ssdnm
środowiskowe
studia doktoranckie
z nauk matematycznych


Bartosz Zaleski

Uniwersytet A. Mickiewicza w Poznaniu


# On some structural properties of random bipartite graphs

Praca semestralna nr 1

(semestr letni 2011/12)

Opiekun pracy: Tomasz Łuczak

# ON SOME STRUCTURAL PROPERTIES OF RANDOM BIPARTITE GRAPHS

BARTOSZ ZALESKI

Abstract. Since 1959, when Erdös and Rényi published their paper about random $G(n, m)$ graph model, several other models of random graphs have been introduced. For example, we can consider bipartite version of $G(n, m)$ and ask for its structural properties. In this survey we deal with various models of random bipartite graphs and some of their structural properties, namely matchings and $k-$cores. The main goal of paper is to present some results and state open problems to solve in the future.

## 1. Formalism

Let us focus on following models of random graphs. Let $B(n, n, m)$ be a bipartite graph with symmetrically divided set of $2n$ vertices and $m$ edges which are randomly chosen (with repetitions) from the set of all edges of complete bipartite graph $K_{n,n}$. Let $B_k(n, m)$ be a bipartite left $k-$regular graph with $V(B_k(n, m)) = L \cup R$, where $|L| = n$, $|R| = m > n$ and each vertex $v \in L$ chooses random $k$ neighbours from the set $R$. Finally, let $B_{k-out}(n, m)$ be a bipartite graph that is obtained by omitting the direction from directed random bipartite graph $B$, with $V(B) = L \cup R$, where $|L| = n$, $|R| = m$, in which each vertex chooses $k$ random neighbours, resulting in out-degree equal to $k$.

By a matching of graph $G$ we mean a subgraph $M \subset G$ in which each vertex has degree 1. If $M$ is a spanning subgraph we call it a perfect matching. For a bipartite graph $G = (V_1 \cup V_2, E)$ matching $M$ is $V_i-$perfect (for $i = 1, 2$) if it contains all vertices from $V_i$. The $k-$core of a graph $G$ is its largest subgraph $H \subset G$ with minimal degree at least $k$ (if there is no such subgraph then we say that $k-$core is empty). The notion of a $k-$core was introduced by Bollobás in [1] where he was looking for large $k-$connected subgraphs of a random graph. There is a very simple algorithm that finds a $k-$core of a graph. One can obtain it by following procedure. Let $G_0 = G$ and $v_0$ be any vertex of degree smaller than $k$ in $G_0$. Having defined $G_i$, for some $0 \leq i \leq n - 1$, we define $v_i$ to be any vertex of degree smaller than $k$ in $G_i$. If it does not exist $G_i$ is the $k-$core of $G$ and we stop the process. Otherwise we let $G_{i+i}$ be $G_i - v_i$. If $G_i$ is empty for any $1 \leq i \leq n - 1$ then we also stop the process setting $k-$core to be empty.

## 2. Matchings

The question of appearance of a perfect matching in random graphs is one of the best studied in this field. In [7] and [8] Erdös and Rényi dealt with this problem in $G(n,m)$ and $B(n,n,m)$ models. They showed, for example, that in non-bipartite model with the threshold value of number of edges is $m = n\ln(n)$, ie. they showed that if $m = n(\ln(n) + c_n)$ then the probability $p_{n,m}$ that the random $G(n,m)$ graph has perfect matching tends to

- 0 if $c_n \to -\infty$,
- $e^{-2e^{-c}}$ if $c_n \to c$, where $c$ is some positive constant,
- 1 if $c_n \to \infty$.

Here we present their result for random bipartite graph (although they solved this problem using the notion of random binary matrices). To do so we recall a well known fact of equivalence of $G(n,p)$ and $G(n,m)$ models for monotone properties.

**Theorem 2.1.** *Let $c(n)$ be a function and $c > 0$ some constant. Let $p = \frac{\ln(n)+c(n)}{n}$. Then the probability $p_{n,n,p}$ that random bipartite graph $B(n,n,p)$ has a perfect matching satisfies*

$$\lim_{n\to\infty} p_{n,n,p} = \begin{cases} 0 & \text{when } c(n) \to -\infty, \\ e^{-2e^{-c}} & \text{when } c(n) \to c, \\ 1 & \text{when } c(n) \to \infty. \end{cases}$$

*Proof.* To show this we shall use Hall's Theorem which states that a bipartite graph $G = (L \cup R, E)$ contains a perfect matching if and only if

$$\forall_{S\subset L} |S| \le |N(S)|.$$

We can change this condition into

$$\forall_{S\subset L, |S|\le \frac{n}{2}} |S| \le |N(S)|,$$

$$\forall_{T\subset R, |T|\le \frac{n}{2}} |T| \le |N(T)|.$$

First observe that $p_{n,n,p}$ is bounded from below by the probability $p_0$ that $G(n,p)$ contains an isolated vertex. Similarly, it is bounded from above by the sum of $p_0$ and 2 times $p_k$, which is the probability of an event that there exist $k \ge 2$ and a subsets $S \subset L$, $T \subset R$ st. $|S| = k$, $|T| = k-1$, $N(S) \subset T$ and with at least $2k-2$ edges between $S$ and $T$. It is possible to put this conditions on size of $T$ and number of edges because of the following argument. Choose $T$ and $S$ so that $|T| + |S|$ is minimal. Until it is possible repeat following two steps

(1) if $|S| > |T| + 1$ remove $|S| - |T| - 1$ vertices from $S$.

(2) if there exists vertex $v \in T$ with less than 2 neighbours in $S$ remove it from $S$ and its neighbour (if there is one) from $T$.

Observe that if there is no isolated vertex in $L$ and graph does not contain a perfect matching this procedure will lead to sets $S$ and $T$ satisfying our conditions. Let $X$ be the number of sets $S$ and $T$ with above-mentioned properties. We have

$$\begin{aligned}\mathbb{E}(X) &\leq \sum_{k=2}^{n/2} \binom{n}{k}\binom{n}{k-1}\binom{k(k-1)}{2k-2} p^{k-2}(1-p)^{k(n-k+1)} \\ &\leq \sum_{k=2}^{n/2} \Big(\frac{en}{k}\Big)^k \Big(\frac{en}{k-1}\Big)^{k-1} \Big(\frac{ek(\ln(n)+c(n))}{2n}\Big)^{2k-2} e^{-pkn(1-k/n)} \\ &\leq 4\sum_{k=2}^{n/2} n\Big(\frac{e^{O(1)}(\ln(n))^2 n^{k/n}}{n}\Big)^k = 4\sum_{k=2}^{n/2} u_k,\end{aligned}$$

where $u_k = n(\frac{e^{O(1)}(\ln(n))^2 n^{k/n}}{n})^k$. Let us consider two cases

(1) $2 \leq k \leq n^{3/4}$, then

$$u_k = ne^{O(k)} n^{-0.99k},$$

which means that, for large enough $n$,

$$4\sum_{k=2}^{n^{3/4}} = O(n^{-0.9}) = o(1),$$

(2) $n^{3/4} < k \leq n/2$, then we can bound $u_k$ by

$$u_k \leq n^{1-k/3},$$

and

$$4\sum_{n^{3/4}}^{n/2} u_k \leq O(n^{1-\frac{n^{3/4}}{3}}).$$

Thus, by the Markov inequality, probability that there is no perfect matching is equal to $p_0 + o(1)$. Let $X_0$ be the number of isolated vertices in $B(n, n, m)$. First we will deal with the case when $c(n) \to \infty$. We have

$$\mathbb{E}(X_0) = 2n(1-p)^n = 2e^{-c(n)} e^{\frac{\log(n)^2}{n}} \approx 2n(1-p)^n = 2e^{-c(n)}.$$

So, by the Markov inequality, $\mathbb{P}(X_0 > 0) \to 0$. Next, let $c(n) \to -\infty$. Here we have to compute also the second moment of $X_0$. So

$$\begin{aligned}\mathbb{E}(X_0^2) &= \sum_{i,j \in V(B(n,n,p))} \mathbb{P}([d(i)=0]|[d(j)=0])\mathbb{P}([d(j)=0]) \\ &= \mathbb{E}(X_0)(1+\sum_{j:j>1} \mathbb{P}([d(j)=0]|[d(1)=0])) \\ &\approx 2e^{-c(n)}(1+2e^{-c(n)})\end{aligned}$$

Using the Strong Second Moment Method we get that

$$p_0 \geq \frac{\mathbb{E}(X_0^2)}{\mathbb{E}^2(X_0)} = \frac{2e^{-c(n)}}{1+2e^{-c(n)}} \to 1.$$

Finally we deal with the case when $c(n) \to c$ for some constant $c$.We will use the Inclusion−exclusion principle. Let (for $i = 1, \ldots, 2n$) $S_i$ be the probability that at least $i$ vertices are isolated. We have

$$S_i = \sum_{k=0}^{n} \binom{n}{k} \binom{n}{i-k} (1-p)^{in-k(i-k)}.$$

One can check that $\lim_{n\to\infty} S_i = \frac{2^i e^{ic}}{i!}$. So

$$\mathbb{P}([X_0 = 0]) = 1 + \sum_{i=1}^{2n} S_i \to e^{-2^{-c}}.$$

□

There is an interesting connection between non-bipartite graph $G(n, m)$ and $B_2(m, n)$ models. The existence of $L-$perfect matching in the latter one can be translated into a question of existence of so called simple connected components ie. components with at most one cycle. We will discuss this in later part of this survey.

Let us focus now on random $B_k(n, m)$ graphs. Deciding whether there is a $L-$perfect matching in this model is more complicated than in previous one. One can easily check that if $n = m$ it is hardly possible to obtain a perfect matching for small $k$. More precisely, $k$ must be of order greater than logarithmic. We shall now present a group of results concerning the connection between $k$ and size of set $R$ that is necessary to obtain high probability of occurrence of one-sided perfect matching. This problem is not purely academic as it arises in analysis of one of hashing algorithms - Cuckoo Hashing. In this setting vertices from $L$ correspond to elements placed in the hash table and those from $R$ to places in hash table and if there is an edge between vertices $v \in L$ and $w \in R$ then position $w$ is one of $k$ possible places for element $v$.

So the perfect matching can be interpreted as the assignment of elements to places without conflicts.

The case of $k = 2$ is very well studied and one can translate it to previously mentioned problem of simple components in random $G(n, m)$ graph. To see this consider following procedure. Given random $B_2(m, n)$ graph consider a graph on the set of vertices equal to $R$ where $v$ and $w$ are adjacent if in $B_2(m, n)$ those vertices had common neighbour (were chosen by the same vertex from $L$). In order to show this equivalence we will use the Hall's Theorem. Observe that if there is a set $S \subset L$ such that $|N(S)| < |S|$ then the induced subgraph of $G(n, m)$, whose set of vertices is $N(S)$, has $|S|$ edges which means that there must be more than one cycle. For above reason let us present following theorem settling the problem of simple components in $G(n, m)$. Like previously we shall shift to $G(n, p)$ model.

**Theorem 2.2.** *Let $p \leq \frac{1}{n} - \frac{\lambda}{n^{4/3}}$ for any growing to infinity function $\lambda = \lambda(n)$. Then, with high probability, $G(n, p)$ consists only of simple components.*

Before presenting the proof let us discuss one inconsistency. In our setting both $G(n, p)$ and $G(n, m)$ models does not allow multiedges while transformation from $B_2(m, n)$ can produce such structures. But this is very unlikely for it requires at least two vertices both choosing the same neighbours.

*Proof.* Suppose that $G_1$ is the component of $G(n, p)$ with at least two cycles $C_1$ and $C_2$. Let $G_1'$ be the "minimal" connected subgraph comprising them. Those cycles are either joint by a path (possibly of length 0) or they share at least one edge. Let $X$ stand for the number of such pairs. Let $P$ be the longest path in $G_1'$. There must be an edge connecting each end of $P$ with some vertices inside $P$. Then there are at most $k!k^2$ such subgraphs on $k$ vertices. Thus

$$\begin{aligned}
\mathbb{E}(X) &\leq \sum_{k=4}^{n} \binom{n}{k} k!k^2 p^{k+1} \\
&\leq \sum_{k=4}^{n} \frac{n^k}{k!} k!k^2 \frac{1}{n^{k+1}} (1 - \frac{\lambda}{n^{1/3}}) \\
&\leq \int_0^\infty \frac{x^2}{n} e^{-\frac{\lambda x}{n^{1/3}}} dx = \frac{2}{\lambda^3} = o(1).
\end{aligned}$$

So, by the Markov inequality, the probability of existence of at least one such pair tends to 0. ◻

In [3] Drmota and Kutzelnigg showed similar result using the generating functions formalism. More precisely, they proved that for any $\varepsilon \in (0, 1)$ the random $G(2n, m)$ graph (here we allow multiedges) consists only of simple components for $m = (1 - \varepsilon n)$ with high probability. Moreover, they showed that for $n = m$ the probability of such event drops to $\sqrt{\frac{2}{3}} + o(1)$.

After observing the requirements for $k = 2$ a natural question arises. How large must be the set $R$ to find $L-$perfect matching for some fixed $k \geq 3$? Or the other way round, for given size of $R$ what value of $k$ is enough for such structure. Answer for this can be found in [9] Fotakis et al. where they proved following theorem

**Theorem 2.3.** *Let $\varepsilon \in (0,1)$ be fixed real number and let $m = (1+\varepsilon)n$. Then for any $d \geq 2(1+\varepsilon)\ln(\frac{e}{\varepsilon})$ the bipartite graph $B_d(n,m)$ contains $L-$perfect matching with high probability.*

*Proof.* Let $X$ be the number of such sets $S \subset L$ that $|N(S)| < |S|$ and let (for $k = 2, 3, \dots, n$) $X_k$ be the number of such sets of size $k$. For $k = 2$ it is easy to see that

$$\mathbb{E}(X_2) = O(n^{3-2d}).$$

We shall bound $\mathbb{E}(X_k)$ (for $k = 3, 4, \dots n$) by the expected number of pairs $A_k \subset L$ and $B_k \subset R$ such that $|A_k| = |B_k| = k$ and $N(A_k) \subset B_k$. First let us bound it for $k < \frac{n}{e^{d/(d-2)}}$. One can see that

$$\begin{aligned} \mathbb{E}(X_k) &\leq \binom{n}{k}\binom{n(1+\varepsilon)}{k}\Big(\frac{k}{(1+\varepsilon)n}\Big)^{dk} \\ &\leq \frac{e^{2k}k^{k(d-2)}}{n^{k(d-2)}} \end{aligned}$$

For such $k$ the last line is non-increasing function of $k$. So it can be bounded by $O(n^{3-2d})$. For $\frac{n}{e^{d/(d-2)}} \leq k \leq n$ we shall use following bound (which proof can be found in [9])

$$\binom{n}{k} \leq \Big(\frac{n}{n-k}\Big)^{n-k}\Big(\frac{n}{k}\Big)^{k}.$$

Let $\mu = \frac{k}{n}$, then

$$\mathbb{E}(X_k) \leq \bigg(\Big(\frac{1}{1-\mu}\Big)^{1-\mu}\Big(\frac{1}{\mu}\Big)^{\mu}\Big(\frac{1+\varepsilon}{1+\varepsilon-\mu}\Big)^{1+\varepsilon-\mu}\Big(\frac{1+\varepsilon}{\mu}\Big)^{\mu}\Big(\frac{\mu}{1+\varepsilon}\Big)^{d\mu}\bigg)^{n}$$

By checking for which values of $d$ the expression inside the brackets is strictly smaller than 1 we get that

$$d > 1 + \frac{\mu\ln(\frac{1}{\mu}) + (1-\mu)\ln(\frac{1}{1-\mu}) + (1+\varepsilon-\mu)\ln(\frac{1+\varepsilon}{1+\varepsilon-\mu})}{\mu\ln(\frac{1+\varepsilon}{\mu})}.$$

To obtain more suitable bound we shall use following relations

- $\ln(\frac{1}{\mu}) < \ln(\frac{1+\varepsilon}{\mu})$,

- for $\varepsilon \in (0,1)$ and $\mu \in (0,1]$ the expression $\frac{(1+\varepsilon-\mu)\ln(\frac{1+\varepsilon}{1+\varepsilon-\mu})}{\mu\ln(\frac{1+\varepsilon}{\mu})}$, as a non-decreasing function of $\mu$, is maximised by $\frac{\varepsilon\ln(\frac{1+\varepsilon}{\varepsilon})}{\ln(1+\varepsilon)}$,
- for $\varepsilon \in (0,1)$ and $\mu \in (0,1]$, $\frac{(1-\mu)\ln(\frac{1}{1-\mu})}{\mu\ln(\frac{1+\varepsilon}{\mu})} \leq \frac{\varepsilon\ln(\frac{1+\varepsilon}{\varepsilon})}{\ln(1+\varepsilon)}$.

Putting this together we get

$$d > 2 + \frac{2\varepsilon\ln(\frac{1+\varepsilon}{\varepsilon})}{\ln(1+\varepsilon)},$$

which is satisfied as long as $d > 2(1+\varepsilon)\ln(\frac{e}{\varepsilon})$. So, by the Markov inequality, we get that the probability that there is no $L-$perfect matching is at most $O(n^{4-2d})$ provided that $d > 2(1+\varepsilon)\ln(\frac{e}{\varepsilon})$. □

The interesting thing about this result is that there is not much space for improvement as the bound on $d$ is rather tight. It can be showed that if $d < (1+\varepsilon)\ln(\frac{1}{\varepsilon})$ there will be no $L-$perfect matching with high probability as there will be more than $\varepsilon n$ isolated vertices in $R$.

To finish with matchings let us focus on the last model - random $k-$out graphs. This model is somehow similar to the previous one. In can be viewed as a sum of two such graphs. A natural question arises. What is the value of $k$ for which, for $n = m$, there is a high probability of perfect matching. The bipartite version was analysed by Walkup in [10]. He was able to show that it is enough to take $k = 2$. Moreover, he established an upper bound of $3\sqrt{n}\left(\frac{2d}{e}\right)^n$ for a probability of existence of perfect matching in $k-$out random bipartite graph for any $k$. Later the same result was obtained by Frieze [2] for non-bipartite model whose proof based on careful examination of its structure using Tutte's Theorem. Here we present the complement of Walkup's result ie. we show that it is not enough to let $k = 1$.

**Lemma 2.4.** *With high probability random $B_{1-out}(n,n)$ graph contains no perfect matching.*

*Proof.* Let $\Lambda$ be a subgraph of $B_{1-out}$ that consists of three vertices $v_1, v_2$ and $w$ where $v_1, v_2$ belong to one set of bipartition of vertices and $w$ to the other. Moreover, suppose that in the process of choosing neighbours both $v_1$ and $v_2$ chose $w$ and they were not chosen by any vertex (besides possibly w). Trivially, if such subgraph exists there cannot be a perfect matching. Let $X_\Lambda$ be the number of such subgraphs. Obviously

$$\mathbb{E}(X_\Lambda) = 2\binom{n}{2}n\left(\frac{1}{n}\right)^2\left(1-\frac{2}{n}\right)^{n-1} \geq (n-1)e^{-2-1/(n-2)}.$$

Moreover, we have

$$\begin{aligned}\mathbb{E}(X_\Lambda^2) &= \sum_{\Lambda_1,\Lambda_2} \mathbb{P}([\Lambda_1 \subset B_{1-out}, \Lambda_2 \subset B_{1-out}]) \\ &= \mathbb{E}(X_\Lambda)\Big(1 + \sum_{\Lambda_1\neq\Lambda^*} \mathbb{P}([\Lambda_1 \subset B_{1-out}]|[\Lambda^* \subset B_{1-out}])\Big),\end{aligned}$$

where $\Lambda^*$ is some fixed copy of $\Lambda$ and the first sum runs over all possible copies of $\Lambda_1$ and $\Lambda_2$ of graph $\Lambda$. Let $S = 1 + \sum_{\Lambda_1\neq\Lambda^*} \mathbb{P}([\Lambda_1 \subset B_{1-out}]|[\Lambda^* \subset B_{1-out}])$. From the Strong Second Moment Method we have that

$$\mathbb{P}([X_\Lambda > 0]) \geq \frac{(\mathbb{E}(X_\Lambda))^2}{\mathbb{E}(X_\Lambda^2)} = \frac{(\mathbb{E}(X_\Lambda))^2}{\mathbb{E}(X_\Lambda)(1+S)} = \frac{1}{\frac{1}{\mathbb{E}(X_\Lambda)} + \frac{S}{\mathbb{E}(X_\Lambda)}}.$$

It is easy to observe that the elements of $S$ that involve copies of $\Lambda$ that are not vertex disjoint with $\Lambda^*$ are of order smaller than others. Thus we need only to take care of those that are disjoint.

$$\begin{aligned}S &\leq \binom{n-2}{2}(n-1)\Big(\frac{1}{n}\Big)^2\Big(1-\frac{2}{n-2}\Big)^{n-1} + \\ &\quad + \binom{n-1}{2}(n-2)\Big(\frac{1}{n-2}\Big)^2\Big(1-\frac{2}{n}\Big)^{n-3} + o(n) \\ &\leq \frac{1}{2}(n-1)e^{-2}\Big(e^{-\frac{1}{n-2}} + e^{\frac{6}{n}}\Big) + o(n)\end{aligned}$$

So

$$\mathbb{P}(X_\Lambda > 0) \to 1.$$

□

It could seem that those results settle down the question of perfect matchings in random $k-$out graphs, but one could ask where the true threshold for this property is. Obviously it is between 1 and 2. To answer this one must define what it means to be a random $(1+x)-$out graph.
One attempt was done by Karoński and Pittel in [5]. They proved that, with high probability, there is a perfect matching in so called $B_{(1+e^{-1})-out}$ random graph ie. a graph obtained from 1−out random graph by letting each vertex of degree 1 choose another random neighbour, thus making the minimal degree at least 2. Obviously such graph would be a subgraph of 2−out random graph. Another approach could be to fix some set of vertices which can choose only one neighbour and let others choose two. This also still remains open.

## 3. K-CORES

In this part of survey we will focus on several questions connected with $k-$cores. Especially we will ask for parameters of random graphs for which

non-empty $k-$cores exist or how they appear and grow in size; whether it happens in a "continuous" fashion or there is a threshold value where an "explosion" occurs. Apparently those problems are not so well studied as perfect matchings. Although there are some results concerning mainly the $G(n,p)$ ($G(n,m)$) or model with fixed sequence of degrees [6]. Models we were dealing with are yet to be studied in this context. Of course one can think of some simple bounds for the degree of a core. For example, if we consider $B_k(n,m)$ graph then the degree is bounded from above by $k$ and from below by the minimum degree obtained in set $R$. Issue of more precise estimations and of size of the core is still wide open. Similarly, in $B_{k-out}(n,m)$ the degree is at least $k$ but nothing more is certain. For those reasons we shall now discuss the $G(n,p)$. Pittel et al. [11] showed that $k-$cores appear suddenly and when they do, they are of linear size. More precisely, by carefully observing some branching process they proved that, for $k \geq 3$, the $G(n,m)$ graph contains giant $k-$core when $m$ reaches $c_k \frac{n}{2}$. Here $c_k = \min_{\lambda>0}(\frac{\lambda}{\pi_k(\lambda)})$ and $\pi_k(\lambda)$ is a probability that a Poisson distributed random variable is at least $k-1$ and by a giant $k-$core we mean one of a linear size.
Due to length and complication we shall present weaker results by Łuczak [4]. Let $v(k;n,p)$ be a random variable equal to 0 if the $k-$core of $G(n,p)$ is empty and to the size of it otherwise. To measure the density of $G(n,p)$ we shall use the average degree of a vertex $c_p = (n-1)p$. First we will show that there is a $0-1$ law concerning the parameter $v(k;n,p)$, but to do so we need following structural lemma.

**Lemma 3.1.** *Let $a = a(n)$ and $c = c(n)$ be functions of $n$ such that $a(n) \geq 1.1n$ and $c(n) = o(an^{1-1/a})$. Then, with high probability, for $p = \frac{c(n)}{n}$ the $G(n,p)$ graph contains no subgraph on $s \leq 0.35\left(\frac{2a}{c}\right)^{a/(a-1)} e^{-2/(a-1)}$ vertices and more than $as$ edges.*

*Moreover, for $c \leq 3$, if there is a subgraph on $s$ vertices and at least $1.5s$ edges then $s \geq 0.006n$.*

**Theorem 3.2.** *If $k \geq 3$ then with high probability either $v(k;n,p) = 0$ or $v(k;n,p) \geq 0.0002n$.*

*Proof.* We shall consider three cases. First let $c(n) \leq 3k$. The number of edges in a $k-$core is at least $k/2$ times greater then the number of its vertices. Then, by lemma the nonempty $k-$core will, with high probability, comprise of $0.35 \cdot 3^{-3} \cdot e^{-4}n > 0.0002n$ vertices. Furthermore, for $c(n) = 3k$, $G(n,p)$ contains $0.9\binom{n}{2}p > 1.1nk > n(k-1)$ edges. Observe that if there are at least $n(k-1)$ edges in a graph then the process of deleting vertices of degree smaller than $k$ cannot decrease the average degree. Thus in this case the $k-$core cannot be empty and so it must consists of at least $0.0002n$ vertices. To settle the remaining case we utilise the fact that containing a subgraph on at least $0.0002n$ vertices and minimal degree greater than $k-1$ is increasing so the lower bound on the size of a core for $c(n) = 3k$ is also valid for $c(n) > 3k$. $\square$

Following theorem shows that if we ask for a core of degree not very much large in relation to the average degree then such core exists and is quite large.

**Theorem 3.3.** *Let $\varepsilon$ be a real number from $(0,1)$. Then there exist constant $d$ such that for $c = c(n) > d$ and $k \leq c - c^{0.5+\varepsilon}$, with high probability, there exist a $k-$core of size $v(k;n,p) \geq n - ne^{-c^{\varepsilon}}$.*

*Proof.* Let us consider an increasing sequence $\{U_i\}_{i=0}^{s}$ of subsets of vertices of $G(n,p)$ in a following way

(1) $U_0$ comprise of vertices of degree at most $k$.
(2) Having constructed $U_0, U_1, \ldots, U_l$ we define $U'_{l+1}$ to be the set of vertices outside the $U_l$ with at least two neighbours in it. If $U'_{l+1}$ is nonempty then $U_{l+1} = U_l \cup v$, where $v$ is any vertex from $U'_{l+1}$. Otherwise we finish the process and set $s = l$.

If $U_s$ is not the whole set of vertices of $G(n,p)$ then the subgraph induced by the set of all vertices outside $U_s$ has minimal degree at least $k$, so it is a subset of a $k-$core. Thus we have to show that $|U_s| < ne^{-c^{\varepsilon}}$. Obviously, if $c(n) \geq n^{1/2}$ then $U_0 = \emptyset$. Thus, let us assume that $c(n) < n^{1/2}$ and let $X_k$ denote the number of vertices of degree at most $k$. We have

$$\begin{aligned}
\mathbb{E}(X_k) &= n\sum_{i=0}^{k}\binom{n-1}{i}p^i(1-p)^{n-i-1} \\
&\leq 2n\left(\frac{ec}{k}\right)^k e^{-c} \\
&\leq 2ne^{-c^{0.5+\varepsilon}}\left(1+\frac{c^{0.5+\varepsilon}}{c-c^{0.5+\varepsilon}}\right) \\
&\leq 0.01ne^{-c^{\varepsilon}}
\end{aligned}$$

In the similar fashion one can obtain that $\mathbb{V}ar(X_k) = o((\mathbb{E}(X_k))^2)$, so by the Chebyshev inequality

$$|U_0| < 0.1ne^{-c^{\varepsilon}}.$$

It follows the definition of the sequence $\{U\}_{i=0}^{s}$ that for $i = 1, 2, \ldots, s$ the subgraph induced by $U_i$ has at least $2(|U_i| - |U_0|)$ edges. So, if for some $i = 1, 2, \ldots, s$ $|U_i| = \lfloor ne^{-c^{\varepsilon}} \rfloor$ then

$$2(|U_i| - |U_0|) \geq 2(|U_i| - 0.2|U_i|) > 1.5|U_i|.$$

This is a contradiction with a lemma. Thus, with high probability, $|U_s| < ne^{-c^{\varepsilon}}$. □

Finally we present the most general theorem that in some way is a combination of both previous ones.

**Theorem 3.4.** *Let $\varepsilon$ be a real number from $(0, 1)$. Then there exists constant $d$ such that for $c = c(n) = np > d$, with high probability, either $k-$core is empty or $v(k; n, p) > n - nc^{-0.5+\varepsilon}$.*

*Proof.* First we shall use the previous theorem to settle the result in case when $k < c - c^{-0.5}$. Suppose then that $k \geq c - c^{-0.5}$ and let $S_k$ be the set of vertices outside the $k-$core. If it is not empty, then from theorem 3.2, we have that $|S_k| < 0.9998$. We shall prove that if $|S_k| > nc^{-0.5+\varepsilon}$ then at least half of the vertices of a $k-$core has more than $0.5c^{0.5+\varepsilon}$ neighbours in $S_k$. If we fix $S_k$ and $v \notin S_k$ then the number of vertices in $S_k$ connected to $v$ has binomial distribution with parameters $|S_k|$ and $p$. One can check that for such random variable following inequalities hold

$$\mathbb{P}([|S_k| > (1+\varepsilon)np]) < e^{-\frac{\varepsilon^2 np}{3}},$$

$$\mathbb{P}([|S_k| > (1-\varepsilon)np]) < e^{-\frac{\varepsilon^2 np}{2}}.$$

And thus, the probability of following

$$|S(v) \cap S_k| \leq 0.5^{0.5\varepsilon} \leq 0.5p|S_k|$$

is less then $e^{-0.1c^{0.5+\varepsilon}}$. And so, the probability that there exists such set $|S_k|$ for at least $0.5(n - |S_k|)$ vertices is at most

$$\sum_{i=nc^{-0.5+\varepsilon}} 0.9998n\binom{n}{s}\binom{n-s}{0.5(n-s)}e^{-0.1c^{0.5+\varepsilon}0.5(n-s)} \leq n2^n2^ne^{-c^{0.5}n} = o(1),$$

where the last inequality holds for large enough $c$. It follows the above, that at least half of the vertices of the $k-$core have more than $0.5c^{0.5+\varepsilon}$ neighbours outside it. Following similar estimations as in previous theorem one can easily obtain that at most $ne^{-c^{0.5\varepsilon}}$ vertices of $G(n, p)$ have degree greater than $c + c^{0.5(1+\varepsilon)}$. So at least one third of vertices from the $k-$core have less than

$$c + c^{0.5(1+\varepsilon)} - 0.5c^{0.5+\varepsilon} < c - 2c^{0.5(1+\varepsilon)}$$

neighbours inside the core. But as we deal with $k > c - c^{0.5(1+\varepsilon)}$, we obtain a contradiction. □

## References


[1] B. Bollobás, *The evolution of sparse graphs*, Graph Theory and Combinatorics (1984), 35–37.

[2] A. Frieze, *Maximum matchings in a class of random graphs*, J. Comb. Theory Ser. B **40** (1986), 196–212.

[3] M. Drmota, R. Kutzelnigg, *A precise analysis of cuckoo hashing*, ACM Trans. Algorithms, 2009.

[4] T. Łuczak, *Size and connectivity of the k-core of a random graph*, Discrete Mathematics **91** (2003), 61–68.

[5] B. Pittel M. Karoński, *Existence of a perfect matching in a random (1+1/e)-out bipartite graph*, J. Comb. Theory Ser. B **88** (2003), 1–16.

[6] D. Fernholz, V. Ramachandran, *The giant k-core of a random graph with a specified degree sequence*, Manuscript available at http://www.cs.utexas.edu/vlr/pubs.html (2003).

[7] P. Erdős, A. Rényi, *On random matrices*, Publ. Math. Inst. Hungar. Acad. Sci. **8** (1964), 455–461.

[8] ———, *On existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368.

[9] D. Fotakis, R. Pagh, P. Sanders, P. Spirakis, *Space efficient hash tables with worst case constant access time*, Theory of Computing Systems **38** (2005), 229–248.

[10] D. Walkup, *Matching in random regular bipartite digraphs*, Discrete Mathematics **31** (1980), 59–64.

[11] B. Pittel, J. Spencer, N. Wormald, *Sudden emergence of a giant k-core in a random graph*, J. Comb. Theory Ser. B **67** (2003), 111–151.


Faculty of Mathematics and Computer Science, Adam Mickiewicz University, ul. Umultowska 87, 61-614 Poznań, Poland