The decomposition threshold for bipartite graphs with minimum degree one

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

Abstract

Let H be a fixed bipartite graph with $\delta(H) = 1$. It is shown that if G is any graph with n vertices and minimum degree at least $\frac{n}{2}(1 + o_n(1))$ and $e(H)$ divides $e(G)$, then G can be decomposed into $e(G)/e(H)$ edge-disjoint copies of H. This is best possible and significantly extends the result of [8] which deals with the case where H is a tree.

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 connected graph. An *H-packing* of a graph G is a coloring of the edges of G such that each color class induces a subgraph which contains a copy of H. The *H-packing number* of G, denoted by $P(H, G)$, is the maximum possible number of colors used in an H-packing of G. Clearly, $P(H, G) \leq e(G)/e(H)$. In case equality holds, we must have that every color class induces a subgraph which is isomorphic to H. In this case we say that G has an *H-decomposition*. Thus, a necessary condition for the existence of an H-decomposition is that $e(H)$ divides $e(G)$. Another obvious necessary condition is that $gcd(H)$ divides $gcd(G)$, where the *gcd* of a graph is the greatest common-divisor of the degrees of its vertices. We say that G has property $\mathcal{H}$ if these two conditions hold. Note that it is a trivial computational task to verify if a graph G has $\mathcal{H}$.

The combinatorial and computational aspects of the H-packing and H-decomposition problems have been studied extensively. Wilson in [7] has proved that whenever $n \geq n_0 = n_0(H)$, and K_n has $\mathcal{H}$, then K_n has an H-decomposition. The H-packing problem for 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

*e-mail: raphy@research.haifa.ac.il World Wide Web: http:\\research.haifa.ac.il\~raphy

is known that the H-decomposition and H-packing problems are, in general, NP-Hard, since Dor and Tarsi [4] showed that deciding if G has an H-decomposition is NP-Complete, where H is any fixed connected graph with at least three edges (so, even for $H = P_4$ or $H = C_3$ or $H = K_{1,3}$ this is hard). In view of Wilson's positive result, and the Dor-Tarsi negative result, the following extremal problem was raised in [9]:

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 has $\mathcal{H}$, then G has an H-decomposition. (Note that it is possible that for some n we might have $f_H(n) = \infty$).

Another version of Problem 1 is to determine the asymptotic behavior:

Problem 1A: Determine $g(H) = \limsup_{n\to\infty} f_H(n)/n$.

Wilson's result proves that $f_H(n)$ exists for all $n \geq n_0(H)$, or, in other words, $f_H(n) \leq n-1$ for all $n \geq n_0(H)$. 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 [5]. 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 has $\mathcal{H}$, then G has an H-decomposition. In other words, $f_H(n) \leq (1-\epsilon(H))n$, for all n sufficiently large, and thus $g(H) \leq 1-\epsilon(H)$. Unfortunately, the $\epsilon(H)$ in Gustavsson's result is a very small number. For example, if $H = C_3$ then $\epsilon(H) \leq 10^{-24}$. In general, $\epsilon(H) \leq 10^{-24}/|H|$. It seems likely, however, that the correct value for $f_H(n)$ is much smaller. In fact, Nash-Williams conjectured in [6] that when $H = C_3$, then $f_H(n) \leq \lceil 3n/4 \rceil$, and he also gave an example showing that this would be best possible, and thus his conjecture is that $f_{C_3}(n) = \lceil 3n/4 \rceil$, and that $g(C_3) = 3/4$. 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 [8] in case the graph H is a tree. 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. Thus, $g(H) \leq 0.5$ for trees. This result is asymptotically best possible as it is also shown in [8] that $f_H(n) \geq \lfloor n/2 \rfloor - 1$ for every connected graph H with at least 3 vertices. Hence, $g(H) = 0.5$ for trees (and we can replace $\limsup$ with $\lim$ in the definition of $g(H)$, in this case).

In this paper we significantly extend the result of [8] and show that it holds for every bipartite graph H with $\delta(H) = 1$. Note that trees are contained in this class of graphs. Our main result is summarized in the statement of the following theorem:

Theorem 1.1 *Let H be a bipartite graph with minimum degree 1. Then,* $\lim_{n\to\infty} f_H(n)/n = 0.5$.

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

Theorem 1.1 is best possible in another sense as well. It cannot be improved to include, say, all fixed bipartite graphs with minimum degree 2. The following example, due to Peter Winkler and

Jeff Kahn demonstrates this fact. Let $H = C_4$, and let $n = 5k$ where $k \equiv 3 \bmod 8$. Consider the graph $G = C_5 * K_k$ (i.e. each vertex of C_5 is blown up to a copy of K_k and each original edge of C_5 is now a complete bipartite graph $K_{k,k}$). G has n vertices, it is regular of degree $3k - 1 = 0.6n - 1$ which, by the choice of k, is an even number. The total number of edges of G is $5k^2 + 5\binom{k}{2}$ which, again by the choice of k, is divisible by 4. Hence, G has property $\mathcal{H}$. However, G does not have a C_4 decomposition. To see this, consider some complete bipartite graph $K_{k,k}$ connecting two blown up original vertices. Every C_4 in G contains an even number of edges of this $K_{k,k}$ (i.e. 0,2 or 4 edges). However, k^2 is odd. One might conjecture that replacing 0.5 with 0.6 in Theorem 1.1 suffices for $H = C_4$. However, a proof of this currently seems beyond reach.

In the following section we make some initial preparations, mainly concerning a construction which we need in the proof of Theorem 1.1. A top-down sketch of the proof appears in Section 3. The detailed proof appears in Section 4. We note that, although Theorem 1.1 extends the result in [8], the method of proof is entirely different. The method of proof is a generalization of that of [9], which is, in fact, a proof of a very special case of Theorem 1.1 where H is a complete bipartite graph together with a vertex of degree one.

2 A preliminary construction

From here onwards we fix a bipartite graph H with $\delta(H) = 1$ and with h edges. We will assume $h \geq 3$ since otherwise the result is trivial. We label the vertices of H as $\{b, a_0, \ldots, a_{r-1}\}$ where b has degree one, a_0 is the unique neighbor of b, and a_1 is in a different vertex class from that of a_0 is a (fixed) bipartition of H. Notice also that $r \geq 3$. We call b the *leaf* of H (although there may be other vertices with degree one, we only designate b as a leaf). We call a_0 the *pivot* of H and we call (b, a_0) the *leaf edge* of H. We denote by H' the bipartite graph obtained from H by deleting the leaf vertex. Note that it is possible that H' has a high minimum degree. The vertex set of H' is $\{a_0, \ldots, a_{r-1}\}$ and H' has $h-1$ edges. We still refer to a_0 as the pivot of H' (although this term may be somewhat misleading w.r.t. H').

In the proof it will be convenient to initially find many edge-disjoint copies of H' in G. However, we will need that the set of these copies will meet several strong conditions regarding the distribution of vertices of G with respect to the role they play in each copy of H' (namely, to which vertex of H' they map in each copy). This is rather inconvenient since H' may be highly non-symmetric. In fact, it may even have a trivial automorphism group.

To overcome this obstacle we first construct a larger *fixed* bipartite graph, denoted H^*, which, in turn, decomposes in some very specific manner into copies of H'. We will then use H^* in the major part of the proof.

Lemma 2.1 *There exists a bipartite graph H^* with the following properties:*

1. *H^* has $r^2(r-1)$ vertices in each vertex class.*

2. *H^* has $2(h-1)r^2(r-1)^2$ edges, and is regular of degree $2(h-1)(r-1)$.*

3. *There exists an H'-decomposition of H^*, denoted D, where $|D| = 2r^2(r-1)^2$, such that every vertex of H^* is the pivot in precisely $r-1$ elements of D.*

4. *For any two distinct vertices of H^* denoted v and u, there is at least one element of D in which v plays the role of the pivot and u does not appear at all.*

Proof: The vertex-set of H^* is composed of $2r(r-1)$ disjoint subsets that we denote by $X(i,j)$ and $Y(i,j)$ where $i = 0,\dots,r-1$ and $j = 1,\dots,r-1$. Each of these subsets contains r vertices. Thus, we denote $X(i,j) = \{x^0_{(i,j)},\dots,x^{r-1}_{(i,j)}\}$ and $Y(i,j) = \{y^0_{(i,j)},\dots,y^{r-1}_{(i,j)}\}$. The union of all $X(i,j)$ will correspond to one vertex class while the union of all $Y(i,j)$ will correspond to the other vertex class. Thus, H^* has precisely $r^2(r-1)$ vertices in each vertex class. It remains to describe the edges between some $X(i,j)$ and some $Y(s,t)$.

The induced subgraph of H^* on the vertices $X(i,j)$ and $Y(s,t)$ will contain two vertex-disjoint copies of H', and thus $2(h-1)$ edges. Hence, the total number of edges in H^* is $2(h-1)r^2(r-1)^2$ and H^* has a decomposition D into $2r^2(r-1)^2$ copies of H'. The $2(h-1)$ edges between $X(i,j)$ and $Y(s,t)$ are defined as follows. For each $w = 0,\dots,r-1$ except $w=1$ or $w=t$ (note that it is possible that $t=1$), map vertex a_w of H' to vertex $x^{(w+s) \bmod r}_{(i,j)}$ of $X(i,j)$. Similarly, for each $w = 0,\dots,r-1$ except $w=1$ or $w=j$, map vertex a_w of H' to vertex $y^{(w+i) \bmod r}_{(s,t)}$ of $Y(s,t)$. For $w=1$, map a_1 to $x^{(t+s) \bmod r}_{(i,j)}$ and to $y^{(j+i) \bmod r}_{(s,t)}$. For $w=t$, map a_t to $x^{(1+s) \bmod r}_{(i,j)}$. For $w=j$, map a_j to $y^{(1+i) \bmod r}_{(s,t)}$.

We have thus defined a bijection between the vertices of H' and $X(i,j)$ and a bijection between the vertices of H' and $Y(s,t)$. Now, for each edge (a_w,a_z) of H' connect an edge between the vertex of $X(i,j)$ mapped to a_w and the vertex of $Y(s,t)$ mapped to a_z. Similarly, connect an edge between the vertex of $X(i,j)$ mapped to a_z and the vertex of $Y(s,t)$ mapped to a_w. Clearly, the induced subgraph of H^* on $X(i,j)$ and $Y(s,t)$ contains $2(h-1)$ edges and is composed of two vertex-disjoint copies of H'. One of this copies contains the vertices of $X(i,j)$ mapped to the first vertex class of H' and the vertices of $Y(s,t)$ mapped to the second vertex class of H'. The other copy contains the vertices of $X(i,j)$ mapped to the second vertex class of H' and the vertices of $Y(s,t)$ mapped to the first vertex class of H'.

Now consider an arbitrary vertex $x^s_{(i,j)}$. By the definition of our bijection, $x^s_{(i,j)}$ is a pivot in one of the two copies of H' in the subgraph induced on $X(i,j)$ and $Y(s,t)$. Since this is true for each $t = 1,\dots,r-1$ we have that $x^s_{(i,j)}$ is a pivot in precisely $r-1$ elements of the decomposition D. The same argument hold for vertices of the form $y^i_{(s,t)}$.

We also need to show that for any vertex $u \in H^*$, distinct from $x^s_{(i,j)}$, at least one of the $r-1$ elements of D that contain $x^s_{(i,j)}$ as a pivot does not contain u at all. Consider first the case where $u = x^{s'}_{(i',j')}$. Trivially, if $(i',j') \neq (i,j)$ then u and $x^s_{(i,j)}$ never appear together in any element of

D. Thus, assume $u = x^{s'}_{(i,j)}$ and hence $s' \neq s$. Let $t = (s' - s) \bmod r$. Consider the subgraph of H^* induced on $X(i,j)$ and $Y(s,t)$. The bijection from the vertices of H' to $X(i,j)$ associated with this induced subgraph maps a_0 to $x^s_{(i,j)}$ and maps a_1 to $x^{(s+t) \bmod r}_{(i,j)} = x^{s'}_{(i,j)} = u$. Recall, however, that a_0 and a_1 are in distinct vertex classes of H' and thus one of the two copies of H' in the subgraph of H^* induced on $X(i,j)$ and $Y(s,t)$ contains $x^s_{(i,j)}$ as the pivot but does not contain u at all (u appears in the other copy and plays the role of a_1 there). Next, assume that $u = y^{s''}_{(s',t)}$. If $s' \neq s$ then, trivially, whenever $x^s_{(i,j)}$ is a pivot, u does not appear at all. Assume, therefore, that $u = y^{s''}_{(s,t)}$. If $t' \neq t$ then the subgraph induced by $X(i,j)$ and $Y(s,t')$ contains a copy of H' in which $x^s_{(i,j)}$ is the pivot, and trivially u does not appear at all in such a copy. Thus, there are $r - 2 > 0$ elements of D in which $x^s_{(i,j)}$ is the pivot and u does not appear at all. Similarly, all of these arguments also apply to vertices of the form $y^i_{(s,t)}$, namely, for each $u \in H^*$, distinct from $y^i_{(s,t)}$, at least one of the $r-1$ elements of D that contain $y^i_{(s,t)}$ as a pivot does not contain u at all.

It remains to show that H^* is regular. Consider, without loss of generality, some vertex $x^s_{(i,j)}$. We already showed that the number of elements of D in which $x^s_{(i,j)}$ plays the role of a_0 is $r-1$. We claim that this is true not only for a_0 but for all other vertices of H' as well. Consider an arbitrary vertex a_w of H'. We show that in precisely $r-1$ elements of D, $x^s_{(i,j)}$ plays the role of a_w. Assume first that $w \neq 1$. Consider the subgraph of H^* induced by $X(i,j)$ and $Y((s-w) \bmod r, t)$, where $t \neq w$. The bijection from the vertices of H' to $X(i,j)$ which corresponds to this subgraph maps a_w to $x^s_{(i,j)}$. Since there are $r-2$ choices for $t \neq w$ we have found $r-2$ elements of D in which $x^s_{(i,j)}$ plays the role of a_w. However, there is still another one. In the subgraph of H^* induced by $X(i,j)$ and $Y((s-1) \bmod r, w)$ the corresponding bijection maps a_w to $x^s_{(i,j)}$. We have thus explicitly described the $r-1$ copies of D in which $x^s_{(i,j)}$ plays the role of a_w (by the definition of our bijections or by the pigeonhole principle, there cannot be more than $r-1$). Now consider a_1. For each $t = 1, \ldots, r-1$, let s' be the unique solution between 0 and $r-1$ to $(s'+t) \bmod r = s$. In the subgraph of H^* induced by $X(i,j)$ and $Y(s',t)$ the corresponding bijection maps a_1 to $x^s_{(i,j)}$. As there are $r-1$ choices for t we have (precisely) $r-1$ elements of D in which $x^s_{(i,j)}$ plays the role of a_1. Since D is a decomposition, every edge incident with $x^s_{(i,j)}$ belongs to precisely one element of D. Thus, the degree of $x^s_{(i,j)}$ is $(r-1)$ times the sum of the degrees of H', which is $2(h-1)$. Since $x^s_{(i,j)}$ was arbitrary and since similar arguments hold for vertices of the form $y^i_{(s,t)}$ we have that H^* is $2(h-1)(r-1)$ regular. $\square$

It is well known that the Turán number of fixed bipartite graphs is $o(n^2)$. There are many results giving upper bounds for such Turán numbers. In essence, they all give a slightly better upper bound than the following:

Lemma 2.2 *Suppose $n >> k$. 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.* $\square$

An example of such a result is that of Znám [10]. See also [2] for more results of this type. Since K_{r^3,r^3} contains H^* we have the following:

Corollary 2.3 *Let $n >> r^3$. If G is a graph with n vertices and at least n^{2-1/r^3} edges then G contains H^* as a subgraph. If B is a bipartite graph with at most n vertices in each vertex class and B has at least $2n^{2-1/r^3}$ edges then B contains H^* as a subgraph.*

3 A top down view of the proof

In order to prove Theorem 1.1 it obviously suffices to prove the following theorem for n sufficiently large as a function of h and r.

Theorem 3.1 *Let $G = (V, E)$ be a graph with n vertices and $\delta(G) \geq n/2 + 40(h-1)n^{1-1/r^3}$. If $|E|$ is a multiple of h then G has an H-decomposition.*

Note that Theorem 3.1 together with the fact that $f_H(n) \geq \lfloor n/2 \rfloor - 1$ mentioned in the introduction (recall this lower bound is valid for every connected graph H with at least three vertices), yield Theorem 1.1.

Here is an outline of the proof of Theorem 3.1: Given a graph $G = (V, E)$ as in the statement of the theorem, put $|E| = mh$. Thus, m is an integer, and our goal is to find in G a set of m edge-disjoint copies of H. Our algorithm consists of several steps.

1. Find $k = \lfloor m/(2r^2(r-1)^2) \rfloor$ edge-disjoint copies of H^*. We will be able do this since the Turán number of H^* is relatively small. However, as will be seen, we shall require several strong properties from our set of k copies of H^*, such as that each vertex of G be a vertex in a significant number of these copies. In fact, we need several other properties (their description is somewhat technical, so we defer it to the actual proof in the next section).

2. Find additional $m - 2kr^2(r-1)^2$ copies of H' in G which are edge-disjoint from the k copies of H^* that were previously found.

3. Using the fact that each H^* can be decomposed into $2r^2(r-1)^2$ copies of H' we have, in fact, a set of $2r^2(r-1)^2k + m - 2kr^2(r-1)^2 = m$ edge-disjoint copies of H' together with the copies of H' found in the previous step. We denote this set of m copies of H' by L.

4. There are exactly $mh - m(h-1) = m$ edges of G that do not belong to any of the m copies of H' designated in the previous step. Denote this set of edges by C. We prove (using all of the properties that we required from the k copies of H^* found in the first step, and using Lemma 2.1) that there is a perfect matching between C and L such that an edge $(x, y) \in C$ is matched to a copy S of H' in L only if exactly one of x or y is the pivot of S, while the other endpoint of the edge does not appear at all in S. Note that each such match introduces a copy of H, and thus a perfect matching corresponds to an H-decomposition of G.

Remark: *From here onwards, we assume, whenever necessary, that n is sufficiently large as a function of the constants r and h.*

4 Proof of the main result

We begin by showing that G contains a spanning subgraph with slightly less than m edges whose expansion properties resemble those of G. These edges, together with some additional ones, will be used as leaf edges in elements of the H-decomposition. The proof of the next lemma resembles the proof of a lemma appearing in [9].

Lemma 4.1 *Let $G = (V, E)$ satisfy $|V| = n$, $|E| = mh$, $\delta(G) \geq n/2 + 40(h-1)n^{1-1/r^3}$. Then, G has a spanning subgraph $G^* = (V, E^*)$ with the following properties:*

1.
$$m - n^{2-1/r^3} \geq |E^*| \geq m - 3n^{2-1/r^3}.$$

2. *For every $v \in V$,*
$$\frac{d_G(v)}{h} - n^{1-1/r^3} \geq d_{G^*}(v) \geq \frac{d_G(v)}{h} - 9n^{1-1/r^3}.$$

3. *Let $W \subset V$ be an arbitrary subset of vertices satisfying $n/(50r^6) \leq |W| \leq 0.9n$. Let w_1 denote the sum of the degrees of the vertices of W in $G' = (V, E \setminus E^*)$. Let w_2 denote the number of edges of E^* with both endpoints in W. Then,*
$$w_1 \geq 9.5n^{2-1/r^3}(h-1) + 2(h-1)w_2.$$

Proof: We will show the existence of G^* using a probabilistic argument. For ease of notation put $t = n^{2-1/r^3}$, and let $p = \frac{m-2t}{mh}$. We first show that $p > 1/(2h)$. This is equivalent to showing that $t < m/4$, and this holds for n sufficiently large since $t = o(n^2)$ and $m = |E|/h = \Theta(n^2)$.

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 mh indicator random variables, it has binomial distribution, so we can use the Chernoff inequality (cf. [1] Appendix A) to bound the deviation of $|E^*|$ from its mean:
$$\Pr[|\ |E^*| - (m-2t)\ | > t] < 2e^{-\frac{2t^2}{mh}} < 2e^{-\frac{2n^{4-2/r^3}}{n^2/2}} = 2e^{-4n^{2-2/r^3}} < \frac{1}{n}.$$
Thus, with probability at least $1 - 1/n$, $m - t \geq |E^*| \geq m - 3t$.

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

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

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

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

For the lower bound this translates to

$$d_{G^*}(v) \geq p \cdot d_G(v) - \sqrt{n \log n} = \frac{d_G(v)}{h} - \frac{2t d_G(v)}{mh} - \sqrt{n \log n} \geq$$

$$\frac{d_G(v)}{h} - \frac{2n^{2-1/r^3} n}{n^2/4} - \sqrt{n \log n} = \frac{d_G(v)}{h} - 8n^{1-1/r^3} - \sqrt{n \log n} >$$

$$\frac{d_G(v)}{h} - 9n^{1-1/r^3}.$$

For the upper bound this translates to

$$d_{G^*}(v) \leq p \cdot d_G(v) + \sqrt{n \log n} = \frac{d_G(v)}{h} - \frac{2t d_G(v)}{mh} + \sqrt{n \log n} \leq$$

$$\frac{d_G(v)}{h} - \frac{2n^{2-1/r^3} n/2}{n^2/2} + \sqrt{n \log n} =$$

$$\frac{d_G(v)}{h} - 2n^{1-1/r^3} + \sqrt{n \log n} \leq \frac{d_G(v)}{h} - n^{1-1/r^3}.$$

3. Now consider a set $W \subset V$ satisfying $n/(50r^6) \leq |W| \leq 0.9n$. Let y denote the number of edges of G with only one endpoint in W, and let y^* denote the number of edges of E^* with only one endpoint in W. Note that $w_1 \geq y - y^*$. If $|W| \leq n/2$ then $y \geq |W|(n/2 + 40(h-1)n^{1-1/r^3} - |W|)$. By elementary calculus, if n is sufficiently large then the minimum for y is obtained when $|W| = n/2$ (recall that $|W| = \Theta(n)$ so it cannot be too small) , and then

$$y \geq \frac{n}{2} 40(h-1)n^{1-1/r^3} = 20t(h-1). \tag{1}$$

If $|W| \geq n/2$ then $y \geq (n - |W|)(n/2 + 40(h-1)n^{1-1/r^3} - (n - |W|))$. Once again, if n is sufficiently large the minimum is obtained when $|W| = n/2$ and thus (1) holds in any case. Clearly, the expectation of y^* is py. Now, by the Chernoff inequality,

$$\Pr[y^* > 3py/2] = \Pr[y^* - py > py/2] < e^{-2(py/2)^2/y} = e^{-p^2 y/2} < e^{-y/(4h^2)} <$$

$$e^{-20t(h-1)/(4h^2)} << e^{-n}.$$

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

We now show that when $y^* \leq 3py/2$ then also $y - y^* \geq 10t(h-1)$. Indeed, according to (1), $0.5y \geq 10t(h-1)$. Also, trivially, $0.5y > 1.5py$ since $p < 1/h \leq 1/3$. Thus, $y = 0.5y + 0.5y > 1.5py + 10t(h-1) \geq y^* + 10t(h-1)$. Now, consider first the case where $e(G[W]) \leq t/4$. In this case,

$$w_1 \geq y - y^* \geq 10t(h-1) \geq 9.5t(h-1) + 2(h-1)e(G[W]) \geq 9.5t(h-1) + 2(h-1)w_2.$$

Now consider the case where $e(G[W]) > t/4$. The expectation of w_2 is $p \cdot e(G[W])$. Using the Chernoff inequality we obtain

$$\Pr[w_2 > \frac{e(G[W])}{h}] = \Pr[w_2 - p \cdot e(G[W]) > \frac{2t}{mh} e(G[W])]$$

$$< e^{-\frac{8t^2 e(G[W])^2}{m^2 h^2 e(G[W])}} = e^{-\frac{8t^2 e(G[W])}{m^2 h^2}} \leq e^{-\frac{2t^3}{m^2 h^2}} < e^{-\frac{2(n^{2-1/r^3})^3}{(n^2/2)^2}} < e^{-8n}.$$

Therefore, with probability at least $1 - (2/e^8)^n$, for all subsets W having $e(G[W]) > t/4$, we have that $w_2 \leq \frac{e(G[W])}{h}$. This implies that

$$w_1 = y - y^* + 2e(G[W]) - 2w_2 \geq 10t(h-1) + 2hw_2 - 2w_2 > 9.5t(h-1) + 2(h-1)w_2.$$

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

$$1 - (2/e^8)^n - (2/e)^n - 2/n - 1/n > 0$$

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

Let G and G^* be as in Lemma 4.1. Put $E' = E \setminus E^*$ and $G' = (V, E')$. Notice that for every $v \in V$

$$d_{G'}(v) = d_G(v) - d_{G^*}(v) \geq d_G(v) - \frac{d_G(v)}{h} + n^{1-1/r^3} > \frac{h-1}{h} d_G(v) > \frac{n}{3}. \quad (2)$$

Lemma 4.2 *G' contains a set L' of $k = \lfloor m/(2r^2(r-1)^2) \rfloor$ edge-disjoint copies of H^* such that each $v \in V$ appears in at least $n/(50r^6)$ elements of L'.*

Proof: Put $q = \lceil n^2/(50r^6) + n \rceil$. We begin by picking a set of q copies of H^* which we denote by $S_1, \ldots, S_q$ as follows: Suppose we have already picked $S_1, \ldots S_z$, where $0 \leq z < q$. We show how S_{z+1} is selected. Let $G_z = (V, E_z)$ be the spanning subgraph of G' consisting of the edges of E' that are not used by $S_1 \cup \ldots \cup S_z$ (initially, $G_0 = G'$). Let v be the vertex which appears the minimum number of times in $Z = \{S_1, \ldots, S_z\}$. We pick S_{z+1} to be a copy of H^* in G_z, which contains v. We must show that, indeed, there exists a copy of H^* in G_z which contains v. Since each S_i has

$2r^2(r-1)$ vertices, there are, in total, $2zr^2(r-1)$ vertices in all the elements of Z (counting with multiplicities), so by the definition of v, we have that v appears in at most $2zr^2(r-1)/n$ elements of Z. Now,

$$\frac{2zr^2(r-1)}{n} \leq \frac{2(q-1)r^2(r-1)}{n} \leq \frac{2r^2(r-1)}{n} \cdot (\frac{n^2}{50r^6}+n) < \frac{n}{24r^3}. \qquad (3)$$

Therefore, v appears in less than $n/(24r^3)$ elements of Z. Let D be the neighborhood of v in G_z, and put $d=|D|$, the degree of v in G_z. Using the fact that H^* is regular of degree $2(h-1)(r-1)$ we have by (3) that $d > d_{G'}(v) - 2(h-1)(r-1) \cdot n/(24r^3) \geq d_{G'}(v) - n/12$. (Note that, trivially, $h-1 \leq r^2$ since H' is a bipartite graph with r vertices and $h-1$ edges.) By (2), $d_{G'}(v) > n/3$, so $d > n/3 - n/12 \geq n/4$.

As in the proof of Lemma 4.1, put $t = n^{2-1/r^3}$. We claim that there are more than $3t+d$ edges of G_z with an endpoint in D. To see this, note that by (2), the sum of degrees of the vertices of D in G' is at least $nd/3 > n^2/12$. There are exactly $2(h-1)r^2(r-1)^2z$ edges which appear in G' and do not appear in G_z. Thus, the sum of the degrees of the vertices of D in G_z is greater than $n^2/12 - 2(h-1)r^2(r-1)^2z$. However,

$$\frac{n^2}{12} - 2(h-1)r^2(r-1)^2z \geq \frac{n^2}{12} - 2(h-1)r^2(r-1)^2(\frac{n^2}{50r^6}+n) \geq$$

$$\frac{n^2}{12} - \frac{n^2}{25} - \Theta(n) > \frac{n^2}{25} >> 8t >> 6t+2d.$$

So, the sum of the degrees of the vertices of D in G_z is greater than $6t+2d$, and thus there are more than $3t+d$ edges of G_z 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_z with both endpoints in D, or there are $2t$ edges in the bipartite subgraph of G_z induced by the vertex classes D and $V \setminus (D \cup \{v\})$. In the first case, by Corollary 2.3, G_z has a copy of H^* whose edges are all in $G_z[D]$, so v may take the role of any one of the vertices of this H^*, thereby proving that v appears in an H^* copy of G_z. In the second case, again by Corollary 2.3, G_z has a copy of H^* with one vertex class in D and the other in $V \setminus (D \cup \{v\})$, so v may take the role of any one of the vertices in the second vertex class of this H^*, thereby proving, again, that v appears in an H^* copy of G_z. We have proved the existence of $S_1, \ldots, S_q$.

Let $v \in V$ be arbitrary. We prove that v must appear in at least $n/(50r^6)$ elements of $\{S_1, \ldots, S_q\}$. By our construction, each S_z must contain a vertex which appears the minimum number of times in $S_1, \ldots, S_{z-1}$. Thus, there is a vertex w which was chosen as minimal at least q/n times. Let z_0 be the last stage in which w was chosen as minimal. w appears at least $q/n-1$ times in $S_1, \ldots, S_{z_0-1}$, and by the minimality of w, every $v \in V$ appears at least $q/n-1$ times in $S_1, \ldots, S_{z_0-1}$. However, $q/n - 1 \geq n/(50r^6)$, proving what we wanted.

In order to complete the proof of the lemma we need only to show how to enlarge $\{S_1, \ldots, S_q\}$ be adding to it $k-q$ additional edge-disjoint copies of H^*. Notice that by Lemma 4.1 we have

$|E'| = |E| - |E^*| = mh - |E^*| \geq mh - (m-t) = m(h-1) + t$. By Corollary 2.3, this means that we can greedily select $\lfloor m(h-1)/e(H^*) \rfloor = k$ edge-disjoint copies of H^* in G'. We selected the first q of them non-greedily as in the above process. We can thus continue selecting $k-q$ additional copies greedily from the remaining edges of E' which do not appear in $\{S_1, \ldots, S_q\}$. $\square$

Using the last lemma, and the properties of H^* shown in Lemma 2.1, and Corollary 2.3 we can now show the following:

Lemma 4.3 *G' contains a set L of m edge-disjoint copies of H' such that for any two distinct vertices $v, u \in V$, there are at least $n/(50r^6)$ elements of L which contain v as the pivot and do not contain u at all. Furthermore, if f_v denotes the total number of edges incident with v and belonging to some element of L and p_v denotes the total number of elements of L in which v is the pivot then $p_v > f_v/(2(h-1)) - r^2(r-1)^2$.*

Proof: Consider the set $L' = \{S_1, \ldots, S_q, S_{q+1}, \ldots, S_k\}$ of edge-disjoint copies of H^* that was constructed in Lemma 4.2. By Lemma 2.1, each S_i can be further decomposed into $2r^2(r-1)^2$ copies of H'. This yields a set L'' of $k \cdot 2r^2(r-1)^2$ edge-disjoint copies of H' in E'. Furthermore, each $v \in V$ appears in at least $n/(50r^6)$ elements of L', and thus, by Lemma 2.1, for any two distinct vertices vertices v and u, there are at least $n/(50r^6)$ elements of L'' which contain v as the pivot and do not contain u at all.

Now, let f''_v denote the total number of edges incident with v and belonging to an element of L''. By Lemma 2.1, H^* is regular of degree $2(h-1)(r-1)$. Thus, v appears in precisely $f''_v/(2(h-1)(r-1))$ elements of L'. By lemma 2.1 we also have that in every decomposition of S_i into $2r^2(r-1)^2$ copies of H', each vertex of S_i appears as a pivot precisely $r-1$ times. Hence, v is a pivot in precisely $f''_v/(2(h-1))$ elements of L''. Thus, $p_v \geq f''_v/(2(h-1))$.

We now show how to enlarge L'' by adding to it $m - k \cdot 2r^2(r-1)^2$ additional edge-disjoint copies of H' from the set of edges of E' not appearing in any element of L''. As in the last part of Lemma 4.2 we use the fact that $|E'| \geq m(h-1) + t$, the fact that $t > n^{2-1/v(H')} = n^{2-1/r}$, the fact that H' has $h-1$ edges, and Lemma 2.2 to obtain that we can always greedily select m edge-disjoint copies of H' in G'. We already selected the first $|L''| = k \cdot 2r^2(r-1)^2$, so we can greedily select the remaining $m - k \cdot 2r^2(r-1)^2$ additional copies of H' from the remaining edges of E' which do not appear in the elements of L''. Denote the enlarged L'' by L. Notice that $|L \setminus L''| = m - k \cdot 2r^2(r-1)^2 < 2r^2(r-1)^2$. Thus, $f_v - f''_v < 2(h-1)r^2(r-1)^2$. Hence, $p_v > f_v/(2(h-1)) - r^2(r-1)^2$. $\square$

There are $|E'| - (h-1)m$ edges of E' which are still not used by any element of L. We add these edges to E^* and obtain a set C with exactly $|E^*| + |E'| - (h-1)m = |E| - (h-1)m = m$ edges. Let R be the bipartite graph whose vertex classes are L and C. We connect an edge between an element $S \in L$ and an element $e = (x, y) \in C$ if exactly one of x or y is the pivot of S (recall that S is isomorphic to H') and the other one does not appear at all in S. Note that if S is adjacent to e in R then $S \cup e$ induces a copy of H. Thus, in order to complete the proof of Theorem 3.1 we need to show:

Lemma 4.4 *R has a perfect matching.*

Proof: By Hall's Theorem (cf.[2]) it suffices to show that for any subset $C' \subset C$, the set of neighbors of C' in R, denoted $N(C')$ satisfies $|N(C')| \geq |C'|$. We shall use the properties of L established in Lemma 4.3 and the properties of $E^* \subset C$ established in Lemma 4.1 to prove that this holds.

Let $v(C') \subset V$ be the set of vertices which are endpoints of at least one edge of C'. Let $A \subset v(C')$ be the subset of vertices which are incident with at least r edges of C', and let $B = v(C') \setminus A$ be the subset of vertices incident with less than r edges of C'.

Claim: If $v \in A$ is a pivot of some element $S \in L$ then $S \in N(C')$.

Proof: Suppose $v \in A$ is the pivot of some $S \in L$. Since $v \in A$ there are r edges of C' which have v as their endpoint. In at least one of these r edges, the other endpoint does not belong to S since S only has $r-1$ vertices other than the pivot v. By definition of R, S is adjacent to this edge of C' in R. This proves the claim.

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

- $|C'| \leq n/(50r^6)$.
  Consider an arbitrary element $(u, v) \in C'$. By Lemma 4.3, there are at least $n/(50r^6)$ elements of L which contain v as the pivot and do not contain u at all. By definition of R, all of these elements belong to $N(C')$. Hence, trivially, $|C'| \leq n/(50r^6) \leq |N(C')|$.

- $|C'| > n/(50r^6)$ and $|A| \leq \frac{\sqrt{n}}{250r^{10}}$.
  Trivially, $|C'| \leq \binom{|v(C')|}{2}$, which implies $|v(C')| > \sqrt{2|C'|} > \frac{\sqrt{n}}{5r^3}$. Thus,

  $$|B| = |v(C')| - |A| \geq \frac{\sqrt{n}}{5r^3} - \frac{\sqrt{n}}{250r^{10}} > \frac{\sqrt{n}}{10r^3}.$$

  Consider a vertex $u \in B$, and let $(u, w) \in C'$ be arbitrary. According to Lemma 4.3, there are at least $n/(50r^6)$ elements of L which contain u as the pivot and do not contain w. All of these elements are neighbors of (u, w) in R. Thus, they are all in $N(C')$. Since this is true for every $u \in B$, we have at least $|B|n/(50r^6)$ elements of L counted in this way, and since no copy of L is counted more than once (there is only one pivot in each copy) we get that

  $$|N(C')| \geq \frac{|B|n}{50r^6}.$$

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

$$\frac{n|A| + r|B|}{2} \le \frac{|B|n}{50r^6}.$$

This is equivalent to showing that

$$|A| \le (\frac{1}{25r^6} - \frac{r}{n})|B|.$$

Indeed,

$$|A| \le \frac{\sqrt{n}}{250r^{10}} \le (\frac{1}{25r^6} - \frac{r}{n})\frac{\sqrt{n}}{10r^3} \le (\frac{1}{25r^6} - \frac{r}{n})|B|.$$

- $|C'| > n/(50r^6)$ and $\frac{\sqrt{n}}{250r^{10}} < |A| < n/(50r^6)$.
  Clearly, $|C'| \le |B|(r-1) + \binom{|A|}{2} < nr + \binom{|A|}{2}$. By Lemma 4.3, every $u \in A$ appears as a pivot in at least $n/(50r^6)$ elements of L. By the claim proved above, if u appears in some $S \in L$ then $S \in N(C')$. Thus, every $u \in A$ contributes at least $n/(50r^6)$ elements to $N(C')$, and every such element S_i is counted at most once by these contributions. Therefore,

  $$|N(C')| \ge \frac{n}{50r^6}|A|.$$

  Now, for $\frac{\sqrt{n}}{250r^{10}} < |A| < n/(50r^6)$ we have

  $$|C'| < nr + \binom{|A|}{2} < \frac{n}{100r^6}|A| + \frac{n}{100r^6}|A| \le \frac{n}{50r^6}|A| \le |N(C')|$$

  as required.

- $|C'| > n/(50r^6)$ and $n/(50r^6) \le |A| \le 0.9n$.
  In this case, we can use Lemma 4.1 applied to A as W. Let z_1 denote the number of elements of C with both endpoints in A, and let w_2 denote the number of edges of E^* with both endpoints in A. Since, by Lemma 4.1, $|C| - |E^*| \le 3n^{2-1/r^3}$ we have that $z_1 \le 3n^{2-1/r^3} + w_2$, and thus

  $$|C'| \le |B|(r-1) + z_1 < rn + z_1 \le rn + 3n^{2-1/r^3} + w_2.$$

  Hence, we must show that $|N(C')| \ge rn + 3n^{2-1/r^3} + w_2$. According to Lemma 4.1, if w_1 denotes the sum of the degrees of the vertices of A in $G' = (V, E')$ then

  $$w_1 \ge 9.5n^{2-1/r^3}(h-1) + 2(h-1)w_2.$$

  Let w^* denote the sum of the degrees of the vertices of A in the union of all the elements of L. Only the edges of $C \setminus E^*$ appear in E' and do not appear in an element of L. Thus, $w^* \ge w_1 - 6n^{2-1/r^3}$, and therefore $w^* \ge 9.5n^{2-1/r^3}(h-1) + 2(h-1)w_2 - 6n^{2-1/r^3}$.

  For $v \in A$, let f_v denote the number of edges incident with v that appear in an element of L. Hence $w^* = \sum_{v \in A} f_v$. Let p_v denote the number of elements of L in which v is a pivot. By

Lemma 4.3, $p_v > f_v/(2(h-1)) - r^2(r-1)^2$. If $v \in A$ is a pivot of an element of L then by our claim on the vertices of A we have that this element is in $N(C')$. Since each element of L only has one pivot we get that $|N(C')| \geq \sum_{v \in A} p_v$. We now have

$$|N(C')| \geq \sum_{v \in A} p_v > \sum_{v \in A} \left(\frac{f_v}{2(h-1)} - r^2(r-1)^2 \right) = \frac{w^*}{2(h-1)} - |A| r^2 (r-1)^2 \geq$$

$$4.75 n^{2-1/r^3} + w_2 - \frac{3n^{2-1/r^3}}{h-1} - nr^2(r-1)^2 \geq rn + 3n^{2-1/r^3} + w_2$$

as required.

- $|C'| > n/(50r^6)$ and $|A| > 0.9n$.
  Let $n' = n - |A|$ be the size of $V \setminus A$. The conditions imply that $n' \leq 0.1n$. If some $S_i \in L$ is not in $N(C')$ then, according to the claim proved above, S_i contains no vertex of A, and therefore all its r vertices are in $V \setminus A$. Thus, $|N(C')| \geq m - \binom{n'}{2}/(h-1)$. We need to show that $|C'| \leq m - \binom{n'}{2}/(h-1)$, or, equivalently, that $|C \setminus C'| \geq \binom{n'}{2}/(h-1)$. Consider a vertex $v \in V \setminus A$. By Lemma 4.1,

$$d_{G^*}(v) \geq \frac{d_G(v)}{h} - 9n^{1-1/r^3} \geq \frac{n}{2h} - 9n^{1-1/r^3} \geq \frac{n}{3h}.$$

  Thus, there are at least $n/(3h)$ edges of C having v as an endpoint. Either $v \notin v(C')$ or $v \in B$. In any case, v is an endpoint of at most $r-1$ edges of C'. Thus, there are at least $n/(3h) - r + 1$ edges of $C \setminus C'$ having v as an endpoint. Since this is true for every $v \in V \setminus A$, we have that there are at least $(n/(3h) - r + 1)n'/2$ edges in $C \setminus C'$. Thus, we must show that $(n/(3h) - r + 1)n'/2 \geq n'(n'-1)/(2h-2)$. This is equivalent to showing that

$$n' \leq \frac{n}{3} - \frac{n}{3h} - r(h-1) + h.$$

  Indeed, this holds for n sufficiently large since $n' \leq 0.1$ and $h \geq 3$. □

Acknowledgments

The author thanks Peter Winkler and Jeff Kahn for the example showing that $f_{C_4}(n) \geq 0.6n(1 + o(1))$ and the referees for extremely valuable comments.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, John Wiley and Sons Inc., New York, 1991.

[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] 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.

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

[6] 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.

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

[8] R. Yuster, *Tree decomposition of graphs*, Random Structures and Algorithms 12 (1998), 237-251.

[9] R. Yuster, *Decomposing large graphs with small graphs of high density*, Journal of Graph Theory 32 (1999), 27-40.

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