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 $G(n, p)$, where p is a constant no greater than $1/2$, asymptotically almost surely $\tau(G)$ is between $n - c(\log_{1/p} n)^{3+\epsilon}$ and $n - (2 + o(1)) \log_{1/(1-p)} n$ for any positive constants c and ϵ.

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

$$\tau(K_n) = n - 1. \tag{1}$$

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

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

where n_+ is the number of positive eigenvalues and n_- is the number of negative eigenvalues of the adjacency 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 centered at vertices in the complement of a largest independent set. It follows immediately that

$$\tau(G) \leq n - \alpha(G). \tag{3}$$

It was mentioned in a 1988 paper [9], by Kratzke, Reznick and West, that Erdős conjectured the equality in (3) holds for almost all graphs $G \in G(n, 1/2)$, although we could not find this conjecture in any other publication on Erdős' problems.

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, (fan@ucsd.edu),
Research is supported in part by ONR MURI N000140810747, and AFSOR AF/SUB 552082.
†Center for Applied Mathematics, Tianjin University, Tianjin, 300072 China, (x2peng@tju.edu.cn).

Let $\beta(G)$ be the size of the largest induced complete bipartite graph in G. Another possible upper bound for $\tau(G)$, as pointed out by Alon [1], is

$$\tau(G) \leq n - \beta(G) + 1.$$

This follows from the fact that edges in G can be partitioned into $n - \beta(G)$ stars and a largest induced complete bipartite graph. Therefore we get

$$\tau(G) \leq \min\{n - \alpha(G), n - \beta(G) + 1\}.$$

A random graph asymptotically 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 asymptotically almost surely the number of positive and negative eigenvalues of the adjacency matrix of G is bounded above by $n/2 + c'\sqrt{n}$. Consequently, the inequality in (2) yields a rather weak lower bound of $\tau(G)$. We will prove the following theorem.

Theorem 1 *For a random graph G in $G(n, 1/2)$, asymptotically almost surely the bipartition number $\tau(G)$ of G satisfies*

$$n - c(\log_2 n)^{3+\epsilon} \leq \tau(G) \leq n - (2 + o(1)) \log_2 n$$

for any positive constants c and ϵ.

Theorem 1 is a special case of the following theorem.

Theorem 2 *For a random graph G in $G(n, p)$, where p is a constant no greater than $1/2$, asymptotically almost surely the bipartition number $\tau(G)$ of G satisfies*

$$n - c(\log_{1/p} n)^{3+\epsilon} \leq \tau(G) \leq n - (2 + o(1)) \log_{1/(1-p)} n$$

for any positive constants c and ϵ.

We remark that our techniques can be extended to the case where $p \leq 1 - c$ for any positive constant c, but we restrict our attention here to the case where $p \leq 1/2$.

Alon [1] disproved Erdős' conjecture by showing asymptotically almost surely $\tau(G) \leq n - \alpha(G) - 1$ for most values of n if $G \in G(n, 1/2)$. Recently, Alon, Bohman, and Huang [2] established a better upper bound which asserts if $G \in G(n, 1/2)$, then asymptotically almost surely $\tau(G) \leq n - (1 + c)\alpha(G)$ for some small positive constant c. This result implies that Erdős' conjecture is even false. For sparser random graphs, Alon [1] proved that there exists some (small) constant c such that for $\frac{2}{n} \leq p \leq c$, the bipartition number for a random graph G in $G(n, p)$ satisfies $\tau(G) = n - \Theta\left(\frac{\log np}{p}\right)$ asymptotically almost surely.

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 [9]. Therefore the problem of determining $\tau(G)$ is also NP-complete. Nevertheless, Theorem 2 implies 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. It is possible that a graph G dose not admit such a partition, then we define $\tau'(G)$ as ∞ in this case; 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 is a constant and $p \leq \frac{1}{2}$.

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 covered by several specified families of complete bipartite subgraphs. In Section 4, we consider the remaining uncovered edges and give corresponding lower bounds that our main theorem needs. In Section 5, we show that asymptotically almost surely the strong bipartition number is at least $1.0001n$ for a random graph on n vertices. In Section 6, we use the lemmas and the strong bipartition theorem to prove Theorem 2. A number of problems and remarks are mentioned in Section 7.

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 $\{u \colon u \in V(G) \text{ and } \{u, v\} \in E(G)\}$ 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 $|\{F \colon v \in F \text{ and } F \in E(H)\}|$. For $U \subseteq V(G)$, let $e(U)$ be the number 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 $E(A, B) = \{\{u, v\} \in E \colon u \in A \text{ and } v \in B\}$. 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 *[4] Let $X_1, \ldots, X_n$ be independent random variables with*

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

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

$$\begin{aligned} &\textit{(Lower tail)} && \Pr(X \leq \mathrm{E}(X) - \lambda) && \leq && e^{-\lambda^2/2\mathrm{E}(X)}, \\ &\textit{(Upper tail)} && \Pr(X \geq \mathrm{E}(X) + \lambda) && \leq && e^{-\frac{\lambda^2}{2(\mathrm{E}(X)+\lambda/3)}}. \end{aligned}$$

Theorem 4 *[3] 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 $t_1, \ldots, t_{i-1}, t'_i$:*

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

then

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

The following lemma on edge density will be useful later.

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

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

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 [5]. We include the statement and a short proof here for the sake of completeness.

Lemma 2 *For $G \in G(n,p)$, where p is a constant no greater than $1/2$, asymptotically almost surely all complete bipartite graphs $K_{A,B}$ in G with $|A| \leq |B|$ satisfy $|A| \leq 2\log_b n$ where $b = 1/p$.*

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 $k \geq 2\log_b n$, we have

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

as p is a constant. The lemma then follows. □

The upper bound in Theorem 2 is an immediate consequence of (3). 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 [7].

Theorem 5 *[7] If p is a constant, $p < 1 - c$, and $G \in G(n,p)$, then asymptotically almost surely $\alpha(G)$ is of order*

$$\alpha(G) = 2\log_{1/(1-p)} n + o(\log n)$$

where c is a positive constant .

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 $\mathcal{A} = \{A_1, A_2, \ldots, A_k\} \subseteq 2^V$ and σ, 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 σ. 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

$$f(G,\mathcal{A}) = \max_{\sigma} \max_{l} \sum_{i=1}^{k} |E(A_i, l(i))|.$$

Basically, for given A_i's, we wish to choose B_i's so that the complete bipartite graphs $K_{A_1,B_1}, \ldots, K_{A_k,B_k}$ cover as many edges in G as possible. An example is illustrated in Figure 1 for $\mathcal{A} = \{A_1, A_2, A_3\}$ with $A_1 = \{a,b\}, A_2 = \{b,c\}$, and $A_3 = \{c,d\}$. Here $f(G,\mathcal{A}) = 4$ is achieved by σ = 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 $f(G,\mathcal{A}) \leq \sum_{i=1}^{k} |E(A_i, V(G,A_i))|$ 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 $f(G,U,\mathcal{A})$ to denote $f(G[U],\mathcal{A})$. Note that $G[U]$ denotes the induced subgraph of G on a subset U of $V(G)$.

Lemma 3 *Suppose $G \in G(n,p)$, $U \subseteq V(G)$, and $\mathcal{A}$ is a family of 2-sets of U with $|\mathcal{A}| \leq 1.0001|U|$. If $p \leq \frac{1}{2}$, then asymptotically almost surely we have*

$$f(G,U,\mathcal{A}) \leq 2p^2|\mathcal{A}||U| + 8|U|^{3/2}\log n$$

for all choices of U and $\mathcal{A}$.

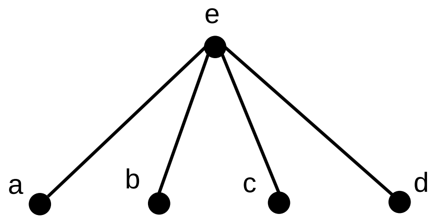

Figure 1: An illustration of $f(G, \mathcal{A})$.

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 $1 \le i \le m$, we consider $T_i \in \{\mathrm{H}, \mathrm{T}\}$ where $T_i = \mathrm{H}$ means e_i is an edge and $T_i = \mathrm{T}$ 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 $1 \le j \le i-1$ we wish to establish an upper bound for

$$|\mathrm{E}(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = \mathrm{H}) - \mathrm{E}(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = \mathrm{T})| . \tag{4}$$

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

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

and

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

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 μ is a bijection and $\Pr(K \in \mathcal{K}_1) = \Pr(\mu(K) \in \mathcal{K}_2)$. Therefore the expression (4) 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 $|f(K, \mathcal{A}) - f(\mu(K), \mathcal{A})| \le 2$ 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 (4) is bounded above by 2.

Now we apply Theorem 4 for $\lambda = 8|U|^{3/2} \log n$ and $c_i = 2$. Then we have

$$\Pr\left(|X - \mathrm{E}(X)| \ge 8|U|^{3/2} \log n\right) \le 2e^{-64|U|^3 \log^2 n / 2\sum_{i=1}^m c_i^2} \le 2e^{-4|U| \log^2 n}, \tag{5}$$

using the fact $m \le \frac{|U|^2}{2}$. To estimate $\mathrm{E}(X)$, we note that $\mathrm{E}(f(G, U, A)) \le 2p^2|U|$ for a fixed $A \in \mathcal{A}$. Therefore,

$$\mathrm{E}(X) \le \sum_{A \in \mathcal{A}} \mathrm{E}\left(f(G, U, A)\right) \le 2p^2 |\mathcal{A}||U|.$$

Thus (5) implies

$$\begin{aligned} \Pr\left(X \ge 2p^2|\mathcal{A}||U| + 8|U|^{3/2} \log n\right) &\le \Pr\left(|X - \mathrm{E}(X)| \ge 8|U|^{3/2} \log n\right) \\ &\le 2e^{-4|U| \log^2 n}. \end{aligned}$$

Recall the assumptions $|\mathcal{A}| \leq 1.0001|U|$ and $|A| = 2$ for each $A \in \mathcal{A}$. For fixed sizes of U and $\mathcal{A}$, the number of choices for U and $\mathcal{A}$ is at most $n^{|U|}|U|^{2|\mathcal{A}|}$ which is less than $n^{3.5|U|}$. Therefore the probability that there are some U and $\mathcal{A}$ which violate the assertion in the lemma is at most $1.0001|U| \times n \times n^{3.5|U|} \times 2e^{-4|U|\log^2 n} < 2e^{-\log^2 n}$ for sufficiently large n, which completes the proof of this lemma. □

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 $G \in G(n,p)$, $\mathcal{A}$ is a family of subsets of $U \subseteq V(G)$ satisfying $|\mathcal{A}| \leq 1.0001|U|$ and $2 \leq |A| \leq 2\log_2 n$ for each $A \in \mathcal{A}$. If $p \leq \frac{1}{2}$, then asymptotically almost surely we have*

$$f(G,U,\mathcal{A}) \leq 2p^2|\mathcal{A}||U| + 8|U|^{3/2}\log^{3/2} n \log^{1/2}|U|$$

for all choices of U and $\mathcal{A}$ satisfying $|U| \geq \log^2 n$.

Proof: We will use the Azuma's inequality. For $e_i = \{u_i, v_i\}$, we define $\mathcal{K}_1$, $\mathcal{K}_2$, and a bijection μ similarly. The only difference is that $|f(K,\mathcal{A}) - f(\mu(K),\mathcal{A})| \leq 2\log_2 n$ for each $K \in \mathcal{K}_1$. This is because e_i and at most other $2\log_2 n - 1$ edges sharing the same endpoint with e_i could by covered by $\mathcal{A}$ in $\mu(K)$ but not in K.

Therefore the corresponding expression for (4) can be upper bounded by $2\log_2 n$. We can then estimate $X = f(G,U,\mathcal{A})$ by applying Theorem 4 with

$$\lambda = 8|U|^{3/2}\log^{3/2} n \log^{1/2}|U| \text{ and } c_i = 2\log_2 n.$$

This leads to

$$\Pr\left(|X - \mathrm{E}(X)| \geq 8|U|^{3/2}\log^{3/2} n \log^{1/2}|U|\right) \leq 2e^{-4|U|\log n \log|U|}. \tag{6}$$

Since $\mathrm{E}(X) \leq \sum_{A\in\mathcal{A}} \mathrm{E}(f(G,U,A)) \leq 2p^2|\mathcal{A}||U|$ and the number of choices for U and $\mathcal{A}$ can be bounded from above by

$$1.0001|U| \times n^{1+|U|} \times |U|^{2.0003|U|\log_2 n} \leq e^{2.5|U|\log n \log|U|}.$$

Here we used the following simple fact $\sum_{1\leq t\leq k}\binom{s}{t} = (1+o(1)))\binom{s}{k}$ if k is much smaller than s. Therefore, the probability that there are some U and $\mathcal{A}$ which violate the lemma is at most $e^{-|U|\log n \log|U|} \leq e^{-\log^2 n}$, which completes the proof of the lemma. □

Lemma 5 *Suppose $G \in G(n,p)$, $\mathcal{A}$ is a family of subsets of $U \subseteq V(G)$ satisfying $|\mathcal{A}| \leq 1.0001|U|$ and $3 \leq |A| \leq 2\log_2 n$ for each $A \in \mathcal{A}$. If $p \leq \frac{1}{2}$, then asymptotically almost surely we have*

$$f(G,U,\mathcal{A}) \leq 3p^3|\mathcal{A}||U| + 8|U|^{3/2}\log^{3/2} n \log^{1/2}|U|$$

for all choices of U and $\mathcal{A}$ satisfying $|U| \geq \log^2 n$.

Proof: The proof is similar to that of Lemma 4. The only difference is that we assume $|A| \geq 3$ and therefore

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

We use Theorem 4 in a similar way as in the proof of Lemma 4 to complete the proof of Lemma 5. □

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

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

for all choices of U and $\mathcal{A}$ satisfying $|U| \geq \log^2 n$.

Proof: We use the assumptions on $|A|$ to derive

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

Then we bound the number of choices for U and $\mathcal{A}$ from above by

$$|U|^{1+\delta} \times n^{1+|U|} \times |U|^{2.01|U|^{1+\delta}\log_2 n} \leq e^{3.5|U|^{1+\delta}\log n \log |U|}.$$

Applying Theorem 4 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. □

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 $V(G)$ provided $G \in G(n,p)$. 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. We note that S and T are determined by $\mathcal{A}$. Furthermore, $|S| \geq |T|$.

Suppose $G \in G(n,p)$. For $u,v \in S$, let $X_{u,v}$ be the indicator random variable such that $\{u,v\} \in E(G)$, $\{u, A_v \setminus \{v\}\} \notin E(G)$, and $\{v, A_u \setminus \{u\}\} \notin E(G)$. Then we define

$$g(G,\mathcal{A},S,T) = \sum_{\{u,v\}\in\binom{S}{2}} X_{u,v}$$

We observe that $g(G,\mathcal{A},S,T)$ indeed gives a lower bound on the number of edges which are not covered by $\mathcal{A}$. Since if $X_{u,v} = 1$ and $\{u,v\}$ is covered by some $A \in \mathcal{A}$, then either $A = A_u$ or $A = A_v$. The former case can not happen because we assume $\{v, A_u \setminus \{u\}\}$ is not an edge. We have the similar argument for the latter case.

Lemma 7 *Suppose $G \in G(n,p)$, $U \subseteq V(G)$, and $\mathcal{A} \subseteq \binom{U}{2}$ with $|\mathcal{A}| \leq 1.0001|U|$. Let S and T be defined as above. If $p \leq \frac{1}{2}$, then asymptotically almost surely we have*

$$g(G,\mathcal{A},S,T) \geq p^3\binom{|S|}{2} - 4|U|^{3/2}\log n$$

for all choices of U and $\mathcal{A}$ with $|U| \geq \log^2 n$.

Proof: We sketch the proof here which is similar to that of Lemma 3. We list edges with endpoints in $S \cup T$ as $e_1, \ldots, e_m$, where $m = \binom{|S \cup T|}{2}$ and $m \leq \binom{|U|}{2}$. For each $e_i = \{u_i, v_i\}$ for $1 \leq i \leq m$, we consider $T_i \in \{\mathrm{H}, \mathrm{T}\}$ where $T_i = \mathrm{H}$ means e_i is an edge and $T_i = \mathrm{T}$ means e_i is not an edge. Let X denote the random variable $g(G, \mathcal{A}, S, T)$ 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

$$|\mathrm{E}(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = \text{ H}) - \mathrm{E}(X|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}, T_i = \text{ T})|\,. \quad (7)$$

For $e_i = \{u_i, v_i\}$, if $u_i, v_i \in T$ then the outcome of T_i does not contribute to (7). If $u_i, v_i \in S$ then the outcome of T_i can change (7) 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 (7) 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 (7) is bounded above by one.

Applying Azuma's theorem as stated in Theorem 4 with $\lambda = 4|U|^{3/2}\log n$ and $c_i = 1$, we have

$$\Pr\left(|X - \mathrm{E}(X)| \geq 4|U|^{3/2}\log n\right) \leq 2e^{-4|U|\log^2 n}, \quad (8)$$

using $m \leq \binom{|U|}{2}$. To estimate $\mathrm{E}(X)$, we note $\Pr(X_{u,v} = 1) = p(1-p)^2 \geq p^3$ as $p \leq \frac{1}{2}$. Thus,

$$\mathrm{E}(X) = \sum_{\{u,v\} \in \binom{S}{2}} \Pr(X_{u,v} = 1) \geq p^3 \binom{|S|}{2},$$

which implies

$$\Pr\left(X \leq p^3\binom{|S|}{2} - 4|U|^{3/2}\log n\right) \leq \Pr\left(X \leq \mathrm{E}(X) - 4|U|^{3/2}\log n\right) \leq 2e^{-4|U|\log^2 n}.$$

We recall S and T are determined by $\mathcal{A}$. For fixed sizes of U and $\mathcal{A}$, the number of choices for U and $\mathcal{A}$ is at most

$$n^{|U|}|U|^{2|\mathcal{A}|}.$$

As $|\mathcal{A}| \leq 1.0001|U|$, the probability that there are some U and $\mathcal{A}$ which violate this lemma is at most

$$1.0001|U| \times n^{1+|U|} \times |U|^{2.0003|U|} \times 2e^{-4|U|\log^2 n} \leq e^{-0.5|U|\log^2 n} < e^{-\log^2 n}.$$

The lemma is proved. □

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 $L\colon W \to 2^{U \setminus W}$ 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 (later we will show that $h(G, U, W, L)$ gives a lower bound for the number of uncovered edges).

Lemma 8 *Suppose $G \in G(n,p)$, p is a constant no greater than $1/2$, $b = \frac{1}{p}$, and $|U| \geq \log^2 n$. Assume $W \subset U$ satisfies $|W| = |U|/\log_b^2|U|$ and L as defined above satisfies $1 \leq |L(w)| \leq c\log_b|U|$ for some positive constant c. Then asymptotically almost surely we have*

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

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 indicator random 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

$$\mathrm{E}(Y) \geq b^{-1-2c \log_b |U|} \binom{|W|}{2} \geq c'|U|^{2-2c}/\log_b^4 |U|,$$

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 3 with $\lambda = 2|U|^{3/2}\sqrt{\log n}$, we have

$$\begin{aligned}\Pr(Y \leq c'|U|^{2-2c}/\log_b^4 |U| - 2|U|^{3/2}\sqrt{\log n}) &\leq \Pr(Y \leq \mathrm{E}(Y) - 2|U|^{3/2}\sqrt{\log n}) \\ &\leq e^{-\frac{4|U|^3 \log n}{2\mathrm{E}(Y)}} \leq e^{-2|U| \log n},\end{aligned}$$

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

$$n^{|U|}|U|^{|U|/\log_b^2 |U|}|U|^{1.1c|U|/\log_b |U|} < n^{|U|}|U|^{2c|U|/\log_b |U|} \leq n^{1.5|U|},$$

when n is sufficiently large. The probability that there is some U, W and L which violate the lemma is at most $ne^{-2|U| \log n} n^{1.5|U|} < e^{-\log^2 n}$ 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 fix the constant $b = \frac{1}{p}$ and we will prove the following theorem.

Theorem 6 *Suppose $G \in G(n, p)$, $U \subseteq V(G)$ is a vertex subset with $|U| \geq c(\log_b n)^{3+\epsilon}$ where $b = 1/p$, c and ϵ are positive constants. For p being a positive constant no greater than $1/2$, asymptotically almost surely for all U we have*

$$\tau'(G[U]) \geq 1.0001|U|.$$

The proof of Theorem 6 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 1, the number of edges in $G[U]$ satisfies $e(G[U]) = (\frac{p}{2} + o(1))u^2$, here $|U| = u$. We will prove Theorem 6 by contradiction. Suppose

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

and $m < \left(1 + \frac{1}{10000}\right) u$ where '$\sqcup$' denotes the disjoint union. We assume further $|A_i| \leq |B_i|$ for each $1 \leq i \leq m$. Lemma 2 implies $|A_i| \leq 2 \log_b n \leq 2 \log_2 n$ for each $1 \leq i \leq m$.

We define $\mathcal{L} = \{A_i : 1 \leq i \leq m\}$. We consider three subsets of $\mathcal{L}$ defined as follows:

$$\begin{aligned}\mathcal{L}_1 &= \{A_i \in \mathcal{L} : |A_i| < \delta_1 \log_b u\} \\ \mathcal{L}_2 &= \{A_i \in \mathcal{L} : |A_i| < \delta_2 \log_b u\} \\ \mathcal{L}_3 &= \{A_i \in \mathcal{L} : |A_i| = 2\},\end{aligned}$$

where

$$\delta_1 = \min\left\{\frac{\epsilon}{4(3+\epsilon)}, \frac{1}{200}\right\} \text{ and } \delta_2 = \frac{\delta_1}{10^4}.$$

We observe that a typical complete bipartite graph $K_{2,B}$ in a random graph $G(n, 1/2)$ roughly contains the same the number of edges as a typical star does. This is the reason why we define the set $\mathcal{L}_3$.

We have the following lemma.

Lemma 9 *If* $|\mathcal{L}_2| \leq \left(\frac{1}{2} + \frac{1}{1500}\right) u$*, then we have* $|\mathcal{L}_3| \geq \left(\frac{1}{2} + \frac{1}{2000}\right) u$*.*

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 3, the number of edges covered by $\mathcal{L}_3$ is at most

$$2p^2|\mathcal{L}_3|u + 8u^{3/2}\log^{1/2} n.$$

By Lemma 5, the number of edges covered by $\mathcal{L}_2 \setminus \mathcal{L}_3$ (i.e., $|A_i| \geq 3$) is at most

$$3p^3|\mathcal{L}_2 \setminus \mathcal{L}_3|u + 8u^{3/2}\log^{3/2} n \log^{1/2} u.$$

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

$$2p^2|\mathcal{L}_3|u + 3p^3|\mathcal{L}_2 \setminus \mathcal{L}_3|u + 8u^{3/2}\log^{1/2} n + 8u^{3/2}\log^{3/2} n \log^{1/2} u.$$

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

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

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

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

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

Applying Lemma 6 with $\delta = \delta_2$, the number of edges covered by $\mathcal{L} \setminus \mathcal{L}_2$ (i.e., $|A_i| \geq \delta_2 \log_b n$) is at most

$$\delta_2|\mathcal{L} \setminus \mathcal{L}_2|u^{1-\delta_2}\log_b u + 8u^{(3+\delta_2)/2}\log^{3/2} n \log^{1/2} u.$$

Since $u^{(3+\delta_2)/2}\log^{3/2} n \log^{1/2} u = o(u^2)$ by the choice of δ_2, in order to cover the remaining edges, we need at least $C_1u^{1+\delta_2/2}$ extra complete bipartite graphs for some positive constant C_1, i.e., $|\mathcal{L} \setminus \mathcal{L}_2| \geq C_1u^{1+\delta_2/2}$. Since $C_1u^{1+\delta_2/2} > \left(1 + \frac{1}{10000}\right) u$ for sufficiently large n, this leads to a contradiction. Thus we have $|\mathcal{L}_3| \geq \left(\frac{1}{2} - \frac{1}{250}\right)$ and the claim is proved.

Now we proceed to prove the lemma using the fact that $|\mathcal{L}_3| \geq \left(\frac{1}{2} - \frac{1}{250}\right) u$. We consider a auxiliary graph U^* whose vertex set is U and edge set is $\mathcal{L}_3$. We partition the vertex set of U^* into three sets U_1, U_2, and U_3, where $U_1 = \{v \in V(U^*) \colon d_{U^*}(v) = 0\}$, $U_2 = \{v \in V(U^*) \colon d_{U^*}(v) = 1\}$, and $U_3 = \{v \in V(U^*) \colon d_{U^*}(v) \geq 2\}$. We will prove the following.
Claim 2: The number of edges not contained in any K_{A_i,B_i} for $A_i \in \mathcal{L}_3$ is at least $p\binom{|U_1|}{2} + p^3\binom{|U_2|/2}{2} + o(u^2)$.

Proof of Claim 2: The first part of the sum follows from Lemma 1. For the second part of the sum, we let U_2' be a maximum subset of 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 7 with $S = U_2'$ and T consisting of neighbors of S' in U^*. To finish the proof of Claim 2, we use the fact $u^{3/2} \log n = o(u^2)$ as $u \geq c(\log_b n)^{3+\epsilon}$.

We will prove Lemma 9 by contradiction. Suppose $|\mathcal{L}_3| \leq \left(\frac{1}{2} + \frac{1}{2000}\right) u$. This implies that the average degree of U^* is at most $1 + \frac{1}{1000}$. We consider the following cases.

Case 1: $|U_3| \geq \frac{1}{5}u$.

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, $|U_3| \geq \frac{1}{5}u$ implies $|U_1| \geq \frac{1}{6}u$. 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 $A_i \in \mathcal{L}_3$ is at least $\left(\frac{p}{72} + o(1)\right) u^2$. By Claim 1 we have $|\mathcal{L}_3| \geq \left(\frac{1}{2} - \frac{1}{250}\right) u$, so the number of additional complete bipartite graphs K_{A_i,B_i} with $A_i \in \mathcal{L}_2 \setminus \mathcal{L}_3$ is at most $\frac{7}{1500}u$ using the assumption $|\mathcal{L}_3| \leq \left(\frac{1}{2} + \frac{1}{2000}\right) u$. These complete bipartite graphs can cover at most $(\frac{7p^3}{500} + o(1))u^2$ edges by Lemma 5. Thus we conclude that the number of edges not covered by any of K_{A_i,B_i} with $A_i \in \mathcal{L}_2$ 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 is constant and $p \leq \frac{1}{2}$. By applying Lemma 6 with $\delta = \delta_2$, the bipartite graphs K_{A_i,B_i} with $A_i \in \mathcal{L} \setminus \mathcal{L}_2$ (i.e., $|A_i| \geq \delta_2 \log_b n$) can cover at most

$$\delta_2 |\mathcal{L} \setminus \mathcal{L}_2| u^{1-\delta_2} \log_b u + 8u^{(3+\delta_2)/2} \log^{3/2} n \log^{1/2} u$$

edges. We note $u^{(3+\delta_2)/2} \log^{3/2} n \log^{1/2} u = o(u^2)$ because of the choice of δ_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 $A_i \in \mathcal{L} \setminus \mathcal{L}_2$ for some positive constant C_1', i.e., $|\mathcal{L} \setminus \mathcal{L}_2| \geq C_1' u^{1+\delta_2/2}$. Since $C_1' u^{1+\delta_2/2} > \left(1 + \frac{1}{10000}\right) u$ for n large enough, we get a contradiction to the assumption $|\mathcal{L}| \leq \left(1 + \frac{1}{10000}\right) u$.

Case 2: $|U_3| < \frac{1}{5}u$.

In this case we have $|U_1| + |U_2| \geq \frac{4}{5}u$. Note that the lower bound given by Claim 2 is minimized when $|U_2| = \frac{4}{5}u$, 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 $\left(\frac{2}{25}p^3 + o(1)\right) u^2$. 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 $(\frac{7p^3}{500} + o(1))u^2$. Now there are at least

$$\left(\frac{33}{500}p^3 + o(1)\right) u^2$$

edges which is not in any K_{A_i,B_i} with $A_i \in \mathcal{L}_2$. We note that $\frac{33}{500}p^3$ is a positive constant as we assume p is a constant. By using Lemma 6 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 $|\mathcal{L}_3| > (\frac{1}{2} + \frac{1}{2000})$. $\square$

Remark: When we apply Lemma 6, we require the error term is in a lower order of magnitude in comparison to the main term. To make this satisfied, we have to assume $|U| = \Omega((\log n)^{3+\epsilon})$.

Lemma 10 *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 $C_2 u$ vertices of H with degree less than $\left(\frac{\delta_1}{2} - \frac{\delta_1}{3000}\right) \log_b u$.*

Proof: We consider several cases.

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

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 $|\mathcal{L}_1| \leq |\mathcal{L}| = m < \left(1 + \frac{1}{10000}\right) u$ and the choice of δ_2.

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

By Lemma 9, $|\mathcal{L}_3| \geq \left(\frac{1}{2} + \frac{1}{2000}\right) u$. 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 $\left(\frac{\delta_1}{2} - \frac{\delta_1}{2000}\right) u \log_b u$. Let U' be the set of vertices with degree at least $\left(\frac{\delta_1}{2} - \frac{\delta_1}{3000}\right) \log_b u$. 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 $\left(\frac{\delta_1}{2} - \frac{\delta_1}{3000}\right) \log_b u$ and $|U \setminus U'| \geq C_2 u$. The lemma is proved. $\square$

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

Lemma 11 *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 $C_3 u^{2-\delta_1+\delta_1/2000}$ 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 10. Let W be the set of vertices with degree less than $\left(\frac{\delta_1}{2} - \frac{\delta_1}{3000}\right) \log_b u$ in H; we have $|W| \geq C_2 u$ for some positive constant C_2 by Lemma 10.

We will use Lemma 8 to prove Lemma 11. In order to apply Lemma 8, 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 $q = |W| / \log_b^2 u$, $W_0 = W$, $Z_0 = \emptyset$ and $H_0 = H$. For each $1 \leq i \leq q$, we recursively define a vertex v_i, a set W_i, a set Z_i 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 $F(v_i) = \{A \in E(H_{i-1}) \colon v_i \in A\}$. By the assumption on the size of sets in $\mathcal{L}_1$ and the degree upper bound for vertices in W, we have $|\cup_{A\in F(v_i)} A| \leq \log_b^2 u/2$. We define $Z_i = \{A \in E(H_{i-1})\colon |A \setminus \left(\cup_{A'\in F(v_i)}A'\right)| = 1\}$. Then $|\cup_{A\in Z_i} A \setminus \left(\cup_{A'\in F(v_i)}A'\right)| \leq \log_b^2 u/2$ since each $A' \in F(v_i)$ can contribute at most $\delta_1 \log_b u$ to the sum and $|F(v_i)| \leq \frac{\delta_1}{2}\log_b u$ because of the degree upper bound for vertices in W. We define $W_i = W_{i-1} \setminus \left(\cup_{A\in Z_i\cup F(v_i)}A\right)$ and H_i to be the new hypergraph with the vertex set $V(H_{i-1}) \setminus \left(\cup_{A\in Z_i\cup F(v_i)}A\right)$. If $A \in E(H_{i-1})$ then $A \setminus \left(\cup_{A\in Z_i\cup F(v_i)}A\right) \in E(H_i)$. Thus $|V(H_i)| = |W_i| \geq |W_{i-1}| - \log_b^2 u$ and $W_{q-1} \neq \emptyset$. 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 $\cup_{A\in Z_i\cup F(v_i)}A$ 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 $F'(v_i) = \cup_{A\in F(v_i)}\{f(A)\}$. It follows from the preceding definitions that $F'(v_i) \cap F'(v_j) = \emptyset$ for $1 \leq i \neq j \leq q$. Now an application of Lemma 8 with $U = V(H)$, $W = W'$, $L(v_i) = F'(v_i)$ for each v_i and $c = \frac{\delta_1}{2} - \frac{\delta_1}{3000}$ will prove the lemma.

We are left to verify the function $h(G, U, W, L)$ indeed gives a lower bound on the number of edges which is not covered by $\mathcal{L}_1$. From the construction , 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 $z \in F'(v_j)$ 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 latter case. □

Remark: We mention here that when we defined the set $W' \subset W$, we did not aim to find the largest one as $|W'| = |W|/\log_b^2 n$ is large enough for proving Theorem 6.

We are ready to prove Theorem 6.

Proof of Theorem 6: Suppose that

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

If $m > \left(1 + \frac{1}{10000}\right)u$, then we are done. Otherwise, Lemma 11 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 6 with $\delta = \delta_1$ which gives an upper bound for the number of edges covered by $\mathcal{L} \setminus \mathcal{L}_1$ (i.e., $|A_i| \geq \delta_1 \log_b u$) :

$$\delta_1 |\mathcal{L} \setminus \mathcal{L}_1| u^{1-\delta_1}\log_b u + 8u^{(3+\delta_1)/2}\log^{3/2} n \log^{1/2} u.$$

Here we note that $u^{(3+\delta_1)/2}\log^{3/2} n \log^{1/2} u = o(u^{2-\delta_1+\delta_1/2000})$ because of the choice of δ_1. Therefore we need at least $C_4 u^{1+\delta_1/2500}$ 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 $C_4 u^{1+\delta_1/2500} > 1.0001u$ when n is sufficiently large and we get a contradiction. Theorem 6 is proved. □

6 Proof of Theorem 2

Before proving Theorem 2, we first state the following lemma. The proof will be omitted as it is very simple.

Lemma 12 *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 $E(G) = \sqcup_{i=1}^{k} E(K_{A_i,B_i})$ 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 $A_j, B_j \subseteq V(G) \setminus \cup_{i=1}^{k_2} A_i$.*

We are ready to prove Theorem 2.

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

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

Therefore we have $\tau(G) \le n - (2+o(1))\log_{1/(1-p)} n$.

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:

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

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

Let $W = \{v_1, \ldots, v_l\}$. If $W = \emptyset$ then Theorem 2 follows from Theorem 6 directly. We need only to consider the case $W \ne \emptyset$. By Lemma 12, we can assume $E(G') = \sqcup_{i=l+1}^{k} E(K_{A_i,B_i})$ where G' is the subgraph induced by $T = V(G) \setminus W$. We get

$$\tau(G) = |W| + \tau'(G'). \tag{9}$$

We will prove $l > n - c(\log_{1/p} n)^{3+\epsilon}$ for any positive constants c and ϵ. Suppose $l \le n - c(\log_{1/p} n)^{3+\epsilon}$ for some c and ϵ. Thus, $|T| \ge c(\log_{1/p} n)^{3+\epsilon}$. By Theorem 6 we have $\tau'(G[T]) \ge \left(1 + \frac{1}{10000}\right)|T|$. Therefore

$$\tau(G) = |W| + \tau'(G') \ge |W| + \left(1 + \tfrac{1}{10000}\right)|T| \ge n,$$

which is a contradiction. Theorem 2 is proved. $\square$

7 Problems and remarks

The results on the bipartite decomposition in this paper lead to many questions, several of which we mention here.

For $G \in G(n, 1/2)$, the lower bound $\tau(G) \ge n - o((\log n)^{3+\epsilon})$ for any positive constant ϵ is given by Theorem 1 in this paper. For the upper bound, the result from [2] gives asymptotically almost surely $\tau(G) \le n - (1+c)\alpha(G)$ for some positive constant c. Similar upper bound is not known for any constant p with $p < 1/2$. We believe the following conjecture is true.

Conjecture 1: For a random graph $G \in G(n,p)$, where p is a constant and $p < 1/2$, asymptotically almost surely we have $\tau(G) = n - (2+o(1))\log_{1/(1-p)} n$.

For sparser random graphs, Alon [1] showed that there exists some (small) constant c such that for $\frac{2}{n} \le p \le c$, a random graph G in $G(n,p)$ satisfies $\tau(G) = n - \Theta\left(\frac{\log np}{p}\right)$. It will be of interest to further sharpen the lower bound.

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

$$\tau(G) = n - (1+o(1))\tfrac{2}{p}\log np.$$

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 $G \in G(n,p)$, where p is a constant no greater than $1/2$, asymptotically almost surely $\tau'(G) \geq 1.0001n$. For the case of $p \leq c$ for some small c, Alon [1] showed that asymptotically almost surely $\tau'(G) \geq 2n$ for $G \in G(n,p)$. A natural question is to improve the lower bound for $\tau'(G)$.

References

[1] N. Alon, Bipartite decomposition of random graphs, *J. Combin. Theory Ser. B*, **113** (2015), 220–235.

[2] N. Alon, T. Bohman, and H. Huang, More on the bipartite decomposition of random graphs, arXiv: 1409.6165[math.CO].

[3] K. Azuma, Weighted sums of certain dependent random variables, *Tohuku Math. Journal*, **19**(3) (1967), 357–367.

[4] H. Chernoff, A note on an inequality involving the normal distribution, *Ann. Probab.*, **9** (1981), 533–535.

[5] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.*, **53** (1947), 292–294.

[6] R. L. Graham and H. O. Pollak, On the addressing problem for loop switching, *Bell Syst. Tech. J.*, **50** (8) (1971), 2495–2519.

[7] G. R. Grimmett and C. J. H. McDirmid, On colouring random graphs, *Math. Proc. Cambridge Phil. Soc.*, **77** (1975), 313–324.

[8] R. Karp, Reducibility among combinatorial problems, *Complexity of Computer Computations* (R. E. Miller, J. W. Thatcher eds.) Plenum, New York, (1972), 85–103.

[9] T. Kratzke, B. Reznick and D. West, Eigensharp graphs: Decomposition into complete bipartite subgraphs, *Trans. Amer. Math. Soc.* , **308** (1988), 637–653.

[10] L. Lovász, On coverings of graphs, *Theory of Graphs (Proc. Conf. Tihany)*, Academic Press, (1969), 231–236.

[11] G. W. Peck, A new proof of a theorem of Graham and Pollak, *Discrete Math.*, **49** (1984), 327–328.

[12] H. Tverberg, On the decomposition of K_n into complete bipartite graphs, *J. Graph Theory*, **6** (1982), 493–494.

[13] S. Vishwanathan, A polynomial space proof of the Graham-Pollak theorem, *J. Combin. Theory Ser. A*, **115** (2008), 674–676.

[14] S. Vishwanathan, A counting proof of the Graham Pollak Theorem, *Discrete Math.*, **313**(6) (2013), 765–766.

[15] W. Yan and Y. Ye, A simple proof of Graham and Pollak's theorem, *J. Combin. Theory Ser. A,* **113**, (2006), 892–893.