Asymptotically optimal tree-packings in regular graphs

Alexander Kelmans *

Rutgers University, New Brunswick, New Jersey and
University of Puerto Rico, San Juan, Puerto Rico
kelmans@rutcor.rutgers.edu

Dhruv Mubayi †

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA
Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399
mubayi@microsoft.com

Benny Sudakov ‡

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
and Institute for Advanced Study, Princeton, NJ 08540, USA
bsudakov@math.princeton.edu

Submitted: February 15, 2001; Accepted: November 21, 2001.
AMS Subject Classifications: 05B40, 05C05, 05C35, 05C70, 05D15
Keywords: *Packing trees, matchings in hypergraphs*

Abstract

Let T be a tree with t vertices. Clearly, an n vertex graph contains at most n/t vertex disjoint trees isomorphic to T. In this paper we show that for every $\epsilon > 0$, there exists a $D(\epsilon, t) > 0$ such that, if $d > D(\epsilon, t)$ and G is a simple d-regular graph on n vertices, then G contains at least $(1-\epsilon)n/t$ vertex disjoint trees isomorphic to T.

1 Introduction

We consider simple undirected graphs. Given a graph G and a family $\mathcal{F}$ of graphs, an *$\mathcal{F}$-packing* of G is a subgraph of G each of whose components is isomorphic to a member of $\mathcal{F}$. The *$\mathcal{F}$-packing problem* is to find an $\mathcal{F}$–packing of the maximum number of vertices. There are various results on the $\mathcal{F}$–packing problem (see e.g. [3, 9, 10, 11, 12, 13, 14, 15]).

*Research supported in part by the National Science Foundation under DIMACS grant CCR 91-19999.

†Research supported in part by the National Science Foundation under grant DMS-9970325.

‡Research supported in part by NSF grants DMS-0106589, CCR-9987845 and by the State of New Jersey.

When $\mathcal{F}$ consists of a single graph F, we abuse notation by writing *F–packing*. The very special case of the F–packing problem when $F = K_2$, a single edge, is simply that of finding a maximum matching. This problem is well-studied, and can be solved in polynomial time (see, for example, [15]). However, if F is a connected graph with at least three vertices then the F-packing problem is known to be NP-hard [13]. The F-packing problem remains NP-hard even for 3-regular graphs if F is a path with at least 3 vertices [11].

There are various directions for studying this generally intractable problem. One possible direction is to try to obtain bounds on the size of the maximum F–packing of various families of graphs, as well as the corresponding polynomial approximation algorithms. The following is an example of such a result. It concerns the P_3–packing problem for 3-regular graphs, where P_3 is the 3-vertex path.

Theorem 1.1. [12] *Suppose that G is a 3-regular graph. Then G contains at least $v(G)/4$ vertex disjoint 3-vertex paths that can be found in polynomial time (and so for 3-regular graphs there is a polynomial approximation algorithm that guarantees at least a 3/4–optimal solution for the P_3–packing problem).*

Another direction is to consider some special classes of graphs in hope to find a polynomial time algorithm for the corresponding F–packing problem. Here is an example of such a result.

Theorem 1.2. [9] *Suppose that G is a claw–free graph (i.e. G contains no induced subgraph isomorphic to $K_{1,3}$). Suppose also that G is connected and has at most two end–blocks (in particular, 2–connected). Then the maximum number of disjoint 3–vertex paths in G is equal to $\lfloor v(G)/3 \rfloor$ vertex disjoint 3-vertex paths. Moreover there is a polynomial time algorithm for finding an optimal P_3–packing in G.*

An asymptotic approach provides another direction for studying this NP-hard problem. There is a series of interesting asymptotic packing results on sufficiently dense graphs. They have beed iniciated by the following deep theorem of Hajnal and Szemerédi.

Theorem 1.3. [8] *If G has n vertices and minimum degree at least $(1 - 1/r)n$, then G contains $\lfloor n/r \rfloor$ vertex-disjoint copies of K_r.*

Theorem 1.3 has been generalized by Alon and Yuster for graphs other than K_r.

Theorem 1.4. [2] *For every $\gamma > 0$ and for every positive integer h, there exists an $n_0 = n_0(\gamma, h)$ such that for every graph H with h vertices and for every $n > n_0$, any graph G with hn vertices and with minimum degree $\delta(G) \geq (1 - 1/\chi(H) + \gamma)hn$ contains n vertex-disjoint copies of H.*

In this paper we consider an asymptotic version of the F–packing problem, where F is a tree. Our main result is the following.

Theorem 1.5. *Let T be a tree on t vertices and let $\epsilon > 0$. Suppose that G is a d-regular graph on n vertices and $d \geq \frac{128t^3}{\epsilon^2} \ln(\frac{128t^3}{\epsilon^2})$. Then G contains at least $(1 - \epsilon)n/t$ vertex disjoint copies of T.*

Both Theorem 1.3 and Theorem 1.4 require G to have $\Omega(n^2)$ edges. Theorem 1.5 differs from these results in that our graphs are not required to be dense. Indeed, d above is only a function of ϵ and the size of the tree and does not depend on n. Consequently, Theorem 1.5 cannot possibly be extended to graphs other than trees, since the Turán number of a cycle of length $2t$ is known to be at least $\Omega(n^{(2t+1)/2t})$ [4], and there exist essentially regular graphs with about this many edges that contain no copy of C_{2t}.

In this paper, we present two approaches for obtaining tree-packing results for regular graphs. First, in Section 2 we give a short proof of an asymptotic version of Theorem 1.5. This proof relies on powerful hypergraph packing results of Frankl and Rödl [7] and Pippenger and Spencer [17]. Next, in Section 3 we present a proof of Theorem 1.5, based on a probabilistic approach. It uses another powerful result called the Lovász Local Lemma (see e.g., [1]). In addition, it provides an explicit dependence of the degree on t and ϵ. Section 4 contains some concluding remarks and an open question.

2 T-packings from matchings in hypergraphs

In this section we present the proof of the following asymptotic version of Theorem 1.5.

Theorem 2.1. *Let T be a tree on t vertices. Let G_n be a d_n-regular graph on n vertices. Suppose that $d_n \to \infty$ when $n \to \infty$. Then G_n contains at least $(1 - o(1))n/t$ (and, obviously, at most n/t) disjoint trees isomorphic to T.*

The proof of this theorem is based on a hypergraph packing result of Pippenger and Spencer [17]. The main idea behind this proof came from a result of Rödl [18] that solved an old packing conjecture of Erdős and Hanani [5]. Rödl's idea, now known as his "nibble", was used by Frankl-Rödl [7] to prove that under certain regularity and local density conditions, a hypergraph has a large matching. Pippenger and Spencer used probabilistic methods to extend and generalize the result in [7].

First we introduce some notions about hypergraphs. All hypergraphs we consider are allowed to have multiple edges. Given a hypergraph $\mathcal{H} = (V, E)$, the *degree* $d(v)$ of a vertex $v \in V$ is the number of edges containing v. For vertices v, w, the *codegree* $cod(v, w)$ of v and w is the number of edges containing both v and w. Let

$$\Delta(\mathcal{H}) = \max_{v \in V} d(v), \qquad \delta(\mathcal{H}) = \min_{v \in V} d(v), \qquad C(G) = \max_{u,v \in V, u \neq v} cod(u, v).$$

A *matching* in $\mathcal{H}$ is a set of pairwise disjoint edges of $\mathcal{H}$. Let $\mu(\mathcal{H})$ be the size of the largest matching in $\mathcal{H}$. A matching M is *perfect* if every vertex of $\mathcal{H}$ is in exactly one edge of M. A hypergraph $\mathcal{H}$ is t-uniform if each of its edges consists of exactly t elements.

Theorem 2.2. [17] *For every $t \geq 2$ and $\varepsilon > 0$, there exist $\varepsilon' > 0$ and n_0 such that if $\mathcal{H}$ is a t-uniform hypergraph on $n(\mathcal{H}) \geq n_0$ vertices with $\delta(\mathcal{H}) \geq (1 - \varepsilon')\Delta(\mathcal{H})$, and $C(\mathcal{H}) \leq \varepsilon'\Delta(\mathcal{H})$, then*

$$\mu(\mathcal{H}) \geq (1 - \varepsilon)n/t.$$

We rephrase Theorem 2.2 in more convenient asymptotic notation.

Theorem 2.1′. *Let $\mathcal{H}_1, \mathcal{H}_2, \ldots$ be sequence of t-uniform hypergraphs, with $|V(\mathcal{H}_k)| \to \infty$. If $\delta(\mathcal{H}_k) \sim \Delta(\mathcal{H}_k)$, and $C(\mathcal{H}_k) = o(\Delta(\mathcal{H}_k))$, then $\mu(\mathcal{H}_k) \sim |V(\mathcal{H}_k)|/t$.*

The above result says that under certain regularity and local density conditions on $\mathcal{H}$, one can find an almost perfect matching M in $\mathcal{H}$, i.e., the number of vertices in no edge of M is negligible. In fact, [17] proves something much stronger, namely that one can decompose almost all the edges of $\mathcal{H}$ into almost perfect matchings, but we need only the weaker statement.

Next we show how Theorem 2.2 can be applied to provide asymptotically optimal tree-packings of regular graphs. For convenience, we omit the subscript k and the use of integer parts in what follows. Our goal is to produce a large T-packing in G. By a copy of T we mean a subgraph isomorphic to T.

Given $u, v \in V(G)$ let $c(v)$ and $c(u,v)$ denote the number of copies of T in G containing v and $\{u,v\}$, respectively (note that different copies may have the same vertex set). The following lemma provides necessary estimates for the numbers $c(v)$ and $c(u,v)$.

Lemma 2.3. *Let T be a tree with t vertices. Suppose that G is a d-regular graph on n vertices. Then*

- *(c1) $c(v) = (1+o(1))c_T d^{t-1}$ $(d \to \infty)$ for every $v \in V(G)$, where c_T depends only on T and does not depend on the choice of v, and*
- *(c2) $c(a,b) = O(d^{t-2})$ for every pair $a, b \in V(G)$, $a \neq b$.*

Proof. We first estimate $c(v)$. Let us consider the rooted tree R obtained from T by specifying a vertex r of T as a root. Let $c_r(v)$ denote the number of copies of R in G in which the vertex $v \in V(G)$ is chosen to be the root r.

It is easy to see that $c(v) = \sum\{c_r(v)/g : r \in V(T)\} = (1+o(1))(t/g)d^{t-1}$, where g is the size of the automorphism group of T. Therefore it suffices to show that $c_r(v) = (1+o(1))d^{t-1}$ for all $r \in V(T)$ and $v \in V(G)$.

Let x_1 be a leaf of R distinct from r, $R_1 = R - x_1$, and y_1 be the vertex in R_1 adjacent to x_1. If (x_i, y_i, R_i) is already defined, let x_{i+1} be a leaf of R_i distinct from r, $R_{i+1} = R_i - x_{i+1}$, and y_{i+1} be the vertex in R_{i+1} adjacent to x_{i+1}. Clearly $r = y_{t-1} = R_{t-1}$. Now we estimate $c_r(v)$ as follows. There is only one way to allocate r in G, namely, to allocate r in v. Since v is of degree d in G and G is simple, there are d ways to allocate x_{t-1} in G. Suppose that R_i, $1 \leq i < t-1$, is already allocated in G, and y_i is allocated in a vertex v_i in G. Since v_i is of degree d in G and G is simple, there are at most d and at least $d - t + i$ ways to allocate x_i in G. Therefore

$$(d-t)^{t-1} < c_r(v) < d^{t-1}. \qquad (*)$$

Since $d \to \infty$, we have: $c_r(v) = (1+o(1))d^{t-1}$ for all $r \in V(T)$ and $v \in V(G)$.

Now we will estimate $c(a,b)$, the number of copies of T in G containing both a and b where $a \neq b$. For $x, y \in V(T)$, let $c_{x,y}(a,b)$ denote the number of copies of T containing

a, b, with a playing the role of x and b playing the role of y. Clearly

$$c(a,b) \le \binom{t}{2} \max_{x,y \in V(T)} c_{x,y}(a,b),$$

because a, b play the role of some pair x, y in each copy of T containing them. Hence it suffices to show that $c_{x,y}(a,b) \le d^{t-2}$.

Split T in two nontrivial trees X and Y where X is rooted at x and Y is rooted at y, $V(X) \cap V(Y) = \emptyset$, and $V(X) \cup V(Y) = V(T)$. This can be done by deleting any edge from the unique path between x and y. By $(*)$, there are at most $d^{|V(X)|-1}$ copies of X in G with a playing the role of x, and at most $d^{|V(Y)|-1}$ copies of Y in G with b playing the role of y. Thus $c_{x,y}(a,b) \le d^{|V(X)|-1} d^{|V(Y)|-1} = d^{t-2}$. $\square$

Proof of Theorem 2.1 Given G, we must find a T-packing of size at least $(1-o(1))n/t$. From G construct the hypergraph $\mathcal{H} = (V, E)$ with $V = V(G)$ and E consisting of vertex sets of copies of T in G (note that $\mathcal{H}$ can have multiple edges). Then claim $(c1)$ of Lemma 2.3 implies $\delta(\mathcal{H}) = \Delta(\mathcal{H}) \sim c_T d^{t-1}$, and claim $(c2)$ of Lemma 2.3 implies $C(\mathcal{H}) = O(d^{t-2}) = o(d^{t-1}) = o(\Delta(\mathcal{H}))$. Hence, by Theorem 2.2, $\mu(\mathcal{H}) \sim |V(H)|/t = n/t$. This clearly yields a T-packing in G of the required size. $\square$

3 T-packings from the Lovász Local Lemma

This section contains a proof of Theorem 1.5 based on a probabilistic approach and the so called Lovász Local Lemma. We use the following symmetric version of the Lovász Local Lemma.

Theorem 3.1. [1] *Let $A_1, \ldots, A_n$ be events in a probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most d, and that $Prob[A_i] \le p$ for all i. If $ep(d+1) \le 1$, then $Prob[\wedge \overline{A_i}] > 0$.*

Here we make no attempt to optimize our absolute constants. First we need the following lemma. Given a partition $V_1, \ldots, V_t$ of the vertex set of a graph G, let $d_i(v)$ denote the number of neighbors of a vertex v of G in V_i.

Lemma 3.2. *Let t be an integer and let G be a d-regular graph satisfying $d \ge 4t^3$. Then there exists a partition of $V(G)$ into t subsets $V_1, \ldots, V_t$ such that*

$$\frac{d}{t} - 4\sqrt{\frac{d}{t}\ln d} \le d_i(v) \le \frac{d}{t} + 4\sqrt{\frac{d}{t}\ln d}$$

for every $v \in V$ and $1 \le i \le t$.

Proof. Partition the set of vertices V into t subsets $V_1, V_2, \ldots, V_t$ by choosing for each vertex randomly and independently an index i in $\{1, \ldots, t\}$ and placing it into V_i. For $v \in V(G)$ and $1 \le i \le t$, let $A_{i,v}$ denote the event that $d_i(v)$ is either greater than

$\frac{d}{t} + 4\sqrt{\frac{d}{t}\ln d}$ or less than $\frac{d}{t} - 4\sqrt{\frac{d}{t}\ln d}$. Observe that if none of the events $A_{i,v}$ holds, then our partition satisfies the assertion of the lemma. Hence it suffices to show that with positive probability no event $A_{i,v}$ occurs. We prove this by applying Theorem 3.1.

Since the number of neighbors of any vertex v in $V_i, i = 1, 2, \ldots, t$, is a binomially distributed random variable with parameters d and $1/t$, it follows by the standard Chernoff's-type estimates for Binomial distributions (cf. , e.g., [16], Theorem 2.3) that for every $v \in V$

$$Pr\left(|d_i(v) - \frac{d}{t}| > a\frac{d}{t}\right) \leq 2e^{-\frac{a^2(d/t)}{2(1+a/3)}}.$$

By substituting a to be $4\sqrt{(t/d)\ln d}$, we obtain that the probability of the event $A_{i,v}$ is at most $2e^{-4\ln d} = 2d^{-4}$. Clearly each event $A_{i,v}$ is independent of all but at most $td(d-1)$ others, as it is independent of all events $A_{j,u}$ corresponding to vertices u whose distance from v is larger than 2. Since $e \cdot 2d^{-4} \cdot (td(d-1)+1) < e \cdot 2d^{-4} \cdot td^2 < 1$, we conclude, by Theorem 3.1, that with positive probability no event $A_{i,v}$ holds. This completes the proof of the lemma. $\square$

Next we prove the following tree-packing result for nearly-regular, t-partite graphs, which is interesting in its own right.

Theorem 3.3. *Let T be a fixed tree with the vertex set $u_1, \ldots, u_t$ and let H be a t-partite graph with parts $V_1, \ldots, V_t$ such that $|V_1| = h$ and for every vertex $v \in V(H)$ and every $1 \leq i \leq t$ the number $d_i(v)$ of neighbors of v in V_i satisfies $(1-\delta)k \leq d_i(v) \leq (1+\delta)k$ for some $k > 0$ and $0 \leq \delta < 1$. Then H contains $(1-2(t-1)\delta)h$ vertex disjoint copies of T with the property that V_i contains the vertex of each copy corresponding to u_i, $1 \leq i \leq t$.*

Proof. We use induction on t. For $t = 1$ the assertion is trivially true. Therefore let $t \geq 2$. Without loss of generality, we can assume that u_t is a leaf adjacent to the vertex u_{t-1}. Let $T' = T - u_t$ and $H' = H - V_t$. Then by the induction hypothesis, we can find at least $(1-2(t-2)\delta)h$ vertex disjoint copies of T' in H' such that in all these copies the vertices, corresponding to u_{t-1}, belong to V_{t-1}. Denote the set of these vertices by S. Consider all the edges between S and V_t. In the resulting bipartite graph B each vertex is of degree at most $(1+\delta)k$. Therefore the edges of B can be covered by $(1+\delta)k$ disjoint matchings. In addition, note that each vertex from S has degree at least $(1-\delta)k$. Since the number of edges in B is at least $(1-\delta)k|S|$, we conclude that B contains a matching of size at least

$$\frac{(1-\delta)k|S|}{(1+\delta)k} = \frac{1-\delta}{1+\delta}|S| \geq (1-2\delta)|S|.$$

By adding the edges of this matching to the appropriate copies of T', we obtain at least $(1-2\delta)|S| = (1-2\delta)(1-2(t-2)\delta)h \geq (1-2(t-1)\delta)h$ vertex disjoint copies of T. This completes the proof of the statement. $\square$

Having finished all necessary preparations, we are now ready to complete the proof of Theorem 1.5.

Proof of Theorem 1.5. Let G be a d-regular graph on n vertices with $d \geq \frac{128t^3}{\epsilon^2}\ln(\frac{128t^3}{\epsilon^2})$ and let T be a tree with t vertices. By Lemma 3.2, we can partition vertices of G into

t parts $V_1, \ldots, V_t$ such that $|V_1| \geq n/t$ (pick V_1 to be the largest part) and for every vertex the number of its neighbors in V_i, $1 \leq i \leq t$, is bounded by $(1 \pm \delta)d/t$, where $\delta = 4\sqrt{(t/d)\ln d} \leq \epsilon/2t$. Thus by Theorem 3.3, G contains at least $(1 - 2(t-1)\delta)|V_1| \geq (1-\epsilon)n/t$ vertex disjoint copies of T. $\square$

4 Concluding remarks

- The regularity requirement in Theorem 1.5 cannot be weakened to a minimum degree requirement. To see this, let G_d be the complete bipartite graph with parts X, Y of sizes d and d^2, respectively. The minimum degree of G_d is $d \to \infty$, but clearly the largest T-packing has size at most $d = o(|V(G_d)|)$. On the other hand, it is easy to see that the proof of Theorem 1.5 remains valid for nearly-regular graphs. More precisely one can show the following.

  Proposition 4.1. *Let T be a tree on t vertices. For all t and $\epsilon > 0$, there exist two positive numbers $\gamma = \gamma(t,\epsilon)$ and $D(t,\epsilon)$ such that the following holds: if $d > D(t,\epsilon)$ and G is a graph on n vertices with $(1-\gamma)d \leq \delta(G) \leq \Delta(G) \leq (1+\gamma)d$, then G contains $(1-\epsilon)n/t$ vertex disjoint copies of T.*

  It is also easy to see that the above results can be extended to d-regular multigraphs provided all multiplicities are bounded.

- The dependency of the degree of the graph on both t and ϵ is needed in the statement of Theorem 1.5. To see this, let G be a regular graph consisting of $\lceil \epsilon n/t \rceil$ disjoint cliques of size k, where $k = \Theta(t/\epsilon)$ is an integer such that $k \equiv t-1 \pmod{t}$. Clearly any packing of G by a tree on t vertices misses at least $t-1$ vertices in each clique. Therefore altogether it will miss at least $(t-1)(\epsilon n/t) = \Omega(\epsilon |V(G)|)$ vertices. This shows that in the statement of Theorem 1.5 the degree of the graph should be at least $\Omega(t/\epsilon)$. Thus there is a big gap between the upper and lower bounds and this leads to the following

  Question. What is the correct dependency of the degree of the graph G on t and ϵ to guarantee $(1-\epsilon)n/t$ vertex disjoint copies of T in G?

Acknowledgments. The first author thanks Michael Krivelevich for very useful remarks. The second author thanks Brendan Nagle for very helpful discussions and for pointing out some relevant references.

References

[1] N. Alon, J. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.

[2] N. Alon, R. Yuster, H-factors in dense graphs, *J. Combin. Theory Ser. B* **66** (1996), no. 2, 269–282.

[3] G. Cornuéjols and D. Hartvigsen, An extension of matching theory, *J. Combin. Theory* **B 40** (1986) 285–296.

[4] P. Erdős, Graph theory and probability, *Canadian J. Math.* **11**, 34–38.

[5] P. Erdős, H. Hanani, On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10** (1963), 10–13.

[6] P. Erdős, H. Sachs, Reguläre graphen gegenbener Taillenweite mit minimaler Knotenzahl, *Wiss. Z. Uni. Halle (Math. Nat.)* **12** (1963), 251–257.

[7] P. Frankl, V. Rödl, Near Perfect Coverings in Graphs and Hypergraphs, *Europ. J. Combin.* **6** (1985), 317–326.

[8] A. Hajnal, E. Szemerédi, Proof of a conjecture of P. Erdős, in: *Combinatorial theory and its applications*, II (Proc. Colloq., Balatonfüred, 1969), pp. 601–623. North-Holland, Amsterdam, 1970.

[9] A. Kaneko, A. Kelmans, and T. Nisimura, On packing 3–vertex paths in a graph, *J. Graph Theory* **36** (2001) 175–197.

[10] A. Kelmans, Optimal packing of induced stars in a graph, *Discrete Mathematics*, **173**, (1997) 97–127.

[11] A. Kelmans, Packing P_k in a cubic graph is NP-hard if $k \geq 3$, in print.

[12] A. Kelmans and D. Mubayi, How many disjoint 2–edge paths must a cubic graph have ?, submitted (see also *DIMACS Research Report 2000–23*, Rutgers University).

[13] D. G. Kirkpatrick, P. Hell, On the complexity of general graph factor problems, *SIAM J. Comput.* **12**, (1983) 601–609.

[14] M. Loebl and S. Poljak, Efficient subgraph packing, *J. Combin. Theory* **B 59** (1993) 106–121.

[15] L. Lovasz, M. Plummer, *Matching Theory*, North-Holland, Amsterdam, 1986.

[16] C. McDiarmid, Concentration, in : *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, Springer, Berlin, 1998.

[17] N. Pippenger, J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs, *J. Combin. Theory Ser. A* **51** (1989), 24–42.

[18] V. Rödl, On a Packing and Covering Problem, *Europ. J. Combin.* **5** (1985), 69–78.