The previous definitions imply that $\mathcal{T}_{n/2}(\xi) \subseteq \mathcal{N}_{T_{1/2}}$.

We now pick a $k > 2$ such that $p_k = \lim n_k/n > 0$; from the definition of the connectivity critical window we know that such a k exists. We want to find a lower bound on

$$N_{T_{1/2}}(k) = \#\{v \in \mathcal{N}_{T_{1/2}} : d_v = k\} \geq \#\{v \in \mathcal{T}_{n/2}(\xi) : d_v = k\}.$$

Before running the exploration, we sequentially locate the half-edges of the vertices of degree k in ξ . We stop this process once we have examined $n/(4k)$ vertices, or when we run out of vertices of degree k . We define the σ -algebra \mathcal{F}_i^k generated by the positions of the half-edges of the first i vertices that we have examined. We then find that, at each step j , thanks to the stopping conditions, there are still at least $n/4$ available spots among the last $n/2$ half-edges in ξ , so that

$$\mathbb{P}(v_j \in \mathcal{T}_{n/2}(\xi) \mid \mathcal{F}_{j-1}^k) \geq \left(\frac{n}{4\ell_n}\right)^k. \tag{4.7}$$

We know that

$$\lim_{n \rightarrow \infty} \left(\frac{n}{4\ell_n}\right)^k = \left(\frac{1}{4d}\right)^k \equiv q_k,$$

so that

$$N_{T_{1/2}}(k) \stackrel{st}{\geq} \text{Bin}\left(\left(p_k \wedge \frac{1}{4k}\right)n, q_k\right), \tag{4.8}$$

where $\stackrel{st}{\geq}$ indicates stochastic domination. By concentration of the binomial distribution (see e.g. [1]), there exists a $c = c(a, q_k)$ such that, uniformly in n ,

$$\mathbb{P}\left(\text{Bin}(an, q_k) \leq \frac{an}{2} q_k\right) \leq e^{-cn} = o(n^{-1}). \tag{4.9}$$

The claim follows by picking

$$\varepsilon < \frac{1}{2} \left(p_k \wedge \frac{1}{4k}\right) q_k. \quad \square$$

We notice that $S_{t+1}^{(v)} - S_t^{(v)} < 0$ only when one of the following events occurs.

- ▷ $A(t) = \{d_{v(e_2(t))} = 1\}$, where $e_2(t)$ is the half-edge to which the t th paired half-edge is paired. In this case $S_{t+1}^{(v)} - S_t^{(v)} = -1$. Thanks to Lemma 4.3, if we define \mathcal{F}_k as the σ -algebra generated by the first k steps of the exploration, then, uniformly for $t \leq T_{1/2}$,

$$\mathbb{P}(A(t) \mid \mathcal{F}_{t-1}) \leq \frac{2\rho_1}{\sqrt{n}}. \tag{4.10}$$

- ▷ $B(t) = \{e_2(t) \in \mathcal{A}_t\}$, where $e_2(t)$ is the half-edge to which the t th paired half-edge is paired. In this case $S_{t+1}^{(v)} - S_t^{(v)} = -2$. From the description of the exploration, we obtain that, uniformly for $t \leq T_{1/2}$,

$$\mathbb{P}(B(t) \mid \mathcal{F}_{t-1}) \leq \frac{S_t^{(v)} - 1}{\ell_n - t - 1} \leq \frac{2S_t^{(v)}}{n}. \tag{4.11}$$

Now we prove three lemmas that together will yield Proposition 4.2.

The first lemma contains a lower bound on the survival time of the process. Indeed, we show that w.h.p. the component of v is at least of polynomial size with respect to n .

Lemma 4.4 (no early hit of zero). *Let $CM_n(\mathbf{d})$ be in the connectivity critical window defined in Condition 2.1. Then,*

$$\mathbb{P}(\exists t \leq n^{1/8} : S_t^{(v)} = 0) = o(n^{-1}). \tag{4.12}$$

Proof. For the process to die out before time $n^{1/8}$, it has to make at least 3 steps down while it is very low. We can characterize the ways in which this can happen in terms of occurrences of the events $A(t)$ (finding a degree 1 vertex) and $B(t)$ (finding an already active half-edge).

We thus need one of the following three events to occur:

$$F_1 = \bigcup_{s_1, s_2, s_3 \leq n^{1/8}} A(s_1) \cap A(s_2) \cap A(s_3) \cap \{S_{s_1}^{(v)}, S_{s_2}^{(v)}, S_{s_3}^{(v)} \leq 3\},$$

$$F_2 = \bigcup_{s_1, s_2 \leq n^{1/8}} A(s_1) \cap B(s_2) \cap \{S_{s_1}^{(v)}, S_{s_2}^{(v)} \leq 3\},$$

$$F_3 = \bigcup_{s_1, s_2 \leq n^{1/8}} B(s_1) \cap B(s_2) \cap \{S_{s_1}^{(v)}, S_{s_2}^{(v)} \leq 4\}.$$

We estimate using (4.10) and (4.11) to obtain

$$\mathbb{P}(F_1) \leq \binom{n^{1/8}}{3} \left(\frac{2\rho_1}{\sqrt{n}}\right)^3 \leq \frac{4\rho_1^3 n^{3/8}}{3 n^{3/2}} = o(n^{-1}), \tag{4.13}$$

$$\mathbb{P}(F_2) \leq \binom{n^{1/8}}{2} \frac{2\rho_1}{\sqrt{n}} \frac{6}{n} \leq \frac{3\rho_1 n^{1/4}}{2 n^{3/2}} = o(n^{-1}), \tag{4.14}$$

$$\mathbb{P}(F_3) \leq \binom{n^{1/8}}{2} \left(\frac{8}{n}\right)^2 \leq 32 \frac{n^{1/4}}{n^2} = o(n^{-1}). \tag{4.15}$$

Applying the union bound proves the claim. □

The next lemma proves instead that when the process is sufficiently low, it is very unlikely to decrease further, since we have few active half-edges with which to create loops.

Lemma 4.5 (unlikely to dip even lower). *Let $CM_n(\mathbf{d})$ be in the connectivity critical window defined in Condition 2.1. Fix v such that $d_v \geq 3$. Then, for every $t \leq T_{1/2}$ and $\gamma > 0$,*

$$\mathbb{P}\left(\sum_{i \leq \gamma n^{1/8}} (S_{t+i+1}^{(v)} - S_{t+i}^{(v)}) \mathbb{1}_{\{S_{t+i+1}^{(v)} < S_{t+i}^{(v)} < 3\gamma n^{1/8}\}} \geq 6\right) = o(n^{-2}). \tag{4.16}$$

Proof. As in the proof of Lemma 4.4, we find some events that must occur in order for the event in the left-hand side of (4.16) to occur. Again we express the possible ways in which the required 6 steps down can occur in terms of the events $A(t)$ and $B(t)$. We start by introducing some notation. For $1 \leq i < j$ and $s_i \geq 0$, we write

$$A_{[i,j]}(t) = A(t + s_i) \cap \dots \cap A(t + s_j), \quad B_{[i,j]}(t) = B(t + s_i) \cap \dots \cap B(t + s_j).$$

Then, for the event on the left-hand side of (4.16) to occur, we need one of the following events to occur:

$$\begin{aligned}
 G_1 &= \bigcup_{s_1, \dots, s_6 \leq \gamma n^{1/8}} A_{[1,6]}(t) \cap \bigcap_{i \leq 6} \{S_{t+s_i}^{(v)} \leq 3\gamma n^{1/8}\}, \\
 G_2 &= \bigcup_{s_1, \dots, s_5 \leq \gamma n^{1/8}} A_{[1,4]}(t) \cap B(t + s_5) \cap \bigcap_{i \leq 5} \{S_{t+s_i}^{(v)} \leq 3\gamma n^{1/8}\}, \\
 G_3 &= \bigcup_{s_1, \dots, s_4 \leq \gamma n^{1/8}} A_{[1,2]}(t) \cap B_{[3,4]} \cap \bigcap_{i \leq 4} \{S_{t+s_i}^{(v)} \leq 3\gamma n^{1/8}\}, \\
 G_4 &= \bigcup_{s_1, \dots, s_3 \leq \gamma n^{1/8}} B_{[1,3]}(t) \cap \bigcap_{i \leq 3} \{S_{t+s_i}^{(v)} \leq 3\gamma n^{1/8}\}.
 \end{aligned}$$

Again we estimate using (4.10) and (4.11) to obtain

$$\mathbb{P}(G_1) \leq \binom{\gamma n^{1/8}}{6} \left(\frac{2\rho_1}{\sqrt{n}}\right)^6 \leq \frac{2^6 \gamma^6 \rho_1^6 n^{6/8}}{6! n^3} = o(n^{-2}), \tag{4.17}$$

$$\mathbb{P}(G_2) \leq \binom{\gamma n^{1/8}}{5} \left(\frac{2\rho_1}{\sqrt{n}}\right)^4 \frac{6\gamma n^{1/8}}{n} \leq \frac{2^5 3\gamma \rho_1^4 \gamma^6 n^{6/8}}{5! n^3} = o(n^{-2}), \tag{4.18}$$

$$\mathbb{P}(G_3) \leq \binom{\gamma n^{1/8}}{4} \left(\frac{2\rho_1}{\sqrt{n}}\right)^2 \left(\frac{6\gamma n^{1/8}}{n}\right)^2 \leq \frac{2^4 3^2 \gamma^6 \rho_1^2 n^{6/8}}{4! n^3} = o(n^{-2}), \tag{4.19}$$

$$\mathbb{P}(G_4) \leq \binom{\gamma n^{1/8}}{3} \left(\frac{6\gamma n^{1/8}}{n}\right)^3 \leq \frac{6^3 \gamma^6 n^{6/8}}{3! n^3} = o(n^{-2}). \tag{4.20}$$

Applying the union bound proves the claim. □

We now show that not only does the exploration survive up to time $n^{1/8}$ but also that we have quite a large number of active half-edges.

Lemma 4.6 (law of large numbers lower bound on exploration). *Fix v such that $d_v \geq 3$. The exploration on $CM_n(\mathbf{d})$ in the connectivity critical window defined in Condition 2.1 satisfies that there exists a $\gamma > 0$ such that*

$$\mathbb{P}(S_{n^{1/8}}^{(v)} < 2\gamma n^{1/8}) = o(n^{-1}). \tag{4.21}$$

Proof. We divide the proof into two cases.

(1) There exists $t < n^{1/8}$ such that $S_t^{(v)} \geq 3\gamma n^{1/8}$. In this case, fix n so large that $3\gamma n^{1/8} - 6 \geq 2\gamma n^{1/8}$. Then, note that in order for $S_{n^{1/8}}^{(v)} < 2\gamma n^{1/8}$ to occur and since $S_{t+1}^{(v)} - S_t^{(v)} \geq -2$, we must have that

$$\sum_{i \leq \gamma n^{1/8}} (S_{t+i+1}^{(v)} - S_{t+1}^{(v)}) \mathbb{1}_{\{S_{t+i+1}^{(v)} < S_{t+i}^{(v)} < 3\gamma n^{1/8}\}} \geq 6,$$

which by Lemma 4.5 implies that $S_{n^{1/8}}^{(v)} \geq 3\gamma n^{1/8} - 6 \geq 2\gamma n^{1/8}$ has probability $o(n^{-2})$.

(2) $S_t^{(v)} < 3\gamma n^{1/8}$ for all $t \leq n^{1/8}$. In this case, we know from Lemma 4.5 that with probability $o(n^{-2})$ the sum of the down steps

$$(S_{t+i}^{(v)} - S_{t+i+1}^{(v)}) \mathbb{1}_{\{S_{t+i+1}^{(v)} < S_{t+i}^{(v)} < 3\gamma n^{1/8}\}}$$

is at most 6. Under this condition, we recall Lemma 4.3 and note that $d_{v_t} \geq 3$ with probability at least ε , for some $\varepsilon > 0$, since $n^{1/8} \leq T_{1/2}$. Thus,

$$S_{n^{1/8}}^{(v)} \stackrel{st}{\geq} \text{Bin}(n^{1/8}, \varepsilon) - 6. \tag{4.22}$$

By concentration of the binomial distribution (see e.g. [1])

$$\mathbb{P}\left(\text{Bin}(n^{1/8}, \varepsilon) \leq \frac{1}{2}\varepsilon n^{1/8}\right) \leq e^{-cn^{1/8}} = o(n^{-2}), \tag{4.23}$$

for sufficiently large n . The claim now follows by choosing $\gamma < \varepsilon/4$. □

Now we know that at time $t = n^{1/8}$, with probability $1 - o(n^{-1})$, $S_t^{(v)} \geq 2\gamma n^{1/8}$. This means that from that point onwards, we need at least $\gamma n^{1/8}$ steps for the process to die. Thus, the following lemma is needed to prove Proposition 4.2.

Lemma 4.7 (process does not go down too much). *Let $\text{CM}_n(\mathbf{d})$ be in the connectivity critical window defined in Condition 2.1. Fix v such that $d_v \geq 3$. Then, for every $\gamma > 0$,*

$$\mathbb{P}(\exists t \in (n^{1/8}, T_{1/2}): S_{t+\gamma n^{1/8}}^{(v)} < S_t^{(v)} < 3\gamma n^{1/8} - 6) = o(n^{-1}). \tag{4.24}$$

Proof. First fix $t \in (n^{1/8}, T_{1/2})$. Again we split the proof into two parts.

(1) There exists $i < \gamma n^{1/8}$ such that $S_{t+i}^{(v)} \geq 3\gamma n^{1/8}$. In this case, we again know from Lemma 4.5 that

$$S_{t+\gamma n^{1/8}}^{(v)} \geq 3\gamma n^{1/8} - 6 \geq 2\gamma n^{1/8}$$

with probability $1 - o(n^{-2})$.

(2) $S_{t+i}^{(v)} < 3\gamma n^{1/8}$ for all $t \leq \gamma n^{1/8}$. In this case we know from Lemma 4.5 that with probability $o(n^{-2})$ the sum of the down steps

$$(S_{t+i}^{(v)} - S_{t+i+1}^{(v)}) \mathbb{1}_{\{S_{t+i+1}^{(v)} < S_{t+i}^{(v)} < 3\gamma n^{1/8}\}}$$

is at most 6. Under this condition we can again write

$$S_{t+\gamma n^{1/8}}^{(v)} - S_t^{(v)} \stackrel{st}{\geq} \text{Bin}(n^{1/8}, \varepsilon) - 6. \tag{4.25}$$

Formula (4.23) proves that the probability that $S_{t+\gamma n^{1/8}}^{(v)} < S_t^{(v)}$ is at most $o(n^{-2})$.

The union bound implies that

$$\mathbb{P}(\exists t \in (n^{1/8}, T_{1/2}): S_{t+\gamma n^{1/8}}^{(v)} < S_t^{(v)} < 3\gamma n^{1/8} - 6) \leq \ell_n o(n^{-2}) = o(n^{-1}). \tag{4.26}$$

□

Now we are ready to complete the proof of Proposition 4.2.

Proof of Proposition 4.2. Lemmas 4.4 and 4.6 show that up to time $n^{1/8}$ the process is very unlikely to die and very likely to grow at least until polynomial size:

$$\mathbb{P}(T_0 > n^{1/8}, S_{n^{1/8}}^{(v)} > 2\gamma n^{1/8}) = 1 - o(n^{-1}). \tag{4.27}$$

Now we define the sequence of random variables $Q_i = S_{(1+\gamma^i)n^{1/8}}^{(v)}$, so that

$$\mathbb{P}(Q_0 < 2\gamma n^{1/8}) = o(n^{-1}).$$

By Lemma 4.7,

$$\mathbb{P}\left(\forall i \leq \frac{T_{1/2}}{\gamma n^{1/8}} : Q_{i+1} \geq Q_i\right) = 1 - o(n^{-1}), \tag{4.28}$$

and consequently

$$\mathbb{P}\left(\forall i \leq \frac{T_{1/2}}{\gamma n^{1/8}} : Q_i \geq 2\gamma n^{1/8}\right) = 1 - o(n^{-1}). \tag{4.29}$$

Since $S_{t+1}^{(v)} - S_t^{(v)} \geq -2$, we know that $S_{t+s}^{(v)} \geq S_t^{(v)} - 2s$, so we conclude that

$$\mathbb{P}(S_t^{(v)} > 0 \ \forall t \leq T_{1/2}) = 1 - o(n^{-1}). \tag{4.30}$$

This completes the proof of Proposition 4.2. □

We can now conclude the proof of Theorem 4.1.

Proof of Theorem 4.1. Proposition 4.2 proves (4.1) in Theorem 4.1. To prove (4.2) in Theorem 4.1, we use that if $|\mathcal{C}(v)| > n/2$, then $v \in \mathcal{C}_{\max}$, to bound

$$\mathbb{E}[\#\{v \in [n] \setminus \mathcal{C}_{\max} : d_v \geq 3\}] \leq \mathbb{E}[\#\{v \in [n] : d_v \geq 3, |\mathcal{C}(v)| < n/2\}] = o(1) \tag{4.31}$$

by Proposition 4.2. □

To show that the size of the graph without the giant component does indeed have bounded expectation, we need a slightly stronger result.

Proposition 4.8 (clusters of vertices of degree at least three outside \mathcal{C}_{\max}). *Let $CM_n(\mathbf{d})$ be in the connectivity critical window defined in Condition 2.1. Then*

$$\mathbb{E}[\#\{v \notin \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}] \rightarrow 0, \tag{4.32}$$

where, for a set of vertices $A \subseteq [n]$, $v \leftrightarrow A$ denotes that there exists $a \in A$ such that v and a are in the same connected component.

Proof. We have already proved that $\mathbb{E}[\#\{v \in [n] \setminus \mathcal{C}_{\max} : d_v \geq 3\}] \rightarrow 0$. We now initialize the exploration starting from a vertex v with $d_v \in \{2, 1\}$. Notice that the probability of the process surviving for $n^{1/8}$ steps without finding vertices of degree 3 is smaller than $e^{-cn^{1/8}}$ for some $c > 0$, since at every step the probability of finding a vertex w with $d_w \geq 3$ is bounded away from 0.

▷ If $d_v = 2$ and our exploration finds a vertex w with $d_w \geq 3$ before time $n^{1/8}$, then for the process to die out before time $n^{1/8}$, we again need one of the events F_1, F_2, F_3 to occur. Then we can apply Lemmas 4.4, 4.6 and 4.7 to complete the proof that

$$\mathbb{E}[\#\{v \in \mathcal{N}_2 \setminus \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}] \rightarrow 0,$$

in the same way as the proof of Proposition 4.2.

▷ In the connectivity critical window, we have that $n_1 = O(\sqrt{n})$. If $d_v = 1$ and our exploration at a certain point finds a vertex w with $d_w \geq 3$, then for the process to die out before time $n^{1/8}$ we need one of the following two events to occur:

$$\begin{aligned} F'_1 &= \bigcup_{s_1, s_2 \leq n^{1/8}} A(s_1) \cap A(s_2) \cap \{S_{s_1}^{(v)}, S_{s_2}^{(v)} \leq 2\}, \\ F'_2 &= \bigcup_{s \leq n^{1/8}} B(s) \cap \{S_s^{(v)} = 2\}. \end{aligned} \tag{4.33}$$

We estimate using (4.10) and (4.11) to obtain

$$\mathbb{P}(F'_1) \leq \binom{n^{1/8}}{2} \left(\frac{2\rho_1}{\sqrt{n}}\right)^2 \leq 2\rho_1^2 \cdot \frac{n^{1/4}}{n} = o(n^{-1/2}), \tag{4.34}$$

$$\mathbb{P}(F'_2) \leq n^{1/8} \cdot \frac{4}{n} = o(n^{-1/2}). \tag{4.35}$$

Now we can apply Lemmas 4.6 and 4.7 to complete the proof that

$$\mathbb{E}[\#\{v \in \mathcal{N}_1 \setminus \mathcal{C}_{\max} : v \leftrightarrow w; d_w \geq 3\}] \rightarrow 0$$

in a similar way to the proof of Proposition 4.2.

Since

$$\begin{aligned} \mathbb{E}[\#\{v \notin \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}] \\ = \mathbb{E}[\#\{v \in [n] \setminus \mathcal{C}_{\max} : d_v \geq 3\}] + \mathbb{E}[\#\{v \in (\mathcal{N}_1 \cup \mathcal{N}_2) \setminus \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}], \end{aligned} \tag{4.36}$$

we obtain the claim. □

5. Proof of the main theorems

We can now finally prove the main theorems, putting together results from the previous two sections.

Proof of Theorem 2.2. We know that

$$\{\text{CM}_n(\mathbf{d}) \text{ is connected}\} = \{C_k(n) = L_k(n) = 0 \forall k\} \cap \{[n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2) \subseteq \mathcal{C}_{\max}\}. \tag{5.1}$$

We have proved in Theorem 4.1 that w.h.p. $[n] \setminus \mathcal{C}_{\max} \subseteq \mathcal{N}_1 \cup \mathcal{N}_2$. Thus,

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) = \mathbb{P}(C_k(n) = L_k(n) = 0 \forall k) + o(1). \tag{5.2}$$

By Theorem 3.3 and the independence of C_k, L_k , for each $j < \infty$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \forall k \leq j) = \prod_{k=1}^j \mathbb{P}(C_k = 0) \prod_{k=2}^j \mathbb{P}(L_k = 0). \tag{5.3}$$

To pass to the limit we use dominated convergence. We compute that

$$\mathbb{E}[L_k(n)] = n_1 \frac{2n_2}{\ell_n - 1} \frac{2n_2 - 2}{\ell_n - 3} \cdots \frac{2n_2 - 2k + 4}{\ell_n - 2k + 3} \frac{n_1 - 1}{\ell_n - 2k + 1} \leq \frac{n_1^2 (2n_2)^{k-2}}{(\ell_n - 2k)^{k-1}}. \tag{5.4}$$

Now

$$\frac{n_1}{\sqrt{n}} \rightarrow \rho_1, \quad \frac{n_2}{n} \rightarrow p_2 \quad \text{and} \quad \frac{\ell_n}{n} \rightarrow d,$$

for each ε and n sufficiently large that

$$\mathbb{E}[L_k(n)] \leq \frac{n_1^2 (2n_2)^{k-2}}{2(\ell_n - 2k)^{k-1}} \leq \frac{(\rho_1^2 + \varepsilon)^2}{2(d - \varepsilon)} \left(\frac{2(p_2 + \varepsilon)}{d - \varepsilon} \right)^{k-2}. \tag{5.5}$$

For ε small enough, $2(p_2 + \varepsilon) < d - \varepsilon$, so the series on the right-hand side of (5.5) is exponentially small in k . Similarly for $C_k(n)$,

$$\mathbb{E}[C_k(n)] = \frac{1}{2k} n_2 \frac{2n_2 - 2}{\ell_n - 2} \cdots \frac{2n_2 - 2k + 4}{\ell_n - 2k + 4} \frac{1}{\ell_n - 2k + 2} \leq \frac{(2n_2)^k}{k(\ell_n - 2k)^k}. \tag{5.6}$$

As before, we have for every $\varepsilon > 0$

$$\mathbb{E}[C_k(n)] \leq \frac{1}{k} \frac{(2n_2)^k}{(\ell_n - 2k)^k} \leq \frac{(2n_2)^k}{k(\ell_n - 2k)^k} \leq \frac{(2p_2 + 2\varepsilon)^k}{k(d - \varepsilon)^k}. \tag{5.7}$$

Again, for $\varepsilon > 0$ small enough, $2(p_2 + \varepsilon) < d - \varepsilon$, so the series on the right-hand side of (5.7) is exponentially small in k .

Since

$$\{C_k(n) = L_k(n) = 0 \ \forall k\} = \bigcap_j \{C_k(n) = L_k(n) = 0 \ \forall k \leq j\},$$

we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k) &\leq \lim_{j \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k \leq j) \\ &= \prod_{k=1}^{\infty} \mathbb{P}(C_k = 0) \prod_{k=2}^{\infty} \mathbb{P}(L_k = 0) \\ &= \exp\left(-\sum_{k=1}^{\infty} \frac{(2p_2)^k}{2kd^k} - \sum_{k=2}^{\infty} \frac{\rho_1^2 (2p_2)^{k-2}}{2d^{k-1}}\right) \\ &= \left(\frac{d - 2p_2}{d}\right)^{1/2} \exp\left(-\frac{\rho_1^2}{2(d - 2p_2)}\right), \end{aligned} \tag{5.8}$$

where we use that $-\sum_{k \geq 1} x^k/k = \log(1 - x)$ for $x \geq 0$.

For the lower bound, we use

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k) &\geq \lim_{j \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k \leq j) \\ &\quad - \limsup_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\exists k > j : C_k(n) + L_k(n) \geq 1). \end{aligned} \tag{5.9}$$

Using Markov’s inequality, we find that

$$\begin{aligned} & \limsup_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\exists k > j : C_k(n) + L_k(n) \geq 1) \\ &= \limsup_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{k > j} (C_k(n) + L_k(n)) \geq 1\right) \\ &\leq \limsup_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{k > j} \mathbb{E}[C_k(n) + L_k(n)] = 0, \end{aligned} \tag{5.10}$$

by (5.5) and (5.7). As a result,

$$\lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k) = \left(\frac{d - 2p_2}{d}\right)^{1/2} \exp\left(-\frac{\rho_1^2}{2(d - 2p_2)}\right). \tag{5.11}$$

From (5.11), we obtain (2.1) using Theorem 4.1.

We next investigate the boundary cases in Remark 1. The result in (2.5) follows in an identical way to the proof of Theorem 2.2. For the result for $d = \infty$ in (2.6), we notice that if $\mathbb{E}[D_n] \rightarrow \infty$ then, for all $k \geq 1$,

$$\lim_{n \rightarrow \infty} \frac{2n_2}{\ell_n - 2k} = 0, \tag{5.12}$$

so that

$$\sum_{k \geq 3} L_k(n) + \sum_{k \geq 1} C_k(n) \xrightarrow{\mathbb{P}} 0$$

by (5.5) and (5.7) and Markov’s inequality. Moreover,

$$\mathbb{P}(L_2(n) = 0) = \prod_{i=1}^{n_1} \frac{\ell_n - n_1 - i + 1}{\ell_n - 2i + 1} = e^{-n_1^2/(2\ell_n(1+o(1)))}, \tag{5.13}$$

so that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) = \lim_{n \rightarrow \infty} \mathbb{P}(L_2(n) = 0) = \lim_{n \rightarrow \infty} e^{-n_1^2/(2\ell_n)}. \tag{5.14}$$

Further, we notice that

$$n - |\mathcal{C}_{\max}| = \sum_{k=1}^{\infty} k(C_k(n) + L_k(n)) + \#\{v \notin \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}. \tag{5.15}$$

From Proposition 4.8 we know that

$$\mathbb{E}[\#\{v \notin \mathcal{C}_{\max} : v \leftrightarrow [n] \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)\}] \rightarrow 0,$$

so that

$$n - |\mathcal{C}_{\max}| = \sum_{k \geq 1} k(C_k(n) + L_k(n)) + o_{\mathbb{P}}(1). \tag{5.16}$$

By (5.5) and (5.7) and dominated convergence, we obtain

$$n - |\mathcal{C}_{\max}| \xrightarrow{d} \sum_{k=1}^{\infty} k(C_k + L_k), \tag{5.17}$$

which completes the proof of (2.2).

Since we have shown convergence of all moments, we also obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[n - |\mathcal{C}_{\max}|] &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} k \mathbb{E}[C_k(n) + L_k(n)] \\ &= \lim_{j \rightarrow \infty} \sum_{k=1}^j k \frac{(2p_2)^k}{2kd^k} + \sum_{k=2}^j k \frac{\rho_1^2(2p_2)^{k-1}}{2d^{k-1}} = \frac{\rho_1^2(2d - p_2)}{2(d - p_2)^2} + \frac{p_2}{d - 2p_2}, \end{aligned} \tag{5.18}$$

as required. □

Proof of Theorem 2.3. If we condition on simplicity, then we already have that $C_1(n) = C_2(n) = 0$. Therefore, we find using the same method as in the previous proof that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(C_k(n) = L_k(n) = 0 \ \forall k \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ = \prod_{k=3}^{\infty} \mathbb{P}(C_k = 0) \prod_{k=2}^{\infty} \mathbb{P}(L_k = 0) = \exp\left(-\sum_{k=3}^{\infty} \frac{(2p_2)^k}{2kd^k} - \sum_{k=2}^{\infty} \frac{\rho_1^2(2p_2)^{k-2}}{2d^{k-1}}\right) \\ = \left(\frac{d - 2p_2}{d}\right)^{1/2} \exp\left(-\frac{\rho_1^2}{2(d - 2p_2)} + \frac{p_2^2 + dp_2}{d^2}\right), \end{aligned} \tag{5.19}$$

from which we obtain (2.7) thanks to Theorem 4.1.

We recall that $\mathcal{N}_n(\mathbf{d})$ denotes the number of simple graphs with degree distribution \mathbf{d} . We know that

$$\mathcal{N}_n(\mathbf{d}) = \exp\left\{-\frac{v}{2} - \frac{v^2}{4}\right\} \frac{(\ell_n - 1)!!}{\prod_{i \in [n]} d_i!} (1 + o(1)). \tag{5.20}$$

Since $\text{CM}_n(\mathbf{d})$ conditioned on being simple has the uniform distribution over all possible simple graphs with degree sequence \mathbf{d} ,

$$\mathcal{N}_n^C(\mathbf{d}) = \mathcal{N}_n(\mathbf{d}) \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}), \tag{5.21}$$

which yields the claim. □

Acknowledgements

The work of LF and RvdH is supported by the Netherlands Organisation for Scientific Research (NWO) through the Gravitation NETWORKS grant 024.002.003. The work of RvdH is also supported by the Netherlands Organisation for Scientific Research (NWO) through VICI grant 639.033.806.

References

- [1] Arratia, R. and Gordon, L. (1989) Tutorial on large deviations for the binomial distribution. *Bull. Math. Biol.* **51** 125–131.
- [2] Bollobás, B. (2001) *Random Graphs*, second edition, Vol. 73 of Cambridge Studies in Advanced Mathematics, Cambridge University Press.
- [3] van der Hofstad, R. (2017) *Random Graphs and Complex Networks*, Vol. 1, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

- [4] Janson, S. (2008) The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.* **18** 1651–1668.
- [5] Janson, S. (2009) The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** 205–225.
- [6] Janson, S. (2014) The probability that a random multigraph is simple, II. *J. Appl. Probab.* **51A** 123–137.
- [7] Janson, S. and Łuczak, M. J. (2009) A new approach to the giant component problem. *Random Struct. Alg.* **34** 197–216.
- [8] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience.
- [9] Łuczak, T. (1992) Sparse random graphs with a given degree sequence. In *Random Graphs (Poznań 1989)*, Vol. 2, Wiley-Interscience, pp. 165–182.
- [10] Molloy, M. and Reed, B. (1995) A critical point for random graphs with a given degree sequence. In *Proc. Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science: Random Graphs '93 (Poznań 1993)*, *Random Struct. Alg.* **6** 161–179.
- [11] Molloy, M. and Reed, B. (1998) The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7** 295–305.
- [12] Wormald, N. C. (1981) The asymptotic connectivity of labelled regular graphs. *J. Combin. Theory Ser. B* **31** 156–167.