Minimum H-Decompositions of Graphs

Oleg Pikhurko*
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
Web: `http://www.math.cmu.edu/~pikhurko`

Teresa Sousa†
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213
E-mail: `tmj@andrew.cmu.edu`

April 25, 2006

Abstract

Given graphs G and H, an *H-decomposition* of G is a partition of the edge set of G such that each part is either a single edge or forms a graph isomorphic to H. Let $\phi_H(n)$ be the smallest number ϕ such that any graph G of order n admits an H-decomposition with at most ϕ parts.

Here we determine the asymptotic of $\phi_H(n)$ for any fixed graph H as n tends to infinity.

The exact computation of $\phi_H(n)$ for an arbitrary H is still an open problem. Bollobás [Math. Proc. Cambridge Philosophical Soc. **79** (1976) 19–24]

*Partially supported by the National Science Foundation, Grant DMS-0457512.
†Partially supported by the Portuguese Science Foundation under grant SFRH/BD/8617/2002.

accomplished this task for cliques. When H is bipartite, we determine $\phi_H(n)$ with a constant additive error and provide an algorithm returning the exact value with running time polynomial in $\log n$.

1 Introduction

Given two graphs G and H, an *H-decomposition* of G is a partition of the edge set of G such that each part is either a single edge or forms an H-subgraph, i.e., a graph isomorphic to H. We allow partitions only, that is, every edge of G appears in precisely one part. Let $\phi_H(G)$ be the smallest possible number of parts in an H-decomposition of G.

It is easy to see that, for non-empty H, $\phi_H(G) = e(G) - p_H(G)(e(H) - 1)$, where $p_H(G)$ is the maximum number of pairwise edge-disjoint H-subgraphs that can be packed into G and $e(G)$ denotes the number of edges in G. Building upon a body of previous research, Dor and Tarsi [5] showed that if H has a component with at least 3 edges then the problem of checking whether an input graph G is perfectly decomposable into H-subgraphs is NP-complete. Hence, it is NP-hard to compute the function $\phi_H(G)$ for such H.

Here we study the function

$$\phi_H(n) = \max\{\phi_H(G) \mid v(G) = n\},$$

which is the smallest number such that any graph G of order n admits an H-decomposition with at most $\phi_H(n)$ parts. Motivated by the problem of representing graphs by set intersections, Erdős, Goodman and Pósa [7] proved that $\phi_{K_3}(n) = t_2(n)$, where K_r denotes the complete graph (clique) of order r, and $t_r(n)$ is the maximum size of an r-partite graph on n vertices. This result was extended by Bollobás [3], who proved that

$$\phi_{K_r}(n) = t_{r-1}(n), \quad \text{for all } n \geq r \geq 3. \tag{1.1}$$

Here we determine the asymptotic of $\phi_H(n)$ for any fixed graph H as $n \to \infty$.

Theorem 1.1. *Let H be any fixed graph of chromatic number $r \geq 3$. Then,*

$$\phi_H(n) = t_{r-1}(n) + o(n^2).$$

The upper bound of Theorem 1.1 is proved in Section 2. The lower bound follows from the trivial inequalities $\phi_n(H) \geq \mathrm{ex}(n,H) \geq t_{r-1}(n)$, where

$$\mathrm{ex}(n,H) = \max\{e(G) \mid v(G) = n,\ H \not\subseteq G\}$$

is the *Turán function.* We make the following conjecture.

Conjecture 1.2. *For any graph H of chromatic number $r \geq 3$ there is $n_0 = n_0(H)$ such that $\phi_H(n) = \mathrm{ex}(n,H)$ for all $n \geq n_0$.*

This conjecture is known to be true for cliques (Bollobás [3]), clique-extensions (Sousa [16]), the cycle of length 5 and some other graphs (Sousa [15]).

For a bipartite graph H it is easy to determine the asymptotic (see Sousa [15]):

Lemma 1.3. *For any non-empty graph H with m edges and any integer n, we have*

$$\phi_H(n) \leq \frac{1}{m}\binom{n}{2} + \frac{m-1}{m}\,\mathrm{ex}(n,H). \tag{1.2}$$

In particular, if H is a fixed bipartite graph with m edges and $n \to \infty$, then

$$\phi_H(n) = \left(\frac{1}{m} + o(1)\right)\binom{n}{2}. \tag{1.3}$$

Proof. To prove (1.2) remove greedily one by one the edge-sets of H-subgraphs of a given graph G and then remove the remaining edges. The bound (1.2) follows as at most $\mathrm{ex}(n,H)$ parts are single edges.

The upper bound in (1.3) follows from (1.2) and the inequality

$$\mathrm{ex}(n,K_{t,t}) = O(n^{2-1/t}), \tag{1.4}$$

of Kővari, Sös and Turán [10]. The lower bound in (1.3) follows from $\phi_H(n) \geq \phi_H(K_n) \geq \frac{1}{m}\binom{n}{2}$. □

We managed to determine $\phi_H(n)$ for any fixed bipartite graph H with an $O(1)$ additive error (see Theorem 1.4 below). Furthermore, our proof gives a simple procedure for computing the exact values of $\phi_H(n)$ for all large n. Although, it should be possible to write a closed formula for the exact value of $\phi_H(n)$ for H bipartite, it seems to be too cumbersome so we do not attempt this here.

For a non-empty graph H, let $\gcd(H)$ denote the greatest common divisor of the degrees of H. For example, $\gcd(K_{6,4}) = 2$ while for any tree T with at least 2 vertices we have $\gcd(T) = 1$. We will prove the following result in Section 3.

Theorem 1.4. *Let H be a bipartite graph with m edges and let $d = \gcd(H)$. Then there is $n_0 = n_0(H)$ such that for all $n \geq n_0$ the following statements hold.*

If $d = 1$, then if $\binom{n}{2} \equiv m - 1 \pmod{m}$,

$$\phi_H(n) = \phi_H(K_n) = \left\lfloor \frac{n(n-1)}{2m} \right\rfloor + m - 1, \tag{1.5}$$

otherwise,

$$\phi_H(n) = \phi_H(K_n^*) = \left\lfloor \frac{n(n-1)}{2m} \right\rfloor + m - 2 \tag{1.6}$$

where K_n^* denotes any graph obtained from K_n after deleting at most $m - 1$ edges in order to have $e(K_n^*) \equiv m - 1 \pmod{m}$. Furthermore, if G is extremal then G is either K_n or K_n^*.

If $d \geq 2$, then

$$\phi_H(n) = \frac{nd}{2m} \left(\left\lfloor \frac{n}{d} \right\rfloor - 1 \right) + \frac{1}{2} n(d-1) + O(1). \tag{1.7}$$

Moreover, there is a procedure with running time polynomial in $\log n$ which determines $\phi_H(n)$ and describes a family $\mathcal{D}$ of n-sequences such that a graph G of order n satisfies $\phi_H(G) = \phi_H(n)$ if and only if the degree sequence of G belongs to $\mathcal{D}$. (It will be the case that $|\mathcal{D}| = O(1)$ and each sequence in $\mathcal{D}$ has $n - O(1)$ equal entries, so $\mathcal{D}$ can be described using $O(\log n)$ bits.)

2 H-Decompositions for a non-bipartite H

In this section we will prove the upper bound in Theorem 1.1. In outline, the proof is the following. First, we apply Szemerédi's Regularity Lemma [17] to the graph G that we want to decompose. The regularity partition of G gives us a weighted graph K with large but bounded number k of vertices. By generalizing the method of Bollobás [3] we decompose K into weighted copies of K_r and K_2 with aggregate weight at most $t_{r-1}(k) + o(k^2)$. Then, we split G into subgraphs that correspond to the cliques from the above decomposition of K. Finally, each of the obtained r-partite subgraphs of G is almost perfectly decomposed into copies of H by using the theorem of Pippenger and Spencer [11].

Before presenting the proof we need to introduce the tools.

Let $G = (V, E)$ be a graph and let A and B be two disjoint non-empty subsets of V. Let $e(A, B)$ denote the number of edges between A and B. The *density of* (A, B) is defined as

$$d(A, B) = \frac{e(A, B)}{|A||B|}.$$

For $\varepsilon > 0$ the pair (A, B) is said to be *ε-regular* if for every $X \subset A$ and $Y \subset B$ satisfying $|X| > \varepsilon|A|$ and $|Y| > \varepsilon|B|$ we have

$$|d(X, Y) - d(A, B)| < \varepsilon.$$

Theorem 2.1 (Regularity Lemma [17])**.** *For every $\varepsilon > 0$ and m there exist two integers $M(\varepsilon, m)$ and $N(\varepsilon, m)$ with the following property: for every graph $G = (V, E)$ with $n \geq N(\varepsilon, m)$ vertices there is a partition of the vertex set into $k + 1$ classes (clusters)*

$$V = V_0 \cup V_1 \cup \cdots \cup V_k$$

such that

(i) $m \leq k \leq M(\varepsilon, m)$,

(ii) $|V_0| < \varepsilon n$,

(iii) $|V_1| = |V_2| = \ldots = |V_k|$,

(iv) all, but at most εk^2, of the pairs (V_i, V_j), $1 \leq i < j \leq k$, are ε-regular. □

Let $\mathcal{H}$ be a *t-uniform hypergraph*, that is, every hyperedge of $\mathcal{H}$ contains exactly t vertices. If v and w are vertices of $\mathcal{H}$, the *codegree* of v and w, denoted by $\operatorname{codeg}(v, w)$, is the number of hyperedges in $\mathcal{H}$ containing both v and w.

We will need the following theorem of Pippenger and Spencer [11], see also Rödl [12]. By $a \pm c$ we mean a real between $a - c$ and $a + c$.

Theorem 2.2. *For every integer t and real $c_2 > 0$, there are $c_3 = c_3(t, c_2) > 0$ and $d_0 = d_0(t, c_2)$ such that for any $n \geq D \geq d_0$ the following holds.*

Every t-uniform hypergraph $\mathcal{H}$ on a set V of n vertices satisfying all of the following conditions

1. *for all vertices $x \in V$ but at most $c_3 n$ of them, $\deg(x) = (1 \pm c_3)D$;*

2. *for all $x \in V$, $\deg(x) \leq D/c_3$;*

3. *for any two distinct $x, y \in V$, $\mathrm{codeg}(x, y) < c_3 D$;*

contains a matching consisting of at least $(1 - c_2)n/t$ hyperedges. □

We will also need the following version of Turán's Theorem, see e.g. [3].

Theorem 2.3 (Turán's Theorem, Min-Degree Version)**.** *If in a graph with n vertices the degree of every vertex is greater than $\left\lfloor \frac{r-2}{r-1} n \right\rfloor$ then the graph contains a K_r.* □

A *weighted graph* is a graph K together with a *weight function* κ that assigns to each edge of K a real number between 0 and 1. By assigning weight 0 to all non-edges, we may assume that K is a complete graph. A *weighted K_r-decomposition* of K is a collection $A_1, \ldots, A_m$ of subsets of $[k]$ and positive reals $\alpha_1, \ldots, \alpha_m$, each A_i consisting of either a K_r or a K_2 such that for any distinct $i, j \in [k]$ we have $\kappa(ij) = \sum_{h: A_h \ni ij} \alpha_h$. The *total weight* of the decomposition is $\sum_{i=1}^{m} \alpha_i$. Thus we are allowed to remove weighted versions of K_r's and K_2's.

Lemma 2.4. *For any integer $r \geq 3$ and a positive real c_1, there are $c_2 > 0$ and k_0 such that any weighted graph K on $k \geq k_0$ vertices admits a weighted K_r-decomposition of total weight at most $t_{r-1}(k) + 2c_1 k^2$ in which every K_r has weight at least c_2.*

Proof. Our proof is built upon the ideas from Bollobás [3]. Given r and c_1 choose, in this order, small c_2, large f and large C.

We will be iteratively updating our weighted graph K, decreasing the edge-weights by a corresponding amount after the removal of any clique in the obvious way, until all edge-weights are zero. Also, we agree that if at any stage the current graph K has an edge ij of weight $\kappa(ij) < c_2$, then we immediately remove this edge (as a 2-clique). Since we do this at most $\binom{k}{2}$ times, the total weight of our decomposition will increase by at most $c_2\binom{k}{2} \le \frac{c_1}{2}\binom{k}{2}$.

Also, whenever we remove a K_r we take the maximal possible weight. Thus each K_r will have weight at least c_2, and the second condition of the lemma is automatically satisfied.

We use induction on k to prove the bound

$$t_{r-1}(k) + c_1k^2 + C, \tag{2.1}$$

on the total weight of our decomposition. If $k \le f$, then the required bound follows from the C term alone since $\binom{k}{2} \le C$. So assume that $k > f$. Let the *weighted degree* of a vertex x be $\kappa(x) = \sum_{y\in\Gamma(x)} \kappa(xy)$, where $\Gamma(x)$ denotes the neighborhood of x. Let x have the smallest weighted degree, call it γ. We want to decompose all edges incident to x.

If $\gamma \le t_{r-1}(k) - t_{r-1}(k-1) + c_1(2k-1)$, then we just remove all single edges at x and decompose the remaining graph of order $k-1$ by induction, obtaining (2.1) as required. So suppose that

$$\gamma > t_{r-1}(k) - t_{r-1}(k-1) + c_1(2k-1). \tag{2.2}$$

Let A consist of all y such that $\kappa(xy) > 0$. Let $\alpha = |A|$. As each edge-weight is at most 1, $\alpha \ge \gamma$. Let us greedily remove maximum weight K_r's through x. Suppose that the removed K_r's have total weight h. Let $B \subset A$ consist of those $y \in A$ for which we still have $\kappa(xy) > 0$. The weighted graph induced by B contains no K_{r-1}. Thus, by the min-degree version of Turán's Theorem, Theorem 2.3, and since each edge-weight is at most 1, for some $y \in B$ we must have $\kappa_B(y) \le \frac{r-3}{r-2}\beta$, where $\beta = |B|$ and

$$\kappa_B(y) = \sum_{z\in\Gamma(y)\cap B} \kappa(yz).$$

We have

$$\beta \ge \gamma - (r-1)h - c_2k,$$

since

$$\gamma = \kappa(x) \le \sum_{z\in B} \kappa(xz) + (r-1)h + c_2k \le \beta + (r-1)h + c_2k$$

and each edge-weight is at most 1. Moreover, those of the removed K_r's that contain y have total weight at most 1, again because each edge-weight is at most 1.

Since initially we had $\kappa(y) \geq \gamma$, we conclude that

$$\gamma \leq \frac{r-3}{r-2}\beta + k - \beta + r - 1 \leq k + r - 1 - \frac{\gamma - (r-1)h - c_2 k}{r-2}.$$

Thus,

$$h \geq \gamma - \frac{r-2}{r-1}k - r + 2 - \frac{c_2 k}{r-1},$$

and the total weight removed through x is at most

$$h + \gamma - (r-1)h = \gamma - (r-2)h \leq \gamma - (r-2)\left(\gamma - \frac{r-2}{r-1}k - r + 2 - \frac{c_2 k}{r-1}\right).$$

The right-hand side is a non-increasing function of γ (recall that $r \geq 3$), so it is maximized when γ attains equality in (2.2), giving at most

$$t_{r-1}(k) - t_{r-1}(k-1) + c_1(2k-1),$$

since $\gamma - \frac{r-2}{r-1}k - r + 2 - \frac{c_2 k}{r-1} \geq 0$ in view of $k > f$.

This proves the bound (2.1) by induction. The lemma clearly follows from (2.1). □

Let us return to Theorem 1.1.

Proof of the upper bound in Theorem 1.1. Let $c_0 > 0$ be arbitrary. We choose, in this order, sufficiently small $c_1 \gg \cdots \gg c_5 > 0$ and then let n_0 be sufficiently large. Let G be any graph of order $n \geq n_0$. We will show that $\phi_H(G) \leq t_{r-1}(n) + c_0 n^2$.

Apply the Regularity Lemma to G to find a $c_4/2$-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$ with $1/c_3 \leq k < 1/c_5$. Remove all edges inside parts, in non-regular pairs and in regular pairs of density less than c_1 — these will be removed as single edges. We removed at most $c_1 n^2 \ll c_0 n^2$ edges.

Let K be the weighted complete graph on $[k]$ where the weight $\kappa(ij)$ is the density of $G[V_i, V_j]$ (after the removals), where $G[V_i, V_j]$ denotes the bipartite graph on $V_i \cup V_j$ consisting of all edges of G between V_i and V_j. As $k \geq 1/c_3$ is large, by Lemma 2.4 we can find a weighted K_r-decomposition of K with total weight at most $t_{r-1}(k) + 2c_1k^2$, where each K_r has weight at least c_2. Let $A_1, \dots, A_m$ be all the K_r's with weights $\alpha_1, \dots, \alpha_m$ respectively.

Perform the following procedure for each pair ij with $\kappa(ij) > 0$. Let $p_{ij,l} = \alpha_l/\kappa(ij)$ for $l \in [m]$ and let $p_{ij,0} = 1 - \sum_{l=1}^m p_{ij,l} \ge 0$. Partition $G[V_i, V_j]$ into bipartite subgraphs $B_{ij,0}, \dots, B_{ij,m}$ with vertex sets $V_i \cup V_j$, where each edge of $G[V_i, V_j]$ is included into $B_{ij,l}$ with probability $p_{ij,l}$, independently of the other edges. For $1 \le l \le m$, the expected density of $B_{ij,l}$ is α_l if $ij \in A_l$ and 0 otherwise.

Let us call a bipartite graph $G[A, B]$ *(c, ε)-regular* if for every $X \subset A$ and $Y \subset B$ satisfying $|X| > \varepsilon|A|$ and $|Y| > \varepsilon|B|$ we have $|d(X, Y) - c| < \varepsilon$. For example, if a bipartite graph is (c, ε)-regular, then it is 2ε-regular (as defined in Section 2).

Claim 1. *With high probability for every i, j, l with $\kappa(ij) > 0$ and $ij \in A_l$ the graph $B_{ij,l}$ is (α_l, c_4)-regular.*

Proof. Recall that $a \pm c$ means a real between $a - c$ and $a + c$. Let $v = |V_i| = |V_j| \ge (1 - c_4/2)n/k$.

Fix any $U_i \subset V_i$ and $U_j \subset V_j$, each of size at least c_4v. By the $c_4/2$-regularity of $G[V_i, V_j]$, the pair U_i, U_j spans $(\kappa(ij) \pm c_4/2)|U_i||U_j|$ edges in G. The number of edges in $B_{ij,l}[U_i, U_j]$ has binomial distribution with parameters $(e(G[U_i, U_j]), p_{ij,l})$.

Using Chernoff's bound [4] we can bound the probability that the pair U_i, U_j violates the (α_l, c_4)-regularity by $e^{-\lambda v^2}$, where λ can be chosen to depend on c_4 only. (Recall that $\alpha_l \ge c_2$.) Hence, for fixed i, j, l, the expected number of pairs U_i, U_j violating the (α_l, c_4)-regularity is at most

$$(2^v)^2 e^{-\lambda v^2} = o(k^{-2}m^{-1}).$$

Markov's inequality implies the claim. □

Fix any choice of $B_{ij,l}$ satisfying the conclusions of Claim 1.

Claim 2. *Let $r \ge 3$ and $\chi(H) = r$. Let $c_2 \gg c_3 \gg c_4 \gg 1/v$. Let $\lambda > c_2$ and G' be an r-partite graph on $V_1 \cup \dots \cup V_r$ with each $|V_i| = v$ such that each $G'[V_i, V_j]$ is (λ, c_4)-regular. Then G' minus at most $c_2e(G')$ edges can be perfectly decomposed into edge disjoint copies of H.*

Proof. Fix a coloring $h : V(H) \to [r]$. Let H have t edges and s vertices.

We will apply Theorem 2.2 to the hypergraph $\mathcal{H}$ whose vertex set consists of all edges of G' and whose hyperedges are the edge-sets of (not necessarily induced) H-subgraphs of G' such that $x \in V(H)$ is embedded into $V_{h(x)}$. Thus $v(\mathcal{H}) = e(G') = (\lambda \pm c_4)v^2\binom{r}{2}$. Let

$$D = v^{s-2}\lambda^{t-1}.$$

First, let us briefly recall the standard argument for counting vertex-labeled H-subgraphs, see e.g. Simonovits and Sós [14, Theorem 5]. It is slightly modified to better suit our purpose. Arbitrarily order the vertices of H as $x_1, \dots, x_s$. For $i \in [s]$ let $U_{i,1} = V_{h(x_i)}$, where $h : V(H) \to [r]$ is the fixed r-coloring of H. We will be constructing the embedding $f : V(H) \to V(G')$ one by one as follows. Suppose we have already embedded $x_1, \dots, x_{j-1}$ and have the current potential sets $U_{1,j}, \dots, U_{s,j}$ where $U_{i,j} = \{f(x_i)\}$ for $i = 1, \dots, j-1$. We are about to embed x_j. For $i > j$ with $x_jx_i \in E(H)$ let the *bad* set $B_{j,i}$ consist of all vertices $x \in U_{j,j}$ such that $|\Gamma(x) \cap U_{i,j}| \neq (\lambda \pm c_4)|U_{i,j}|$. (For all other i's, we let $B_{j,i} = \emptyset$ for convenience.)

If we assume that

$$|U_{i,j}| \geq c_4 v, \tag{2.3}$$

then $|B_{j,i}| \leq 2c_4v$. Indeed, let X (resp. Y) consist of those $x \in U_{j,j}$ that have more than $(\lambda + c_4)|U_{i,j}|$ (resp. less than $(\lambda - c_4)|U_{i,j}|$) neighbors in $U_{i,j}$. The (λ, c_4)-regularity of $G'[V_{h(x_i)}, V_{h(x_j)}]$ implies that $|X| \leq c_4v$ and $|Y| \leq c_4v$. Since $B_{j,i} = X \cup Y$, the claim follows.

Hence, in total there are at most $2c_4sv$ bad vertices in $U_{j,j}$. For $f(x_j)$ choose any vertex of $U_{j,j}$ that is not bad. Update:

$$U_{i,j+1} = \begin{cases} \{f(x_i)\}, & i \leq j, \\ U_{i,j} \setminus \{f(x_j)\}, & i > j \text{ and } x_jx_i \notin E(H), \\ (U_{i,j} \setminus \{f(x_j)\}) \cap \Gamma(f(x_j)), & i > j \text{ and } x_jx_i \in E(H). \end{cases}$$

For any $i > j$ we have $|U_{i,j+1}| \geq (\lambda - c_4)^t v \geq c_4v$, so (2.3) and all above estimates are valid by induction on j.

Recall that $c_4 \ll c_3 \ll \lambda$. Rather crudely, it follows that the number of the above embeddings is

$$(\lambda \pm c_4 \pm 2c_4s)^t(v \pm 2c_4sv)^s = (1 \pm c_3)v^s\lambda^t.$$

In all other embeddings that preserve the coloring h, we have to use a *bad* vertex (that is, a vertex in a bad set given the fixed ordering $x_1, \ldots, x_s$) at least once. Hence, the number of the remaining embeddings is at most

$$2c_4 s^2 v^s \ll (1 \pm c_3) v^s \lambda^t.$$

Now call an edge xy, with say $x \in V_i$ and $y \in V_j$, of G' *good* if

- x has $(\lambda \pm c_4)(v-1)$ neighbors in $V_j \setminus \{y\}$,
- y has $(\lambda \pm c_4)(v-1)$ neighbors in $V_i \setminus \{x\}$,
- for any $g \in [r] \setminus \{i, j\}$, each of x, y has $(\lambda \pm c_4)v$ neighbors in V_g while their common neighborhood in V_g has size $(\lambda \pm c_4)^2 v$.

The above argument gives that all but at most

$$\binom{r}{2} \left(2c_4 v(r-1) \times v + v \times 2c_4(2r-3)\right) < c_3 e(G')$$

edges of G' are good and that any good edge belongs to $(1 \pm c_3) v^{s-2} \lambda^{t-1} = (1 \pm c_3) D$ vertex-labelled copies of H. This shows that $\mathcal{H}$ satisfies Condition (1.) of Theorem 2.2.

For any edge, there are at most $v^{s-2} < D/c_3$ H-subgraphs containing it. For any two edges, there are at most $v^{s-3} < c_3 D$ H-subgraphs containing both of them. Hence, all assumptions of Theorem 2.2 are satisfied.

Therefore $\mathcal{H}$ contains a matching consisting of at least $(1-c_2)v(\mathcal{H})/t$ hyperedges, that is, our graph G' contains at least $(1-c_2)e(G')/t$ edge disjoint copies of H. We are left with at most $c_2 e(G')$ edges of G' not decomposed. So Claim 2 holds. $\square$

This shows that for each $l \in [m]$, we can find at least

$$(1-c_2)\alpha_l \binom{r}{2} / t \times ((1 - c_4/2)n/k)^2$$

pairwise edge disjoint H-subgraphs in B_l, where B_l is the union of bipartite graphs $B_{ij,l}$, $ij \in \binom{[k]}{2}$. All the remaining edges of our graph G are removed one by one as single edges.

Let $\alpha = \sum_{i=1}^{m} \alpha_i$ and $\kappa(K) = \sum_{ij \in E(K)} \kappa(ij)$. Since $t \geq \binom{r}{2}$, $e(G) \leq \kappa(G)n^2/k^2 + c_1 n^2$, and $\alpha + \kappa(K) - \binom{r}{2}\alpha \leq t_{r-1}(k) + 2c_1 k^2$, the total number of parts in our decomposition of G is at most

$$\begin{aligned}
&\alpha(1-c_2)\binom{r}{2}\frac{n^2}{tk^2} + e(G) - t\alpha(1-c_2)\binom{r}{2}\frac{n^2}{tk^2} \\
&\leq \left(\alpha - \binom{r}{2}\alpha + \kappa(K) + (t-1)c_2\alpha\right)\frac{n^2}{k^2} + c_1 n^2 \\
&\leq \left(t_{r-1}(k) + 2c_1 k^2\right)\frac{n^2}{k^2} + 2c_1 n^2 \\
&\leq t_{r-1}(n) + c_0 n^2
\end{aligned}$$

as required. This finishes the proof of Theorem 1.1. □

3 H-decompositions for a bipartite H

In this section we will prove Theorem 1.4. Before we start with the proof, we provide some auxiliary results.

Lemma 3.1. *For any bipartite graph H with bipartition (V_1, V_2) and any $A \subset V_1$ with $a \geq 1$ elements, there are integers C and n_0 such that the following holds. In any graph G of order $n \geq n_0$ with minimum degree $\delta(G) \geq \frac{2}{3}n$ there is a family of edge disjoint copies of H such that the vertex subsets corresponding to $A \subset V(H)$ are disjoint and cover all but at most C vertices of G. One can additionally ensure that each vertex of G belongs to at most $3(v(H))^2$ copies of H.*

Proof. Let $|V_1| = h_1$, $|V_2| = h_2$ and let $t = 2\lceil h_1/a \rceil h_2 a$. Let K be the complete 3-partite graph with t vertices in each color class. Let n_0 be sufficiently large. Let G be a graph with $n \geq n_0$ vertices and minimum degree at least $\frac{2}{3}n$.

A theorem of Shokoufandeh and Zhao [13] (see also Alon and Yuster [2] and Komlós, Sárközy, and Szemerédi [9]) implies that, in G, we can find vertex disjoint K-subgraphs covering all but at most C vertices, where C is a constant. Therefore, it suffices to prove that K contains $3t/a$ edge disjoint copies of H having vertex disjoint sets A.

Claim. The complete bipartite graph $K_{t,t}$ contains t/a edge disjoint copies of H with vertex disjoint sets A in one part.

Proof of Claim. Let (X,Y) be a bipartition of $K_{t,t}$. For $1 \le i \le t/a$ define $X_i = \{(i-1)a+1, \dots, (i-1)a+h_1\}$ and $A_i = \{(i-1)a+1, \dots, ia\}$ where the elements are taken modulo t.

Consider the graph $\mathcal{G}$ with vertex set $X_1, \dots, X_{t/a}$ and $\{X_i, X_j\}$ is an edge if and only if $X_i \cap X_j \neq \emptyset$. For $i = 1, \dots, t/a$, $\deg X_i$ is at most the number of other sets, not equal to X_i, that contain an endpoint of the interval X_i. Thus, $\Delta(\mathcal{G}) \le 2\left(\lceil h_1/a \rceil - 1\right)$. Color the vertices of $\mathcal{G}$ using at most $\Delta(\mathcal{G})+1$ colors.

Let $I_1, \dots, I_{t/h_2}$ be disjoint subsets of Y of size h_2. We pair all color-k vertices of $\mathcal{G}$ with I_k. All X_i get paired since the number of colors is at most t/h_2. Observe that a pair X_i and I_j induces a copy of K_{h_1,h_2}. Inside this graph choose an arbitrary H-subgraph so that $A_i \subset X_i$ corresponds to $A \subset V_1$. Since I_j is paired with pairwise disjoint subsets of X, the obtained copies of H are edge disjoint. This completes the proof of the claim. $\square$

Returning to the proof of the lemma, let (X,Y,Z) be a 3-partition of K. Apply the Claim to the complete bipartite graphs with bipartitions (X,Y), (Y,Z) and (Z,X). To complete the proof observe that each vertex of K appears in at most

$$2\left\lceil \frac{h_1}{a} \right\rceil + \frac{t}{a} \le 2\left\lceil \frac{h_1}{a} \right\rceil + 2h_2\left\lceil \frac{h_1}{a} \right\rceil \le 2v(H) + 2(v(H))^2 \le 3(v(H))^2$$

copies of H. $\square$

The following results appearing in Alon, Caro and Yuster [1, Theorem 1.1, Corollary 3.4, Lemma 3.5] which follow with some extra work from the powerful decomposition theorem of Gustavsson [8], are crucial to the proof of Theorem 1.4.

Recall that for a non-empty graph H, $\gcd(H)$ denotes the greatest common divisor of the degrees of H.

Lemma 3.2. *For any non-empty graph H with m edges, there are $\gamma > 0$ and N_0 such that the following holds. Let $d = \gcd(H)$. Let G be a graph of order $n \ge N_0$ and of minimum degree $\delta(G) \ge (1-\gamma)n$.*

If $d = 1$, then

$$p_H(G) = \left\lfloor \frac{e(G)}{m} \right\rfloor. \tag{3.1}$$

If $d \geq 2$, let $\alpha_u = d \lfloor \frac{\deg(u)}{d} \rfloor$ for $u \in V(G)$ and let X consist of all vertices whose degree is not divisible by d. If $|X| \geq \frac{n}{10d^3}$, then

$$p_H(G) = \left\lfloor \frac{1}{2m} \sum_{u \in V(G)} \alpha_u \right\rfloor. \tag{3.2}$$

If $|X| < \frac{n}{10d^3}$, then

$$p_H(G) \geq \frac{1}{m} \left(e(G) - \frac{n}{5d^2} \right). \tag{3.3}$$

□

Proof of Theorem 1.4. Given H, let $\gamma(H)$ and N_0 be given by Lemma 3.2. Assume that $\gamma \leq \gamma(H)$ is sufficiently small and that $n_0 \geq N_0$ is sufficiently large to satisfy all the inequalities we will encounter. Let $n \geq n_0$ and let G be any graph of order n with $\phi_H(G) = \phi_H(n)$.

Let $G_n = G$. Repeat the following at most $\lfloor n/\log n \rfloor$ times: If the current graph G_i has a vertex x_i of degree at most $(1-\gamma/2)i$, let $G_{i-1} = G_i - x_i$ and decrease i by 1.

Suppose we stopped after s repetitions. Then, either $\delta(G_{n-s}) \geq (1-\gamma/2)(n-s)$ or $s = \lfloor n/\log n \rfloor$. Let us show that the later cannot happen. Otherwise, we have

$$e(G) \leq \binom{n-s}{2} + \left(1 - \frac{\gamma}{2}\right) \sum_{i=n-s+1}^{n} i < \binom{n}{2} - \frac{\gamma n^2}{4 \log n}. \tag{3.4}$$

Let t satisfy $K_{t,t} \supset H$. Using (1.2), (1.4), and (3.4) we obtain

$$\phi_H(G) \leq \frac{1}{m} \left(\binom{n}{2} - \frac{\gamma}{4} \frac{n^2}{\log n} \right) + \frac{m-1}{m} c n^{2-1/t} < \frac{1}{m} \binom{n}{2} \ \leq\ \phi_H(K_n),$$

which contradicts our assumption on G. Therefore, $s < \lfloor n/\log n \rfloor$ and we have $\delta(G_{n-s}) \geq (1-\gamma/2)(n-s)$.

Let $\alpha = 2\gamma$. We will have another pass over the vertices $x_n, \ldots, x_{n-s+1}$, each time decomposing the edges incident to x_i by H-subgraphs and single edges. It will be the case that each time we remove the edges incident to the current vertex x_i, the degree of any other vertex drops by at most $3h^4$, where $h = v(H)$. Here is a formal description. Initially, let $G'_n = G$ and $i = n$. If in the current graph G'_i we have $\deg_{G'_i}(x_i) \le \alpha n$, then we remove all G'_i-edges incident to x_i as single edges and let $G'_{i-1} = G'_i - x_i$.

Suppose that $\deg_{G'_i}(x_i) > \alpha n$. Then, the set

$$X_i = \{y \in V(G_{n-s}) : x_i y \in E(G'_i)\},$$

has at least $\alpha n - s + 1$ vertices. The minimum degree of $G[X_i]$ is

$$\delta(G[X_i]) \ge |X_i| - s - \frac{\gamma n}{2} - s \times 3h^4 \ge \frac{2}{3}|X_i|.$$

Let $y \in V(H)$, $A = \Gamma_H(y)$ and $a = |A|$. By Lemma 3.1 there is a constant C such that all but at most C vertices of $G[X_i]$ can be covered by edge disjoint copies of $H - y$ each of them having vertex disjoint sets A. Therefore, all but at most C edges between x_i and X_i can be decomposed into copies of H. All other edges incident to x_i are removed as single edges. Let G'_{i-1} consist of the remaining edges of $G'_i - x_i$ (that is, those edges that do not belong to an H-subgraph of the above x_i-decomposition). This finishes the description of the case $\deg_{G'_i}(x_i) > \alpha n$.

Consider the sets $S = \{x_n, \dots, x_{n-s+1}\}$, $S_1 = \{x_i \in S : \deg_{G'_i}(x_i) \le \alpha n\}$, and $S_2 = S \backslash S_1$. Let their sizes be s, s_1, and s_2 respectively, so $s = s_1 + s_2$.

Let F be the graph with vertex set $V(G_{n-s}) \cup S_2$, consisting of the edges coming from the removed H-subgraphs when we processed the vertices in S_2. We have

$$\phi_H(G) \le \phi_H(G'_{n-s}) + \frac{e(F)}{m} + s_1 \alpha n + s_2 C + \binom{s}{2}. \tag{3.5}$$

We know that $\phi_H(G'_{n-s}) = e(G'_{n-s}) - p_H(G'_{n-s})(m-1)$. The last statement of Lemma 3.1 guarantees that $\delta(G'_{n-s}) \ge (1-\gamma)(n-s)$. Thus, $p_H(G'_{n-s})$ can be estimated using Lemma 3.2.

Consider first the case $d = 1$. Using the inequalities $e(F) \le (1 - \gamma/2)ns_2$ and $\alpha \le (2-\gamma)/2m$, we obtain

$$\begin{aligned}
\phi_H(G) &\le e(G'_{n-s}) - \left\lfloor \frac{e(G'_{n-s})}{m} \right\rfloor (m-1) + \frac{e(F)}{m} + s_1 \alpha n + s_2 C + \binom{s}{2} \\
&\le \left(\frac{1}{m} \binom{n-s}{2} + m - 1 \right) + \frac{2-\gamma}{2m} s_2 n + s_1 \alpha n + s_2 C + \binom{s}{2} \\
&\le \frac{1}{m} \binom{n}{2} - \frac{(n-1)s}{m} + \frac{s(s-1)}{2m} + \frac{2-\gamma}{2m} sn + \binom{s}{2} + m - 1
\end{aligned}$$

Observe that, if $S \neq \emptyset$ then, in view $1/m > (2-\gamma)/2m$, we have

$$\frac{s}{m} + \frac{s(s-1)}{2m} + \binom{s}{2} + m - 1 < \left(\frac{1}{m} - \frac{2-\gamma}{2m}\right) ns = \frac{\gamma}{2m} ns,$$

since $s < \frac{n}{\log n}$ and n is sufficiently large. Therefore, $\phi_H(G) < \frac{1}{m}\binom{n}{2} \leq \phi_H(K_n)$ which contradicts our assumption on G. Thus, $S = \emptyset$ and

$$\phi_H(G) = e(G) - (m-1)\left\lfloor \frac{e(G)}{m} \right\rfloor, \tag{3.6}$$

is a function of $e(G)$ alone. By the optimality of G we cannot increase the right-hand side of (3.6) by increasing $e(G)$ by 1 or by m. Thus $e(G)$ is $\binom{n}{2}$ or the largest integer below $\binom{n}{2}$ congruent to $m-1$ modulo m. (In fact, the optimal value for $e(G)$ is unique unless $m = 2$ and $\binom{n}{2}$ is even when both of the above values give the maximum.) This proves the theorem for the case $d = 1$.

Consider the case $d \geq 2$. To prove the lower bound in (1.7) we consider a graph L of order $n \geq n_0$, which is r-regular (except at most one vertex of degree $r-1$) where $r \in [n-d, n-1]$ has residue $d-1$ modulo d. (Such a graph L exists, which can be seen either directly or from Erdös and Gallai's result [6].)

Let $r = qd + d - 1$. Then

$$\phi_H(L) = e(L) - p_H(L)(m-1) \geq \frac{1}{2}n(qd+d-1) - \frac{1}{2} - \frac{ndq}{2m}(m-1),$$

giving the required lower bound in view of $q = \lfloor n/d \rfloor - 1$.

We will now prove the upper bound in (1.7).

Assume first that (3.3) holds. Then, by (3.5)

$$\begin{aligned}
\phi_H(G) &\leq e(G'_{n-s}) - \frac{1}{m}\left(e(G'_{n-s}) - \frac{n-s}{5d^2}\right)(m-1) + \frac{e(F)}{m} + s_1\alpha n + s_2 C + \binom{s}{2} \\
&\leq \frac{1}{m}\binom{n-s}{2} + \frac{m-1}{m}\frac{n-s}{5d^2} + \frac{2-\gamma}{2m}s_2 n + s_1\alpha n + s_2 C + \binom{s}{2} \\
&\leq \frac{1}{m}\binom{n}{2} - \frac{(n-1)s}{m} + \frac{s(s-1)}{2m} + \frac{m-1}{m}\frac{n-s}{5d^2} + \frac{2-\gamma}{2m}sn + s_2 C + \binom{s}{2}.
\end{aligned}$$

For $s > \frac{2(m-1)}{5\gamma d^2}$ we have $\frac{\gamma}{2m} - \frac{m-1}{5md^2 s} > 0$. Thus, for n sufficiently large

$$\frac{s}{m} + \frac{s(s-1)}{2m} - \frac{m-1}{m}\frac{s}{5d^2} + \binom{s}{2} + s_2 C < \left(\frac{1}{m} - \frac{2-\gamma}{2m} - \frac{m-1}{5md^2 s}\right) ns.$$

That is, $\phi_H(G) < \frac{1}{m}\binom{n}{2} \le \phi_H(K_n)$ which contradicts the optimality of G. Otherwise, s is bounded by a constant independent of n, and the terms of order n^2 and n alone give us the contradiction $\phi_H(G) < \phi_H(L)$, where L is the (almost) r-regular graph from the lower bound on $\phi_H(n)$. In fact, the coefficient of sn is $-\frac{1}{m} + \frac{2-\gamma}{2m} < 0$, so to get a contradiction it is enough to show

$$\frac{1}{m}\binom{n}{2} + \frac{n}{5d^2} \le \frac{nd}{2m}\left(\frac{n}{d} - 2\right) + \frac{1}{2}n(d-1),$$

that is,

$$\frac{n}{5d^2} \le \frac{1-2d}{2m}n + \frac{1}{2}n(d-1).$$

The worst case is when $m = 4$ (note $m \ge 4$ since $d \ge 2$). Therefore, it suffices to show that

$$\frac{8n}{5d^2} \le (2d-3)n,$$

which holds as $d \ge 2$.

Finally, assume that (3.2) holds. It follows that $p_H(G)$ and thus $\phi_H(G)$, depends only on the degree sequence $d_1, \ldots, d_n$ of G. Namely, the packing number $\ell = p_H(G)$ equals $\lfloor \frac{1}{2m}\sum_{i=1}^n r_i \rfloor$, where $r_i = d\lfloor d_i/d \rfloor$ is the largest multiple of d not exceeding d_i.

Thus, is enough for us to prove the upper bound in (1.7) on $\phi_{\max}$, the maximum of

$$\phi(d_1, \ldots, d_n) = \frac{1}{2}\sum_{i=1}^n d_i - (m-1)\left\lfloor \frac{1}{2m}\sum_{i=1}^n \left\lfloor \frac{d_i}{d} \right\rfloor d \right\rfloor, \tag{3.7}$$

over all (not necessarily graphical) sequences $d_1, \ldots, d_n$ of integers with $0 \le d_i \le n-1$.

Let $d_1, \ldots, d_n$ be an optimal sequence attaining the value $\phi_{\max}$. For $i = 1, \ldots, n$ let $d_i = q_i d + r_i$ with $0 \le r_i \le d-1$. Then, $\ell = \left\lfloor \frac{(q_1+\cdots+q_n)d}{2m} \right\rfloor$.

Let $n = qd + r$ with $0 \le r \le d-1$ and $q = \lfloor n/d \rfloor$. Define $R = qd - 1$ to be the maximum integer which is at most $n-1$ and is congruent to $d-1$ modulo d. Let $C_1 = \{i \in [n] : r_i = d-1 \text{ and } d_i < R\}$ and $C_2 = \{i \in [n] : d_i = n-1\}$ if $n - 1 \ne R$ and $C_2 = \emptyset$ otherwise.

Since $d_1, \ldots, d_n$ is an optimal sequence, we have that if $r_i \neq d-1$ then $d_i = n-1$ for all $i \in [n]$. Also, $|C_1| \leq \frac{2m}{d} - 1$ and $|C_2| \leq 2m-1$. We have

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^{n} d_i &= \frac{1}{2}(n - |C_1 \cup C_2|)R + \frac{1}{2}\sum_{i \in C_1} d_i + \frac{1}{2}|C_2|(n-1) \\
&\leq \frac{1}{2}nd(q-1) + \frac{1}{2}n(d-1) - \frac{d}{2}\sum_{i \in C_1}(q-1-q_i) + O(1), \\
\ell &\geq \left(\frac{1}{2m}\sum_{i=1}^{n}\left\lfloor\frac{d_i}{d}\right\rfloor d\right) - 1 \\
&\geq \frac{1}{2m}nd(q-1) - \frac{d}{2m}\sum_{i \in C_1}(q-1-q_i) + O(1).
\end{aligned}$$

These estimates give us the required bound:

$$\phi_{\max} = \frac{1}{2}\sum_{i=1}^{n} d_i - (m-1)\ell \leq \frac{1}{2m}nd(q-1) + \frac{1}{2}n(d-1) + O(1). \tag{3.8}$$

If we want to compute the function $\phi_H(n)$ exactly we proceed as follows. From the obtained lower and upper bounds it follows that $\delta(G) \geq n - O(1)$ and $|C_1 \cup C_2| = O(1)$. Our algorithm generates all such sequences, eliminates those which are not graphical using the characterization of Erdős and Gallai [6] and computes $\phi(d_1, \ldots, d_n)$ using (3.7).

To finish the proof it remains to obtain a contradiction if $S \neq \emptyset$ holds. Let $\bar{d}_1, \ldots, \bar{d}_n$ be the degree sequence of the graph with vertex set $V(G)$ and edge set $E(G'_{n-s}) \cup E(F)$. Consider the new sequence of integers

$$d'_i = \begin{cases} \bar{d}_i, & \text{if } x_i \notin S, \\ \bar{d}_i + \left\lceil \frac{(1-3\gamma)}{m} n \right\rceil m, & \text{if } x_i \in S_1, \\ \bar{d}_i + \left\lceil \frac{\gamma}{4m} n \right\rceil m, & \text{if } x_i \in S_2. \end{cases}$$

Each d'_i lies between 0 and $n-1$, so $\phi(d'_1, \ldots, d'_n) \leq \phi_{\max}$. We obtain

$$\begin{aligned}
\phi_H(G) &\leq \phi(\bar{d}_1, \ldots, \bar{d}_n) + s_1\alpha n + s_2 C + \binom{s}{2} \\
&< \phi(d'_1, \ldots, d'_n) - \frac{1-3\gamma}{2m}s_1 n - \frac{\gamma}{8m}s_2 n + s_1\alpha n + s_2 C + \binom{s}{2} \\
&\leq \phi_{\max} - \frac{\gamma}{10m}sn,
\end{aligned}$$

which contradicts the already established facts that the right-hand side of (1.7) is at most $\phi_H(G)$ by the optimality of G and is at least $\phi_{\max}$ by (3.8). $\square$

References

[1] N. Alon, Y. Caro, and R. Yuster, *Packing and covering dense graphs*, J. Combin. Designs **6** (1998), 451–472.

[2] N. Alon and R. Yuster, *H-factors in dense graphs*, J. Combin. Theory (B) **66** (1996), 269–282.

[3] B. Bollobás, *On complete subgraphs of different orders*, Math. Proc. Camb. Phil. Soc. **79** (1976), 19–24.

[4] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.*, Ann. Math. Statistics **23** (1952), 493–507.

[5] D. Dor and M. Tarsi, *Graph decomposition is NP-complete: a complete proof of Holyer's conjecture*, SIAM J. Computing **26** (1997), 1166–1187.

[6] P. Erdős and T. Gallai, *Graphs with prescribed degree of vertices*, Mat. Lapok **11** (1960), 264–274.

[7] P. Erdős, A. W. Goodman, and L. Pósa, *The representation of a graph by set intersections*, Can. J. Math. **18** (1966), 106–112.

[8] T. Gustavsson, *Decompositions of large graphs and digraphs with high minimum degree*, Ph.D. thesis, Univ. of Stockholm, 1991.

[9] J. Komlós, G. N. Sárkőzy, and E. Szemerédi, *Proof of the Alon-Yuster conjecture*, Discrete Math. **235** (2001), 255–269.

[10] P. Kővari, V. T. Sös, and P. Turán, *On a problem of K. Zarankiewicz*, Colloq. Math. **3** (1954), 50–57.

[11] N. Pippenger and J. Spencer, *Asymtotic behavior of the chromatic index for hypergraphs*, J. Combin. Theory (A) **51** (1989), 24–42.

[12] V. Rödl, *On a packing and covering problem*, Europ. J. Combin. **5** (1985), 69–78.

[13] A. Shokoufandeh and Y. Zhao, *Proof of a tiling conjecture of Komlós*, Random Struct. Algorithms **23** (2003), 180–205.

[14] M. Simonovits and V. T. Sós, *Szemerédi's partition and quasirandomness*, Random Struct. Algorithms **2** (1991), 1–10.

[15] T. Sousa, *Decompositions of graphs into 5-cycles and other small graphs*, Electronic J. Combin. **12** (2005), 7pp.

[16] T. Sousa, *Decompositions of graphs into a given clique-extension*, Submitted, 2005.

[17] E. Szemerédi, *Regular partitions of graphs*, Proc. Colloq. Int. CNRS, Paris, 1976, pp. 309–401.