Families of trees decompose the random graph in an arbitrary way

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

Abstract

Let $F = \{H_1, \ldots, H_k\}$ be a family of graphs. A graph G is called *totally F-decomposable* if for *every* linear combination of the form $\alpha_1 e(H_1) + \cdots + \alpha_k e(H_k) = e(G)$ where each α_i is a nonnegative integer, there is a coloring of the edges of G with $\alpha_1 + \cdots + \alpha_k$ colors such that exactly α_i color classes induce each a copy of H_i, for $i = 1, \ldots, k$. We prove that if F is any fixed nontrivial family of trees then $\log n/n$ is a sharp threshold function for the property that the random graph $G(n,p)$ is totally F-decomposable. In particular, if H is a tree with more than one edge, then $\log n/n$ is a sharp threshold function for the property that $G(n,p)$ contains $\lfloor e(G)/e(H) \rfloor$ edge-disjoint copies of H.

1 Introduction

All graphs considered here are finite, undirected and have no loops or multiple edges. For the standard terminology used the reader is referred to [2]. For the standard terminology used in Random Graph Theory the reader is referred to [3]. Let H and G be two graphs. An H-packing of G is a collection of edge-disjoint subgraphs of G, each being isomorphic to H. The *H-packing number of* G, denoted $\nu_H(G)$, is the maximum size of an H-packing of G. Clearly, $\nu_H(G) \le \lfloor e(G)/e(H) \rfloor$. If equality holds, we say that G has an *optimal H-packing*. If, in addition, $e(H)$ divides $e(G)$ and $\nu_H(G) = e(G)/e(H)$ then we say that G has an *H-decomposition*.

Packing and decomposition theory is a central topic in Graph Theory and Design Theory. We shall mention here the following general results. If $G = K_n$, and n is sufficiently large, Wilson [9] gave necessary and sufficient conditions for the existence of an H-decomposition of K_n. For graphs H with at most 5 vertices, necessary and sufficient conditions for an H-decomposition are known for all n (cf. [6]). Caro and Yuster [5] gave a closed formula for $\nu_H(K_n)$, for n sufficiently large. The formula only depends on the degree sequence of H, and on n. It follows that for n sufficiently

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

large, there are necessary and sufficient conditions for the existence of an optimal H-packing. For arbitrary graphs G, and for fixed graphs H other than trees, almost nothing is known. We mention the result of Gustavsson [8], giving necessary and sufficient conditions for an H-decomposition of a graph G, where G is "almost" complete. In case H is a tree, it has been proved [11] that every graph G with minimum degree $d(G) \geq \lfloor n/2 \rfloor$ (n sufficiently large), has an optimal H-packing. This result is sharp for all trees with at least two edges. A related result concerning trees, appearing in [10], shows that every graph G which is a good expander has an optimal H-packing. By "good expander" we mean that G is $C_H\sqrt{n \log n}$ edge-expanding for some constant C_H depending only on H, and n is sufficiently large as a function of H.

Let $F = \{H_1, \ldots, H_k\}$ be a family of graphs. A graph G is called *totally F-decomposable* if for *every* linear combination of the form $\alpha_1 e(H_1) + \cdots + \alpha_k e(H_k) = e(G)$ where each α_i is a nonnegative integer, there is a coloring of the edges of G with $\alpha_1 + \cdots + \alpha_k$ colors such that exactly α_i color classes induce each a copy of H_i, for $i = 1, \ldots, k$. In other words, G is totally F-decomposable if we can decompose it into elements of F in any arbitrary way. Clearly, if H is a graph then, considering the very special case of the family $F = \{H, K_2\}$, we have that G is totally F-decomposable if and only if G has an optimal H-packing.

Let $G = G(n,p)$ denote, as usual, the random graph with n vertices and edge probability p. In the extensive study of the properties of random graphs, many researchers observed that there are sharp *threshold functions* for various natural graph properties. For a graph property A and for a function $p = p(n)$, we say that $G(n,p)$ satisfies A *almost surely* if the probability that $G(n,p(n))$ satisfies A tends to 1 as n tends to infinity. We say that a function $f(n)$ is a *sharp threshold function* for the property A if there are two positive constants c and C so that $G(n, cf(n))$ almost surely does not satisfy A and $G(n,p)$ satisfies A almost surely for all $p \geq Cf(n)$. (There is a sharper definition of sharp thresholds, where c and C are replaced with $c(1 - o(1))$ and $c(1 + o(1))$ respectively.) It is well known that all monotone graph properties have a sharp threshold (see [4] and [7]).

In this paper we consider the property of being totally F-decomposable where F is any fixed family of trees. To avoid the trivial case we assume $F \neq \{K_2\}$. The property of being totally F-decomposable is not monotone. In fact, even the very special case of the property of having an optimal H-packing is not monotone for every tree H with at least two edges. Let m be a positive integer, and let G be any graph having $e(H) \cdot m - 1$ edges, and having an optimal H-packing. Add two isolated vertices to G, and denote the new graph by G'. G' also has an optimal H-packing. Now add to G' an edge between the two isolated vertices. The new graph has $e(H) \cdot m$ edges, but, obviously, does not have an H-decomposition. One may claim that the non-connectivity of G' caused the non-monotonicity. However, it is not difficult to show that if H has three edges or more, there exist connected graphs G with an optimal H-packing, and such that it is possible to add an edge to G and obtain a graph which does not have an optimal H-packing. For example, let $H = K_{1,k}$ where $k \geq 3$ and let $G = K_{1,sk-1}$ where $s \geq 1$. G contains $s - 1$ edge-disjoint copies of

H and therefore has an optimal H-packing. Add to G an edge connecting two non-root vertices. The new graph has sk edges but does not contain s edge-disjoint copies of H.

It is easy to show that for a sufficiently large constant C'_H, if $p = C'_H \sqrt{\frac{\log n}{n}}$ then $G(n,p)$ almost surely is $C_H\sqrt{n \log n}$ edge-expanding, and thus, by the result in [10] mentioned above, $G(n,p)$ almost surely has an optimal H-packing. On the other hand, it is well-known that if $p = c \log n/n$ where c is a sufficiently small constant, then $G(n,p)$ has isolated vertices and many small components. Thus, trivially, we almost surely do not have $\lfloor e(G)/e(H) \rfloor$ edge-disjoint copies of H in G, for any fixed tree H with at least two edges. Consequently, *if there exists* a sharp threshold function $p(n)$ for the property of containing an optimal H-packing then it must be within these bounds. In this paper we prove that, indeed, such a sharp threshold function exists. In fact, we show something much stronger:

Theorem 1.1 *Let F be a family of trees. Then, $p(n) = \frac{\log n}{n}$ is a sharp threshold function for the property of being totally F-decomposable.*

By considering $F = \{H, K_2\}$ we have the following immediate corollaries:

Corollary 1.2 *Let H be a fixed tree with at least two edges. Then, $p(n) = \frac{\log n}{n}$ is a sharp threshold function for the property of having an optimal H-packing.*

Corollary 1.3 *Let H be a fixed tree with at least two edges. Then, there are absolute positive constants c and C such that if $p = c\frac{\log n}{n}$ then $G(n,p)$ almost surely does not have an H-decomposition, and if $1 - 1/n \ge p \ge C\frac{\log n}{n}$ then $G(n,p)$ has an H-decomposition with probability approaching $1/e(H)$, as $n \to \infty$.*

The next section contains the proof of Theorem 1.1. The final section contains some concluding remarks.

2 Proof of the main result

Let $F = \{H_1, \ldots, H_k\}$ be a family of trees, and let $h_i = e(H_i)$ denote the number of edges of H_i. Put $c = h_1 + \ldots + h_k$. Notice that $c \ge 2$ as we assume $F \ne \{K_2\}$. Let $C = (33c)^{20}$. We prove the following.

Lemma 2.1 *Let $p \ge C\frac{\log n}{n}$. Then, $G(n,p)$ is almost surely totally F-decomposable.*

The proof of Theorem 1.1 follows immediately from Lemma 2.1 and the trivial fact that for a sufficiently small constant c, $G(n, c\log n/n)$ is almost surely not totally F-decomposable. We note here that the constant $(33c)^{20}$ can easily be improved. We make no attempt to optimize it.

In the rest of this paper we assume n is sufficiently large, whenever necessary. The first part of our proof does not concern random graphs. We show that if G is an n-vertex graph, that has several "semi-random" properties (to be stated in the following lemma), then G is totally F-decomposable. Since $G(n,p)$ will almost surely have these semi-random properties, the result will follow. Let $f(n)$ be any function of n satisfying $n \geq f(n) \geq C \log n$.

Lemma 2.2 *Let G be a graph with n vertices such that*

1. $\Delta(G) \leq 1.5f(n)$.
2. *Every subset X of vertices with $|X| \leq n/2$ is incident with at least $0.46|X|f(n)$ edges whose other endpoint is not in X.*
3. *For every subset X of vertices with $|X| \geq n/2$ there are at least $0.05nf(n)$ edges with both endpoints in X.*

Then, G is totally F-decomposable.

The proof of Lemma 2.2 is based, in part, on the following lemma, which establishes similar "semi-random" conditions for being H-decomposable.

Lemma 2.3 *Let H be a tree with $h \geq 2$ edges. Let $n \geq f(n) \geq (10h)^{10} \log n$. Suppose G is an n-vertex graph with mh edges where m is an integer. Furthermore, suppose that*

1. $\Delta(G) \leq 1.5f(n)$.
2. *Every subset X of vertices with $|X| \leq n/2$ is incident with at least $0.42|X|f(n)$ edges whose other endpoint is not in X.*

Then, G has an H-decomposition.

The proof of Lemma 2.3 is very similar to the proof of the main result appearing in [10]. We say that a graph $G = (V, E)$ is 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 $V \setminus X$. The following result is proved in [10].

Lemma 2.4 *Let H be any tree with $h \geq 1$ edges. Let G be a graph on n vertices and mh edges where m is an integer. If G is $10h^4\sqrt{n \log n}$ edge-expanding then G has an H-decomposition.*

Notice the difference between Lemma 2.4 and Lemma 2.3. Lemma 2.4 requires a large edge expansion, namely $\Theta(\sqrt{n \log n})$, but there are no constraints placed on the maximum degree. Theorem 2.3 proves a seemingly stronger result, where the expansion needed is only $\Theta(\log n)$ but there is also a maximum-degree constraint. It is not too difficult to modify the (rather complicated) proof of Lemma 2.4 to use a smaller edge expansion under the assumption that the maximum degree is bounded as in Lemma 2.3. This modified proof yielding Lemma 2.3 appears in Appendix A.

Proof of Lemma 2.2: Fix an n-vertex graph $G = (V, E)$ satisfying the conditions of Lemma 2.2. Let $\alpha_1, \ldots, \alpha_k$ be nonnegative integers satisfying $\alpha_1 h_1 + \ldots + \alpha_k h_k = |E|$. We must show how to decompose G into α_i copies of H_i for $i = 1, \ldots, k$. Define

$$t_i = \left\lfloor \frac{100c\alpha_i}{\alpha_1 + \ldots + \alpha_k} \right\rfloor.$$

Notice that $0 \le t_i \le 100c$ for $i = 1, \ldots, k$. Given any set of trees, we can *concatenate* them into one tree by choosing one vertex from each tree, and identifying all the chosen vertices. The concatenated tree is, by definition, decomposable to its originators. Let H denote the tree obtained by concatenating t_i copies of H_i for each $i = 1, \ldots, k$. Note that H has exactly $h = t_1 h_1 + \ldots + t_k h_k$ edges and that $h \le 100c^2$. Now define:

$$q = \left\lfloor \frac{\alpha_1 + \ldots + \alpha_k}{100c} \right\rfloor.$$

Since for any nonnegative reals α, β we have $0 \le \alpha\beta - \lfloor\alpha\rfloor\lfloor\beta\rfloor = \alpha(\beta - \lfloor\beta\rfloor) + (\alpha - \lfloor\alpha\rfloor)\lfloor\beta\rfloor \le \alpha + \beta$ we have

$$0 \le \alpha_i - qt_i \le \frac{\alpha_1 + \ldots + \alpha_k}{100c} + \frac{100c\alpha_i}{\alpha_1 + \ldots + \alpha_k}.$$

Define $b_i = \alpha_i - qt_i$. Thus, $b_i \ge 0$ and

$$b_1 h_1 + \ldots + b_k h_k \le \frac{\alpha_1 + \ldots + \alpha_k}{100} + \frac{100c|E|}{\alpha_1 + \ldots + \alpha_k} \le 0.01|E| + 100c^2 \qquad (1)$$

where in the last inequality we used $|E|/c \le \alpha_1 + \ldots + \alpha_k \le |E|$.

Our next goal is to find in G a spanning subgraph G' with the property that G' has a decomposition in which there are exactly b_i copies of H_i for each $i = 1, \ldots, k$, and $\Delta(G') \le 0.04f(n)$. We use the following procedure. Assume that we have already found a subgraph $\hat{G}$ of G with $\Delta(\hat{G}) \le 0.04f(n)$ and which contains a decomposition into b_i copies of each H_i, $i = 1, \ldots, k-1$ and $b_k - 1$ copies of H_k (completing the last element is, clearly, the most difficult part in the construction, as we may assume h_k is the size of the largest tree in F). We wish to add a copy of H_k to $\hat{G}$ such that the edges of H_k are taken from $E \setminus e(\hat{G})$, and such that the resulting graph G' has $\Delta(G') \le 0.04f(n)$. Indeed, by (1),

$$e(\hat{G}) < h_1 b_1 + \ldots + h_k b_k \le 0.01|E| + 100c^2.$$

It follows that $\hat{G}$ has at least $\lceil n/2 \rceil$ vertices with degrees not exceeding $(0.04|E| + 400c^2)/n$. Let X be such a set of $\lceil n/2 \rceil$ vertices. Consider the graph induced by the vertices of X and the edges of $E \setminus e(\hat{G})$. We denote this graph by X as well. By the third condition in Lemma 2.2, we have that the number of edges of X is at least $0.05nf(n) - 0.01|E| - 100c^2 > 0.04nf(n)$. Hence X has a subgraph with minimum degree $\Omega(\log n) \gg h_k$ and thus, we can find in X a copy of H_k. Joining the edges of a copy of H_k in X to $\hat{G}$ we obtain the graph G' which, by construction, is a subgraph of G and, furthermore,

$$\Delta(G') \leq \max\left\{\Delta(\hat{G})\ ,\ \frac{0.04|E|+400c^2}{n}+\Delta(H_k)\right\} \leq$$

$$\max\left\{0.04f(n)\ ,\ 0.03f(n)+\frac{400c^2}{n}+h\right\} = 0.04f(n).$$

Having constructed the graph G' we now come to the final stage of the proof. Denote by G^* the spanning subgraph of G obtained by deleting the edges of G'. We claim that G^* has an H-decomposition, and the number of elements in this decomposition is q. We prove this using Lemma 2.3. First, we must show that $e(G^*) = q \cdot e(H) = qh$. This is true since:

$$e(G^*) = |E| - e(G') = \sum_{i=1}^{k}\alpha_i h_i - \sum_{i=1}^{k} b_i h_i = q\sum_{i=1}^{k} h_i t_i = qh.$$

Next, we show that G^* and H satisfy the other conditions of Lemma 2.3 with $m = q$. First notice that $f(n) \geq C\log n > (10h)^{10}\log n$. Also, $\Delta(G^*) \leq \Delta(G) \leq 1.5f(n)$. Finally, by the second condition in Lemma 2.2, and by the fact that $\Delta(G') \leq 0.04f(n)$ we have that every subset of vertices X with $|X| \leq n/2$ has at least

$$0.46|X|f(n) - 0.04|X|f(n) = 0.42|X|f(n)$$

edges connecting X and $V\backslash X$ in G^*. It now follows from lemma 2.3 that G^* has an H-decomposition into q copies.

Since every copy of H is decomposable into t_i copies of H_i for each $i = 1,\ldots,k$, we have that G^* has a decomposition into qt_i copies of H_i for each $i = 1,\ldots,k$. It is now easy to see that G has a decomposition into α_i copies of H_i for $i = 1,\ldots,k$. First note that by our construction, G' and G^* are edge disjoint and their edges union is E. Finally notice that the decomposition of G' has b_i copies of H_i and the decomposition of G^* has qt_i copies of H_i. Together, this gives $b_i + qt_i = \alpha_i$ copies of H_i. □

We now prove that $G(n,p)$ almost surely satisfies the properties stated in Lemma 2.2.

Lemma 2.5 *Let $p = f(n)/n$. Almost surely,*

1. $\Delta(G) \leq 1.5f(n)$.

2. *Every subset X of vertices with $|X| \leq n/2$ is incident with at least $0.46|X|f(n)$ edges whose other endpoint is not in X.*

3. *For every subset X of vertices with $|X| \geq n/2$ there are at least $0.05nf(n)$ edges with both endpoints in X.*

Proof: The expected degree of a vertex v in G, is $f(n)(1-1/n)$. By standard large deviation estimates (cf. [1] Appendix A),

$$\Pr\left[d_G(v) > 1.5f(n)\right] \le \Pr\left[d_G(v) - E[d_G(v)] > 0.5f(n)\right]$$
$$< exp\left(-\frac{0.25f(n)^2}{2f(n)} + \frac{0.125f(n)^3}{1.99f(n)^2}\right) < \frac{1}{n^2}.$$

Thus, with probability at least $1-1/n$ the first part of the lemma holds.
Let V denote the set of all vertices, and consider a nonempty $X \subset V$ with $|X| \le n/2$. Let $out(X)$ denote the number of edges connecting a vertex of X and a vertex of $V \setminus X$. The expectation of $out(X)$ is $p|X|(n-|X|) \ge \frac{1}{2}|X|f(n)$. Using large deviation once again we get:

$$\Pr\left[out(X) - p|X|(n-|X|) < -0.08p|X|(n-|X|)\right] < \exp\left(-\frac{0.08^2}{2}p|X|(n-|X|)\right) < \frac{1}{n^2\binom{n}{|X|}}.$$

Since there are $\binom{n}{|X|}$ subsets of size $|X|$, and since there are $n/2$ sizes to consider, we have that with probability at least $1-1/n$, for every $X \subset V$ with $|X| \le n/2$,

$$out(X) \ge p|X|(n-|X|) - 0.08p|X|(n-|X|) = 0.92p|X|(n-|X|) \ge 0.46|X|f(n),$$

which means that the second part of the lemma holds.
Finally, consider $X \subset V$ with $|X| \ge n/2$. Let $in(X)$ denote the number of edges with both endpoints in X. The expectation of $in(X)$ is $p\frac{|X|(|X|-1)}{2}$. Hence,

$$\Pr\left[in(X) - p\frac{|X|(|X|-1)}{2} < -\frac{1}{2}p\frac{|X|(|X|-1)}{2}\right] < \exp\left(-\frac{(\frac{1}{2})^2}{2}p\frac{|X|(|X|-1)}{2}\right) < \frac{1}{n^2\binom{n}{|X|}}.$$

Since there are $\binom{n}{|X|}$ subsets of size $|X|$, and since there are $n/2$ sizes to consider, we have that with probability at least $1-1/n$, for every $X \subset V$ with $|X| \ge n/2$,

$$in(X) \ge p\frac{|X|(|X|-1)}{2} - \frac{1}{2}p\frac{|X|(|X|-1)}{2} = \frac{1}{2}p\frac{|X|(|X|-1)}{2} > \frac{1}{2}p\frac{n^2}{9} > 0.05nf(n)$$

which means that the third part of the lemma holds. □

Now, lemma 2.1 clearly follows from Lemma 2.2 and Lemma 2.5. □

3 Concluding remarks

- The proof of Lemma 2.1 is algorithmic. Namely, given as input a graph taken from the probability distribution $G(n,p)$, and given $\alpha_1,\ldots,\alpha_k$ such that $\alpha_1 h_1 + \ldots + \alpha_k h_k = e(G)$, the algorithm *almost surely* finds a decomposition of G into α_i copies of H_i for $i = 1,\ldots,k$. This follows from the fact, proved in Lemma 2.5, that the input graph almost surely satisfies the conditions of Lemma 2.2, and from the fact that all the details in Lemma 2.2, except for the final part which uses Lemma 2.3, can be implemented by a *deterministic* polynomial time algorithm. Indeed, we only need to count degrees, and to find trees in graphs whose minimum degree is higher than the number of vertices of the tree. These computational tasks are easy to perform in polynomial time. Finally, Lemma 2.3 has a randomized polynomial time algorithm (exactly as shown for Lemma 2.4 in [10]).

- It may be interesting to find other families of graphs for which sharp threshold functions can be determined for the property of being totally decomposable. More specifically, say Q is the four vertex graph consisting of a triangle and an additional edge. Let $F = \{Q, K_2\}$. Can one determine a sharp threshold function for being F decomposable? Notice that for trivial divisibility reasons, some families do not have an associated nontrivial threshold function. For example, suppose $C_3 \in F$. Every graph $G = (V, E)$ which is totally F-decomposable must either have $|E| \neq 0 \bmod 3$ or else have all its degrees even. For every nontrivial p, $G(n, p)$ does not satisfy these two requirements with probability very close to 1/3.

- A proof similar to that of Lemma 2.2, combined with the main result of [10] yields the following theorem, whose proof is omitted.

  Theorem 3.1 *Let F be a family of trees. Then, for n sufficiently large, every graph with minimum degree $0.5n(1 + o_n(1))$ is totally F-decomposable.*

  Notice that for the very special case of $F = \{H, K_2\}$, the result in [11] states that Theorem 3.1 is true even without the error term $o_n(1)$. It would be interesting to determine if this stronger version of Theorem 3.1 holds for every fixed family of trees.

Acknowledgment

The author wishes to thank the referee for many helpful suggestions.

References

[1] N. Alon and J. H. Spencer, **The Probabilistic Method**, *Second Edition*, Wiley, New York, 2000.

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

[3] B. Bollobás, *Random Graphs*, Academic Press, 1985.

[4] B. Bollobás and A. Thomason, *Threshold functions*, Combinatorica 7 (1986), 35–38.

[5] Y. Caro and R. Yuster, *Packing Graphs: The packing problem solved*, Elect. J. Combin. 4 (1997), #R1.

[6] C.J. Colbourn and J.H. Dinitz, **CRC Handbook of Combinatorial Design**, CRC press 1996.

[7] E. Friedgut and G. Kalai, *Every monotone graph property has a sharp threshold*, Proc. Amer. Math. Soc. 124 (1996), 2993–3002 .

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

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

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

[11] R. Yuster, *Packing and decomposition of graphs with trees*, J. Combin. Theory, Ser. B 78 (2000), 123–140.

A Proof of Lemma 2.3

For the rest of this appendix we assume $G = (V, E)$ is an n-vertex graph with mh edges, where m is an integer and which satisfies the conditions of Lemma 2.3. Namely, $f(n) \geq (10h)^{10} \log n$, $\Delta(G) \leq 1.5f(n)$, and $out(X) \geq 0.42|X|f(n)$ for every $X \subset V$ with $|X| \leq n/2$, where $out(X)$ is the number of edges between X and $V \setminus X$. In particular, notice that $\delta(G) \geq 0.42f(n)$ and therefore $m \geq \frac{(10h)^{10}}{5h} n \log n$. It will be convenient to denote $C = (10h)^{10}$.

Lemma A.1 *E can be partitioned into h subsets $E_1, \ldots, E_h$, each having size m, such that each of the spanning subgraphs $G_i = (V, E_i)$ of G, and each vertex v have the following properties:*

1. $$\left| d_i(v) - \frac{d(v)}{h} \right| \leq 0.05 \frac{d(v)}{h^2},$$
   where $d_i(v)$ denotes the degree of v in G_i.

2. *$out_i(X) > \frac{1}{5h}|X|f(n)$, for every $X \subset V$ with $|X| \leq n/2$. Here $out_i(X)$ denotes the number of edges between X and $V \setminus X$ in G_i.*

Proof: Each edge of E chooses a random integer between 0 and h, where 0 is chosen with probability $\beta = n^{-1/2}$ and the other numbers are chosen with probability $\alpha = (1 - \beta)/h$. All the choices are independent. For $i = 0, \ldots, h$ let F_i denote the set of edges which selected i. Let $d'_i(v)$ denote the number of edges of F_i incident with v. The expectation of $|F_i|$ is $\alpha|E| = m(1 - \beta)$, for $i \neq 0$. Using a large deviation inequality of Chernoff we get that for $i \neq 0$:

$$\Pr[|F_i| > m] = \Pr[|F_i| - m(1 - \beta) > m\beta] < \exp\left(-\frac{2m^2\beta^2}{mh}\right) = \tag{2}$$

$$\exp\left(-\frac{2m}{nh}\right) \leq \exp\left(-\frac{2\frac{C}{5h} n \log n}{nh}\right) < \frac{1}{n}.$$

For all $i = 1, \ldots, h$ and for all $v \in V$ we have:

$$\Pr\left[\left|d'_i(v) - \alpha d(v)\right| > 0.02\frac{d(v)}{h^2}\right] < 2\exp\left(-\frac{0.0008\frac{d(v)^2}{h^4}}{d(v)}\right) < 2\exp\left(-\frac{0.0003C \log n}{h^4}\right) < \frac{1}{n^2}. \tag{3}$$

Similarly, for $i = 0$ we have:

$$\Pr\left[|d'_0(v) - \beta d(v)| > 0.02\frac{d(v)}{h^2}\right] < \frac{1}{n^2}. \tag{4}$$

From equation (2), (3) and (4) we get that with probability at least $1 - h/n - hn/n^2 - n/n^2 > 0.9$ all of the following events happen simultaneously:

1. $|F_i| \leq m$ for $i = 1, \ldots, h$.

2. $|d'_i(v) - \alpha d(v)| \le 0.02\frac{d(v)}{h^2}$ for all $i = 1, \ldots, h$ and for all $v \in V$.

3. $|d'_0(v) - \beta d(v)| \le 0.02\frac{d(v)}{h^2}$ for all $v \in V$.

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

$$d_i(v) \ge d'_i(v) \ge \alpha d(v) - 0.02\frac{d(v)}{h^2} = \frac{d(v)}{h} - \frac{d(v)}{\sqrt{n}h} - 0.02\frac{d(v)}{h^2} \ge \frac{d(v)}{h} - 0.03\frac{d(v)}{h^2}. \qquad (5)$$

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

$$d_i(v) \le d'_i(v) + d'_0(v) \le \alpha d(v) + \beta d(v) + 0.04\frac{d(v)}{h^2} =$$

$$\frac{d(v)}{h} - \frac{d(v)}{\sqrt{n}h} + 0.04\frac{d(v)}{h^2} + \frac{d(v)}{\sqrt{n}} \le \frac{d(v)}{h} + 0.05\frac{d(v)}{h^2}. \qquad (6)$$

It now follows from inequalities (5) and (6) that $|d_i(v) - \frac{d(v)}{h}| \le 0.05\frac{d(v)}{h^2}$.
It remains to show that with probability greater than $1 - 0.9 = 0.1$, the requirements regarding $out_i(X)$ are met for each $i = 1, \ldots, h$ and each $X \subset V$ with $|X| \le n/2$. Since $E_i \supset F_i$ it suffices to show that for each such X, the number of edges between X and $V \setminus X$ in the subgraph induced by F_i, denoted $out'_i(X)$, is at least $\frac{1}{5h}|X|f(n)$. Consider a subset X. The expectation of $out'_i(X)$ is $\alpha \cdot out(X)$. Applying large deviation we get:

$$\Pr[|out'_i(X) - \alpha \cdot out(X)| > \alpha \cdot out(X)/2] < 2\exp(-\frac{2 \cdot out(X)^2\alpha^2/4}{out(X)}) = 2\exp(-out(X)\alpha^2/2) \le$$

$$2\exp(-out(X)/(3h^2)) \le 2\exp(-0.42|X|f(n)/(3h^2)) \ll \frac{1}{nh\binom{n}{|X|}}.$$

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 $i = 1, \ldots, h$ and for all sets $X \subset V$ with $|X| \le n/2$,

$$|out'_i(X) - \alpha \cdot out(X)| \le \alpha \cdot out(X)/2.$$

In particular this means that

$$out_i(X) \ge out'_i(X) \ge \alpha \cdot out(X)/2 \ge \frac{1 - 1/\sqrt{n}}{h}0.21|X|f(n) \ge \frac{1}{5h}|X|f(n).$$

□

We call a partition of E into the subsets E_i having the properties guaranteed by Lemma A.1 a *feasible partition.* Given a feasible partition, 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 $i = 1, \ldots, h$. 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 sequential search of the tree, like Breadth-First Search or Depth First Search. Let $e_1, \ldots, e_h$ be the oriented edges of $H(q)$, in the order they are discovered by the search. Note that for $i = 2, \ldots, h$, 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 $i = 2, \ldots, h$, 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 $i = 2, \ldots, h$, and $|d_i^+(v) - d_i^-(v)| \le i \cdot 0.1d(v)/h^2$, for all $i = 1, \ldots, h$. Note that the second requirement implies also that $|d_i^+(v) - d_i(v)/2| \le 0.05d(v)/h$ and, similarly, $|d_i^-(v) - d_i(v)/2| \le 0.05d(v)/h$.

Lemma A.2 *Every feasible partition has a feasible orientation. Furthermore, in every feasible orientation*

$$d_i^+(v) \ge 0.17\frac{f(n)}{h} \tag{7}$$

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

Proof: We show how to construct our orientation in h 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 $|d_1^+(v) - d_1^-(v)| \le 1 \le 0.1d(v)/h^2$. Assume now that we have oriented all the subsets E_j for $1 \le j < i$, 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 $|d_i^+(v) - d_i^-(v)| \le i \cdot 0.1d(v)/h^2$. Our second goal is to show that such an orientation exists. The following inequality achieves the first goal:

$$|d_i^+(v) - d_i^-(v)| = |2c_v - d_i(v)| = |2d_j(v) - 2d_j^+(v) - d_i(v)| \le |2d_j^+(v) - d_j(v)| + |d_j(v) - d_i(v)| =$$

$$|d_j^+(v) - d_j^-(v)| + |d_j(v) - d_i(v)| \le j \cdot 0.1\frac{d(v)}{h^2} + |d_j(v) - \frac{d(v)}{h}| + |d_i(v) - \frac{d(v)}{h}| \le$$

$$j \cdot 0.1\frac{d(v)}{h^2} + 0.1\frac{d(v)}{h^2} \le i \cdot 0.1\frac{d(v)}{h^2}.$$

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, $S = \{v^{(k)} \mid v \in V,\ 1 \le k \le c_v\}$. 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 $S' \subset S$, $|N(S')| \ge |S'|$ where $N(S') \subset E_i$ are the neighbors of S' in B. Fix $\emptyset \ne S' \subset S$. Let $V' = \{v \in V \mid v^{(k)} \in S'\}$. Put $V' = \{v_1, \ldots, v_t\}$. Clearly, $|S'| \le \sum_{l=1}^{t} c_{v_l}$. 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 $t \le n/2$. By Lemma A.1 we have $out_i(V') = t_1 \ge \frac{1}{5h} t f(n)$. Therefore,

$$|N(S')| = t_1 + t_2 = \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{t_1}{2} \ge \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{1}{10h} t f(n) >$$

$$\sum_{l=1}^{t} \left(\frac{d_i(v_l)}{2} + \frac{1.5 f(n)}{15h} \right) \ge \sum_{l=1}^{t} \left(\frac{d_i(v_l)}{2} + \frac{d(v_l)}{15h} \right) \ge \sum_{l=1}^{t} c_{v_l} \ge |S'|.$$

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 $|V''| \le n/2$ we have $t_1 \ge \frac{1}{5h}(n-t) f(n)$. Now,

$$|N(S')| = t_1 + t_2 = \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{t_1}{2} \ge \sum_{l=1}^{t} \frac{d_i(v_l)}{2} + \frac{1}{10h}(n-t) f(n) >$$

$$m - \sum_{l=t+1}^{n} \left(\frac{d_i(v_l)}{2} - \frac{1.5 f(n)}{15h} \right) \ge m - \sum_{l=t+1}^{n} c_{v_l} = \sum_{l=1}^{t} c_{v_l} \ge |S'|.$$

Finally, we need to show that (7) holds. We use the fact that $|d_i^+(v) - d_i(v)/2| \le 0.05 d(v)/h$ and Lemma A.1 which states that $|d_i(v) - d(v)/h| \le 0.05 d(v)/h^2$ and the fact that $h \ge 2$ to obtain that

$$|d_i^+(v) - \frac{d(v)}{2h}| \le 0.0625 \frac{d(v)}{h}.$$

Thus,

$$d_i^+(v) \ge \frac{d(v)}{2h} - 0.0625 \frac{d(v)}{h} = 0.4375 \frac{d(v)}{h} \ge 0.175 \frac{f(n)}{h}.$$

□

A feasible orientation defines a decomposition of the edges of G into a set L^* of m edge-disjoint connected graphs, each graph having h 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 $D_i^+(v) \subset E_i^*$ 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 $i = 2, \ldots, h$ 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 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 $h+1$ vertices) then it *is* isomorphic to $H(q)$.

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

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

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. Before proceeding with the proof we require a few definitions.

For a member $T \in L^*$, and for $i = 1, \ldots, h$, 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 edges are good. For $1 \le i \le j \le h$, let

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

Clearly, $|N(v,i,j)| \le d_i^+(v)$. Our next goal is to show that if all the $n(h-1)$ perfect matchings $B_i(v)$ are selected randomly and independently, then with high probability, $|N(v,i,j)|$ is significantly smaller than $d_i^+(v)$.

Lemma A.3 *If all the perfect matchings $B_i(v)$ are selected randomly and independently, then with probability at least* 0.9*, for all $i = 1, \ldots, h$, for all $j = i, \ldots, h$ and for all $v \in V$, $|N(v, i, j)| \leq f(n)/C$.*

Proof: 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 $n(h-2)$ 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 $U = \{(x_1, u_1), (x_2, u_2), \ldots, (x_k, u_k)\}$ of k edges, where for $t = 1, \ldots, k$, $(x_t, u_t) \in D^-_{p(j)}(u_t)$, 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 $t = 1, \ldots, k$, (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

$$\Pr[(x_t, u_t) \ \textit{is matched in} \ B_j(u_t) \ \textit{to a bad edge}] \leq \frac{h}{d^+_j(u_t)}.$$

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 $1 \leq s < t$, is at most $h/(d^+_j(u_t) - (t-1))$. Thus,

$$\Pr[U \ \textit{is bad}] < \Pi_{t=1}^{k} \frac{h}{d^+_j(u_t) - t + 1}.$$

Assuming $k \leq d^+_j(u_t)/2$, and using (7) we have

$$\Pr[U \ \textit{is bad}] < \left(\frac{12h^2}{f(n)}\right)^k.$$

Consequently,

$$\Pr[|N(v, i, j)| \geq k \mid N^*] < \binom{d^+_i(v)}{k} \left(\frac{12h^2}{f(n)}\right)^k.$$

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

$$\Pr[|N(v,i,j)| \geq k] < \binom{d_i^+(v)}{k} \left(\frac{12h^2}{f(n)}\right)^k.$$

Now put $k = \lceil f(n)/C \rceil$ and note that indeed, $k \leq d_j^+(u_t)/2$. We may therefore estimate the last inequality as follows:

$$\Pr[|N(v,i,j)| \geq f(n)/C] < 2^{d_i^+(v)} \left(\frac{12h^2}{f(n)}\right)^{f(n)/C} < 2^{f(n)} \left(\frac{12h^2}{f(n)}\right)^{f(n)/C} =$$

$$\left(\frac{2^C \cdot 12h^2}{f(n)}\right)^{f(n)/C} < \frac{1}{10nh^2}.$$

Thus, with probability at least $1 - nh^2/(10nh^2) \geq 0.9$, for all $1 \leq i \leq j \leq h$, and for all $v \in V$, $|N(v,i,j)| \leq f(n)/C$. □

For two vertices $u, v \in V$ (not necessarily distinct) and for two indices $0 \leq j, i \leq h$ 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 $D_0^-(x)$ is undefined, so we define $D_0^-(x) = D_1^+(x)$ and $d_0^-(x) = d_1^+(x)$ in this case only. Notice also that $L([u,i],[v,i]) = \emptyset$ when $u \neq v$.

Lemma A.4 *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 $0 \leq j < i \leq h$,*

$$|L([u,j],[v,i])| \leq \frac{f(n)}{10h^5}. \tag{8}$$

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, $|L([u,j],[v,i])| \leq 1$ in this case, so (8) clearly holds.

We may now assume $i > 1$ and $j \neq p(i)$. Let $k = p(i)$, so we must have $k \neq j$. 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 $|L([u,j],[x,k],[v,i])| \leq |L([x,k],[v,i])| \leq 1$ 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 $E[R_x] = \Pr[R_x = 1] = f_x/d_i^+(x)$, while for $(x,v) \notin D_i^-(v)$ we have $R_x = 0$. Note that if $(x,v) \in D_i^-(v)$ and $x \neq y$ 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 (7) that $d_i^+(x) \geq 0.17\frac{f(n)}{h}$. Therefore,

$$E[R] \leq \frac{d_j^-(u)h}{0.17f(n)}. \tag{9}$$

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 $\alpha > 0$:

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

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

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

and it now follows from (9) that with probability at least $1 - 1/(4h^2n^2)$,

$$R \leq \frac{d_j^-(u)h}{0.17f(n)} + \sqrt{d_j^-(u)\log(2hn)}.$$

However,

$$d_j^-(u) = d_j(u) - d_j^+(u) \leq d_j(u) - 0.17\frac{f(n)}{h} \leq \frac{d(u)}{h} + 0.05\frac{d(u)}{h^2} - 0.17\frac{f(n)}{h} < 1.5\frac{f(n)}{h}$$

and therefore, with probability at least $1 - 1/(4h^2n^2)$,

$$R \le 9 + \sqrt{1.5\frac{f(n)}{h}\log(2hn)} \ll \frac{f(n)}{10h^5}. \tag{10}$$

Note that the estimation for R in (10) does not depend on the f_x's. Thus, with probability at least $1 - 1/(4h^2n^2)$,

$$|L([u,j],[v,i])| \le \frac{f(n)}{10h^5}.$$

Consequently, with probability at least $1 - h^2n^2/(4h^2n^2) = 3/4$, (8) holds for all $u, v \in V$ and for $0 \le j < i \le h$. □

Completing the proof: According to Lemmas A.3 and A.4 we know that with probability at least 0.65, we can obtain a decomposition L^* with the properties guaranteed by Lemmas A.3 and A.4. We therefore fix such a decomposition, and denote it by L'. We let each member $T \in L'$ choose an integer $c(T)$, where $1 \le c(T) \le h$. Each value has equal probability $1/h$. 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 $E[c(v,i)] = d_i^+(v)/h$. Since the choices are independent, we know that

$$\Pr\left[c(v,i) < \frac{d_i^+(v)}{h+1}\right] = \Pr\left[c(v,i) - E[c(v,i)] < -\frac{d_i^+(v)}{h(h+1)}\right] <$$

$$\exp\left(-\frac{2d_i^+(v)^2}{(h+1)^2h^2d_i^+(v)}\right) < \exp\left(-\frac{2d_i^+(v)}{(h+1)^4}\right) \le \exp\left(-\frac{0.17\frac{C}{h}\log n}{(h+1)^4}\right) < \frac{1}{2nh}.$$

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

$$c(v,i) \ge \frac{d_i^+(v)}{h+1}. \tag{11}$$

We therefore fix the choices $c(T)$ for all $T \in L'$ such that (11) 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 fewer 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 $0 \le j \le h$, 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 A.4,

$$|L([u, j], [v, p(i)])| \le \frac{f(n)}{10h^5}.$$

Hence,

$$| \cup_{u \in U} \cup_{j=0}^{h-1} L([u, j], [v, p(i)])| < h^2 \frac{f(n)}{10h^5} = \frac{f(n)}{10h^3}.$$

Let $C'(v, i)$ be the set of members of $C(v, i)$ which satisfy the third requirement. By (11), (7) and the last inequality,

$$|C'(v, i)| \ge c(v, i) - \frac{f(n)}{10h^3} \ge \frac{d_i^+(v)}{h+1} - \frac{f(n)}{10h^3} \ge$$
$$0.17 \frac{f(n)}{h(h+1)} - \frac{f(n)}{10h^3} > (h+1)\frac{f(n)}{C}.$$

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 $\sum_{j=i}^h |N(v, i, j)|$ 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 A.3, $|N(v, i, i)| + \sum_{j=i}^h |N(v, i, j)| \le (h+1)f(n)/C$. Since $|C'(v, i)| > (h+1)f(n)/C$, 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. $\square$