Extremal graphs with local covering conditions

Debsoumya Chakraborti∗ and Po-Shen Loh†

September 12, 2019

Abstract

We systematically study a natural problem in extremal graph theory, to minimize the number of edges in a graph with a fixed number of vertices, subject to a certain local condition: each vertex must be in a copy of a fixed graph $H$. We completely solve this problem when $H$ is a clique, as well as more generally when $H$ is any regular graph with degree at least about half its number of vertices. We also characterize the extremal graphs when $H$ is an Erdős-Rényi random graph. The extremal structures turn out to have the same form as the conjectured extremal structures for a well-studied but elusive problem of similar flavor with local constraints: to maximize the number of copies of a fixed clique in graphs in which all degrees have a fixed upper bound.

1 Introduction

1.1 Problem and motivation

Extremal graph theory considers problems of maximizing or minimizing certain graph parameters in graph classes of interest. The most classical example is the Turán problem of maximizing the number of edges in a graph with a given number of vertices, subject to the condition of being $H$-free (having no subgraph isomorphic to a fixed graph $H$). Turán solved the problem completely when $H$ is a complete graph in [14], and for general $H$ with chromatic number at least 3, the Erdős-Stone-Simonovits Theorem [6] provides an asymptotic answer. When $H$ has chromatic number 2, the problem is more intricate, with many longstanding open questions remaining (see, e.g., the surveys [9] and [13]).

A variety of questions of similar flavor have been the focus of significant research attention. When one generalizes the problem to maximizing the number of cliques of order $t \geq 3$ in an $n$-vertex $H$-free graph, even the basic question where $H = K_{1,D}$ is a star (translating into a maximum degree condition of $D-1$) is not completely understood. Indeed, if $n = qD + r$ ($q \in \mathbb{N}$ and $1 \leq r \leq D$), the maximum number of cliques $K_t$ is conjectured to be obtained by the disjoint union of $q$ many $K_D$ and one $K_r$. Although for $q = 1$ this conjecture is resolved

∗Department of Mathematical Sciences, Carnegie Mellon University. Email: dchakrab@cmu.edu. Research supported in part by National Science Foundation CAREER Grant DMS-1455125.
†Department of Mathematical Sciences, Carnegie Mellon University. Email: ploh@cmu.edu. Research supported in part by National Science Foundation CAREER Grant DMS-1455125.
in [10], the general problem is still wide open despite substantial effort (see, eg., [3]). It appears to be difficult to prove extremality of this type of structure, with many disjoint copies of the same clique, and one residual graph which depends on the residue class of \( n \).

An equivalent form of the above conjecture was asked by Engbers and Galvin in [4]. In this equivalent form, we seek the maximum number of independent sets of order \( t \) in an \( n \)-vertex graph with minimum degree at least \( \delta \). It is easy to see the equivalence between this statement and the aforementioned conjecture by considering the complement of a graph. This is now the same as asking for the maximum number of independent sets of order \( t \) in an \( n \)-vertex graph where every vertex \( v \) is in a copy of \( K_{1,\delta} \) with \( v \) being the root of \( K_{1,\delta} \). It is natural to consider the same problem with \( K_{1,\delta} \) replaced by a general graph \( H \). In this paper, we will consider a weaker condition where we do not require that each vertex \( v \) in \( G \) is a specific vertex in a copy of \( H \). Note that this does not change the problem when \( H \) is a complete graph. It turns out that even the case \( t = 2 \) for this problem is non-trivial and has some surprising results, which will be the main topic of this article. Note that maximizing the number of independent sets of order 2 is equivalent to minimizing the number of edges in a graph.

On the topic of \( H \)-free graphs, the question of minimizing the number of edges becomes interesting when one adds the further property that adding one more edge to \( G \) will create a copy of \( H \). This motivates the definition of \( \text{sat}(n, H) \), which is the minimum number of edges in an \( n \)-vertex \( H \)-free graph \( G \), with that property. Erdős, Hajnal, and Moon [5] answered this question when \( H \) is a graph on \( p \) vertices then \( \text{sat}(n, H) \leq \text{sat}(n, K_p) = O(n) \), but it is not even known if \( \lim_{n \to \infty} \frac{\text{sat}(n,H)}{n} \) exists for every graph \( H \) (Tuza’s conjecture [15]). This class of problems is extensively surveyed in [7].

In this paper, we consider a natural minimization problem which has similar flavor. We study the problem of minimizing the number of edges (or more generally, copies of \( K_t \)) in a graph \( G \) with \( n = qD + r \) vertices, subject to the condition that it is \( H \)-covered by some fixed graph \( H \), i.e., each vertex in \( G \) is in a copy of \( H \). We completely solve the problem of determining the unique \( K_D \)-covered graph with any given number of vertices that minimizes the number of cliques of order \( t \). It turns out that the edge-minimal graph has almost the same structure as in the conjecture in [10], mentioned at the beginning of this paper. This time, the residual graph is not \( K_r \), but instead is a union of two copies of \( K_D \) overlapping in \( D - r \) vertices, as stated formally below.

### 1.2 Main results

**Proposition 1.1.** For any positive integers \( q, n, t \) with \( 2 \leq t \leq n \), and any integer \( N = qn + r \) with \( 0 \leq r < n \), the graph consisting of the union of 2 copies of \( K_n \) sharing \( n - r \) vertices, together with the disjoint union of \( q - 1 \) many \( K_n \), has the least number of copies of \( K_t \) among all \( K_n \)-covered graphs on \( N \) vertices. Moreover, this is the unique such graph.

The first natural question to ask is the following. Are there other graphs \( H \) that give rise to the same structure, with many disjoint copies of \( H \), and one residual graph? We first show a simple example of a graph which does not. Consider the graph \( H \) consisting of the union of \( K_n \) and \( K_2 \) overlapping in one vertex, i.e., \( K_n \) with one pendant vertex. A moment’s reflection will convince the reader that an \( H \)-covered graph \( G \) is edge-minimal if and only
if $G$ consists of a copy of $K_n$, together with $N - n$ more edges which connect the rest of the $N - n$ vertices to the $K_n$ with a single edge each, i.e., $K_n$ with $N - n$ pendant vertices. One can also prove this by using Theorem 1.3 which will be stated towards the end of this introduction. In order to generalize Proposition 1.1 to other graphs, we need a couple of definitions to describe the residual “last piece” (the analogue of the two overlapping copies of $K_n$).

**Definition 1.** Suppose $H$ and $T$ are graphs with $A \subseteq V(H)$ where $V(H)$ denotes the vertex set of $H$. Let $a^H_T(A)$ denote the number of subgraphs isomorphic to $T$ in $H$ which have at least one vertex in $A$. Also, let $a^H_T(k)$ denote the minimum value of $a^H_T(A)$ over all $k$-vertex subsets $A \subseteq V(H)$. For convenience, we define $a^H_T(0) = 0$. When $H$ is understood from the context, we will drop $H$ from the notation. Also, when $T = K_2$, we will drop $T$ from the notation.

When generalizing the “last piece” of Proposition 1.1 it turns out that there may be several options. Intuitively the last piece is just 2 copies of $H$ overlapping in the most efficient way to minimize the total number of edges. We write $\sqcup$ to denote disjoint union throughout this paper.

**Definition 2.** Suppose $T$ is a complete graph, $H$ is a graph with $n$ vertices and $N = n + r$ with $0 \leq r < n$. Define $L^H_{T,N}$ to be the collection of graphs $G$ on the vertex set $V \sqcup V_1 \sqcup V_2$ with $|V| = n - r$ and $|V_1| = |V_2| = r$ satisfying the following. There are no edges between $V_1$ and $V_2$. The set $V$ induces a subgraph of $H$ with $n - r$ vertices which has the maximum possible number of subgraphs isomorphic to $T$, and $V_1 \sqcup V$ and $V_2 \sqcup V$ both induce $H$. When $T = K_2$, we will drop $T$ from the notation.

Note that in Definition 2 the number of subgraphs isomorphic to $T$ in $V$ is $a^H_T(n) - a^H_T(r)$. Hence the number of subgraphs isomorphic to $T$ in $L^H_{T,N}$ is $a^H_T(n) + a^H_T(r)$. Observe that when $H = K_n$, $L^H_n$ is just the last piece of the graph in Figure 1. We now define a notion that intuitively represents any graph which gives rise to similar structures to what the complete graph gave for the edge minimization problem of $H$-covered graphs.

**Definition 3.** Suppose $H$ is a graph with $n$ vertices and $r$ is an integer with $0 \leq r < n$. We call a graph $H$ **ideal in remainder class $r$** if for all $N = qn + r$ with $q \in \mathbb{Z}^+$, a graph $G$ is edge-minimal among $N$-vertex $H$-covered graphs if and only if $G$ consists of a graph in $L^H_{n+r}$, together with the disjoint union of $q - 1$ copies of $H$. For any subset $S \subseteq \{0, 1, \ldots, n - 1\}$, we call a graph $H$ **ideal in remainder class $S$** if it is ideal in remainder class $r$ for all $r \in S$. We call a graph $H$ **ideal** if it is ideal in remainder class $\{0, 1, \ldots, n - 1\}$.
Proposition [1.1] states that $K_n$ is an ideal graph. There are, however, graphs that are not ideal, as we saw before. As a more general example, we can show that for any graph $H$ whose two lowest-degree vertices have degrees $d_1$ and $d_2$ with $d_1 \geq d_2 + 2$, this type of structure does not minimize the number of edges. For such a graph $H$, any graph in $\mathcal{L}^H_{n+2}$ has at least $e(H) + d_1 + d_2 - 1$ many edges, where $e(H)$ denotes the number of edges in $H$. On the other hand, consider the graph $M$ created by the union of a copy of $H$ and 2 more vertices which mimic the minimum degree in $H$, i.e., these 2 vertices are adjacent to a vertex $v$ of $H$ if and only if $v$ is adjacent to the vertex with minimum degree $d_1$. It can easily be seen that $M$ is $H$-covered and has $e(H) + 2d_1$ many edges, which is less than the number of edges in graphs in $\mathcal{L}^H_{n+2}$. Hence for $N = qn + 2$ with $q \in \mathbb{N}$, we will have strictly less edges if we use $M$ as the last piece instead of any graph in $\mathcal{L}^H_{n+2}$, so $H$ is not ideal in remainder class $\{2\}$. Also, for a similar reason, if the two lowest-degree vertices have degrees $d_2 = d_1 + 1$, then the structure in Definition 3 may not always be the unique graph for our problem. Hence it is quite natural to study regular graphs, and we have the following theorem.

**Theorem 1.2.** Every $n$-vertex $d$-regular graph with $d \geq \frac{n-1}{2}$ is ideal. This is tight in a sense that there exist non-ideal $d$-regular $n$-vertex graphs when $d$ is the largest even integer less than $\frac{n-1}{2}$.

The structures for the edge minimization problem with respect to non-regular graphs $H$ can be very different in nature, and it seems that there is no systematic way to generalize and classify them. But surprisingly, we can characterize the problem for random graphs $H$ quite well. We consider the Erdős-Rényi random graph model $G_{n,p}$, for constant $p$, where each potential edge among $n$ vertices independently appears with probability $p$. Throughout this paper, we say that event $A_n$ occurs **with high probability** if the probability of that event ($\mathbb{P}[A_n]$ in notation) tends to one as $n \to \infty$. Although $G_{n,p}$ is not ideal (in particular $G_{n,p}$ is not ideal in remainder class $\{2\}$ because of the well known fact that with high probability the two lowest degrees differ by much more than 2), it turns out that with high probability $G_{n,p}$ is ideal in large remainder classes, and we can get another interesting kind of structure for small remainder classes. Motivated by the structure that appeared when the difference between the two lowest degrees of $H$ was 2, we define a couple of notions to help study the the problem for small remainder classes when $H$ is a random graph. Throughout the paper, we use the standard notation $[n]$ to represent the set $\{1, 2, \ldots, n\}$.

**Definition 4.** Suppose $H$ is a graph with $n$ vertices and $n \leq N < 2n$. Define $\mathcal{M}^H_N$ to be the collection of graphs $G$ satisfying the following: $G$ is a graph on the vertex set $[N]$, where for all $n \leq i \leq N$, the vertex subset $[n-1] \cup \{i\}$ induces a copy of $H$ and the number of edges in $G$ is the minimum possible.

In Definition 3 of $\mathcal{M}^H_N$, clearly there is no edge in $G$ between $i, j$ for all $n \leq i < j \leq N$, and for all $n \leq i \leq N$, the degree of $i$ in $G$ is the minimum degree in $H$. Note that $\mathcal{M}^H_n = \{H\}$. In Figure 2 $H$ is a graph on vertex set $[n]$ ($n \geq 3$) with $n$ being a vertex with minimum degree. We obtained the graph in the figure by adding 2 more vertices $n+1, n+2$ and mimicking the vertex $n$, i.e., for each $u \in [n-1]$, adding edges $(n+1, u)$ and $(n+2, u)$ if and only if there is an edge $(n, u)$. Note that if there is more than one vertex with minimum degree, the extra vertices (e.g., $n+1, n+2$ in Figure 2) can choose to mimic any one of
them. On the other hand, if $H$ has a unique vertex with minimum degree, then $\mathcal{M}^H_n$ has only one graph for each $n \leq N < 2n$ and the graph will look like the one in Figure 2.

**Definition 5.** Suppose $H$ is a graph with $n$ vertices and $r$ be an integer with $0 \leq r < n$. We call a graph $H$ elementary in remainder class $r$ if for all $N = qn + r$ with $q \in \mathbb{Z}^+$, a graph $G$ is edge-minimal among $N$-vertex $H$-covered graphs if and only if $G$ is the disjoint union of $q$ many graphs, each from $\mathcal{M}^H_n$ for some $n \leq b_i < 2n$, such that $\sum_{i=1}^{q} b_i = N$. For any subset $S \subseteq \{0,1,\ldots,n-1\}$, we call a graph $H$ elementary in remainder class $S$ if it is elementary in remainder class $r$ for all $r \in S$. We call a graph $H$ elementary if it is elementary in remainder class $\{0,1,\ldots,n-1\}$.

Note that for remainder class $\{0,1\}$, $H$ is elementary if and only if $H$ is ideal, because $\mathcal{M}^H_n = \mathcal{L}^H_n$ and $\mathcal{M}^H_{n+1} = \mathcal{L}^H_{n+1}$, where $n$ is the number of vertices of $H$.

**Theorem 1.3.** Let $0 < p < 1$ be a constant. With high probability, $G_{n,p}$ is elementary in remainder class $\{0,1,\ldots,\gamma\}$ and is ideal in remainder class $\{\beta, \beta+1,\ldots,n-1\}$ for some $\gamma, \beta$ with $\gamma = \Theta(\sqrt{n \log n})$ and $\beta = \gamma + 1$ or $\beta = \gamma + 2$.

**Remark.** Since $\gamma < \beta \leq \gamma + 2$, with high probability there exists at most one $r = \Theta(\sqrt{n \log n})$ such that $G_{n,p}$ is neither elementary nor ideal in remainder class $r$. Furthermore it turns out that with high probability, if such $r$ exists, then for the edge minimization problem of $G_{n,p}$-covered graphs, both the graph structures in Definition 3 and Definition 5 work, and those are the only graphs as well. We will also give the condition on when such $r$ exists, but we postpone the detailed analysis until the proof of Theorem 1.3 in Section 5.

Consider the more general problem of minimizing the number of cliques of an arbitrary fixed order instead of edges ($K_2$) among $H$-covered graphs with a given number of vertices. We will formulate this general problem in terms of an integer program which will be useful to prove Theorem 1.2 and Theorem 1.3.

**Theorem 1.4.** Fix positive integers $N$ and $t$, and a graph $H$ with $n$ vertices. The minimum number of subgraphs isomorphic to $K = K_t$ among all $H$-covered graphs with $N$ vertices, denoted by $\text{cov}_K(N,H)$, is equal to the solution to the following integer program:
Proof. The function \( a^H_k(x) \) is a contradiction.

Lemma 2.1. We start with a well-known result. We will use the following standard graph theoretic notation throughout this paper. For a graph \( G \) and \( v \in V(G) \), \( N_G(v) \) denotes the set of neighbors of \( v \) in \( G \) and \( d_G(v) = |N_G(v)| \), i.e., \( d_G(v) \) is the degree of \( v \). We write \( \delta(G) \) to denote the minimum degree of \( G \).

This paper is organized as follows. In Section 2, we prove Proposition 1.1 and Theorem 1.4. In Section 3, we prove Theorem 1.2 and construct graphs to show its tightness. In Section 4, we discuss some facts and lemmas about random graphs as a preparation to prove Theorem 1.3. Finally, we prove it in Section 5 and end with some concluding remarks in Section 6.

2 Minimizing the number of cliques of a fixed order

We start with a well-known result.

Lemma 2.1. Suppose \( n, k, r, \) and \( m \) are non-negative integers with \( r < n \) and \( m > k \). Over all choices of values \( x_i \in \{0, 1, 2, ..., n\} \) for each \( i \in [m] \) such that \( \sum_{i=1}^{m} x_i = kn + r \), the expression \( \sum_{i=1}^{m} \binom{x_i}{2} \) achieves its maximum if and only if \( x_i = n \) for \( k \) many values of \( i \in [m] \), \( x_i = r \) for one \( i \), and \( x_i = 0 \) for the rest of the values of \( i \).

Proof. The function \( f : \mathbb{Z} \to \mathbb{R} \) defined by \( f(x) = \frac{x(x-1)}{2} \) is strictly convex, so \( f(a) + f(b) < f(a-1) + f(b+1) \) whenever \( a < b \). Hence, if the values of \( x_i \) are not as in Lemma 2.1, we can increase the value of \( \sum_{i=1}^{m} \binom{x_i}{2} \) by changing 2 of \( x_i \) while keeping \( \sum_{i=1}^{m} x_i \) constant, which is a contradiction.

Warm up for Proposition 1.1 when \( t = 2 \). Fix \( q, n \) and \( r \) with \( r < n \), and let \( N = qn + r \). Let \( F \) denote the graph which consists of the union of 2 copies of \( K_n \) overlapping in \( n-r \) vertices, together with the disjoint union of \( q-1 \) many \( K_n \). First note that the number of edges in the graph \( F \) is \( (q+1)\binom{n}{2} - \binom{n-r}{2} \). Now our goal is to show that any \( N \)-vertex \( K_n \)-covered graph has at least \( (q+1)\binom{n}{2} - \binom{n-r}{2} \) many edges.

Consider a \( K_n \)-covered graph \( G \) with \( N \) vertices. Since \( v \) in is a copy of \( K_n \) for all \( v \in V(G) \), choose a set \( S(v) \) of \( n \) vertices of \( G \) such that \( v \in S(v) \) and \( S(v) \) induces \( K_n \). Now in order to lower-bound the number of edges in \( G \), do the following:
Algorithm 1.

1. Set $V_1 = V(G)$, $E_1 = E(G)$.

2. In the $i$-th step, pick an arbitrary vertex $v_i \in V_i$. Consider the subgraph $A(v_i) \cong K_n$ induced by $S(v_i)$. Now set $V_{i+1} = V_i \setminus S(v_i)$ and $E_{i+1} = E_i \setminus E(A(v_i))$. Repeat the procedure until $V_{i+1} = \emptyset$.

Note that in each step $i$, $E_i \supseteq E(G) \setminus E(V \setminus V_i)$. So, the number of edges in $A(v_i) \cap E_i$ is at least $\binom{n}{2} - \binom{k_i}{2}$, where $k_i = |(V(G) \setminus V_i) \cap S(v_i)|$. Let $m$ be the number of iterations, i.e. $V_{m+1} = \emptyset$. Observe that $q \leq m \leq N$. Clearly $N = \sum_{i=1}^{m} (n - k_i)$. and the number of edges in $G$ is at least $\sum_{i=1}^{m} \left( \binom{n}{2} - \binom{k_i}{2} \right)$. So, the only remaining thing is to determine the maximum value of $\sum_{i=0}^{m} k_i = mn - N = (m - q - 1)n + (n - r)$.

By using Lemma 2.1, $\sum_{i=1}^{m} \left( \binom{n}{2} - \binom{k_i}{2} \right)$ attains its maximum if and only if $k_i = n$ for $m - q - 1$ many values of $i$, $k_i = n - r$ for some other $i$, and $k_i = 0$ for the rest of the values of $i$. Hence, the number of edges in $G$ is at least $(q + 1)\left( \binom{n}{2} - \binom{n-r}{2} \right)$, proving the required bound. Now in order to prove the uniqueness of the extremal graph $F$, let us have a closer look at Algorithm 1.

As we have already noticed that $F$ has $(q + 1)\left( \binom{n}{2} - \binom{n-r}{2} \right)$ many edges, if we assume that $G$ has the minimum number of edges among $N$-vertex $K_n$-covered graphs, then $E_{m+1} = \emptyset$. By Lemma 2.1, the values of $k_i$’s are unique up to permutation, and they are exactly same as what we have seen in the beginning of this paragraph. Hence, it is straightforward to see that $F$ is the only graph achieving the minimum number of edges among $N$-vertex $K_n$-covered graphs.

Proof of Theorem 1.4 Let the minimum value of the integer program (1.1) be $z$. We first show that there is an $N$-vertex $H$-covered graph with $z$ many copies of $K = K_t$. Let $x_k$ be non-negative integer for all $k \in [n]$ such that $x_n \geq 1$ and $\sum_{k=1}^{n} kx_k = N$. We will construct an $N$-vertex $H$-covered graph with the number of copies of $K$ being $\sum_{k \in [n]} a_K(k)x_k$, where $a_K$ is a short hand for $a_H^K$.

Let $y_k = x_k$ for $k < n$ and let $y_n = x_n - 1$. Consider a graph $G$ on the vertex set $V \cup (\bigcup_{k=1}^{n} V_k)$, where $V_k = \bigcup_{i=1}^{y_k} V_{i,k}$, $V$ has $n$ vertices and each $V_{i,k}$ has $k$ vertices. There are no edges between $V_{i,k}$ and $V_{j,l}$ in $G$ when $(i, k) \neq (j, l)$. $V$ induces $H$ and $V \cup V_{i,k}$ induces a graph in $\mathcal{L}_H^K$. From the discussion after Definition 2, remember that the number of subgraphs isomorphic to $K$ in $\mathcal{L}_H^K$ is $a_K(n) + a_K(k)$, hence the number of copies of $K$ in $G$ is $\sum_{k \in [n]} a_K(k)x_k$.

At this point, to finish the proof of Theorem 1.4, it will be enough to show that the number of copies of $K$ of an $N$-vertex $H$-covered graph is at least $z$. We will do a similar treatment as it was done in the proof of Proposition 1.1 where $H$ was a complete graph and $K$ was $K_2$. Consider an $N$-vertex $H$-covered graph $G$. So, $v$ is in a copy of $H$ for all $v \in V(G)$, choose $S(v)$ a set of $n$ vertices of $G$ such that $v \in S(v)$ and $H$ is a subgraph of the induced subgraph on $S(v)$. Now we follow the following steps to lower-bound the number of copies of $K$ in $G$, which is similar to Algorithm 1 in the proof of Proposition 1.1 for $t = 2.$
Algorithm 2.

1. \( V_1 = V(G), \ E_1 = E(G) \).

2. In the \( i \)-th step, pick an arbitrary vertex \( v_i \in V_i \). Consider the subgraph \( A(v_i) \supseteq H \) induced by \( S(v_i) \). Now set \( V_{i+1} = V_i \setminus S(v_i) \) and \( E_{i+1} = E_i \setminus E(A(v_i)) \). Repeat the procedure if \( V_{i+1} \neq \emptyset \).

By using Definition 1, the number of subgraphs isomorphic to \( K \) in \( A(v_i) \cap E_i \) is at least \( a_K(k_i) \), where \( k_i = |V_i \cap S(v_i)| \). Note that \( k_1 = n \). Let \( m \) be the number of iterations, i.e. \( V_{m+1} = \emptyset \). Note that \( q \leq m \leq N \). Clearly \( \sum_{i=1}^{m} k_i = N \) and the number of subgraphs isomorphic to \( K \) in \( G \) is at least \( \sum_{i=1}^{m} a_K(k_i) \). Hence, the number of subgraphs isomorphic to \( K \) in \( G \) is at least the minimum value of the integer program (1.1).

Proof of Corollary 1.5. Let \( c = \min_{k \in [n]} \frac{a_H^K(k)}{k} \). From the integer program (1.1), one can observe that
\[
cN \leq \text{cov}_{K_t}(N, H) \leq a_H^K(n) + cN,
\]
which finishes the proof of Corollary 1.5.

Proof of Proposition 1.1. Note that \( a_{K_t}^n(k) = \binom{n}{k} - \binom{n-k}{k} \). So, by using the fact that \( \binom{x}{y} + \binom{y}{z} > \binom{y}{z} + \binom{y}{w} \) for all \( x, y, z, w \in \mathbb{N} \) such that \( x + y = z + w \) and \( x > \max(z, w) \), we get the following 2 equations for \( k, l > 0 \).

\[
a_{K_t}^n(k) + a_{K_t}^n(l) > a_{K_t}^n(k + l), \quad \text{for } k + l \leq n \tag{2.1}
\]

\[
a_{K_t}^n(k) + a_{K_t}^n(l) > a_{K_t}^n(n) + a_{K_t}^n(b), \quad \text{for } n < k + l = n + b < 2n \tag{2.2}
\]

Hence, if \( r \neq 0 \), then the integer program (1.1) in Theorem 1.4 has one unique solution: \( x_n = q, \ x_r = 1, \) and \( x_k = 0 \) for other values of \( k \in [n] \). For \( r = 0 \), the unique solution is just \( x_n = q \) and \( x_k = 0 \) for \( k \neq n \). Now for an \( N \)-vertex \( K_n \)-covered graph \( G \), by analyzing the steps of Algorithm 2 similarly to the warm up proof of Proposition 1.1 for \( t = 2 \), we can conclude that \( G \) minimizes the number of subgraphs isomorphic to \( K_t \) among \( N \)-vertex \( K_n \)-covered graphs if and only if \( G \) is the extremal graph in Proposition 1.1.

3 Regular graphs

3.1 Proof of Theorem 1.2

Throughout this section, let \( H \) be a fixed graph on \( n \) vertices so that we may write \( a(k) \) instead of \( a_H^K(k) \). We first warm up by proving the following interesting result.

Proposition 3.1. For any positive integers \( q, n, N = qn \) and any \( n \)-vertex graph \( H \) with \( \delta(H) \geq \frac{n-1}{2} \), \( G \) is the edge-minimal \( N \)-vertex \( H \)-covered graph if and only if \( G \) is the disjoint union of \( q \) many copies of \( H \).

Remark. In contrast with Theorem 1.2, Proposition 3.1 relaxes the regularity condition on \( H \), but imposes a divisibility condition on \( N \) so that \( n \) divides \( N \).
Proposition 3.1 routinely follows from Theorem 1.4 and Lemma 3.2 below.

Lemma 3.2. If $\delta(H) \geq \frac{n-1}{2}$, then $\frac{a(k)}{k}$ is minimized if and only if $k = n$.

Proof. For $k \leq \frac{n}{2}$, fix $A \subset V(H)$ with $|A| = k$. As $\delta(H) \geq \frac{n-1}{2}$,

$$\sum_{v \in A} d_H(v) \geq k \cdot \frac{n-1}{2}.$$ 

Clearly, each edge in the induced subgraph on $A$ contributes 2 to $\sum_{v \in A} d_H(v)$. So,

$$a(A) \geq k \cdot \frac{n-1}{2} - \binom{k}{2} = k \cdot \frac{n-k}{2}. \quad (3.1)$$

Similarly, for $\frac{n}{2} < k < n$, fix $A \subset V(H)$ with $|A| = k$. As $\delta(H) \geq \frac{n-1}{2}$, for $v \in V(H) \setminus A$,

$$d_A(v) \geq \frac{n-1}{2} - (n-k-1).$$

Hence, the number of edges between $A$ and $V(H) \setminus A$ is at least $(n-k) \left( \frac{n-1}{2} - (n-k-1) \right) = (n-k) \left( \frac{n-1}{2} \right)$. Now because $\sum_{v \in A} d_H(v) \geq k \cdot \frac{n-1}{2}$,

$$a(A) \geq (n-k) \left( k - \frac{n-1}{2} \right) + \frac{1}{2} \left( k \cdot \frac{n-1}{2} - (n-k) \left( k - \frac{n-1}{2} \right) \right)$$

$$= \frac{1}{2} (n-k) \left( k - \frac{n-1}{2} \right) + \frac{1}{2} \cdot k \cdot \frac{n-1}{2}$$

$$> \frac{1}{2} (n-k) \left( k - \frac{n-1}{2} \right) + \frac{1}{2} \cdot (n-k) \cdot \frac{n-1}{2}$$

$$= k \cdot \frac{n-k}{2}. \quad (3.2)$$

Hence, by using (3.1) and (3.2), we can conclude that,

$$a(k) \geq k \cdot \frac{n-k}{2}. \quad (3.3)$$

Note that $a(n)$ is just the number of edges in $H$ and the number of edges in $H$ is at most $a(A) + \binom{n-|A|}{2}$ for any $A \subseteq V(H)$. Hence for any $k$,

$$a(n) \leq a(k) + \binom{n-k}{2}. \quad (3.4)$$

Now from (3.1) and (3.3), we can say that

$$a(n) < a(k) + (n-k) \cdot \frac{a(k)}{k}.$$ 

So,

$$\frac{a(k)}{k} > \frac{a(n)}{n}.$$
Now in order to prove Theorem 1.2 we need a technical lemma. For $A \subseteq V(H)$, let $e(A)$ denote the number of edges in the induced subgraph on $A$, and let $e(k)$ denote the maximum value of $e(A)$ among all $k$-vertex subsets $A \subseteq V(H)$.

**Lemma 3.3.** If $\delta(H) \geq \frac{n-1}{2}$, then for all positive integers $k$ and $l$ less than $n$, with $k+l \leq n$, we have $e(k) + e(l) < e(k+l)$.

Before proving Lemma 3.3, let us first state an equivalent form of it. Remember that $a(0) = 0$ from Definition 1.

**Corollary 3.4.** If $\delta(H) \geq \frac{n-1}{2}$, then for all positive integers $k$ and $l$ less than $n$, with $k+l \geq n$, we have $a(k) + a(l) > a(n) + a(r)$, where $k+l = n+r$.

**Proof.** Note that for any $A \subseteq V(H)$, $e(A) + a(V(H) \setminus A)$ is just the number of edges in $H$. So for each $k \leq n$, $e(k) + a(n-k) = e(n) = a(n)$. Hence, we have the following:

$$e(k) + e(l) < e(k + l) \text{ for all } k, l \leq n \text{ and } k + l \leq n$$

$$\iff (a(n) - e(k)) + (a(n) - e(l)) > 2a(n) - e(k + l) \text{ for all } k, l \leq n \text{ and } k + l \leq n$$

$$\iff a(n-k) + a(n-l) > a(n) + a(n-k-l) \text{ for all } k, l \leq n \text{ and } k + l \leq n$$

$$\iff a(k) + a(l) > a(n) + a(k + l - n) \text{ for all } k, l \leq n \text{ and } k + l \geq n.$$ 

\[ \square \]

**Proof of Lemma 3.3.** Without loss of generality, assume that $k \leq l$.

**Case 1: $k \leq l \leq \frac{n}{2}$.** Suppose $A \subseteq V(H)$ is an $l$-vertex subset such that $e(A) = e(l)$. Our strategy is to choose a $k$-vertex subset $B \subseteq V(H) \setminus A$ uniformly at random and show that the expected number of edges inside $A \cup B$ exceeds $e(k) + e(l)$. As $\delta(H) \geq \frac{n-1}{2}$, the number of edges in $H$ that are not inside the induced subgraph on $A$ is at least $\frac{n(n-1)}{4} - \binom{l}{2}$. Also, each vertex in $A$ can have at most $l-1$ neighbors in $A$, so the number of edges with one endpoint in $A$ and another endpoint in $V(H) \setminus A$ is at least $\frac{(n-1)}{2} - (l-1)l$. Now if we choose a uniformly random subset $B \subseteq V(H) \setminus A$ of size $k$, the edges with one endpoint in $A$ and another endpoint in $V(H) \setminus A$ will be in $A \cup B$ with probability $\frac{k}{n-l}$, and the edges with both endpoints in $V(H) \setminus A$ will be in $A \cup B$ with probability $\frac{\binom{k}{2}}{\binom{n-l}{2}}$. Note that $\frac{k}{n-l} \geq \frac{k}{\binom{n}{2}}$.

So, we have the following lower bound on the expected number of edges in $A \cup B$, which we denote by $\mathbb{E}[e(A \cup B)]$.

\[
\mathbb{E}[e(A \cup B)] \\
\geq e(A) + \binom{n-1}{2} - (l-1)l \cdot \frac{k}{n-l} + \left(\frac{n(n-1)}{4} - \binom{l}{2} - \left(\frac{n-1}{2} - (l-1)l\right)\right) \cdot \frac{\binom{k}{2}}{\binom{n}{2}} \\
\geq e(A) + \frac{1}{2}(n-2l+1)(n-l) \cdot \frac{k^2}{(n-l)^2} + \left(\frac{n-1}{4}(n-2l) + \frac{l(l-1)}{2}\right) \cdot \frac{\binom{k}{2}}{\binom{n-l}{2}} \\
\geq e(A) + \frac{\binom{k}{2}}{\binom{n-l}{2}} \left(\frac{1}{2}(n-l)(n-2l+1) + \frac{n-1}{4}(n-2l) + \frac{l(l-1)}{2}\right)
\]
Now if we choose a uniformly random subset $B \subset V(H) \setminus A$ such that $e(A \cup B) > e(k) + e(l)$ where $|A \cup B| = k + l$, proving that $e(k + l) > e(k) + e(l)$.

**Case 2:** $k \leq \frac{n}{2} < l$. We proceed similarly to the first case. Suppose $A \subset V(H)$ is an $l$-vertex subset such that $e(A) = e(l)$. Clearly, $|V(H) \setminus A| = n - l$, and of course $e(V(H) \setminus A) \leq \binom{n - l}{2}$. The number of edges with one endpoint in $A$ and another endpoint in $V(H) \setminus A$ is at least $(n - l) \frac{n - l - 1}{2} - 2 \cdot e(V(H) \setminus A)$. Now, the total number of edges in $H$ that are not inside the induced subgraph on $A$ is at least

$$e(V(H) \setminus A) + (n - l) \frac{n - 1}{2} - 2 \cdot e(V(H) \setminus A) \geq (n - l) \frac{n - 1}{2} - \binom{n - l}{2} > \binom{n - l}{2}.$$

Now if we choose a uniformly random subset $B \subset V(H) \setminus A$ of size $k$, the edges with one endpoint in $A$ and another endpoint in $V(H) \setminus A$ will be in $A \cup B$ with probability $\frac{k}{n - l} \geq \frac{\binom{k}{k}}{\binom{n - l}{k}}$, and the edges with both endpoints in $V(H) \setminus A$ will be in $A \cup B$ with probability $\frac{k}{\binom{n - l}{k}}$. Hence,

$$\mathbb{E}[e(A \cup B)] > e(A) + \frac{k}{n - l} \binom{n - l}{2} = e(A) + \frac{k}{2}. $$

So, we are done.

Now armed with the last lemma we are ready to prove a technical theorem with a more general condition than Theorem 1.2 which will imply Theorem 1.2.

**Theorem 3.5.** Suppose $H$ is a graph with $n$ vertices and minimum degree at least $\frac{n - 1}{2}$, and $a(k) + a(l) > a(k + l)$ for all positive integers $k$ and $l$ with $k + l \leq n$. Then, $H$ is an ideal graph.

**Proof.** Fix positive integers $q, n$ and $N = qn + r$ with $0 \leq r < n$. The assumption that $a(k) + a(l) > a(k + l)$ is the same as equation (2.1) in Proposition 1.1, and the conclusion of Corollary 3.4 is the same as equation (2.2) in Proposition 1.1. So, the integer program (1.1) in Theorem 1.4 has the same unique minimum solution as in Proposition 1.1 i.e., $x_n = q$, $x_r = 1$, and $x_k = 0$ for all other $k$. This implies that the extremal graph in Definition 3 has the minimum number of edges among $N$-vertex $H$-covered graphs. Now the only remaining thing that we need to show is the uniqueness of the extremal graph. This proof is very similar to the uniqueness proof in Proposition 1.1, which is routine to verify.
Proof of Theorem 1.2. Consider an arbitrary \( d \)-regular graph \( H \). Observe that for any \( A \subseteq V(H) \), we have \( a(A) + e(A) = d|A| \). Hence,
\[
a(k) + e(k) = dk \text{ for } k \leq n. \tag{3.5}
\]
If \( d \geq \frac{n-1}{2} \), Lemma 3.3 implies that \( e(k+l) > e(k) + e(l) \) for all \( k, l > 0 \) with \( k + l \leq n \). Now by using (3.5), we can conclude that \( a(k) + a(l) > a(k + l) \) for all \( k, l > 0 \) with \( k + l \leq n \). Hence, Theorem 1.2 is now implied by Theorem 3.5.

3.2 Construction showing that Theorem 1.2 is tight

In this subsection, we construct graphs to show the tightness of Theorem 1.2 in a strong sense. For every odd \( n \), we will construct non-ideal regular graphs with \( n \) vertices and degree as close as possible but less than \( \frac{n-1}{2} \). We need a few notations for the ease of describing the graphs.

Notation.

- Suppose \( G_i \) is a graph for all \( i \in [n] \) and \( a_i \) is a non-negative integer for all \( i \in [n] \). We denote the disjoint union of \( a_i \) many \( G_i \) for all \( i \in [n] \) by \( \oplus_{i=1}^{n} a_i G_i \). We simply write \( G \) instead of \( 1G \).

- Suppose \( S_i \) is a set of graphs for all \( i \in [n] \). The set \( \oplus_{i=1}^{n} S_i \) denotes the set \( \{ \oplus_{i=1}^{n} H_i : H_i \in S_i \} \).

Let \( n = 2l + 1 \) and \( d \) be an even number such that \( d < \frac{n-1}{2} \). This is the highest we can consider because there are no graphs with an odd number of vertices and an odd degree of regularity. Consider a connected \( d \)-regular graph \( H_1 \) on \( l \) vertices and a connected \( d \)-regular graph \( H_2 \) on \( l + 1 \) vertices. Consider the \( d \)-regular graph \( H = H_1 \oplus H_2 \), which has \( n \) vertices.

Lemma 3.6. \( \frac{a(k)}{k} \) is minimized if and only if \( k \in \{l, l+1, 2l+1\} \).

Proof. As \( H \) is \( d \)-regular, \( a(k) \geq \frac{dk}{2} \). Now \( a(V(H_1)) = \frac{dl}{2} \), \( a(V(H_2)) = \frac{dl+1}{2} \) and \( a(V(H)) = \frac{d(2l+1)}{2} \). Hence, for \( k \in \{l, l+1, 2l+1\} \), \( \frac{a(k)}{k} \) achieves its minimum. For \( A \subset V(H) \) with \( |A| \notin \{l, l+1, 2l+1\} \), there is at least one edge between \( A \) and \( V(H) \setminus A \), which implies that \( a(A) \geq \frac{d|A|}{2} \). So, \( \frac{a(k)}{k} > \frac{d}{2} \) for \( k \notin \{l, l+1, 2l+1\} \).

Proposition 3.7. \( H \) is not ideal in any remainder class.

Proof. Let \( N \geq n \) and \( G \) be an \( N \)-vertex \( H \)-covered graph. Clearly, every vertex of \( G \) has degree at least \( d \) because every vertex of \( G \) is in a copy of \( H \), which is a \( d \)-regular graph. Hence, the number of edges of \( G \) is at least \( \frac{dN}{2} \). Now fix \( r \in \{0, 1, ..., n-1\} \). Choose \( a, b \in \mathbb{N} \) such that \( a > 1 \) and \( an + r = n + bl \). This choice is possible because \( n = 2l + 1 \). Let \( N = an + r = n + bl \). Consider the graph \( G = H \oplus BH_1 = H_2 \oplus (b + 1)H_1 \), clearly every vertex of \( G \) is in a copy of \( H \). Notice that the number of edges in \( G \) is \( \frac{dN}{2} \), because the degrees of all vertices are \( d \). Observe that \( G \) is not isomorphic to any graph in \( (a - 1)H \oplus \mathcal{L}_n^H \), because there are at least 2 disjoint copies of \( H \) in any graph in \( (a - 1)H \oplus \mathcal{L}_n^H \). So, \( H \) is not ideal.

Remark. Moreover, due to Lemma 3.6 the number of edges in any graph in \( (a - 1)H \oplus \mathcal{L}_n^H \) is strictly greater than \( \frac{dN}{2} \) when \( r \notin \{l, l+1, 2l+1\} \).
4 Random graph tools for proving Theorem 1.3

This section will be full of facts and lemmas about $G_{n,p}$, where $0 < p < 1$ is a fixed constant. Throughout this paper we will only focus on constant probability $p$ in $G_{n,p}$ (in other words, $p$ will not depend on $n$). In this section and the next section, all logarithms are in base $e$. We will begin by mentioning some of the theorems and lemmas about the degree sequence of $G_{n,p}$ from [2] and [8], which will be useful in the proof of Theorem 1.3. Let the degree sequence of $G_{n,p}$ be $d_1 \leq d_2 \leq ... \leq d_n$. The abbreviation w.h.p. stands for with high probability, which means that the probability of the event under consideration tends to one as $n \to \infty$, as it was defined in the introduction.

**Lemma 4.1** (Lemma 3.7 in [8]). Let $\epsilon = 1/10$, and let $q = 1 - p$. If
\[
d_\pm = (n - 1)p - (1 \mp \epsilon)\sqrt{2(n - 1)pq \log n}
\]
then w.h.p.

- The minimum degree $\delta(G_{n,p}) \geq d_-.$
- There are $\Omega(n^{2\epsilon(1-\epsilon)})$ vertices of degree at most $d_+$.
- There are no $u \neq v$ such that $d_u, d_v \leq d_+$ and $|d_u - d_v| \leq 10$.

**Corollary 4.2** (Consequence of Lemma 4.1). W.h.p. $G_{n,p}$ has the property that $d_n - d_1 < 4\sqrt{(n - 1)pq \log n}$.

**Theorem 4.3** (Theorem 12 in [2]). Suppose $m \to \infty$ and $m = o(n)$. Put
\[
K(m, n) = pn - \sqrt{2pqn \log(n/m)} + (\log \log(n/m) + \log 4\pi)\sqrt{\frac{pqn}{8 \log(n/m)}}
\]
Then w.h.p. $G_{n,p}$ satisfies $|d_m - K(m, n)| = O\left(\sqrt{\frac{n}{\log(n/m)}}\right)$.

**Lemma 4.4** (Corollary 15 in [2]). If $m = o(n^{1/4})/((\log n)^{1/4})$, then w.h.p. $G_{n,p}$ has the property $d_1 < d_2 < ... < d_m$.

**Lemma 4.5** (Lemma 18 in [2]). For any $\epsilon > 0$, w.h.p. $G_{n,p}$ has the property that whenever $m > n^\epsilon$ every subgraph of order $m$ has $p\binom{m}{2} + o(m^2)$ edges.

**Corollary 4.6** (Consequence of Lemma 4.5). For any $\epsilon > 0$, w.h.p. $G_{n,p}$ has the property that whenever $m \geq l > n^\epsilon$, $A \in \binom{V(G_{n,p})}{m}$, $B \in \binom{V(G_{n,p})}{l}$, and $A \cap B = \emptyset$, then the number of edges with one endpoint in $A$ and another in $B$ is $pml + o(m^2)$.

**Proof Sketch.** The above corollary is an easy consequence of Lemma 4.5. If $A$ and $B$ are disjoint vertex subsets of $G_{n,p}$, then the number of edges with one endpoint in $A$ and another in $B$ is exactly $e(A \cup B) - e(A) - e(B)$, where $e(V)$ denotes the number of edges in the induced subgraph on $V$. Now apply Lemma 4.5 on $A$, $B$, and $A \cup B$ to conclude Corollary 4.6. 

13
Proof. It is enough to prove that w.h.p. for all $A \subset V(G_{n,p})$ with $0 < |A| < n$, there exists $v \in V(G_{n,p}) \setminus A$ such that the number of neighbors of $v$ in $A$ is more than $(1 - \epsilon)p|A|$. From Lemma 4.1, it can be seen that the probability that there exists $A \subset V(G_{n,p})$ with $1 \leq |A| \leq \frac{n}{2}$, and let $m = |A|$. Now by the Chernoff Bound (4.1), the probability that there are less than $(1 - \epsilon)pm(n - m)$ edges between $A$ and $V(G_{n,p}) \setminus A$ is less than $e^{-m(n - m)e_{\epsilon}/2}$. So, the probability that there exists $A$ such that $1 \leq |A| \leq \frac{n}{2}$ and the number of edges between $A$ and $V(G_{n,p}) \setminus A$ is less than $(1 - \epsilon)p|A|(n - |A|)$ is less than

$$\sum_{m=\lfloor \frac{n}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} \left( \frac{ne}{m} \cdot e^{-\frac{n^2}{4m}} \right)^m = o(1).$$

$\square$

The next two lemmas are not tight, but give us what will suffice to prove our results.

Lemma 4.8. Let the degree sequence of $G_{n,p}$ be $d_1 \leq d_2 \leq \ldots \leq d_n$. Then there exists $\epsilon > 0$ such that w.h.p. for all $2 \leq k \leq \epsilon \sqrt{n} \log n$, $\sum_{i=1}^{k}(d_i - d_1) > \binom{k}{2}$.

Proof. Let $m = n^{1/3}$. By Lemma 4.4 w.h.p. $d_1 < d_2 < d_3 < \ldots < d_{3m}$. Note that we should have used $[3m]$ or $[3m]$ instead of $3m$ but it is true in both cases, hence for asymptotic results we omit rounding when it is not essential. From Lemma 4.1, it can be seen that w.h.p. $d_2 - d_1 \geq 2$. Note that $\binom{k}{2} = \sum_{i=1}^{k-1}i$, hence we have proven the statement of Lemma 4.8 for $k \leq 3m$. Now using Theorem 4.3, we can conclude that for some $\epsilon > 0$, $d_m - d_1 \geq d_m - d_{\sqrt{m}} \geq \epsilon \sqrt{n} \log n$ w.h.p. Now for $3m < k < \epsilon \sqrt{n} \log n$, at least half of the $i \in [k]$ have $d_i \geq d_m$, so $\sum_{i=1}^{k}(d_i - d_1) > \binom{k}{2}(d_m - d_1) \geq k \frac{k}{2} \epsilon \sqrt{n} \log n > \binom{k}{2}$.

$\square$

Lemma 4.9. Let the degree sequence of $G_{n,p}$ be $d_1 \leq d_2 \leq \ldots \leq d_n$. Then w.h.p. for all $n^{1/3} < k < \frac{1}{2}n^{1/3} \log n$, $\sum_{i=1}^{k}(d_{2k} - d_i) = o(k^2)$.

Proof.

$$\sum_{i=1}^{k} d_{2k} - d_i \leq \sum_{i=1}^{k} (d_{\sqrt{n} \log n} - d_i) + \sum_{i=\sqrt{n} \log n}^{k} (d_{\sqrt{n} \log n} - d_i)$$

$$\leq \sqrt{n} \log n \cdot 4 \sqrt{pq} \log n + k(d_{\sqrt{n} \log n} - d_{\sqrt{n} \log n})$$

$$\leq o(n) + k \sqrt{2pq \log \left( \frac{n \log n}{\sqrt{n}} \right)} - \sqrt{2pq \log \left( \frac{n \log n}{\sqrt{n}} \log n \right)} + o(\sqrt{n})$$

$$= o(k^2).$$
In the above calculation, in step (4.2), we have used the concentration of the degree sequence of $G_{n,p}$ (Corollary [4.2]), and in step (4.3), we have used Theorem 4.3.

Now we will switch to a natural problem of independent interest that turns out to be very useful to our problem. We need to use and build upon a couple of definitions from Section 3.1 to describe the problem.

**Definition 6.**

- For a graph $G$ and a subset $A$ of its vertex set, let $e(A)$ denote the number of edges in the induced subgraph on $A$.

- For a graph $G$, $e_G(k)$ is defined to be the maximum value of $e(A)$ among all $k$-vertex subsets $A \subseteq V(G)$. When $G$ is clear from the context, we will write $e(k)$ instead of $e_G(k)$. In all random graph situations in this paper, $p$ will be a fixed constant as $n$ grows. With $p$ being understood as given, we will write $e_n(k)$ to denote the random variable $e_{G_{n,p}}(k)$.

Our aim is to obtain a concentration-type result for $e_n(k)$, where the following inequality will become handy. This version of Talagrand’s Inequality can be found in Section 21.8 of [8].

**Theorem 4.10** (Talagrand’s Inequality). Let $\Omega = \prod_{i=1}^{n} \Omega_i$, where each $\Omega_i$ is a probability space and $\Omega$ has the product measure. Let $A \subseteq \Omega$ and let $x = (x_1, x_2, ..., x_n) \in \Omega$.

For $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{R}^n$ we let

$$d_\alpha(A, x) = \inf_{y \in A} \sum_{i : y_i \neq x_i} \alpha_i.$$

Then we define

$$\rho(A, x) = \sup_{|\alpha| = 1} d_\alpha(A, x),$$

where $|\alpha|$ denotes the Euclidean norm $\sqrt{\alpha_1^2 + \alpha_2^2 + ... + \alpha_n^2}$.

We then define, for $t \geq 0$,

$$A_t = \{x \in \Omega : \rho(A, x) \leq t\}.$$

Then

$$\mathbb{P}[A](1 - \mathbb{P}[A_t]) \leq e^{-\frac{t^2}{2}}.$$

**Lemma 4.11.** For any function $f(n)$ such that $f(n)$ tends to infinity as $n$ tends to infinity, w.h.p. the difference between $e_n(k)$ and $e_{\frac{n}{2}}(k)$ is at most $k^2$ for all $f(n) \leq k \leq \frac{n}{2}$.

**Proof.** We first mention a brief outline of the proof. We will use Talagrand’s Inequality [4.10] to show that $e_{\frac{n}{2}}(k)$ is concentrated in an interval of length $k^2$ for all $c \log n \leq k \leq n$ with (very) high probability. Then we will show that there are $V_1, V_2, ..., V_l \in \binom{V(G_{n,p})}{n/2}$ (where $l$ is not very big) such that any $k$-subset of $V(G_{n,p})$ will be totally inside at least one $V_i$. Note that the induced graph on $V_i$ is just distributed as $G_{\frac{n}{2},p}$, hence $\mathbb{P}[e_n(k) > x] \leq l \cdot \mathbb{P}[e_{n/2}(k) > x]$.
by the union bound. This is the strategy to show a small gap between \( e_n(k) \) and \( e_{n/2}(k) \) for fixed \( k \), and of course to prove the same for all \( f(n) \leq k \leq n \), we use the union bound.

**Concentration of \( e_2(k) \).** Let \( m = \binom{n/2}{2} \). The \( G_{2,p} \) model is the same as \( \Omega = \prod_{i=1}^{m} \Omega_i \), where the \( \Omega_i \)'s are independent Bernoulli Distributions with success probability \( p \). Let

\[
x_k = \min\{x \in \mathbb{N} : \mathbb{P}[e_{n/2}(k) \leq x] \geq q_k\}, \quad \text{where } q_k = e^{-\frac{k^3/2}{8}}. \tag{4.4}
\]

Let \( \mathcal{A} = \{G \in \Omega : e_G(k) \leq x_k\} \). Clearly

\[
\mathbb{P}[\mathcal{A}] \geq q_k. \tag{4.5}
\]

For \( G = (G_1, G_2, \cdots, G_m) \in \Omega \), let

\[
\rho(A, G) = \sup_{|\alpha| = 1} \inf_{H \in \mathcal{A}} \sum_{i : G_i \not\in H_i} \alpha_i.
\]

For \( t > 0 \), let

\[
\mathcal{A}_t = \{G \in \Omega : \rho(A, G) \leq t\}.
\]

If \( G \in \Omega \) such that \( e_G(k) \geq x_k + d_k \), then there exists a \( k \)-vertex subset \( A \subseteq V(G) \) such that \( e(A) \geq x_k + d_k \). So, by using \( \alpha_i \)'s such that those corresponding to pairs of vertices in \( A \) are \( \alpha_i = \frac{1}{\sqrt{2}} \) and all other \( \alpha_i = 0 \), it is easy to see that \( G \not\in \mathcal{A}_{\frac{d_k}{4}} \).

Using Talagrand’s Inequality (Theorem 4.10), \( \mathbb{P}[\mathcal{A}] (1 - \mathbb{P}[\mathcal{A}_t]) \leq e^{-\frac{t^2}{4}} \), and using (4.5), we can conclude that \( q_k \mathbb{P}[e_{n/2}(k) \geq x_k + d_k] \leq e^{-\frac{d_k^2}{4k^2}} \). Now choosing \( d_k = k^{7/2} \) and using (4.4), we have \( \mathbb{P} \left[ e_{n/2}(k) \geq x_k + k^{7/2} \right] \leq e^{-\frac{k^3/2}{8}}. \) Hence,

\[
\mathbb{P} \left[ e_{n/2}(k) \notin \left[ x_k, x_k + k^{7/2} \right] \right] \leq 2e^{-\frac{k^3/2}{8}}. \tag{4.6}
\]

Since \( f(n) \to \infty \) as \( n \to \infty \), the probability that there exists \( f(n) \leq k \leq n \) such that \( e_{n/2}(k) \notin [x_k, x_k + k^{7/2}] \) is at most \( \sum_{k=f(n)}^{\infty} 2e^{-\frac{k^3/2}{8}} = o(1). \)

**Covering of \( V \) (\(|V| = n\)) by subsets of size \( \frac{n}{2}\).** We will show that there exists \( l < 2^{2k} \) and \( V_1, V_2, \ldots, V_l \in \left( \begin{array}{c} n \vspace{-0.2cm} \end{array} \right) \) such that for all \( A \in \left( \begin{array}{c} V \vspace{-0.2cm} \end{array} \right) \) there exists \( i \in [l] \) with \( A \subset V_i \). First partition \( V \) into \( 2k + 1 \) almost-equal sets \( U_1, U_2, \ldots, U_{2k+1} \) such that \( |U_i| \) equals to either \( \left\lceil \frac{n}{2k+1} \right\rceil \) or \( \left\lfloor \frac{n}{2k+1} \right\rfloor \) for all \( i \in [2k+1] \). Now for all \( S \in \left( \begin{array}{c} \binom{2k+1}{k} \end{array} \right) \), choose \( V_S \in \left( \begin{array}{c} V \vspace{-0.2cm} \end{array} \right) \) such that \( V_S \supseteq \bigcup_{i \in S} U_i \), which is clearly possible because \( \sum_{i \in S} |U_i| < k \left( \frac{n}{2k+1} + 1 \right) < \frac{n}{2} \) for large enough \( n \). Note that \( |S| = \binom{2k+1}{k} < 2^{2k} \). Now we show that for \( A \in \left( \begin{array}{c} V \vspace{-0.2cm} \end{array} \right) \) there exists \( S \in \left( \begin{array}{c} \binom{2k+1}{k} \end{array} \right) \) such that \( A \subseteq V_S \). For a fixed \( A \in \left( \begin{array}{c} V \vspace{-0.2cm} \end{array} \right) \), consider \( S' = \{i : U_i \cap A \neq \emptyset\} \); clearly \( |S'| \leq k \) and so any \( S \in \left( \begin{array}{c} \binom{2k+1}{k} \end{array} \right) \) with \( S \supseteq S' \) will obviously give \( A \subset V_S \).

**Completing proof of Lemma 4.11.** Note that the distribution of an induced subgraph of \( G_{n,p} \) on vertex set \( V \subseteq V(G_{n,p}) \) is \( G_{|V|, p} \). For a graph \( G \) with \( n \) vertices using the above
covering of vertex set $V(G)$, we know that there exist at most $2^{2k}$ induced subgraphs on $\frac{n}{2}$ vertices such that if all of those subgraphs do not have any subgraph of size $k$ with at least $M$ edges then $G$ also does not have a subgraph of size $k$ with at least $M$ edges. Hence, by using (4.6),

$$\Pr[e_n(k) > x_k + k\frac{7}{4}] \leq 2^{2k} \Pr[e_{n/2}(k) > x_k + k\frac{7}{4}] \leq 2^{2k} 2e^{-\frac{k^{3/2}}{8}} \leq e^{-\frac{k^{3/2}}{8}}.$$  

It is also trivial to see that

$$\Pr[e_n(k) < x_k] \leq \Pr[e_{n/2}(k) < x_k] \leq e^{-\frac{k^{3/2}}{9}}.$$  

Since $f(n) \to \infty$ as $n \to \infty$, by a similar summation over $k$ as was done for $e_{n/2}(k)$, one can see that w.h.p. $e_n(k)$ also takes a value in the interval $[x_k, x_k + k\frac{7}{4}]$ for all $f(n) \leq k \leq n$. Hence, we are done.

5 Proof of Theorem 1.3

Let $0 < p < 1$ be a fixed constant. Fix a graph $H$ sampled from the distribution of random graphs $G_{n,p}$. Throughout this section, let $d_1 \leq d_2 \leq \ldots \leq d_n$ be the degree sequence of $H$, and we will use $a(A)$ and $a(k)$ from Definition 1. Let $\beta'$ denote the minimum $k \in \{2, 3, \ldots, n\}$ for which $a(k) \leq ka(1)$. Note that this inequality is trivially true for $k = 1$, but w.h.p. false for $k = 2$ since w.h.p. $d_2 - d_1 \geq 2$ by Lemma 4.1. Lemma 5.1 will show that w.h.p. $\beta'$ exists. Let $\gamma = \beta' - 1$ and let us define $\beta$ as follows.

- If $a(\beta') = \beta'a(1)$, then define $\beta = \beta' + 1$.
- Otherwise define $\beta = \beta'$.

Our aim is to show that this choice of $\gamma$ and $\beta$ works in Theorem 1.3. From the definition of $\gamma$ we have the following for all $2 \leq k \leq \gamma$:

$$a(k) > ka(1).$$  

(5.1)

Lemma 5.1. W.h.p. $\gamma = \Theta(\sqrt{n \log n})$.

Proof. By using Lemma 4.8 there exists $\epsilon > 0$ such that w.h.p. for all $2 \leq k \leq \epsilon \sqrt{n \log n}$,

$$a(k) \geq \sum_{i=1}^{k} d_i - \binom{k}{2} > kd_1 = ka(1).$$  

(5.2)
Now there exists a constant $L$ such that w.h.p. for $k = L \sqrt{n \log n}$,

$$a(k) < \sum_{i=1}^{k} d_i - \frac{p}{2} \binom{k}{2} \tag{5.3}$$

$$< kd_1 + k\Theta(\sqrt{n \log n}) - \frac{p}{2} \binom{k}{2} \tag{5.4}$$

$$< kd_1 \tag{5.5}$$

$$= ka(1) \tag{5.6}$$

where (5.3) is true because w.h.p. if $A \in \binom{V(H)}{k}$ consists of the $k$ vertices with smallest degrees, then $a(A) < \sum_{i=1}^{k} d_i - \frac{p}{2} \binom{k}{2}$, where the $\frac{p}{2} \binom{k}{2}$ part comes from Lemma 4.5. Step (5.4) comes from the fact that w.h.p. $d_n - d_1 \leq \Theta(\sqrt{n \log n})$, which is true because of Corollary 4.2. One can choose a constant $L$ satisfying (5.5). Now combining (5.2) and (5.6), we conclude that w.h.p. $\gamma = \Theta(\sqrt{n \log n})$.

**Lemma 5.2.** W.h.p. for all positive integers $k$ and $l$ such that $n^{\frac{1}{2}} < k < l \leq n$, it holds that

$$\frac{a(k)}{k} > \frac{a(l)}{l}.$$

**Proof.** It is enough to show that w.h.p. for all $k > n^{\frac{1}{2}}$, it holds that $\frac{a(k)}{k} > \frac{a(k+1)}{k+1}$. By Lemma 4.5, w.h.p. for an arbitrary $k$-vertex subset $A \subset V(H)$, we have $e(A) < \frac{2}{3} pk^2$. So,

$$a(A) = \sum_{v \in A} d_v - e(A) > \sum_{i=1}^{k} d_i - \frac{2}{3} pk^2.$$ 

Hence,

$$a(k) > \sum_{i=1}^{k} d_i - \frac{2}{3} pk^2. \tag{5.7}$$

Let us first prove that w.h.p. $\frac{a(k)}{k} > \frac{a(k+1)}{k+1}$ for all $n^{\frac{1}{2}} < k < \frac{1}{2} n^{\frac{1}{2}} \log n$. Let $A \in \binom{V(H)}{k}$ such that $a(k) = a(A)$. Let $B$ be the set of the $k$ smallest degree (in $H$) vertices outside $A$. Now by using Lemma 4.6, there are more than $\frac{2}{3} pk^2$ many edges between $A$ and $B$. Hence, there is $v \in B$ such that $v$ has more than $\frac{2}{6} pk$ neighbors in $A$. So,

$$a(A \cup \{v\}) = a(A) + d_v - N_A(v) < a(k) + d_{2k} - \frac{5}{6} pk.$$ 

Hence,

$$a(k + 1) < a(k) + d_{2k} - \frac{5}{6} pk. \tag{5.8}$$

Now adding inequality (5.7) to inequality (5.8) multiplied by $k$, we get

$$ka(k + 1) < (k + 1)a(k) + \sum_{i=1}^{k} (d_{2k} - d_i) - \frac{1}{6} pk^2.$$ 

So, in order to prove that w.h.p. $\frac{a(k)}{k} > \frac{a(k+1)}{k+1}$, it is enough to show that $\sum_{i=1}^{k} (d_{2k} - d_i) < \frac{1}{6} pk^2$, which is w.h.p. true because of Lemma 4.9.
For the case $k \geq \frac{1}{2}n^{\frac{1}{2}} \log n$, instead of needing to carefully pick which vertex to add to $A$, we can pick an arbitrary vertex. Let $A \in \binom{V(H)}{k}$ such that $a(k) = a(A)$. Now use Lemma 4.7 to conclude that w.h.p. there is $v \in V(H) \setminus A$ such that $N_A(v) > \frac{5}{6}pk$. So, similar to the other case but with $d_n$ instead of $d_{2k}$ because we pick an arbitrary vertex, it is sufficient to show that $\sum_{i=1}^{k} (d_n - d_i) < \frac{1}{6}pk^2$, which is w.h.p. true because of the fact that w.h.p. $d_n - d_1 = \Theta(\sqrt{n \log n})$ from Corollary 1.2.

From Lemma 5.1 and Lemma 5.2, we can say that w.h.p. for $k \geq \beta$,

$$a(k) < ka(1).$$

(5.9)

**Lemma 5.3.** W.h.p. for all positive integers $k$ and $l$ with $\beta \leq k + l \leq n$, we have $a(k) + a(l) > a(k + l)$.

**Proof.** First we claim that for any positive integers $k'$ and $l'$ with $k' < l'$ and $\beta \leq l'$, it holds that $\frac{a(k')}{k'} > \frac{a(l')}{l'}$. This clearly holds when $k' > n^{\frac{1}{2}}$ by Lemma 5.2. Otherwise $k' \leq \gamma$ because $\gamma = \Theta(\sqrt{n \log n})$ by Lemma 5.1 and in this case by (5.1) and (5.9), we have $\frac{a(k')}{k'} \geq a(1) > \frac{a(l')}{l'}$. Now by using this claim, for any positive integers $k$ and $l$ with $\beta \leq k + l \leq n$,

$$a(k) + a(l) > \frac{k}{k + l} \cdot a(k + l) + \frac{l}{k + l} \cdot a(k + l) = a(k + l).$$

\[\square\]

**Lemma 5.4.** W.h.p. for all positive integers $k$ and $l$ less than $n$ with $k + l \geq n$, we have $a(k) + a(l) > a(n) + a(r)$, where $k + l = n + r$.

**Remark.** If $p > \frac{1}{2}$, then w.h.p. the minimum degree of $G_{n,p}$ is at least $\frac{n}{2}$, and in this case Corollary 3.4 already implies Lemma 5.4. Regardless, the proof below works for all $0 < p < 1$.

**Proof.** As we defined in Definition 6, $e(k) = \max_{A \in \binom{V(H)}{k}} e(A)$, where $e(A)$ denotes the number of edges in the induced subgraph on $A$. Just as in Corollary 3.4 to prove Lemma 5.4, it is enough to show that $e(k) + e(l) < e(k + l)$ whenever $k + l \leq n$. Without loss of generality, let us assume that $k \geq l$.

**Case 1:** $k \geq \frac{L}{p}$. Let $A \in \binom{V(H)}{k}$ such that $e(k) = e(A)$. Let $A_0 = A$, and define $A_{i+1}$ recursively by adding some $v \in V(H) \setminus A_i$ to $A_i$ such that $N_{A_i}(v) > \frac{p}{2} |A_i|$, which we know exists by Lemma 4.7. Now

$$e(A_i) > e(k) + \frac{p}{2} kl > e(k) + \left(\frac{l}{2}\right) \geq e(k) + e(l).$$

Note that $|A_i| = k + l$, and hence $e(k) + e(l) < e(k + l)$.

**Case 2:** $k < \frac{L}{p}$ and $k \geq n^{\frac{1}{2}}$. Lemma 5.4 is trivial for this case by Lemma 4.5.

**Case 3:** $k < \frac{L}{p}$ and $k < n^{\frac{1}{2}}$. Let $b = \frac{1}{p}$. It is well-known that the largest clique in $G_{n,p}$ has order $\approx 2 \log_b n$ (see, e.g., [12] or Theorem 7.3 in [8]). So, it is enough to handle this case only when $k + l \geq \log_b n$. Fix $k, l \in \mathbb{N}$ such that $k < \frac{L}{p}$, $k < n^{\frac{1}{2}}$ and $k + l \geq \log_b n$. Partition
the vertex set $V(H)$ into two almost equal sets $V_1, V_2$ such that $||V_1|−|V_2|| ≤ 1$. Now the strategy is to first reveal the edges (edges are there with probability $p$ independently) in the induced subgraphs on $V_1$ and $V_2$, then find the densest $k$-vertex subset $A_1 \subseteq V_1$ and $l$-vertex subset $A_2 \subseteq V_2$, and finally after revealing all the edges between $A_1$ and $A_2$ we argue that the number of edges between $A_1$ and $A_2$ cannot be too small. Let $H_i$ denote the induced subgraph on $V_i$ for $i = 1, 2$. Clearly, the distribution of the induced subgraph on $V \subseteq V(G_{n,p})$ is same as $G_{|V|,p}$. From Lemma 4.11, we know that for $j = k, l$,

$$e_{H}(j) − e_{H_1}(j) ≤ j^2$$ for $i = 1, 2$. (5.10)

Let $A_1 \in \binom{V_1}{k}$ such that $e_{H_1}(k) = e(A_1)$, and let $A_2 \in \binom{V_2}{l}$ such that $e_{H_2}(l) = e(A_2)$. By the Chernoff Bound (4.1),

$$\mathbb{P} \left( e(A_1, A_2) ≤ \frac{1}{2} klp \right) \leq e^{-\frac{klp}{8} ≤ e^{-\frac{\sigma^2}{8}} ≤ e^{-\frac{p^2(\log n)^2}{32}}},$$ (5.11)

where $e(A_1, A_2)$ denotes the number of edges between $A_1$ and $A_2$. If $e(A_1, A_2) > \frac{1}{2} klp > \frac{\sigma^2}{2} k^2$, then

$$e_{H}(k + l) ≥ e_{H}(A_1 \cup A_2) = e(A_1) + e(A_2) + e(A_1, A_2) > e_{H_1}(k) + e_{H_2}(l) + \frac{p^2}{2} k^2$$

$$≥ e_{H}(k) − k^2 + e_{H}(l) − l^2 + \frac{p^2}{2} k^2$$ (5.12)

$$> e_{H}(k) + e_{H}(l),$$

where step (5.12) uses (5.10). Now by using (5.11), the probability that there exist $k, l$ in this case with $e(k) + e(l) ≥ e(k + l)$ is at most $ne^{-\frac{p^2(\log n)^2}{32}} = o(1)$. So, we are done. □

Completing proof of Theorem 1.3. By using (5.1), (5.9), Lemma 5.3 and Lemma 5.4, the solutions of the integer program (1.1) in Theorem 1.3 for the graph $H$ (which was sampled from $G_{n,p}$) are given by

- If $N = qn + r$ with integers $q > 0$ and $0 ≤ r ≤ \gamma$, then

  $$x_n = q, x_1 = r, \text{ and all other } x_k = 0.$$ (5.13)

- If $N = qn + r$ with integers $q > 0$ and $\beta ≤ r ≤ n − 1$, then

  $$x_n = q, x_r = 1, \text{ and all other } x_k = 0.$$ (5.14)

- If $a(\gamma + 1) = (\gamma + 1)a(1)$, and $N = qn + r$ with integers $q > 0$ and $r = \gamma + 1$, then both (5.13) and (5.14) are solutions, as it was remarked after the statement of Theorem 1.3 in Section 1.

Now by using Theorem 1.4 and analyzing Algorithm 2 for $N$-vertex $H$-covered graphs like we did for Proposition 1.11, we can show the uniqueness of the extremal graphs, finishing the proof of Theorem 1.3. □
6 Concluding remarks

- It will be interesting to consider the problem of maximizing the number of independent sets of order $t > 2$ in an $n$-vertex $H$-covered graph. We have some initial observations that the structure of the optimal graph might be drastically different for $t > 2$. It seems to be a fairly different problem than the case of $t = 2$, which is one of the reasons we did not include any of the analysis in this paper.

- It is not hard to see that trees and cycles are ideal by a straightforward application of Theorem 1.4. It will be interesting to classify all graphs that are ideal. In Theorem 1.2, we have seen that $n$-vertex $d$-regular graphs with $d \geq \frac{n-1}{2}$ are ideal. It might be worth searching for other nice classes of ideal graphs.

- On the generalized problem of minimizing the number of copies of a graph $T$ in an $N$-vertex $H$-covered graph, it will be interesting to see what kind of arguments can be made when $T$ is not a complete graph. For example, does Corollary 1.5 still hold? The integer program like (1.1) in Theorem 1.4 will still give a lower bound, but may not be the correct answer because there is the possibility of extra copies of $T$ in $L^H_{T,N}$, i.e., in Definition 2 the graph $L^H_{T,N}$ may have some extra copies of $T$ having vertices both in $V_1$ and $V_2$, which did not happen when $T$ was complete.

- In the context of Theorem 1.2, it will be nice to investigate if the random $d$-regular graph $G_{n,d}$ for $d < \frac{n-1}{2}$ is ideal with high probability. It is not hard to notice that $G_{n,2}$ is not ideal w.h.p. by using the fact that $G_{n,2}$ is just the union of disjoint cycles of at least two different orders. It is worth mentioning that there exist connected $d$-regular graphs that are not ideal.

- An interesting task will be to consider Theorem 1.3 when $p$ is a function of $n$ in $G_{n,p}$. Although our proof extends for some range of $p$ tending to zero, we leave this open for further investigation.

References

[1] B. Bollobás, Random Graphs, Cambridge University Press 2001.

[2] B. Bollobás, Degree sequences of random graphs, Discrete Math. 33 (1981), 1–19.

[3] J. Cutler and A.J. Radcliffe, The maximum number of complete subgraphs of fixed size in a graph with given maximum degree, Journal of Graph Theory, 84(2) (2017), 134–145.

[4] J. Engbers and D. Galvin, Counting independent sets of a fixed size in graphs with a given minimum degree, Journal of Graph Theory 76(2) (2014), 149–168.

[5] P. Erdős, A. Hajnal, and J. W. Moon, A problem in graph theory, In Amer. Math. Monthly 71 (1964), 1107–1110.
[6] P. Erdős and A. Stone, On the structure of linear graphs, *Bulletin of the American Mathematical Society** 52 (1946), 1087–1091.

[7] J. R. Faudree, R. J. Faudree, and J. R. Schmitt, A Survey of Minimum Saturated Graphs, *The Electronic Journal of Combinatorics* 18 (2011), #DS19.

[8] A. Frieze and M. Karoński, Introduction to Random Graphs, *Cambridge University Press* 2015.

[9] Z. Füredi and M. Simonovits, The history of degenerate (bipartite) extremal graph problems, *Erdős centennial* 169–264, Bolyai Soc. Math. Stud., 25, Janós Bolyai Math. Soc., Budapest, 2013.

[10] W. Gan, P. Loh, and B. Sudakov, Maximizing the number of independent sets of a fixed size, *Combinatorics, Probability and Computing* 24 (2015), 521–527.

[11] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58 (1963), 13–30.

[12] D. Matula, The largest clique size in a random graph, *Technical Report, Department of Computer Science, Southern methodist University* (1976).

[13] A. Sidorenko, What we know and what we do not know about Turán numbers, *Graphs Combin.* 11(2) (1995), 179–199.

[14] P. Turán, On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok* 48 (1941), 436–452.

[15] Z. Tuza, Extremal problems on saturated graphs and hypergraphs, *Ars Combin.* 25(B) (1988), 105–113. Eleventh British Combinatorial Conference (London, 1987).