Decomposition of random graphs into complete bipartite graphs

Fan Chung∗ Xing Peng∗

Abstract

We consider the problem of partitioning the edge set of a graph $G$ into the minimum number $\tau(G)$ of edge-disjoint complete bipartite subgraphs. We show that for a random graph $G$ in $G(n, p)$, for $1/2 \geq p = \Omega(1)$, almost surely $\tau(G)$ is between $n - c_1 (\log n)^{3+\epsilon}$ and $n - c_2 \log n$ for any positive constant $\epsilon$ and some positive constants $c_i$.

1 Introduction

For a graph $G$, the bipartition number, denoted by $\tau(G)$, is the minimum number of complete bipartite subgraphs that are edge-disjoint and whose union is the edge set of $G$. In 1971, Graham and Pollak [6] proved that

$$\tau(K_n) = n - 1. \quad (1)$$

In particular, they showed that for a graph $G$ on $n$ vertices, the bipartition number $\tau(G)$ is bounded below as follows:

$$\tau(G) \geq \max \{ n_+, n_- \} \quad (2)$$

where $n_+$ is the number of positive eigenvalues and $n_-$ is the number of negative eigenvalues of the distance matrix of $G$. Then, (1) follows from (2). Since then, there have been a number of alternative proofs for (1) by using linear algebra [10, 11, 12] or by using matrix enumeration [13, 14].

Let $\alpha(G)$ denote the independence number of $G$, which is the maximum number of vertices so that there are no edges among some set of $\alpha(G)$ vertices in $G$. A star is a special bipartite graph in which all edges share a common vertex which we call the center of the star. For a graph $G$ on $n$ vertices, the edge set of $G$ can obviously be decomposed into $n - \alpha(G)$ stars. It follows immediately that

$$\tau(G) \leq n - \alpha(G). \quad (3)$$

Erdős conjectured (see [1]) that for a random graph $G$ in $G(n, 1/2)$, the equality in (3) almost surely holds:

A conjecture of Erdős: Almost all graphs $G$ on $n$ vertices satisfy

$$\tau(G) = n - \alpha(G). \quad (4)$$

∗Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, (fan@ucsd.edu, x2peng@ucsd.edu). Research is supported in part by ONR MURI N000140810747, and AFSOR AF/SUB 552082.
We remark that the above conjecture by Erdős implies that a decomposition by stars is the best possible for almost all graphs.

A random graph almost surely has an independent set of order $c \log n$ and therefore $\tau(G) \leq n - c \log n$. For the lower bound, for a random graph $G$, it is well known that almost surely the number of positive and negative eigenvalues of the adjacency matrix of $G$ is bounded above by $n/2 + c' \sqrt{n}$. The distance matrix $M$ of $G$ almost surely satisfies $M = A + 2\hat{A} = 2J - 2I + A$ where $J, I, A$ and $\hat{A}$ denote the all 1’s matrix, the identity matrix, the adjacency matrix of $G$, and the adjacency matrix of the complement graph $\bar{G}$ of $G$, respectively. The distribution of eigenvalues of $M$, with the exception of one eigenvalue, is basically the same as the distribution for $A$ up to a linear shift of $-2$. Wigner’s semi-circle law implies that any interval of length 1 contains at most $c' \sqrt{n}$ eigenvalues. Therefore $n_-, n_+ < n/2 + c'' \sqrt{n}$ for some constant $c''$. Consequently, the inequality in (2) yields a rather weak lower bound of $\tau(G) \geq n/2 + c' \sqrt{n}$. We will prove the following theorem which lends support to Erdős’ conjecture [4].

**Theorem 1** For a random graph $G$ in $G(n, p)$ with $p \leq \frac{1}{2}$ and $p = \Omega(1)$, almost surely the bipartition number $\tau(G)$ of $G$ satisfies

$$n - o((\log n)^{3+\epsilon}) \leq \tau(G) \leq n - 2 \log n$$

for $b = \frac{1}{p}$ and any positive constant $\epsilon$.

For sparse random graphs, we have the following theorem.

**Theorem 2** For a random graph $G$ in $G(n, p)$ with $p = o(1)$ and $p = \omega \left( \frac{\log^2 n}{\sqrt{n}} \right)$, almost surely the bipartition number $\tau(G)$ of $G$ satisfies

$$n - o \left( \frac{\log^3 n}{p^2} \right)^{1+\eta} \leq \tau(G) \leq n - \frac{2 - \epsilon}{p} \log(np)$$

for any positive constants $\eta$ and $\epsilon$. Here $\log$ denotes the natural logarithm.

We follow the standard notation that by $g(x) = \omega(f(x))$ we mean $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$ and we write $g(x) = \Omega(f(x))$ if $f(x) \leq cg(x)$ for some positive constant $c$.

We remark that the difficulty for computing $\tau(G)$ is closely related to the intractability of computing $\alpha(G)$. In general, the problem of determining $\alpha(G)$ is an NP-complete problem, as one of the original 21 NP-complete problems in Karp [8]. If $G$ does not contain a 4-cycle, then $\tau(G) = n - \alpha(G)$. Schrijver showed that the problem of determining $\alpha(G)$ for the family of $C_4$-free graphs $G$ remains NP-complete [2]. Therefore the problem of determining $\tau(G)$ is also NP-complete. Nevertheless, Theorems 1 and 2 imply that for almost all graphs $G$, we can bound $\tau(G)$ within a relatively small range.

We also consider a variation of the bipartition number by requiring an additional condition that no complete bipartite graph in the partition is a star. We define the strong bipartition number, denoted by $\tau'(G)$, to be the minimum number of complete bipartite graphs (which are not stars) needed to partition the edge set of $G$. If there is no such a partition then we define $\tau'(G)$ as $\infty$; if $|V(G)| \leq 2$ then we define $\tau'(G)$ to be zero. We will show that for a random graph $G \in G(n, p)$, the strong bipartition number satisfies $\tau'(G) \geq 1.0001n$ if $p = 1/2$ and $p \leq \frac{1}{2}$, and $\tau'(G) \geq n$ if $p = o(1)$. This provides further evidence for Erdős’ conjecture in [4] that the best bipartition decomposition consists of all stars for almost all graphs. However, we can not rule out yet the case of the existence of a mixed bipartition decomposition achieving $\tau(G)$ but consisting of some stars and some complete bipartite graphs which are not stars.
The paper is organized as follows: In the next section, we state some definitions and basic facts that we will use later. In Section 3 we establish upper bounds for the number of edges incident to several specified families of complete bipartite subgraphs. In Section 4 we consider the remaining uncovered edges and give corresponding lower bounds that our main theorems need. In Section 5 we show that almost surely the strong bipartition number is at least $1.0001n$ for a random graph on $n$ vertices. In Section 6 we use all the lemmas and the strong bipartition theorem to prove Theorem 1. Also Theorem 2 is proved in Section 7. A number of problems and remarks are mentioned in Section 8.

2 Preliminaries

Let $G = (V, E)$ be a graph. For a vertex $v \in V(G)$, the neighborhood $N_G(v)$ of $v$ is the set \{u: u \in V(G) and (u, v) \in E(G)\} and the degree $d_G(v)$ of $v$ is $|N_G(v)|$. For a hypergraph $H = (V, E)$ and $v \in V(H)$, we define the degree $d_H(v)$ to be $|\{F: v \in F and F \in E(H)\}|$. For $U \subseteq V(G)$, let $e(U)$ be the set of edges of $G$ with both endpoints in $U$ and $G[U]$ be the subgraph induced by $U$. Furthermore, $2^U$ denotes the power set of $U$. For two subsets $A$ and $B$ of $V$, we define $E(A, B) = \{(u, v) \in E: u \in A and v \in B\}$. We say $A$ and $B$ form a complete bipartite graph if $A \cap B = \emptyset$ and $(u, v) \in E(G)$ for all $u \in A$ and $v \in B$.

We will use the following versions of Chernoff’s inequality and Azuma’s inequality.

**Theorem 3** [3] Let $X_1, \ldots, X_n$ be independent random variables with

$$
\Pr(X_i = 1) = p_i, \quad \Pr(X_i = 0) = 1 - p_i.
$$

We consider the sum $X = \sum_{i=1}^n X_i$ with expectation $E(X) = \sum_{i=1}^n p_i$. Then we have

- **(Lower tail)** \(\Pr(X \leq E(X) - \lambda) \leq e^{-\lambda^2/2E(X)}\),
- **(Upper tail)** \(\Pr(X \geq E(X) + \lambda) \leq e^{-\lambda^2/2E(X)}\).

**Theorem 4** [2] Let $X$ be a random variable determined by $m$ trials $T_1, \ldots, T_m$, such that for each $i$, and any two possible sequences of outcomes $t_1, \ldots, t_i$ and $t_1, \ldots, t_{i-1}, t'_i$:

$$
|E(X|T_1 = t_1, \ldots, T_i = t_i) - E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = t'_i)| \leq c_i
$$

then

$$
\Pr(|X - E(X)| \geq \lambda) \leq 2e^{-\lambda^2/2\sum_{i=1}^m c_i^2}.
$$

The following lemma on edge density will be useful later.

**Lemma 1** Almost surely a random graph $G$ in $G(n, p)$ satisfies, for all $U \subset V(G)$ with $|U| \geq \sqrt{\log n}$,

$$
|e(U) - \frac{D}{2}|U|^2| \leq C|U|^3/2 \log^{1/2} n
$$

where $C$ is some positive constant.

The lemma follows from Theorem 3.

The following lemma is along the lines of a classical result of Erdős for random graphs [4]. We include the statement and a short proof here for the sake of completeness.

**Lemma 2** Almost surely all complete bipartite graphs $K_{A, B} G \in G(n, p)$ with $|A| \leq |B|$, $p \leq \frac{1}{2}$ and $p = \Omega(1)$ satisfy $|A| \leq 2 \log_b n$ where $b = 1/p$. 

3
Proof: For two subsets $A$ and $B$ of $V(G)$, with $|A| = |B| = k$, the probability that $A$ and $B$ form a complete bipartite graph in $G(n, p)$ is at most $p^{k^2}$. There are at most $\binom{n}{k} \binom{n}{k}$ choices for $A$ and $B$. For $k \geq 2 \log_n n$, we have

$$\binom{n}{k}^2 p^{k^2} = o(1)$$

as $p = \Omega(1)$. The lemma then follows. \qed

We note that Erdős’ result on Ramsey’s theorem \cite{Erdos1959} states that every 2-coloring of the edges of the complete graph $K_n$ contains a monochromatic clique of order $\frac{1}{2} \log_2 n$. It is not hard to show along the same lines that $G(n, 1/2)$ contains a complete bipartite graph $K_{A,B}$ with $|A| = |B| = \frac{1}{2} \log_2 n$ and the bound in Lemma \ref{lemma:complete_bipartite} is tight up to a constant factor.

The upper bounds in Theorems \ref{theorem:random_graph} and \ref{theorem:independence_number} are immediate consequences of \cite{Bollobas1982} and the classical results on the independence number $\alpha(G)$ of a random graph. An upper bound for $\alpha(G)$ can be found in \cite{Bollobas1982}. The problem of determining the independence number for a random graph has been extensively studied in the literature. The asymptotic order of $\alpha(G)$ for $G$ in $G(n, p)$ was determined in \cite{Bollobas1982, Janson2011}.

**Theorem 5** For $G \in G(n, p)$ the independence number $\alpha(G)$ satisfies the following:

1. \cite{Bollobas1982} If $p < 1 - \epsilon$ and $p = \Omega(1)$, almost surely $\alpha(G)$ is of order

   $$\alpha(G) = 2 \log_b n + o(\log n)$$

   where $b = 1/p$ and $\epsilon > 0$.

2. \cite{Bollobas1982} If $p = o(1)$, $\alpha(G)$ almost surely satisfies

   $$\alpha(G) \geq \frac{2}{p} \log np - (1 + o(1)) \log \log np.$$

   where $\log$ denotes the natural logarithm.

3 Edges covered by a given family of subsets

For a graph $G = (V, E)$ and $A \subset V$, we define

$$V(G, A) = \{v : v \in V(G) \setminus A \text{ and } \{u, v\} \in E \text{ for all } u \in A\}.$$

It immediately follows that $A$ and $B$ form a complete bipartite graph if $B$ is contained in $V(G, A)$, namely, $B \subseteq V(G, A)$. We say an edge $(u, v) \in E$ is covered by $A$ if either $u \in A$ and $v \in V(G, A)$ or $v \in A$ and $u \in V(G, A)$.

For $A = \{A_1, A_2, \ldots, A_k\} \subseteq 2^V$ and $\sigma$, a linear ordering of $[k]$, we define a function $l$ as follows. For notational convenience, we use $i$ to denote the $i$-th element under the ordering $\sigma$. For each $1 \leq i \leq k$, we define $G_i$ and $l(i)$ recursively. We let $G_1 = G$ and let $l(1)$ be an arbitrary subset of $V(G_1, A_1)$. Given $G_{i-1}$, we let $G_i$ be a new graph with the vertex set $V(G)$ and the edge set $E(G_{i-1}) \setminus E(A_{i-1}, l(i-1))$. We set $l(i)$ to be an arbitrary subset of $V(G_i, A_i)$. We define

$$f(G, A) = \max_{\sigma} \max_{l} \sum_{i=1}^{k} |E(A_i, l(i))|.$$
for $A = \{A_1, A_2, A_3\}$ with $A_1 = \{a, b\}, A_2 = \{b, c\}$, and $A_3 = \{c, d\}$. Here $f(G, A) = 4$ is achieved by $\sigma = \text{identity}$, $l(1) = \{e\}$, $l(2) = \emptyset$, and $l(3) = \{e\}$, or $\sigma = (213)$, $l(1) = \emptyset$ and $l(2) = l(3) = \{e\}$. We observe $f(G, A) \leq \sum_{i=1}^{k} |E(A_i, V(G, A_i))|$ as $l(i) \subseteq V(G, A_i)$. When $U \subset V(G)$ and $A \subset U$ for each $A \in \mathcal{A}$, we use $f(G, U, A)$ to denote $f(G[U], A)$. Note that $G[U]$ denotes the induced subgraph of $G$ on a subset $U$ of $V(G)$.

**Lemma 3** Suppose that for $G \in G(n, p)$ and $U \subseteq V(G)$, $\mathcal{A}$ is a family of 2-sets of $U$ with $|\mathcal{A}| \leq |U|$. Then almost surely we have

$$f(G, U, \mathcal{A}) \leq 2p^2|\mathcal{A}||U| + 8|U|\sqrt{|U|\log n}$$

for $p \leq \frac{1}{2}$.

**Proof:** We list edges with both endpoints in $U$ as $e_1, e_2, \ldots, e_m$ where $m = \binom{|U|}{2}$. For each $e_i = (u_i, v_i)$ for $1 \leq i \leq m$, we consider $T_i \in \{H, T\}$ where $T_i = H$ means $e_i$ is an edge and $T_i = T$ means $e_i$ is not an edge. To simplify the notation we use $X$ to denote the random variable $f(G, U, \mathcal{A})$ and notice that $X$ is determined by $T_1, \ldots, T_m$. Given the outcome $t_j$ of $T_j$ for each $1 \leq j \leq i - 1$ we wish to establish an upper bound for

$$|E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = H) - E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = T)|. \quad (5)$$

Let $\mathcal{K}_1$ be the set of graphs over $U$ such that $e_j$ is given by $t_j$ for each $1 \leq j \leq i - 1$ and $e_i$ is a non-edge. Similarly, let $\mathcal{K}_2$ be the set of graphs over $U$ such that $e_j$ is given by $t_j$ for each $1 \leq j \leq i - 1$ and $e_i$ is an edge. We have $|\mathcal{K}_1| = |\mathcal{K}_2|$. Thus we get

$$E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = H) = \sum_{K \in \mathcal{K}_1} f(K, \mathcal{A})\Pr(K \in \mathcal{K}_1)$$

and

$$E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = T) = \sum_{K \in \mathcal{K}_2} f(K, \mathcal{A})\Pr(K \in \mathcal{K}_2).$$

Define a mapping $\mu : \mathcal{K}_1 \to \mathcal{K}_2$ such that $E(K)$ and $E(\mu(K))$ differ only by $e_i$ for each $K \in \mathcal{K}_1$. We get $\mu$ is a bijection and $\Pr(K \in \mathcal{K}_1) = \Pr(\mu(K) \in \mathcal{K}_2)$. Therefore the expression $(5)$ can be bounded from above by

$$\sum_{K \in \mathcal{K}_1} |f(K, \mathcal{A}) - f(\mu(K), \mathcal{A})|\Pr(K \in \mathcal{K}_1).$$

Notice that each edge can be covered by at most once. We observe $|f(K, \mathcal{A}) - f(\mu(K), \mathcal{A})| \leq 2$ because $e_i$ and the other edge sharing one endpoint with $e_i$ could be covered by $\mathcal{A}$ in $\mu(K)$ but not in $K$. Therefore $(5)$ is bounded above by 2.
Now we apply Theorem 4 for \( \lambda = 8|U| \sqrt{|U| \log n} \) and \( c_i = 2 \). Then we have
\[
\Pr \left( |X - E(X)| \geq 8|U| \sqrt{|U| \log n} \right) \leq 2e^{-64|U|^3 \log n / 2 \sum_{i=1}^m c_i^2} \leq 2e^{-4|U| \log n},
\]
using the fact \( m \leq \frac{|U|^2}{2} \). To estimate \( E(X) \), we note that \( E(f(G, U, A)) \leq 2p^2|U| \) for a fixed \( A \in \mathcal{A} \). Therefore,
\[
E(X) \leq \sum_{A \in \mathcal{A}} E(f(G, U, A)) \leq 2p^2|\mathcal{A}| |U|.
\]
Thus (6) implies
\[
\Pr \left( X \geq 2p^2|\mathcal{A}| |U| + 8|U| \sqrt{|U| \log n} \right) \leq \Pr \left( |X - E(X)| \geq 8|U| \sqrt{|U| \log n} \right)
\leq 2e^{-4|U| \log n}.
\]
Recall the assumptions \( |\mathcal{A}| \leq |U| \) and \( |A| = 2 \) for each \( A \in \mathcal{A} \). For a fixed size of \( U \), the number of choices for \( U \) and \( \mathcal{A} \) is at most \( n|U||U|^2 \) which is less than \( n^3|U| \). Therefore the probability that there are some \( U \) and \( \mathcal{A} \) which violate the assertion in the lemma is at most \( ne^{-4|U| \log n^3} < \frac{1}{2} \) for sufficiently large \( n \) as \( |U| \geq 2 \). The lemma is proved. \( \square \)

The following lemmas for other families of sets \( \mathcal{A} \) have proofs which are quite similar to the proof of Lemma 3. We will sketch proofs here.

**Lemma 4** Suppose that for \( G \in G(n, p) \), \( \mathcal{A} \) is a family of subsets of \( U \subseteq V(G) \) satisfying \( |\mathcal{A}| \leq |U| \) and \( 2 \leq |A| \leq 2 \log_2 n \) for each \( A \in \mathcal{A} \). Then almost surely we have
\[
f(G, U, \mathcal{A}) \leq 2p^2|\mathcal{A}| |U| + 8|U| \log_2 n \sqrt{|U| \log |U|}
\]
for \( p \leq \frac{1}{2} \).

**Proof:** We will use the Azuma’s inequality. For \( e_i = (u_i, v_i) \), we define \( K_1, K_2, \) and a bijection \( \mu \) similarly. The only difference is that \( |f(K, A) - f(\mu(K), A)| \leq 2 \log_2 n \) for each \( K \in K_1 \). This is because \( e_i \) and at most other \( 2 \log_2 n - 1 \) edges sharing the same endpoint with \( e_i \) could be covered by \( A \) in \( \mu(K) \) but not in \( K \).

Therefore the corresponding expression for (6) can be upper bounded by \( 2 \log_2 n \). We can then estimate \( X = f(G, U, \mathcal{A}) \) by applying Theorem 4 with
\[
\lambda = 8|U| \log_2 n \sqrt{|U| \log_2 n \log |U|}
\]
and \( c_i = 2 \log_2 n \).

This leads to
\[
\Pr \left( |X - E(X)| \geq 8|U| \log_2 n \sqrt{|U| \log_2 n \log |U|} \right) \leq 2e^{-4|U| \log_2 n \log |U|},
\]
Since \( E(X) \leq \sum_{A \in \mathcal{A}} E(f(G, U, A)) \leq 2p^2|\mathcal{A}| |U| \) and the number of choices for \( U \) and \( \mathcal{A} \) can be bounded from above by
\[
n^{1+|U||U|^2} \log_2 n,
\]
we can bound the probability in (7) by \( 2e^{-4|U| \log_2 n \log |U| n^{1+|U||U|^2} \log_2 n} \leq \frac{1}{n} \) as \( |U| \geq 2 \). This completes the proof of the lemma. \( \square \)

**Lemma 5** Suppose that for \( G \in G(n, p) \), \( \mathcal{A} \) is a family of subsets of \( U \subseteq V(G) \) satisfying \( |\mathcal{A}| \leq |U| \) and \( 3 \leq |A| \leq 2 \log_2 n \) for each \( A \in \mathcal{A} \). Then almost surely we have
\[
f(G, U, \mathcal{A}) \leq 3p^3|\mathcal{A}| |U| + 8|U| \log_2 n \sqrt{|U| \log |U| \log_2 n}
\]
for \( p \leq \frac{1}{2} \).
Proof: The proof is similar to that of Lemma \[4\] The only difference is that we assume \(|A| \geq 3\) and therefore

\[
E(f(G, U, A)) \leq \sum_{A \in \mathcal{A}} E(f(G, U, A)) \leq 3p^3 |A||U|.
\]

We use Theorem \[3\] in a similar way as in the proof of Lemma \[4\] to complete the proof of Lemma \[5\].

**Lemma 6** Suppose that for \(G \in G(n, p)\), \(\mathcal{A}\) is a family of subsets of \(U \subseteq V(G)\) satisfying \(|A| \leq |U|^{1+\delta}\) and \(\delta \log_b |U| \leq |A| \leq 2 \log_2 n\) for each \(A \in \mathcal{A}\), where \(b = \frac{1}{p}\) and \(\delta\) is some positive constant. Then almost surely we have

\[
f(G, U, A) \leq \delta |A||U|^{1-\delta} \log_b |U| + 8|U|^{(3+\delta)/2} \log_2 n \sqrt{\log |U|} \log_2 n
\]

provided \(p \leq \frac{1}{2}\).

**Proof:** We use the assumptions on \(|A|\) to derive

\[
E(f(G, U, A)) \leq \sum_{A \in \mathcal{A}} E(f(G, U, A)) \leq \delta \log_b |U|p^{\delta \log_b |U|} |A||U| \leq \delta |A||U|^{1-\delta} \log_b |U|.
\]

Then we bound the number of choices for \(U\) and \(A\) from above by

\[
n^{1+|U|/2|U|^{1+\delta} \log_2 n}
\]

Applying Theorem \[3\] for \(\lambda = 8|U|^{(3+\delta)/2} \log_2 n \sqrt{\log |U|} \log_2 n\) and \(c_i = 2 \log_2 n\), the lemma then follows.

We remark that we only use Lemma \[6\] when \(p = \Omega(1)\) for proving the main theorem.

## 4 Bounding uncovered edges

In order to prove the bipartite decomposition theorem, we also need to establish lower bounds for the number of uncovered edges for a given family \(\mathcal{A}\) of subsets.

First, we will derive a lower bound on the number of uncovered edges for a collection \(\mathcal{A}\) of \(2\)-sets of \(V(G)\). Let \(S_0\) be the set of \(u \in V(G)\) such that \(u\) is in only one \(A \in \mathcal{A}\). For \(u\) in \(S_0\), we denote the only \(2\)-set containing \(u\) by \(A_u\). Our goal is to give a lower bound on the number of uncovered edges with both endpoints in \(S_0\). To simplify the estimate, we impose some technical restrictions and work on a subset \(S\) of \(S_0\). To do so, we will lose at most a factor of \(2\) in the lower bound estimate (which is tolerable). To form \(S\), for each \(A_u = \{u, v\}\) with \(u, v \in S_0\), we delete one of \(u\) and \(v\) arbitrarily from \(S_0\). Let \(T = \cup_{u \in S} (A_u \setminus u)\). Clearly \(S\) and \(T\) are disjoint. Furthermore, \(|S| \geq |T|\).

We define \(E'\) to be the set of edges \((u, v)\) with \(u \in S\) and \(v \in T\) and \(E''\) to be the set of edges \((u, v)\) with \(u, v \in S\). We assume \(|E'| = p\). For a fixed ordering \(\sigma\) of edges in \(E'\) and an ordering \(\tau\) of edges in \(E''\), we define sets \(E'_i\) and \(E''_i\) for each \(0 \leq i \leq \tau\) recursively. Let \(E'_0 = E'\) and \(E''_0 = E''\). Given \(E'_{i-1}\) and \(E''_{i-1}\) for each \(1 \leq i \leq \tau\), we assume the first edge in \(E'_{i-1}\) is \(e = (u, v)\) with \(u \in S\) and \(v \in T\). Let \((u', v') \in E''_{i-1}\) be the first edge such that \(u = u'\) and \(A_{u'} = \{v, v'\}\). We note edges \((u, v)\) and \((u', v')\) are covered by \(\{v, v'\}\). We define \(E'_i = E'_{i-1} \setminus (u, v)\) and \(E''_i = E''_{i-1} \setminus (u', v')\). If there is no such an edge \((u', v')\) then we define \(E'_i = E'_{i-1} \setminus (u, v)\) and \(E''_i = E''_{i-1}\). Finally we define

\[
g(G, S, T) = \min_{\sigma} \min_{\tau} |E''_0|,
\]

where \(\sigma\) and \(\tau\) range over all orderings of edges in \(E(S, T)\) and \(E(S)\), respectively.
Lemma 7 Suppose that a graph $G$ and each edge of $G$ is contained in at most one of the complete bipartite graphs $K_{A_i,B_i}$ with $|A_i| = 2 \leq |B_i|$. For $\mathcal{A} = \{ A : A = A_i \text{ for some } i \}$, let $S$ and $T$ denote two disjoint subsets of $V(G)$ satisfying the properties that (i) each $u \in S$ is in a unique $A \in \mathcal{A}$, (ii) $S$ does not contain any $A$ in $\mathcal{A}$, (iii) $T = \{ v : \{u,v\} \in \mathcal{A} \text{ for some } u \in S \}$. Then $g(G,S,T)$ is a lower bound for the number of uncovered edges by $\mathcal{A}$ with both endpoints in $S$.

Proof: We can choose an arbitrary order $\sigma$ on edges in $E(S,T)$. Suppose an edge $e = (u,v)$, with $u \in S$ and $v \in T$, is uniquely covered by $A$ in $\mathcal{A}$. If $A = \{w,v\}$ for some $w$ in $S$, we associate with $e$ the edge $(u,w) \in E(S,S)$. In addition, we choose the order $\tau$ on $E(S,S)$ to be consistent with $\sigma$ in the sense that the associated edges in $E(S,S)$ maintain the same order. If $A = \{w',v\}$ for some $w'$ not in $S$ or $e$ is not covered by any $A$, then we do not associate any edge to $e$.

We note that if each edge $e' = (u',v') \in E(S,S)$ is contained in a unique bipartite graph $K_{A',B'}$, then $A'$ is in $\mathcal{A}$ from the assumptions (i) $\sim$ (iii). From the definition of $g$, $e'$ will be removed in the process. Thus, $g(G,S,T)$ is a lower bound for the number of uncovered edges by $\mathcal{A}$. \qed

Lemma 8 For $G \in G(n,p)$ and disjoint subsets $S$ and $T$ in $V(G)$, almost surely we have

$$g(G,S,T) \geq p^3 \left( \frac{|S|}{2} \right) - 5|S|\sqrt{|S|\log n}$$

for all choices of $S$ and $T$, provided $p \leq \frac{1}{2}$ and $|S| \geq 4$.

Proof: We sketch the proof here which is similar to that of Lemma 7. We list edges with endpoints in $S \cup T$ as $e_1, \ldots, e_m$, where $m = \binom{|S| + |T|}{2}$. For each $e_i = (u_i, v_i)$ for $1 \leq i \leq m$, we consider $T_i \in \{H,T\}$ where $T_i = H$ means $e_i$ is an edge and $T_i = T$ means $e_i$ is not an edge. Let $X$ denote the random variable $g(G,S,T)$ for $G \in G(n,p)$. We note that $X$ is determined by $T_1, \ldots, T_m$. For the fixed outcome $t_j$ of $T_j$ for $1 \leq j \leq i - 1$, we consider

$$|E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = H) - E(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = T)|. \quad (8)$$

For $e_i = (u_i, v_i)$, if $u_i, v_i \in T$ then the outcome of $T_i$ does not contribute to $X$. If $u_i, v_i \in S$ then the outcome of $T_i$ can change by at most one depending on whether $e_i$ is covered or not. If $u_i \in S$ and $v_i \in T$ then the outcome of $T_i$ could effect $X$ by at most one. This is because $e_i$ could make another edge $(u_i, w)$ covered by the 2-set $\{v_i, w\}$. Thus $X$ is bounded above by one.

Applying Azuma’s theorem as stated in Theorem 3 with $\lambda = 5|S|\sqrt{|S|\log n}$ and $c_i = 1$, we have

$$\Pr \left( |X - E(X)| \geq 5|S|\sqrt{|S|\log n} \right) \leq 2e^{-6|S|\log n}, \quad (9)$$

using $m \leq 2|S|^2$. To estimate $E(X)$, we define $X_{u,v}$, for $u,v \in S$, to be the event $(u,v) \in E(G)$, $(u, A_u \setminus \{v\}) \not\subset E(G)$, and $(v, A_v \setminus \{u\}) \not\subset E(G)$. Here $G \in G(n,p)$ and $A_v$ (resp. $A_u$) is the only 2-set containing $v$ (resp. $u$). Let $I_{u,v}$ denote the random indicator variable for $X_{u,v}$. We note $\Pr(I_{u,v} = 1) = p^3$ and

$$E(X) \geq \sum_{u,v \in S} \Pr(I_{u,v} = 1) = p^3 \left( \frac{|S|}{2} \right).$$

Thus

$$\Pr \left( X \leq p^3 \left( \frac{|S|}{2} \right) - 5|S|\sqrt{|S|\log n} \right) \leq \Pr \left( X \leq E(X) - 5|S|\sqrt{|S|\log n} \right) \leq 2e^{-6|S|\log n}.$$
For a given size of $S$, there are at most $n^{2|S|}$ choices for $S$ and $T$ and $|S|^{2|S|}$ choices for $\sigma$ and $\tau$. By \cite{4}, the probability that there are some $S$ and $T$ which violate the lemma is at most $2n^{3|S|+1}e^{-6|S|\log n} < \frac{1}{n}$ provided $n$ is sufficiently large and $|S| \geq 4$. This proves the lemma.

Next, we wish to establish a lower bound on the number of uncovered edges for general cases of $\mathcal{A}$.

For $W \subset U \subset V(G)$, we consider $L: W \to 2^U \setminus W$ with the property $L(w) \cap L(w') = \emptyset$ for $w, w' \in W$. We define $h(G, U, W, L)$ to be the number of edges $(w, w')$ in $G$ such that $w, w' \in W$, $(w, z) \notin E(G)$ for each $z \in L(w')$, and $(w', z') \notin E(G)$ for each $z' \in L(w)$. We will use the following Lemma (late we will show that $h(G, U, W, L)$ gives a lower bound for the number of uncovered edges).

**Lemma 9** For $G \in G(n, p)$, $p = \Omega(1)$, $p \leq 1/2$, $b = \frac{1}{p}$, $|U| \geq 4$, suppose $W \subset U$ satisfies $|W| = |U|/\log^2 |U|$ and $L$ as defined above satisfies $1 \leq |L(w)| \leq c\log_b |U|$ for some positive constant $c$. Then almost surely we have

$$h(G, U, W, L) \geq c' |U|^{2-2c}/\log^4_b |U| - 2|U|^{3/2} \sqrt{\log n}$$

for all choices of $U$, $W$ and $L$, where $c'$ is some positive constant.

**Proof:** For $u, v \in W$, let $X_{u,v}$ denote the event that $(u, v) \in E(G)$, $(u, w) \notin E(G)$ for each $w \in L(v)$, and $(v, z) \notin E(G)$ for each $z \in L(u)$. Here $G \in G(n, p)$. The random indicator variable for $X_{u,v}$ is written as $I_{u,v}$. From the definition of $h$, we have $h(G, U, W, L) = \sum_{u,v \in W} I_{u,v} = Y$ for $G \in G(n, p)$. Since $\Pr(I_{u,v} = 1) \geq b^{-1-2c\log_b |U|}$, we have

$$E(Y) \geq b^{-1-2c\log_b |U|} \left( \frac{|W|}{2} \right) \geq c' |U|^{2-2c}/\log^4_b |U|,$$

for some constant $c'$. From the definition of $L$, we have $X_{u,v}$ are independent of one another. By applying the Chernoff’s bound for the lower tail in Theorem \cite{3} with $\lambda = 2|U|^{3/2} \sqrt{\log n}$, we have

$$\Pr(Y \leq c' |U|^{2-2c}/\log^4_b |U| - 2|U|^{3/2} \sqrt{\log n}) \leq \Pr(Y \leq E(Y) - 2|U|^{3/2} \sqrt{\log n}) \leq e^{-\frac{4|U|^{3/2} \log n}{2E(Y)}} \leq e^{-2|U| \log n},$$

using the fact that $E(Y) \leq |U|^2$. For a given size $|U|$ of $U$, it is straightforward to bound the number of choices for $U$, $W$ and $L$ from above by

$$n^{|U||U|^{|U|^2|U|\log^2 |U|}|U|^{2|U|\log |U|}} < n^{4|U| |U|^{2|U|\log |U|}} \leq n^{1.5|U|},$$

when $n$ is sufficiently large. The probability that there is some $U$, $W$ and $L$ which violate the lemma is at most $ne^{-2|U| \log n 1.5|U|} < \frac{1}{n}$ if $n$ is large enough. This completes the proof of the lemma.

## 5 A theorem on strong bipartition decompositions

Recall the strong bipartition number $\tau'(G)$ is the minimum number of complete bipartite graphs whose edges partition the edge set of $G$ and none of them is a star. We will prove the following theorem.

**Theorem 6** Suppose that for $G \in G(n, p)$, $U \subseteq V(G)$ is a vertex subset with $|U| \geq \eta(\log n)^{3+b}$ where $b = 1/p$, $\eta$ and $\epsilon$ are positive constants. For $p = \Omega(1)$ and $p \leq \frac{1}{3}$, almost surely we have

$$\tau'(G[U]) \geq 1.0001|U|.$$
The proof of Theorem \( \text{[5]} \) is based on several lemmas which we will first prove. In this section, we may assume that \( G \in G(n, p) \) satisfies the statements in all lemmas in the preceding sections. By Lemma \( \text{[1]} \) the number of edges in \( G[U] \) satisfies \( e(G[U]) = (\frac{e}{2} + o(1))u^2 \), here \( |U| = u \). We will prove Theorem \( \text{[6]} \) by contradiction. Suppose

\[
E(G[U]) = \bigcup_{i=1}^{m} E(K_{A_i, B_i})
\]

and \( m < \left(1 + \frac{1}{1000}\right) u \) where ‘\( \sqcup \)’ denotes the disjoint union. We assume further \( |A_i| \leq |B_i| \) for each \( 1 \leq i \leq m \). Lemma \( \text{[2]} \) implies \( |A_i| \leq 2 \log_2 n \) for each \( 1 \leq i \leq m \).

We define \( \mathcal{L} = \{ A_i : 1 \leq i \leq m \} \) and note that \( \mathcal{L} \) could be a multi-set. We consider three subsets of \( \mathcal{L} \) defined as follows:

\[
\mathcal{L}_1 = \{ A_i \in \mathcal{L} : |A_i| < \delta_1 \log_6 u \} \\
\mathcal{L}_2 = \{ A_i \in \mathcal{L} : |A_i| < \delta_2 \log_6 u \} \\
\mathcal{L}_3 = \{ A_i \in \mathcal{L} : |A_i| = 2 \},
\]

where

\[
\delta_1 = \min \left\{ \frac{e}{4(3 + e)} + \frac{1}{200} \right\} \quad \text{and} \quad \delta_2 = \frac{\delta_1}{10^4}.
\]

We have the following lemma.

**Lemma 10**: If \( |\mathcal{L}_2| \leq \left(\frac{1}{2} + \frac{1}{1500}\right) u \), then we have \( |\mathcal{L}_3| \geq \left(\frac{1}{2} + \frac{1}{3000}\right) u \).

**Proof**: We will first prove the following claim:

**Claim 1**: \( |\mathcal{L}_3| \geq \left(\frac{1}{2} - \frac{1}{250}\right) u \).

**Proof of Claim 1**: Suppose the contrary. By Lemma \( \text{[3]} \) the number of edges covered by \( \mathcal{L}_3 \) is at most

\[
2p^2 |\mathcal{L}_3| u + 8u \sqrt{u \log n}.
\]

By Lemma \( \text{[4]} \) the number of edges covered by \( \mathcal{L}_2 \setminus \mathcal{L}_3 \) (i.e., \( |A_i| \geq 3 \)) is at most

\[
3p^3 |\mathcal{L}_2 \setminus \mathcal{L}_3| u + 8u \log_2 n \sqrt{u \log u \log_2 n}.
\]

Therefore the total number of edges covered by \( \mathcal{L}_2 \) is at most

\[
2p^2 |\mathcal{L}_3| u + 3p^3 |\mathcal{L}_2 \setminus \mathcal{L}_3| u + 8u \sqrt{u \log n} + 8u \log_2 n \sqrt{u \log u \log_2 n}.
\]

Thus the number of edges which are not in any \( K_{A_i, B_i} \) with \( A_i \in \mathcal{L}_2 \) is at least

\[
\left(\frac{p}{2} + o(1)\right) u^2 - 2p^2 |\mathcal{L}_3| u - 3p^3 |\mathcal{L}_2 \setminus \mathcal{L}_3| u - 8u \sqrt{u \log n} - 8u \log_2 n \sqrt{u \log u \log_2 n}.
\]

Observe that the expression above is a decreasing function if we view \( |\mathcal{L}_3| \) as the variable. From the assumptions \( |\mathcal{L}_3| < \left(\frac{1}{2} - \frac{1}{250}\right) u \) and \( |\mathcal{L}_2| \leq \left(\frac{1}{2} + \frac{1}{1500}\right) u \), the number of edges which are not contained in any \( K_{A_i, B_i} \), with \( A_i \in \mathcal{L}_2 \) is at least

\[
\frac{p}{2} u^2 - \left(\frac{1}{2} - \frac{1}{250}\right) 2p^2 u^2 - 7 \frac{p^3}{500} u^2 + o(u^2) \geq \left(\frac{p^2}{125} - \frac{7p^3}{500} + o(1)\right) u^2
\]

when \( n \) is large enough. Here we note that \( u \log_2 n \sqrt{u \log u \log_2 n} = o(u^2) \) as we assume \( u \geq \eta(\log_6 n)^{3+\epsilon} \). Since \( p \leq \frac{1}{2} \) and \( p = \Omega(1) \), we get that \( \frac{p^2}{125} - \frac{7p^3}{500} \) is a positive constant.

Applying Lemma \( \text{[3]} \) with \( \delta = \delta_2 \), the number of edges covered by \( \mathcal{L} \setminus \mathcal{L}_2 \) (i.e., \( |A_i| \geq \delta_2 \log_6 n \)) is at most

\[
\delta_2 |\mathcal{L} \setminus \mathcal{L}_2| u^{1-\delta_2} \log_6 u + 8u^{(3+\delta_2)/2} \log_2 n \sqrt{\log u \log_2 n}.
\]
Since \( u^{(3+\delta_2)/2} \log n \sqrt{\log u \log_2 n} = o(u^2) \) by the choice of \( \delta_2 \), in order to cover the remaining edges, we need at least \( C_1 u^{1+\delta_2/2} \) extra complete bipartite graphs for some positive constant \( C_1 \), i.e., \( |C \setminus L_2| \geq C_1 u^{1+\delta_2/2} \). Since \( C_1 u^{1+\delta_2/2} > (1 + \frac{10000}{\sqrt{n}}) \) for sufficiently large \( n \), this leads to a contradiction. Thus we have \( |L_3| \geq (\frac{1}{2} - \frac{1}{2000}) u \) and the claim is proved.

Now we proceed to prove the lemma using the fact that \( |L_3| \geq (\frac{1}{2} - \frac{1}{2000}) u \). We consider a auxiliary graph \( U^* \) whose vertex set is \( U \) and edge set is \( L_3 \). It is possible that \( U^* \) has multiple edges. We partition the vertex set of \( U^* \) into three sets \( U_1, U_2, \) and \( U_3 \), where \( U_1 = \{ v \in V(U^*) : d_{U^*}(v) = 0 \} \), \( U_2 = \{ v \in V(U^*) : d_{U^*}(v) = 1 \} \), and \( U_3 = \{ v \in V(U^*) : d_{U^*}(v) \geq 2 \} \). We will prove the following.

**Claim 2:** The number of edges not contained in any \( K_{A_i,B_i} \) for \( A_i \in L_3 \) is at least \( p(\frac{|U_1|}{2}) + p^3(\frac{|U_2|}{2}) + o(u^2) \).

*Proof of Claim 2:* The first part of the sum follows from Lemma 11. For the second part of the sum, we let \( U_2' \subseteq U_2 \) such that for each \( v \in U_2' \), the neighbor of \( v \) in \( U^* \) is not in \( U_2 \). We have \( |U_2'| \geq |U_2|/2 \). Then we apply Lemma 5 with \( S = U_2' \) and \( T \) consisting of neighbors of \( S' \) in \( U^* \). To finish the proof of Claim 2, we use the facts \( u = \Omega(\log^3 n) \) and \( |S| \geq \Omega(\log n) \) and \( u = o(u^2) \).

We will prove Lemma 10 by contradiction. Suppose \( |L_3| \geq (\frac{1}{2} + \frac{1}{1000}) u \). This implies that the average degree of \( U^* \) is at most \( 1 + \frac{1}{1000} \). We consider the following cases.

**Case 1:** \( |U_3| \geq \frac{1}{4} u \).

By considering the total sum of degrees of \( U^* \), we have
\[
2|U_3| + (u - |U_1| - |U_3|) \leq \left( 1 + \frac{1}{1000} \right) u.
\]

Thus, \( |U_3| \geq \frac{1}{4} u \) implies \( |U_1| \geq \frac{1}{4} u \). Claim 2 together with this lower bound on \( |U_1| \) implies that the number of edges not in any \( K_{A_i,B_i} \) with \( A_i \in L_3 \) is at least \( (\frac{1}{4} + o(1)) u^2 \). By Claim 1 we have \( |L_3| \geq (\frac{1}{2} - \frac{1}{2000}) u \), so the number of additional complete bipartite graphs \( K_{A_i,B_i} \) with \( A_i \in L_2 \setminus L_3 \) is at most \( \frac{10000}{\sqrt{n}} u \) using the assumption \( |L_3| \geq (\frac{1}{2} - \frac{1}{2000}) u \). These complete bipartite graphs can cover at most \( (\frac{p^3}{500} + o(1)) u^2 \) edges by Lemma 5. Thus we conclude that the number of edges not covered by any of \( K_{A_i,B_i} \) with \( A_i \in L_2 \) is at least
\[
\left( \frac{p}{72} - \frac{7p^3}{500} + o(1) \right) u^2.
\]

Note that \( \frac{p}{72} - \frac{7p^3}{500} \) is a positive constant when \( p = \Omega(1) \) and \( p \leq \frac{1}{2} \). By applying Lemma 6 with \( \delta = \delta_2 \), the bipartite graphs \( K_{A_i,B_i} \) with \( A_i \in L \setminus L_2 \) (i.e., \( |A_i| \geq \delta_2 \log_2 n \)) can cover at most
\[
\delta_2 |L \setminus L_2| u^{1-\delta_2} \log_2 u + 8u^{(3+\delta_2)/2} \log_2 n \sqrt{\log u \log_2 n}
\]
edges. We note \( u^{(3+\delta_2)/2} \log_2 n \sqrt{\log u \log_2 n} = o(u^2) \) because of the choice of \( \delta_2 \). To cover the remaining edges, we need at least \( C'_1 u^{1+\delta_2/2} \) extra complete bipartite graphs \( K_{A_i,B_i} \) with \( A_i \in L \setminus L_2 \) for some positive constant \( C_1 \), i.e., \( |L \setminus L_2| \geq C'_1 u^{1+\delta_2/2} \). Since \( C'_1 u^{1+\delta_2/2} > (1 + \frac{10000}{\sqrt{n}}) u \) for \( n \) large enough, we get a contradiction to the assumption \( |L| \leq (1 + \frac{1}{10000}) u \).

**Case 2:** \( |U_3| < \frac{1}{4} u \).

In this case we have \( |U_1| + |U_2| \geq \frac{1}{4} u \). Note that the lower bound given by Claim 2 is minimized when \( |U_2| = \frac{1}{4} u \), i.e., the number of edges not contained in any \( K_{A_i,B_i} \)
with \( A_i \in \mathcal{L}_3 \) is at least \((\frac{33}{300}p^3 + o(1))u^2\). By the same argument as in Case 1 we can show the number of edges in \( K_{A_i,B_i} \) with \( A_i \in \mathcal{L}_2 \setminus \mathcal{L}_3 \) is at most \((\frac{33}{300}p^3 + o(1))u^2\). Now there are at least
\[
\left( \frac{33}{300}p^3 + o(1) \right) u^2
\]
edges which is not in any \( K_{A_i,B_i} \) with \( A_i \in \mathcal{L}_2 \). We note that \( \frac{33}{300}p^3 \) is a positive constant under the assumption \( p = \Omega(1) \). By using Lemma 10 with \( \delta = \delta_2 \), the bipartite graphs \( K_{A_i,B_i} \) with \( A_i \in \mathcal{L} \setminus \mathcal{L}_2 \) (i.e., \(|A_i| \geq \delta_2 \log_b n\)) can cover at most
\[
\delta_2 |\mathcal{L} \setminus \mathcal{L}_2| u^{1-\delta_2} \log_b u + o(u^2)
\]
edges. As in Case 1, we consider the number of extra bipartite graphs \( K_{A_i,B_i} \) with \( A_i \in \mathcal{L} \setminus \mathcal{L}_2 \) needed to cover the remaining edges, leading to the same contradiction to the assumption on \( \mathcal{L} \).

Therefore we have proved \(|\mathcal{L}_3| > (\frac{1}{2} + \frac{1}{2000})u\).

**Lemma 11** Let \( H \) be a hypergraph with the vertex set \( U \) and the edge set \( \mathcal{L}_1 \). There is some positive constant \( C_2 \) such that there are \( C_2u \) vertices of \( H \) with degree less than \((\frac{1}{2} - \frac{1}{3000}) \log_b u\).

**Proof:** We consider several cases.

**Case a:** \(|\mathcal{L}_2| > (\frac{1}{2} + \frac{1}{1500})u\).

The sum of degrees in \( H \) is less than
\[
\delta_2 |\mathcal{L}_2| \log_b u + \delta_1 |\mathcal{L}_1 \setminus \mathcal{L}_2| \log_b u \leq \delta_2 \left( \frac{1}{2} + \frac{1}{1500} \right) u \log_b u + \delta_1 \left( \frac{1}{2} + \frac{1}{10000} - \frac{1}{1500} \right) u \log_b u
\]
\[
\leq \left( \frac{\delta_1}{2} - \frac{\delta_1}{2000} \right) u \log_b u.
\]

Here we used the assumption \(|\mathcal{L}_1| \leq |\mathcal{L}| = m < (1 + \frac{1}{1000})u \) and the choice of \( \delta_2 \).

**Case b:** \(|\mathcal{L}_2| \leq (\frac{1}{2} + \frac{1}{1500})u\).

By Lemma 10 \(|\mathcal{L}_3| \geq (\frac{1}{2} + \frac{1}{2000})u\). The sum of degrees is at most
\[
2 |\mathcal{L}_3| + \delta_1 |\mathcal{L}_1 \setminus \mathcal{L}_3| \log_b u \leq 2 \left( \frac{1}{2} + \frac{1}{2000} \right) u + \delta_1 \left( \frac{1}{2} + \frac{1}{10000} - \frac{1}{2000} \right) u \log_b u
\]
\[
\leq \left( \frac{\delta_1}{2} - \frac{\delta_1}{2000} \right) u \log_b u.
\]

We have proved that the sum of degrees of \( H \) is less than \((\frac{\delta_1}{2} - \frac{\delta_1}{3000}) u \log_b u \). Let \( U' \) be the set of vertices with degree at least \((\frac{\delta_1}{2} - \frac{\delta_1}{3000}) \log_b u \). We consider
\[
|U'| \left( \frac{\delta_1}{2} - \frac{\delta_1}{3000} \right) \log_b u \leq \left( \frac{\delta_1}{2} - \frac{\delta_1}{2000} \right) u \log_b u,
\]
which yields \(|U'| \leq (1 - C_2)u\) for some positive constant \( C_2 \). Each vertex in \( U \setminus U' \) has degree less than \((\frac{\delta_1}{2} - \frac{\delta_1}{3000}) \log_b u \) and \(|U \setminus U'| \geq C_2 u\). The lemma is proved.

We recall that \( G[U] \) is the subgraph of \( G \) induced by \( U \). We have the following lemma.

**Lemma 12** The number of edges in \( G[U] \) which are not contained in any \( K_{A_i,B_i} \) with \( A_i \in \mathcal{L}_1 \) is at least \( C_3 u^{2 - \delta_1 - \delta_2 / 2000} \) for some positive constant \( C_3 \).
Proof: We consider the hypergraph $H$ with the vertex set $U$ and the edge set $\mathcal{L}_1$ as defined in Lemma 11. Let $W$ be the set of vertices with degree less than $\left(\frac{1}{2} - \frac{3}{2000}\right) \log_2 u$ in $H$; we have $|W| \geq C_2 u$ for some positive constant $C_2$ by Lemma 11.

We will use Lemma 9 to prove Lemma 12. In order to apply Lemma 9, we will first find a subset $W'$ of $W$ such that for any $u, v \in W'$ there is no $A_i \in \mathcal{L}_1$ containing $u$ and $v$. Also we will associate each $w \in W'$ with a set $L(w) \subset U \setminus W'$ satisfying the property that $L(w) \cap L(w') = \emptyset$ for each $w \neq w' \in W'$.

To do so, we consider an arbitrary linear ordering of vertices in $W$. Let $q = |W|/\log_2 u$, $W_0 = W$, $Z_0 = \emptyset$ and $H_0 = H$. For each $1 \leq i \leq q$, we recursively define a vertex $v_i$, a set $W_i$, a set and a hypergraph $H_i$ as follows: For given $W_{i-1}$ and $H_{i-1}$, we let $v_i$ be the first vertex in $W_{i-1}$ and define $F(v_i) = \{A \in E(H_{i-1}) : v_i \in A\}$. By the assumption on the size of sets in $\mathcal{L}_1$ and the degree upper bound for vertices in $W$, we have $|\bigcup_{A \in F(v_i)} A| \leq \log_2 u/2$. We define $Z_i = \{A \in E(H_{i-1}) : |A \setminus (\bigcup_{A' \in F(v_i)} A')| = 1\}$. Then $|\bigcup_{A \in Z_i, A \setminus (\bigcup_{A' \in F(v_i)} A')}| \leq \log_2 u/2$ since each $A' \in F(v_i)$ can contribute at most $\delta_1 \log_2 u$ to the sum and $|F(v_i)| \leq \frac{1}{8} \log_2 u$ because of the degree upper bound for vertices in $W$. We define $W_i = W_{i-1} \setminus \left(\bigcup_{A \in Z_i} A\right)$ and $H_i$ to be the new hypergraph with the vertex set $V(H_i) = \left(\bigcup_{A \in Z_i} A\right)$ and $E(H_i) = E(H_{i-1}) \cup \bigcup_{A \in Z_i} E(A)$. If $A \in E(H_{i-1})$, then $|E(H_i)| = |E(H_{i-1})| + \log_2 u$ and $W_{i-1} \setminus \log_2 u$. Therefore, $v_i$ is well-defined for $1 \leq i \leq q$. We write $W' = \{v_1, v_2, \ldots, v_q\}$.

For each $A \in F(v_i)$ and $A' \in F(v_j)$ with $i < j$ we have $A \cap A' = \emptyset$ as we delete the set $\bigcup_{A \in Z_i} A$ in step $i$. For each $v_i \in W'$ and each $A \in F(v_i)$, we let $f(A)$ be an arbitrary vertex other than $v_i$ from $A$ and $F'(v_i) = \bigcup_{A \in F(v_i)} f(A)$. It follows from the preceding definitions that $F'(v_i) \cap F'(v_j) = \emptyset$ for $1 \leq i \neq j \leq q$. Furthermore, for each $v_i$ and each $A \in \mathcal{L}_1$ containing $v_i$, either $A$ is in $F(v_i)$ or a subset of $A$ with size at least two is in $F(v_i)$. Hence, $A \cap F'(v_i) \neq \emptyset$. For an edge $(v_i, v_j)$, if $(v_i, z)$ is a non-edge for each $z \in F'(v_i)$ and $(v_j, z')$ is a non-edge for each $z' \in F'(v_i)$, then the edge $(v_i, v_j)$ is uncovered by the family of sets $\mathcal{L}_1$. Suppose $(v_i, v_j)$ is in $K_{A,B}$ for some $A \in \mathcal{L}_1$. We have either $v_i \in A$ or $v_j \in A$. In the former case we get $A \cap L(v_i) \neq \emptyset$ by the definition of $L(v_i)$. Let $z \in A \cap L(v_i)$. Then $A$ and $B$ does not form a complete bipartite graph since $(v_j, z)$ is not an edge by the assumption. We get a contradiction and we have a similar argument for the later case. Therefore the function $h(G, U, W, L)$ gives a lower bound for the number of uncovered edges in $G$ for the given family of sets $\mathcal{L}_1$.

Now we apply Lemma 9 with $U = V(H)$, $W = W'$, $L(v_i) = F'(v_i)$ for each $v_i$ and $c = \frac{1}{2} - \frac{3}{2000}$. The lemma then follows.

We are ready to prove Theorem 6.

Proof of Theorem 6. Suppose that

$$E(G[U]) = \bigcup_{i=1}^m E(K_{A_i,B_i}).$$

If $m > (1 + \frac{1}{10000}) u$, then we are done. Otherwise, Lemma 12 implies that there are at least $C_3 u^{2 - \delta_1 + \delta_1/2000}$ edges uncovered after we delete the edges in $K_{A_i,B_i}$ for each $A_i \in \mathcal{L}_1$. We then apply Lemma 9 with $\delta = \delta_1$ which gives an upper bound for the number of edges covered by $\mathcal{L} \setminus \mathcal{L}_1$ (i.e., $|A_i| \geq \delta_1 \log_2 u$) : $$(\delta_1 |\mathcal{L} \setminus \mathcal{L}_1| u^{1 - \delta_1} \log_2 u + 8u^{3+\delta_1/2}/\log_2 n \sqrt{\log u \log_2 n}.$$ Here we note that $u^{3+\delta_1/2} \log_2 n \sqrt{\log u \log_2 n} = o(u^{2-\delta_1 + \delta_1/2000})$ because of the choice of $\delta_1$. Therefore we need at least $C_4 u^{1+\delta_1/2000}$ additional complete bipartite graphs $K_{A_i,B_i}$ with $A_i \in \mathcal{L} \setminus \mathcal{L}_1$ to cover the remaining edges, where $C_4$ is some positive constant. Since $C_4 u^{1+\delta_1/2000} > 1.0001 u$ when $n$ is sufficiently large and we get a contradiction. Theorem 6 is proved.
6 Proof of Theorem \[14\]

Before proving Theorem 1, we first prove the following lemma.

**Lemma 13** Suppose that edges of $G$ can be decomposed into $k_1$ complete bipartite graphs, of which $k_2$ complete bipartite graphs are stars for some $k_2 \leq k_1$. Then $G$ has an edge decomposition $E(G) = \bigcup_{i=1}^{k} E(K_{A_i,B_i})$ with $k \leq k_1$ such that for $i \leq k_2$, $K_{A_i,B_i}$ are stars and for $j > k_2$, we have $A_j, B_j \subseteq V(G) \setminus \bigcup_{i=1}^{k_1} A_i$.

**Proof:** For an edge decomposition $B = \{K_{A_1,B_1}, \ldots, K_{A_{k_1},B_{k_1}}\}$, we can modify $B$ by the following algorithm.

**Algorithm A**

Input $G$ and $B$.

**Step 1:** Set $G' = G$, $V' = \emptyset$ and $B' = \emptyset$.

**Step 2:** If none of $K_{A_i,B_i} \in B$ is a star, then stop and output $B$. Otherwise go to Step 3.

**Step 3:** For $i = 1, \ldots, k_1$, if $K_{A_i,B_i}$ is a star, add the center of the star to $V'$ and add $K_{A_i,B_i}$ to $B'$.

**Step 4:** For each $K_{A_i,B_i} \in B \setminus B'$, replace it by $K_{A_i',B_i'}$ where $A_i' = A_i \setminus V'$ and $B_i' \setminus V'$. For each star $S = K_{A_i,B_i}$ in $B'$ with $A_i = \{v_i\}$, define $S_i$ to be the star centered at $v_i$ containing all edges incident to $v_i$ in $G \setminus \{v_1, \ldots, v_{i-1}\}$. Then we replace $S = K_{A_i,B_i}$ by $S'$ in $B$.

**Step 5:** Output $B$.

We note that the cardinality of $B$ does not increase throughout Algorithm A, although it is possible that some member of $B$ might have no edge left. In that case, $|B|$ decreases. \[\square\]

**Lemma 14** For a graph $G$, there is a subset $T \subseteq V(G)$ such that

$$\tau(G) = n - |T| + \tau'(G[T]).$$

**Proof:** We use the notation in the proof of Lemma 13. Note that for $k_1 = \tau(G)$, we have $k \geq k_1$ so that the output of Algorithm A have size $k = k_1$. We choose $T = V \setminus V'$. Then all non-stars in the decomposition in $B$ are an edge decomposition for $G[T]$. Therefore, we have $\tau(G) = k = |V'| + \tau'(G[T])$. \[\square\]

We are ready to prove Theorem 1.

**Proof of Theorem 1** The upper bound follows from the well known fact (see Theorem 5) that almost surely a random graph $G \in G(n,p)$ has an independent set $I$ with size $2 \log_b n$ where $b = 1/p$ and $p = \Omega(1)$. We consider vertices $v_1, \ldots, v_m$ with $m = n - 2 \log_b n$, which are not contained in $I$. For each $1 \leq i \leq m$ we define a star $K_{A_i,B_i}$ with $A_i = \{v_i\}$ and $B_i = \{v_j \colon j > i \text{ and } (v_i,v_j) \in E(G)\}$. We have

$$E(G) = \bigcup_{i=1}^{m} E(K_{A_i,B_i}).$$

Therefore we have $\tau(G) \leq n - 2 \log_b n$.

For the lower bound, we may assume that $G \in G(n,p)$ satisfies all statements in the lemmas in the preceding sections. Suppose $G$ has an edge decomposition:

$$E(G) = \bigcup_{i=1}^{k} E(K_{A_i,B_i}),$$

where
with $k = \tau(G)$ and assume that for some $l \leq k$, we have $A_i = \{v_i\}$ for $1 \leq i \leq l$.

Let $W = \{v_1, \ldots, v_l\}$. If $W = \emptyset$ then Theorem 1 follows from Theorem 3 directly. We need only to consider the case $W \neq \emptyset$. By Algorithm A we can assume $E(G') = \bigcup_{i=l+1}^k E(K_{A_i, B_i})$ where $G'$ is the subgraph induced by $T = V(G) \setminus W$. We get

$$\tau(G) = |W| + \tau'(G').$$

(10)

We will prove $l > n - \eta(\log_b n)^{3+\epsilon}$ for some positive constants $\eta$ and $\epsilon$. Suppose $l \leq n - \eta(\log_b n)^{3+\epsilon}$. Thus, $|T| \geq \eta(\log_b n)^{1+\epsilon}$. By Theorem 3 we have $\tau'(G[T]) \geq (1 + \frac{1}{10000})|T|$. Therefore

$$\tau(G) = |W| + \tau'(G') \geq |W| + (1 + \frac{1}{10000})|T| \geq n,$$

which is a contradiction. Theorem 1 is proved. \hfill \Box

7 Proof of Theorem 2

It remains to deal with the sparse case that $p = o(1)$. By using the lemmas previously stated in preceding sections, the proof for the sparse case is shorter.

Proof of Theorem 2 For the lower bound, we prove the following:

Claim A: For $G \in G(n, p)$, $U \subset V(G)$ with $|U| \geq \zeta \left(\frac{\log^2 u}{p^2}\right)^{1+\eta}$, and $p = o(1)$, where $\zeta$ and $\eta$ are positive constants, almost surely we have $\tau(G[U]) \geq |U|$.

Proof of Claim A: We write $|U| = u$ and assume the following partition

$$E(G[U]) = \bigcup_{i=1}^m E(K_{A_i, B_i})$$

(11)

for some $m < u$. We can assume further $|A_i| \leq |B_i|$. By Lemma 2 with $p = \frac{1}{2}$, almost surely we have $|A_i| \leq 2 \log n$ for $G \in G(n, p)$ when $p = o(1)$. By the assumption of the size of $U$ and Lemma 3 we have $e(U) = \left(\frac{p}{2} + o(1)\right) u^2$. Applying Lemma 3 with $A = \{A_1, \ldots, A_m\}$ we get

$$|\bigcup_{i=1}^m E(K_{A_i, B_i})| \leq 2p^2 mu + 8u \log n \sqrt{u \log u \log \log n}.$$

We note $2p^2 mu = o(u^2)$ as $p = o(1)$ and $m < u$; also $u \log n \sqrt{2u \log u \log n} < \frac{mu^2}{2}$ by the assumption on $u$. We obtain a contradiction to (11) and we have proved Claim A.

The remaining proof for the lower bound is quite similar to the proof of Theorem 1 and will be omitted.

For the upper bound, a result by Frieze 1 states that in $G(n, p)$, with $p = o(1)$, almost surely there is an independence set with size at least $\frac{2+\epsilon}{p} \log np$ for any small positive constant $\epsilon$. This completes the proof of Theorem 2. \hfill \Box

8 Problems and remarks

The conjecture of Erdős, that almost all graphs $G$ satisfies $\tau(G) = n - \alpha(G)$ as stated in 4, remains unsolved. Here we state the following slightly weaker conjectures:

Conjecture 1: For a random graph $G \in G(n, p)$, with $1/2 \geq p = \Omega(1)$, almost surely

$$\tau(G) = n - (2 + o(1)) \log_b n$$
where \( b = 1/p \).

**Conjecture 2:** For a random graph \( G \in G(n, p) \), with \( p = o(1) \), almost surely

\[
\tau(G) = n - (1 + o(1)) \frac{2}{p} \log np,
\]

where \( \log \) denotes the natural logarithm.

**Conjecture 3:** For a random graph \( G \in G(n, p) \), with \( 1/2 \geq p = \Omega(1) \), suppose that an edge decomposition \( E(G) = \bigcup_{i=1}^{k} E(K_{A_i, B_i}) \) achieves \( \tau(G) = k \). Then at least \( n - o(n) \) of the bipartite subgraphs \( K_{A_i, B_i} \) are stars.

In this paper, we have given rather crude estimates for the constants involved. In particular, for the strong bipartition number \( \tau'(G) \), a consequence of Theorem 6 states that for \( G \in G(n, 1/2) \), we have \( \tau'(G) \geq 1.0001n \). A natural question is to improve the constant here.

In the other direction, it is of interest to characterize graphs with specified upper bounds for \( \tau' \).

**Problem 4:** Characterize graphs \( G \) such that \( \tau'(G) \leq n \).

**References**


