Optimal Factorizations of Families of Trees

Raphael Yuster
Department of Mathematics
University of Haifa-ORANIM
Tivon 36006, Israel.
e-mail: raphy@mofetsrv.macam98.ac.il

Abstract

Let $\{T_1, \ldots, T_k\}$ be a set of trees which is K_h-packable. It is shown that every n-vertex graph $G = (V, E)$ with $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$ has k subgraphs $S_1, \ldots S_k$ with the following properties:

1. S_i is a set of $\lfloor n/h \rfloor$ vertex-disjoint copies of T_i.
2. The subgraphs $S_1, \ldots, S_k$ are edge-disjoint.
3. $S_1 \cup \ldots \cup S_k$ has maximum degree at most $h - 1$.

There are many interesting special cases of this result. To name just two:

- If H is a tree with h vertices and $G = (V, E)$ is a graph with n vertices, h divides n, and $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$, then G has an H-factor.
- If h divides n, and $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$, then G has a set S of n star subgraphs, where for each $i = 1, \ldots, h$, there are exactly n/h stars in S having i vertices, any two members of S having the same size are vertex-disjoint, and the union of all the members of S is an $h - 1$ regular spanning subgraph of G.

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 [4]. An *h-packing* of a set of graphs $\mathcal{F} = \{H_1, \ldots, H_k\}$ is a coloring of the edges of K_h with k colors, such that the subgraph induced by color i contains H_i as a subgraph. It should be noted that $\mathcal{F}$ is allowed to contain isomorphic members. Clearly, if there exists an h-packing of $\mathcal{F}$, then h must be at least as large as the largest (w.r.t. vertices) member of $\mathcal{F}$. There are many results concerning h-packings, among the famous ones are [7] and [5]. An h-packing is called an *h-decomposition* if there are $\binom{h}{2}$ edges in all the members of $\mathcal{F}$ together, or in other words, each subgraph induced by color i is *isomorphic* to H_i. There are many results concerning h-decompositions, mainly in the area of designs (cf. [3] for a good source on Design Theory). A simple example of a family of graphs having an h-decomposition is the family $\{S_2, \ldots, S_h\}$ where S_i is a star with i vertices. Another example is the family of paths having i vertices, for $i = 2, \ldots, h$. It was conjectured by Gyárfás and Lehel [6] that every family of trees $\{T_2, \ldots, T_h\}$, where T_i is an arbitrary tree with i vertices, has an h-decomposition. This conjecture is still open. A special case of an h-packing or h-decomposition occurs when all the members of $\mathcal{F}$ have h vertices. For example, two identical paths on four vertices have a 4-decomposition, since we can color the edges of K_4 with two colors such that each color induces a path with three edges and four vertices.

Let H be a connected graph with h vertices. An H-factor of a graph G is a spanning subgraph of G where each connected component is isomorphic to H. Note that the number of vertices of G, denoted by n, is assumed to be a multiple of h. Assume now that $\mathcal{F} = \{H_1, \ldots H_k\}$ is a set of k graphs, each having h vertices, which has an h-decomposition. We can ask whether G has an H_i-factor for each $i = 1, \ldots, k$. Can we also insist that all the k factors be edge-disjoint? If so, consider the union of the factors. It contains $n(h-1)/2$ edges. Thus, the average degree is $h-1$. Can we insist that this also be the maximum degree? If all this occurs we say that G has an *optimal factorization of $\mathcal{F}$*. The purpose of this paper is to give sufficient conditions which guarantee that a graph has an optimal factorization of $\mathcal{F}$ in case all the members of $\mathcal{F}$ are trees. In fact, we prove a much more general result which is the following:

Theorem 1.1 *Let $\mathcal{F} = \{T_1, \ldots, T_k\}$ be a set of trees which has an h-packing. If $G = (V, E)$ is a graph with n vertices and $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$, then G has k subgraphs $S_1, \ldots, S_k$ with the following properties:*

1. *S_i is a set of $\lfloor n/h \rfloor$ vertex-disjoint copies of T_i.*
2. *The subgraphs $S_1, \ldots, S_k$ are edge-disjoint.*
3. *$S_1 \cup \ldots \cup S_k$ has maximum degree at most $h - 1$.*

There are many interesting special cases which can be solved by applying Theorem 1.1. We mention just a few:

1. Suppose all the members of $\mathcal{F}$ have exactly h vertices, and suppose $\mathcal{F}$ has an h-decomposition. If $G = (V, E)$ satisfies the conditions of Theorem 1.1 and h divides n, then S_i is, in fact, a T_i-factor. It now follows from Theorem 1.1 that $\mathcal{F}$ has an optimal factorization. To summarize:

Theorem 1.2 *Let $\mathcal{F} = \{T_1, \ldots, T_{h/2}\}$ be a set of $h/2$ trees on h vertices each, having an h-decomposition. Then, if $G = (V, E)$ has $\delta(G) \geq \frac{|V|}{2} + 3h\sqrt{|V| \log |V|}$ and h divides $|V|$ then G has an optimal factorization of $\mathcal{F}$.*

2. Another special case occurs when $\mathcal{F}$ is a set consisting of only one tree, H. Trivially, if H has h vertices, then it has an h-packing. If $G = (V, E)$ satisfies the conditions of Theorem 1.1 and h divides n, then the Theorem states that G has an H-factor. In other words, we have the following:

   Theorem 1.3 *Let H be a tree with h vertices. If $G = (V, E)$ has $\delta(G) \geq \frac{|V|}{2} + 3h\sqrt{|V| \log |V|}$ and h divides $|V|$ then G has an H-factor.*

   Unlike Theorem 1.2, Theorem 1.3 is not new. Minimum degree requirements guaranteeing the existence of H-factors when H is an arbitrary fixed graph have been studied by several researchers. We mention just the result in [2], which shows, among other things, that if H is any bipartite graph, then a minimum degree of $\delta(G) \geq n/2 + \epsilon(H)n$ suffices.

3. Another special case which follows from Theorem 1.1 is the following:

   Theorem 1.4 *Let $\mathcal{F} = \{T_1, \ldots, T_k\}$ be a set of trees which has an h-decomposition. If $G = (V, E)$ is a graph with n vertices and $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$, and h divides n, then G has k subgraphs $S_1, \ldots, S_k$ with the following properties:*

   (a) S_i is a set of n/h vertex-disjoint copies of T_i.

   (b) The subgraphs $S_1, \ldots, S_k$ are edge-disjoint.

   (c) $S_1 \cup \ldots \cup S_k$ is an $h - 1$ regular spanning subgraph of G.

   Clearly, Theorem 1.4 applies in the special case when $\mathcal{F}$ is the set $\{S_2, \ldots, S_h\}$ where S_i is the star with i vertices, which gives the result mentioned in the abstract.

Theorem 1.1 is best possible up to the error term $3h\sqrt{n \log n}$, since there are examples where a minimum degree of $n/2$ does not suffice even for the existence of an H-factor of some trees on h vertices. This also shows that Theorems 1.2, 1.3 and 1.4 are also best possible, up to the sublinear error term.

The rest of this paper contains the proof of Theorem 1.1 in Section 2, and some concluding remarks and open problems in Section 3. Throughout this paper, all logarithms are natural.

2 Proof of the main result

In this section we prove Theorem 1.1. Let $\mathcal{F} = \{T_1, \ldots, T_k\}$ be a set of trees having an h-packing. We may therefore assume that the members of $\mathcal{F}$ are edge-disjoint trees on the *same* vertex set $\{1, \ldots, h\}$. Let $G = (V, E)$ be an n-vertex graph with $\delta(G) \geq n/2 + 3h\sqrt{n \log n}$. If h does not divide n we may delete at most $h - 1$ vertices from G in order to obtain a graph whose number of vertices is divisible by h. Thus, we may assume that

$$n > \delta(G) \geq n/2 + 3h\sqrt{n \log n} - h \geq n/2 + 2h\sqrt{n \log n}, \qquad (1)$$

and h divides n.

Our first task is to show that V can be partitioned into h equal parts, such that each two parts have sufficiently many edges between them. This is achieved by the following lemma.

Lemma 2.1 *There exists a partition of V into h parts, $V_1, \ldots, V_h$, of size n/h each, such that every vertex has at least $n/(2h)$ neighbors in each of the parts.*

Proof: We let each vertex $v \in V$ choose a random integer between 0 and h, where 0 is chosen with probability $\beta = h\sqrt{\log n}/\sqrt{n}$ (note that $\beta < 1$ by (1)) and the other numbers are chosen with probability $\alpha = (1 - \beta)/h$. All the choices are independent. For $i = 0, \ldots, h$, let $W_i \subset V$ be the set of vertices which selected i. For $v \in V$, Let $w_i(v)$ be the number of neighbors of v in W_i. Clearly, for $i > 0$, the expected size of W_i is $E[|W_i|] = \alpha n = \frac{n}{h}(1 - \beta)$, and the expected value of $w_i(v)$ is $E[w_i(v)] = \alpha d(v)$, where $d(v)$ is the degree of v in G. We may use the large deviation result of Chernoff (cf., e.g. [1] Appendix A) to derive that for $i > 0$

$$\mathrm{Prob}[|W_i| > \frac{n}{h}] = \mathrm{Prob}[|W_i| - \frac{n}{h}(1 - \beta) > \beta\frac{n}{h}] < \exp(-\frac{2n^2\beta^2/h^2}{n}) = \frac{1}{n^2}. \qquad (2)$$

Similarly, we have that for each $i = 1, \ldots, h$ and for each $v \in V$

$$\mathrm{Prob}[|w_i(v) - \alpha d(v)| > \sqrt{d(v) \log n}] < 2\exp(-2d(v) \log n/d(v)) = \frac{2}{n^2}. \qquad (3)$$

Since, by (1),

$$h \cdot \frac{1}{n^2} + nh \cdot \frac{2}{n^2} < 0.5$$

we have by inequalities (2) and (3) that with probability greater than 0.5, all of the following events hold:

1. $|W_i| \leq n/h$ for $i = 1, \ldots, h$.

2. $|w_i(v) - \alpha d(v)| \leq \sqrt{d(v) \log n}$ for each $i = 1, \ldots, h$ and for each $v \in V$.

Consider, therefore, a partition of E into $W_0, \ldots, W_h$ in which all of these events hold. Since $|W_i| \leq n/h$, for $i = 1, \ldots, h$, we may partition W_0 into h subsets $X_1, \ldots, X_h$, where $|X_i| = n/h - |W_i|$. Put $V_i = W_i \cup X_i$ for $i = 1, \ldots, h$. Note that $|V_i| = n/h$ and $V_i \cap V_j = \emptyset$ for $1 \leq i < j \leq h$. Let $d_i(v)$ be the number of neighbors of v in V_i. Clearly,

$$d_i(v) \geq w_i(v) \geq \alpha d(v) - \sqrt{d(v) \log n} = \frac{d(v)}{h} - \beta \frac{d(v)}{h} - \sqrt{d(v) \log n} \geq$$

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

□

Consider a partition of V into $V_1, \ldots, V_h$, as guaranteed by Lemma 2.1. Using this partition, we can now show that there exists a spanning subgraph of G with the following structural properties:

Lemma 2.2 *There exists an $h-1$-regular spanning subgraph of G, with the property that for each $v \in V$, if $v \in V_i$, then v has exactly one neighbor in each V_j for $j \neq i$.*

Proof: It suffices to show that each pair of distinct vertex classes V_i and V_j, have a perfect matching with edges of G, since the union of all these $\binom{h}{2}$ matchings yields the required subgraph. To see that V_i and V_j have a perfect matching we may use Hall's Theorem (cf. [4]). We need to show that each $X \subset V_i$ has $|N(X)| \geq |X|$, where $N(X)$ is the set of vertices of V_j adjacent to at least one vertex of X. Indeed, if $0 < |X| \leq n/(2h)$ then, by Lemma 2.1, every vertex of X has at least $n/2h$ neighbors in V_j, and so $|N(X)| \geq n/(2h) \geq |X|$. If $n/h \geq |X| > n/(2h)$ then, by the fact that each vertex of V_j has at least $n/(2h)$ neighbors in V_i it follows that $N(X) = V_j$ and so $|N(X)| = n/h \geq |X|$. Thus, Hall's condition ensuring a perfect matching is satisfied. □

Let R denote the spanning subgraph of G whose existence is guaranteed in Lemma 2.2. For $1 \leq i < j \leq h$, let $R(i,j)$ be the n/h edges of R which connect V_i to V_j. By Lemma 2.2, $R(i,j)$ is a perfect matching between V_i and V_j.

We must now construct, for each $i = 1, \ldots, k$ a subgraph S_i of G consisting of n/h vertex-disjoint copies of the tree T_i. In fact, these subgraphs will only use edges of R, and each edge of R will be used in at most one of the S_i's. This guarantees that $S_1, \ldots, S_k$ are k edge-disjoint subgraphs, and that the union $S_1 \cup \ldots S_k$ has maximum degree at most $h-1$, since it is a subgraph of R, and R is $h-1$-regular.

We construct S_i as follows: The edges of S_i are simply the union of all the sets $R(s,t)$ where (s,t) is an edge of T_i. This definition is proper since, by the remark in the beginning of the section, the vertex-set of T_i is $\{1, \ldots, h\}$, so $s, t \in \{1, \ldots, h\}$. Now S_i is simply the subgraph induced by this set of edges. Note that S_i is, in fact, a subgraph of R. Now, since T_i is a tree (this is crucial!), we claim that S_i is a set of n/h vertex-disjoint copies of T_i. This follows from the fact that each path in S_i is isomorphic to a path of T_i, and since there are no cycles in T_i, each connected component of S_i contains *exactly* one edge from each $R(s,t)$ for $(s,t) \in T_i$. Now, the obvious isomorphism between the vertex classes $V_1, \ldots, V_h$ and the vertices of T_i shows that each connected component is isomorphic to T_i. Finally, the fact that for $i \neq j$, S_i and S_j are edge-disjoint, follows from the fact that T_i and T_j are edge-disjoint trees on the same vertex-class $\{1, \ldots, h\}$. $\square$

3 concluding remarks and open problems

1. As mentioned in the introduction, the minimum degree requirement in Theorem 1.1 is mandatory, up to the sub-linear error term $3h\sqrt{n \log n}$. In fact, it is shown in [2] that there are bipartite graphs H, where for arbitrary large n, a minimum degree of $n/2$ for G does not suffice in order to guarantee even the existence of an H-factor, let alone the much stronger requirements in Theorem 1.1. For example, consider the star S_h on $h > 2$ vertices, where h is even. If $n = kh$ where k is any odd positive integer, and G is the complete bipartite graph with $n/2$ vertices in each vertex class then G cannot have an S_h-factor, although $\delta(G) = n/2$ and h divides n. Similar examples involving other types of trees also exist.

2. Theorem 1.2 applies whenever $\mathcal{F}$ is a set of $h/2$ trees with h vertices each, which has an h-decomposition. There are many such families of trees. The smallest nontrivial example is when $h = 4$ and $\mathcal{F}$ consists of two paths on four vertices. This is the only example for $h = 4$. The case $h = 6$ already contains examples where the three members of $\mathcal{F}$ are not all isomorphic to each other. In fact, since the number of non-isomorphic trees with the same size h grows exponentially with h, so does the number of different sets $\mathcal{F}$ to which Theorem 1.2 applies.

3. Theorem 1.1 has an obvious randomized algorithm. Lemma 2.1 is the only random part, and can clearly be performed in $O(n^2)$ time. The probability of achieving success in the obtained partition of V constructed in Lemma 2.1 is proved there to be greater than 0.5. By letting each vertex know its class, we can verify in $O(n^2)$ time if, in fact, the random partition satisfies the requirements of the Lemma. Lemma 2.2 can be done in $O(n^{2.5})$ using any one of the well-known algorithms for bipartite matching. Having done this, the construction of the sets S_i is performed by a one time pass on the edges of R, namely in $O(n)$ time. The overall running time is, therefore, $O(n^{2.5})$. In fact, since the number of events we need to control in Lemma 2.1 is polynomial ($h + nh$ events, to be precise), we can use the standard derandomization technique of conditional probabilities (cf. [1]) to obtain a polynomial deterministic algorithm.

4. As mentioned in the introduction, results guaranteeing the existence of H-factors for trees (and other graphs) provided the minimum degree is $n/2 + o(n)$ (in the case of trees) are known (cf. e.g., [2]). However, all of these results use the Szemerédi Regularity Lemma ([8]) and therefore have horrible constants, which require that n be very large with respect to $|H|$, where "very large" is a tower function of $|H|$. On the other hand, Theorem 1.3 only requires that n be quadratic in h, (as can be seen from inequality (1)). This is advantageous if one needs to obtain H-factors of graphs G with a reasonable number of vertices.

References

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

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

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

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

[5] B. Bollobás and S.E. Eldridge, *Packings of graphs and applications to computational complexity*, J. Combin. Theory, Ser B 24 (1978), 1-17.

[6] A. Gyárfás and J. Lehel, *Packing trees of different order into K_n*, Coll. Math. Soc. J. Bolyai 18 Combinatorics (1978), 463-469.

[7] N. Sauer and J. Spencer, *Edge-disjoint placement of graphs*, J. Combin. Theory, Ser B 25 (1979).

[8] E. Szemerédi, *Regular partitions of graphs*, in: Proc. Colloque Inter. CNRS (J. -C. Bermond, J. -C. Fournier, M. Las Vergnas and D. Sotteau eds.) (1978), 399-401.