A Phase Transition for Avoiding a Giant Component

Tom Bohman Jeong Han Kim

July 26, 2004

Abstract

Let c be a constant and $(e_1, f_1), (e_2, f_2), \ldots, (e_{cn}, f_{cn})$ be a sequence of ordered pairs of edges from the complete graph K_n chosen uniformly and independently at random. We prove that there exists a constant c_2 such that if $c > c_2$ then **whp** every graph which contains at least one edge from each ordered pair (e_i, f_i) has a component of size $\Omega(n)$ and if $c < c_2$ then **whp** there is a graph containing at least one edge from each pair that has no component with more than $O\left(n^{1-\epsilon}\right)$ vertices, where ϵ is a constant that depends on $c_2 - c$. The constant c_2 is roughly 0.97677.

1 Introduction

Let c be a constant and $(e_1, f_1), (e_2, f_2), \ldots, (e_{cn}, f_{cn})$ be a sequence of ordered pairs of edges on vertex set $[n] = \{1, 2, \ldots, n\}$ chosen uniformly and independently at random. Our goal is to choose exactly one edge from each pair to form a graph that contains no giant component (i.e. a component of size $\Omega(n)$). For example, if we choose the first edge from every pair, it is well known that there is no giant component if and only if $c < 1/2$ [8]. Of course, we expect that the introduction of a choice will allow us to form graphs with cn edges and no giant component for c larger than $1/2$.

This question was introduced by Achlioptas, who posed an on-line version. In this version the pairs of edges are presented sequentially and the choice of an edge from the presented pair (e_i, f_i) is made on-line, that is, without knowledge of the pairs that have not yet been presented. Thus, the evolution of the graph is guided by the algorithm we use to make the choices. We will refer to any such on-line process as an *Achlioptas process*. Bohman and Frieze introduced an Achlioptas process and proved that with high probability it produces a graph with no giant component for any $c < 0.535$ [3]. Spencer and Wormald recently proved that $c = 0.83$ can be achieved on-line (and gave results that indicate that even $c = 0.89$ can be achieved on-line) [10]. Bohman, Frieze and Wormald established an upper bound on the number of edges in a graph that is produced by an Achlioptas process and has no giant component [4], which we state precisely for reference. Let c_1 be the root of the equation

$$1 - c - 5e^{-4c}/3 - (2c - 9/8)e^{-8c} - e^{-12c}/2 + e^{-16c}/24 = 0.$$

Note that c_1 is roughly 0.96445.

Theorem 1 (Bohman, Frieze, and Wormald). *If c is a constant, A is an algorithm for the on-line choice of one edge from each presented pair in a sequence $(e_1, f_1), (e_2, f_2), \ldots, (e_{cn}, f_{cn})$ of pairs of edges on vertex set $[n]$ chosen uniformly and independently at random, and $c > c_1$ then* **whp** *the graph G produced by A has a giant component.*

In [4] Bohman, Frieze and Wormald also considered a version of the question where the edges do not appear in pairs: If F is a collection of $2cn$ random edges from the complete graph K_n, does there exist a set $E \subseteq F$ such that $|E| = cn$ and the graph $([n], E)$ has no giant component? They determined a threshold for this property, which we now state explicitly. For $c > 1/4$ define $t = t(c) < 1$ by $te^{-t} = 4ce^{-4c}$. Let c_3 be the unique solution to the equation

$$\frac{t^2}{8c} + 1 - \frac{t}{4c} - c = 0.$$

Note that c_3 is roughly 0.97926.

Theorem 2 (Bohman, Frieze and Wormald). *Let c be a constant and let F be a collection of $2cn$ edges on vertex set $[n]$ chosen uniformly and independently at random.*

1. *If $c > c_3$ then* **whp** *all sets $E \in \binom{F}{cn}$ produce graphs $([n], E)$ that have components of size $\Omega(n)$.*

2. *Let $\epsilon > 0$ and $c = c_3 - \epsilon$. There exists a constant C (depending on ϵ) such that* **whp** *there exists $E \in \binom{F}{cn}$ such that the graph $([n], E)$ has no component with more than C vertices.*

They also considered the corresponding on-line question: Suppose $2cn$ random edges are presented sequentially (*not* in pairs) and we use them to form a graph by making an on-line decision whether or not to use each edge as it is presented, with the condition that at least cn edges are accepted by the end of the process. We would expect such processes to form larger graphs without giant components than the Achlioptas processes because here we may choose not to use several consecutive bad edges. They showed that there exists an on-line algorithm that succeeds in producing a graph that **whp** has no giant component for c slightly larger than c_1 [4]. Whether or not the threshold described in Theorem 2 can actually be achieved on-line remains open.

In this paper we consider the off-line version of Achlioptas' original question. Suppose $(e_1, f_1), (e_2, f_2), \ldots, (e_{cn}, f_{cn})$ is a sequence of pairs of edges on vertex set $[n]$ chosen uniformly and independently at random. Let ρ be the solution of the following equation:

$$\frac{\rho(1-e^{-\rho})}{4(1-e^{-\rho}-\rho e^{-\rho})} = 1, \tag{1.1}$$

and set

$$c_2 = \frac{\rho}{4(1-e^{-\rho})^3} \tag{1.2}$$

Note that c_2 is roughly 0.97677. We are now ready to state our main result.

Theorem 3. *Let c be a constant and let $(e_1, f_1), (e_2, f_2), \ldots, (e_{cn}, f_{cn})$ be a sequence of ordered pairs of edges on vertex set $[n]$ chosen independently and uniformly at random.*

(a) If $c > c_2$ then **whp** *every collection of edges E such that $|E \cap \{e_i, f_i\}| \geq 1$ for all i gives a graph $([n], E)$ that has a component of size $\Omega(n)$.*

(b) If $c < c_2$ then **whp** *there exists an edge set E such that $|E \cap \{e_i, f_i\}| \geq 1$ for all i for which the graph $([n], E)$ has no component with more then $O(n^{1-\epsilon})$ vertices where ϵ is a constant that depends on $c_2 - c$.*

Note that Theorems 1 and 3 establish a clear distinction between the on-line and off-line versions of Achlioptas' question: there are values of c (to be precise, $c_1 < c < c_2$) for which the on-line version has a giant while the off-line version does not. As is noted above, this is not known to be the case (and the evidence suggests that it might not be the case) for the problem considered in [4] in which edges appear one at a time rather than in pairs. It is also interesting to note that the transition given by Theorem 3 is not nearly as sharp as that given by Theorem 2, where we go from components of constant size on one side of the threshold to a giant component on the other. The behavior of the off-line version of the Achlioptas process below the critical value c_2 remains an intriguing open question. While there are many possible behaviors, the following seems a natural first question:

Question 4. *Does there exist a function $w(n)$ that tends to infinity as $n \to \infty$ and a constant $c < c_2$ such that* **whp** *every collection of edges E such that $|E \cap \{e_i, f_i\}| \geq 1$ for all i gives a graph $([n], E)$ that has a component of size $\Omega((\log(n))^{w(n)})$?*

We prove Theorem 3 by reducing to a question about edge density in random 4-uniform hypergraphs. Let $\mathcal{H}(n, m; 4)$ be a collection of m 4-element subsets of $[n]$ chosen independently and uniformly at random. Note that, since the expected number of indices i such that $e_i \cap f_i \neq \emptyset$ is a constant, **whp** we have $|\{i : e_i \cap f_i \neq \emptyset\}| \leq \omega(n)$ for any function $\omega(n)$ that tends to infinity as n goes to infinity. It follows that we can simply remove these 'bad' rounds from consideration and assume the distribution of the collection

$$\mathcal{H}(cn) = \{e_1 \cup f_1, e_2 \cup f_2, \ldots, e_{cn} \cup f_{cn}\}$$

is identical to the distribution of $\mathcal{H}(n, cn; 4)$. The *2-core* of a hypergraph $\mathcal{F}$, which we denote $\mathrm{cr}_2(\mathcal{F})$, is the subgraph obtained by iteratively removing vertices of degree 1. In other words, the 2-core is the largest induced sub-hypergraph $\mathcal{G}$ of $\mathcal{F}$ such that the degree of every vertex in $\mathcal{G}$ is at least 2. We define the *density* of a hypergraph $\mathcal{F}$ to be

$$d(\mathcal{F}) = \frac{|\mathcal{F}|}{|V(\mathcal{F})|}.$$

Recall that there is a threshold κ_2 for the existence of a nonempty 2-core in $\mathcal{H}(n, cn; 4)$, and note that the expected density of $\mathrm{cr}_2(\mathcal{H}(n, cn; 4))$ is continuous and increasing for $c > \kappa_2$.

The constant c_2 is the value of c for which

$$E[d(\mathrm{cr}_2(\mathcal{H}(n, cn; 4)))] = 1. \tag{1.3}$$

This follows, for example, from an application of the Poisson cloning technique recently introduced by the second author [9] (or from recent results of Cooper [7]). A key ingredient of our proof of Theorem 3 is showing that c_2 is the critical value for the property that $\mathcal{H}(n, cn; 4)$ has an induced subgraph $\mathcal{F}$ such that $d(\mathcal{F}) \geq 1$.

Theorem 5. *If $c < c_2$ then* **whp** *we have*

$$\frac{|\mathcal{F}|}{|\cup_{e \in \mathcal{F}} e|} < 1$$

for all $\mathcal{F} \subseteq \mathcal{H}(n, cn; 4)$.

We prove Theorem 5 in Section 3. In Section 4 we use Theorem 5 to prove part (b) of Theorem 3. The proof of part (a) of Theorem 3 follows directly from a lemma in [4]. For the sake of completeness, this Lemma is given in the next Section.

Finally, we note that the ideas introduced in Section 4 extend to a class of related problems. Suppose, for example, that we are given a sequence of k element sets $A_1, A_2, \ldots, A_{cn}$ chosen uniformly and independently at random from $[n]$ and we form a graph G by choosing one edge from $\binom{A_i}{2}$ for $i = 1, \ldots, cn$. A simple modification of Section 4 (and Lemma 6) show that the critical value for existence of a giant in G is the threshold for the existence of a sub-hypergraph of density 1 in the random hypergraph $\mathcal{H}(n, cn; k)$. For the sake of brevity, we do not state a full generalization.

2 Upper Bound

Note that if $c > c_2$ then **whp** the density of the 2-core of the hypergraph $\mathcal{H}(cn)$ is strictly greater than 1. In other words, there exists $S \subseteq [n]$ such that

$$|\{A \in \mathcal{H}(cn) : A \subseteq S\}| > |S|(1 + \epsilon)$$

for some $\epsilon > 0$. Therefore for any choice of edges the graph G formed is a subset of the random graph $G_{n,2cn}$ with the property that there is a set of vertices S such that $|E(G[S])| > (1+\epsilon)|S|$. It follows from the following Lemma from [4] that $G[S]$ contains a giant component. The proof of the Lemma is reproduced here for the sake of completeness.

Let H be the random graph $G_{n,m}$. For constants $\epsilon, \delta > 0$ let $\mathcal{A}_{\epsilon,\delta}$ be the event that there exists $S \subseteq [n]$ such

(a) $|S| < \delta n$,

(b) the graph $H[S]$ contains more than $(1+\epsilon)|S|$ edges.

Lemma 6. *If $\epsilon > 0$, $\delta = \delta(\epsilon) = 2\epsilon\,(4ce)^{-1-1/\epsilon}$ and $m > 2cn$ then $\Pr(\mathcal{A}_{\epsilon,\delta}) = o(1)$.*

Proof. Since the property in question is monotone increasing we can work within the independent model $H' = G_{n,4c/n}$ (see Theorem 2.2 in [6]). Also, we can assume without loss of generality that $H'[S]$ is connected.

We bound the probability of the existence of S in one of two ways, depending on $s := |S|$. First, assume $4 \le s \le (\log n)/6$ and let A_s be the event that there exists $S \in \binom{[n]}{s}$ such that $H'[S]$ is a spanning connected graph containing at least $s+1$ edges. We have

$$
\begin{aligned}
\Pr(A_s) &\le \binom{n}{s} s^{s-2} \binom{\binom{s}{2}}{2} p^{s+1} \\
&\le \left(\frac{ne}{s}\right)^s s^{s-2} \frac{s^4}{8} \left(\frac{4c}{n}\right)^{s+1} \\
&= \frac{cs^2}{2n} (4ce)^s \\
&= o(n^{-1/2}).
\end{aligned}
$$

For $\log n/6 < s < \delta n$ let A_s be the event that there exists $S \in \binom{[n]}{s}$ such that $H'[S]$ is a spanning connected graph containing at least $(1+\epsilon)s$ edges.

$$
\begin{aligned}
\Pr(A_s) &\le \binom{n}{s} s^{s-2} \binom{\binom{s}{2}}{\epsilon s} p^{s(1+\epsilon)} \\
&\le \frac{(ne)^s}{s^2} \left(\frac{se}{2\epsilon}\right)^{\epsilon s} \left(\frac{4c}{n}\right)^{(1+\epsilon)s} \\
&= \frac{1}{s^2} \left[4ce \left(\frac{4ce}{2\epsilon}\right)^{\epsilon} \left(\frac{s}{n}\right)^{\epsilon}\right]^s \\
&< \frac{1}{s^2}.
\end{aligned}
$$

Thus

$$\sum_{s=4}^{\delta n} \Pr(A_s) = o(1)$$

as required. □

3 Density of Induced Sub-hypergraphs

A nonempty subset U of the vertex set of $\mathcal{H}(cn)$ is called 1-*dense* if the number of edges of $\mathcal{H}(cn)$ contained in U is at least $|U|$. In this section we show that if $c < c_2 = 0.97677\cdots$, then there is no 1-dense subset in $\mathcal{H}(cn)$.

The 2-core and Poisson cloning. The 2-core $\mathrm{cr}_2(\mathcal{H}(cn))$ of $\mathcal{H}(cn)$ is the largest subhypergraph with minimum degree at least 2. The vertex set of $\mathrm{cr}_2(\mathcal{H}(cn))$ is denoted by $V_2(cn)$, or simply V_2 if the identity of c is clear. The 2-core sometimes refers to $V_2(cn)$. This causes no confusion since $\mathrm{cr}_2(\mathcal{H}(cn))$ is the subhypergraph of $\mathcal{H}(cn)$ induced by $V_2(cn)$. If such a graph does not exist, $V_2(cn)$ is empty. To obtain the 2-core, one iteratively removes vertices of degree 1 or 0 or takes the union of all subhypergraphs with the property.

Suppose $\mathcal{H}(cn)$ contains a 1-dense subset. Then each vertex in a minimal 1-dense subset is contained in at least two edges of it. Otherwise, one might remove a vertex of degree 1 or 0 to obtain a smaller 1-dense subset. Thus the 2-core contains all minimal 1-dense subsets. Therefore, it is enough to show that for $c < c_2$

$$\Pr\Big(\text{the 2-core of } \mathcal{H}(cn) \text{ contains a 1-dense subset}\Big) = o(1). \tag{3.1}$$

In [9], the second author introduced the Poisson cloning model $\mathcal{H}_{PC}(n,p\,;k)$ for random k-uniform hypergraphs, $k \geq 2$, and proved that the model is essentially the same as the classical random k-uniform hypergraph model $\mathcal{H}(n,p\,;k)$ as long as the expected average degree is bounded. $\mathcal{H}_{PC}(n,p\,;k)$ is generated in three steps. First, the vertex degrees are chosen (without any hypergraph yet in play). For every vertex v, its degree $d(v)$ is an independent Poisson $\lambda := p\binom{n-1}{k-1}$ random variable. Then, for each vertex v, $d(v)$ copies, or *clones*, of v are introduced. If the sum of all degrees is divisible by k, we choose a perfect k-matching on the set of all clones uniformly at random (i.e. a set of $\frac{1}{k}\sum_{v\in V} d(v)$ pairwise disjoint sets of k clones). If the sum is not divisible by k, one may define a perfect k-matching as $\lfloor \sum_{v\in V} d(v)/k \rfloor$ pairwise disjoint edges of k clones and one additional defected edge consisting of the remaining clones. Or, one may avoid this case by conditioning on the event that $\sum_{v\in V} d(v)$ is divisible by k. Here, it is convenient to use the conditional model. The random (multi)hypergraph $\mathcal{H}_{PC}(n,p;k)$ is the projection, or contraction, of the clones of v to v of the uniform random perfect k-matching. Note that the last two steps in this model (i.e. the choice of a random k-matching on the set of all clones followed by the contraction of the clones of v to v) is the same as the configuration model which produces a random hypergraph with a fixed degree sequence, see e.g. [1, 2, 5, 11].

Two theorems, one for the essential equivalence and the other for properties of the 2-core of $\mathcal{H}_{PC}(n,p\,;k)$, will be needed. More details about the model and proofs of the theorems can be found in [9]. Though it is possible, we do not have to introduce similar theorems for $\mathcal{H}(cn)$ because $\mathcal{H}(cn) \subseteq \mathcal{H}(n,p\,;4)$ **whp** as long as $p\binom{n}{4} \geq cn+\sqrt{n\log n}$, or equivalently $\frac{p}{4}\binom{n-1}{3} \geq c+\sqrt{\log n/n}$.

Theorem 7. *Let $k \geq 2$ be a fixed integer. If $p\binom{n-1}{k-1} = O(1)$, then there are positive constants b_1 and b_2 so that for any collection $\mathcal{H}$ of simple k-uniform hypergraphs*

$$\begin{aligned} b_1 \Pr(\mathcal{H}_{PC}(n,p\,;k) \in \mathcal{H}) - e^{-\Omega(pn^k)} &\leq \Pr(\mathcal{H}(n,p\,;k) \in \mathcal{H}) \\ &\leq b_2\Big(\Pr(\mathcal{H}_{PC}(n,p\,;k) \in \mathcal{H})\Big)^{1/k} + e^{-\Omega(pn^k)}. \end{aligned}$$

We apply Theorem 7 to the collection $\mathcal{H}$ of all 4-uniform hypergraphs containing 1-dense subhypergraphs. In other words, we will show that

$$\Pr(\mathcal{H}_{PC}(n,p\,;4) \in \mathcal{H}) = o(1) \quad \text{for all} \quad c := \frac{p}{4}\binom{n-1}{3} < c_2.$$

Theorem 7 then yields the desired result.

Let

$$\rho(c) = 4\theta^3 c, \quad \text{where } \theta = \theta(c) \text{ is the largest solution of } \ \theta = 1 - e^{-4\theta^3 c}.$$

It is easy to check $\theta(c) > 0$ if and only if

$$c \geq \kappa_2 := \inf_{\rho>0} \frac{\rho}{4(1-e^{-\rho})^3} \approx 0.77228\cdots \quad ,$$

and $\theta(c) = 0$ otherwise. Let $\mathbf{n}_2(c)$ be the number of vertices in the 2-core of $\mathcal{H}_{PC}(n,p\,;4)$, $\mathbf{a}(c)$ be the ratio $\mathbf{n}_2(c)/n$ and $\mathbf{a}_i(c)n$ be the number of vertices of degree i in the 2-core. It is known that **whp**

$$\mathbf{a}(c) = \Pr\Big(\mathrm{Poi}(\rho(c)) \geq 2\Big) + O\Big(\frac{\log n}{n^{1/2}}\Big) = 1-(1+\rho(c))e^{-\rho(c)} + O\Big(\frac{\log n}{n^{1/2}}\Big)$$

and

$$\mathbf{a}_i(c) = \Pr\Big(\mathrm{Poi}(\rho(c)) = i\Big) + O\Big(\frac{\log n}{n^{1/2}}\Big) = \frac{\rho^i(c)e^{-\rho(c)}}{i!} + O\Big(\frac{\log n}{n^{1/2}}\Big)$$

for all $i = 2, 3, \ldots$, and the number $\mathbf{m}(c)$ of edges in the 2-core satisfies

$$\mathbf{m}(c) = \frac{n}{4}E\Big(\mathrm{Poi}(\rho(c))\mathbf{1}_{\mathrm{Poi}(\rho(c))\geq 2}\Big) + O\Big(n^{1/2}\log n\Big) = \frac{\rho(c)n}{4}(1-e^{-\rho(c)}) + O\Big(n^{1/2}\log n\Big).$$

Let

$$r(c) := \frac{\rho(c)(1-e^{-\rho(c)})}{4-4(1+\rho(c))e^{-\rho(c)}} \quad .$$

Then the density $\frac{\mathbf{m}(c)}{\mathbf{n}_2(c)}$ of the 2-core is $r(c) + o(1)$ **whp**. Notice that $r(c)$ strictly increases as $c \geq \kappa_2$ increases and c_2 is the unique solution of $r(c_2) = 1$. Numerically, $c_2 = 0.97676\cdots$ and corresponding $\theta(c_2) = 0.97250\cdots$, $\rho(c_2) = 3.59351\cdots$. Let

$$\alpha = 1 - (1 + \rho(c_2))e^{-\rho(c_2)}, \quad \text{and} \quad \alpha_i = \frac{\rho^i(c_2)e^{-\rho(c_2)}}{i!}.$$

So, **whp** we have

$$\mathbf{a}(c_2) = \alpha + O\Big(\frac{\log n}{n^{1/2}}\Big) \quad \text{and} \quad \mathbf{a}_i(c_2) = \alpha_i + O\Big(\frac{\log n}{n^{1/2}}\Big).$$

Numerically,

$$\alpha = 0.87367\cdots, \quad \beta_2 := \alpha_2/\alpha = 0.20324\cdots, \quad \beta_3 := \alpha_3/\alpha = 0.24345\cdots$$
$$\text{and} \quad \beta_4 := \alpha_4/\alpha = 0.21871\cdots.$$

Note that the number of edges $\mathbf{m}(c_2)$ in the 2-core is $(\alpha + o(1))n$ **whp**.

The second theorem we need is as follows.

Theorem 8. *For fixed $c = c_2 - \delta$ and $\mathbf{b}_i(c) = \mathbf{a}_i(c)/\mathbf{a}(c)$,* **whp** *in $\mathcal{H}_{PC}(n, p; 4)$ with $p = 4c\binom{n-1}{3}^{-1}$ we have*

$$\mathbf{a}(c) = \alpha - \Theta(\delta), \quad \frac{\mathbf{m}(c)}{\mathbf{n}_2(c)} = 1 - \Theta(\delta), \tag{3.2}$$

and

$$\mathbf{b}_i(c) = \beta_i + O(\delta) \;\; \forall\, 2 \leq i \leq 10\log n, \quad \mathbf{b}_i(c) = 0 \;\; \forall\, i = 0, 1 \text{ or } i \geq 10\log n. \tag{3.3}$$

In particular, the density of the 2-core is less than 1. Moreover, if $\mathbf{a}(c)$ and $\mathbf{b}_i(c)$ are given, the 2-core is isomorphic to the cloning model conditioned on a degree sequence with $\mathbf{b}_i(c)\mathbf{n}_2$ vertices of degree i.[1]

Remark Notice that $\mathbf{m}(c) = \frac{1}{4}\sum_{i\geq 2}\mathbf{b}_i(c)\mathbf{n}_2$ and (3.2) yield

$$\frac{\mathbf{m}(c)}{n} = \alpha - \Theta(\delta).$$

We henceforth assume that $\mathbf{a}(c) = \alpha(c)$, and $\mathbf{b}_i(c) = \beta_i(c)$ are given and that they satisfy (3.2) and (3.3). Then $\mathbf{n}_2(c) = \alpha(c)n$ and $\mathbf{m}(c) = \frac{1}{4}\sum_{i\geq 2}\beta_i(c)\alpha(c)n$ are also given. For the sake of simplicity, we write n_2, m for $\mathbf{n}_2(c), \mathbf{m}(c)$, respectively, and V_2 for the vertex set of the 2-core. In the rest of this section, we will analyze the cloning model conditioned on a fixed degree sequence with $\beta_i(c)n_2$ vertices of degree i (so, $V_2 = \cup_{i=2}^{10\log n} B_i$ where $|B_i| = \beta_i(c)n_2$, all vertices in B_i have degree i, and the 2-core

[1]We recently learned that Cooper [7] also proved a similar result that yields this theorem.

is chosen uniformly at random from the collection of all 4-uniform hypergraphs on V_2 with these fixed degrees). In terms of the new notation, (3.2) and (3.3) mean

$$\alpha(c) = \alpha - \Theta(\delta), \quad \frac{m}{n_2} = 1 - \Theta(\delta), \tag{3.4}$$

and

$$\beta_i(c) = \beta_i + O(\delta) \ \ \forall\ 2 \le i \le 10 \log n, \quad \beta_i(c) = 0 \ \ \forall\ i \ge 10 \log n\,. \tag{3.5}$$

Bins and balls. Let U be a maximal 1-dense subset of the 2-core. Since U cannot be the entire vertex set of the 2-core **whp** by (3.4), the number of edges in U must be exactly $|U|$. Otherwise one may find a larger 1-dense subset by adding any vertex not in U. Moreover, if there is an edge e which contains exactly three vertices in U, then a larger 1-dense subset may be obtained by adding the vertex in e not in U, but this is not possible since U is maximal. Therefore, if U is a maximal 1-dense subset of V_2, then

$$|U \cap f| \neq 3 \quad \text{for all edges } f \text{ in } V_2. \tag{3.6}$$

For a subset W of V_2, let $P(W)$ denote the probability that its complement $U = V_2 \setminus W$ is a maximal 1-dense subset of V_2. If W has $w_i n_2$ vertices of degree i (i.e. if $|W \cap B_i| = w_i n_2$), then there are $\sum_{i=2}^{\infty} i w_i n_2$ clones in W. Define $w = \sum_{i\ge 2} w_i = |W|/n_2$,

$$p_W = \frac{n_2}{4m} \sum_{i=2}^{\infty} i w_i \ \text{ and } \ q_W = 1 - p_W. \tag{3.7}$$

So, W contains $4p_W m$ clones, and the probability that a clone chosen at random lies in W is p_W.

Now, we color clones in W blue and the other clones red. A random perfect 4-matching on the set of clones may be generated by throwing $4p_W m$ blue clones and $4q_W m$ red clones into m bins so that each bin has exactly 4 clones. Let $K(m, p_W)$ be the number of such assignments. If U is maximal 1-dense, then there are exactly $|U| = (1-w)n_2$ bins containing 4 red clones. Furthermore, (3.6) implies that there is no bin containing exactly three red clones. Let $L(m, p_W, (1-w)n_2)$ be the number of assignments that respect these two conditions. Then we have

$$P(W) \le \frac{L(m, p_W, (1-w)n_2)}{K(m, p_W)}.$$

Lemma 9. *Let Z_j's be i.i.d random variables with* $\Pr(Z_1 = 2) = 6/11, \Pr(Z_1 = 3) = 4/11$ *and* $\Pr(Z_1 = 4) = 1/11$. *Then*

$$\frac{L(m,p,un_2)}{K(m,p)} = (1+o(1))(2\pi m)^{1/2} \binom{m}{un_2} p^{4pm} q^{4qm} 11^{m-un_2} \Pr\Big(\sum_{j=1}^{m-un_2} Z_j = 4pm\Big).$$

In particular,

$$\frac{L(m,p,un_2)}{K(m,p)} \le (1+o(1))(2\pi m)^{1/2}\binom{m}{un_2} p^{4pm} q^{4qm} 11^{m-un_2}.$$

Proof. We will express $K = K(m,p)$ and $L = L(m,p,un_2)$ in terms of independent $X = \mathrm{Poi}(4pm)$ and $Y = \mathrm{Poi}(4qm)$. We choose X blue clones and Y red clones and independently throw each clone into a bin chosen uniformly at random among m bins. Conditioning on $X = 4pm$ and $Y = 4qm$, there are m^{4m} different assignments and all are equally likely. Thus counting symmetries we have

$$K = \frac{m^{4m}}{m!} \Pr[X_i + Y_i = 4 \ \ \forall\, i = 1, ..., m | X = 4pm, Y = 4qm],$$

where X_i and Y_i are the numbers of blue and red clones in bin i, respectively. Similarly,

$$L = \frac{m^{4m}}{m!} \Pr[X_i + Y_i = 4,\ Y_i \neq 3 \ \ \forall\, i \ \text{ and } \ |\{i : Y_i = 4\}| = un_2 | X = 4pm, Y = 4qm].$$

Using $\Pr[A|B] = \Pr[B]^{-1}\Pr[A]\Pr[B|A]$ and $\Pr[A \cap C|B] = \Pr[B]^{-1}\Pr[A]\Pr[B \cap C|A]$, we have that

$$\frac{L}{K} = \frac{\Pr[X = 4pm, Y = 4qm, \quad Y_i \neq 3\ \forall i \ \text{ and } \ |\{i : Y_i = 4\}| = un_2 | X_i + Y_i = 4 \ \ \forall\, i]}{\Pr[X = 4pm, Y = 4qm | X_i + Y_i = 4 \ \ \forall\, i]}.$$

It is well-known that all X_i, Y_i are independent and $X_i = \mathrm{Poi}(4p)$ and $Y_i = \mathrm{Poi}(4q)$. Furthermore, conditioned $X_i + Y_i = 4$ for all i, X_i's are i.i.d with $X_1 = \mathrm{Bin}(4,p)$. Since $X_i + Y_i = 4\ \forall i$ yields $X + Y = 4m$, and then $X = 4pm$ if and only if $Y = 4qm$,

$$\begin{aligned}
\Pr\left[X = 4pm, Y = 4qm \middle| X_i + Y_i = 4 \ \ \forall\, i\right] &= \Pr\left[X = 4pm \middle| X_i + Y_i = 4 \ \ \forall i\right] \\
&= \Pr\left[\sum_{i=1}^{m} X_i^* = 4pm\right] \\
&= \frac{1+o(1)}{(2\pi m)^{1/2}},
\end{aligned}$$

where X_i^*'s are i.i.d $\mathrm{Bin}(4,p)$.

For the numerator, suppose there are B_j bins containing j blue clones, $j = 0, ..., 4$. Since B_0 and B_1 are required to be un_2 and 0, respectively, $B_2 + B_3 + B_4 = m - un_2$ and $2B_2 + 3B_3 + 4B_4 = 4pm$. So, we have that the probability is

$$\sum_{(b_2,b_3,b_4)}^{*} \Pr\left[B_0 = un_2, B_2 = b_2 m, B_3 = b_3 m, B_4 = b_4 m \middle| X_i + Y_i = 4 \ \ \forall\, i\right],$$

where the $*$ in the sum means that it is taken for all triples (b_2, b_3, b_4) such that $b_j m$'s are nonnegative integers and satisfy

$$b_2 m + b_3 m + b_4 m = m - un_2 \ \text{ and } \ 2b_2 m + 3b_3 m + 4b_4 m = 4pm. \tag{3.8}$$

Let

$$P(b_2,b_3,b_4) = \Pr\left[B_0 = un_2, B_2 = b_2m, B_3 = b_3m, B_4 = b_4m \middle| X_i + Y_i = 4 \ \ \forall\ i\right]$$

Since $X_i \in \mathrm{Bin}(4,p)$, $Y_i = 4 - X_i \in \mathrm{Bin}(4,q)$ under the condition $X_i + Y_i = 4$ and all X_i are independent,

$$\begin{aligned}
P(b_2,b_3,b_4) &= \binom{m}{un_2}\binom{m-un_2}{b_2m\ b_3m\ b_4m} q^{4un_2}(6p^2q^2)^{b_2m}(4p^3q)^{b_3m}p^{4b_4m} \\
&= \binom{m}{un_2} p^{4pm}q^{4qm}\binom{m-un_2}{b_2m\ b_3m\ b_4m} 6^{b_2m}4^{b_3m} \\
&= \binom{m}{un_2} p^{4pm}q^{4qm}11^{m-un_2}\binom{m-un_2}{b_2m\ b_3m\ b_4m}\left(\frac{6}{11}\right)^{b_2m}\left(\frac{4}{11}\right)^{b_3m}\left(\frac{1}{11}\right)^{b_4m}.
\end{aligned}$$

It is easy to check that

$$\sum_{(b_2,b_3,b_4)}^{*}\binom{m-un_2}{b_2m\ b_3m\ b_4m}\left(\frac{6}{11}\right)^{b_2m}\left(\frac{4}{11}\right)^{b_3m}\left(\frac{1}{11}\right)^{b_4m} = \Pr\left[\sum_{j=1}^{m-un_2} Z_j = 4pm\right].$$

$\square$

Notice that (3.8) implies that

$$L(m,p,un_2) = 0 \ \text{ unless } \ pm \le m - un_2.$$

Since $u = 1 - w$, this fact together with $n_2 = (1+\Theta(\delta))m$ (which follows from the second equality in (3.4)) implies that

$$P(W) = 0 \ \text{ unless } \ p_W \le w.$$

On the other hand, the same equality yields

$$p_W = \frac{n_2}{4m}\sum_{i\ge 2} iw_i \ge \frac{1}{2}\sum_{i\ge 2} w_i = w/2.$$

Let $\mathcal{W}(w,p)$ be the collection of all subsets W of the 2-core with $|W| = wn_2$ and $p_W = p$, and $N(w,p) = |\mathcal{W}(w,p)|$. Then

$$\Pr\left(\exists \text{ 1-dense subset in the 2-core}\right) \le \sum_{\substack{w,p \\ \frac{w}{2}\le p\le w}} \sum_{W\in\mathcal{W}(w,p)} P(W).$$

Lemma 9 yields the following corollary.

Corollary 10. *For $|W| = wn$ and $p_W = p$,*

$$P(W) \le (1+o(1))(2\pi m)^{1/2}\binom{n_2}{wn_2} p^{4pm} q^{4qm} 11^{wm}, \tag{3.9}$$

and for $P(w,p) = \Pr(\exists$ 1-dense subset of size $(1-w)n_2$ with $4(1-p)m$ clones$)$,

$$P(w,p) \le (1+o(1))(2\pi m)^{1/2} N(w,p)\binom{n_2}{wn_2} p^{4pm} q^{4qm} 11^{wm}. \tag{3.10}$$

Small 1-dense subsets. Let $|U| = un$ with $4/n \le u \le 0.5$. Since the probability that each random edge in $\mathcal{H}(cn)$ is contained in U is $\binom{un}{4}/\binom{n}{4} \le u^4$, we have

$$\sum_{U:|U|=un} \Pr[U \text{ is 1-dense}] \le \binom{n}{un}\binom{cn}{un} u^{4un} \le \binom{n}{un}^2 u^{4un} c^{un}.$$

Clearly,

$$\Pr(\exists \text{ 1-dense } U \text{ with } |U| \le 0.5n) \le \sum_{u:4/n\le u\le 0.5} \binom{n}{un}^2 u^{4un} c^{un} = O(n^{-8}).$$

One may easily show the same bound for $\mathcal{H}_{PC}(n,p;4)$ with $p = 4c\binom{n-1}{3}^{-1}$. Since $0.5n \ge 0.57n_2$,

$$\Pr\left(\exists \text{ 1-dense subset in the 2-core of size} \le 0.57n_2\right) = O(n^{-8}).$$

Large 1-dense subsets of the 2-core. Suppose w, p are fixed. Since $m \le (1-\Omega(\delta))n_2$ by (3.4), and the number of possible values of (w,p) is $O(n_2^2)$, it is enough by (3.10) to show that

$$N(w,p)\binom{n_2}{wn_2} p^{4pm} q^{4qm} 11^{wm} \le e^{-\Omega(\delta)n_2}$$

for any fixed $\Omega(\delta) \le w \le 0.43$ and $0.5w \le p \le w$, or equivalently

$$I(w,p) := H(w) - 4rH(p) + rw\log 11 + \frac{1}{n_2}\log N(w,p) \le -\Omega(\delta),$$

where $r = m/n_2 = 1-\Theta(\delta)$ and $H(w) = -w\log w - (1-w)\log(1-w)$. We find that only elementary calculus is needed if we consider the following five cases separately;

Case (0) $0.7w \le p \le w$,
Case (1) $0.65w \le p \le 0.7w$,
Case (2) $0.6w \le p \le 0.65w$,

Case (3) $0.55w \le p \le 0.6w$, and
Case (4) $0.5w \le p \le 0.55w$.

For Case (0), using

$$N(w,p) \le \binom{n_2}{wn_2} \quad \text{or} \quad \log N(w,p) \le H(w)n_2,$$

we have

$$I(w,p) \le 2H(w) - 4rH(p) + rw\log 11.$$

Since $r = 1 - \Theta(\delta)$ and $H(p)$ decreases as $p \le w \le 0.43$ decreases,

$$I(w,p) \le 2H(w) - 4H(0.7w) + w\log 11 + O(\delta w \log(1/w)).$$

Let

$$\phi_0(w) := 2H(w) - 4H(0.7w) + w\log 11.$$

For $\Omega(\delta) \le w \le 0.43$, $\phi_0(w)$ is convex since

$$\phi_0''(w) = -\frac{2}{w} - \frac{2}{1-w} + \frac{4\cdot 0.7}{w} + \frac{4\cdot 0.7^2}{1-0.7w} > 0,$$

and hence $\phi_0(w)$ attains its maximum at one of the two end points. Since $\phi_0(w) = (0.8 - O(\delta))(w + o(1))\log w$ as w tends to 0 and $\phi_0(0.43) = -0.04911\cdots$,

$$I(w,p) \le \phi_0(w) + O(\delta w\log(1/w)) \le -\Omega(\delta).$$

For all the other cases, we need to analyze $N(w,p)$ more carefully. First,

$$N(w,p) = \sum_{\substack{(w_i)_{i=2,3,\ldots,10\log n} \\ \sum w_i = w, \sum iw_i = 4rp}} \prod \binom{\beta_i(c)n_2}{w_i n_2} \le n_2^{10\log n} \max_{\substack{(w_i)_{i=2,3,\ldots,10\log n} \\ \sum w_i = w, \sum iw_i = 4rp}} \prod \binom{\beta_i(c)n_2}{w_i n_2}.$$

Since $4p = r^{-1}\sum_{i=2}^{\infty} iw_i \ge \sum_{i=2}^{\infty} iw_i$,

$$4p \ge 2w_2 + 3\sum_{i=3}^{\infty} w_i = 3w - w_2 \quad \text{or} \quad w_2 \ge 3w - 4p. \tag{3.11}$$

Similarly,

$$4p \ge 2w_2 + 3w_3 + 4\sum_{i=4}^{\infty} w_i \ge 4w - 2w_2 - w_3, \quad \text{or} \quad w_3 \ge 4(w-p) - 2w_2, \tag{3.12}$$

which, in particular, gives

$$w_2 + w_3 \ge 2(w-p). \tag{3.13}$$

For Case (1), since $\beta_2(c) + \beta_3(c) = \beta_2 + \beta_3 + O(\delta) \le 0.45 + O(\delta)$ by (3.5), $w_2 + w_3 \ge 2(w - 0.7w) = 0.6w$ yields

$$\begin{aligned} N(w,p) &\le \binom{(\beta_2(c)+\beta_3(c))n_2}{0.6wn_2}\binom{(1-\beta_2(c)-\beta_3(c))n_2}{0.4wn_2} e^{O(\log^2 n)} \\ &= \binom{(\beta_2+\beta_3)n_2}{0.6wn_2}\binom{(1-\beta_2-\beta_3)n_2}{0.4wn_2} e^{O(\log^2 n + \delta w n_2)} \end{aligned}$$

or

$$\frac{1}{n_2}\log|N(w,p)| \le (\beta_2+\beta_3)H\Big(\frac{0.6w}{\beta_2+\beta_3}\Big) + (1-\beta_2-\beta_3)H\Big(\frac{0.4w}{1-\beta_2-\beta_3}\Big) + O(\delta w).$$

Thus we have

$$\begin{aligned} I(w,p) &\le H(w) - 4rH(0.65w) + rw\log 11 + \psi_1(w) + O(\delta w) \\ &= H(w) - 4H(0.65w) + w\log 11 + \psi_1(w) + O(\delta w \log(1/w)), \end{aligned}$$

where $\psi_1(w) = (\beta_2+\beta_3)H(0.6w/(\beta_2+\beta_3)) + (1-\beta_2-\beta_3)H(0.4w/(1-\beta_2-\beta_3))$.

Let

$$\phi_1(w) := H(w) - 4H(0.65w) + w\log 11 + \psi_1(w).$$

Then, for $\Omega(\delta) \le w \le 0.43$,

$$\begin{aligned} \phi_1''(w) &= -\frac{1}{w} - \frac{1}{1-w} + \frac{4\cdot 0.65}{w} + \frac{4\cdot 0.65^2}{1-0.65w} - \frac{0.6}{w} \\ &\qquad - \frac{0.36}{\beta_2+\beta_3-0.6w} - \frac{0.4}{w} - \frac{0.16}{1-\beta_2-\beta_3-0.4w} \\ &= \frac{0.6}{w} + \frac{1.69}{1-0.65w} - \frac{1}{1-w} - \frac{0.36}{\beta_2+\beta_3-0.6w} - \frac{0.16}{1-\beta_2-\beta_3-0.4w}, \end{aligned}$$

and it decreases as w increases up to 0.43 (consider the first two terms together). Notice that $\phi_1''(0.39) = 0.0677\cdots > 0$ and $\phi_1''(0.4) = -0.03127\cdots < 0$. This means that $\phi_1(w)$ is convex for $\Omega(\delta) \le w \le 0.39$, and, for $0.39 \le w \le 0.43$, $\phi_1'(w)$ initially increases and then keeps decreasing after a certain point between 0.39 and 0.4, which in turn implies that the minimum of $\phi_1'(w)$ over $0.39 \le w \le 0.43$ is attained at either $w = 0.39$ or $w = 0.43$. Since both of $\phi'(0.39) = 0.3537\cdots$ and $\phi'(0.43) = 0.3484\cdots$ are positive, we have $\phi_1'(w) > 0$ on the interval. Hence $\phi_1(w)$ is increasing for $0.39 \le w \le 0.43$. Since $\phi_1(w) = (0.6 + o(1))w\log w$ as $w \to 0$, and $\phi_1(0.43) = -0.00835\cdots + O(\delta)$, we conclude that

$$I(w,p) \le \phi_1(w) + O(\delta w \log(1/w)) \le -\Omega(\delta).$$

It is convenient to express Cases (i), $i = 2,3,4$, as

$$\text{Case (i): } (0.7 - 0.05i)w \le p \le (0.75 - 0.05i)w.$$

For Case (i), we have

$$\begin{aligned} w_2 &\geq 3w - 4(0.75-0.05i)w = 0.2iw, \quad \text{and} \\ w_2 + w_3 &\geq 2(w - (0.75-0.05i)w) = (0.5+0.1i)w, \end{aligned}$$

which yields

$$\begin{aligned} N(w,p) &\leq \binom{\beta_2(c)n_2}{0.2iwn_2}\binom{\beta_3(c)n_2}{(0.5-0.1i)wn_2}\binom{(1-\beta_2(c)-\beta_3(c))n_2}{(0.5-0.1i)wn_2} e^{O(\log^2 n)} \\ &= \binom{\beta_2 n_2}{0.2iwn_2}\binom{\beta_3 n_2}{(0.5-0.1i)wn_2}\binom{(1-\beta_2-\beta_3)n_2}{(0.5-0.1i)wn_2} e^{O(\log^2 n + w\delta n_2)} \end{aligned}$$

or

$$\begin{aligned} \frac{1}{n_2}\log|N(w,p)| \leq \beta_2 H\Big(\frac{0.2iw}{\beta_2}\Big) + \beta_3 H\Big(\frac{(0.5-0.1i)w}{\beta_3}\Big) \\ + (1-\beta_2-\beta_3) H\Big(\frac{(0.5-0.1i)w}{1-\beta_2-\beta_3}\Big) + O(\delta w). \end{aligned}$$

Thus we have

$$I(p,w) \leq H(w) - 4H\Big((0.7-0.05i)w\Big) + w\log 11 + \psi_i(w) + O(\delta w \log(1/w)),$$

where

$$\psi_i(w) = \beta_2 H\Big(\frac{0.2iw}{\beta_2}\Big) + \beta_3 H\Big(\frac{(0.5-0.1i)w}{\beta_3}\Big) + (1-\beta_2-\beta_3) H\Big(\frac{(0.5-0.1i)w}{1-\beta_2-\beta_3}\Big).$$

Let

$$\phi_i(w) := H(w) - 4H\Big((0.7-0.05i)w\Big) + w\log 11 + \psi_i(w).$$

Then

$$\begin{aligned} \phi_i''(w) &= \frac{0.8-0.2i}{w} + \frac{4\cdot(0.7-0.05i)^2}{1-(0.7-0.05i)w} - \frac{1}{1-w} - \frac{(0.2i)^2}{\beta_2 - 0.2iw} \\ &\quad - \frac{(0.5-0.1i)^2}{\beta_3-(0.5-0.1i)w} - \frac{(0.5-0.1i)^2}{1-\beta_2-\beta_3-(0.5-0.1i)w}. \end{aligned}$$

For instance, if $i=2$ then

$$\phi_2''(w) = \frac{0.4}{w} + \frac{1.44}{1-0.6w} - \frac{1}{1-w} - \frac{0.16}{\beta_2-0.4w} - \frac{0.09}{\beta_3-0.3w} - \frac{0.09}{1-\beta_2-\beta_3-0.3w}.$$

Since $\phi_i''(w)$ decreases as w increases, or $\phi_i'(w)$ is concave, $\phi_i(w)$ attains its maximum at the starting point or the unique $w < 0.43$ such that

$$\phi_i'(w) = 0 \ \text{ and } \ \phi_i''(w) < 0. \tag{3.14}$$

If such a w does not exist, it may attains its maximum at $w = 0.43$.

For Case (2), the mean value theorem together with $\phi_2'(0.36) > 0$ and $\phi_2'(0.37) < 0$ gives $\phi_2'(w_2) = 0$ for some w_2 with $0.36 \le w_2 \le 0.37$. Taylor theorem also yields

$$\phi_2(w_2) = \phi_2(0.36) + \phi_2'(w^*)(w_2 - 0.36)$$

for some w^* between 0.36 and w_2. Clearly, $w_2 - 0.36 \le 0.01$. Moreover, for all $w \ge 0.36$ including w_2,

$$\phi_2''(w) \le \phi_2''(0.36) < 0, \tag{3.15}$$

which implies that $\phi'(w)$ decreases as $w \ge 0.36$ increases, especially

$$\phi'(w^*) \le \phi'(0.36) \le 0.02$$

and hence

$$\phi_2(w_2) = \phi_2(0.36) + \phi_2'(w^*)(w_2 - 0.36) < -0.0075 + 0.0002 < -0.007.$$

Since (3.15) implies that w_2 satisfies (3.14), $\phi_2(w) = (0.4 + o(1))w \log w$ as $w \to 0$ gives

$$I(w, p) \le \phi_2(w) + O(\delta w \log(1/w)) \le -\Omega(\delta).$$

Similarly, the mean value theorem together with $\phi_3'(0.1) > 0$ and $\phi_3'(0.11) < 0$ gives $\phi_3'(w_3) = 0$ for some w_3 with $0.1 \le w_3 \le 0.11$. Taylor theorem yields

$$\phi_3(w_3) = \phi_3(0.1) + \phi_3'(w_3^*)(w_3 - 0.1)$$

for some w_3^* between 0.1 and w_3. Clearly, $w_3 - 0.1 \le 0.01$. Moreover, for all $w \ge 0.1$ including w_3,

$$\phi_3''(w) \le \phi_3''(0.1) < 0, \tag{3.16}$$

which implies that $\phi_3'(w)$ decreases as $w \ge 0.1$ increases, especially

$$\phi_3'(w_3^*) \le \phi_3'(0.1) \le 0.0002$$

and hence

$$\phi_3(w_3) = \phi_3(0.1) + \phi_3'(w_3^*)(w_3 - 0.1) < -0.0085 + 2 \cdot 10^{-6} < -0.008.$$

Since (3.16) implies that w_3 satisfies (3.14), $\phi_3(w) = (0.2 + o(1))w \log w$ as $w \to 0$ gives

$$I(w, p) \le \phi_3(w) + O(\delta w \log(1/w)) \le -\Omega(\delta).$$

Case (4) is different simply because $\phi_4(w) > 0$ for some w, e.g. $w = 0.09$. However, since $\phi_4(0.1), \phi_4'(0.1)$ and $\phi_4''(w)$ for any $w > 0$ are all negative, $\phi_4(w)$ decreases as $w \ge 0.1$ increases. In particular,

$$\phi_4(w) \le \phi_4(0.1) < -0.001 \quad \text{for all } w \ge 0.1.$$

We have to take a different approach when $\Omega(\delta) \le w \le 0.1$. For a sequence (w_i), let $N((w_i))$ be the number of W that contains $w_i n_2$ vertices of degree i and $P((w_i))$ be the upper bound given in (3.9) without the first two factors, that is,

$$N((w_i)) = \prod_{i\ge 2} \binom{\beta_i(c)n_2}{w_i n_2}, \quad P((w_i)) = \binom{n_2}{wn_2} p^{4pm} q^{4qm} 11^{wm},$$

where $w = \sum w_i$, $p = (4r)^{-1}\sum i w_i$, $q = 1-p$. For w with $\Omega(\delta) \le w \le 0.1$, take (w_i) with $\sum w_i = w$ that maximizes $N((w_i))P((w_i))$. Then by increasing $w_2 n_2$ by 1 and decreasing $w_j n_2$ by 1, if possible, we obtain

$$\Big(1+O((\delta n_2)^{-1})\Big)\Big(\frac{p}{q}\Big)^{2-j}\binom{\beta_2(c)n_2}{w_2 n_2+1}\binom{\beta_j(c)n_2}{w_j n_2-1} \le \binom{\beta_2(c)n_2}{w_2 n_2}\binom{\beta_j(c)n_2}{w_j n_2},$$

or

$$\frac{(\alpha_2(c)-w_2)n_2}{w_2 n_2+1}\cdot\frac{w_j n_2}{(\alpha_j(c)-w_j)n_2+1} \le \Big(1+O((\delta n_2)^{-1})\Big)\Big(\frac{p}{q}\Big)^{j-2},$$

which implies

$$w_j \le \Big(1+O((\delta n_2)^{-1})\Big)\frac{\beta_j}{\beta_2-0.1}\Big(\frac{p}{q}\Big)^{j-2} w.$$

Since $w_2 n_2 \le w n_2 \le 0.1 n_2 \le \beta_2 n_2/2$ we may always increase $w_2 n_2$ by 1. So, the only case we may not apply the above argument is when $w_j = 0$, in which case, however, the inequality remains valid.

Hence $p \le 0.55w \le 0.055$ implies that

$$w_3 \le \Big(1+O((\delta n_2)^{-1})\Big)\frac{\beta_3}{\beta_2-0.1}\Big(\frac{0.055}{1-0.055}\Big) w \le 0.14w$$

and

$$\begin{aligned}
\sum_{i\ge 4} w_i &\le \Big(1+O((\delta n_2)^{-1})\Big)\sum_{j\ge 4}\frac{\beta_j}{\beta_2-0.1}\Big(\frac{p}{q}\Big)^{j-2} w \\
&\le \Big(1+O((\delta n_2)^{-1})\Big)\sum_{j\ge 4}\frac{0.22}{\beta_2-0.1}\Big(\frac{p}{q}\Big)^{j-2} w \\
&= \frac{0.22\Big(1+(\delta n_2)^{-1})\Big)(p/q)^2 w}{(\beta_2-0.1)(1-p/q)} \\
&\le 0.01w,
\end{aligned}$$

and finally $w_2 \ge 0.85w$. Thus for any (w_i) with $\sum w_i = w \le 0.1$,

$$N((w_i))P((w_i)) \le \binom{\beta_2 n_2}{0.85w}\binom{\beta_3 n_2}{0.14w}\binom{(1-\beta_2-\beta_3)n_2}{0.01w}\binom{n_2}{wn_2} p^{4pm} q^{4qm} 11^{wm} e^{O(\delta w n_2)}$$

and

$$I(w,p) \le H(w) - 4H(0.5w) + w\log 11 + \psi_5(w) + O(\delta w \log(1/w)),$$

where

$$\psi_5(w) = \beta_2 H\Big(\frac{0.85w}{\beta_2}\Big) + \beta_3 H\Big(\frac{0.14w}{\beta_3}\Big) + (1-\beta_2-\beta_3)H\Big(\frac{0.01w}{1-\beta_2-\beta_3}\Big).$$

As usual, let

$$\phi_5(w) = H(w) - 4H(0.5w) + w\log 11 + \psi_5(w).$$

Then

$$\phi_5''(w) = \frac{1}{1-0.5w} - \frac{1}{1-w} - \frac{(0.85)^2}{\beta_2 - 0.85w} - \frac{(0.14)^2}{\beta_3 - 0.14w} - \frac{(0.01)^2}{1-\beta_2-\beta_3-0.01w} < 0.$$

Since

$$\begin{aligned} \phi_5'(w) &= -2\log 2 - 0.85\log\Big(\frac{0.85}{\beta_2}\Big) - 0.14\log\Big(\frac{0.14}{\beta_3}\Big) \\ &\qquad -0.01\log\Big(\frac{0.01}{(1-\beta_2-\beta_3)}\Big) + O(w) \\ &\sim -0.087\cdots + O(w) \end{aligned}$$

as $w \to 0$, $\phi''(w) < 0$ yields $\phi'(w) \le -0.087 < 0$ for all $0 < w \le 0.1$. Thus $\phi(w)$ decreases as w increases and

$$I(w,p) \le \phi(w) + O(\delta w \log(1/w)) = -\Omega(\delta).$$

4 Lower Bound

Let $c < c_2$ be fixed. Let $\epsilon > 0$ such that $c' = c + \epsilon < c_2$. We generate the sequence of pairs of random edges $(e_1, f_1), \dots, (e_{cn}, f_{cn})$ in two steps. First, we consider a collection of $c'n$ pairs of random edges $(e_1, f_1), \dots, (e_{c'n}, f_{c'n})$. We form a graph G' by choosing one edge from each of these pairs. We then delete ϵn random edges from the graph G' to form a graph G. These ϵn random edges in G correspond to ϵn pairs from $(e_1, f_1), \dots, (e_{c'n}, f_{c'n})$ chosen uniformly at random. Thus, the graph G formed in this way contains one edge from each of a set of cn pairs of edges chosen uniformly and independently at random.

It follows from Theorem 5 that we may assume that the edge density of any sub-hypergraph of the hypergraph $\mathcal{H}(c'n)$ has density less than one. It then follows from Hall's Theorem that $\mathcal{H}(c'n)$, regarded as simply a collection of sets, has a system of distinct representatives; that is, there exist *distinct* vertices $v_1, \dots, v_{c'n}$ such that $v_i \in e_i \cup f_i$ for $i = 1, \dots, c'n$. Indeed, consider the bipartite graph B with parts $\mathcal{H}$ and

$[n]$ and an edge between $A \in \mathcal{H}$ and $v \in [n]$ if and only if $v \in A$. This bipartite graph has an $\mathcal{H}$-perfect matching (which gives our system of distinct representatives) if and only if the size of the neighborhood in B of $\mathcal{F}$ is larger than $\mathcal{F}$ for every $\mathcal{F} \subseteq \mathcal{H}$. This condition is equivalent to the condition that every sub-hypergraph of $\mathcal{H}$ has density at most 1.

We use this system of distinct representatives to determine which edge from the pair (e_i, f_i) is used in our graph: For $i = 1, \ldots, c'n$ we use the edge that contains the vertex v_i. Let G' be the graph formed by these edges. Note that the number of edges in any connected component of G' is at most the number of vertices in that component (since each vertex 'accounts' for at most one of its incident edges). Thus, the connected components of G' are isolated vertices, trees or unicyclic graphs.

Before proceeding with the deletion step, we note a structural property of G'. We say that the *depth* of a rooted tree is the maximum distance between any vertex in the tree and the root.

Lemma 11. *Let H be the random graph $G_{n,m}$ where $m = \lambda n$ for some constant $\lambda > 0$. Let K be a constant such that $\log K > 2 + (1 + 1/K)\log(3\lambda)$. With high probability all (not necessarily induced) rooted trees T in H of depth at most $\frac{\log n}{K}$ have at most $n^{\frac{\log(2K)}{K}}$ vertices.*

Proof. Since the property in question is monotone and $G_{n,m} \subseteq G_{n,3\lambda/n}$ **whp** it suffices to show that the Lemma holds in the independent model $H' = G_{n,2\lambda/n}$.

Let $\mathcal{B}_{\alpha,\beta}$ be the event that there are disjoint sets $A, B \in [n]$ such that $|A| = \alpha$, $|B| = \beta$ the induced graph $H[A]$ contains a path and for each $b \in B$ there exists $a \in A$ such that $\{a, b\}$ is an edge of H. Applying the union bound we have

$$\begin{aligned}
\Pr(\mathcal{B}_{\alpha,\beta}) &\leq n^\alpha \binom{n}{\beta} \alpha^\beta \left(\frac{3\lambda}{n}\right)^{\alpha+\beta-1} \\
&\leq ne^\beta \left(\frac{\alpha}{\beta}\right)^\beta (3\lambda)^{\alpha+\beta-1}
\end{aligned}$$

Setting $\alpha = \frac{\log n}{K}$ and $\beta = \log n$ we have

$$Pr(\mathcal{B}_{\alpha,\beta}) \leq \exp\left\{\log n\big(2 - \log K + (1 + 1/K)\log(3\lambda)\big)\right\}.$$

By the condition we imposed on K, the probability of $\mathcal{B}_{(\log n)/K, \log n}$ is $o(1)$.

Suppose $\mathcal{B}_{(\log n)/K, \log n}$ does not hold. Let $T = T_0$ be a rooted tree of depth $\alpha = \frac{\log n}{K}$ in H. Let r_0 be the root of T_0. Let T_1 be the largest tree in the forest obtained by deleting r_0 from T_0. We have

$$|T_1| \geq \frac{|T_0| - 1}{d_{T_0}(r_0)}.$$

Let r_1 be the vertex in T_1 that was connected to r_0 in T. We now iterate this construction through the tree. For $i = 1, \ldots, \alpha - 2$ let T_{i+1} be the largest tree in the forest obtained by removing r_i from T_i. We have

$$|T_{i+1}| \geq \frac{|T_i| - 1}{d_{T_i}(r_i)} \geq \frac{|T_0| - \sum_{j=0}^{i-1} \prod_{s=0}^{j} d_{T_s}(r_s) - 1}{\prod_{s=0}^{i} d_{T_s}(r_s)}$$

Since $T_{\alpha-1}$ is simply a star, we have

$$1 + d_{T_{\alpha-1}}(r_{\alpha-1}) = |T_{\alpha-1}| \geq \frac{|T_0| - \sum_{j=0}^{\alpha-3} \prod_{s=0}^{j} d_{T_s}(r_s) - 1}{\prod_{s=0}^{\alpha-2} d_{T_s}(r_s)}.$$

Therefore

$$1 + \sum_{s=0}^{\alpha-1} \prod_{i=1}^{s} d_{T_i}(r_i) \geq |T|,$$

and

$$\prod_{i=0}^{\alpha-1} d_{T_i}(r_i) \geq \frac{|T|}{\alpha}.$$

By Jensen's inequality (taking the logarithm) we have,

$$\sum_{i=0}^{\alpha-1} d_{T_i}(r_i) \geq \alpha \left[\frac{|T|}{\alpha}\right]^{\frac{1}{\alpha}}.$$

On the other hand, since the event $\mathcal{B}_{\alpha,\beta}$ does not hold, we have

$$\sum_{i=0}^{\alpha-1} (d_{T_i}(r_i) - 1) < \beta = \log n.$$

It follows that

$$|T| < \alpha \left[\frac{\log n}{\alpha} + 1\right]^{\alpha} < (2K)^{\frac{\log n}{K}}$$

for n sufficiently large. □

Now, we choose a set of ϵn edges from G' at random and delete them to form the graph G. For the purposes of the following argument, we assume each component in G' is a tree by ignoring one edge per cycle of in G'. Under this assumption, we prove that **whp** every component in G has at most $n^{1-\epsilon}$ vertices. It follows (since we only ignored one edge per component) that twice this number is a true upper bound on the component size of G.

For each component of G' we choose a root. Let U be the set of vertices who distance from their respective root is greater than $\log(n)/K$. For a vertex $u \in U$ let $P(u)$ be the first $\log(n)/K$ edges in path from u to the root of the component containing u. We say u *survives* if no edge in $P(u)$ is deleted (i.e. if $P(u) \subseteq E(G)$). Let S be the set of vertices in U that survive. We have

$$E[|S|] < n\left(1 - \frac{\epsilon}{c'}\right)^{\log(n)/K} \leq n^{1-\epsilon/K}.$$

It then follows from Markov's inequality that **whp**

$$|S| \leq n^{1-\epsilon/K+\delta} \tag{4.1}$$

for any $\delta > 0$.

Now, consider a component C of G. Let C' be the component of G' that contains C. Let v be the vertex of C closest to the root of C'. Note that $C \setminus S$ is a tree of depth at most $\log(n)/K$ rooted at v. Applying Lemma 11 and (4.1), we have

$$|C| \leq |C \setminus S| + |S| \leq n^{\log(2K)/K} + n^{1-\epsilon/K+\delta}.$$

with high probability.

Acknowledgment. We thank the anonymous referee for many helpful comments and suggestions.

References

[1] A. Békéssy, P. Békéssy and J. Komlós, Asymptotic enumeration of regular matrices, *Studia Sci. Math. Hungar.*, **7** (1972), 343–353.

[2] E. Bender and R. Canfield, The asymptotic number of labeled graphs with given degree sequences, *J. Combinatorial Theory Ser. A* **24** (1978), 296–307.

[3] T. Bohman and A. Frieze, Avoiding a giant component, *Random Structures and Algorithms*, 19 (2001), 75-85.

[4] T. Bohman, A Frieze and N. Wormald, Avoiding a giant component in half the edge set of a random graph, *Random Structures and Algorithms*, to appear.

[5] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.* **1** (1980), 311–316.

[6] B. Bollobás, *Random Graphs, Second Edition*, Academic Press(2001).

[7] C. Cooper, The sizes of the cores of random hypergraphs with a given degree sequence, *Random Structures and Algorithms*, to appear.

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. math. Inst. Hungar. Acad. Sci.*, 5 (1960), 17-61.

[9] J. H. Kim, Poisson cloning model for random graphs. In preparation.

[10] J. Spencer and N. Wormald, Birth control for giants. manuscript.

[11] N.C. Wormald, Models of random regular graphs, In *Surveys in Combinatorics, 1999*, J.D. Lamb and D.A. Preece, eds, pp. 239–298.