

Almost-2-regular random graphs

LORENZO FEDERICO

LUISS — Data Lab, Dipartimento di Scienze Politiche
Viale Pola 12, Rome
Italy
lfederico@luiss.it

Abstract

We study a special case of the configuration model, in which almost all the vertices of the graph have degree 2. We show that the graph has a very peculiar and interesting behaviour; in particular, when the graph is made up of a vast majority of vertices of degree 2 and a vanishing proportion of vertices of higher degree, the giant component contains $n(1 - o(1))$ vertices, but the second component can still grow as a power of n . On the other hand, when almost all the vertices have degree 2, except for $o(n)$ which have degree 1, there is no component of linear size.

1 Introduction

In this paper, we study the behaviour of random graphs in which the majority of vertices have degree 2, but there exists a vanishing proportion of vertices of either higher or lower degree. The study of random graphs with a prescribed degree sequence has been a very popular topic for decades. The usual way to construct them has been the procedure known as the *configuration model*, which yields a random multigraph over n vertices with a fixed degree sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$. Conditioning on simplicity, the configuration model has uniform distribution over all the labeled graphs with such degree sequence. In the configuration model each vertex v_i , $i \in [n]$ is given d_i half-edges, and all the half-edges are paired uniformly at random to create the edges of the graph. This model is very flexible, allowing for any arbitrary degree sequence, while remaining tractable, thanks to its high level of symmetry. For these reasons it has become widely popular, both among applied researchers trying to fit the model to empirical degree distributions observed in real-world networks, such as power laws, and theoretical mathematicians who were looking for a tractable model that could exhibit a wide variety of behaviours. The study of connectivity of random graphs with a prescribed degree sequence started with the work of Bollobás [3] and Wormald [19], who proved that if the minimum degree is at least 3, the graph is connected with high probability (w.h.p.), and then it was further developed by Luczak [16], and the author and van der Hofstad [9], who analyzed the asymptotic

connectivity probability when vertices of lower degree are allowed. Moreover, it is known that the configuration model, like many other random graphs, presents a phase transition. The first investigation of this phenomenon is due to Molloy and Reed [17], who proved that the critical point for the phase transition is identified by $\lim_{n \rightarrow \infty} \sum_{i=1}^n d_i(d_i - 2)/n = 0$. Above such threshold there exists w.h.p. a connected component that contains a positive fraction of the vertices; below, instead, the size of the largest component is determined by the highest degree vertex up to a constant, as proved by Janson [13]. For the study of the behaviour close to criticality see the work by Dhara et al. [5, 6] and van der Hofstad et al. [11].

Usually the situation in which vertices of degree 2 make up almost the entirety of the graph has been left out as a boundary case that was hard to address with the usual techniques and only orders of magnitude of component sizes have been derived in [16]. Here, we will analyze in detail what happens under such an assumption, showing that indeed these graphs show very peculiar properties that are rarely found in other settings.

Notation.

All limits in this paper are taken as $n \rightarrow \infty$ unless stated otherwise. For asymptotic statements we use the following notation:

- Given a sequence of events $(\mathcal{A}_n)_{n \geq 1}$ we say that \mathcal{A}_n happens *with high probability* (*w.h.p.*) if $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$.
- Given two random variables X and Y we write $X \preceq Y$ to indicate that Y *stochastically dominates* X .
- Given the 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 sequences of (possibly quasi-deterministic) random variables $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$, we write $X_n = O(Y_n)$ if the sequence $(X_n/Y_n)_{n \geq 1}$ is bounded almost surely; $X_n = o(Y_n)$ if $X_n/Y_n \rightarrow 0$ almost surely; $X_n = \Theta(Y_n)$ if $X_n = O(Y_n)$ and $Y_n = O(X_n)$.
- Similarly, for sequences $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$ of (possibly degenerate) random variables, we write $X_n = O_{\mathbb{P}}(Y_n)$ if the sequence $(X_n/Y_n)_{n \geq 1}$ is tight; $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \xrightarrow{\mathbb{P}} 0$; and $X_n = \Theta_{\mathbb{P}}(Y_n)$ if $X_n = O_{\mathbb{P}}(Y_n)$ and $Y_n = O_{\mathbb{P}}(X_n)$.
- $\text{Poi}(\lambda)$ denotes a Poisson distributed random variable with mean λ and $\text{Bin}(n, p)$ denotes a random variable with binomial distribution with n trials each with probability of success p .

We will use the standard abbreviations i.i.d. for independent identically distributed and a.s. for almost surely.

2 Main Results

We first recall that the configuration model $\text{CM}_n(\mathbf{d})$ is constructed from a degree sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$ such that $\sum_{i=1}^n d_i =: \ell$ is even, giving d_i half-edges to each vertex v_i for any $i \in [n]$ and pairing them uniformly at random to form the edges. We consider the pairing as a sequential process, i.e., we pick one of the ℓ half-edges, and we pair it with another one chosen uniformly at random among the $\ell - 1$ remaining to form an edge, then we pick another half edge and pair it with a uniformly chosen one among the $\ell - 3$ remaining ones to build another edge, and we iterate until all the half-edges have been paired. In this case, we allow for self-loops and multi-edges to be present in the graph. The specific rule according to which the new half-edge is chosen at the beginning of each new step is not relevant for the distribution of the final outcome, something that we will exploit in many proofs. We study separately two cases depending on whether the degree sequence allows for vertices of degree larger than 2 and for vertices of degree 1. We always assume that there are no vertices of degree 0 since they do not interact with the rest of the graph and are thus uninteresting. For each $j \in \mathbb{N}$ we define \mathcal{N}_j as the set of vertices of degree j in \mathbf{d} and n_j as its cardinality. We write the total number of half-edges as

$$\ell := \ell(n) = \sum_{i=1}^n d_i = \sum_{j=1}^{\infty} j n_j,$$

and \mathcal{L} for the set of all the half-edges. Moreover, we define the total number of half-edges attached to vertices of degree different from 2 as

$$\ell_{\neq 2} := \sum_{j \neq 2} j n_j,$$

and $\mathcal{L}_{\neq 2}$ for the set of all such half-edges. Note that if ℓ is even, then also $\ell_{\neq 2}$ is even, since $\ell_{\neq 2} = \ell - 2n_2$. We study separately the two cases in which the degree sequence allows for vertices of degree higher or lower than 2, as their behaviour is completely different. We define the conditions under which we call a sequence of configuration models $\text{CM}_n(\mathbf{d})$ a sequence of *upper almost-two-regular graphs*.

Condition 2.1 (Upper almost-two-regular graph). *We define a sequence of random graphs as a sequence of upper almost-two-regular graphs if it is distributed as a sequence $(\text{CM}_n(\mathbf{d}))_{n \in \mathbb{N}}$ and the following conditions are satisfied as $n \rightarrow \infty$:*

- $\ell/n \rightarrow 2$,
- $n_0, n_1 = 0$,
- $\ell_{\neq 2} \rightarrow \infty$.

This conditions can be described as the degree sequence having a vast majority of vertices of degree 2 and a diverging but sublinear number of vertices of higher degree.

As computed originally by Łuczak in [16], under strict conditions on the maximum degree of the graph, the giant component contains almost all the vertices, but the largest components outside of the giant are of order of magnitude $\Theta(n/\ell_{\neq 2})$. This shows the possibility for a supercritical random graph to have components outside the giant of every intermediate order of magnitude between 1 and n , choosing an appropriate asymptotic of $\ell_{\neq 2}$. We generalize and strengthen this result in the main theorem of this paper, which describes the asymptotic behaviour of the largest component \mathcal{C}_{\max} and the i -th largest component \mathcal{C}_i , for every fixed $i \in \mathbb{N}$:

Theorem 2.2. *Consider a sequence $(\text{CM}_n(\mathbf{d}))_{n \geq 1}$ that satisfies Condition 2.1. Then as $n \rightarrow \infty$,*

$$n - |\mathcal{C}_{\max}| = O_{\mathbb{P}}(n/\ell_{\neq 2}), \tag{2.1}$$

and for every $i \geq 2$

$$|\mathcal{C}_i| \frac{\ell_{\neq 2}}{n} \xrightarrow{d} Y_i,$$

where Y_i is a random variable with support on \mathbb{R}^+ and cumulative distribution function

$$F_{Y_i}(a) = \mathbb{P}\left(\text{Poi}\left(-\int_a^\infty \frac{e^{-r/2}}{2r} dr\right) \leq i - 2\right). \tag{2.2}$$

By the definition we gave of the upper-almost-two-regular random graph, we know that the random variable D_n representing the degree of a randomly chosen vertex in $\text{CM}_n(\mathbf{d})$, converges in distribution to $\mathfrak{2}$, which we define as the random variable such that $\mathbb{P}(\mathfrak{2} = 2) = 1$. We know [1, 14] that, as long as the second moment of D converges, $\text{CM}_n(\mathbf{d})$ is simple with non-vanishing probability, and that conditionally on simplicity, it is a uniformly chosen simple graph with the degree sequence \mathbf{d} . Under these stricter conditions we can show that Theorem 2.2 holds also conditionally on simplicity, and thus holds for a uniform simple graph with degree sequence \mathbf{d} .

Theorem 2.3. *Consider a sequence of simple graphs $(\widetilde{\text{CM}}_n(\mathbf{d}))_{n \geq 1}$ such that for each n , $\widetilde{\text{CM}}_n(\mathbf{d})$ is distributed as $\text{CM}_n(\mathbf{d})$ from a sequence $(\text{CM}_n(\mathbf{d}))_{n \geq 1}$ that satisfies Condition 2.1 and such that $\mathbb{E}[D_n^i] \rightarrow 2^i, i = 1, 2, 3$, conditioned on simplicity. Then as $n \rightarrow \infty$, equations (2.1)-(2.2) hold.*

In light of more recent results, this stands up as a counterexample to the universality of the “no middle ground” property, i.e. that in a supercritical random graph the connected components outside the giant are always very small, typically with the size of the second largest component expressed as a polynomial in $\log n$, which holds for the most famous random graph models, as we will discuss in Section 3.

We also define the *lower almost-two-regular graphs* as sequences of configuration models satisfying the following conditions as $n \rightarrow \infty$:

Condition 2.4 (Lower almost-two-regular graph). *We define a sequence of random graphs as a sequence of lower almost-two-regular graphs if it is distributed as a sequence $(\text{CM}_n(\mathbf{d}))_{n \in \mathbb{N}}$ and the following conditions are satisfied as $n \rightarrow \infty$:*

- $\ell/n \rightarrow 2$,
- $n_0 = 0$,
- $\ell_{\neq 2} = n_1 \rightarrow \infty$.

In this case we are instead considering degree sequences in which the vast majority of vertices have degree 2, and the only other ones are a sublinear number of degree 1 vertices. In this case the behavior of the graph is radically different, as expressed in the following theorem. We define \mathcal{C}_j as the j -th largest cluster, then we obtain:

Theorem 2.5. *Consider a sequence $(\text{CM}_n(\mathbf{d}))_{n \geq 1}$ that satisfies Condition 2.4. Then, for both sequences, as $n \rightarrow \infty$, for every $j \in \mathbb{N}$*

$$|\mathcal{C}_j| = \frac{2n \log n_1}{n_1} (1 + o_{\mathbb{P}}(1)). \quad (2.3)$$

We see how this model closely resembles subcritical random graphs, with the largest components being all very similar to each other and with no component of linear size. The fact that the same holds for a lower-almost-two-regular graph conditioned on simplicity is a simple corollary:

Corollary 2.6. *Consider a sequence of simple graphs $(\widetilde{\text{CM}}_n(\mathbf{d}))_{n \geq 1}$ such that for each n , $\widetilde{\text{CM}}_n(\mathbf{d})$ is distributed as $\text{CM}_n(\mathbf{d})$ from a sequence $(\text{CM}_n(\mathbf{d}))_{n \geq 1}$ that satisfies Condition 2.4, conditioned on simplicity. Then (2.3) holds.*

We do not investigate in this paper what happens when the vast majority of vertices have degree 2 but there are both vertices of higher and lower degree. That would require a very detailed case analysis depending on the relative scaling of n_i for different i as $n \rightarrow \infty$ which we think could be interesting for a future work.

3 Discussion of the results

In this section, we present some relevant consequences and some interesting observations related to our main theorems. The first interesting fact to point out is that the upper almost-two-regular graph is a counterexample to the very general “no middle ground” property of supercritical random graphs. It was first observed by Erdős and Rényi [8] that the random graph $G(n, m)$ (the uniform random graph with n vertices and m edges), in the supercritical phase, i.e., when $m > (1 + c)n/2$, presents a big gap between the largest component, which contains a positive fraction of the vertices, and the second one, which contains $\Theta_{\mathbb{P}}(\log n)$ vertices. Such a property has been established to hold for a huge class of random graphs such as inhomogeneous random graphs [4], most of the other cases of the configuration model [17] and percolation on the Hamming graph [12]. For percolation on a large box or torus on \mathbb{Z}^d , the second largest component in the supercritical phase is of order $\Theta_{\mathbb{P}}(\log n^{d/(d-1)})$ [15]. In the case presented in this paper instead, we see from Theorem 2.2 that the upper

almost-two-regular graph shows a clear supercritical behaviour, that is, there exists a unique giant component, that has a clearly different structure from all the other components and includes a positive fraction (actually, almost all) of the vertices, but the second largest component can be of every possible order of magnitude such that $1 \ll |\mathcal{C}_2| \ll n$, choosing the right scaling between n and $\ell_{\neq 2}$. One of the possible

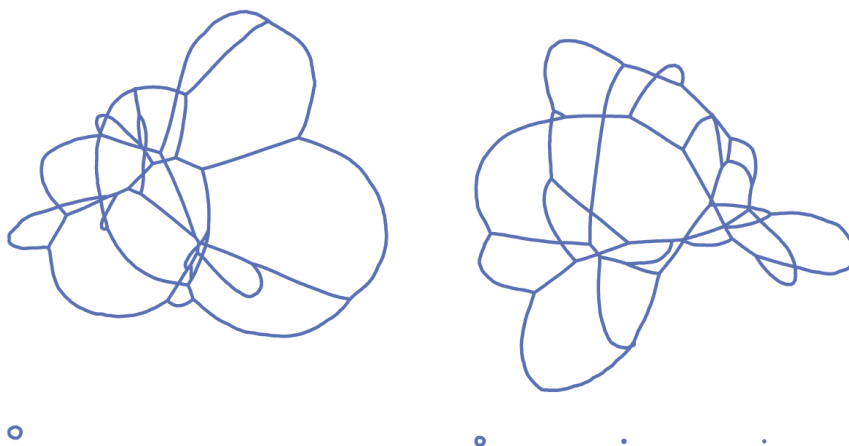


Figure 1: Two samples of the Configuration model over 10000 vertices with $n_2 = 9970$, $n_3 = 30$. It is easy to observe that the graph is mostly made by long linear sections.

reasons for this phenomenon is the *one-dimensional nature* of the graph. We can observe that the *local weak limit* of the graph is \mathbb{Z} , i.e., the neighbourhood of a uniformly chosen vertex resembles w.h.p. a straight path up to any finite distance (see Figure 1). The reasons for which a large second component in a supercritical graph is unlikely is that large sets tend to have large boundaries, and thus are hard to separate from the giant component. In a one-dimensional model instead, connected sets with large boundaries are impossible to achieve, and thus it is relatively easy to separate even quite large components from the giant.

The lower almost-two-regular graph instead, shows clear signs of being a subcritical graph, i.e., the largest component is in no way “special”, but the k largest components are almost indistinguishable for every fixed k . Still, also in this case we see the peculiar nature of this model, as the size of the largest component is in no way determined by the highest degree up to a constant or at most a logarithmic term, as it is in most cases, like $G(n, m)$ [7], the Random Intersection Graph [2] or many instances of the Configuration Model [13] but we see that the largest component can be of any order of magnitude such that $1 \ll |\mathcal{C}_{\max}| \ll n$ in a subcritical graph with highest degree equal to 2.

4 Proof of Theorem 2.2

In this section, we prove the main results about the upper almost-2-regular graph, both seen as a multigraph, as summarized in Theorem 2.2, and conditioned on simplicity as in Theorem 2.3. Our proof will consist of two main steps: first, we prove that w.h.p. all the connected components except the giant are cycles (i.e. components that contain only vertices of degree 2), second, we find the distribution of the size of the largest cycle component in the graph. The first proof uses an argument similar to the one presented by Łuczak in [16] and by Pittel and Wormald in [18], by which the configuration model after the removal of the vertices of degree 2 is equivalent to a configuration model with a different degree sequence. The second instead will require the use of enumerative combinatorics to find the probability of having cycles of a given length. In order to understand the large-scale structure of $\text{CM}_n(\mathbf{d})$, we define the notion of *kernel* of the graph, adapted from [16].

Definition 4.1 (Kernel of the Configuration Model). *We define the kernel $K(\text{CM}_n(\mathbf{d}))$ of the configuration model as the graph obtained from $\text{CM}_n(\mathbf{d})$ after running the following algorithm:*

Initialize *Sample $\text{CM}_n(\mathbf{d})$.*

Step *At every step, choose a vertex v_i such that $d_i = 2$ according to any arbitrary rule and remove it from $\text{CM}_n(\mathbf{d})$. **If** v_i has a self loop, do nothing more, **else** pair the two half-edges which had been paired with the half-edges of v_i , even if this procedure would create a self loop or a multi-edge.*

Terminate *Stop the process when there are no more vertices of degree 2.*

From the definition, it follows that $K(\text{CM}_n(\mathbf{d}))$ is a multigraph with a degree sequence \mathbf{d}' such that, defining n'_i as the number of vertices of degree i in \mathbf{d}' , we have $n'_2 = 0$, $n'_i = n_i$ for all $i \neq 2$. Moreover, in the next lemma we prove that $K(\text{CM}_n(\mathbf{d}))$ is distributed as $\text{CM}_{n-n_2}(\mathbf{d}')$.

Lemma 4.2. *Consider the graph $\text{CM}_n(\mathbf{d})$ with any arbitrary degree sequence \mathbf{d} and its kernel $K(\text{CM}_n(\mathbf{d}))$. Then*

$$K(\text{CM}_n(\mathbf{d})) \stackrel{d}{=} \text{CM}_{n-n_2}(\mathbf{d}'),$$

where \mathbf{d}' is such that $n'_2 = 0$, $n'_i = n_i$ for all $i \neq 2$.

Proof. We show that however we choose a half-edge e_1 in $\mathcal{L}_{\neq 2}$, e_1 is paired in $K(\text{CM}_n(\mathbf{d}))$ with another half-edge e_2 chosen uniformly in $\mathcal{L}_{\neq 2}$. We do so through an algorithm that sequentially builds $\text{CM}_n(\mathbf{d})$. Define the time variable of the algorithm as a pair (i, j) , $i \geq 1, j \geq -1$, and the sets of unexplored half-edges at any time, $\mathcal{U}(i, j)$. We describe in the following pseudocode how the algorithm can sample the edge set of $K(\text{CM}_n(\mathbf{d}))$ as a set of pairs of half edges $(\{e_1(i), e_2(i)\})_{i=1}^{\ell_{\neq 2}/2}$, while building $\text{CM}_n(\mathbf{d})$ at the same time.

Initialize Set $i = 1, j = -1$. Set $\mathcal{U}(1, -1) = \mathcal{L}$.

- Step** (a) **If** $\mathcal{L}_{\neq 2} \cap \mathcal{U}(i, -1) = \emptyset$, pair the remaining half-edges in $\mathcal{U}(i, -1)$ uniformly at random and terminate the algorithm.
Else, choose $e \in \mathcal{L}_{\neq 2} \cap \mathcal{U}(i, -1)$ according to any arbitrary rule. Set $j = 0, \mathcal{U}(i, 0) = \mathcal{U}(i, -1) \setminus \{e\}$ and $e_1(i) = e$
- (b) Choose a half-edge e' uniformly at random in $\mathcal{U}(i, j)$ and pair it with e
- (c) **If** $e' \in \mathcal{L}_{\neq 2}$, then fix $e_2(i) = e'$ (i.e., $\{e, e'\}$ is an edge of the kernel), increase i by 1, set $j = -1$ and $\mathcal{U}(i + 1, -1) = \mathcal{U}(i, j) \setminus \{e'\}$. Go back to Step (a).
Else, set e equal to the other half-edge e'' incident to the same vertex as e' . Increase j by 1 and set $\mathcal{U}(i, j + 1) = \mathcal{U}(i, j) \setminus \{e', e''\}$. Go back to Step (b).

This algorithm produces the edge set of $K(\text{CM}_n(\mathbf{d}))$ as the set of all the couples $\{e_1(i), e_2(i)\}$, since every time a half edge $e_1(i)$ is paired with a vertex v of degree 2, such vertex is removed, and $e_1(i)$ is paired with the other half-edge connected to v , exactly as in the construction of $K(\text{CM}_n(\mathbf{d}))$. Moreover, every time the edge e' is chosen uniformly over $\mathcal{U}(i, j)$, and consequently, conditioning on $e' \in \mathcal{L}_{\neq 2}$, its distribution is uniform over the set

$$\mathcal{U}(i, j) \cap \mathcal{L}_{\neq 2} = \mathcal{L}_{\neq 2} \setminus \left(\{e_1(i)\} \cup \bigcup_{h < i} \{e_1(h), e_2(h)\} \right).$$

From this we obtain that the pairing $(\{e_1(i), e_2(i)\})_{i=1}^{\ell_{\neq 2}/2}$ is a uniform pairing over $\mathcal{L}_{\neq 2}$, and thus $K(\text{CM}_n(\mathbf{d})) \stackrel{d}{=} \text{CM}_{n-n_2}(\mathbf{d}')$. □

We next define the usual exploration process of the configuration model (see e.g. [17, Section 1]) that we are going to use multiple times in the paper.

Definition 4.3 (Exploration of the Configuration Model). *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 run the following process:*

Initialize *Pick a vertex $v \in [n]$ according to any arbitrary (deterministic or stochastic) rule and set all its half-edges as active. Set all the other half-edges as neutral.*

- Step** *At each step t , pick a half-edge $e_1(t)$ in \mathcal{A}_t uniformly at random, and pair it with another half-edge $e_2(t)$ chosen uniformly at random in $\mathcal{A}_t \cup \mathcal{N}_t$. Set $e_1(t), e_2(t)$ as dead.
If $e_2(t) \in \mathcal{N}_t$, then find the vertex $v(e_2(t))$ incident to $e_2(t)$ and activate all its other half-edges.*

Terminate *If $\mathcal{A}_t = \emptyset$, terminate the process.*

We use this algorithm to compute the number of vertices that are contained in cycle components; the same proof works for both the lower and upper almost-two regular random graphs:

Proposition 4.4. *Consider a sequence satisfying Condition 2.1 or 2.4. Define $\mathbf{C}(n)$ as the number of vertices of $\text{CM}_n(\mathbf{d})$ in cycle components. Then*

$$\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{C}(n) \geq tn_2/\ell_{\neq 2}) = 0.$$

Proof. We run the exploration process described in Definition 4.3 starting from a uniformly chosen vertex of degree 2. We run the process until the first of the two following stopping times:

$$T_{\neq 2} = \min\{t : e_2(t) \in \mathcal{L}_{\neq 2}\}, \quad T_C = \min\{t : e_2(t) \in \mathcal{A}_t\}.$$

It is impossible that $T_{\neq 2} = T_C$ because for every $t \leq T_{\neq 2}$, $\mathcal{A}_t \cap \mathcal{L}_{\neq 2} = \emptyset$. Moreover, v is in a cycle component, if and only if $T_C < T_{\neq 2}$. We know that for every $t \leq \min\{T_{\neq 2}, T_C\}$, $|\mathcal{A}_t| = 2$ and $|\mathcal{L}_{\neq 2} \cap \mathcal{N}_t| = \ell_{\neq 2}$. Consequently we have, for every $t \geq 1$

$$\mathbb{P}(T_{\neq 2} = t \mid \min\{T_{\neq 2}, T_C\} > t - 1) = \ell_{\neq 2} \mathbb{P}(T_C = t \mid \min\{T_{\neq 2}, T_C\} > t - 1),$$

so that

$$\mathbb{P}(T_{\neq 2} = t, T_C > T_{\neq 2}) = \ell_{\neq 2} \mathbb{P}(T_C = t, T_{\neq 2} > T_C),$$

and thus

$$\begin{aligned} \mathbb{P}(T_{\neq 2} > T_C) &= \sum_{t \geq 1} \mathbb{P}(\min\{T_{\neq 2}, T_C\} = t) \mathbb{P}(T_{\neq 2} > T_C \mid \min\{T_{\neq 2}, T_C\} = t) \\ &= \sum_{t \geq 1} \mathbb{P}(\min\{T_{\neq 2}, T_C\} = t) \frac{1}{\ell_{\neq 2} + 1} = \frac{1}{\ell_{\neq 2} + 1}. \end{aligned}$$

Consequently,

$$\mathbb{E}[\mathbf{C}(n)] = \sum_{v \in \mathcal{N}_2} \mathbb{P}(v \text{ is in a cycle component}) = n_2 \mathbb{P}(T_{\neq 2} > T_C) = \frac{n_2}{\ell_{\neq 2} + 1}.$$

We thus write, by the first moment method,

$$\mathbb{P}(\mathbf{C}(n) > tn_2/\ell_{\neq 2}) \leq \frac{n_2}{\ell_{\neq 2} + 1} \frac{\ell_{\neq 2}}{tn_2} < \frac{1}{t},$$

and the claim follows. □

We can now prove that under Condition 2.1, the largest component contains w.h.p. all the vertices of degree at least 3, and is of size $n - o_{\mathbb{P}}(n)$.

Proposition 4.5. *Consider a sequence $\text{CM}_n(\mathbf{d})$ that satisfies Condition 2.1. Then as $n \rightarrow \infty$,*

$$\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(n - |\mathcal{C}_{\max}| \geq tn_2/\ell_{\neq 2}) = 0.$$

Moreover, $[n] \setminus \mathcal{C}_{\max}$ contains only vertices of degree 2 w.h.p.

Proof. Every two vertices v, w such that $d_v, d_w \neq 2$ are in the same component in $\text{CM}_n(\mathbf{d})$ if and only if they are in the same component in $K(\text{CM}_n(\mathbf{d}))$. Thus, if $K(\text{CM}_n(\mathbf{d}))$ is connected, there exists a component $\tilde{\mathcal{C}}$ of $\text{CM}_n(\mathbf{d})$ such that $[n] \setminus \tilde{\mathcal{C}}$ contains only vertices of degree 2. We will show that said component exists and takes up the vast majority of the graph w.h.p.

By Lemma 4.2, $K(\text{CM}_n(\mathbf{d}))$ is distributed as $\text{CM}_{n-n_2}(\mathbf{d}')$, where \mathbf{d}' has minimum degree at least 3. Consequently, $\text{CM}_{n-n_2}(\mathbf{d}')$ is connected w.h.p. by the main theorem of [19]. If $K(\text{CM}_n(\mathbf{d}))$ is connected, the number of vertices outside the largest component is thus at most the number of vertices in cycle components. We thus write, for every $t \geq 1$

$$\begin{aligned} \mathbb{P}(n - |\mathcal{C}_{\max}| > tn_2/\ell_{\neq 2}) &\leq \mathbb{P}(\mathbf{C}(n) > tn_2/\ell_{\neq 2}) + \mathbb{P}(K(\text{CM}_n(\mathbf{d})) \text{ is disconnected}) \\ &< \frac{1}{t} + o(1). \end{aligned}$$

□

This is enough to complete the proof of (2.1). In order to prove (2.2), we need to analyze the distribution of the number and size of the cycle components.

From the previous analysis we see that w.h.p. all the components outside the giant are cycles, and that

$$\# \text{vertices in cycle components} = O_{\mathbb{P}}\left(\frac{n_2}{\ell_{\neq 2}}\right).$$

We know that if we remove the giant from a configuration model, the rest of the graph is distributed as a configuration model with the degree sequence of the remaining vertices. By Proposition 4.5 in this case what is left is w.h.p. a random 2-regular graph.

In a random 2-regular graph, we know that $|\mathcal{C}_{\max}|/n$ converges in distribution to a random variable with no points with positive mass. Thus, we expect the second largest component of $\text{CM}_n(\mathbf{d})$ (i.e. w.h.p. its largest cycle component) to be of size $O_{\mathbb{P}}\left(\frac{n_2}{\ell_{\neq 2}}\right)$ and to have a random size. To prove this we analyse the distribution of the number of cycle components that are larger than $an_2/\ell_{\neq 2}$ for any fixed constant $a \in (0, \infty)$.

Define $C_n(k)$ as the number of cycle components of size k in $\text{CM}_n(\mathbf{d})$. We will next analyse the random variables

$$S_n^{(a)}(k) := \begin{cases} 0 & \text{if } k \leq an_2/\ell_{\neq 2}, \\ \sum_{j=an_2/\ell_{\neq 2}}^k C_n(j) & \text{if } k > an_2/\ell_{\neq 2}. \end{cases} \tag{4.1}$$

In particular we are interested in proving the following proposition, from which we can deduce (2.2):

Proposition 4.6. *Consider a sequence $\text{CM}_n(\mathbf{d})$ satisfying $\ell_{\neq 2} \rightarrow \infty$ and $\ell_{\neq 2}/\ell \rightarrow 0$ and the associated processes $(S_n^{(a)}(k))_{k>0}$ as defined in (4.1). Then for every $a < t < \infty$*

$$S_n^{(a)}(tn_2/\ell_{\neq 2}) \xrightarrow{d} \text{Poi}\left(\int_a^t \lambda(r)dr\right),$$

where $\lambda(t) = \frac{e^{-t/2}}{2t}$.

Proof. We prove the result through the method of moments. We define the factorial moments

$$(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h = S_n^{(a)}(tn_2/\ell_{\neq 2})(S_n^{(a)}(tn_2/\ell_{\neq 2}) - 1) \cdots (S_n^{(a)}(tn_2/\ell_{\neq 2}) - h + 1).$$

To prove the Poisson convergence we need to prove that

$$\mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h] \rightarrow \left(\int_a^t \lambda(r)dr\right)^h,$$

for every $t > a$; $h \in \mathbb{N}$ (see e.g. [10, Section 2.1]). We define, for every $j \leq n_2$,

$$\mathcal{C}(j) := \binom{\mathcal{N}_2}{j},$$

i.e., $\mathcal{C}(j)$ is the set of all possible sets of vertices that can be arranged in a cycle of length j . Given an element $A \in \mathcal{C}(j)$, we define C_A as the event that the vertices in A form a cycle and I_{C_A} as its indicator variable. We now write

$$\mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h] = \sum_{j_1, \dots, j_h \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}}],$$

where \sum^* indicates sum over distinct elements, that is, that we cannot have $A_a = A_b$ for any $a \neq b$, $a, b \leq j$.

We need to prove that

$$\begin{aligned} & \max_{j_1, \dots, j_h \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}}] \left(\frac{n_2}{\ell_{\neq 2}}\right)^h \\ & \quad - \lambda(j_1 n_2/\ell_{\neq 2}) \lambda(j_2 n_2/\ell_{\neq 2}) \cdots \lambda(j_h n_2/\ell_{\neq 2}) \rightarrow 0. \end{aligned} \tag{4.2}$$

We rewrite it as

$$\begin{aligned} & \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}}] \\ & = \sum_{A_1 \in \mathcal{C}(j_1)} \cdots \sum_{A_{h-1} \in \mathcal{C}(j_{h-1}) \setminus \{A_1, \dots, A_{h-2}\}} \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_{h-1}}}] \\ & \quad \times \sum_{A_h \in \mathcal{C}(j_h) \setminus \{A_1, \dots, A_{h-1}\}} \mathbb{E}[I_{C_{A_h}} \mid C_{A_1}, \dots, C_{A_{h-1}}]. \end{aligned} \tag{4.3}$$

We note immediately that, if there exist distinct $i, g \leq h$ such that $A_i \cap A_l \neq \emptyset$, then $I_{C_{A_i}} I_{C_{A_g}} = 0$ deterministically. Consequently, we can restrict the sum to the case in which the sets A_1, \dots, A_h are all mutually disjoint. Define $m_i := \sum_{l=1}^{i-1} j_l$, that is, as the number of vertices used to create the first $i - 1$ cycles, all of which have degree 2. We note that for all $1 \leq i \leq h$, as long as A_i does not overlap with any of the previous sets, $\mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}]$ is the probability that the vertices in A_i form a cycle component in a configuration model $\text{CM}_{n-m_i}(\tilde{\mathbf{d}})$, with $\tilde{\mathbf{d}}$ satisfying $\tilde{n}_2 = n_2 - m_i$ and $\tilde{n}_l = n_l$ for all $l \neq 2$, where \tilde{n}_k is the number of vertices of degree k in $\tilde{\mathbf{d}}$. We further note that $\tilde{\mathbf{d}}$ does not depend on the precise choice of A_1, \dots, A_{i-1} , as long as $A_g \in \mathcal{C}(j_g)$ for every g . Thus, for all possible disjoint collections A_1, \dots, A_h , for every $i \leq h$,

$$\begin{aligned} & \mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}] \\ &= \frac{2j_i - 2}{\ell - 2m_i - 1} \frac{2j_i - 4}{\ell - 2m_i - 3} \cdots \frac{2}{\ell - 2m_i - 2j_i + 3} \frac{1}{\ell - 2m_i - 2j_i + 1} \\ &= 2^{j_i-1} (j_i - 1)! \prod_{g=0}^{j_i-1} \frac{1}{\ell - 2m_i - 2g - 1}. \end{aligned}$$

There exists $\binom{n_2 - m_i}{j_i}$ distinct choices for A_i that do not overlap with any A_k for any $k < i$, so we obtain

$$\begin{aligned} \sum_{A_i \in \mathcal{C}(j_i)} \mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}] &= \binom{n_2 - m_i}{j_i} 2^{j_i-1} (j_i - 1)! \prod_{g=0}^{j_i-1} \frac{1}{\ell - 2m_i - 2g - 1} \\ &= \frac{1}{2j_i} \prod_{g=0}^{j_i-1} \frac{2(n_2 - m_i - g)}{\ell - 2m_i - 2g - 1}. \end{aligned}$$

We use that, for every $g \geq 0$,

$$\frac{2(n_2 - m_i - g)}{\ell - 2m_i - 2g - 1} \leq \frac{2n_2}{\ell - 1}.$$

For a lower bound we write, recalling that $m_i + g \leq ht\ell/\ell_{\neq 2}$, for all $i \leq h, g \leq j_1 - 1$

$$\begin{aligned} \frac{2n_2}{\ell} - \frac{2(n_2 - m_i - g)}{\ell - 2m_i - 2g - 1} &\leq \frac{2n_2}{\ell} - \frac{2n_2 - 2(m_i + g + 1)}{\ell - 2(m_i + g + 1)} \\ &= \frac{2n_2(\ell - 2(m_i + g + 1)) - \ell(2n_2 - 2(m_i + g + 1))}{\ell(\ell - 2(m_i + g + 1))} \\ &= \frac{2(m_i + g + 1)(\ell - 2n_2)}{\ell^2(1 - o(1))} \leq \frac{ht}{\ell}. \end{aligned}$$

Thus we obtain that

$$\begin{aligned} \prod_{g=0}^{j_i-1} \frac{2(n_2 - m_i - g)}{\ell - 2m_i - 2g - 1} &\geq \left(\frac{2n_2 - ht}{\ell}\right)^{j_i} = \left(1 - \frac{\ell_{\neq 2} + ht}{\ell}\right)^{j_i} \\ &= \left(1 - \frac{\ell_{\neq 2} + ht}{\ell}\right)^{\frac{\ell}{\ell_{\neq 2} + ht} \frac{j_i(\ell_{\neq 2} + ht)}{\ell}}, \end{aligned}$$

and

$$\begin{aligned} \prod_{g=0}^{j_i-1} \frac{2(n_2 - m_i - g)}{\ell - 2m_i - 2g - 1} &\leq \left(\frac{2n_2}{\ell - 1}\right)^{j_i} = \left(1 - \frac{\ell_{\neq 2} - 1}{\ell - 1}\right)^{j_i} \\ &= \left(1 - \frac{\ell_{\neq 2} - 1}{\ell - 1}\right)^{\frac{\ell-1}{\ell_{\neq 2}-1} \frac{j_i(\ell_{\neq 2}-1)}{\ell-1}}. \end{aligned}$$

We thus write

$$\begin{aligned} -\frac{j_i \ell_{\neq 2}}{\ell - 1} (1 + o(1)) &\leq \log \left(2j_i \sum_{A_i \in \mathcal{C}(j_i)} \mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}] \right) \\ &\leq -\frac{j_i(\ell_{\neq 2} + ht)}{\ell} (1 + o(1)), \end{aligned}$$

where the $o(1)$ term is independent of j_i . We can thus conclude that

$$\max_{j_i \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \max_{A_1, \dots, A_{i-1}} \left| 2j_i \sum_{A_i \in \mathcal{C}(j_i)} \mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}] - e^{-j_i \ell_{\neq 2}/\ell} \right| \rightarrow 0.$$

Recall that

$$\frac{e^{-j_i \ell_{\neq 2}/(2n_2)}}{2j_i} = \lambda(j_i \ell_{\neq 2}/n_2) \frac{\ell_{\neq 2}}{n_2}.$$

We thus obtain,

$$\max_{j_i \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \max_{A_1, \dots, A_{i-1}} \left| \sum_{A_i \in \mathcal{C}(j_i)} \mathbb{E}[I_{C_{A_i}} \mid C_{A_1}, \dots, C_{A_{i-1}}] \frac{n_2}{\ell_{\neq 2}} - \lambda(j_i \ell_{\neq 2}/n_2) \right| \rightarrow 0. \tag{4.4}$$

Iterating the substitution of (4.4) in (4.3) for all the other $j_i, i < h$, we obtain (4.2). By the uniform convergence we obtain

$$\begin{aligned} &\sum_{j_1, \dots, j_h \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}}] \\ &= \left(\frac{\ell_{\neq 2}}{n_2}\right)^h (1 + o(1)) \sum_{j_1, \dots, j_h \in \{an_2/\ell_{\neq 2}; tn_2/\ell_{\neq 2}\}}^* \lambda(j_1 \ell_{\neq 2}/n_2) \cdots \lambda(j_h \ell_{\neq 2}/n_2) \\ &\rightarrow \left(\int_a^t \lambda(r) dr\right)^h. \end{aligned}$$

This, by the method of moments, yields the claim. □

We next prove that, if we assume convergence of the third moment of the degree distribution, the same limit in distribution for the number of large components holds even if we condition on the graph being simple.

Proposition 4.7. *Consider a sequence $\text{CM}_n(\mathbf{d})$ satisfying $\ell_{\neq 2} \rightarrow \infty$ and $\ell_{\neq 2}/\ell \rightarrow 0$ and $\mathbb{E}[D_n^i] \rightarrow 2^i, i = 1, 2, 3$ and the associated processes $(S_n^{(a)}(k))_{k>0}$ as defined in (4.1). Then, for every $a < t < \infty$*

$$(S_n^{(a)}(tn_2/\ell_{\neq 2}) \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \xrightarrow{d} \text{Poi}\left(\int_a^t \lambda(r)dr\right).$$

Proof. We again use the method of factorial moments, in particular we need to prove that, for each $t > a, h \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h \mid \text{CM}_n(\mathbf{d}) \text{ is simple}] &= \frac{\mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h I_{\{\text{CM}_n(\mathbf{d}) \text{ is simple}\}}]}{\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is simple})} \\ &\rightarrow \left(\int_a^t \lambda(r)dr\right)^h, \end{aligned} \tag{4.5}$$

We know by the results in [1, Theorem 1.1] that

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is simple}) \rightarrow \exp\left\{-\frac{\mathbb{E}[2^2] - \mathbb{E}[2]}{2} - \frac{(\mathbb{E}[2^2] - \mathbb{E}[2])^2}{4}\right\} = e^{-1}. \tag{4.6}$$

We rewrite the expression in the numerator on the left-hand side of (4.5) as

$$\begin{aligned} &\mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h I_{\{\text{CM}_n(\mathbf{d}) \text{ is simple}\}}] \\ &= \sum_{j_1, \dots, j_h \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}} I_{\{\text{CM}_n(\mathbf{d}) \text{ is simple}\}}]. \end{aligned}$$

We now compute

$$\sum_{\substack{j_1, \dots, j_h \in \\ [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]}} \sum_{A_1 \in \mathcal{C}(j_1), \dots, A_h \in \mathcal{C}(j_h)}^* \mathbb{E}[I_{C_{A_1}} \cdots I_{C_{A_h}}] \mathbb{E}[I_{\{\text{CM}_n(\mathbf{d}) \text{ is simple}\}} \mid C_{A_1}, \dots, C_{A_h}]. \tag{4.7}$$

Conditionally on C_{A_1}, \dots, C_{A_h} , as long as A_1, \dots, A_h are all disjoint and thus $I_{C_{A_1}} \cdots I_{C_{A_h}}$ is not 0 almost surely, the probability that $\text{CM}_n(\mathbf{d})$ is simple is the same as the probability that a random graph $\widehat{\text{CM}}_n(\mathbf{d}, j_1, \dots, j_{h_1})$, whose degree sequence is obtained from the original one by removing the $j_1 + j_2 + \dots + j_h$ vertices of degree 2 appearing in C_{A_1}, \dots, C_{A_h} is simple. This is because for C_{A_1}, \dots, C_{A_h} to happen, it is necessary that no self-loops or multi-edges appear among the vertices in A_1, \dots, A_h . We note that $j_1 + j_2 + \dots + j_h \leq htn_2/\ell_{\neq 2}$. We define \tilde{D}_n as the degree of a uniformly chosen vertex from the degree sequence \mathbf{d} obtained removing $htn_2/\ell_{\neq 2}$ vertices from \mathbf{d} . We obtain that for the degree D'_n of a uniformly chosen vertex in $\widehat{\text{CM}}_n(\mathbf{d}, j_1, \dots, j_{h_1})$, $D_n \preceq D'_n \preceq \tilde{D}_n$, where the relation \preceq indicates stochastic domination. Since $htn_2/\ell_{\neq 2} = o(n)$, it follows that $D' \xrightarrow{d} 2$, and $\mathbb{E}[D'] \rightarrow 2, \mathbb{E}[D'^2] \rightarrow 4$ and $\mathbb{E}[D'^3] \rightarrow 8$, uniformly in j_1, j_2, \dots, j_{h_1} and thus, by [1, Theorem 1.1]

$$\lim_{n \rightarrow \infty} \max_{j_1, \dots, j_h \in [an_2/\ell_{\neq 2} \dots tn_2/\ell_{\neq 2}]} |\mathbb{P}(\widehat{\text{CM}}_n(\mathbf{d}, j_1, \dots, j_{h_1}) \text{ is simple}) - e^{-1}| = O(1/\ell).$$

Substituting this result in (4.7) and applying Proposition 4.6 we obtain

$$\mathbb{E}[(S_n^{(a)}(tn_2/\ell_{\neq 2}))_h I_{\{\text{CM}_n(\mathbf{d}) \text{ is simple}\}}] \rightarrow e^{-1} \left(\int_a^t \lambda(r) dr \right)^h.$$

From this using (4.6) we obtain (4.5) and the claim follows by the method of moments. \square

We now put together the results obtained in this section to prove Theorems 2.2 and 2.3.

Proof of Theorem 2.2. Proposition 4.5 directly implies (2.1), so what we are left to prove is (2.2).

Define $\mathcal{C}_{i-1}^{\text{cycle}}$ as the $(i - 1)$ -th largest cycle component. By Proposition 4.5

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{C}_{i-1}^{\text{cycle}} = \mathcal{C}_i) = 1.$$

By Proposition 4.6, for any $a \in (0, \infty)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{i-1}^{\text{cycle}}| \leq an_2/\ell_{\neq 2}) &\leq \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(S_n^{(a)}(tn_2/\ell_{\neq 2}) \leq i - 2) \\ &= \lim_{t \rightarrow \infty} \mathbb{P}\left(\text{Poi}\left(\int_a^t \lambda(r) dr\right) \leq i - 2\right). \end{aligned} \tag{4.8}$$

For the matching lower bound we write, using Proposition 4.4,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{i-1}^{\text{cycle}}| \leq an_2/\ell_{\neq 2}) &\geq \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(S_n^{(a)}(tn_2/\ell_{\neq 2}) \leq i - 2) \\ &\quad - \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{C}(n) \geq tn_2/\ell_{\neq 2}) \\ &= \lim_{t \rightarrow \infty} \mathbb{P}\left(\text{Poi}\left(\int_a^t \lambda(r) dr\right) \leq i - 2\right) + o(1). \end{aligned} \tag{4.9}$$

Combining (4.8) and (4.9) and the fact that $n_2 = n(1 - o(1))$, the claim follows. \square

Proof of Theorem 2.3. From Proposition 4.5 and (4.6) we obtain:

$$\begin{aligned} &\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(n - |\mathcal{C}_{\max}| \geq tn_2/\ell_{\neq 2} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ &\leq \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{P}(n - |\mathcal{C}_{\max}| \geq tn_2/\ell_{\neq 2})}{\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is simple})} \rightarrow 0. \end{aligned} \tag{4.10}$$

This concludes the proof that (2.1) holds, and furthermore from it follows that,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{C}_{i-1}^{\text{cycle}} = \mathcal{C}_i \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) = 1.$$

By Proposition 4.7, for any $a \in (0, \infty)$,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{i-1}^{\text{cycle}}| \leq an_2/\ell_{\neq 2} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ &\leq \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(S_n^{(a)}(tn_2/\ell_{\neq 2}) \leq i - 2 \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ &= \lim_{t \rightarrow \infty} \mathbb{P}\left(\text{Poi}\left(\int_a^t \lambda(r) dr\right) \leq i - 2\right). \end{aligned} \tag{4.11}$$

For the matching lower bound we write, using (4.10) and Proposition 4.4,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{i-1}^{cycle}| \leq an_2/\ell_{\neq 2} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ & \geq \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(S_n^{(a)}(tn_2/\ell_{\neq 2}) \leq i - 2 \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \\ & \quad - \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{C}(n) \geq tn_2/\ell_{\neq 2} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}) \tag{4.12} \\ & = \lim_{t \rightarrow \infty} \mathbb{P}\left(\text{Poi}\left(\int_a^t \lambda(r)dr\right) \leq i - 2\right) + o(1). \end{aligned}$$

Combining (4.11) and (4.12) and the fact that $n_2 = n(1 - o(1))$, the claim follows. □

5 Proof of Theorem 2.5

In this section we analyze the lower almost-2-regular graph, and prove Theorem 2.5. Since Condition 2.4 allow only for vertices of degree 1 and 2, we know that all the components in $\text{CM}_n(\mathbf{d})$ are either paths (components made of 2 vertices of degree 1 connected by vertices of degree 2) or cycles. Here we consider a vertex of degree 2 with a self-loop as a cycle of length 1, and a pair of double edges between a pair of vertices of degree 2 is considered to be a cycle of length 2. Thus

$$|\mathcal{C}_{\max}| = \max\{|\mathcal{C}_{\max}^{cycle}|, |\mathcal{C}_{\max}^{line}|\},$$

where $\mathcal{C}_{\max}^{cycle}$ is the largest cycle component and $\mathcal{C}_{\max}^{line}$ is the largest path component. Moreover, define \mathcal{C}_j^{line} as the j -th largest path component.

We next prove that the size of the largest path components concentrates and it is w.h.p. larger than that of $\mathcal{C}_{\max}^{cycle}$, which we can estimate using Proposition 4.4.

Lemma 5.1. *Consider $\text{CM}_n(\mathbf{d})$ satisfying Conditions 2.4. Then, for every $j \in \mathbb{N}$,*

$$\frac{n_1 |\mathcal{C}_j^{line}|}{n \log n_1} \xrightarrow{\mathbb{P}} 2.$$

Proof. We start by proving the upper bound. For every $\alpha \in (0, \infty)$ we define

$$N(1, \alpha) := \#\{v \in \mathcal{N}_1 : |\mathcal{C}(v)| > \alpha n \log(n_1)/n_1\},$$

where for every $v \in [n]$, $\mathcal{C}(v)$ is the connected component that contains v . To find bounds on $N(1, \alpha)$ we run the exploration from Definition 4.3 starting from a uniformly chosen vertex $v \in \mathcal{N}_1$. Since there are no vertices of degree larger than 2, at every step t , $|\mathcal{A}_t| \in \{0, 1\}$, and the exploration of $\mathcal{C}(v)$ ends as soon as $e_2(t) \in \mathcal{L}_{\neq 2}$, i.e., the first time the exploration finds another vertex of degree 1. Thus, the only way the process can survive up to time $\alpha n \log(n_1)/n_1$ is that $e_2(t) \notin \mathcal{L}_{\neq 2}$, for all

$t \leq \alpha n \log(n_1)/n_1$. We write the probability for this to happen as

$$\begin{aligned} \pi_\alpha &:= \mathbb{P}(|\mathcal{C}(v)| > \alpha n \log(n_1)/n_1) = \prod_{t=0}^{\alpha n \log(n_1)/n_1} \left(\frac{2(n_2 - t)}{\ell - 2t - 1} \right) \\ &= \prod_{k=0}^{\alpha n \log(n_1)/n_1} \left(1 - \frac{n_1 - 1}{\ell - 2t - 1} \right). \end{aligned}$$

We bound this quantity by

$$\begin{aligned} \pi_\alpha &\leq \left(1 - \frac{n_1 - 1}{\ell} \right)^{\alpha n \log(n_1)/n_1} = e^{-\alpha \log n_1/2(1+o(1))} \\ \pi_\alpha &\geq \left(1 - \frac{n_1 - 1}{\ell - 2\alpha n \log(n_1)/n_1} \right)^{\alpha n \log(n_1)/n_1} = e^{-\alpha \log n_1/2(1+o(1))}. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{E}[N(1, \alpha)] &= \sum_{v \in \mathcal{N}_1} \mathbb{P}(|\mathcal{C}(v)| > \alpha n \log(n_1)/n_1) \\ &= n_1 \pi_\alpha = n_1 e^{-\alpha \log n_1/2(1+o(1))} = n_1^{(2-\alpha+o(1))/2(1+o(1))}. \end{aligned} \tag{5.1}$$

By the first moment method,

$$\mathbb{P}\left(\frac{n_1 |\mathcal{C}_{\max}^{Line}|}{n \log n_1} \geq \alpha\right) \leq \mathbb{E}[N(1, \alpha)]/2,$$

so that, for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{n_1 |\mathcal{C}_{\max}^{Line}|}{n \log n_1} \geq 2 + \varepsilon\right) = 0.$$

Next we prove a sharp lower bound on the size of $|\mathcal{C}_j^{Line}|$ for every $j \geq 1$. Using the Chebyshev inequality we write

$$\mathbb{P}\left(|\mathcal{C}_j^{Line}| \leq \frac{\alpha n \log n_1}{n_1}\right) = \mathbb{P}(N(1, \alpha) \leq 2j) \leq \frac{\mathbb{E}[N(1, \alpha)^2] - \mathbb{E}[N(1, \alpha)]^2}{(\mathbb{E}[N(1, \alpha)] - 2j)^2}. \tag{5.2}$$

By (5.1), we know that for every $\alpha < 2$, $\mathbb{E}[N(1, \alpha)] \rightarrow \infty$. Thus, it is enough to prove that for every $\alpha < 2$,

$$\frac{\mathbb{E}[N(1, \alpha)^2]}{\mathbb{E}[N(1, \alpha)]^2} \rightarrow 1.$$

We write

$$\begin{aligned} \mathbb{E}[N(1, \alpha)^2] &= \sum_{v, w \in \mathcal{N}_1} \mathbb{P}(|\mathcal{C}(v)|, |\mathcal{C}(w)| > \alpha n \log(n_1)/n_1) \\ &= \sum_{v \in \mathcal{N}_1} \mathbb{P}(|\mathcal{C}(v)| > \alpha n \log(n_1)/n_1) \\ &\quad + \sum_{v, w \in \mathcal{N}_1, v \neq w} \left(\mathbb{P}(|\mathcal{C}(v)|, |\mathcal{C}(w)| \geq \alpha n \log(n_1)/n_1, \mathcal{C}(v) \neq \mathcal{C}(w)) \right. \\ &\quad \left. + \mathbb{P}(|\mathcal{C}(v)| \geq \alpha n \log(n_1)/n_1, \mathcal{C}(v) = \mathcal{C}(w)) \right). \end{aligned} \tag{5.3}$$

We know that for every $v \in \mathcal{N}_1$, almost surely $|(\mathcal{N}_1 \cap \mathcal{C}(v)) \setminus \{v\}| = 1$, so that,

$$\begin{aligned} \sum_{v,w \in \mathcal{N}_1, v \neq w} \mathbb{P}(|\mathcal{C}(v)| \geq \alpha n \log(n_1)/n_1, \mathcal{C}(v) = \mathcal{C}(w)) \\ = \sum_{v \in \mathcal{N}_1} \mathbb{E}[|(\mathcal{N}_1 \cap \mathcal{C}(v)) \setminus \{v\}| I_{\{|\mathcal{C}(v)| \geq \alpha n \log(n_1)/n_1\}}] \\ = \sum_{v \in \mathcal{N}_1} \mathbb{P}(|\mathcal{C}(v)| > \alpha n \log(n_1)/n_1) = \pi_\alpha n_1. \end{aligned} \tag{5.4}$$

To bound the probability that v and w are in distinct large components, we now run two copies of the exploration process from Definition 4.3, starting from two different vertices v and w in \mathcal{N}_1 . We first explore starting from v , and we let the exploration run up to time $\alpha n \log(n_1)/n_1$. If the exploration has survived, then we know that $|\mathcal{C}(v)| > \alpha n \log(n_1)/n_1$, and that $\mathcal{A}_{\alpha n \log(n_1)/n_1} = \{e'\}$ for some half-edge $e' \notin \mathcal{L}_{\neq 2}$. In this case, we start running a new exploration from w . We know that $|\mathcal{C}(w)| > \alpha n \log(n_1)/n_1$, $\mathcal{C}(w) \neq \mathcal{C}(v)$ if the exploration starting from w survives up to time $\alpha n \log(n_1)/n_1$ finding neither e' nor any half-edge in $\mathcal{L}_{\neq 2}$.

We bound this probability by

$$\prod_{k=0}^{\alpha n \log(n_1)/n_1} \left(1 - \frac{n_1 - 2}{\ell - 2k - 2\alpha n \log(n_1)/n_1}\right) \leq \prod_{k=0}^{\alpha n \log(n_1)/n_1} \left(1 - \frac{n_1 - 1}{\ell - 2k - 1}\right) = \pi_\alpha,$$

so that

$$\mathbb{P}(|\mathcal{C}(v)|, |\mathcal{C}(w)| \geq \alpha n \log(n_1)/n_1, \mathcal{C}(v) \neq \mathcal{C}(w)) \leq \pi_\alpha^2. \tag{5.5}$$

We thus obtain, substituting (5.1), (5.4) and (5.5) into (5.3)

$$\mathbb{E}[N(1, \alpha)^2] \leq 2\pi_\alpha n_1 + \pi_\alpha^2 n_1 (n_1 - 1).$$

We further compute, recalling that $n_1 \pi_\alpha = \mathbb{E}[N(1, \alpha)]$,

$$\mathbb{E}[N(1, \alpha)^2] \leq 2\mathbb{E}[N(1, \alpha)] + \mathbb{E}[N(1, \alpha)]^2.$$

If $\alpha = 2 - \varepsilon$, then, by (5.1), $\mathbb{E}[N(1, \alpha)] \rightarrow \infty$, so we obtain

$$\mathbb{E}[N(1, \alpha)^2] = \mathbb{E}[N(1, \alpha)]^2(1 + o(1)).$$

Consequently, by (5.2), for every $j \geq 1$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{n_1 |\mathcal{C}_j^{line}|}{n \log n_1} \leq 2 - \varepsilon\right) = \lim_{n \rightarrow \infty} \mathbb{P}(N(1, 2 - \varepsilon) \leq 2j) = 0.$$

□

We can now finally prove Theorem 2.5.

Proof of Theorem 2.5. By Lemma 5.1, for every $\alpha < 2$, $j \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(|\mathcal{C}_j| \geq \frac{\alpha n \log n_1}{n_1}\right) \geq \lim_{n \rightarrow \infty} \mathbb{P}\left(|\mathcal{C}_j^{Line}| \geq \frac{\alpha n \log n_1}{n_1}\right) = 1.$$

On the other hand, by Proposition 4.4 and Lemma 5.1, for every $\alpha > 2$, $j \in \mathbb{N}$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(|\mathcal{C}_j| \geq \frac{\alpha n \log n_1}{n_1}\right) \\ & \leq \lim_{n \rightarrow \infty} \left(\mathbb{P}\left(|\mathcal{C}_{\max}^{Line}| \geq \frac{\alpha n \log n_1}{n_1}\right) + \mathbb{P}\left(|\mathcal{C}_{\max}^{cycle}| \geq \frac{\alpha n \log n_1}{n_1}\right) \right) = 0. \end{aligned}$$

By the upper and lower bound, we obtain the claim. □

From Theorem 2.5, using the fact that $\text{CM}_n(\mathbf{d})$ is simple with non-vanishing probability, we obtain Corollary 2.6

Proof of Corollary 2.6. Under Condition 2.4 we obtain that, recalling that D is the distribution of the degree of a uniformly chosen random vertex, $D \xrightarrow{d} \mathfrak{2}$ and further $\mathbb{E}[D] \rightarrow \mathbb{E}[\mathfrak{2}] = 2$, $\mathbb{E}[D^2] \rightarrow \mathbb{E}[\mathfrak{2}^2] = 4$. From this follows (see e.g [14], [1]) that

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is simple}) \rightarrow \exp\left\{-\frac{\mathbb{E}[\mathfrak{2}^2] - \mathbb{E}[\mathfrak{2}]^2}{2} - \frac{(\mathbb{E}[\mathfrak{2}^2] - \mathbb{E}[\mathfrak{2}]^2)^2}{4}\right\} = e^{-1}.$$

From this, and Theorem 2.5, it follows that for every $j \in \mathbb{N}$ and $\varepsilon > 0$:

$$\begin{aligned} & \mathbb{P}\left(|\mathcal{C}_j| \geq \frac{(2 + \varepsilon)n \log n_1}{n_1} \vee |\mathcal{C}_j| \leq \frac{(2 - \varepsilon)n \log n_1}{n_1} \mid \text{CM}_n(\mathbf{d}) \text{ is simple}\right) \\ & \leq \frac{\mathbb{P}\left(|\mathcal{C}_j| \geq \frac{(2 + \varepsilon)n \log n_1}{n_1} \vee |\mathcal{C}_j| \leq \frac{(2 - \varepsilon)n \log n_1}{n_1}\right)}{\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is simple})} \rightarrow 0. \end{aligned}$$

□

Acknowledgments

The work in this paper is supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation-grant NETWORKS-024.002.003 and by the European Research Council (ERC) through Starting Grant Random Graph, Geometry and Convergence 639046.

References

- [1] O. ANGEL, R. VAN DER HOFSTAD AND C. HOLMGREN, Limit laws for self-loops and multiple edges in the configuration model, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 55 (2019), 1509–1530.
- [2] M. BEHRISCH, Component evolution in random intersection graphs, *Electron. J. Combin.* 14 (2007), #R17.
- [3] B. BOLLOBÁS, *Random graphs*, Cambridge Studies in Advanced Mathematics Vol. 73, Cambridge University Press, Cambridge, 2nd ed. (2001).
- [4] B. BOLLOBÁS, S. JANSON AND O. RIORDAN, The phase transition in inhomogeneous random graphs, *Random Structures Algorithms* 31 (2007), 3–122.
- [5] S. DHARA, R. VAN DER HOFSTAD, J. H. V. LEEUWAARDEN AND S. SEN, Critical window for the configuration model: finite third moment degrees, *Electron. J. Probab.* 22 (2017), #16.
- [6] S. DHARA, R. VAN DER HOFSTAD, J. H. V. LEEUWAARDEN AND S. SEN, Heavy-tailed configuration models at criticality, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56 (2020), 1515–1558.
- [7] P. ERDŐS AND A. RÉNYI, On random graphs I, *Publ. Math. Debrecen* 6 (1959), 290–297.
- [8] P. ERDŐS AND A. RÉNYI, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl* 5 (1960), 17–61.
- [9] L. FEDERICO AND R. VAN DER HOFSTAD, Critical window for connectivity in the configuration model, *Combin. Probab. Comput.* 26 (2017), 660–680.
- [10] R. VAN DER HOFSTAD, *Random graphs and complex networks, Vol. 1*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, (2017).
- [11] R. VAN DER HOFSTAD, S. JANSON AND M. LUCZAK, Component structure of the configuration model: Barely supercritical case, *Random Structures & Algorithms* 55 (2019), 3–55.
- [12] R. VAN DER HOFSTAD, M. J. LUCZAK AND J. SPENCER, The second largest component in the supercritical 2D Hamming graph, *Random Structures Algorithms* 36 (2010), 80–89.
- [13] S. JANSON, The largest component in a subcritical random graph with a power law degree distribution, *Ann. Appl. Probab.* 18 (2008), 1651–1668.
- [14] S. JANSON, The probability that a random multigraph is simple, *Combin. Probab. Comput.* 18 (2009), 205–225.

- [15] H. KESTEN AND Y. ZHANG, The probability of a large finite cluster in supercritical Bernoulli percolation, *Ann. Probab.* 18 (1990), 537–555.
- [16] T. ŁUCZAK, Sparse random graphs with a given degree sequence, in: “Random graphs”, Vol. 2 (Poznań, 1989), Wiley-Intersci. Publ., Wiley, New York, (1992), 165–182.
- [17] M. MOLLOY AND B. REED, A critical point for random graphs with a given degree sequence, in: *Proc. Sixth Int. Seminar on Random Graphs and Probabilistic Methods in Combin. and Comp. Sci*, “Random Graphs ’93” (Poznań, 1993) 6, (1995), 161–179.
- [18] B. PITTEL AND N. C. WORMALD, Counting connected graphs inside-out, *J. Combin. Theory Ser. B* 93 (2005), 127–172.
- [19] N. C. WORMALD, The asymptotic connectivity of labelled regular graphs, *J. Combin. Theory Ser. B* 31 (1981), 156–167.

(Received 14 Oct 2021; revised 15 May 2022, 24 Mar 2023)