Decomposing large graphs with small graphs of high density

Raphael Yuster *
Department of Mathematics
University of Haifa-ORANIM
Tivon 36006, Israel

Abstract

It is shown that for every positive integer h, and for every $\epsilon > 0$, there are graphs $H = (V_H, E_H)$ with at least h vertices and with density at least $0.5 - \epsilon$ with the following property: If $G = (V_G, E_G)$ is any graph with minimum degree at least $\frac{|V_G|}{2}(1 + o(1))$ and $|E_H|$ divides $|E_G|$ then G has an H-decomposition. This result extends the results of Wilson [8], Gustavsson [6] and Yuster [9].

1 Introduction

All graphs considered here are finite, undirected, and have no loops or multiple edges. For the standard graph-theoretic notations the reader is referred to [2]. Let H be a graph without isolated vertices. An *H-packing* of a graph G is a set $L = \{G_1, \ldots, G_s\}$ of edge-disjoint subgraphs of G, where each subgraph is isomorphic to H. The *H-packing number* of G, denoted by $P(H, G)$, is the maximum cardinality of an H-packing of G. G has an *H-decomposition* if it has an H-packing with the property that every edge of G appears in *exactly* one member of the H-packing. Note that in order for G to have an H-decomposition, two necessary conditions must hold. The first is that $e(H)$ divides $e(G)$. The second is that $gcd(H)$ divides $gcd(G)$ where the *gcd* of a graph is the greatest common-divisor of the degrees of its vertices. Note that for any pair of graphs G and H, we can verify in polynomial time if G satisfies these two conditions. We call these conditions the "necessary H-decomposition conditions".

The combinatorial and computational aspects of the H-packing and H-decomposition problems have been studied extensively. Wilson in [8] has proved that if $G = K_n$ where $n \geq n_0 = n_0(H)$, and K_n satisfies the necessary H-decomposition conditions, then K_n has an H-decomposition.

*e-mail: raphy@macam98.ac.il   World Wide Web: http:\\www.math.tau.ac.il\~raphy

Recently, the H-packing problem for $G = K_n$ ($n \geq n(H)$) was solved [3], by giving a closed formula for computing $P(H, K_n)$. In case the graph G is not complete, it is known that the H-decomposition and H-packing problems are, in general, NP-Hard, since Dor and Tarsi [5] showed that deciding if G has an H-decomposition is NP-Complete, where H is any fixed connected graph with at least three edges. In view of Wilson's positive result, and the Dor-Tarsi negative result, the following extremal problem is naturally raised:

Problem 1: Determine $f_H(n)$, the smallest possible integer, such that whenever G has n vertices, and $\delta(G) \geq f_H(n)$, and G satisfies the necessary H-decomposition conditions, then G has an H-decomposition.

Wilson's result proves that $f_H(n)$ exists, for $n \geq n_0(H)$, or, in other words, $f_H(n) \leq n - 1$ for $n \geq n_0(H)$. It turns out that estimating $f_H(n)$ is extremely difficult for general H. The first, and only, nontrivial general upper bound for $f_H(n)$ was obtained in 1991 by Gustavsson [6]. He has shown that if $\delta(G) \geq (1 - \epsilon(H))n$, where $\epsilon(H)$ is some small positive constant depending on H, and G satisfies the necessary H-decomposition conditions, then G has an H-decomposition. In other words, $f_H(n) \leq (1 - \epsilon(H))n$, for n sufficiently large. Unfortunately, the $\epsilon(H)$ in Gustavsson's result is a very small number. For example, if H is a triangle then $\epsilon(H) \leq 10^{-24}$. In general, $\epsilon(H) \leq 10^{-24}/|H|$ (in fact, it is much smaller). It is believed, however, that the correct value for $f_H(n)$ is much smaller. In fact, Nash-Williams conjectured in [7] that when H is a triangle, then $f_H(n) \leq \lceil 3n/4 \rceil$, and he also gives an example showing that this would be be best possible, and thus his conjecture is that $f_H(n) = \lceil 3n/4 \rceil$. However, the best result still known for triangles is Gustavsson's asymptotic result.

The first significant improvement over Gustavsson's result was obtained by the author in [9] in case the graph H is a tree (or a forest). It is shown there that $f_H(n) \leq n/2 + h^4\sqrt{n \log n}$, where h is the number of vertices of the tree H. This result is asymptotically best possible as it is also shown in [9] that $f_H(n) \geq \lfloor n/2 \rfloor - 1$ for every connected graph H with at least 3 vertices.

Prior to this article, trees are the only graphs for which $f_H(n)/n$ is asymptotically known. If H has a cycle, then the best estimate that was known is Gustavsson's result. The purpose of this paper is to construct a family of graphs which are much more dense than trees, for which $f_H(n)/n$ can also be asymptotically determined. Recall that the density of a graph H with m edges and n vertices is $d(H) = m/\binom{n}{2}$. Thus, trees on n vertices have very low density, namely $2/n$, while complete graphs have the maximum possible density, namely 1. We will show that there is an infinite family of graphs with the property that for every $\epsilon > 0$ and for every positive integer h, there is a graph H in the family with at least h vertices and with density at least $0.5 - \epsilon$, for which $f_H(n)/n \to 0.5$ when $n \to \infty$. We summarize our exact result in the following Theorem:

Theorem 1.1 *Let h be a positive integer, and let $\epsilon > 0$. There exists a graph $H = (V_H, E_H)$ with $|V_H| \geq h$ and with $d(H) \geq 0.5 - \epsilon$ having $f_H(n)/n \to 0.5$ when $n \to \infty$.*

Theorem 1.1 can be extended easily to show that if $G = (V_G, E_G)$ has minimum degree $\frac{n}{2}(1 + o(1))$, but does not satisfy the necessary H-decomposition conditions then there is an optimal packing in the sense that $P(H, G) = \lfloor |E_G|/|E_H| \rfloor$.

In the following section we describe the graphs H of Theorem 1.1 and mention some of their properties. The proof of Theorem 1.1, which requires some probabilistic arguments, appears in Section 3. Section 4 contains some concluding remarks and open problems.

A word about notation used in this sequel. $d_G(v)$ is used to denote the degree of a vertex v in the graph $G = (V, E)$. $e(G)$ is used to denote the number of edges of G. For $X \subset V$, we denote by $G[X]$ the subgraph induced by X. $e(X, Y)$ denotes the number of edges between X and Y, and $e(X)$ denotes the number of edges in $G[X]$. $d(v, X)$ denotes the number of neighbors of v in X. Finally, we note that all logarithms mentioned are natural.

2 Graphs with high density which decompose nicely

Given ϵ and h, we now show how to choose the graph H for which Theorem 1.1 is applied. Let h_0 be the smallest odd number satisfying

$$h_0 \geq \max\{h\ ,\ 5\ ,\ 1/(2\epsilon)\}.$$

and put $h_0 = 2k + 1$. Note that $k \geq 2$. Let H_k be the graph obtained from $K_{k,k}$ by adding to it a new vertex which is connected by an edge to some vertex of $K_{k,k}$. Clearly, H_k has h_0 vertices and $k^2 + 1$ edges, and its density is

$$d(H_k) = \frac{1}{2} - \frac{1}{2h_0} + \frac{2}{h_0(h_0 - 1)} \geq \frac{1}{2} - \epsilon.$$

Also, $gcd(H_k) = 1$, so the only necessary H_k-decomposition condition is that the number of edges of the large graph be divisible by $k^2 + 1$. We will prove Theorem 1.1 for $H = H_k$, and for the remainder of this section, and Section 3, we fix k and H_k.

In the proof of Theorem 1.1 we need to use the fact that $K_{k,k}$ has a low Turán number, and a low Bipartite Turán number. This was proved by several researchers, and we shall use the result of Znám [10] (also in [2]):

Lemma 2.1 [Znám [10]]

1. *If G has n vertices and G does not contain $K_{k,k}$ as a subgraph, then the number of edges of G is less than*
$$\frac{1}{2}((k-1)^{1/k}n^{2-1/k} + \frac{k-1}{2}n).$$
2. *If B is a bipartite graph with at most n vertices in each vertex class and B does not contain $K_{k,k}$ as a subgraph, then the number of edges of B is less than*
$$(k-1)^{1/k}n^{2-1/k} + \frac{k-1}{2}n. \quad \square$$

Corollary 2.2 *Suppose $n \geq k$. If G is a graph with n vertices not containing $K_{k,k}$ then G has less than $n^{2-1/k}$ edges. If B is a bipartite graph with at most n vertices in each vertex class, and B does not contain $K_{k,k}$, then B has less than $2n^{2-1/k}$ edges.* $\square$

The corollary follows from Lemma 2.1 by observing that whenever $n \geq k$, $2n^{2-1/k} \geq (k-1)^{1/k}n^{2-1/k} + \frac{k-1}{2}n$.

3 Proof of the main result

In order to prove Theorem 1.1, we will show that for n sufficiently large (as a function of k)
$$f_{H_k}(n) \leq \frac{n}{2} + 20k^2n^{1-1/(k+1)}.$$
This, together with the lower bound $f_H(n) \geq \lfloor n/2 \rfloor - 1$ mentioned in the introduction (and which applies to any connected graph H with at least 3 vertices, in particular, it applies to H_k) shows that $f_{H_k}(n)/n \to 0.5$ when $n \to \infty$.

Given a graph $G = (V, E)$ on n vertices, $e(G) = m(k^2+1)$ edges, having $\delta(G) \geq n/2 + 20k^2n^{1-1/(k+1)}$, we must show that G has an H_k-decomposition. As mentioned before, whenever necessary, we shall assume that n is sufficiently large as a function of k.

Put $t = n^{2-1/(k+1)}$. Our initial step is to show that G contains a spanning subgraph G^* with at most $m - t$ edges, and at least $m - 3t$ edges, whose expansion properties resemble those of G. This is done in the following lemma:

Lemma 3.1 *G has a spanning subgraph $G^* = (V, E^*)$ with the following properties:*

1.
$$m - t \geq |E^*| \geq m - 3t.$$
2. *For every $v \in V$,*
$$\frac{d_G(v)}{k^2+1} - t/n \geq d_{G^*}(v) \geq \frac{d_G(v)}{k^2+1} - 9t/n.$$

3. *Let $X_0 \subset V$ be an arbitrary subset of vertices satisfying $n/(1000k^3) \le |X_0| \le 3n/4$. Let z denote the number of edges not in G^* which have an endpoint in X_0. Then:*

$$z \ge 4tk^2 + k^2 e(G^*[X_0])$$

where $e(G^*[X_0])$ is the number of edges of G^* with both endpoints in X_0.

Proof: We will show the existence of G^* using a probabilistic argument. Let $p = \frac{m-2t}{m(k^2+1)}$. We first show that $p > 1/(2k^2+2)$. This is equivalent to showing that $t < m/4$, and this holds since

$$t = n^{2-1/(k+1)} < \frac{n^2}{16(k^2+1)} = \frac{m}{4}\frac{n^2/4}{e(G)} < \frac{m}{4}.$$

(In the last inequality, we have used here the fact that n is sufficiently large as a function of k, and we have also used the obvious fact that $e(G) > n^2/4$). Each edge of G chooses to be in G^* by flipping a biased coin with probability p for being in G^*. All the choices of all the edges are independent. We now show that with high probability, the three conditions required of G^* hold.

1. The expected number of edges of G^* is exactly $m - 2t$. Since $|E^*|$, the number of edges of G^*, is the sum of $m(k^2+1)$ indicator random variables, it has binomial distribution, so we can use the Chernoff inequality (c.f. [1]) to bound the deviation of $|E^*|$ from its mean:

$$\text{Prob}[|\ |E^*| - (m-2t)\ | > t] < 2e^{-\frac{2t^2}{m(k^2+1)}} < 2e^{-\frac{2n^{4-2/(k+1)}}{n^2/2}} = 2e^{-4n^{2-2/(k+1)}} < 1/n.$$

   Thus, with probability at least $1 - 1/n$, $m - t \ge |E^*| \ge m - 3t$.

2. Consider a vertex v. The expected degree of v in G^* is exactly $p \cdot d_G(v)$. Once again, $d_{G^*}(v)$ has binomial distribution, so according to the Chernoff inequality, we know that

$$\text{Prob}[|d_{G^*}(v) - p \cdot d_G(v)| > \sqrt{n \log n}] < 2e^{-2n\log n/d_G(v)} < 2e^{-2\log n} = \frac{2}{n^2}.$$

   Thus, with probability at least $1 - 2/n$, we have that for *every* $v \in V$,

$$|d_{G^*}(v) - p \cdot d_G(v)| \le \sqrt{n \log n}.$$

   This translates to

$$d_{G^*}(v) \ge p \cdot d_G(v) - \sqrt{n\log n} = \frac{d_G(v)}{k^2+1} - \frac{2td_G(v)}{m(k^2+1)} - \sqrt{n \log n} \ge$$

$$\frac{d_G(v)}{k^2+1} - \frac{2n^{2-1/(k+1)}n}{n^2/4} - \sqrt{n\log n} = \frac{d_G(v)}{k^2+1} - 8n^{1-1/(k+1)} - \sqrt{n \log n} >$$

$$\frac{d_G(v)}{k^2+1} - 9n^{1-1/(k+1)} = \frac{d_G(v)}{k^2+1} - 9t/n.$$

and

$$d_{G^*}(v) \le p \cdot d_G(v) + \sqrt{n \log n} = \frac{d_G(v)}{k^2+1} - \frac{2td_G(v)}{m(k^2+1)} + \sqrt{n \log n} \le$$

$$\frac{d_G(v)}{k^2+1} - \frac{2n^{2-1/(k+1)}n/2}{n^2/2} + \sqrt{n \log n} =$$

$$\frac{d_G(v)}{k^2+1} - 2n^{1-1/(k+1)} + \sqrt{n \log n} \le \frac{d_G(v)}{k^2+1} - n^{1-1/(k+1)} = \frac{d_G(v)}{k^2+1} - t/n.$$

3. Now consider a set $X_0 \subset V$ satisfying $n/(1000k^3) \le |X_0| \le 3n/4$. Let $y = e_G(X_0, V \setminus X_0)$ denote the number of edges of G with only one endpoint in X_0, and let $y^* = e_{G^*}(X_0, V \setminus X_0)$. If $|X_0| \le n/2$ then $y \ge |X_0|(n/2 + 20k^2 n^{1-1/(k+1)} - |X_0|)$. By elementary calculus, if n is sufficiently large then the minimum for y is obtained when $|X_0| = n/2$, and then

$$y \ge \frac{n}{2} 20k^2 n^{1-1/(k+1)} = 10tk^2. \tag{1}$$

If $|X_0| \ge n/2$ then $y \ge (n - |X_0|)(n/2 + 20k^2 n^{1-1/(k+1)} - (n - |X_0|))$. Once again, if n is sufficiently large the minimum is obtained when $|X_0| = n/2$ and thus (1) holds in any case. Clearly, the expectation of y^* is py. Now, by the Chernoff inequality,

$$\text{Prob}[y^* > 3py/2] = \text{Prob}[y^* - py > py/2] < e^{-2(py/2)^2/y} = e^{-p^2 y/2} < e^{-y/(8(k^2+1)^2)} <$$

$$e^{-10tk^2/20k^4} = e^{-t/(2k^2)} = e^{-n^{2-1/(k+1)}/(2k^2)} < e^{-n}.$$

As there are 2^n possible subsets of V (and even less possible subsets which may correspond to X_0), we have that with probability at least $1 - (2/e)^n$, for all X_0, the corresponding y^* satisfies $y^* \le 3py/2$.

We now show that when $y^* \le 3py/2$ then also $y - y^* \ge 5tk^2$. Indeed, according to (1), $y/2 \ge 5tk^2$. Also, trivially, $y/2 > 1.5py$ since $p < 1/(k^2+1) < 1/3$. Thus, $y = y/2 + y/2 > 1.5py + 5tk^2 \ge y^* + 5tk^2$. Consider first the case where $e(G[X_0]) \le t$. In this case,

$$z \ge y - y^* \ge 5tk^2 \ge 4tk^2 + k^2 e(G[X_0]) \ge 4tk^2 + k^2 e(G^*[X_0]).$$

Now consider the case where $e(G[X_0]) > t$. The expectation of $e(G^*[X_0])$ is $p \cdot e(G[X_0])$. Using the Chernoff inequality we obtain

$$\text{Prob}[e(G^*[X_0]) > \frac{e(G[X_0])}{k^2+1}] = \text{Prob}[e(G^*[X_0]) - p \cdot e(G[X_0]) > \frac{2t}{m(k^2+1)} e(G[X_0])]$$

$$< e^{-\frac{8t^2 e(G[X_0])^2}{m^2(k^2+1)^2 e(G[X_0])}} = e^{-\frac{8t^2 e(G[X_0])}{m^2(k^2+1)^2}} \le e^{-\frac{8t^3}{e(G)^2}} < e^{-\frac{8(n^{2-1/(k+1)})^3}{(n^2/2)^2}} < e^{-32n}.$$

Therefore, with probability at least $1-(2/e^{32})^n$, for all subsets X_0 having $e(G[X_0]) > t$, we have that $e(G^*[X_0]) \le \frac{e(G[X_0])}{k^2+1}$. This implies that

$$z = y-y^*+e(G[X_0])-e(G^*[X_0]) \ge 5tk^2+(k^2+1)e(G^*[X_0])-e(G^*[X_0]) > 4tk^2+k^2e(G^*[X_0]).$$

Summing up all the probabilities, we have that with probability at least

$$1-(2/e^{32})^n-(2/e)^n-2/n-1/n > 0$$

all of the properties required from G^* in parts 1,2 and 3 of the lemma hold. □

We now fix G^* with the properties guaranteed by Lemma 3.1. Let $G' = (V, E')$ be the spanning subgraph of G obtained by removing from G the edges of G^*. Hence $E' = E \setminus E^*$. By lemma 3.1, $m - 3t \le |E^*| \le m - t$, and therefore

$$k^2m + 3t \ge |E'| \ge k^2m + t. \tag{2}$$

By Lemma 3.1, $d_{G^*}(v) < d_G(v)/(k^2+1)$ for every $v \in V$. Thus,

$$\frac{n}{2} < d_G(v) = d_{G'}(v) + d_{G^*}(v) < d_{G'}(v) + \frac{d_G(v)}{k^2+1}.$$

Hence, since $k \ge 2$,

$$\forall v \in V \ : \ d_{G'}(v) \ge d_G(v)\frac{k^2}{k^2+1} \ge \frac{n}{2}\cdot\frac{4}{5} = 0.4n. \tag{3}$$

Our next step is to find in G' a set L of exactly m edge-disjoint copies of $K_{k,k}$. L can clearly be constructed in a greedy way, using the fact that $|E'| \ge k^2m + t$, for as long as there are t edges of E' which are not assigned to any copy of $K_{k,k}$, we can find an additional copy using Corollary 2.2, since $t > n^{2-1/k}$. However, we do not want to create L by an entirely greedy procedure, since we want L to satisfy two requirements (the reason for insisting on these requirement will be made apparent later). The requirements are:

1. **property A:** Every vertex appears in at least $n/(600k^2)$ members of L.
2. **property B:** For every two distinct vertices a and b, there are at least $n/(600k^3)$ members of L which contain a and do not contain b.

In the following paragraph we describe how to pick the m copies of $K_{k,k}$ in such a way that both requirements are met.

Put $q = \lceil n^2/(100k^2) + n \rceil$. We begin by picking a set of q copies of $K_{k+1,k+1}$ which we denote by $T_1, \ldots, T_q$ as follows: Suppose we have already picked $T_1, \ldots T_r$, where $0 \le r < q$. We show how T_{r+1} is selected. Let $G_r = (V, E_r)$ be the spanning subgraph of G' consisting of the edges of E' that are not used by $T_1 \cup \ldots \cup T_r$ (initially, $G_0 = G'$). Let v be the vertex which appears the minimum number of times in $\{T_1, \ldots, T_r\}$. We pick T_{r+1} to be a copy of $K_{k+1,k+1}$ in G_r, which contains v. We must show that, indeed, there exists a copy of $K_{k+1,k+1}$ in G_r which contains v. This is proved in the following lemma:

Lemma 3.2 *If v appears the minimum number of times in $T_1, \ldots, T_r$, then there exists a $K_{k+1,k+1}$ in G_r which contains v.*

Proof: Let $X = \{T_1, \ldots T_r\}$, and suppose the number of members of X having v as a vertex is minimum. Each T_i has $2(k+1)$ vertices. Overall there are $2(k+1)r$ vertices in all the members of X, so v appears in at most $2(k+1)r/n$ members of X. Now,

$$\frac{2(k+1)r}{n} \leq \frac{2(k+1)(q-1)}{n} \leq 2(k+1)(\frac{n}{100k^2}+1) = \frac{n}{50k} + \frac{n}{50k^2} + 2k + 2 < \frac{n}{30k}. \qquad (4)$$

(The fact that $n/(50k) + n/(50k^2) + 2k + 2 < n/(30k)$ follows from the fact that $k \geq 2$ and n is sufficiently large.) Therefore, v appears in less than $n/(30k)$ members of X. Let D be the neighborhood of v in G_r, and put $d = |D|$, the degree of v in G_r. By (4) $d > d_{G'}(v) - (k+1) \cdot (n/30k)$. By (3), $d_{G'}(v) \geq 0.4n$, so $d > 0.4n - 3n/60 > 0.3n$.
We claim that there are more than $3t + d$ edges of G_r with an endpoint in D. To see this, note that by (3), the sum of degrees of the vertices of D in G' is at least $0.4nd > 0.4n \cdot 0.3n = 0.12n^2$. There are $(k+1)^2 r$ edges which appear in G' and do not appear in G_r. Thus, the sum of the degrees of the vertices of D in G_r is greater than $0.12n^2 - 2(k+1)^2 r$. However,

$$0.12n^2 - 2(k+1)^2 r \geq 0.12n^2 - 2(k+1)^2(\frac{n^2}{100k^2} + n) > 0.07n^2 > 8n^{2-1/(k+1)} = 6t + 2t > 6t + 2d.$$

So, the sum of the degrees of the vertices of D in G_r is greater than $6t + 2d$, and thus there are more than $3t + d$ edges of G_r with an endpoint in D.
Excluding from the edges with an endpoint in D the d edges connected to v, we still remain with more than $3t$ edges. Thus, either there are t edges of G_r with both endpoints in D, or there are $2t$ edges in the bipartite subgraph of G_r induced by the vertex classes D and $V \setminus (D \cup \{v\})$. In the first case, by Corollary 2.2, G_r has a copy of $K_{k+1,k+1}$ whose edges are all in $G_r[D]$, so v may be joined to either vertex class of this $K_{k+1,k+1}$, thereby proving that v appears in a $K_{k+2,k+1}$ copy of G_r (which is even better than what we need). In the second case, again by Corollary 2.2, G_r has a copy of $K_{k+1,k+1}$ with one vertex class in D and the other in $V \setminus (D \cup \{v\})$, so v may be joined to the second vertex class, thereby proving, again, that v appears in a $K_{k+2,k+1}$ copy of G_r. $\Box$

After picking $T_1, \ldots, T_q$, we *randomly* select from each T_i, one vertex from each vertex class and delete it from T_i, thereby forming a $K_{k,k}$ which we denote by S_i. All the $2q$ random selections are independent. Note that the probability that a vertex $v \in T_i$ does not appear in S_i is exactly $1/(k+1)$. Let G'' be the subgraph of G' from which the edges of $S_1 \cup \ldots \cup S_q$ have been deleted. By (2) G'' has $|E'| - k^2 q \geq k^2(m-q) + t$ edges, so we can continue selecting, greedily, $m - q$ edge-disjoint additional copies of $K_{k,k}$ which we denote by $S_{q+1}, \ldots, S_m$.

Having produced the set $L = \{S_1, \ldots, S_m\}$ by the process described above, we claim that L is guaranteed to have, with positive probability, properties A and B. This is formally proved in the next two lemmas.

Lemma 3.3 *With probability at least* $1 - 1/n$*, every* $v \in V$ *appears in at least* $n/(600k^2)$ *members of* L*.*

Proof: Let $v \in V$ be arbitrary. We first show that v must appear in at least $n/(100k^2)$ copies of $X = \{T_1, \ldots, T_q\}$. Recall the process which created X. Each T_r must contain a vertex which appears the minimum number of times in $T_1, \ldots, T_{r-1}$. Thus, there is a vertex w which was chosen as minimal at least q/n times. Let r_0 be the last stage in which w was chosen as minimal. w appears at least $q/n - 1$ times in $T_1, \ldots, T_{r_0-1}$. and by the minimality of w, every $v \in V$ appears at least $q/n - 1$ times in $T_1, \ldots, T_{r_0-1}$. However, $q/n - 1 \geq n/(100k^2)$, proving what we wanted. Let $Y_v = \{i \mid v \in T_i\}$. By the first part above, $|Y_v| \geq n/(100k^2)$. Let $Z_v = \{i \mid v \in S_i,\ i \leq q\}$. $|Z_v|$ is a random variable which is the sum of $|Y_v|$ independent indicator variables whose probability of success is $k/(k+1)$. Thus, the expectation of $|Z_v|$ is $k|Y_v|/(k+1)$, and by the Chernoff inequality:

$$\text{Prob}[|Z_v| - \frac{k}{k+1}|Y_v| < -\frac{|Y_v|}{k}] < e^{-\frac{2|Y_v|^2}{k^2|Y_v|}} = e^{-2\frac{|Y_v|}{k^2}} < e^{-n/(50k^4)} < 1/n^2.$$

Thus, with probability greater than $1 - 1/n^2$, $|Z_v|$ is at least $k|Y_v|/(k+1) - |Y_v|/k \geq |Y_v|/6 \geq n/(600k^2)$, and therefore, with probability $1 - n \cdot 1/n^2 = 1 - 1/n$ this holds for every $v \in V$. $\square$

Lemma 3.4 *With probability at least 0.5 the following holds: for every two distinct vertices* a *and* b*, there there are at least* $n/(600k^3)$ *members of* L *which contain* a *and do not contain* b*.*

Proof: Fix two distinct vertices a and b. Using the same notation of Lemma 3.3, let $Y_a = \{i \mid a \in T_i\}$. By the proof of Lemma 3.3, $|Y_a| \geq n/(100k^2)$. Since the $\{T_1, \ldots, T_q\}$ are edge-disjoint, there is at most one member which contains a and b in distinct vertex classes (in other words, the edge (a,b), if it exists, appears in at most one of the T_i). Let $f(a,b)$ be the number of members of $\{S_1, \ldots, S_q\}$ which contain a and do not contain b. For each $i \in Y_a$, (except for, maybe, at most one member of Y_a that contains the edge (a,b)) the probability that it contains a and does not contain b is either $1/(k+1)$ if a and b are in the same vertex class of T_i, or $k/(k+1)$ if b does not appear in T_i. In any case, $f(a,b)$ is the sum of $|Y_a|$ (or $|Y_a| - 1$) independent indicator random variables having probability of success at least $1/(k+1)$. Denoting by μ the expectation of $f(a,b)$ we have:

$$\text{Prob}[f(a,b)-\mu < -|Y_a|/2k] < e^{-\frac{2|Y_a|^2}{4k^2|Y_a|}} = e^{-\frac{|Y_a|}{2k^2}} < e^{-n/(200k^4)} << 1/(2n^2).$$

Thus, with probability greater than $1-1/(2n^2)$, $f(a,b) \geq \mu - |Y_a|/2k \geq |Y_a|/(k+1) - |Y_a|/(2k) \geq |Y_a|/(6k) \geq n/(600k^3)$. Therefore, with probability $1-n(n-1)/(2n^2) > 0.5$ this holds for all ordered pairs a and b. $\Box$

By lemmas 3.3 and 3.4 we have that with probability at least $0.5 - 1/n > 0.4$, L satisfies both properties A and B. We therefore fix a set L satisfying both of these properties. Let M denote the set of edges which do not appear in any member of L. Clearly, $M \supset E^*$, $|M| = m$. Let F be those edges of G' that do not appear in any member of L. Thus, $M = E^* \cup F$. Our goal is to match the m edges of M with the m members of L such that $(a,b) \in M$ is matched to some $S_i \in L$ if and only if exactly one of a or b is a vertex of S_i. Such a matching shows that G has m edge-disjoint copies of H_k, and thus an H_k-decomposition, as required. For this purpose we define a bipartite graph B with two vertex classes of size m each. The left vertex class is M and the right vertex class is L. A vertex of the left vertex class (namely some edge $(a,b) \in M$), is connected in B to a vertex of the right vertex class (namely some $S_i \in L$), if and only if exactly one of a and b appears in S_i. Our goal is, therefore, to show that B has a perfect matching. For this purpose, we will use Hall's Theorem (cf. e.g. [2]). Let $M' \subset M$ be an arbitrary nonempty subset. We need to show that $N(M')$, the neighborhood of M' in B, satisfies $|N(M')| \geq |M'|$.

Let $X \subset V$ be the set of vertices which are endpoints of at least one edge of M'. Let $X_0 \subset X$ be the subset of vertices which are incident with at least k edges of M', and let $X_1 = X \setminus X_0$ be the subset of vertices of X incident with less than k edges of M'. An important observation about X_0 is the following:

Claim: If $v \in X_0$ appears in S_i then $S_i \in N(M')$.

Proof: The vertex class of S_i which contains v has $k-1$ vertices other than v. However, since $v \in X_0$, we know that there are at least k members of M' which have v as their endpoint. Therefore, there exists some $(v,w) \in M'$ such that w is not in the same vertex class of S_i as v. Also, w is not in the other vertex class of S_i since $(v,w) \notin S_i$ because $(v,w) \in M' \subset M$. Thus, w does not appear at all in S_i, and therefore, (v,w) is connected in B to S_i. Hence, $S_i \in N(M')$. This proves the claim.

In order to prove that $|N(M')| \geq |M'|$ we distinguish between several cases, according to the sizes of M' and X_0.

- $|M'| \leq n/(600k^3)$.

Consider an arbitrary member $(a.b) \in M'$. By Lemma 3.4, there are at least $n/(600k^3)$ members of L which contain a and do not contain b. It follows that $|N(M')| \geq |N(\{a,b\})| \geq n/(600k^3) \geq |M'|$, as required.

- $|M'| > n/(600k^3)$ and $|X_0| \leq \frac{\sqrt{n}}{(6k)^{5.5}}$.
  Trivially, $|M'| \leq \binom{|X|}{2}$, which implies $|X| > \sqrt{2|M'|} > \frac{\sqrt{n}}{18k\sqrt{k}}$. Thus,

$$|X_1| = |X| - |X_0| \geq \frac{\sqrt{n}}{18k\sqrt{k}} - \frac{\sqrt{n}}{(6k)^{5.5}} > \frac{\sqrt{n}}{20k\sqrt{k}}.$$

  Consider $v \in X_1$, and let $(v,w) \in M'$ be arbitrary. According to Lemma 3.4, there are at least $n/(600k^3)$ members of L which contain v and do not contain w. All of these members are neighbors of (v,w) in B. Thus, they are all in $N(M')$. Since this is true for every $v \in X_1$, we have at least $|X_1|n/(600k^3)$ members of L counted in this way, and since no copy of L is counted more than $2k$ times ($2k$ is the number of vertices of $K_{k,k}$), we get that

$$|N(M')| \geq \frac{|X_1|n}{1200k^4}.$$

  Note that, obviously, $(n-1)|X_0| + (k-1)|X_1| \geq 2|M'|$ (the l.h.s. bounds from above the sum of the degrees in the subgraph induced by M') which implies $|M'| < \frac{n|X_0|+k|X_1|}{2}$. Thus, it suffices to show that

$$\frac{n|X_0| + k|X_1|}{2} \leq \frac{|X_1|n}{1200k^4}.$$

  This is equivalent to showing that

$$|X_0| \leq (\frac{1}{600k^4} - \frac{k}{n})|X_1|.$$

  Indeed,

$$|X_0| \leq \frac{\sqrt{n}}{(6k)^{5.5}} \leq (\frac{1}{600k^4} - \frac{k}{n})\frac{\sqrt{n}}{20k\sqrt{k}} \leq (\frac{1}{600k^4} - \frac{k}{n})|X_1|.$$

- $|M'| > n/(600k^3)$ and $\sqrt{n}/(6k)^{5.5} < |X_0| < n/(1000k^3)$.
  Clearly, $|M'| \leq |X_1|(k-1) + \binom{|X_0|}{2} < nk + \binom{|X_0|}{2}$. By Lemma 3.3, every $v \in X_0$ appears in at least $n/(600k^2)$ members of L. By the claim proved above, if v appears in S_i then $S_i \in N(M')$. Thus, every $v \in X_0$ contributes at least $n/(600k^2)$ members to $N(M')$, and every such member S_i is counted at most $2k$ times by these contributions (since S_i has $2k$ vertices). Therefore,

$$|N(M')| \geq \frac{n}{600k^2}|X_0|\frac{1}{2k}.$$

Now, for $\sqrt{n}/(6k)^{5.5} < |X_0| < n/(1000k^3)$ we have

$$|M'| < nk + \binom{|X_0|}{2} < \frac{n}{1200k^3}|X_0| \le |N(M')|$$

as required.

- $|M'| > n/(600k^3)$ and $n/(1000k^3) \le |X_0| \le 2n/3 + k^3$.
  In this case, we can use Lemma 3.1 applied to $|X_0|$ (note that $2n/3 + k^3 < 3n/4$ for n sufficiently large). Let x be the number of members of L having a vertex of X_0. Clearly, $x \le |N(M')|$. Let z_1 denote the number of members of M with both endpoints in X_0, and let z_2 denote the number of edges of E^* with both endpoints in X_0. Since $M = E^* \cup F$ and since, by (2) $|F| = |E'| - k^2 m \le 3t$, we get that $z_1 \le 3t + z_2$ and therefore

  $$|M'| \le |X_1|(k-1) + z_1 < kn + z_1 \le kn + 3t + z_2.$$

  Hence, it suffices to show that $x \ge kn + 3t + z_2$. According to Lemma 3.1, z, the number of edges of E' with at least one endpoint in X_0 satisfies $z \ge 4tk^2 + k^2 z_2$. These z edges, except for the edges of F, all appear in the members of L. Thus, the number of edges in members of L which have at least one endpoint in X_0 is at least $4tk^2 + k^2 z_2 - 3t$. Since each member of L has k^2 edges we get that

  $$x \ge 4t + z_2 - \frac{3t}{k^2} \ge 3t + z_2 + \frac{t}{4} > kn + 3t + z_2.$$

- $|M'| > n/(600k^3)$ and $|X_0| > 2n/3 + k^3$.
  Let $n' = n - |X_0|$ be the size of $V \setminus X_0$. The conditions imply that $n' \le n/3 - k^3$. If some $S_i \in L$ is not in $N(M')$ then, according to the claim proved above, S_i contains no vertex of X_0, and therefore all its $2k$ vertices are in $V \setminus X_0$. Thus, $N(M') \ge m - \binom{n'}{2}/k^2$. We need to show that $|M'| \le m - \binom{n'}{2}/k^2$, or, equivalently, that $|M \setminus M'| \ge \binom{n'}{2}/k^2$. Consider a vertex $v \in V \setminus X_0$. By Lemma 3.1,

  $$d_{G^*}(v) \ge \frac{d_G(v)}{k^2+1} - 9t/n \ge \frac{n}{2k^2+2} - 9n^{1-1/(k+1)} \ge \frac{n}{3k^2}.$$

  Thus, since $M \supset E^*$, there are at least $n/(3k^2)$ edges of M having v as an endpoint. Either $v \notin X$ or $v \in X_1$. In any case, v is an endpoint of at most $k-1$ edges of M'. Thus, there are at least $n/(3k^2) - k + 1$ edges of $M \setminus M'$ having v as an endpoint. Since this is true for every $v \in V \setminus X_0$, we have that there are at least $(n/(3k^2) - k + 1)n'/2$ edges in $M \setminus M'$. Thus, we must show that $(n/(3k^2) - k + 1)n'/2 \ge n'(n'-1)/(2k^2)$. Indeed, this holds for $n' \le n/3 - k^3$. $\square$

4 Concluding remarks and open problems

1. The graphs H_k which are used to prove Theorem 1.1 have density which is arbitrary close to $1/2$. In particular, k^2+1, the number of edges of H_k, is a quadratic function of the number of vertices, which is $2k+1$. Thus, the graphs H_k are *dense*. However, the minimum degree of H_k is 1. It is an open problem whether there exist graphs H with arbitrary high minimum degree for which the statement of Theorem 1.1 holds. Namely, is it possible to determine the limit of $f_H(n)/n$ for graphs H with arbitrary high minimum degree. A somewhat less ambitious problem, but still an open one, is to find a graph H with $\delta(H)=2$, for which the limit of $f_H(n)/n$ can be determined.

2. Another obstacle in extending Theorem 1.1 to other families of graphs is the chromatic number. The graphs H_k are bipartite, and so are all trees, for which the limit of $f_H(n)/n$ is determined in [9]. It will be interesting to find graphs H with arbitrary high chromatic number, for which $f_H(n)/n$ can be asymptotically determined. We do not even know of a 3-Chromatic graph for which this can be done.

3. Although Theorem 1.1 determines the asymptotic behavior of $f_{H_k}(n)$, i.e. $f_{H_k}(n)=\frac{n}{2}(1+o(1))$ there is still a sublinear gap between the lower bound of $\lfloor n/2 \rfloor - 1$ and the upper bound of $n/2+O(n^{1-1/(k+1)})$. It would be interesting to close this gap.

4. As mentioned in the introduction, a lower bound of $\lfloor n/2 \rfloor - 1$ for $f_H(n)$ is described in [9], and applies to all connected graphs with at least three vertices (this lower bound is valid for $n \ge n_0(H)$ since when n is small there is some noise). For the sake of completeness, we describe it here for $H=H_k$. We will assume that $n \ge 4k^2$ is even, although a similar argument holds when $n > 4k^2$ is odd. It suffices to show the existence of a graph $G=(V,E)$ with n vertices, and $\delta(G)=n/2-2$ where k^2+1 divides $|E|$, but still there is no H_k-decomposition of G. Put $n=2x$ and let $d=x(x-1) \bmod (k^2+1)$, where $0 \le d \le k^2$. If $d \ne 0$ consider the graph G obtained from the vertex-disjoint union of K_x and $P_{x,d}$ where $P_{x,d}$ is the complete graph on x vertices from which d independent edges have been removed. (We can remove d independent edges since $2d \le 2k^2 \le n/2 = x$). G has $2x=n$ vertices, $|E|=x(x-1)-d$ edges, and so k^2+1 divides $|E|$. Also $\delta(G)=x-2=n/2-2$. However, G does not have an H_k-decomposition since k^2+1 does not divide $\binom{x}{2}$, which is the number of edges of the connected component K_x of G. If $d=0$ we can take G to be the union of $P_{x,1}$ and P_{x,k^2}, and once again G has $|E|=x(x-1)-(k^2+1)$ edges, and so k^2+1 divides $|E|$. $\delta(G)=x-2=n/2-2$, and G does not have an H_k decomposition since k^2+1 does not

divide $\binom{x}{2} - 1$, which is the number of edges of the component $P_{x,1}$.

5. As mentioned in the introduction, Theorem 1.1 can be extended to show that if G has minimum degree $\frac{n}{2}(1 + o(1))$ but does not have the necessary H_k-decomposition conditions (i.e. the number of edges of G is not divisible by $k^2 + 1$) then G still has an optimal packing, namely, there exist $\lfloor |E_G|/(k^2+1) \rfloor$ edge-disjoint copies of H_k in G. This follows from Theorem 1.1, since if $\delta(G) \geq n/2 + 20k^2 n^{1-1/(k+1)} + 1$, and if $d = |E_G| \bmod (k^2 + 1)$ where $1 \leq d \leq k^2$, then by deleting from G an arbitrary set of d independent edges we remain with a subgraph G^α with n vertices, $\delta(G^\alpha) = \delta(G) - 1 \geq n/2 + 20k^2 n^{1-1/(k+1)}$, and G^α does satisfy the necessary H_k-decomposition conditions, so by Theorem 1.1 G^α has an H_k-decomposition, and the number of members in this decomposition is $\lfloor |E_G|/(k^2 + 1) \rfloor$.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley.

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press, 1978.

[3] Y. Caro and R. Yuster, *Packing graphs: The packing problem solved*, Elect. J. Combin. (1997), #R1.

[4] Y. Chang, *A bound for Wilson's Theorem - III*, J. Combinatorial Designs 4 (1996), 83-93.

[5] D. Dor and M. Tarsi, *Graph decomposition is NPC - A complete proof of Holyer's conjecture*, Proc. 20th ACM STOC, ACM Press (1992), 252-263.

[6] T. Gustavsson, *Decompositions of large graphs and digraphs with high minimum degree*, Doctoral Dissertation, Dept. of Mathematics, Univ. of Stockholm, 1991.

[7] C. St. J. A. Nash-Williams, *An unsolved problem concerning decomposition of graphs into triangles*, Combinatorial Theory and its Applications III. ed. P. Erdös, P. Rényi and V.T. Sós. North Holland (1970), 1179-1183.

[8] R. M. Wilson, *Decomposition of complete graphs into subgraphs isomorphic to a given graph*, Congressus Numerantium XV (1975), 647-659.

[9] R. Yuster, *Tree packing of graphs*, Random Structures and Algorithms, to appear.

[10] S. Znám, *On a combinatorial problem of K. Zarankiewicz*, Colloq. Math. 11 (1963) 81-84.