A note on the component structure in random intersection graphs with tunable clustering

Andreas N. Lagerås*
Mathematical Sciences and Centre for Theoretical Biology,
Chalmers University of Technology and Göteborg University, 412 96 Gothenburg, Sweden

Mathias Lindholm†
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden

Submitted: Sep 10, 2007; Accepted: Apr 4, 2008; Published: Apr 10, 2008
Mathematics Subject Classification: 05C80

Abstract

We study the component structure in random intersection graphs with tunable clustering, and show that the average degree works as a threshold for a phase transition for the size of the largest component. That is, if the expected degree is less than one, the size of the largest component is a.a.s. of logarithmic order, but if the average degree is greater than one, a.a.s. a single large component of linear order emerges, and the size of the second largest component is at most of logarithmic order.

1 Introduction

The random intersection graph, denoted $\mathcal{G}_{m,p}^{(n)}$, with a set of vertices $\mathcal{V} = \{v_1, \ldots, v_n\}$ and a set of edges $\mathcal{E}$ is constructed from a bipartite graph $\mathcal{B}_{m,p}^{(n)}$ with two sets of vertices: $\mathcal{V}$, identical to those of $\mathcal{G}_{m,p}^{(n)}$, and $\mathcal{A} = \{a_1, \ldots, a_m\}$, which we call auxiliary vertices. Edges in $\mathcal{B}_{m,p}^{(n)}$ between vertices and auxiliary vertices are included independently with probability $p \in [0, 1]$. An edge between two vertices v_i and v_j in $\mathcal{G}_{m,p}^{(n)}$ is only present in $\mathcal{E}$ if both v_i and v_j are adjacent to some auxiliary vertex a_k in $\mathcal{B}_{m,p}^{(n)}$. Along the lines of Karoński et al. [5] we set $m := \lfloor \beta n \rfloor$ and $p := \gamma n^{-(1+\alpha)/2}$, where $\alpha, \beta, \gamma \geq 0$, to obtain an interesting graph structure and bounded average vertex degree. For random (multi)graphs, the vertex degree distribution is defined as the distribution of the degree, i.e. the number of adjacent

*supported by the Faculty of Science, Göteborg University.
†supported by the Swedish Foundation for Strategic Research (SSF).

edges, of a vertex chosen uniformly at random. As has been shown by Stark [7], the vertex degree distribution of the random intersection graph is highly dependent on the value of α, but as shown by Deijfen and Kets [3], the clustering is *tunable* only when $\alpha \equiv 1$. In a recent paper by Behrisch [2], the component structure of the random intersection graph is studied for $\alpha \neq 1$ and $\beta = 1$, and the aim of the present note is to describe the component structure when $\alpha = 1$. We will henceforth keep β and γ fixed and positive, and sometimes suppress the dependency on these parameters in the notation: $\mathcal{G}^{(n)}$.

2 The degree distribution

We define $D(m,n,p)$ to be a random variable with the vertex degree distribution of the graph $\mathcal{G}^{(n)}_{m,p}$. Stark [7, Thm. 1] showed that the distribution of $D(m,n,p)$ has the following generating function

$$g_{D(m,n,p)}(z) := \mathbb{E}\left[z^{D(m,n,p)}\right] = \sum_{j=0}^{n-1} \binom{n-1}{j} z^j (1-z)^{n-1-j} \left[1-p+p\,(1-p)^{n-1-j}\right]^m .$$

This distribution is from here onwards denoted $\mathrm{RIG}(m,n,p)$. Let us define a certain compound Poisson random variable Z by its generating function

$$g_Z(s) := \mathbb{E}\left[s^Z\right] = \exp\left\{\lambda'\left(e^{\lambda''(s-1)}-1\right)\right\},$$

and write $Z \in \mathrm{CPoisson}(\lambda', \lambda'')$. Here $\mathbb{E}[Z] = \lambda'\lambda''$. Another result by Stark [6, Thm. 2], here slightly generalised, is

Lemma 1. *If n' and n'' are functions of n such that $\lfloor \beta n \rfloor \geq n'$, $n'/n = \beta + o(1)$, $n \geq n''$, $n''/n = 1 + o(1)$, then $D(n', n'', \gamma/n) \xrightarrow{d} \mathrm{CPoisson}(\beta\gamma, \gamma)$ as $n \to \infty$.*

This can be shown by inspecting the generating functions. In particular

$$\begin{aligned}
\mathbb{E}[D(m,n,p)] &= g'_{D(m,n,p)}(1) = (n-1)[1-(1-p^2)^m] \\
\mathbb{E}[D(m,n,p)(D(m,n,p)-1)] &= g''_{D(m,n,p)}(1) \\
&= (n-1)(n-2)[1-2(1-p^2)^m + (1-p^2(2-p))^m],
\end{aligned}$$

and with n' and n'' as in the lemma we can deduce $\mathbb{E}[D(n', n'', \gamma/n)] = \mu + o(1) = O(1)$, where $\mu := \beta\gamma^2$, and $\mathbb{E}[D(n', n'', \gamma/n)^2] = \mu(1+\mu+\gamma) + o(1) = O(1)$. We write

$$g(s) := \exp\left\{\beta\gamma\left(e^{\gamma(s-1)}-1\right)\right\}$$

for the generating function of the limiting distribution $\mathrm{CPoisson}(\beta\gamma, \gamma)$. Finally, let us define ρ to be the smallest non-negative root of $\rho = g(\rho)$.

3 Results

Theorem 2. *Let $\mu := \beta\gamma^2$, i.e. the asymptotic expected degree of a randomly chosen vertex of $\mathcal{G}^{(n)}$.*

(i) If $\mu < 1$, then there is a.a.s. no connected component in $\mathcal{G}^{(n)}$ with more than $O(\log n)$ vertices.

(ii) If $\mu > 1$, then $0 < \rho < 1$ and there exists a unique giant component of size $(1 - \rho + o_p(1))n$, and the size of the second largest connected component is a.a.s. no larger than $O(\log n)$.

With $W_n = o_p(a_n)$ we mean that $W_n/a_n \to 0$ in probability as $n \to \infty$. As mentioned in the introduction, Behrisch has investigated the component structure for the random intersection graph when $\alpha \neq 1$, see [2, Thm. 1]. It is worth noting that the results in Theorem 2 in this note are closer to the results that Behrisch obtained for the case $\alpha > 1$, than for $\alpha < 1$. For $\alpha < 1$ the size of the giant component is no longer linear in n.

4 Proof of Theorem 2

For the remainder of this note we will follow the notation and steps of the proof of Theorem 5.4 in [4, Ch. 5.2]. Therefore most of the details that have not been altered from the original proof will be omitted. The proof is based on choosing a vertex at random from $\mathcal{V}$, say v, and exploring its component, say $\mathcal{C}(v)$. We start by *visiting* the chosen vertex v and *identifying* its neighbours. Then we proceed by visiting an identified but unvisited vertex, if any remains, and identify its neighbours, and repeat this procedure until all vertices in the component have been visited. Let X_i denote the number of newly identified vertices at the ith step of this exploration process. The event $\{|\mathcal{C}(v)| = k\}$ is equivalent to $\sum_{i=1}^{k} X_i = k - 1$, and it is thus important to understand the growth of this partial sums process. The random variables $X_1, X_2, \ldots$ are not i.i.d. but the partial sums process can nevertheless be related to other partial sums processes with i.i.d. summands so that we obtain bounds on events of the type above. We will need the following result for these partial sums processes.

Lemma 3. *Let $\delta > 0$ and $\tilde{X} := \tilde{X}_1 + \cdots + \tilde{X}_k$, where $\tilde{X}_1, \ldots$ are i.i.d. as $D(n', n'', \gamma/n)$ of Lemma 1. Then, for large enough n, there exists a positive constant $C := C(\beta, \gamma, \delta)$ such that $\mathbb{P}(\tilde{X} \geq (1 + \delta)\mu k) \leq e^{-Ck}$ and $\mathbb{P}(\tilde{X} \leq (1 - \delta)\mu k) \leq e^{-Ck}$.*

Remark. This bound on the tail probabilities works since $\tilde{X}$ is a sum of k independent random variables. As $n \to \infty$, the RIG-distribution of the summands does not change much: It is more or less CPoisson$(\beta\gamma, \gamma)$, which is a "well behaved" distribution, and as k increases, we expect an exponential decay of probabilities away from the mean of the sum. Since C is not further specified, this bound is only useful as k tends to infinity, which it may or may not do as a function of n.

Before we prove the lemma, note that we can construct a *multigraph* $\mathcal{H}^{(n)}_{m,p}$ from the same bipartite graph $\mathcal{B}^{(n)}_{m,p}$ as we used in the construction of $\mathcal{G}^{(n)}_{m,p}$, by letting the number of edges between v_i and v_j equal the number of auxiliary vertices a_k that are adjacent to both v_i and v_j. We denote with $\mathrm{RIMG}(m,n,p)$ the degree distribution of $\mathcal{H}^{(n)}_{m,p}$. $\mathrm{RIMG}(m,n,p)$ clearly dominates $\mathrm{RIG}(m,n,p)$, as we can obtain $\mathcal{G}^{(n)}_{m,p}$ from $\mathcal{H}^{(n)}_{m,p}$ by coalescing multiple edges between vertices into one single edge. $\mathrm{RIMG}(m,n,p)$ is a compound binomial distribution with generating function

$$h(z) = (1-p+p(1-p+pz)^{n-1})^m,$$

since, by construction, a vertex $v_i \in \mathcal{H}^{(n)}_{m,p}$ is connected to a Binomial(m,p) number of auxiliary vertices, each of which being connected to an independent Binomial$(n-1,p)$ number of vertices in $\mathcal{V} \setminus \{v_i\}$.

The expected value of $\mathrm{RIMG}(\lfloor \beta n \rfloor, n, \gamma/n)$ is thus $\lfloor \beta n \rfloor (n-1)\gamma^2/n^2 = \mu + O(1/n) = \mathbb{E}[D(\lfloor \beta n \rfloor, n, \gamma/n)] + O(1/n)$, so the expected difference in vertex degree between the multigraph and the ordinary graph is only $O(1/n)$. With $\eta^{(n)}$ defined as the difference in the total number of edges in $\mathcal{H}^{(n)}_{\lfloor \beta n \rfloor, \gamma/n}$ and $\mathcal{G}^{(n)}_{\lfloor \beta n \rfloor, \gamma/n}$, we have $\mathbb{E}[\eta^{(n)}] = O(1)$, by summing over all vertices. This will be used in the proof of Theorem 2(ii).

Proof of Lemma 3. Note that $\mathbb{E}[e^{\theta \tilde{X}}] = \mathbb{E}[e^{\theta \tilde{X}_1 + \cdots + \theta \tilde{X}_k}] = \mathbb{E}[e^{\theta \tilde{X}_1}]^k$. Let $s > 0$. We have

$$\begin{aligned}
\mathbb{P}(\tilde{X} \le (1-\delta)\mu k) &= \mathbb{P}(e^{-s\tilde{X}} \ge e^{-s(1-\delta)\mu k}) \\
&\le e^{s(1-\delta)\mu k}\mathbb{E}[e^{-s\tilde{X}}] = \left(e^{s\mu - s\delta\mu}\mathbb{E}[e^{-s\tilde{X}_1}]\right)^k, \qquad (1) \\
\mathbb{P}(\tilde{X} \ge (1+\delta)\mu k) &= \mathbb{P}(e^{s\tilde{X}} \ge e^{s(1+\delta)\mu k}) \\
&\le e^{-s(1+\delta)\mu k}\mathbb{E}[e^{s\tilde{X}}] = \left(e^{-s\mu - s\delta\mu}\mathbb{E}[e^{s\tilde{X}_1}]\right)^k, \qquad (2)
\end{aligned}$$

by Markov's inequality. Since $e^{-s\tilde{X}_1} \le 1 - s\tilde{X}_1 + \frac{1}{2}s^2\tilde{X}_1^2$ for $s > 0$,

$$\begin{aligned}
\mathbb{E}[e^{-s\tilde{X}_1}] &\le 1 - sE[\tilde{X}_1] + \tfrac{1}{2}s^2\mathbb{E}[\tilde{X}_1^2] = \exp\{\log(1 - sE[\tilde{X}_1] + \tfrac{1}{2}s^2\mathbb{E}[\tilde{X}_1^2])\} \\
&= \exp\{-s\mathbb{E}[\tilde{X}_1] + O(s^2)\} = \exp\left\{-s\left(\mu + o(1) + O(s)\right)\right\}.
\end{aligned}$$

The right hand side of (1) is thus $\exp\left\{-s\left(\delta\mu + o(1) + O(s)\right)k\right\}$, and we can fix a small s, such that for large enough n, $s(\delta\mu + o(1) + O(s))$ is positive (regardless of the value of k), and thus $\mathbb{P}(\tilde{X} \le (1-\delta)\mu k) \le e^{-C'k}$ for some positive C'.

For the second part of the proof, let $\hat{X} \in \mathrm{RIMG}(n', n'', \gamma/n)$, so that $\hat{X} \ge_d \tilde{X}_1$.

$$\begin{aligned}
\mathbb{E}[e^{s\tilde{X}_1}] \le \mathbb{E}[e^{s\hat{X}}] &= \left(1 - \tfrac{\gamma}{n} + \tfrac{\gamma}{n}\left(1 - \tfrac{\gamma}{n} + \tfrac{\gamma}{n}e^s\right)^{n''-1}\right)^{n'} \\
&< \exp\left\{\gamma\tfrac{n'}{n}\left(e^{\gamma\frac{n''-1}{n}(e^s-1)} - 1\right)\right\} \\
&= \exp\left\{\gamma(\beta + o(1))\left(e^{\gamma(1+o(1))s(1+O(s))} - 1\right)\right\} \\
&= \exp\left\{\gamma\beta(1+o(1))\left(e^{\gamma s(1+o(1)+O(s))} - 1\right)\right\} \\
&= \exp\left\{\mu s(1 + o(1) + O(s))\right\}.
\end{aligned}$$

The right hand side of (2) is thus less than $\exp\{-s(\delta\mu + o(1) + O(s))k\}$, and we can fix a small s, such that for large enough n, $s(\delta\mu + o(1) + O(s))$ is positive (regardless of the value of k), and thus $\mathbb{P}(\tilde{X} \geq (1+\delta)\mu k) \leq e^{-C''k}$ for some positive C''. We conclude the proof of the lemma by letting $C = \min\{C', C''\}$. □

Proof of Theorem 2(i). The process of exploring vertices that was briefly described in the beginning of Section 4, implies that X_1, the number of neighbours of the initially picked vertex, has distribution $\mathrm{RIG}(\lfloor \beta n \rfloor, n, \gamma/n)$. This, together with the fact that vertices only can be newly identified once, implies that $\sum_{i=1}^{k} X_i \leq_d \sum_{i=1}^{k} X_i^+$ for all k, where $X_1^+, \ldots$ are i.i.d. $\mathrm{RIG}(\lfloor \beta n \rfloor, n, \gamma/n)$. Thus

$$\mathbb{P}(\exists i : |\mathcal{C}(v_i)| \geq k) \leq \sum_{i=1}^{n} \mathbb{P}(|\mathcal{C}(v_i)| \geq k) = n\mathbb{P}(|\mathcal{C}(v)| \geq k) \leq n\mathbb{P}\left(\sum_{j=1}^{k} X_j^+ \geq k-1\right).$$

Now we take $k := k(n)$ increasing to infinity with n. Since all X_i^+ are i.i.d. $\mathrm{RIG}(\lfloor \beta n \rfloor, n, \gamma/n)$, Lemma 3 applies to $X^+ := \sum_{j=1}^{k} X_j^+$, which gives us

$$\begin{aligned}\mathbb{P}(\exists i : |\mathcal{C}(v_i)| \geq k) &\leq n\mathbb{P}\left(X^+ \geq k-1\right) = n\mathbb{P}(X^+ \geq (1+2\delta)\mu k - 1)\\ &\leq n\mathbb{P}(X^+ \geq (1+\delta)\mu k) \leq n\exp\{-Ck\},\end{aligned}$$

where $\mu < 1$, $\delta = (1/\mu - 1)/2 > 0$, C is defined as in Lemma 3, and the penultimate inequality follows from $\delta\mu k > 1$ for large enough k. That is, if $k(n) := \lceil (1+\epsilon)\log n / C \rceil$, $\epsilon > 0$, then $\mathbb{P}(\exists i : |\mathcal{C}(v_i)| \geq k) \leq n^{-\epsilon} \to 0$ as $n \to \infty$, and the first part of Theorem 2 is proved. □

Proof of Theorem 2(ii). We will first show that there with probability tending to one are no clusters of size k with $O(\log n) = k_-(n) \leq k \leq k_+(n) := n^{2/3}$. From now on, let $k_- \leq k \leq k_+$, where $k_-(n)$ will be specified shortly. The construction used in the proof is similar but more involved than the one of the first part of the proof.

For the remainder of the proof we will implicitly condition on the event $\{\eta^{(n)} \leq \sqrt{n}\}$, whose probability tends to one when $n \to \infty$, by Markov's inequality and the fact that $\mathbb{E}[\eta^{(n)}] = O(1)$. Our construction fails on the complementary event, but this is of no consequence for the proof, since the probability of this event tends to zero.

Let $A(v)$ be the event that the exploration process, initiated at v, at step k_+ has not terminated and that it at that step has identified fewer than $(\mu-1)k_+/2$ vertices that have not yet been visited, i.e. $A(v) = \{k_+ \leq \sum_{j=1}^{k_+} X_j \leq k_+ - 1 + (\mu-1)k_+/2\}$. Let $B(v)$ be the event $\{\sum_{j=1}^{k_+} X_j \leq k_+ - 1 + (\mu-1)k_+/2\}$. We will prove that the probability that the exploration process terminates after k steps or that $A(v)$ holds for some v, tends to zero. Note that $\{|\mathcal{C}(v)| = k\} \subseteq B(v)$ for each k, and in particular $\{|\mathcal{C}(v)| = k_+\} \cup A(v) \subseteq B(v)$. We also have $\{|\mathcal{C}(v)| = k\} \subseteq \{\sum_{j=1}^{k} X_j \leq k - 1 + (\mu-1)k/2\}$ for each k.

On the set $B(v) \cap \{\eta^{(n)} \leq \sqrt{n}\}$, the exploration process has at step k identified vertices in $\mathcal{V}$, that are adjacent to fewer than $(\mu+1)k_+/2 + \sqrt{n}$ auxiliary vertices in $\mathcal{B}_{m,p}^{(n)}$. We claim that

$$\mathbb{P}\bigg(\sum_{i=1}^{k} X_i \leq k-1+\frac{\mu-1}{2}k\bigg) \leq \mathbb{P}\bigg(\sum_{i=1}^{k} X_i^- \leq k-1+\frac{\mu-1}{2}k\bigg) \tag{3}$$

holds with $X_1^-,\ldots,$ i.i.d. $\mathrm{RIG}(\lfloor \beta n-(\mu+1)k_+/2-\sqrt{n}\rfloor, \lfloor n-(\mu+1)k_+/2\rfloor, \gamma/n)$. Note that $\sum_{i=1}^k X_i^-$ is *not* a lower stochastic bound on $\sum_{i=1}^k X_i$ in the same way as $\sum_{i=1}^k X_i^+$ is an upper bound since the distribution of X_1^- depends on k_+. The claim follows by a slight adaptation of the arguments of the proof of Theorem 4.3 in [6, Ch. 4.2]: We compare our exploration process with another exploration process, which does not follow vertices that belong to a group of *forbidden* vertices, or that are reached through edges generated by a group of forbidden auxiliary vertices. Both groups of forbidden vertices and auxiliary vertices are adjusted (diminished) after each step so that there are $(\mu+1)k_+/2$ vertices that are forbidden or identified, and $(\mu+1)k_+/2+\sqrt{n}$ auxiliary vertices that are forbidden or have generated an edge to an identified verte x. These adjustments are possible until the process has identified $(\mu+1)k_+/2$ vertices, which is long enough to deduce whether fewer than $k-1+(\mu-1)k/2$ vertices have been identified after step k. Furthermore, since we keep the number of forbidden vertices and auxiliary vertices fixed, the number of newly identified vertices by the modified exploration process will in each step be i.i.d. $\mathrm{RIG}(\lfloor \beta n-(\mu+1)k_+/2-\sqrt{n}\rfloor, \lfloor n-(\mu+1)k_+/2\rfloor, \gamma/n)$.

Using (3), assuming that $\{\eta^{(n)} \leq \sqrt{n}\}$ holds, gives us that

$$\begin{aligned}
\mathbb{P}(\exists i : \{k_- \leq |\mathcal{C}(v_i)| \leq k_+\} \cup A(v_i)) &\leq n\mathbb{P}(\{k_- \leq |\mathcal{C}(v)| \leq k_+\} \cup A(v)) \\
&= n\bigg(\sum_{k=k_-}^{k_+-1} \mathbb{P}(|\mathcal{C}(v)|=k) + \mathbb{P}(|\mathcal{C}(v)|=k_+) + \mathbb{P}(A(v))\bigg) \\
&\leq n\sum_{k=k_-}^{k_+} \mathbb{P}\bigg(\sum_{j=1}^{k} X_j \leq k-1+\frac{\mu-1}{2}k\bigg) \\
&\leq n\sum_{k=k_-}^{k_+} \mathbb{P}\bigg(\sum_{j=1}^{k} X_j^- \leq k-1+\frac{\mu-1}{2}k\bigg).
\end{aligned}$$

We apply Lemma 3 to $X^- := \sum_{j=1}^k X_j^-$, which yields

$$\begin{aligned}
\mathbb{P}(\exists i : \{k_- \leq |\mathcal{C}(v_i)| \leq k_+\} \cup A(v_i)) &\leq n\sum_{k=k_-}^{k_+} \mathbb{P}\bigg(X^- \leq k-1+\frac{\mu-1}{2}k\bigg) \\
&= n\sum_{k=k_-}^{k_+} \mathbb{P}\Big(X^- \leq (1-\delta)\mu k - 1\Big) \\
&\leq n\sum_{k=k_-}^{k_+} \mathbb{P}(X^- \leq (1-\delta)\mu k) \leq nk_+ \exp\{-Ck_-\},
\end{aligned}$$

where $\mu > 1$, $\delta = (1-1/\mu)/2 > 0$ and C is defined as in Lemma 3. Therefore, if $k_-(n) := \lceil (5/3+\epsilon)\log n/C \rceil$, $\epsilon > 0$ and $k_+(n) := \lfloor n^{2/3} \rfloor$ then $\mathbb{P}(\exists i : \{k_- \le |\mathcal{C}(v_i)| \le k_+\} \cup A(v_i)) \le n^{-\epsilon} \to 0$ as $n \to \infty$.

From Section 1 we know that two vertices in $\mathcal{G}^{(n)}$ are not connected if they avoid being adjacent to the same auxiliary vertex. Thus the probability that two vertices are not connected is $(1-\gamma^2/n^2)^{\lfloor \beta n \rfloor}$. Furthermore we know from the previous calculations that the probability that $A(v)$ holds for some v tends to zero as n tends to infinity, i.e. if there exist two different components of size k_+, they will each have at least $(\mu-1)k_+/2$ identified but not yet visited vertices. This implies that the probability that two components each of size k_+ are disjoint after visiting their additional vertices is less than

$$\begin{aligned}\left((1-\gamma^2/n^2)^{\lfloor \beta n \rfloor}\right)^{((\mu-1)k_+/2)^2} &\le \exp\left\{-\frac{\gamma^2}{n^2}\lfloor \beta n \rfloor \left(\frac{\mu-1}{2}\lfloor n^{2/3} \rfloor\right)^2\right\} \\ &\le \exp\left\{-\frac{\mu(\mu-1)^2}{4} O(n^{1/3})\right\} = o(1/n^2).\end{aligned}$$

That is, with probability tending to one, either vertices belong to connected components of size less than k_-, or to a *unique* component of size at least k_+.

To show that the size of the largest component grows linearly in n with high probability, we need to show that the number of vertices that belong to small components, i.e. components of size k_- or less, is strictly less than n, implying the remaining vertices belong to the giant component. Let $L_i := \{|\mathcal{C}(v_i)| \le k_-\}$, $Y_i := \mathbf{1}_{L_i}$, and set $Y := \sum_{i=1}^n Y_i$, so that $\mathbb{E}[Y] = n\mathbb{E}[Y_1] = n\mathbb{P}(L_1)$. By the same reasoning we use above, we can sandwich $\mathbb{P}(L_1)$ between $\mathbb{P}(C^+ \le k_-)$ and $\mathbb{P}(C^- \le k_-)$ where C^+ and C^- are the total sizes of branching processes with offspring distributed as X_1^+ and X_1^-, respectively. Lemma 1 implies that both offspring distributions tend to the same limit, CPoisson$(\beta\gamma, \gamma)$, as n tends to infinity. By standard results in branching process theory, see Athreya and Ney [1, Thm. I.5.1], both probabilities $\mathbb{P}(C^+ \le k_-)$ and $\mathbb{P}(C^- \le k_-)$ tend to the ρ that we defined as the smallest non-negative root of $g(\rho) = \rho$, since $k_-(n)$ tends to infinity with n and ρ is the probability that the branching process with offspring distribution CPoisson$(\beta\gamma, \gamma)$ has finite total size. It also holds that $0 < \rho < 1$, since $\mu > 1$. Due to this, $\mathbb{E}[Y] = (\rho + o(1))n$, which implies that the expected size of the largest component is $(1-\rho+o(1))n$, and the proof that Y is concentrated around ρn follows the last part of the proof of Theorem 1.(2) in Behrisch [2, Sec. 4.2, p. 8] verbatim. □

Acknowledgements

We thank an anonymous referee for careful reading of the manuscript and for pointing out errors.

References

[1] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer Verlag, 1972.

[2] M. Behrisch. Component evolution in random intersection graphs. *The Electronic Journal of Combinatorics*, 14, #R17, 2007.

[3] M. Deijfen and W. Kets. Random intersection graphs with tunable degree distribution and clustering. *Stockholm University Research Reports in Mathematical Statistics*, 2007:1, 2007.

[4] S. Janson, T. Łuczak and A. Ruciński. *Random Graphs*. John Wiley & Sons, 2000.

[5] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.

[6] R. van der Hofstad. *Random Graphs and Complex Networks*. Lecture notes, in preparation, 2008. `http://www.win.tue.nl/~rhofstad`

[7] D. Stark. The vertex degree distribution of random intersection graphs. *Random Structures and Algorithms*, 24(3):249–258, 2004.