Giant Components in Kronecker Graphs

Paul Horn[*] Mary Radcliffe[†]

Abstract

Let $n \in \mathbb{N}$, $0 < \alpha, \beta, \gamma < 1$. Define the random Kronecker graph $K(n, \alpha, \gamma, \beta)$ to be the graph with vertex set $\mathbb{Z}_2^n$, where the probability that $\mathbf{u}$ is adjacent to $\mathbf{v}$ is given by $p_{u,v} = \alpha^{\mathbf{u}\cdot\mathbf{v}}\gamma^{(\mathbf{1}-\mathbf{u})\cdot(\mathbf{1}-\mathbf{v})}\beta^{n-\mathbf{u}\cdot\mathbf{v}-(\mathbf{1}-\mathbf{u})\cdot(\mathbf{1}-\mathbf{v})}$. This model has been shown to obey several useful properties of real-world networks. We establish the asymptotic size of the giant component in the random Kronecker graph.

1 Introduction

Suppose n is fixed. Fix probability $0 < \alpha, \beta, \gamma < 1$. A random Kronecker graph $K(n, \alpha, \gamma, \beta)$ on $N = 2^n$ vertices is a graph whose vertex set is the elements of $\mathbb{Z}_2^n$, and the probability that two vertices $\mathbf{u}$ and $\mathbf{v}$ are adjacent is given by

$$p_{u,v} = \alpha^{\mathbf{u}\cdot\mathbf{v}}\gamma^{(\mathbf{1}-\mathbf{u})\cdot(\mathbf{1}-\mathbf{v})}\beta^{n-\mathbf{u}\cdot\mathbf{v}-(\mathbf{1}-\mathbf{u})\cdot(\mathbf{1}-\mathbf{v})}.$$

This model was originally proposed by Leskovec et al. in [4] as one that adequately models many real-world network properties. In particular, Kronecker graphs have a heavy-tailed degree distribution, and follow the densification power law [4]. Fitting the Kronecker graph model to several real-world graphs has proven to be very successful, as seen in [3]. However, Leskovec et al. primarily focus on a deterministic version of the model, rather than the stochastic version studied here. We develop several results regarding the emergence and size of the giant component in random Kronecker graphs. We note that Mahdian and Xu proved the following necessary and sufficient condition for the emergence of a giant component in the random Kronecker graph in [5]:

Theorem 1. *Let $G = K(n, \alpha, \gamma, \beta)$ be a random Kronecker graph, with $\alpha \geq \beta \geq \gamma$. A necessary and sufficient condition for G to have a giant component of size $\Theta(N)$ a.a.s. is that $(\alpha + \beta)(\beta + \gamma) > 1$.*

[*]Department of Mathematics and Computer Science, Emory University
[†]Department of Mathematics, University of California at San Diego

We will prove the following two theorems. The first generalizes the above theorem to the case where we do not have $\alpha \geq \beta \geq \gamma$ and the second establishes sharp bounds on the asymptotic size of the giant component in $K(n, \alpha, \gamma, \beta)$, when one exists.

Theorem 2. *Let $G = K(n, \alpha, \gamma, \beta)$ be a random Kronecker graph. A necessary and sufficient condition for G to have a giant component of size $\Theta(N)$ a.a.s. is that $(\alpha + \beta)(\beta + \gamma) > 1$.*

Theorem 3. *Suppose $\alpha, \beta, \gamma \in (0, 1)$ with $(\alpha + \beta)(\gamma + \beta) > 1$ and $\alpha \geq \gamma$. By Theorem 2, $K(n, \alpha, \gamma, \beta)$ has a giant component a. a. s.. Suppose, moreover, that $\beta + \gamma < 1$. Let X denote the set of vertices of $K(n, \alpha, \gamma, \beta)$ that are not in the giant component. Then a.a.s.*

$$|X| = \Theta\left(\binom{n}{mn}\right)$$

where

$$m = \frac{-\log(\beta + \gamma)}{-\log(\beta + \gamma) + \log(\alpha + \beta)}$$

We note that in Theorem 3, the assumption that $\beta + \gamma < 1$ is not very restrictive; if $\beta + \gamma \geq 1$ then the graph is connected a.a.s. [5]. Furthermore, in this instance, the constant m is negative, so one could omit the restriction by replacing m with the maximum of the stated value or 0.

The remainder of this paper is organized as follows: In section 2, we prove some basic lemmas needed for the proof of the upper bound in Theorem 3 and the sufficiency in Theorem 2, as well as derive some essential facts about certain intersection graphs which are important in the proof. In section 3 we complete the proof of Theorem 2 and establish the upper bound in Theorem 3. Section 4 is devoted to the lower bound of Theorem 3.

2 Basic Facts

For this section we assume that $0 < \alpha, \beta, \gamma < 1$ are real numbers satisfying $(\alpha + \beta)(\beta + \gamma) > 1$ and $\alpha > \gamma$ (by symmetry, this last assumption is only for convenience). The approach to establishing Theorem 3 will be to find a section of the graph $K(n, \alpha, \gamma, \beta)$ with good expansion, and show this is contained in a giant component. In so doing, we gain structural information about the giant component itself.

For a vertex $\mathbf{v} \in \mathbb{Z}_2^n$, define the weight of $\mathbf{v}$, denoted $w(\mathbf{v})$, to be the number of coordinates which are equal to 1, that is, $w(\mathbf{v}) = \sum_{i=1}^{n} \mathbf{v}_i$.

Let

$$k = \frac{\alpha+\beta}{\alpha+\gamma+2\beta}n,$$

and let H denote the subgraph of $G = K(n,\alpha,\gamma,\beta)$ consisting of vertices of weight k. For a vertex $\mathbf{v} \in H$, we restrict our attention to (potential) edges which swap precisely

$$l = \frac{\beta}{\alpha+\gamma+2\beta}n = \frac{\beta}{\alpha+\beta}k$$

1's for 0's (and thus must also swap l 0's for 1's).

Note that while these parameters may seem quite mysterious, they actually are quite natural. We say an edge from a vertex $\mathbf{v}$ of weight k is of type (l,t) if it involves switching l 1's to 0's, and t 0's to 1's. Then the expected number of neighbors of $\mathbf{v}$ of type (l,t) is

$$\binom{k}{l}\binom{n-k}{t}\beta^l\alpha^{k-l}\beta^t\gamma^{n-k-t}.$$

We first find l and t that maximize this expression in terms of k, and then find k so that $l = t$, resulting in the above parameters.

Lemma 1. *For a vertex $\mathbf{v} \in H$, consider its neighbors of type (l,l). The expected number of such neighbors of $\mathbf{v}$ is*

$$\binom{k}{l}\binom{n-k}{l}\beta^l\alpha^{k-l}\beta^l\gamma^{n-k-l} > (1+o(1))c^n.$$

for some $c > 1$.

Proof. Recall the entropy bound (see, for example, Thm. 2.6 in [6])

$$\binom{n}{pn} > \frac{e^{nH(p)}}{e\sqrt{2\pi np(1-p)}},$$

where $H(p) = -p\log p - (1-p)\log(1-p)$. Note that $l/k = \frac{\beta}{\alpha+\beta}$ and $l/(n-k) = \frac{\beta}{\gamma+\beta}$, so the entropy bound gives us

$$\begin{aligned}
&\binom{k}{l}\binom{n-k}{l}\beta^l\alpha^{k-l}\beta^l\gamma^{n-k-l} \\
&> \frac{(\alpha+\beta)(\beta+\gamma)}{2e^2\pi k\beta\sqrt{\alpha\gamma}}\left(\frac{\alpha+\beta}{\beta}\right)^l\left(\frac{\alpha+\beta}{\alpha}\right)^{k-l}\left(\frac{\beta+\gamma}{\beta}\right)^l\left(\frac{\beta+\gamma}{\gamma}\right)^{n-k-l}\beta^l\alpha^{k-l}\beta^l\gamma^{n-k-l} \\
&= \frac{(\alpha+\beta)(\beta+\gamma)}{2e^2\pi k\beta\sqrt{\alpha\gamma}}(\alpha+\beta)^k(\beta+\gamma)^{n-k} \\
&= \frac{1}{2e^2\pi k\beta\sqrt{\alpha\gamma}}((\alpha+\beta)(\beta+\gamma))^{n-k+1}(\alpha+\beta)^{2k-n}. \qquad (1)
\end{aligned}$$

Notice that $\alpha > \gamma$ and $(\alpha+\beta)(\beta+\gamma) > 1$ implies that $\alpha+\beta > 1$ and also that $k > \frac{n}{2}$. Thus (1) is clearly exponential in n as desired. Note that, while it won't strictly be necessary, the c we obtain is

$$c = ((\alpha+\beta)(\beta+\alpha))^{\frac{\alpha+\beta}{\alpha+\gamma+2\beta}}(\alpha+\beta)^{\frac{\alpha-\gamma}{\alpha+\gamma+2\beta}} > 1.$$

$\square$

Lemma 1 implies that the expected degree of each vertex in H is very large (indeed, exponential in n).

Let $G(n,k,l)$ denote the graph on $\binom{n}{k}$ vertices, where each vertex is a k-subset of an n-set, and two vertices are adjacent if they intersect in exactly $k-l$ points. Note that our graph H, restricted to the swaps of type (l,l), is a percolated version of this graph where edges are taken independently with probability $\beta^l\alpha^{k-l}\beta^l\gamma^{n-k-l}$. In order to show that H is a.a.s. connected, we need to derive some information on $G(n,k,l)$.

Lemma 2. *Suppose k and l are as above. Then*

$$diam(G(n,k,l)) = \Theta(1).$$

Proof. It suffices to show that there is a path between two arbitrary vertices $\mathbf{v}$ and $\mathbf{v}'$. For a set X we define the measure

$$i(X) = \frac{(\alpha+\beta)|X|}{k}.$$

and for two vertices we define

$$i(\mathbf{v},\mathbf{v}') = \frac{(\alpha+\beta)|\mathbf{v}\cap\mathbf{v}'|}{k}.$$

(where $\mathbf{v}$ and $\mathbf{v}'$ are thought of as sets). Note that two vertices are adjacent if $|\mathbf{v}\cap\mathbf{v}'| = k-l = \frac{\alpha}{\alpha+\beta}k$, so two vertices are adjacent if $i(\mathbf{v},\mathbf{v}') = \alpha$. Further note that $i([n]) = \alpha+\gamma+2\beta$.

We prove a series of claims:

Claim 1: If $i(\mathbf{v},\mathbf{v}') \geq \alpha+\beta-\gamma$ and $i(\mathbf{v},\mathbf{v}') \geq \alpha$, then there exists a vertex $\mathbf{v}''$ such that $\mathbf{v} \sim \mathbf{v}''$ and $\mathbf{v}' \sim \mathbf{v}''$.

Note that here $|\mathbf{v}\cup\mathbf{v}'| \leq \frac{\alpha+\beta+\gamma}{\alpha+\gamma+2\beta}n$, in particular there is a set X of size $\frac{\beta}{\alpha+\gamma+2\beta}n = l$ (that is to say, a set X such that $i(X) = \beta$) completely disjoint from $\mathbf{v}\cup\mathbf{v}'$. Let $Y \subseteq \mathbf{v}\cap\mathbf{v}'$ with $|Y| = k-l$. Consider $\mathbf{v}'' = X \cup Y$. Then $i(\mathbf{v},\mathbf{v}'') = i(\mathbf{v}',\mathbf{v}'') = \alpha$, and the proof of Claim 1 is complete.

Claim 2: Suppose $i(\mathbf{v},\mathbf{v}') \leq \alpha$. Then there exists a vertex $\mathbf{v}''$ such that $\mathbf{v} \sim \mathbf{v}''$ and $i(\mathbf{v}'',\mathbf{v}') = i(\mathbf{v},\mathbf{v}') + \beta$.

Here, let $X \subseteq \mathbf{v}' \setminus \mathbf{v}$ with $|X| = l$ and let $Y \subseteq \mathbf{v}$ such that $|Y| = k - l$ and $\mathbf{v} \cap \mathbf{v}' \subseteq Y$. Then $\mathbf{v}'' = X \cup Y$ has the desired property, proving Claim 2.

Claim 3: Suppose $i(\mathbf{v}, \mathbf{v}') = \alpha + x$, where $x > 0$. Then there exists a vertex $\mathbf{v}''$ such that $\mathbf{v} \sim \mathbf{v}''$ and $i(\mathbf{v}'', \mathbf{v}') = \alpha + \beta - x$.

Here, let $X \subseteq \mathbf{v} \cap \mathbf{v}'$ with $|X| = k - l$. Note that $i(\mathbf{v} \cup \mathbf{v}') = \alpha + x + 2(\beta - x) = \alpha + 2\beta - x$, so $i([n] \setminus (\mathbf{v} \cup \mathbf{v}')) = \gamma + x$. Let $Y \subseteq [n] \setminus (\mathbf{v} \cup \mathbf{v}')$ with $i(Y) = x$ and $Z = \mathbf{v}' \setminus \mathbf{v}$, so that $i(Z) = \beta - x$. Then $\mathbf{v}'' = X \cup Y \cup Z$ has the desired properties, proving Claim 3.

Claim 4: Suppose $i(\mathbf{v}, \mathbf{v}') = \alpha + x$, with $0 < x < \beta - \gamma$. Then there exists a vertex $\mathbf{v}''$ with$\mathbf{v}'' \sim \mathbf{v}$ and $i(\mathbf{v}', \mathbf{v}'') = \alpha + \beta - x - 2\gamma$..

Let $X \subseteq \mathbf{v} \cap \mathbf{v}'$ with $i(X) = \alpha - \gamma$. Let $Y \subseteq \mathbf{v} \setminus \mathbf{v}'$ with $i(Y) = \gamma$. Let $Z \subseteq \mathbf{v}' \setminus \mathbf{v}$ with $i(Z) = \beta - x - \gamma$, and let $W = [n] \setminus (\mathbf{v} \cup \mathbf{v}')$, so that $i(W) = \gamma + x$. Let $\mathbf{v}'' = X \cup Y \cup Z \cup W$. Note that $i(\mathbf{v}'') = \alpha - \gamma + \gamma + \beta - x - \gamma + \gamma + x = \alpha + \beta$, and $i(\mathbf{v}, \mathbf{v}'') = i(X) + i(Y) = \alpha$ so that $\mathbf{v}'' \sim \mathbf{v}$. Moreover, $i(\mathbf{v}', \mathbf{v}'') = i(X) + i(Z) = \alpha + \beta - x - 2\gamma$, as desired. Therefore, Claim 4 is proven.

Claim 5: Suppose $i(\mathbf{v}, \mathbf{v}') = \alpha + x$, with $0 < x < \beta - 2\gamma$. Then there exists a vertex $\mathbf{v}''$ with $d(\mathbf{v}, \mathbf{v}'') = 2$ and $i(\mathbf{v}', \mathbf{v}'') = \alpha + x + 2\gamma$.

Claim 4 implies that there exists a $\mathbf{v}''' \sim \mathbf{v}$ with $i(\mathbf{v}', \mathbf{v}''') = \alpha + \beta - x - 2\gamma$. By assumption $\beta - x - 2\gamma > 0$, so we can apply Claim 3 using vertices $\mathbf{v}'$ and $\mathbf{v}'''$, so there there exists $\mathbf{v}'' \sim \mathbf{v}'''$ with $i(\mathbf{v}', \mathbf{v}'') = \alpha + x + 2\gamma$. Therefore, Claim 5 is established.

From here, the proof is simple. Suppose $\mathbf{v}, \mathbf{v}'$ are arbitrary vertices. After at most $\lceil \frac{\alpha+\beta}{\beta} \rceil$ applications of Claim 2, we have a vertex $\mathbf{v}''$ such that $i(\mathbf{v}', \mathbf{v}'') = \alpha + x$ for some $0 < x \leq \beta$. If $x = \beta$ we are done. Otherwise, either Claim 2 applies, and $d(\mathbf{v}', \mathbf{v}'') = 1$, or $0 \leq \beta - 2\gamma \leq x < \beta - \gamma$, and $d(\mathbf{v}', \mathbf{v}'') \leq 3$ by applying Claim 4 followed by Claim 1 followed by Claim 2, or $x < \beta - 2\gamma$. In the final case, after at most $\lceil \frac{\beta}{2\gamma} \rceil$ applications of Claim 5, we must have a vertex $\mathbf{v}'''$ with $i(\mathbf{v}', \mathbf{v}''') = \alpha + y$ for some y satisfying $y \geq \beta - 2\gamma$, and by our above arguments $d(\mathbf{v}''', \mathbf{v}') \leq 3$.

In total we have that

$$diam(G(n,k,l)) \leq \left\lceil \frac{\alpha+\beta}{\beta} \right\rceil + \left\lceil \frac{\beta}{2\gamma} \right\rceil + 3 = \Theta(1).$$

□

It is easy to observe that $G(n,k,l)$ is edge transitive (permutations of $[n]$ are automorphisms of $G(n,k,l)$ and it is easy to construct a permutation which maps one edge to any other.) Recall that the Cheeger constant of a graph H is

$$h_H = \min_{\substack{S \subseteq H \\ \mathrm{vol}(S) \leq \mathrm{vol}(H)/2}} \frac{e(S, \bar{S})}{\mathrm{vol}(S)}$$

where here $\mathrm{vol}(S) = \sum_{\mathbf{v} \in S} \deg(\mathbf{v})$. Theorem 7.1 in Chung [2] asserts that for a edge transitive graph Γ with diameter D,

$$h_\Gamma \geq \frac{1}{2D}.$$

In particular this implies

Lemma 3. *For $\tilde{H} = G(n, k, l)$, $h_{\tilde{H}} \geq K$ for some constant K. In particular, for a set $S \subseteq \tilde{H}$ with $|S| = t \leq |\tilde{H}|/2$, we have that*

$$e(S, \bar{S}) \geq K \mathrm{vol}(S) = Kt \binom{k}{l} \binom{n-k}{l}.$$

We now prove the following:

Theorem 4. *Let H, as described above, be the subgraph of $K(n, \alpha, \gamma, \beta)$ consisting of vertices of weight k and edges of type (l, l). Then H is connected a.a.s.*

Proof. As observed above H is a percolated version of $\tilde{H} = G(n, k, l)$, where each edge is chosen independently with probability $\beta^l \alpha^{k-l} \beta^l \gamma^{k-l}$. In order to establish the theorem, we prove the following claim:

Claim 1: For all sets $S \subseteq H$ with $|S| \leq |H|/2$, $e(S, \bar{S}) > 0$ a.a.s..

By Lemma 3, the number of edges leaving S with $|S| = t$ in $\tilde{H}$ is at least $Kt\binom{k}{l}\binom{n-k}{l}$. Thus

$$\mathbb{E}[e(S, \bar{S})] \geq Kt \binom{k}{l} \binom{n-k}{l} \beta^l \alpha^{k-l} \beta^l \gamma^{k-l} \geq t \cdot \Theta(c^n)$$

for some constant $c > 1$, by applying Lemma 1. As $e(S, \bar{S})$ is binomially distributed, the Chernoff bounds (see, for example, [1]) imply that

$$\mathbb{P}(e(S, \bar{S}) \leq \frac{1}{2}\mathbb{E}[e(S, \bar{S})]) \leq \exp\left(-\frac{1}{8}\mathbb{E}[e(S, \bar{S})]\right) \leq \exp(-t \cdot \Theta(c^n)).$$

Let $\mathcal{A}$ denote the event that some set S has $e(S, \bar{S}) = 0$. Then by the union bound

$$\begin{aligned}
\mathbb{P}(\mathcal{A}) &\leq \sum_{t=1}^{|H|/2} \sum_{S,|S|=t} \mathbb{P}(e(S,\bar{S})=0) \\
&\leq \sum_{t=1}^{|H|/2} \binom{\binom{n}{k}}{t} \exp(-t \cdot \Theta(c^n)) \\
&\leq \sum_{t=1}^{|H|/2} \left(\binom{n}{k} \exp(-\Theta(c^n))\right)^t \\
&\leq \sum_{t=1}^{|H|/2} (o(1))^t = o(1).
\end{aligned}$$

Note that we used the fact that $\binom{n}{k}\exp(-\Theta(c^n)) = o(1)$. This is easily verified by taking logs: $\log\binom{n}{k} = o(n\log n)$ while $\log(\exp(\Theta(c^n))) = \Theta(c^n)$. This completes the proof of Claim 1, and hence of Theorem 4. □

3 The Giant Component in $K(n,\alpha,\gamma,\beta)$

In this section we complete the proof of Theorem 2, and establish the upper bound in Theorem 3.

We first establish the necessity of the condition in Theorem 2 with the following theorem. We note that a similar technique is used in [5], although we include a proof here for the sake of completeness.

Theorem 5. *Suppose $G = K(n,\alpha,\gamma,\beta)$, where $(\alpha+\beta)(\beta+\gamma) < 1$. Then G contains $N - o(N)$ isolated vertices a.a.s..*

Proof. Let ϵ be such that $(\alpha+\beta)(\beta+\gamma) = 1-\epsilon$. Let v have weight $w \leq n/2+n^{2/3}$. The expected degree of v is

$$\begin{aligned}
(\alpha+\beta)^w(\beta+\gamma)^{n-w} &\leq (\alpha+\beta)^{n/2+n^{2/3}}(\beta+\gamma)^{n/2-n^{2/3}} \\
&\leq ((\alpha+\beta)(\beta+\gamma))^{n/2}\left(\frac{\alpha+\beta}{\beta+\gamma}\right)^{n^{2/3}} \\
&\leq (1-\epsilon)^{\frac{1}{2}\log N}\left(\frac{\alpha+\beta}{\beta+\gamma}\right)^{(\log N)^{2/3}} = o(1)
\end{aligned}$$

Thus, v is isolated a.a.s.. Moreover, the proportion of vertices with weight at most $n/2+n^{2/3}$ is at least $1-e^{-(n/2+n^{2/3})^2/2n} = 1-o(1)$ by the Chernoff bound [1]. Therefore, there are at least $N-o(N)$ isolated vertices in G a.a.s., and thus G has no giant component a.a.s.. □

We now turn our attention to establishing the sufficiency of the condition in Theorem 2. For the remainder of this section, we assume $K(n, \alpha, \gamma, \beta)$ has $(\alpha+\beta)(\beta+\gamma) > 1$, and $\alpha \geq \gamma$, as in Section 2. Also, we set $k = \frac{\alpha+\beta}{\alpha+\gamma+2\beta} n$ as in Section 2. Given $0 \leq s \leq n$, let G_s denote the set of vertices of $K(n, \alpha, \gamma, \beta)$ with weight s.

Theorem 6. *Suppose $s \neq k$, with*

$$s \geq \frac{2\log(n) - n\log(\gamma+\beta) - \log\left(\frac{(\alpha+\beta)(\gamma+\beta)}{2e^2\pi\beta n\sqrt{\alpha\gamma}}\right)}{\log(\frac{\alpha+\beta}{\beta+\gamma})}. \tag{2}$$

Then for every $\mathbf{v} \in G_s$, a.a.s. there exists r with $|r-k| < |s-k|$ such that there is a vertex $v' \in G_r$ with $v \sim v'$.

The precise bound on s in the statement of Theorem 6 is quite technical, and falls out from the proof. Note that

$$s \geq \frac{-\log(\beta+\gamma)n}{-\log(\beta+\gamma)+\log(\alpha+\beta)} + \Theta(\log n).$$

As $(\alpha+\beta)(\beta+\gamma) > 1$, we have that

$$m = \frac{-\log(\beta+\gamma)}{-\log(\beta+\gamma)+\log(\alpha+\beta)} < \frac{1}{2}. \tag{3}$$

In particular, Theorem 6 holds for all vertices with weight at least $\frac{n}{2}$, and hence shows the existence of a giant component of size at least $\frac{N}{2}$. Thus, Theorem 6 is sufficient to complete the proof of Theorem 2.

If $\beta+\gamma > 1$, then note that all non-negative s satisfy (2), that is, the graph is connected (this was proven in [5]).

Proof of Theorem 6. Suppose $\mathbf{v} \in G_s$. The expected number of neighbors of $\mathbf{v}$ of type (l, t) is

$$\binom{s}{l}\binom{n-s}{t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t}.$$

Note that this is (roughly) maximized when $l = \frac{\beta}{\beta+\alpha}s$ and $t = \frac{\beta}{\beta+\gamma}(n-s)$. Setting l and t as such, we note that the weight of a neighbor of $\mathbf{v}$ obtained in such a way is

$$f(s) = \frac{\alpha}{\beta+\alpha}s + \frac{\beta}{\beta+\gamma}(n-s).$$

Note that

$$\begin{array}{ccccl} k & > & f(s) & > & s \text{ when } s < k. \\ k & < & f(s) & < & s \text{ when } s > k. \\ & & f(s) & = & s \text{ when } s = k. \end{array}$$

Therefore, a neighbor of $\mathbf{v}$ of type (l,t) has weight r, with $|r-k| < |s-k|$. Using the entropy bound again, the expected number of such neighbors is

$$\begin{aligned}
&\binom{s}{l}\binom{n-s}{t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t} \\
&> \frac{(\alpha+\beta)(\gamma+\beta)}{2e^2\pi\beta\sqrt{s(n-s)\alpha\gamma}}\left(\frac{\alpha+\beta}{\beta}\right)^l\left(\frac{\alpha+\beta}{\alpha}\right)^{s-l}\left(\frac{\beta+\gamma}{\beta}\right)^t\left(\frac{\beta+\gamma}{\gamma}\right)^{n-s-t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t} \\
&= \frac{(\alpha+\beta)(\gamma+\beta)}{2e^2\pi\beta\sqrt{s(n-s)\alpha\gamma}}(\alpha+\beta)^s(\gamma+\beta)^{n-s} \geq n^2. \qquad (4)
\end{aligned}$$

The lower bound on s in the statement of the theorem is chosen precisely so that (4) holds.

The number of neighbors with weight r is binomially distributed, so the Chernoff bounds imply that for a vertex $\mathbf{v}$ the probability that it has no neighbors with weight r is bounded by

$$\begin{aligned}
\mathbb{P}(e(\mathbf{v},G_r)=0) &\leq \frac{1}{2}\mathbb{E}[e(\mathbf{v},G_r)] \\
&\leq \exp\left(-\frac{1}{8}\mathbb{E}[e(\mathbf{v},G_r)]\right) \\
&\leq \exp\left(-\frac{n^2}{8}\right).
\end{aligned}$$

Note that there are 2^n vertices in $K(n,\alpha,\gamma,\beta)$ so by the union bound

$$\mathbb{P}(\exists\mathbf{v}:\mathbb{E}(\mathbf{v},G_r)=0) \leq 2^n\exp\left(-\frac{n^2}{8}\right) = o(1),$$

completing the proof of the theorem. □

As noted above, this completes the proof of Theorem 2. Moreover, this establishes the upper bound in Theorem 3, as the only vertices not in the giant component a.a.s. have weight at most $mn + o(n)$.

4 The Size of the Giant Component

In this section we complete the proof of Theorem 3. In order to derive the precise result, we must examine the vertices with weight $mn + o(n)$ more closely.

Again, we consider the set G_s of vertices with weight s. For a vertex $\mathbf{v} \in G_s$, its expected degree is

$$\sum_{l=0}^{s}\sum_{t=0}^{n-s}\binom{s}{l}\binom{n-s}{t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t} - \alpha^s\gamma^{n-s} = (\alpha+\beta)^s(\gamma+\beta)^{n-s} - \alpha^s\gamma^{n-s}. \qquad (5)$$

(The $\alpha^s\gamma^{n-s}$ term corresponds to a 'self loop' at a vertex).

In the previous section we used the fact that the sum in (5) is roughly maximized when $l = \frac{\beta}{\alpha+\beta}s$ and $t = \frac{\beta}{\gamma+\beta}s$. In order to establish Theorem 3 we need a more precise understanding of the summation.

Lemma 4. *Let $\epsilon > 0$ be small enough that $(1+\epsilon)\frac{\beta}{\beta+\alpha} < 1$. Moreover, suppose that $s = s(n)$, and there exists r with $0 < r < 1$, and $\frac{s}{n} \to r$. Then*

$$\sum_{l=(1+\epsilon)\frac{\beta}{\beta+\alpha}s}^{s} \sum_{t=0}^{(1-\epsilon)\frac{\beta}{\beta+\gamma}(n-s)} \binom{s}{l}\binom{n-s}{t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t} = o((\alpha+\beta)^s(\gamma+\beta)^{n-s}).$$

Proof. Let

$$g(l,t) = \binom{s}{l}\binom{n-s}{t}\beta^l\alpha^{s-l}\beta^t\gamma^{n-s-t}.$$

Then

$$\frac{g(l+1,t)}{g(l,t)} = \frac{s-l}{l+1}\frac{\beta}{\alpha}.$$

For $l \geq (1+\epsilon)\frac{\beta}{\beta+\alpha}s$,

$$\begin{aligned}
\frac{g(l+1,t)}{g(l,t)} &= \frac{s-(1+\epsilon)\frac{\beta}{\beta+\alpha}s}{(1+\epsilon)\frac{\beta}{\beta+\alpha}s+1}\cdot\frac{\beta}{\alpha} \\
&\leq \frac{\beta+\alpha}{\alpha(1+\epsilon)} - \frac{\beta}{\alpha} \\
&\leq \frac{\alpha-\epsilon\beta}{\alpha(1+\epsilon)} \leq \frac{1}{1+\epsilon}.
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{l=(1+\epsilon)\frac{\beta}{\beta+\alpha}s}^{s} g(l,t) &\leq \sum_{l=(1+\epsilon)\frac{\beta}{\beta+\alpha}s}^{s} (1+\epsilon)^{l-(1+\epsilon)\frac{\beta}{\beta+\alpha}s} g\left((1+\epsilon)\frac{\beta}{\beta+\alpha}s, t\right) \\
&\leq Cg((1+\epsilon)\frac{\beta}{\beta+\alpha}s, t),
\end{aligned}$$

where C is obtained by summing the geometric series.

A similar bound on $\frac{g(l,t-1)}{g(l,t)}$ allows us to derive that

$$\sum_{l=(1+\epsilon)\frac{\beta}{\beta+\alpha}s}^{s} \sum_{t=0}^{(1-\epsilon)\frac{\beta}{\beta+\gamma}s} g(l,t) \leq C'g\left((1+\epsilon)\frac{\beta}{\beta+\alpha}s, (1-\epsilon)\frac{\beta}{\beta+\gamma}s\right). \quad (6)$$

Note that

$$\sum_{(1+\epsilon)\frac{\beta}{\beta+\alpha}s-\log(s)}^{(1+\epsilon)\frac{\beta}{\beta+\alpha}s} \sum_{t=(1-\epsilon)\frac{\beta}{\beta+\gamma}(n-s)}^{(1-\epsilon)\frac{\beta}{\beta+\gamma}(n-s)+\log s} g(l,t) = \omega\left(g\left((1+\epsilon)\frac{\beta}{\beta+\alpha}s, (1-\epsilon)\frac{\beta}{\beta+\gamma}\right)\right).$$

as the sums are bounded below by a geometric series with ratio greater than one. Together with equation (6), this completes the proof. $\square$

We next use the following lemma in order to establish the lower bound in Theorem 3.

Lemma 5. *Let X and m be as in Theorem 3. Then*

$$|X| = \Omega\left(\binom{n}{mn}\right).$$

a.a.s.

Proof. Let $\ell = mn-1$. Then the expected degree of a vertex $\mathbf{v}$ with weight ℓ is

$$\begin{aligned}
\mathbb{E}[\deg(\mathbf{v})] &= (\alpha+\beta)^{mn-1}(\beta+\gamma)^{n-mn+1} \\
&= \frac{\beta+\gamma}{\alpha+\beta}\left((\alpha+\beta)^m(\beta+\gamma)^{1-m}\right)^n \\
&= \frac{\beta+\gamma}{\alpha+\beta} =: q < 1
\end{aligned}$$

by the definition of m. Take $p = \mathbb{P}(\deg(\mathbf{v}) = 0) \geq 1-q > 0$.

Let Y denote the set of isolated vertices with weight ℓ. Then $E[|Y|] = p\binom{n}{\ell}$. We write $|Y| = \sum z_{\mathbf{v}}$ where $z_{\mathbf{v}}$ is the indicator that $\mathbf{v}$ is isolated. The $z_{\mathbf{v}}$ are not independent, however it is easy to observe that

$$\mathbb{E}[z_{\mathbf{v}}z_{\mathbf{u}}] = \mathbb{E}[z_{\mathbf{v}}](\mathbb{E}[z_{\mathbf{u}}] - \mathbb{P}(\mathbf{v}\sim\mathbf{u})) = \mathbb{E}[z_{\mathbf{v}}]\mathbb{E}[z_{\mathbf{u}}] - o(\mathbb{E}[z_{\mathbf{v}}]).$$

Thus

$$\mathrm{Cov}(z_{\mathbf{v}}, z_{\mathbf{u}}) = o(\mathbb{E}[z_{\mathbf{v}}]).$$

Furthermore,

$$\mathrm{Var}(z_{\mathbf{v}}) = p(1-p).$$

Thus

$$\begin{aligned}
\mathrm{Var}(|Y|^2) &= \sum_{\mathbf{v}} \mathrm{Var}(z_{\mathbf{v}}) + \sum_{\mathbf{v}}\sum_{\mathbf{u}\neq\mathbf{v}} \mathrm{Cov}(z_{\mathbf{v}}, z_{\mathbf{u}}) \\
&\leq p(1-p)\binom{n}{\ell} + o\left(p\binom{n}{\ell}^2\right) = o\left(\binom{n}{\ell}^2\right).
\end{aligned}$$

Since $\mathrm{Var}(|Y|^2) = o(\mathbb{E}[|Y|^2])$, Chebyshev's inequality (see, for example, [1]) implies that for any $c > 0$,

$$\mathbb{P}\left(\left||Y| - p\binom{n}{\ell}\right| \geq c\binom{n}{\ell}\right) \leq \frac{o\left(\binom{n}{\ell}^2\right)}{c^2\binom{n}{\ell}^2} = o(1).$$

Therefore, a.a.s. we have that $|Y| = p\binom{n}{\ell} + o\left(\binom{n}{\ell}\right)$, and thus $|X| \geq c\binom{n}{\ell} \geq c'\binom{n}{mn}$ a.a.s., establishing the result. $\square$

Using Lemma 5, we can derive from Theorem 6 that

$$\binom{n}{mn} \ll |X| \ll \binom{n}{mn + C\log(n)}$$

for some absolute constant C, but these differ by a factor polynomial in n (and hence by a poly-logarithmic factor in N.) Here, the $\ll$ symbol is in the traditional number theoretic sense, that is, $f(x) \ll g(x)$ if $f(x) = O(g(x))$.

Proof of Theorem 3. By Lemma 5, it suffices to show that $|X| = O(\binom{n}{mn})$ a.a.s. Let f be as in the proof of Theorem 6.

Suppose $s = mn + O(\log(n))$. Choose $\epsilon > 0$ and small enough that

$$\frac{\alpha - \epsilon\beta}{\beta + \alpha}s + \frac{\beta - \epsilon\gamma}{\beta + \alpha}(n - s) > (m + \epsilon)n;$$

such exists by our observation on $f(s)$.

Note that, by Theorem 6, for n sufficiently large, if $s' \geq (m + \epsilon)n$, then all vertices in $G_{s'}$ are in the giant component a.a.s. Thus, a vertex in G_s which is not in the giant component has no edges into $G_{s'}$ for $s' \geq (m + \epsilon)n$.

Consider a vertex $\mathbf{v}$ in G_s. We say that an edge from $\mathbf{v}$ is *good* if it involves no more than $(1+\epsilon)\frac{\beta}{\beta+\alpha}s$ swaps from 1 to 0 and no fewer than $(1-\epsilon)\frac{\beta}{\beta+\gamma}(n-s)$ swaps from 0 to 1. (Note that when we say an edge $\mathbf{uv}$ incident to $\mathbf{v}$ is good, we are assuming the swaps are from $\mathbf{v}$ to $\mathbf{u}$. In this way, an edge may be good when considered from $\mathbf{v}$ but not from $\mathbf{u}$). Let Y be the set of vertices with no incident good edges. It is easy to check that if $\mathbf{v}$ has an incident good edge, then it is connected to a $G_{s'}$ with $s' \geq (m + \epsilon)n$, so every vertex not in Y is in the giant component, hence $|Y| \geq |X|$.

Let $z_{\mathbf{v}}$ denote the number of good edges incident to $\mathbf{v}$. By Lemma 4

$$\mathbb{E}[z_{\mathbf{v}}] = (1 + o(1))(\alpha + \beta)^s(\gamma + \beta)^{n-s}.$$

Since $z_{\mathbf{v}}$ is the sum of independent indicator functions, we can write:

$$\begin{aligned}
\mathbb{P}(z_{\mathbf{v}} = 0) &= \prod_{\mathbf{v}':e(\mathbf{v},\mathbf{v}')\ good} (1 - \mathbb{P}(\mathbf{v} \sim \mathbf{v}')) \\
&\leq \exp\left(- \sum_{\mathbf{v}':e(\mathbf{v},\mathbf{v}')\ good} \mathbb{P}(\mathbf{v} \sim \mathbf{v}') \right) = \exp(-\mathbb{E}[z_{\mathbf{v}}]).
\end{aligned}$$

Thus, for n sufficiently large,

$$\mathbb{P}(\mathbf{v} \in Y) = \mathbb{P}(z_{\mathbf{v}} = 0) \leq \exp\left(-\frac{1}{2}(\alpha+\beta)^s(\gamma+\beta)^{n-s} \right).$$

By Theorem 6, there exists a C such that if $s > mn + C\log(n)$, then all vertices in G_s are in the giant component a.a.s.. Thus

$$|X| \leq \sum_{s=0}^{mn+C\log(n)} |Y \cap G_s|.$$

Choose t to be the least integer such that

$$\exp\left(-\frac{1}{2}\frac{\alpha-\gamma}{\beta+\gamma}\left(\frac{\alpha+\beta}{\beta+\gamma}\right)^t \right) \frac{1-m}{m} < \frac{1}{2}.$$

Define:

$$g(k) = \exp\left(-\frac{1}{2}(\alpha+\beta)^{mn+t+k}(\gamma+\beta)^{n-mn-t-k} \right) \binom{n}{mn+t+k}.$$

We have chosen t so that for $k \geq 0$,

$$\frac{g(k+1)}{g(k)} < \frac{1}{2}.$$

Consider

$$\begin{aligned}
\mathbb{E}\left[\sum_{s=mn+t}^{mn+C\log(n)} |Y \cap G_s| \right] &\leq \sum_{k=0}^{C\log(n)-t} g(k) \\
&\leq \sum_{k=0}^{C\log(n)-t} 2^{-k} g(0) \\
&\leq 2g(0).
\end{aligned}$$

As $\sum |Y \cap G_s|$ can be written as the sum of independent indicator functions, it is tightly concentrated by the Chernoff bounds and hence a.a.s.

$$\sum_{s=mn+t}^{mn+C\log(n)} |Y \cap G_s| \leq (1+o(1))2g(0) = \Theta\left(\binom{n}{mn+t}\right) = O\left(\binom{n}{mn}\right).$$

Note that

$$\sum_{s=0}^{mn+t} |Y \cap G_s| \leq \sum_{s=0}^{mn+t} \binom{n}{s} = \Theta\left(\binom{n}{mn+t}\right) = \Theta\left(\binom{n}{mn}\right).$$

Thus

$$|X| \leq \sum_{s=0}^{mn+C\log n} |Y \cap G_s| = \Theta\left(\binom{n}{mn}\right) + O\left(\binom{n}{mn}\right) = \Theta\left(\binom{n}{mn}\right),$$

a.a.s., completing the proof. □

5 Conclusions and Open Questions

In this paper, we investigated the critical threshold for the giant component in the random graph model as studied by Mahdian and Xu [5]. There are several interesting related questions still open. In particular, it would be of interest to study the emergence of the giant component where α, β, and γ are allowed the vary with n, and $(\alpha+\beta)(\beta+\gamma) = 1 + o(1)$. The general machinery we build in this paper should be useful for such a study.

In fact, in certain regimes we can see that there is a giant component when $(\alpha+\beta)(\beta+\gamma) = 1 - o(1)$. Under the condition that

$$((\alpha+\beta)(\beta+\alpha))^{\frac{\alpha+\beta}{\alpha+\gamma+2\beta}}(\alpha+\beta)^{\frac{\alpha-\gamma}{\alpha+\gamma+2\beta}} > 1$$

the proof of Theorem 4 will still hold; this can hold even if $(\alpha+\beta)(\beta+\gamma) < 1$. So long as

$$\frac{-\log(\beta+\gamma)}{-\log(\beta+\gamma)+\log(\alpha+\beta)} = \frac{1}{2} - O\left(\frac{1}{\sqrt{n}}\right),$$

one may observe that Theorem 6 will still imply the existence of a giant component.

It would be of interest to identify the precise conditions under which a giant component exists, in particular, in the regime where $(\alpha+\beta)(\beta+\gamma) = 1 + o(1)$ and the graph is sparse; in the case studied here the average degree is polynomial in N.

References

[1] N. Alon, J. Spencer, *The Probabilistic Method* (3rd edition), John Wiley & Sons, 2008.

[2] F. Chung, *Spectral Graph Theory.* CBMS Vol. 92, AMS Publications, 1997.

[3] J. Leskovec, C. Faloutsos, Scalable modeling of real graphs using Kronecker multiplication, International Conference on Machine Learning (ICML) (2007), 497-504.

[4] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication, PKDD (2005), 133-145.

[5] M. Mahdian, Y. Xu, Stochastic Kronecker graphs, Proceedings of the 5th Workshop on Algorithms and Models for the Web-Graph (2007).

[6] P. Stanica, Good lower and upper bounds on binomial coefficients, J. Inequal. Pure Appl. Math. **2**, no. 3, (2001), Article 30.