H-Factors in Dense Graphs

Noga Alon *
and
Raphael Yuster
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel

Abstract

The following asymptotic result is proved. For every $\epsilon > 0$, and for every positive integer h, there exists an $n_0 = n_0(\epsilon, h)$ 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 $d \geq (\frac{\chi(H)-1}{\chi(H)} + \epsilon)hn$ contains n vertex disjoint copies of H. This result is asymptotically tight and its proof supplies a polynomial time algorithm for the corresponding algorithmic problem.

1 Introduction

All graphs considered here are finite, undirected and simple. If H is a graph on h vertices and G is a graph on hn vertices, we say that G has an *H-factor* if it contains n vertex disjoint copies of H. Thus, for example, a K_2-factor is simply a perfect matching, whereas a C_4-factor is a spanning subgraph of G every connected component of which is a cycle of length 4.

Let H be a graph on h vertices and let G be a graph on hn vertices. There are several known results that show that if the minimum degree $d = d(G)$ of G is sufficiently large, then G contains an H-factor. Indeed, if H is a path of length $h-1$ then, by Dirac's Theorem on Hamiltonian cycles (cf. [4]), $d \geq hn/2$ suffices for the existence of an H-factor. Corrádi and Hajnal [5] proved that for $H = K_3$, $d = 2n$ suffices and Hajnal and Szemerédi [7] proved that for $H = K_h$, $d = (h-1)n$ guarantees an H-factor. All these results are easily seen to be best possible.

A conjecture of Erdös and Faudree [6] asserts that $d = 2n$ suffices for a C_4-factor. If true, this value is clearly optimal. In this paper we come close to proving the conjecture, and show that a

*Research supported in part by the Fund for Basic Research administered by the Israel Academy of Sciences

minimum degree of $d = (2 + o(1))n$ suffices. In fact, we prove a much more general result, that shows that $d = (1 - 1/\chi(H) + o(1))hn$ suffices for the existence of an H-factor, for any fixed graph H. (As usual, $\chi(H)$ denotes the chromatic number of H). The exact statement of the result is the following.

Theorem 1.1 *For every $\epsilon > 0$ and for every positive integer h, there exists an $n_0 = n_0(\epsilon, h)$ 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 $d(G) \geq (1 - 1/\chi(H) + \epsilon)hn$ has an H-factor.*

The proof of Theorem 1.1 is based on an asymptotic result proved in [3], as well as on some probabilistic arguments, and various combinatorial ideas. In section 2 we introduce the necessary lemmas that are used in the proof. A short outline of the proof is presented in section 3. The proof itself is described in detail in section 4. The tightness of the results, their algorithmic aspects and some concluding remarks and open problems are discussed in the final section.

2 Preliminary lemmas

In order to prove our theorems, we require an asymptotic result obtained in [3] which enables us to cover almost all of the vertices of a (large) graph L with vertex disjoint copies of a (small) graph S.

Lemma 2.1 *([3]) For every $\delta > 0$ and for every integer s_0, there exists an $l_0 = l_0(\delta, s_0)$ such that for every graph S with $s \leq s_0$ vertices and chromatic number $\chi(S)$ and for every $l > l_0$, any graph L with l vertices and minimum degree $d \geq \frac{\chi(S)-1}{\chi(S)} l$ contains at least $(1-\delta)l/s$ vertex disjoint copies of S.* $\square$

The following probabilistic lemma is used in order to partition a dense graph M into three dense induced subgraphs. The (simple) proof uses a large deviation result of Chernoff.

Lemma 2.2 *Let $\gamma > 0$. There exists an $m_0 = m_0(\gamma)$ such that for every $m > m_0$, if $m_1+m_2+m_3 = m$ and M is a graph with m vertices and minimum degree $d = d(M)$, then its vertices can be partitioned into three pairwise-disjoint sets M_1, M_2, M_3 whose sizes are m_1, m_2, m_3 respectively, such that every vertex has at least $m_i d/m - \gamma m$ neighbors in M_i for $i = 1, 2, 3$.*

Proof Let m_0 be chosen such that for all $m \geq m_0$, $(3m + 6)e^{-\gamma^2 m/2} < 0.5$. Let M be a graph on $m \geq m_0$ vertices. Let $M = M'_1 \cup M'_2 \cup M'_3$ be a random partition of M into three subsets obtained by letting each vertex, independently, be a member of M'_i with probability m_i/m. Let X_i be the number of vertices in M_i'. Clearly, $E[X_i] = m_i$. By the large deviation result of Chernoff (see, e.g., [2], Appendix A, Theorem A.4) we have

$$Prob(|X_i - m_i| > \gamma m/2) < 2e^{-\gamma^2 m/2}.$$

For a vertex v, Let X_i^v be the number of neighbors of v in M_i'. Clearly, $E[X_i^v] \geq dm_i/m$. Once again, by the Chernoff estimates:

$$Prob(X_i^v < dm_i/m - \gamma m/2) < e^{-\gamma^2 m/2}.$$

By our choice of m_0, with probability at least 0.5 we have that

$$|X_i - m_i| \leq \gamma m/2$$

and

$$X_i^v \geq dm_i/m - \gamma m/2$$

for all vertices $v \in M$ and for $i = 1, 2, 3$. We can therefore transfer at most $\gamma m/2$ vertices to or from each M_i' such that the adjusted sets, denoted by M_i, have m_i vertices for $i = 1, 2, 3$ and clearly, every vertex has at least $m_i d/m - \gamma m$ neighbors in M_i. □

The next simple lemma gives a minimum degree bound which guarantees that a bipartite graph contains a perfect star-matching. We note that although the lemma can easily be proved by Hall's Theorem, we prefer the following proof since it can be implemented efficiently.

Lemma 2.3 *Let B be a bipartite graph with c vertices in the first vertex class and cr vertices in the second vertex class. Suppose also that every vertex is connected to at least half of the vertices of the opposite class. Then B contains c vertex-disjoint stars on $r+1$ nodes, where the root of each star is a vertex of the first vertex class.*

Proof We reduce this problem to a perfect matching problem. Replace each vertex v of the first vertex class by r copies, and connect each copy to the original neighbors of v. The new graph is bipartite, has cr vertices in each vertex class and minimum degree at least $rc/2$. It clearly suffices to show that this new graph has a perfect matching. Let F be a maximum matching in it. If $|F| = rc$ we are done. Otherwise let u and v be two unmatched vertices, that lie in distinct vertex classes. Let $N(u)$ and $N(v)$ be the neighbor sets of u and v respectively. Since F is maximal, every vertex of $N(u) \cup N(v)$ is matched. Since $N(u) \geq rc/2$ and $N(v) \geq rc/2$ we must have an edge $f = (u', v') \in F$ with $u' \in N(u)$ and $v' \in N(v)$. This, however, contradicts the maximality of F since we can replace f with the two edges (u, u') and (v, v'), and obtain a larger matching. □

The final lemma in this section is a somewhat technical variant of the previous lemma.

Lemma 2.4 *Let F be a bipartite graph on the classes of vertices C and B, where $|C| = c$, $|B| = b \geq rc$, and r is a positive integer. If the degree of each B-vertex is at least a, and the degree of each C-vertex is at least rc then F contains a spanning subgraph F' consisting of a union of stars which satisfies the following.*

1. *The degree of each B-vertex in F' is precisely* 1.
2. *The degree of each C-vertex in F' is at least r and at most $(b + rc)/a$.*
3. *The degrees in F' of all but at most $(r - 1)c/(a + r - 2)$ C-vertices are* $0 \bmod r$.

Proof We begin by greedily assigning to each C-vertex r B-vertices among its neighbors. This can be done since the degree of each C-vertex is at least rc. We are now left with a set B' of unassigned B-vertices, where $|B'| = b - rc$. Pick an assignment of the elements of B' to their neighbors, so as to maximize the number of C-vertices to which $0 \bmod r$ B-vertices have been assigned. If there is more than one such assignment, we pick the one having the maximum possible number of B-vertices assigned to *good* C-vertices, i.e., to C-vertices to which $0 \bmod r$ B-vertices have been assigned. Let d_0 denote the number of good C-vertices, which we denote by $C_1, \ldots, C_{d_0}$ and let the other C-vertices be $C_{d_0+1}, \ldots, C_c$. The maximality of d_0 implies that each non-good C_i has been assigned at least one B-vertex B_i that is not adjacent to any other non-good C-vertex. Therefore, B_i is adjacent to at least $a - 1$ good C-vertices, for $i = d_0 + 1, \ldots, c$. Hence, there exists a good C-vertex which is adjacent to at least $(c - d_0)(a - 1)/d_0$ of the B_i's. By the maximality of the number of B-vertices assigned to good C-vertices we must have $(c - d_0)(a - 1)/d_0 \leq r - 1$ which implies $c - d_0 \leq (r - 1)c/(a + r - 2)$.

It remains to show that it is possible to ensure that less than $(b+rc)/a$ B-vertices will be assigned to any C-vertex. Put $x = (b + rc)/a$. Assume that there exists a C-vertex C_i to which x or more B-vertices have been assigned. We will show how to transfer exactly r B-vertices from C_i to another C-vertex C_j to which less than $x - r$ B-vertices have been assigned. We may then repeat this process until there are no more C-vertices to which x or more B-vertices have been assigned. Let t denote the number of C-vertices to which $x - r$ or more B-vertices have been assigned. Clearly $t(x - r) \leq b$, so $t \leq b/(x - r)$. Each B-vertex assigned to C_i can be assigned to at least $a - t$ C-vertices to which less than $x - r$ B-vertices have been assigned. Hence there is a C-vertex among these, which is adjacent to at least $(a - t)x/(c - t)$ B-vertices of C_i. It remains to show that $(a - t)x/(c - t) \geq r$. Since $c \geq a$, we can replace t with its upper bound $b/(x - r)$ and hence we must show that $(a - b/(x - r))x/(c - b/(x - r)) \geq r$. This, indeed, is always true for $x = (b + rc)/a$. □

3 Outline of the proof

Given graphs H and G as in the theorem, we first omit several vertex disjoint copies of H from G (which exist by some standard results from Extremal Graph Theory), in order to make sure that the number of missing copies of H is divisible by some large integer, chosen appropriately. Let $k = \chi(H)$ denote the chromatic number of H and let K denote the complete k-partite graph with h-vertices in each color class. Let S denote the complete k-partite graph with h/ϵ^4 vertices in each color class. Note that K contains an H-factor and hence it suffices to show that the remainder of G contains a K-factor. The existence of such a K-factor is proven by covering almost all the remainder of G by pairwise vertex disjoint copies of S to which we carefully connect the remianing vertices in a way that enables us to split each copy of S together with the vertices attached to it into vertex disjoint copies of K.

In order to obtain the above mentioned required copies of S we apply Lemma 2.2 to split the remaining vertices of G into three parts M_1, M_2 and M_3 of specified sizes, where each vertex has sufficiently many neighbors in each part. Applying Lemma 2.1 to the large part M_1, we cover most of this part by vertex-disjoint copies of S. Let these copies be denoted by $S_1, \ldots, S_c$. The vertices of M_1 that do not lie in any of the S_i's are transferred to M_2. Next, we assign the vertices of (the adjusted) M_2 (considered as B-vertices in the language of Lemma 2.4) to the copies of S found in M_1 (which play the role of the C-vertices) using Lemma 2.4, where a vertex is adjacent to a copy S_i in the bipartite graph considered in the lemma if it has sufficiently many neighbors in each color class of S_i. By Lemma 2.4 (with $r = (k-1)kh$) almost all the copies are assigned 0 mod $(k-1)kh$ vertices of M_2. Finally we assign the vertices of M_3 to the S_i's in such a way that each vertex of M_3 assigned to S_i has sufficiently many neighbors in all color classes of S_i but at most one. This is done using Lemma 2.3, where the specific choice of the vertices of M_3 which are being assigned to each S_i is chosen carefully in order to guarantee that the total number of vertices assigned to each S_i is divisible by kh and that the vertices from M_3 have enough neighbors in the required color classes of S_i.

To complete the proof we show that each S_i together with the vertices assigned to it from M_2 and M_3 contains a K-factor. To do so we repeatedly remove copies of K from S_i and the vertices assigned to it by picking such an assigned vertex together with h of its neighbors in all color classes of S_i but one, and together with $h-1$ vertices of S_i in this remaining color class. The vertices of M_2 assigned to S_i are helpful in performing this task successfully, as they have enough neighbors in *all* color classes of S_i. Therefore, we can choose from which class to take $h-1$ vertices in order to guarantee that the remaining part of S_i, after removing the copies of K containing all vertices that have been attached to it, will contain the same number of vertices in each color class and will thus contain a K-factor.

The final required H-factor of G consists of the intial copies of H omitted from G together with all the H-factors of the K-factor of the remainder of G.

The detailed proof requires a careful choice of various parameters and some tedius computation, and is described in the next section.

4 The proof of the main result

Let $h > 1$ be an integer and let $\epsilon > 0$ be a small real number satisfying $\epsilon < 1/(800h)$. We may clearly assume that $1/\epsilon$ is an integer. Let $\delta = \epsilon^6/2$, $\gamma = \epsilon^5/16$. Let $m_0 = m_0(\gamma)$ be chosen as in Lemma 2.2, let $l_0 = l_0(\delta, h^2/\epsilon^4)$ be chosen as in Lemma 2.1 and denote

$$g = \frac{1}{\epsilon^4} + 1 + \frac{8}{\epsilon}.$$

We prove the theorem with

$$n_0 = n_0(\epsilon, h) = \max\{\frac{4l_0}{h}, \frac{m_0}{h} + \frac{2h}{\epsilon}g, \frac{512h^3}{\epsilon^6}\}. \tag{1}$$

Let H be a graph with h vertices and $\chi(H) = k \leq h$. Let G be a graph with hn vertices, $n > n_0$, and minimum degree $d = d(G) \geq (\frac{k-1}{k} + \epsilon)hn$. We must show that G contains an H-factor.

We denote by K the complete k-partite graph consisting of h vertices in each color class. Note that K has an H-factor. We denote by S the complete k-partite graph consisting of h/ϵ^4 vertices in each color class. Note that S contains a K-factor, and thereby also an H-factor. Let $s = kh/\epsilon^4$ be the number of vertices of S. Let $p = n \bmod (kg)$.

Claim 1: *G contains p vertex disjoint copies of H.*

Proof Since S contains k/ϵ^4 pairwise-disjoint copies of H, it suffices to show that G contains two disjoint copies of S. Note that $hn > hn_0 > l_0$. We may therefore apply Lemma 2.1 with $L = G$ and obtain that G contains at least $(1-\delta)hn/s > 0.5hn/s \geq 0.5n\epsilon^4/k > 1$ copies of S. This concludes the proof of the claim. (Note that the claim can also be deduced from the known Turán type results since the graph G contains enough edges, much more than the number needed to guarantee two disjoint copies of S).

By the last claim, we may delete from G a set of p vertex-disjoint copies of H, together with their vertices. We denote the remaining graph by M. Clearly, it suffices to prove that M contains a K-factor. M contains

$$m = hn - hp = h(n-p) = chkg \tag{2}$$

vertices, for some positive integer c. We claim that the minimum degree of M satisfies $d(M) \geq (\frac{k-1}{k} + \epsilon/2)m$. Indeed, by our choice of n_0,

$$d(M) \geq (\frac{k-1}{k} + \epsilon)hn - hp > (\frac{k-1}{k} + \epsilon)hn - hk(\frac{1}{\epsilon^4} + 1 + \frac{8}{\epsilon}) \geq (\frac{k-1}{k} + \frac{\epsilon}{2})hn \geq (\frac{k-1}{k} + \frac{\epsilon}{2})m.$$

We also note that $m > hn - hkg > h(m_0/h + kg) - hkg = m_0$. We may therefore apply Lemma 2.2 to the graph M. The values of m_1, m_2 and m_3 that we use are:

$$m_1 = \lceil \frac{ckh}{\epsilon^4}(1 + \epsilon^6) \rceil$$

$$m_2 = \lfloor \frac{8ckh}{\epsilon}(1 - \frac{\epsilon^3}{8}) \rfloor$$

$$m_3 = ckh.$$

We obtain a partition of the vertices of M into three disjoint sets, M_1, M_2 and M_3 as guaranteed by the lemma. Replacing $d(M)$ with the lower estimate of $(\frac{k-1}{k} + \epsilon/2)m$, and γ with $\epsilon^5/16$ we have

$$d_{M_1}(v) \geq \frac{m_1 d(M)}{m} - \gamma m \geq (1 + \epsilon^6)\frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/2) - \frac{\epsilon^5}{16}m > \quad (3)$$

$$\frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/2) - \frac{\epsilon^5}{16}m \geq \frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/4) + \frac{ckh}{4\epsilon^3} - \frac{\epsilon^5}{16}ckh(\frac{1}{\epsilon^4} + 1 + \frac{8}{\epsilon}) \geq$$

$$\frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/4) + \frac{ckh}{4\epsilon^3} - \frac{\epsilon^5}{8}\frac{ckh}{\epsilon^4} \geq \frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/4) + ckh$$

Similar computations show that

$$d_{M_2}(v) > \frac{8ckh}{\epsilon}(\frac{k-1}{k} + \epsilon/4) + ckh, \quad (4)$$

$$d_{M_3}(v) \geq ckh(\frac{k-1}{k} + \epsilon/4). \quad (5)$$

Note that $d_{M_i}(v)$ denotes the number of neighbors of a vertex v of M in M_i, for $i = 1, 2, 3$. Let L be the subgraph of M induced on M_1. Our first goal is to obtain c vertex disjoint copies of S in L. This will be done by applying Lemma 2.1 to L. In order to do so we need to show that $m_1 > l_0$ and that $d(L) \geq \frac{k-1}{k}m_1$. Indeed, by our choice of n_0,

$$m_1 > \frac{ckh}{\epsilon^4} > m/2 > hn/4 > hn_0/4 > l_0.$$

By (3) we have

$$d(L) \geq \frac{ckh}{\epsilon^4}(\frac{k-1}{k} + \epsilon/4) + ckh \geq \frac{ckh}{\epsilon^4}\frac{k-1}{k} + (\epsilon^2 ckh + 1)\frac{k-1}{k} \geq \frac{k-1}{k}m_1$$

Having shown the above, we obtain by Lemma 2.1 that L contains at least $(1-\delta)m_1/s$ disjoint copies of S. We need to show that $(1-\delta)m_1/s \geq c$. By our choice of δ, we know that $(1-\delta)(1+2\delta) > 1$. Hence,

$$(1-\delta)m_1/s \geq (1-\delta)\frac{ckh}{\epsilon^4}(1+\epsilon^6)/s = (1-\delta)c(1+\epsilon^6) = (1-\delta)c(1+2\delta) > c.$$

Having obtained at least c disjoint copies of S in L, we pick c such copies denoted by $S_1, \ldots, S_c$. We denote the vertex classes of S_i by S_i^j for $j = 1, \ldots, k$. Let C_1 be the set of vertices of all these copies. Note that $C_1 \subset M_1$ and $|C_1| = cs = ckh/\epsilon^4$. Put $C_2 = M_2 \cup (M_1 \setminus C_1)$. Clearly, $|C_2| = 8ckh/\epsilon$. For the sake of completeness, we put $C_3 = M_3$. We claim that for every vertex v:

$$d_{C_i}(v) \geq (\frac{k-1}{k} + \epsilon/4)|C_i| \tag{6}$$

for $i = 1, 2, 3$. To see this, note first that for $i = 3$, this follows from (5). For $i = 1$, this follows from (3) and from the fact that $|M_1 \setminus C_1| = \lceil ckh\epsilon^2 \rceil < ckh$. For $i = 2$ this follows from (4) and from the fact that $|C_2 \setminus M_2| = \lceil ckh\epsilon^2 \rceil < ckh$.

For the rest of the proof, let $q = 100h^2k/\epsilon^2$. We say that a vertex $v \in C_2$ is *strongly assignable* to S_i if v has at least q neighbors in each vertex class of S_i. We say that v is *weakly assignable* to S_i if v has at least q neighbors in each vertex class of S_i, but at most one.

Claim 2: *Every vertex $v \in C_2$ is strongly assignable to at least $ck\epsilon/8$ of the S_i's in C_1. Furthermore, for each S_i, $i = 1, \ldots, c$ there is a set of at least ck^2h vertices of C_2 that are strongly assignable to it.*

Proof Let $v \in C_2$, and let x denote the number of S_i's in C_1 to which v is strongly assignable. v has at most $xs = xkh/\epsilon^4$ neighbors in the S_i's to which it is strongly assignable. It has at most $(c-x)((k-1)h/\epsilon^4 + q)$ neighbors in the other S_i's. By (6) we have

$$x\frac{kh}{\epsilon^4} + (c-x)[\frac{(k-1)h}{\epsilon^4} + \frac{100kh^2}{\epsilon^2}] \geq (\frac{k-1}{k} + \frac{\epsilon}{4})\frac{ckh}{\epsilon^4}.$$

This inequality is equivalent to

$$x \geq c\frac{k\epsilon/4 - 100kh\epsilon^2}{1 - 100\epsilon^2 kh}$$

This, in turn, implies that $x \geq ck\epsilon/8$ since $\epsilon < 1/(800h)$.

For the second part of the claim, fix S_i and denote by y the number of vertices $v \in C_2$ that are strongly assignable to S_i. Once again, by (6), for $i = 2$, we have

$$y\frac{kh}{\epsilon^4} + (\frac{8ckh}{\epsilon} - y)[\frac{(k-1)h}{\epsilon^4} + \frac{100kh^2}{\epsilon^2}] \geq (\frac{k-1}{k} + \frac{\epsilon}{4})\frac{8ckh}{\epsilon}\frac{kh}{\epsilon^4}.$$

As before, we obtain $y \geq ck^2h$ for $\epsilon < 1/(800h)$. This completes the proof of the claim.

We may now apply Lemma 2.4 where the C-vertices are $S_1, \ldots, S_c$, the B-vertices are the vertices of C_2, $r = (k-1)kh$ and $a = ck\epsilon/8$. Here a member of C_2 is connected to an S_i if it is strongly assignable to it. Note that by the last claim, the degree of each B-vertex is at least a, and the degree of each C-vertex is at least $ck^2h \geq rc$. By Lemma 2.4 we obtain an assignment of the vertices of C_2 to the graphs $S_1, \ldots, S_c$ with the following properties (a_i denotes the number of vertices of C_2 that are assigned to S_i):

$$(k-1)kh \leq a_i \leq \frac{c(k-1)kh + 8ckh/\epsilon}{ck\epsilon/8} \leq \frac{72h}{\epsilon^2}. \tag{7}$$

$$a_i = 0 \bmod (k-1)kh, \ \ for \ \ i = t+1, \ldots, c. \tag{8}$$

$$t \leq \frac{((k-1)kh - 1)c}{ck\epsilon/8 + (k-1)kh - 2} \leq \frac{8kh}{\epsilon}. \tag{9}$$

Note that $c - t$ denotes the number of C-vertices to which $0 \bmod (k-1)kh$ B-vertices have been assigned, and that without loss of generality we assume here that $S_1, \ldots, S_t$ are the C-vertices that have not been assigned $0 \bmod (k-1)kh$ vertices of C_2.

Our next mission is to assign the vertices of C_3 to the graphs S_i. This will be done in three stages. Put $u_i = a_i \bmod kh$, $w_i = kh - u_i$, $z_i = w_i \bmod k$ for $i = 1, \ldots, t$. We say that a vertex $v \in C_3$ is l-*assignable* to S_i if v has at least q neighbors in each vertex class of S_i except, maybe, S_i^l.

We now proceed with the first stage. In this stage we assign vertices of C_3 to the S_i's as follows. For all $i = 1, \ldots, t$ and for all $l = 1, \ldots, k$ if $l \leq z_i$ we pick a set of $\lceil w_i/k \rceil$ vertices of C_3 that have not yet been assigned and that are l-assignable to S_i. If $l > z_i$ we pick $\lfloor w_i/k \rfloor$ such vertices. Note that the total number of vertices that are assigned in this process is $\sum_{i=1}^{t} w_i \leq tkh \leq 8k^2h^2/\epsilon$, where the last inequality follows from (9). Therefore, in order to show that this process can be completed it is enough to show that there are at least $8k^2h^2/\epsilon$ vertices of C_3 that are l-assignable to S_i, for $i = 1, \ldots, t$ and $l = 1, \ldots, k$. Indeed, let x be the number of vertices of C_3 having this property. Then by (6) we have

$$x\frac{kh}{\epsilon^4} + (ckh - x)[(k-1)\frac{h}{\epsilon^4} + q] \geq (\frac{k-1}{k} + \frac{\epsilon}{4})ckh\frac{kh}{\epsilon^4}$$

which implies, as in the proof of Claim 2, that $x \geq ck^2h\epsilon/8$. However, by our choice of n_0,

$$c = \frac{m}{khg} > \frac{n\epsilon^4}{4kh} \geq \frac{128h}{\epsilon^2}.$$

Hence, $x \geq 16k^2h^2/\epsilon$. (Note that $8k^2h^2/\epsilon$ vertices suffice here, but the remaining ones will be used in the second stage.)

Having assigned $\sum_{i=1}^{t} w_i$ vertices of C_3 to the graphs $S_1, \dots S_t$ in the first stage, we now proceed to the second stage. Note that $\sum_{i=1}^{t} w_i = 0 \bmod kh$. This is true since, clearly, $|C_2| = 8ckh/\epsilon = 0 \bmod kh$ and hence $\sum_{i=1}^{t} a_i = 0 \bmod kh$ which, in turn, implies that $\sum_{i=1}^{t} u_i = 0 \bmod kh$. Put $t' = \frac{\sum_{i=1}^{t} u_i}{kh}$. Since $u_i \leq kh$ for all i, $t' \leq t$. For all $i = 1, \dots, t'$ and all $l = 1, \dots k$ we assign exactly h vertices of C_3 that have not yet been assigned and that are l-assignable to S_i. The total number of vertices we need to assign in this second stage is $\sum_{i=1}^{t} u_i = t'kh \leq tkh \leq 8k^2h^2/\epsilon$. We have already shown that the number of vertices x of C_3 that are l-assignable to S_i is at least $x \geq 16k^2h^2/\epsilon$, and that at most $8k^2h^2/\epsilon$ of them have been assigned in the first stage. Hence, the second stage can be completed successfully.

The total number of vertices of C_3 that have been assigned until now is $\sum_{i=1}^{t}(u_i + w_i) = tkh$. In the third and final stage we assign the remaining $(c-t)kh$ vertices of C_3 (denoted by C_3') to the graphs $S_{t+1}, \dots, S_c$. We will assign exactly kh vertices to each S_i for $i = t+1, \dots c$, where each assigned vertex has the property that it is weakly assignable to S_i. In order to show that this is possible it is enough to show that each vertex of C_3' is weakly assignable to at least $(c-t)/2$ of the graphs $S_{t+1}, \dots, S_c$ and that each such graph has at least $(c-t)kh/2$ vertices of C_3' that are weakly assignable to it. This will enable us to apply Lemma 2.3 and obtain the desired assignment. (One side of the bipartite graph is C_3' and the other side are the graphs $S_{t+1}, \dots, S_c$, its edges denote weak-assignability.) In fact we will show something slightly stronger; namely, that every $v \in C_3$ is weakly assignable to at least $(c+t)/2$ of the graphs $S_1, \dots, S_c$ and that each such graph has at least $(c+t)kh/2$ vertices in C_3 that are weakly assignable to it.

For $v \in C_3$ let x denote the number of graphs S_i to which v is weakly assignable. By (6) we have:

$$x\frac{kh}{\epsilon^4} + (c-x)[(k-2)\frac{h}{\epsilon^4} + 2q] \geq (\frac{k-1}{k} + \frac{\epsilon}{4})\frac{khc}{\epsilon^4}.$$

This implies that $x \geq c(0.5 + \epsilon k/10) \geq c/2 + t/2$. Similar calculations show that if y is the number of vertices that are weakly assignable to S_i then $y \geq (c+t)kh/2$.

Finally, we need to show that for each $i = 1, \dots, c$, the graph S_i together with the a_i vertices of C_2 assigned to it and the vertices assigned to it from C_3 in the three stages, contains a K-factor. Consider first an S_i where $1 \leq i \leq t'$. S_i has been assigned $a_i = u_i + y_i kh$ vertices of C_2, each of these vertices being strongly assignable to S_i, and therefore they are also l-assignable to it for all $l = 1, \dots, k$. For each $l = 1, \dots, z_i$, S_i has been assigned $\lceil w_i/k \rceil + h$ vertices of C_3 that are l-assignable to it. (Recall that $\lceil w_i/k \rceil$ vertices have been assigned in the first stage and h in the second stage). For $l = z_i + 1, \dots, k$, S_i has been assigned $\lfloor w_i/k \rfloor + h$ vertices of C_3 that are l-assignable to it. Hence the set of assigned vertices to S_i, whose size is $a_i + w_i + kh = (y_i + 2)kh$ can be partitioned into k sets $T_i^1, \dots, T_i^k$ each of size $(y_i + 2)h$, where each element of T_i^l is l-assignable to S_i. (Here we used the fact that $\lceil w_i/k \rceil \le (y_i+2)h$, since $w_i \le (y_i+2)kh$.) We now perform the following process. For each $l = 1, \ldots, k$ and for each $v \in T_i^l$, we select h neighbors of v in S_i^j that have not yet been selected, for $j = 1, \ldots, l-1, l+1, \ldots, k$, and we select $h-1$ vertices of S_i^l that have not yet been selected (these vertices need not be neighbors of v). Clearly, if we manage to do this, v together with these $kh-1$ selected vertices, form a copy of K. To see that this process can be completed, we note that when we handle v, at most $((y_i+2)kh-1)h$ vertices have already been selected from S_i^l, and, if $v \notin T_i^l$, we may need to select h neighbors of v in S_i^l which have not yet been selected. Since v has at least $q = 100h^2k/\epsilon^2$ neighbors in S_i^l, it suffices to show that $100h^2k/\epsilon^2 \ge (y_i+2)kh^2$. This, however is true since by (7), $y_i + 2 < a_i/h \le 72/\epsilon^2$. After completing this process, we have selected exactly $(kh^2-h)(2+y_i)$ vertices from S_i^l. Hence the set of non-selected vertices of S_i form a complete k-partite graph with $h/\epsilon^4 - (kh^2-h)(2+y_i)$ vertices in each color class. Since this number is a multiple of h, we have a K-factor in the remainder of S_i.

The case where $t'+1 \le i \le t$ is similar to the above (and even easier, since S_i has only been assigned vertices of C_3 in the first stage, and hence y_i+2 should be replaced by y_i+1 in all of the computations above).

We remain with the case $t+1 \le i \le c$. In this case, S_i has been assigned a_i vertices of C_2 where $a_i = 0 \bmod (k-1)kh$ and $a_i > 0$. Hence it has been assigned y_ikh vertices of C_2 where $y_i \ge k-1$ and $y_i = 0 \bmod (k-1)$, where these vertices are strongly assignable to S_i. S_i has also been assigned exactly kh vertices of C_3 that are weakly assignable to it. These kh vertices may be partitioned into k sets $T_i^1, \ldots, T_i^k$ where each element of T_i^l is l-assignable to S_i. Note that $|T_i^l| \le kh$. We may now arbitrarily assign the y_ikh strongly assignable vertices to the sets T_i^l in such a way that $|T_i^l| = (y_i+1)h \ge kh$. We remain with the property that each element of T_i^l is l-assignable to S_i. We may now proceed as in the two cases above to obtain the K-factor in S_i and its assigned vertices. $\square$

5 Concluding remarks and open problems

1. Theorem 1.1 is essentially best possible in the sense that the quantity $1 - 1/\chi(H)$ appearing there cannot be replaced by any smaller constant. This is easily seen by letting G be a complete k-partite graph with non-equal color classes where H is any uniquely-colorable k-partite graph with equal color classes. Furthermore, some error term is needed in the statement of Theorem 1.1, i.e., the statement of the theorem becomes false if we omit the ϵ. To see this, let G be the graph obtained from two vertex disjoint complete graphs on $hn/2+1$ vertices each by identifying two vertices of the first with two vertices of the second. Then in

G all the degrees are at least $hn/2$. Let H be a 3-connected bipartite graph on $h = 2l$ vertices (e.g., any complete bipartite graph $K_{a,b}$, where $a + b = 2l, a \geq b \geq 3$), and suppose that $n = 2s + 1$, for some integer s. Clearly, every copy of H in G must be contained completely in one of the two complete graphs forming G. However, by the assumptions $\frac{hn}{2} = l \bmod h$ and hence h does not divide $hn/2 - 1, hn/2$ or $hn/2 + 1$, implying that G does not have an H-factor. Note that in this example H does not have to be a complete bipartite graph and may have color classes of different sizes.

Another example showing that some error term is needed in Theorem 1.1 is the following; let H be the complete bipartite graph $K_{l,l}$, where $l \geq 3$ is odd, and let G be the graph obtained from the complete bipartite graph with color classes of sizes $l(2s+1)+1$ and $l(2s+1)-1$ by adding a perfect matching on the vertices of the larger color class. Here, again, the number of vertices of H, which is $h = 2l$ divides the number of vertices of G, which is $(2s+1)2l = hn$, and the minimum degree in G is $hn/2$. It is, however, easy to check, that G does not have an H-factor. This example can be easily extended to show that some error term is needed in Theorem 1.1 for certain graphs H of any desired chromatic number.

2. By the above remark, some error term is needed in any strengthening of Theorem 1.1. The following strengthening seems true.

   Conjecture 5.1 *For every integer h there exists a constant $c(h)$ such that for every graph H with h vertices, any graph G with hn vertices and with minimum degree $d \geq (1-1/\chi(H))hn + c(h)$ contains an H-factor.*

3. By the Hajnal-Szemerédi result stated in the introduction, the error term in Theorem 1.1 is not needed in case H is a complete graph. As mentioned in the introduction this is also trivially the case if H is a path. It would be interesting to find additional nontrivial graphs H for which no error term is needed. A possible interesting example is the case $H = C_4$, as conjectured by Erdös and Faudree [6].

4. The proof of Theorem 1.1 is algorithmic. That is, for any fixed ϵ and any fixed graph H we can *find* an H-factor in a graph G with $hn > hn_0$ vertices and $d(G) \geq (1 - 1/\chi(H) + \epsilon)hn$ in $O(n^{2.376})$ time. This is true since it is shown in [1] that Lemma 2.1 is algorithmic. (Lemma 2.1 uses the Regularity Lemma of Szemerédi [10], for which an algorithmic version is described in [1]). In particular, the $(1 - \delta)l/s$ copies of S can be found in $O(l^{2.376})$ time. Lemma 2.2 clearly has an $O(m^2)$ randomized algorithm, since one needs only to compute the degrees in the resulting M_1', M_2' and M_3' graphs. The expected number of trials until the conditions

in the Lemma are met is constant. By a standard application of the method of conditional expectations (see, e.g., [9]), it can be shown that the randomized algorithm used in Lemma 2.2 can be derandomized, thereby obtaining an $o(m^{2.376})$ deterministic algorithm for Lemma 2.2. Lemma 2.3 can be implemented in linear (that is $O((cr)^2))$ time, with the appropriate data structures. Lemma 2.4 can also be implemented in $O(b^2)$ time, where b is the number of B-vertices. Combining all of these together, the $O(n^{2.376})$ algorithm follows.

5. The value of l_0 in Lemma 2.1 is a rather huge function of δ and s_0. (In this case, $\log^* l_0$ is a polynomial function of $1/\delta$ and s_0). This is due to the large constants that appear in the proof of the Regularity Lemma [10]. However, in the special case when the graph H in Theorem 1.1 is a tree, one can avoid using Lemma 2.1 in the proof, and hence the dependency of n_0 on ϵ and h is moderate. (In fact, in this case n_0 is a polynomial function of $1/\epsilon$ and h). We omit the details. For trees, however, a result much stronger than Theorem 1.1 can be proved. In [8] it is proved that for every positive integer Δ and any real $\epsilon > 0$, there is a constant $n_0 = n_0(\Delta, \epsilon)$ such that every graph G with $n > n_0$ vertices and $d(G) \geq (0.5 + \epsilon)n$ contains every tree with n vertices and maximum degree Δ as a spanning subgraph. This result cannot, however, be extended to general bounded degree bipartite graphs. Even for $\Delta = 3$ this is not true. There are examples of 3-regular bipartite graphs on n vertices that are not subgraphs of the graph G on n vertices with $d(G) \geq 0.51n$ that is obtained by identifying $0.02n$ vertices of two vertex disjoint cliques of size $n/2 + 0.01n$ each. This is because there are 3-regular expanders (see, e.g., [2]) on n vertices in which for every set Y of $0.49n$ vertices, the set consisting of all vertices of Y and their neighbors is of cardinality strictly greater than $0.51n$.

References

[1] N. Alon, R.A. Duke, H. Lefmann, V. Rödl and R. Yuster, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms 16 (1994), 80-109.

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

[3] N. Alon and R. Yuster, *Almost H-factors in dense graphs*, Graphs and Combinatorics 8 (1992), 95-102.

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

[5] K. Corrádi and A. Hajnal, *On the maximum number of independent circuits in a graph*, Acta Math. Acad. Sci. Hungar. 14 (1963), 423-439.

[6] P. Erdös, *Some recent combinatorial problems*, preprint, November 1990.

[7] A. Hajnal and E. Szemerédi, *Proof of a conjecture of Erdös*, In: Combinatorial theory and its applications, Vol. II (P. Erdös, A. Renyi and V.T. Sós eds.), Colloq. Math. Soc. J. Bolyai 4 (1970), 601-623.

[8] J. Komlós, G. N. Sárközy and E. Szemerédi, *Proof of a packing conjecture of Bollobás*, Combinatorics, Probability and Computing, to appear.

[9] P. Raghavan, *Probabilistic construction of deterministic algorithms: approximating packing integer programs*, Journal of Computer and System Sciences, 37:130–143, 1988.

[10] 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.