Tree Decomposition of Graphs

Raphael Yuster
Department of Mathematics
University of Haifa-ORANIM
Tivon 36006, Israel.
e-mail: raphy@math.tau.ac.il

Abstract

Let H be a tree on $h \geq 2$ vertices. It is shown that if $G=(V,E)$ is a graph with $\delta(G) \geq \frac{|V|}{2}+10h^4\sqrt{|V|\log|V|}$, and $h-1$ divides $|E|$ then there is a *decomposition* of the edges of G into copies of H. This result is asymptotically best possible for all trees with at least three vertices.

1 Introduction

All graphs considered here are finite, undirected, and have no loops or multiple edges, unless otherwise noted. For the standard graph-theoretic notations the reader is referred to [2]. Let H be a connected graph. We say that a graph G has an *H-decomposition* if there exists a set L of subgraphs of G, which are isomorphic to H, such that every edge of G appears in exactly one member of L. 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 graph G (assuming H is a fixed graph) we can verify in polynomial time if G satisfies these two conditions. For convenience, we say that G has the property $P(H)$, if G satisfies these necessary conditions.

The combinatorial and computational aspects of the H-decomposition problem have been studied extensively. The focus of the combinatorial research is to find naturally-expressible *sufficient* conditions that guarantee that a graph G satisfying these conditions, and having $P(H)$, has an H-decomposition. Indeed, Wilson has proved in [8] that if $G=K_n$ where $n \geq n_0 = n_0(H)$, and G has $P(H)$, then G has an H-decomposition. Wilson's result can be thought of as a *minimum degree* result, where the minimum degree is the highest possible, i.e. $n-1$. Following Wilson, Gustavsson has shown in [6] that if G is an n-vertex graph, $\delta(G) \geq (1-\epsilon(H))n$, where $\epsilon(H)$ is some small positive constant depending on H, and G has $P(H)$, then G has an H-decomposition. However, 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, $\epsilon(H) \leq 10^{-24}/h$. It is believed, however, that the correct value for $\epsilon(H)$ is much larger. In fact, Nash-Williams conjectured in [7] that when H is a triangle, then $\epsilon(H) = 1/4$, and he also gives an example showing that this would be best possible. The general problem can therefore be expressed as follows:

Problem 1: Determine $f_H(n)$, the smallest possible integer, such that whenever G has n vertices (where $n \geq n_0(H)$), and $\delta(G) \geq f_H(n)$, and G has $P(H)$, then G also has an H-decomposition.

It is not difficult to show that $f_H(n) > n/2 - 2$ for every connected graph H with at least 3 vertices (if H is a single edge, the decomposition problem becomes trivial). To see this, consider the graph on n vertices, where $n = 2x$ is even and $e(H)$ divides $x(x-1)$. Let G be the graph on two vertex-disjoint K_x's. G has n-vertices and $\delta(G) = x - 1$. If $e(H)$ does not divide $x(x-1)/2$, then G does not have an H-decomposition. Otherwise, delete from the first clique one edge, and delete from the second clique $e(H) - 1$ independent edges (this can be done if, say, $x \geq 2(e(H) - 1)$. The obtained graph G' has $\delta(G') = x - 2$, $e(H)$ divides $e(G')$, but G' does not have an H-decomposition. This shows that $f_H(n) \geq n/2 - 1$ whenever $gcd(H) = 1$ and n is even. It is also easy to construct similar examples when $gcd(H) > 1$, or when n is odd (or both).

The purpose of this paper is to solve Problem 1, at least asymptotically, in case H is a tree. The result is summarized in the following theorem:

Theorem 1.1 *Let H be any tree with $h \geq 2$ vertices. Let G be a graph on n vertices satisfying $P(H)$ and $\delta(G) \geq n/2 + 10h^4\sqrt{n \log n}$. Then G has an H-decomposition.*

Note that whenever H is a tree, $gcd(H) = 1$, so $P(H)$ reduces to having $h-1$ divide $e(G)$. Stated in the language of Problem 1, Theorem 1.1 shows that $f_H(n) \leq n/2 + 10h^4\sqrt{n \log n}$, and we therefore have that for every tree on at least three vertices, $n/2 - 2 < f_H(n) \leq n/2 + 10h^4\sqrt{n \log n}$. Thus, $f_H(n)/n$ is asymptotically determined for trees. We note that the previously best known result for general trees was Gustavsson's result, mentioned above.

Theorem 1.1 is a special case of a more general theorem, which states that graphs having good expansion properties and have $P(H)$, also have an H-decomposition, in case H is a tree. A graph $G = (V, E)$ is called r *edge-expanding* if for every nonempty $X \subset V$ with $|X| \leq |V|/2$, there are at least $r|X|$ edges between X and $V \setminus X$.

Theorem 1.2 *Let H be any tree with $h \geq 2$ vertices. Let G be a graph on n vertices having $P(H)$ and which is $10h^4\sqrt{n \log n}$ edge-expanding. Then G has an H-decomposition.*

Note that Theorem 1.1 follows from Theorem 1.2 since, clearly, every graph on n vertices having $\delta(G) \geq n/2 + r$ is r edge-expanding.

In the following section we prove two lemmas which are needed for the proof of Theorem 1.2. The proof of Theorem 1.2 appears in Section 3. Section 4 contains some concluding remarks, mainly dealing with the algorithmic aspect of Theorem 1.2, and an open problem. Most of the proofs apply probabilistic arguments. Throughout this paper all the logarithms are natural.

2 The lemmas

For the rest of this paper, let H be a fixed tree on $h \geq 3$ vertices. A graph $G = (V, E)$ is called *feasible* if it satisfies the conditions of Theorem 1.2. Namely, $|V| = n$, $|E| = m(h-1)$ where m is a positive integer, and G is $10h^4\sqrt{n \log n}$ edge-expanding.

Lemma 2.1 *Let $G = (V, E)$ be a feasible graph. Then E can be partitioned into $h-1$ subsets $E_1, \ldots, E_{h-1}$, such that $|E_i| = m$ and if the degree of a vertex $v \in V$ in $G_i = (V, E_i)$ is denoted by $d_i(v)$, then for every $v \in V$ we have*

$$|d_i(v) - \frac{d(v)}{h-1}| \leq 2.5\sqrt{d(v) \log n}.$$

($d(v)$ denotes the degree of v in G.) Furthermore, each spanning subgraph G_i is $5h^3\sqrt{n \log n}$ edge-expanding.

Proof: First note that an r-expanding graph must have minimum degree at least r, so for each $v \in V$ we have $n > d(v) \geq 10h^4\sqrt{n \log n}$. Therefore, we also have

$$n^{1/8} > (\frac{n}{\log n})^{1/8} > h \geq 3.$$

We let each edge $e \in E$ choose a random integer between 0 and $h-1$, where 0 is chosen with probability $\beta = n^{-1/2}$ and the other numbers are chosen with probability $\alpha = (1-\beta)/(h-1)$. All the choices are independent. For $i = 0, \ldots, h-1$, let $F_i \subset E$ be the set of edges which selected i. Let $d'_i(v)$ be the number of edges adjacent to v which belong to F_i. Clearly, $E[|F_i|] = \alpha|E| = m(1-\beta)$, for $i \neq 0$. We may use the large deviation result of Chernoff (cf., e.g. [1] Appendix A) to derive that for $i \neq 0$

$$\text{Prob}[|F_i| > m] = \text{Prob}[|F_i| - m(1-\beta) > m\beta] < \exp(-\frac{2m^2\beta^2}{m(h-1)}) = \tag{1}$$

$$\exp(-\frac{2m}{n(h-1)}) \leq \exp(-\frac{10h^4\sqrt{n \log n}}{(h-1)^2}) < 1/n.$$

Similarly, we have that for all $i \neq 0$ and for all $v \in V$

$$\text{Prob}[|d'_i(v) - \alpha d(v)| > \sqrt{d(v)\log n}] < 2\exp(\frac{-2d(v)\log n}{d(v)}) = \frac{2}{n^2}. \tag{2}$$

Similarly, for $i = 0$ we get

$$\text{Prob}[|d'_0(v) - \beta d(v)| > \sqrt{d(v)\log n}] < 2/n^2. \tag{3}$$

Since

$$(h-1)\cdot(1/n)+n(h-1)\cdot(2/n^2)+n\cdot(2/n^2) = (h-1)/n+2h/n < 3h/n < 3h/h^8 = 3/h^7 \leq 1/3^6 < 0.1$$

we have by inequalities (1) (2) and (3) that with probability greater than 0.9, all of the following events hold:

1. $|F_i| \leq m$ for $i = 1, \ldots, h-1$.
2. $|d'_i(v) - \alpha d(v)| \leq \sqrt{d(v)\log n}$ for all $i = 1, \ldots, h-1$ and for all $v \in V$.
3. $|d'_0(v) - \beta d(v)| \leq \sqrt{d(v)\log n}$ for all $v \in V$.

Consider, therefore, a partition of E into $F_0, \ldots, F_{h-1}$ in which all of these events hold. Since $|F_i| \leq m$, we may partition F_0 into $h-1$ subsets $Q_1, \ldots, Q_{h-1}$, where $|Q_i| = m - |F_i|$. Put $E_i = F_i \cup Q_i$ for $i = 1, \ldots, h-1$. Note that $|E_i| = m$ and $E_i \cap E_j = \emptyset$ for $1 \leq i < j \leq h-1$. Put $G_i = (V, E_i)$ and let $d_i(v)$ be the degree of v in G_i. Clearly,

$$d_i(v) \geq d'_i(v) \geq \alpha d(v) - \sqrt{d(v)\log n} = \frac{d(v)}{h-1} - \frac{d(v)}{\sqrt{n}(h-1)} - \sqrt{d(v)\log n} \geq \tag{4}$$

$$\frac{d(v)}{h-1} - \frac{\sqrt{d(v)}}{h-1} - \sqrt{d(v)\log n} \geq \frac{d(v)}{h-1} - \sqrt{d(v)}(2\sqrt{\log n}).$$

We also need to bound $d_i(v)$ from above:

$$d_i(v) \leq d'_i(v) + d'_0(v) \leq \alpha d(v) + \beta d(v) + 2\sqrt{d(v)\log n} =$$

$$\frac{d(v)}{h-1} - \frac{d(v)}{\sqrt{n}(h-1)} + 2\sqrt{d(v)\log n} + \frac{d(v)}{\sqrt{n}} \leq \tag{5}$$

$$\frac{d(v)}{h-1} + 2\sqrt{d(v)\log n} + \frac{d(v)}{\sqrt{n}} \leq \frac{d(v)}{h-1} + 2\sqrt{d(v)\log n} + \sqrt{d(v)} \leq \frac{d(v)}{h-1} + 2.5\sqrt{d(v)\log n}.$$

It now follows from inequalities (4) and (5) that $|d_i(v) - \frac{d(v)}{h-1}| \leq 2.5\sqrt{d(v)\log n}$.

Consider the partition of E into $F_0, \ldots, F_{h-1}$. We have already shown that with probability greater than 0.9 this partition is good in the sense that one may obtain the desired partition into the subsets

E_i by transferring vertices from F_0 to the F_i's. This, however, does not guarantee that the graphs $G_i = (V, E_i)$ are $5h^3\sqrt{n \log n}$ edge-expanding, as required. Since r edge-expansion is a monotone-increasing property, it suffices to show that with probability at least $1 - 0.9 = 0.1$, all of the graphs $G'_i = (V, F_i)$ are $5h^3\sqrt{n \log n}$ edge-expanding. We prove this as follows: Let $X \subset V$ with $|X| \le n/2$. Let $n_i(X)$ denote the number of edges between X and $V \setminus X$ in G'_i. Our aim is to show that $n_i(X) \ge 5|X|h^3\sqrt{n \log n}$ for all $i = 1, \ldots, h-1$ and for all X, with probability at least 0.1. Let $n(X)$ be the number of edges between X and $V \setminus X$ in G. Since G is $10h^4\sqrt{n \log n}$ expanding we have that

$$n(X) \ge 10|X|h^4\sqrt{n \log n}.$$

Clearly, $E[n_i(X)] = \alpha n(X)$. Applying the large deviation bound once again we have

$$\text{Prob}[|n_i(X) - \alpha n(X)| > \alpha n(X)/2] < 2\exp(-\frac{2n(X)^2\alpha^2/4}{n(X)}) = 2\exp(-n(X)\alpha^2/2) \le$$

$$2\exp(-n(X)/(2h^2)) \le 2\exp(-5|X|h^2\sqrt{n \log n}) < \frac{2}{n^{|X|h^2}} < \frac{1}{n(h-1)\binom{n}{|X|}}$$

with lots of room to spare in the last part of this inequality. Since there are $\binom{n}{|X|}$ sets of size $|X|$, and since there are $n/2$ possible sizes to consider, we get from the last inequality that with probability at least $0.5 > 0.1$, for all $i = 1, \ldots, h-1$ and for all sets $X \subset V$ with $|X| \le n/2$,

$$|n_i(X) - \alpha n(X)| \le \alpha n(X)/2.$$

In particular this means that

$$n_i(X) \ge \alpha n(X)/2 \ge \frac{1 - 1/\sqrt{n}}{h-1} 5|X|h^4\sqrt{n \log n} \ge 5|X|h^3\sqrt{n \log n}.$$

□

We call a partition of E into the subsets E_i having the properties guaranteed by Lemma 2.1 a *feasible partition*. Given a feasible partition of a feasible graph, our next goal is to orient the edges of every E_i, such that the oriented sets, denoted by E_i^* have certain properties. Let $d_i^+(v)$ and $d_i^-(v)$ denote the outdegree and indegree of v in E_i^*, respectively. Clearly, $d_i(v) = d_i^+(v) + d_i^-(v)$ for all $v \in V$ and $i = 1, \ldots, h-1$. In order to define the properties which we require from our orientation, we need several definitions.

Let q be a leaf of H. Fix a rooted orientation $H(q)$ of H where the root of H is q. Such an orientation can be obtained by performing a Breadth-First Search (BFS) (cf. [3]) of H which originates from q. Let $e_1, \ldots, e_{h-1}$ be the oriented edges of $H(q)$, in the order they are discovered by the BFS. Note that for $i = 2, \ldots, h-1$, the edge $e_i = (x, y)$ has a unique *parent-edge*, which is the unique edge e_j entering x. (Thus, $e_j = (z, x)$ for some z). The edge e_1 is the only edge which has no parent, since it is the only edge emanating from q. For $i = 2, \ldots, h-1$, let $p(i) = j$ if e_j is the parent of e_i. Note that $p(i) < i$. We say that j is a *descendent* of i if $j = i$ or if $p(j)$ is a descendent of i. Note that this definition is recursive.

An orientation of a feasible partition is called a *feasible orientation* if for all $v \in V$, $d^-_{p(i)}(v) = d^+_i(v)$, where $i = 2, \ldots, h-1$, and $|d^+_i(v) - d^-_i(v)| \leq 5i\sqrt{n \log n}$, for all $i = 1, \ldots, h-1$. Note that the second requirement implies also that $|d^+_i(v) - d_i(v)/2| \leq 2.5h\sqrt{n \log n}$ and, similarly, $|d^-_i(v) - d_i(v)/2| \leq 2.5h\sqrt{n \log n}$.

Lemma 2.2 *Every feasible partition of a feasible graph has a feasible orientation. Furthermore, in every feasible orientation*

$$d^+_i(v) \geq 4h^3\sqrt{n \log n} \tag{6}$$

holds for all $v \in V$ and for all $i = 2, \ldots, h-1$.

Proof: We show how to construct our orientation in $h-1$ stages, where in stage i we orient the edges of E_i and form E^*_i. We begin by orienting E_1. It is well-known by Euler's Theorem (cf. [2]), that the edges of every undirected graph can be oriented such that the indegree and outdegree of every vertex differ by at most 1. Such an orientation is called Eulerian. We therefore let E^*_1 be any Eulerian orientation of E_1. Thus $|d^+_1(v) - d^-_1(v)| \leq 1 \leq 5\sqrt{n \log n}$. Assume now that we have oriented all the subsets E_j for $1 \leq j < i$, such that the conditions of a feasible orientation hold for j. We show how to orient the edges of E_i, such that the conditions also hold for i. Let $j = p(i)$, and put $c_v = d^-_j(v)$. We are required to orient the edges of E_i such that for every $v \in V$, $d^+_i(v) = c_v$. Our initial goal is to show that $|d^+_i(v) - d^-_i(v)| \leq 5i\sqrt{n \log n}$. Our second goal is to show that such an orientation exists. The following inequality achieves the first goal:

$$|d^+_i(v) - d^-_i(v)| = |2c_v - d_i(v)| = |2d_j(v) - 2d^+_j(v) - d_i(v)| \leq |2d^+_j(v) - d_j(v)| + |d_j(v) - d_i(v)| =$$

$$|d^+_j(v) - d^-_j(v)| + |d_j(v) - d_i(v)| \leq 5j\sqrt{n \log n} + |d_j(v) - \frac{d(v)}{h-1}| + |d_i(v) - \frac{d(v)}{h-1}| \leq$$

$$5j\sqrt{n \log n} + 5\sqrt{d(v) \log n} \leq 5i\sqrt{n \log n}.$$

We now need to show that the desired orientation exists. Note that $\sum_{v \in V} c_v = m$ and hence the desired orientation exists if every vertex v can select c_v edges from the $d_i(v)$ edges adjacent to v, and such that every edge of E_i is selected by exactly one of its endpoints. To prove this is possible we define a bipartite graph B as follows. B has two vertex classes of size m each. One vertex class is E_i, while the other vertex class, denoted by S, contains c_v copies of each v. Thus,

$S = \{v^{(k)} \mid v \in V,\ 1 \leq k \leq c_v\}$. The edges of B are defined as follows. A member $v_k \in S$ is connected to $e \in E_i$ if v is an endpoint of e. Clearly, our aim is to show that B has a perfect matching. By Hall's Theorem (cf. [2]), it suffices to show that for every set $S' \subset S$, $|N(S')| \geq |S'|$ where $N(S') \subset E_i$ are the neighbors of S' in B. Fix $\emptyset \neq S' \subset S$. Let $V' = \{v \in V \mid v^{(k)} \in S'\}$. Put $V' = \{v_1, \ldots, v_t\}$. Clearly, $|S'| \leq \sum_{l=1}^{t} c_{v_l}$. Note that $N(S')$ contains all the edges of E_i which have an endpoint in V'. Let $T_1 \subset E_i$ be the set of edges having only one endpoint in V' and let $T_2 = N(S') \setminus T_1$ be the set of edges of E_i having both endpoints in V'. Put $t_1 = |T_1|$ and $t_2 = |T_2|$. Clearly, $t_1 + 2t_2 = \sum_{l=1}^{t} d_i(v_l)$. We first consider the case $t \leq n/2$. Since $G_i = (V, E_i)$ is $5h^3\sqrt{n \log n}$ edge-expanding and since $|V'| = t \leq n/2$, we have $t_1 \geq 5h^3 t\sqrt{n \log n}$. Now,

$$|N(S')| = t_1 + t_2 = \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{t_1}{2} \geq (\sum_{l=1}^{t} \frac{d_i(v_l)}{2}) + 2.5h^3 t\sqrt{n \log n} >$$

$$\sum_{l=1}^{t} (\frac{d_i(v_l)}{2} + 2.5h\sqrt{n \log n}) \geq \sum_{l=1}^{t} c_{v_l} \geq |S'|.$$

The case where $t > n/2$ is proved as follows. Put $V'' = V \setminus V' = \{v_{t+1}, \ldots, v_n\}$. Note that T_1 is the set of edges connecting V' with V''. Since G_i is $5h^3\sqrt{n \log n}$ edge-expanding and since $|V''| \leq n/2$ we have $t_1 \geq 5h^3(n-t)\sqrt{n \log n}$. Now,

$$|N(S')| = t_1 + t_2 = \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{t_1}{2} \geq (\sum_{l=1}^{t} \frac{d_i(v_l)}{2}) + 2.5h^3(n-t)\sqrt{n \log n} >$$

$$m - \sum_{l=t+1}^{n} (\frac{d_i(v_l)}{2} - 2.5h\sqrt{n \log n}) \geq m - \sum_{l=t+1}^{n} c_{v_l} = \sum_{l=1}^{t} c_{v_l} \geq |S'|.$$

Finally, we need to show that (6) holds. We use the fact that $|d_i^+(v) - d_i(v)/2| \leq 2.5h\sqrt{n \log n}$ and Lemma 2.1 which states that

$$|d_i(v) - \frac{d(v)}{h-1}| \leq 2.5\sqrt{d(v) \log n}$$

and the fact that $h \geq 3$ to obtain that

$$|d_i^+(v) - \frac{d(v)}{2(h-1)}| \leq 3h\sqrt{n \log n}.$$

Thus,

$$d_i^+(v) \geq \frac{d(v)}{2(h-1)} - 3h\sqrt{n \log n} \geq \frac{10h^4\sqrt{n \log n}}{2(h-1)} - 3h\sqrt{n \log n} \geq 4h^3\sqrt{n \log n}.$$

$\square$

3 The proof of the main result

We begin this section by showing that a feasible orientation defines a decomposition of the edges of a feasible graph G into a set L^* of m edge-disjoint connected graphs, each graph having $h-1$ edges, one from each E_i. Furthermore, each of these graphs is homomorphic to $H(q)$ (and, thus, to H), in the sense that every member of L^* which happens to be a tree, is *isomorphic* to H. Unfortunately, not all the members of L^* are necessarily trees, and we will need to mend L^* in order to obtain our desired decomposition.

We now describe the process which creates L^*. Fix a feasible orientation of G, and let E_i^* denote the oriented edges of E_i. Let $D_i^+(v) \subset E_i^*$ denote those edges of E_i^* which emanate from v, and let $D_i^-(v) \subset E_i^*$ be the edges of E_i^* which enter v. For $i = 2, \ldots, h-1$ and for all $v \in V$ we know that $|D_{p(i)}^-(v)| = |D_i^+(v)| = d_i^+(v)$. Therefore, let $B_i(v)$ be a *perfect matching* between $D_{p(i)}^-(v)$ and $D_i^+(v)$. (Note that there are $d_i^+(v)!$ different ways to select $B_i(v)$, so we pick one arbitrarily). The members of $B_i(v)$ are, therefore, pairs of edges in the form $((x,v),(v,y))$ where $(x,v) \in D_{p(i)}^-(v)$ and $(v,y) \in D_i^+(v)$. We say that (x,v) and (v,y) are *matched* if $((x,v),(v,y)) \in B_i(v)$ for some i. The transitive closure of the "matched" relation defines an equivalence relation where the equivalence classes are connected directed graphs, each having $h-1$ edges, one from each E_i^*, and which are homomorphic to $H(q)$, by the homomorphism which maps the edge e_i of $H(q)$ to the edge belonging to E_i^* in an equivalence class. Thus, L^* is the set of all of these graphs, (or, in set theoretical language, the quotient set of the equivalence relation). Note that although each $T \in L^*$ is homomorphic to $H(q)$, it is not necessarily isomorphic to $H(q)$ since T may contain cycles. For a simple example, consider the case where $H(q)$ is a directed path on 3 edges (q,a,b,c). It may be the case that T is composed of the edges $(x,y) \in E_1^*$, $(y,z) \in E_2^*$ and $(z,x) \in E_3^*$. Thus T is a directed triangle, but not a directed path on 3 edges. It is clear, however, that if T happens to be a tree, (or, equivalently, if T contains h vertices) then it *is* isomorphic to $H(q)$.

As noted, there are many ways to create L^*. In fact, there are

$$\Pi_{i=2}^{h-1}\Pi_{v\in V} d_i^+(v)!$$

different ways to create the decomposition L^*. Our goal is to show that in at least one of these decompositions, all the members of L^* are, in fact, trees, and this will conclude Theorem 1.2. Before proceeding with the proof of Theorem 1.2, we require a few definitions.

For a member $T \in L^*$, and for $i = 1, \ldots, h-1$, let T_i be the subgraph of T which consists only of the first i edges, namely those belonging to $E_1^* \cup \ldots \cup E_i^*$. Note that T_i is a connected subgraph of T. Let $T(i)$ be the edge of T which belongs to E_i^*. Note that for $i > 1$, T_i is obtained from T_{i-1} by adding the edge $T(i)$. Now, suppose T_{i-1} is a tree, and T_i is not a tree. Let $T(i) = (v,u)$. (Note that $(v,u) \in D_i^+(v)$ in this case). It follows that u already appears in T_{i-1}. We therefore call an edge $T(i)=(v,u)$ *bad* if u already appears in T_{i-1}. Otherwise, the edge is called *good*. Clearly, T is a tree iff all its $h-1$ edges are good. Let

$$N(v,i) = \{T \in L^* \ |T(i) \in D_i^+(v), T(j) \ \ is \ \ bad \ \ for \ \ some \ \ j \geq i\}.$$

Clearly, $|N(v,i)| \leq d_i^+(v)$. Our first goal is to show that if all the $n(h-2)$ perfect matchings $B_i(v)$ are selected randomly and independently, then with high probability, $|N(v,i)|$ is significantly smaller than $d_i^+(v)$.

Lemma 3.1 *If all the perfect matchings $B_i(v)$ are selected randomly and independently, then with probability at least* 0.9*, for all $i = 1,\dots,h-1$ and for all $v \in V$, $|N(v,i)| \leq h\sqrt{d_i^+(v)}$.*

Proof: For $j \geq i$, let

$$N(v,i,j) = \{T \in L^* \ |T(i) \in D_i^+(v), T(j) \ \ is \ \ bad\}.$$

Clearly, $|N(v,i)| \leq \sum_{j=i}^{h-1} |N(v,i,j)|$. We will therefore estimate the $|N(v,i,j)|$'s. Since the perfect matchings are selected randomly and independently, we may assume that the n matchings $B_j(u)$ for all $u \in V$ are selected *after* all the other $n(h-3)$ matchings $B_k(u)$, for $k \neq j$, are selected. Prior to the selection of the last n matchings, the transitive closure of the "matched" relation defines two sets M^* and N^* each having m members. Each member in M^* is a subgraph containing the edges of an equivalence class, with exactly one edge from each E_r^* where r is a descendent of j. Each member of N^* is a subgraph containing the edges of an equivalence class, with exactly one edge from each E_r^* where r is *not* a descendent of j (note that if $j=1$ then $i=1$ and since $N(v,1,1)=0$ always, we may assume $j>1$, and thus N^* is not empty). Note that the matchings $B_j(u)$ for all $u \in V$ match the members of M^* with the members of N^*, and each such match produces a member of L^*. Let us estimate $|N(v,i,j)|$ given that we know exactly what N^* contains; i.e. we shall estimate $\{|N(v,i,j)| \ \mid \ N^*\}$. Consider a set $U = \{(x_1,u_1),(x_2,u_2),\dots,(x_k,u_k)\}$ of k edges, where for $t=1,\dots,k$, $(x_t,u_t) \in D_{p(j)}^-(u_t)$, and (x_t,u_t) belongs to a member T^t of N^* containing an edge of $D_i^+(v)$. The last requirement is valid since all the edges of $D_i^+(v)$ belong to members of N^* because i is not a descendent of j. Similarly, the edges of $D_{p(j)}^-(u_t)$ belong to members of N^* since $p(j)$ is not a descendent of j. We call U bad, if for all $t=1,\dots,k$, (x_t,u_t) is matched in $B_j(u_t)$ to an edge $(u_t,y_t) \in D_j^+(u_t)$ where y_t already appears in T^t. (Note that the edges in $D_j^+(u_t)$ belong to members of M^*). Since there are less than h vertices in T^t, and since $B_j(u_t)$ is selected at random, we have that

$$Prob[(x_t,u_t) \ \ is \ \ matched \ \ in \ \ B_j(u_t) \ \ to \ \ a \ \ bad \ \ edge] < \frac{h}{d_j^+(u_t)}.$$

Similarly, the probability that (x_t, u_t) is matched in $B_j(u_t)$ to a bad edge, *given that* (x_s, u_s) is matched in $B_j(u_s)$ to a bad edge, for all $1 \le s < t$, is less than $h/(d_j^+(u_t) - (t-1))$. Thus,

$$\text{Prob}[U \;\; is \;\; bad] < \Pi_{t=1}^{k} \frac{h}{d_j^+(u_t) - t + 1}.$$

Assuming $k \le d_j^+(u_t)/2$, and using (6) we have

$$\text{Prob}[U \;\; is \;\; bad] < (\frac{1}{2h^2\sqrt{n \log n}})^k.$$

Consequently,

$$\text{Prob}[|N(v,i,j)| \ge k \;\; | \;\; N^*] < \binom{d_i^+(v)}{k} (\frac{1}{2h^2\sqrt{n \log n}})^k.$$

Note that the estimation in the last inequality does not depend on N^*, and thus,

$$\text{Prob}[|N(v,i,j)| \ge k] < \binom{d_i^+(v)}{k} (\frac{1}{2h^2\sqrt{n \log n}})^k.$$

Let $k = \lfloor \sqrt{d_i^+(v)} \rfloor$. (Note that this choice of k still satisfies $k \le \sqrt{n} \le d_j^+(u_t)/2$). Using the fact that $\binom{x}{\lfloor \sqrt{x} \rfloor} \le (e\sqrt{x})^{\sqrt{x}}$ we have that

$$\text{Prob}[|N(v,i,j)| \ge \sqrt{d_i^+(v)}] \le (e\sqrt{d_i^+(v)})^{\sqrt{d_i^+(v)}} (\frac{1}{2h^2\sqrt{n \log n}})^{\sqrt{d_i^+(v)}} \le$$

$$(\frac{e}{2h^2})^{\sqrt{d_i^+(v)}} \le 2^{-n^{1/4}} < \frac{1}{10nh^2}.$$

Thus, with probability at least $1 - nh^2/(10nh^2) \ge 0.9$, for all $1 \le i \le j \le h-1$, and for all $v \in V$, $|N(v,i,j)| \le \sqrt{d_i^+(v)}$. In particular, $|N(v,i)| \le h\sqrt{d_i^+(v)}$. $\Box$

For two vertices $u, v \in V$ (not necessarily distinct) and for two indices $0 \le j < i \le h-1$ let $L([u,j],[v,i])$ denote the set of members of L^* which contain an edge of $D_i^-(v)$ and also contain an edge of $D_j^-(u)$. Note that when $j = 0$, $D_j^-(u)$ is undefined, so we define $D_0^-(u) = D_1^+(u)$ and $d_0^-(u) = d_1^+(u)$ in this case only. For the sake of symmetry, define $L([v,i],[u,j]) = L([u,j],[v,i])$, and define $L([u,i],[v,i]) = 0$, when $u \ne v$.

Lemma 3.2 *If all the perfect matchings $B_i(v)$ are selected randomly and independently, then, with probability at least 3/4, for every $u, v \in V$ and for $0 \le j < i \le h-1$,*

$$|L([u,j],[v,i])| \le 2\sqrt{n \log n}. \tag{7}$$

Proof: Consider first the case where $j = p(i)$ or $i = 1$ and $j = 0$. In this case, $L([u,j],[v,i])$ is simply the set of members of L^* which contain (u,v) as their edge from E_i^*. Trivially, this set is empty if $(u,v) \notin E_i^*$ and contains exactly one element if $(u,v) \in E_i^*$. Thus, $|L([u,j],[v,i])| \le 1$ in this case, so (7) clearly holds.

We may now assume $i > 1$ and $j \ne p(i)$. Let $k = p(i)$, so we must have $k \ne j$. Suppose that we know, *for all* $x \in V$, that $|L([u,j],[x,k])| = f_x$ (i.e. we know all these n values). We wish to estimate the value of $|L([u,j],[v,i])|$ given this knowledge. This is done as follows. Let $L([u,j],[x,k],[v,i])$ be the subset of $L([u,j],[v,i])$ consisting of the members having an edge of $D_k^-(x)$. Note that $|L([u,j],[x,k],[v,i])| \le |L([x,k],[v,i])| \le 1$ according to the previous case, since $k = p(i)$. More precisely, if $(x,v) \notin D_i^-(v)$ then $|L([u,j],[x,k],[v,i])| = 0$. If, however, $(x,v) \in D_i^-(v)$ then

$$E[|L([u,j],[x,k],[v,i])| \quad | \quad |L([u,j],[x,k])| = f_x] = f_x/d_i^+(x),$$

since the matching $B_i(x)$ is selected at random and $f_x/d_i^+(x)$ is the probability that (x,v) is matched to one of the f_x members of $D_k^-(x)$ which are edges of members of $L([u,j],[x,k])$. Thus, if we put

$$R_x = \{|L([u,j],[x,k],[v,i])| \quad | \quad |L([u,j],[x,k])| = f_x\}$$

then for $(x,v) \in D_i^-(v)$ we have that R_x is an indicator random variable with $E[R_x] = \text{Prob}[R_x = 1] = f_x/d_i^+(x)$, while for $(x,v) \notin D_i^-(v)$ we have $R_x = 0$. Note that if $(x,v) \in D_i^-(v)$ and $x \ne y$ then R_x is independent from R_y, since the value of R_x depends only on the matching $B_i(x)$, which is independent from the matching $B_i(y)$. Let

$$R = \{|L([u,j],[v,i])| \quad | \quad \forall x \in V, |L([u,j],[x,k])| = f_x\}.$$

According to the definition of R, we have

$$R = \sum_{x \in V} R_x = \sum_{(x,v) \in D_i^-(v)} R_x.$$

Thus, R is the sum of independent indicator random variables. By linearity of expectation,

$$E[R] = \sum_{(x,v) \in D_i^-(v)} E[R_x] = \sum_{(x,v) \in D_i^-(v)} f_x/d_i^+(x).$$

On the other hand, we know that $\sum_{x \in V} f_x = d_j^-(u)$, since this sum equals to the number of copies of L^* having an edge of $D_j^-(u)$, and this number is exactly $d_j^-(u)$. We also know from (6) that $d_i^+(x) \ge 4h^3\sqrt{n \log n}$. Therefore,

$$E[R] \le \frac{d_j^-(u)}{4h^3\sqrt{n \log n}}. \tag{8}$$

Note that if $f_x = 0$ for some $x \in V$, then $R_x = 0$, and the term R_x can be eliminated from the sum which yields R. Since $\sum_{x \in V} f_x = d_j^-(u)$ this means that R is the sum of at most $d_j^-(u)$ independent indicator random variables. We can now apply the Chernoff bounds for R, and obtain, for every $\alpha > 0$:

$$\text{Prob}[R - E[R] > \alpha] < \exp(-\frac{2\alpha^2}{d_j^-(u)}).$$

In particular, for $\alpha = \sqrt{d_j^-(u)\log(2hn)}$,

$$\text{Prob}[R - E[R] > \sqrt{d_j^-(u)\log(2hn)}] < \exp(-\frac{2d_j^-(u)\log(2hn)}{d_j^-(u)}) = \frac{1}{4h^2n^2},$$

and it now follows from (8) that with probability at least $1 - 1/(4h^2n^2)$,

$$R \le \frac{d_j^-(u)}{4h^3\sqrt{n\log n}} + \sqrt{d_j^-(u)\log(2hn)} < \frac{\sqrt{n}}{2} + \sqrt{2n\log n} \le 2\sqrt{n\log n}. \tag{9}$$

Note that the estimation for R in (9) does not depend on the f_x's. Thus, with probability at least $1 - 1/(4h^2n^2)$,

$$|L([u, j], [v, i])| \le 2\sqrt{n\log n}.$$

Consequently, with probability at least $1 - h^2n^2/(4h^2n^2) = 3/4$, (7) holds for all $u, v \in V$ and for $0 \le j < i \le h - 1$. $\square$

Proof of Theorem 1.2: According to Lemmas 3.1 and 3.2 we know that with probability at least 0.65, we can obtain a decomposition L^* with the properties guaranteed by Lemmas 3.1 and 3.2. We therefore fix such a decomposition, and denote it by L'. We let each member $T \in L'$ choose an integer $c(T)$, where $1 \le c(T) \le h - 1$. Each value has equal probability $1/(h - 1)$. All the m choices are independent. Let $C(v, i)$ be the set of members of T which selected i as their value and they contain an edge of $D_i^+(v)$. Put $|C(v, i)| = c(v, i)$. Clearly, $0 \le c(v, i) \le d_i^+(v)$, and $E[c(v, i)] = d_i^+(v)/(h - 1)$. Since the choices are independent, we know that

$$\text{Prob}[c(v,i) < \frac{d_i^+(v)}{h}] < \exp(-\frac{2d_i^+(v)^2}{h^2(h-1)^2d_i^+(v)} \le \exp(-\frac{2d_i^+(v)}{h^4}) \le \exp(-\frac{8h^3\sqrt{n\log n}}{h^4}) < \frac{1}{2nh}.$$

Thus, with positive probability (in fact, with probability at least 0.5), we have that for all $v \in V$ and for all $i = 1, \ldots, h - 1$,

$$c(v, i) \ge \frac{d_i^+(v)}{h}. \tag{10}$$

We therefore fix the choices $c(T)$ for all $T \in L'$ such that (10) holds.

We are now ready to mend L' into a decomposition L consisting only of trees. Recall that each member of L' is homomorphic to $H(q)$. We shall perform a process which, in each step, reduces the overall number of bad edges in L' by at least one. Thus, at the end, there will be no bad edges, and all the members are, therefore, trees. Our process uses two sets L_1 and L_2 where, initially,

$L_1 = L'$ and $L_2 = \emptyset$. We shall maintain the invariant that, in each step in the process, $L_1 \cup L_2$ is a decomposition of G into subgraphs homomorphic to $H(q)$. Note that this holds initially. We shall also maintain the property that $L_1 \subset L'$. Our process halts when no member of $L_1 \cup L_2$ contains a bad edge, and by putting $L = L_1 \cup L_2$ we obtain a decomposition of G into copies of H, as required. As long as there is a $T^\alpha \in L_1 \cup L_2$ which contains a bad edge, we show how to select a member $T^\beta \in L_1$, and how to create two subgraphs T^γ and T^δ which are also homomorphic to $H(q)$ with $E(T^\alpha) \cup E(T^\beta) = E(T^\gamma) \cup E(T^\delta)$, such that the number of bad edges in $E(T^\gamma) \cup E(T^\delta)$ is *less* than the number of bad edges in $E(T^\alpha) \cup E(T^\beta)$. Thus, by deleting T^α and T^β from $L_1 \cup L_2$ and inserting T^γ and T^δ both into L_2, we see that $L_1 \cup L_2$ is a *better* decomposition since it has less bad edges. It remains to show that this procedure can, indeed, be done.
Let i be the maximum number such that there exists a member $T^\alpha \in L_1 \cup L_2$ where $T^\alpha(i)$ is bad. Let $T^\alpha(i) = (v, w)$. Consider the subgraph T^ϵ of T^α consisting of all the edges $T^\alpha(j)$ where j is a descendent of i. Our aim is to find a member $T^\beta \in L_1$, which satisfies the following requirements:

1. $c(T^\beta) = i$.

2. $T^\beta(i) \in D_i^+(v)$.

3. No vertex of T^α, except v, appears in T^β.

We show that such a T^β can always be found. The set $C(v, i)$ is exactly the set of members of L' which meet the first two requirements (although some of them may not be members of L_1). Let U be the set of vertices of T^α, except v. For $u \in U$, and for all $0 \le j \le h-1$, all the members of $L([u, j], [v, p(i)])$ are not allowed to be candidates for T^β. This is because each member of $L([u, j], [v, p(i)])$ contains an edge of $D^-_{p(i)}(v)$, and thus an edge of $D_i^+(v)$, but it also contains the vertex u, which we want to avoid in T^β, according to the third property required. According to Lemma 3.2,

$$|L([u, j], [v, p(i)])| \le 2\sqrt{n \log n}.$$

Hence,

$$| \cup_{u \in U} \cup_{j=0}^{h-1} L([u, j], [v, p(i)])| < 2h^2\sqrt{nlogn}.$$

Let $C'(v, i)$ be the set of members of $C(v, i)$ which satisfy the third requirement. By (10), (6) and the last inequality,

$$|C'(v, i)| \ge c(v, i) - 2h^2\sqrt{n \log n} \ge \frac{d_i^+(v)}{h} - 2h^2\sqrt{n \log n} \ge$$
$$4h^2\sqrt{n \log n} - 2h^2\sqrt{n \log n} = 2h^2\sqrt{n \log n}.$$

We need to show that at least one of the members of $C'(v,i)$ is also in L_1. Each member $T \in C(v,i)$ that was removed from L' in a prior stage was removed either because it had a bad edge $T(j)$ where $j \ge i$ (this is due to the maximality of i), or because it was chosen as a T^β counterpart of some prior T^α, having a bad edge $T^\alpha(i) = (v,z)$ for some z. There are at most $|N(v,i)|$ members $T \in C(v,i)$ which have a bad edge $T(j)$ where $j \ge i$, and there are at most $|N(v,i,i)|$ members $T \in C(v,i)$ having $T(i)$ as a bad edge. According to Lemma 3.1, $|N(v,i)| + |N(v,i,i)| \le (h+1)\sqrt{d_i^+(v)}$. Since $|C'(v,i)| \ge 2h^2\sqrt{n \log n} > (h+1)\sqrt{d_i^+(v)}$, we have shown that the desired T^β can be selected.
Let T^π be the subgraph of T^β consisting of all the edges $T^\beta(j)$ where j is a *descendent* of i. T^γ is defined by taking T^α and replacing its subgraph T^ϵ with the subgraph T^π. Likewise, T^δ is defined by taking T^β and replacing its subgraph T^π with the subgraph T^ϵ. Note that T^γ and T^δ are both still homomorphic to $H(q)$, and that $E(T^\alpha) \cup E(T^\beta) = E(T^\gamma) \cup E(T^\delta)$, so by deleting T^α and T^β from $L_1 \cup L_2$, and by inserting T^γ and T^δ to L_2 we have that $L_1 \cup L_2$ is still a valid decomposition into subgraphs homomorphic to $H(q)$. The crucial point however, is that every edge of $E(T^\alpha) \cup E(T^\beta)$ that was good, remains good due to requirement 3 from T^β, and that the edge $T^\alpha(i)$ which was bad, now plays the role of $T^\delta(i)$, and it is now a good edge due to requirement 3. Thus, the overall number of bad edges in $L_1 \cup L_2$ is reduced by at least one. $\Box$

4 Concluding remarks and open problems

1. The proof of Theorem 1.2 can also be implemented as a randomized algorithm. That is, given a feasible graph G, one can produce an H-decomposition of G with constant positive probability. To see this, note that Lemma 2.1 is algorithmic, as the partition into the F_i's having the required properties can be done with probability of success at least 0.9, and the F_i's can be checked to have the required properties in polynomial time. The correction of the F_i's into the E_i's which are $5h^3\sqrt{n \log n}$ edge-expanding can be done in polynomial time with probability of success at least 0.5. The orientations in Lemma 2.1 can be performed by using any polynomial time algorithm for bipartite matching. If we fail to obtain one of the perfect matchings, this means that one of the graphs $G_i = (V, E_i)$ is not $5h^3\sqrt{n \log n}$ edge-expanding (but this can only happen with probability at most 0.5, as stated above). After choosing the $n(h-2)$ perfect matchings $B_i(v)$ randomly and independently, one can compute in polynomial time that the obtained L^* satisfies the conditions in Lemmas 3.1 and 3.2. This happens with probability at least 0.65, according to these lemmas. If this is the case, the choices for $c(v,i)$ in Theorem 1.2 can be checked to comply with (10) in polynomial time, and (10) holds with probability at least 0.5. The final step of mending L' into the desired decomposition L is a

purely sequential, non-randomized process, which can be done in polynomial time. We note here that it is known, as a special case of the result of Dor and Tarsi in [4], that if $h \geq 4$, deciding whether a general graph G has an H-decomposition where H is a tree on h vertices, is NP-complete (cf. [5] for a definition of this complexity class).

2. As mentioned in the introduction, Theorem 1.1 states that $f_H(n) \leq n/2 + 10h^4\sqrt{n \log n}$. We conjecture, however, that the dependency on $\sqrt{n \log n}$ can be eliminated.

   Conjecture 4.1 *For every tree H with at least 3 vertices, there exists a constant $c(H)$ such that $f_H(n) \leq n/2 + c(H)$.*

3. The constant $10h^4$ appearing in Theorem 1.2 can be somewhat improved, but this is not crucial since our proofs cannot improve upon the more significant $\sqrt{n \log n}$ factor appearing there, for arbitrary H.

5 Acknowledgment

The author wishes to thank Noga Alon for useful discussions.

References

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

[2] J. A. Bondy and U.S. R. Murty, *Graph Theory with Applications*, Macmillan Press, London, 1976.

[3] T.H. Cormen, C.E. Leiserson and R.L. Rivest, *Introduction to algorithms*, The MIT Press, 1990.

[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] M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, 1979.

[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.