DISCRETE MATHEMATICS

Discrete Mathematics 235 (2001) 255–269

www.elsevier.com/locate/disc

Proof of the Alon–Yuster conjecture

János Komlós[a,*], Gábor N. Sárközy[b], Endre Szemerédi[a,c]

[a] *Department of Medicine, Rutgers University, Hill Center, New Brunswick, NJ 08903, USA*
[b] *Worcester Polytechnic Institute, Worcester, Boston, MA, USA*
[c] *Hungarian Academy of Sciences, Hungary*

Abstract

In this paper we prove the following conjecture of Alon and Yuster. Let H be a graph with h vertices and chromatic number k. There exist constants $c(H)$ and $n_0(H)$ such that if $n \geqslant n_0(H)$ and G is a graph with hn vertices and minimum degree at least $(1-1/k)hn + c(H)$, then G contains an H-factor. In fact, we show that if H has a k-coloring with color-class sizes $h_1 \leqslant h_2 \leqslant \cdots \leqslant h_k$, then the conjecture is true with $c(H) = h_k + h_{k-1} - 1$.

1. Introduction

1.1. Notations and definitions

For basic graph concepts see the monograph of Bollobás [3].

$V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G. (A,B,E) denotes a bipartite graph with color-classes A,B and edge set E. For a graph G and a subset U of its vertices, $G|_U$ is the restriction to U of G. $N(v) = N_G(v)$ is the set of neighbors of $v \in V$. Hence $|N(v)| = \deg(v) = \deg_G(v)$, the degree of v. $\delta(G)$ stands for the minimum, and $\Delta(G)$ for the maximum degree in G. $v_i(G)$ denotes the size of a maximum set of vertex disjoint i-stars (stars with i leaves) in G. (Thus $v_1(G) = v(G)$ is the size of a maximum matching.) $K(n_1, n_2, \ldots, n_k)$ is the complete k-partite graph with color-class sizes $n_1, n_2, \ldots, n_k$. When A,B are subsets of $V(G)$, we write $e(A,B) = \#\{(x,y)\colon x \in A,\ y \in B,\ \{x,y\} \in E\}$. In particular, we write $\deg(v,U) = e(\{v\},U)$ for the number of edges from v to U. For a bipartite graph $G = (A,B,E)$, $\delta(A,B)$ and $\Delta(A,B)$ denote the minimum and maximum degrees from A to B. In $G = (A,B,E)$, $v_i(A,B)$ is the size of a maximum set of vertex disjoint i-stars with roots in A. For

* Corresponding author.
E-mail address: komlos@math.rutgers.edu (J. Komlós).

0012-365X/01/$ - see front matter
PII: S0012-365X(00)00279-X

non-empty A and B,

$$d(A,B) = \frac{e(A,B)}{|A||B|}$$

is the *density* of the graph between A and B. In particular, we write $d(A) = d(A,A) = 2|E(G|_A)|/|A|^2$.

Definition 1. The bipartite graph $G = (A,B,E)$ is ε-*regular* if

$$X \subset A,\ Y \subset B,\ |X| > \varepsilon|A|,\ |Y| > \varepsilon|B| \text{ imply } |d(X,Y) - d(A,B)| < \varepsilon,$$

otherwise it is ε-irregular.

We will often say simply that 'the pair (A,B) is ε-regular' with the graph G implicit.

Definition 2. (A,B) is (ε,δ)-*super-regular* if it is ε-regular and

$$\delta(A,B) > \delta|B|, \qquad \delta(B,A) > \delta|A|.$$

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. For example, a K_2-factor is simply a perfect matching.

1.2. H-factors in graphs

Let H be a graph with h vertices and chromatic number k, and let G be a graph on hn vertices. There are several results that show the existence of an H-factor assuming that $\delta(G) \geqslant (1-1/k)hn$. If H is a path of length $h-1$, then a classical result of Dirac [5] says that $\delta(G) \geqslant hn/2$ implies that G contains a Hamiltonian cycle, and thus also an H-factor. Corrádi and Hajnal [4] proved that for $H = K_3$ the condition $\delta(G) \geqslant 2n$ suffices, and Hajnal and Szemerédi [7] extended this to show that for $H = K_h$ the condition $\delta(G) \geqslant (h-1)n$ guarantees an H-factor. A conjecture of Erdős and Faudree [6] asserts that $\delta(G) \geqslant 2n$ implies the existence of a C_4-factor.

For a general H, Alon and Yuster [1] showed that if $\delta(G) \geqslant (1-1/k)hn$, then G contains $(1-\mathrm{O}(1))n$ vertex disjoint copies of H (an 'almost H-factor'). Later in [2] they showed that for every $\varepsilon > 0$ and h there exists an $n_0 = n_0(\varepsilon, h)$ such that if $n \geqslant n_0$ and $\delta(G) \geqslant ((1-1/k)+\varepsilon)hn$, then G has an H-factor. They indicated that some error term is needed here in the minimum degree, i.e. the statement is false for $\varepsilon = 0$. They gave the following two examples to see this. In the first example, let h be even and n be odd, 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 in the second, and let H be any 3-connected bipartite graph on h vertices (e.g., $K_{l,l}$ with $l \geqslant 3$). Then $\delta(G) = hn/2 = (1-1/k)hn$ but clearly G does not have an H-factor.

We modify their second example as follows: Let $l \geqslant 3$ be odd, $h = 2l$, and let n be odd and sufficiently large. Let H be $K_{l,l}$, and let G be the graph obtained from the complete bipartite graph with color-class sizes $hn/2 - 1$ and $hn/2 + 1$ by adding an $(l - 1)$-factor in the larger color-class and an $(l - 3)$-factor in the smaller color-class, such that neither of these factors contain a $K_{2,2}$. (It is not hard to see that such factors exist.) Now $\delta(G) = hn/2 + (l - 2) = (1 - 1/k)hn + (l - 2)$, but it is easy to see that G does not have an H-factor.

They also conjectured in [2] that εhn is not the best possible error term, and a constant $c(H)$ would suffice. In this paper we prove this conjecture.

Theorem 1. *Let H be a graph with h vertices and chromatic number k, and assume that H has a k-coloring with color-class sizes $h_1 \leqslant h_2 \leqslant \cdots \leqslant h_k$. There is a threshold $n_0(H)$ such that if $n \geqslant n_0(H)$ and G is any graph with hn vertices and minimum degree*

$$\delta(G) \geqslant \left(1 - \frac{1}{k}\right) hn + h_k + h_{k-1} - 1 \quad (1)$$

then G contains an H-factor.

Remark. The second example above shows that the theorem is not true with $c(H) = h_k - 2$.

2. The main tools

In the proof the Regularity Lemma of the third author plays a central role. Here we will use the following variation of the lemma.

Lemma 2 (Regularity Lemma — Degree form). *For every $\varepsilon > 0$ there is an $M = M(\varepsilon)$ such that if $G = (V, E)$ is any graph and $d \in [0, 1]$ is any real number, then there is a partition of the vertex-set V into $l + 1$ sets (so-called clusters) $V_0, V_1, \ldots, V_l$, and there is a subgraph $G' = (V, E')$ with the following properties:*

- $l \leqslant M$,
- $|V_0| \leqslant \varepsilon |V|$,
- *all clusters V_i, $i \geqslant 1$, are of the same size $L \leqslant \lceil \varepsilon |V| \rceil$.*
- $\deg_{G'}(v) > \deg_G(v) - (d + \varepsilon)|V|$ *for all* $v \in V$,
- $G'|_{V_i} = \emptyset$ *(V_i are independent in G'),*
- *all pairs $G'|_{V_i \times V_j}$, $1 \leqslant i < j \leqslant l$, are ε-regular, each with a density* 0 *or exceeding d.*

This form can easily be obtained by applying the original Regularity Lemma (with a smaller value of ε), adding to the exceptional set V_0 all clusters incident to many irregular pairs, and then deleting all edges between any other clusters, where the edges either do not form a regular pair or they do but with a density at most d.

Our other main tool is the above-mentioned Hajnal–Szemerédi theorem.

Lemma 3 (Hajnal, Szemerédi [7]). *Let G be a graph on $n = sk$ vertices. If $\delta(G) \geqslant (k - 1)n/k$ then G contains s vertex-disjoint cliques of order k.*

In fact, we are going to use the following easy consequence of this lemma.

Lemma 4. *Let G be a graph on n vertices. If $\delta(G) \geqslant ((k-1)/k)n - x$ for some natural number x, then apart from at most $k(k-1)x + (k-1)^2$ exceptional vertices, $V(G)$ can be covered by vertex-disjoint cliques of order k.*

Indeed, add kx extra vertices to G and possibly a few ($\leqslant k-1$) more to achieve that the new number of vertices is divisible by k. Connect the new vertices to all other vertices. Denote the resulting graph by $\tilde{G}$ and the new number of vertices by $\tilde{n}$. It is easy to see that $\delta(\tilde{G}) \geqslant (k-1)\tilde{n}/k$, therefore using Lemma 3 we can cover $\tilde{G}$ by vertex-disjoint cliques of order k. The number of vertices in $V(G)$ which are in cliques containing at least one extra vertex ($\in V(\tilde{G}) \setminus V(G)$) is at most $k(k-1)x + (k-1)^2$.

We also use the Blow-up Lemma (see [10,12]):

Lemma 5. *Given a graph R of order r and positive parameters δ, Δ, there exists an $\varepsilon = \varepsilon(\delta, \Delta, r) > 0$ such that the following holds. Let $n_1, n_2, \ldots, n_r$ be arbitrary positive integers, and let us replace the vertices $v_1, v_2, \ldots, v_r$ of R with pairwise disjoint sets $V_1, V_2, \ldots, V_r$ of sizes $n_1, n_2, \ldots, n_r$ (blowing up). We construct two graphs on the same vertex-set $V = \bigcup V_i$. The first graph R_b is obtained by replacing each edge $\{v_i, v_j\}$ of R with the complete bipartite graph between the corresponding vertex-sets V_i and V_j. A sparser graph G is constructed by replacing each edge $\{v_i, v_j\}$ of R arbitrarily with some (ε, δ)-super-regular pair between V_i and V_j. If a graph H with $\Delta(H) \leqslant \Delta$ is embeddable into R_b then it is already embeddable into G.*

When using the Blow-up Lemma, we typically also need the following strengthened version: Given $c > 0$, there are positive functions $\varepsilon = \varepsilon(\delta, \Delta, r, c)$ and $\alpha = \alpha(\delta, \Delta, r, c)$ such that the Blow-up Lemma remains true if for every i there are certain vertices x to be embedded into V_i whose images are a priori restricted to certain sets $C_x \subset V_i$ provided that

(i) each C_x within a V_i is of size at least $c|V_i|$,
(ii) the number of such restrictions within a V_i is not more than $\alpha|V_i|$.

Finally we are going to use the following simple fact about the existence of stars.

Lemma 6. (a) *For any positive integers i, δ there exists an $n_0 = n_0(i, \delta)$ such that if G is a graph on $n \geqslant n_0$ vertices with $\delta(G) \geqslant \delta$, then we have*

$$v_i(G) \geqslant \max\left(\delta - i + 1, \frac{(\delta - i + 1)n}{(i+1)(\delta + \Delta(G))}\right). \tag{2}$$

(b) *For any bipartite graph $G = (A, B, E)$ we have*

$$v_i(A, B) \geqslant \frac{(\delta - i + 1)|A|}{\delta(A, B) + i\Delta(B, A)}.$$

Indeed, to prove (a) let $v_i(G) = m$, and consider m vertex disjoint i-stars in G. Then for the number of edges E between the stars and the remaining vertices we get

$$(n - m(i+1))(\delta - (i-1)) \leqslant E$$
$$\leqslant \min(m(n - m(i+1) + i(i-1)), m(i+1)\Delta(G)),$$

which proves the fact. The proof of (b) is similar.

3. Outline of the proof

In a series of papers [8–13] we have developed a general method based on the Regularity Lemma and the Blow-up Lemma for embedding problems in dense graphs. In this paper we use this method again, so the proof follows a similar rough outline as the proof in [13] for example, however, several new ideas are needed.

We will assume throughout the paper that H is fixed and n is sufficiently large. We will use the following main constants:

$$\varepsilon \ll d \ll \alpha \ll 1, \tag{3}$$

where $a \ll b$ means that a is sufficiently small compared to b. We will not compute the actual dependencies.

We apply Lemma 2 for G with ε and d as in (3). We get a partition of $V(G)$ into clusters $V_0, V_1, \ldots, V_l$. We define the following *reduced graph* G_r: The vertices of G_r are the clusters V_i, $i \geqslant 1$, and we have an edge between two clusters if they form an ε-regular pair in G' with density exceeding d. Since in G', $\delta(G') \geqslant ((k-1)/k - (d + \varepsilon))hn$, an easy calculation shows that in G_r we have

$$\delta(G_r) \geqslant \left(\frac{k-1}{k} - 3d\right) l. \tag{4}$$

Let us apply Lemma 4 for G_r to get a covering of most of the vertices in G_r by vertex disjoint cliques of size k. More precisely we can cover the vertices of G_r apart from an exceptional set of size at most $3k(k-1)dl + (k-1)^2 \leqslant 4k(k-1)dl$. Let us put the vertices of these exceptional clusters into the exceptional set V_0. For simplicity V_0 still denotes the resulting set. Then

$$|V_0| \leqslant 4k(k-1)dlL + \varepsilon hn \leqslant 5k(k-1)dhn. \tag{5}$$

In the proof first we assume until Section 6 that neither of the following two extremal conditions holds for G:

Extremal condition 1 (EC1). *$k = 2$ and there exists a partition $V(G) = A \cup B$ such that*

- $|A| = \lfloor \frac{hn}{2} \rfloor$, *and*
- $d(A, B) < \alpha$.

Extremal condition 2 (EC2). *There exists an* $A \subset V(G)$ *such that*

- $|A| = \lfloor \frac{hn}{k} \rfloor$, *and*
- $d(A) < \alpha$.

We show later in Sections 6 and 7 that if either of these conditions hold, then we can find an H-factor. First in the next section we show that under the assumption that these extremal cases do not hold we can slightly modify the clique covering; we can achieve that a constant proportion of the cliques are $(k+1)$-cliques and the rest are k-cliques. This will significantly simplify the adjustment procedure in Section 5. These cliques will be denoted by $K_1, K_2, \dots, K_s$ and we denote the clusters in K_i by $V_1^i, V_2^i, \dots, V_k^i$ $(, V_{k+1}^i$ if K_i is a $(k+1)$-clique).

The rough idea of the proof in the non-extremal case is to reduce the problem of finding an H-factor to the cliques K_i, which can be done with the use of the Blow-up Lemma. For this purpose, first in Section 5 we will take care of the various exceptional vertices and make some adjustments.

We define a sequence of k-partite graphs $H^1, H^2, \dots, H^k$ in the following way. $H^1 = H$ and in general H^i is the union of i vertex disjoint copies of H with the coloring, where the difference between the sizes of the largest and the smallest color-class is as small as possible. Denote the color-classes of H^i by $h_1^i, h_2^i, \dots, h_k^i$, where $h_1^i \leqslant h_2^i \leqslant \cdots \leqslant h_k^i$. It is easy to see that for any $1 \leqslant i \leqslant k$, we have $|h_k^i - h_1^i| \leqslant h_k - h_1$. Also $h_1^k = h_2^k = \cdots = h_k^k = h_1 + h_2 + \cdots + h_k$. We can get H^k for example as the vertex disjoint union of k copies of H, where the ith copy has $h_i, h_{i+1}, \dots, h_k, h_1, \dots, h_{i-1}$ vertices in color-classes $1, 2, \dots, k$ (where $h_0 = h_k$).

4. Modifying the clique cover

We remove $(\sqrt{d})l/k$ (for simplicity we assume that this number is an integer) k-cliques from the clique cover. Let us denote the number of remaining k-cliques by s. Our goal in this section is to show that by slightly changing the remaining cliques and by redistributing the removed clusters, we can get a new clique cover in which $(\sqrt{d})lk/k = \sqrt{d}l$ of the cliques are $(k+1)$-cliques and the remaining $s - \sqrt{d}l$ cliques are k-cliques.

Let us consider an arbitrary removed cluster C. If there is a k-clique K in the current cover (C might not be the first cluster we redistribute) such that we have $(C, C') \in E(G_r)$ for every $C' \in K$, then we just add C to K, we have one more $(k+1)$-clique and we can move to the next removed cluster. Thus we may assume that there is no k-clique K with this property. Using this facts, (4) and (5), and an easy calculation shows that the number of k-cliques K, for which

$$|\{C' \mid C' \in K,\ (C, C') \in E(G_r)\}| = k - 1$$

is at least $(1 - d^{1/3})s$. We consider only these k-cliques and in these cliques we call the cluster that is not a neighbor of C, a C-exchangable cluster. Indeed, these clusters are exchangable with C. Let us denote the set of C-exchangable clusters by S. Similarly as above, if we have a $C' \in S$ and k-cliques K, K' such that $C' \in K$ and $(C', C'') \in E(G_r)$ for every $C'' \in K'$, then again we are done, since we remove C' from K and we add it to K', we add C to K and thus we have one more $(k+1)$-clique. Hence we may assume that there is no C' with this property.

However, in this case the fact that EC2 does not hold, (3), (4) and some computation imply that we can find cliques K, K' with $C_1 = K \cap S$, $C_2 = K' \cap S$ such that

- $(C_1, C_2) \in E(G_r)$.
- There exists a cluster $C_3 \in (K \backslash C_1)$ with $(C_2, C_3) \notin E(G_r)$.
- $N_{G_r}(C_2) \cap K = K \backslash C_3$, $N_{G_r}(C_3) \cap K' = K' \backslash C_2$.

Here we also used the fact that C_3 is C-exchangable in 2 steps. Indeed, we remove C_2 from K' and we add C to it, we remove C_3 from K and we add C_2 to it, and now C_3 plays the role of C.

But then we exchange C_2 and C_3 among K and K', we add C to K' and thus creating one more $(k+1)$-clique again. By repeating this procedure we obtain a clique sequence $K_1, K_2, \ldots, K_s$ where the first $s' = \sqrt{d}\, l$ cliques are $(k+1)$-cliques and the others are k-cliques.

5. Adjustments and the handling of the exceptional vertices

We already have an exceptional set V_0 of vertices in G. We add some more vertices to V_0 to achieve super-regularity. From a cluster V_j^i in a clique K_i we remove all vertices v for which there exists an j' with $1 \leqslant j' \leqslant k$ ($k+1$ if $1 \leqslant i \leqslant s'$), $j' \neq j$ such that

$$\deg(v, V_{j'}^i) \leqslant (d - \varepsilon)|V_{j'}^i|.$$

ε-regularity guarantees that at most $k\varepsilon|V_j^i| \leqslant k\varepsilon L$ such vertices exist in each cluster V_j^i.

At this point we may have a small discrepancy among the number of remaining vertices in each clique K_i. By removing extra vertices from certain clusters (and put them into the exceptional set V_0) we achieve that each cluster has exactly L' vertices where L' is divisible by h. (We will still use the notation V_0 for the enlarged exceptional set.) We still have

$$|V_0| \leqslant 6k(k-1)dhn. \tag{6}$$

Next we take care of the vertices in V_0. We group the vertices in V_0 into blocks of h vertices (note that $h / |V(G)| = hn$).

Consider the first block of vertices $v_1, v_2, \ldots, v_h$. First we show that we may assume that these vertices all came from the same cluster, unless $k = 2$ and we have our first extremal case (EC1). Consider first v_1 and v_2. For every clique K_i, $1 \leqslant i \leqslant s$ we determine a label $(x_1^{K_i}, x_2^{K_i})$ in the following way. $x_j^{K_i}$, $j = 1, 2$ is the number of clusters

$C \in K_i$ for which

$$\deg(v_j, C) \geqslant d|C|. \tag{7}$$

Let us assume first that

$$x_1^{K_i} + x_2^{K_i} \geqslant 2k - 1 \quad \text{for an } i > s'. \tag{8}$$

If we have an equality in (8), then say (7) is not true for v_2 and for cluster $C \in K_i$, otherwise $C \in K_i$ is arbitrary. Then we may exchange v_1 and v_2 for two vertices in C, so now they came from the same cluster. When we add v_1 and v_2 to this cluster C, we immediately eliminate these two vertices by removing two copies of H^k (see Section 3) from K_i, one containing v_1 and the other containing v_2. (Here and later naturally the color-classes of an H^k come from different clusters in K_i.) Thus we still have the same number of vertices left in the clusters in K_i and this number is still divisible by h.

Similarly, if

$$x_1^{K_i} + x_2^{K_i} \geqslant 2k \quad \text{for an } 1 \leqslant i \leqslant s', \tag{9}$$

then we can exchange v_1 and v_2 for two vertices from the same cluster. Again, when we add v_1 and v_2 to this cluster, we immediately eliminate these two vertices by removing two copies of H^k from K_i, one containing v_1 and the other containing v_2. Finally, (1), (6), the fact that EC1 does not hold and some computation show that we have at least $d^{2/3}s$ cliques that satisfy either (8) or (9).

Thus we may assume that v_1 and v_2 came from the same cluster C. Then we do the same procedure for C and v_3, i.e. we define $x_1^{K_i} = \deg_{G_r}(C, K_i)$. By repeating this procedure we may assume that $v_1, v_2, \ldots, v_h$ all came from the same cluster C. For this cluster we find a K_i such that if $i > s'$, then

$$(C, C') \in E(G_r) \quad \text{for all } C' \in K_i$$

and if $1 \leqslant i \leqslant s'$, then there exist k clusters $C' \in K_i$ (denote their clique by K_i') such that

$$(C, C') \in E(G_r).$$

(4) and (6) show that we have at least $d^{2/3}s$ such cliques for the cluster C.

Assume first that for this clique K_i we have $i > s'$. We redistribute the vertices $v_1, v_2, \ldots, v_h$ among the clusters in K_i, we add h_j vertices to V_j^i for $1 \leqslant j \leqslant k$. In K_i we find h copies of H^k such that each H^k copy contains exactly one of the added h vertices. It is easy to see that this can be done. Furthermore, in the remaining vertices in K_i we find a copy of H, such that it contains h_j vertices from V_j^i for $1 \leqslant j \leqslant k$. Removing this H copy and the H^k copies, all clusters have the same number of remaining vertices and this number is divisible by h.

In case $1 \leqslant i \leqslant s'$ for this K_i, then we follow the same procedure with K_i' instead of K_i. However, we might have a discrepancy in the number of remaining vertices in the clusters of K_i. We can eliminate this discrepancy with the following process. First, the number of remaining vertices in the clusters of K_i is divisible by h. We always remove the cluster with the least number of remaining vertices from K_i and from the remaining k-clique we remove a copy of H^k. By repeating this procedure we can achieve that in all clusters in K_i we have the same number of remaining vertices and this number is divisible by h.

Next we handle the second block of h vertices in V_0, etc. Unfortunately, because $|V_0|$ is quite large, we cannot just repeat this procedure for all vertices in V_0, since we might hurt the super-regularity. Note that we never hurt the ε-regularity. Therefore we do the following. We define κ as $\varepsilon \ll \kappa \ll d$. We maintain another set Q beside V_0. Initially $Q = \emptyset$. After handling $\lfloor \kappa n \rfloor$ vertices from V_0, we update Q in the following way. From a cluster V_j^i in a clique K_i we remove all vertices v and add them to Q for which there exists a j' with $1 \leqslant j' \leqslant k$ ($k+1$ if $1 \leqslant i \leqslant s'$), $j' \neq j$ such that $\deg(v, V_{j'}^i) \leqslant (d-\varepsilon)|V_{j'}^i|$. Here we only consider the remaining vertices in a cluster. We also remove some extra vertices to make sure that we have the same number of vertices left in the clusters in K_i and this number is divisible by h. ε-regularity guarantees that we added at most $k\varepsilon hn$ vertices to Q. We add some more vertices to Q from V_0 to guarantee that the number of vertices in Q is divisible by h (if V_0 is empty then $|Q|$ is divisible by h already). Then we handle the vertices in Q exactly the same way as the exceptional vertices above. Next we handle the next $\lfloor \kappa n \rfloor$ vertices of V_0, after this we update Q and we handle the new vertices in Q, etc.

It is not hard to see that we can achieve that we are left with the following situation. In each clique K_i we have the same number of remaining vertices in each cluster and this number is divisible by h (and it is $\geqslant \frac{3}{4}L$ say). For the cliques K_i, $i > s'$, Lemma 5 finds an H^k-factor and thus an H-factor. For the cliques $K_i, 1 \leqslant i \leqslant s'$ we do the following. We redistribute the vertices in V_{k+1}^i among the other k clusters in K_i in the following way. First we add $\lfloor |V_{k+1}^i|/kh \rfloor h$ vertices to each of the k clusters. Then for the remaining rh $(0 \leqslant r < k)$ vertices, we add rh_j vertices to V_j^i for $1 \leqslant j \leqslant k$. We eliminate these rh vertices by removing rh copies of H^k from these k clusters, each containing exactly one of the rh vertices. Then we remove r copies of H, each containing h_j vertices V_j^i for $1 \leqslant j \leqslant k$. In the leftover in each of the k clusters we have the same number of vertices left and this number is divisible by h. Lemma 5 finds an H^k-factor and thus an H-factor in them, finishing the proof in the non-extremal case.

6. The extremal cases

6.1. EC1

In this section we assume that the first extremal case (EC1) is satisfied so $k=2$ and we have a partition $V(G) = A_1 \cup A_2$ with $|A_1| = \lfloor hn/2 \rfloor$ and $d(A_1, A_2) < \alpha$. Let

$$hn = \left\lfloor \frac{n}{2} \right\rfloor 2h + rh \quad \text{where } r = 0 \text{ or } 1.$$

In A_1 (and similarly for A_2) we can have at most $\alpha^{2/3}|A_1|$ exceptional vertices $v\in A_1$ for which we have

$$\deg(v,A_2)\geqslant\alpha^{1/3}|A_2|. \tag{10}$$

We call these exceptional vertices in A_1 1-bad. We have to handle the bad vertices first.

More precisely, we have to eliminate a special type of bad vertices; for a vertex $v\in A_1$ (and similarly in A_2) we say that it is *exceptional*, if

$$\deg(v,A_1)\leqslant\frac{\alpha^{1/3}}{2}|A_1|.$$

Note that if a vertex $v\in A_1$ is exceptional then it is 1-bad.

First we have to eliminate the exceptional vertices. The other bad vertices are not causing any further complications.

We may assume that we have either no 1-bad vertices, or there are no exceptional vertices in A_2. Indeed, otherwise we could exchange an 1-bad vertex in A_1 with an exceptional vertex in A_2 and this way we decreased the number of 1-bad vertices. By iterating this procedure we can achieve that either we have no more 1-bad vertices, or there are no more exceptional vertices left in A_2. Thus we can have exceptional vertices only in at most one of the sets A_1 and A_2. Assume first that we have exceptional vertices in A_1. Then we have no 2-bad vertices. We remove the exceptional vertices from A_1 and we add them to A_2. For simplicity we still denote the sets by A_1 and A_2. Let $|A_1|=\lfloor|A_1|/h\rfloor h+x$, where $0\leqslant x<h$. Using (1) and Lemma 6.b we find a set of x h_1-stars in $G|_{A_1\times A_2}$ which are vertex disjoint from each other and the exceptional vertices added to A_2. Here we used $c(H)\geqslant h_1-1$. We remove the roots of these stars from A_1 and add them to A_2. Hence now we have $h/|A_1|$ and $h/|A_2|$. We remove x copies of H from $G|_{A_2}$ such that each copy contains exactly one root. Then trivially (the densities are close to 1 and there are no exceptional vertices) there is an H-factor in $G|_{A_1}$ and in $G|_{A_2}$.

Assume now that we have no exceptional vertices in A_1 and in A_2. In case n is even, we are done, so let us assume that n is odd. If we have $\lfloor h/2\rfloor$ 1-bad vertices, or $\lceil h/2\rceil$ 2-bad vertices, then we can move these vertices to the other set and we are done again. Hence we may assume that we have $x_1<\lfloor h/2\rfloor$ 1-bad vertices and $x_2<\lceil h/2\rceil$ 2-bad vertices. Using (1) and Lemma 6.b we can find $\lfloor h/2\rfloor-x_1$ h_1-stars in $G|_{A_1\times A_2}$ which are vertex disjoint from each other and the bad vertices. Here we used $c(H)\geqslant h_1+h_2-1\geqslant\lceil h/2\rceil+h_1-1\geqslant x_2+h_1$ (this is the only place where we used the extra h_{k-1} term in $c(H)$). We remove the 1-bad vertices and the roots of these stars from A_1 and add them to A_2. Hence now we have $h/|A_1|$ and $h/|A_2|$. Thus again there is an H-factor in $G|_{A_1}$ and in $G|_{A_2}$.

6.2. EC2

In this section we assume that the second extremal case (EC2) is satisfied so we have an $A\subset V(G)$ with $|A|=\lfloor hn/k\rfloor$ and $d(A)<\alpha$. The ideas are going to be similar to the ones used in EC1. We define m as the largest integer for which $h_m \leqslant \lfloor h/k \rfloor$, so in particular H have equal color-classes ($h_1 = h_2 = \cdots = h_k = h/k$) if and only if $m = k$. Also, let

$$hn = \left\lfloor \frac{n}{k} \right\rfloor kh + rh \quad \text{where } 0 \leqslant r < k. \tag{11}$$

First let us assume that we have the following special case: there exists a partition

$$V(G) = A_1 \cup A_2 \cup \cdots \cup A_k,$$

with $|A_i| = \lfloor hn/k \rfloor$ or $|A_i| = \lceil hn/k \rceil$ for $1 \leqslant i \leqslant k$ and $d(A_i) < \alpha$ for $1 \leqslant i \leqslant k$. In each A_i we can have at most $\alpha^{2/3}|A_i|$ exceptional vertices $v \in A_i$ for which we have

$$\deg(v, A_i) \geqslant \alpha^{1/3}|A_i|. \tag{12}$$

Again we call these exceptional vertices in A_i i-bad. For simplicity let us assume first that we have no i-bad vertices for any $1 \leqslant i \leqslant k$. In this case if $m = k$, then using the Blow-up Lemma (Lemma 5) we can find an H-factor. If $m < k$, then we have to move some vertices around before we apply the Blow-up Lemma, so we do the following. For each $A_i, 1 \leqslant i \leqslant m$ we find a set S_i of

$$s_i = |A_i| - \left\lfloor \frac{n}{k} \right\rfloor h - h_i^r$$

vertex disjoint h_1-stars in $G|_{A_i}$. (1) and Lemma 6.a make this possible. Here we used $c(H) \geqslant h_1$. We remove the roots of these stars from the corresponding A_i, and we add them to the A_i-s with $m < i \leqslant k$ such that now we have

$$|A_i| = \left\lfloor \frac{n}{k} \right\rfloor h + h_i^r \quad \text{for every } 1 \leqslant i \leqslant k.$$

(11) implies that this is possible. We remove $\sum_{i=1}^{m} rs_i$ copies of H^k (again for an H^k every color-class of size h comes from a different A_i) such that each added vertex is contained in exactly one of these H^k-s (the stars make this possible). Then we remove a copy of H^r containing h_i^r vertices for $A_i, 1 \leqslant i \leqslant k$. In the leftover in each $A_i, 1 \leqslant i \leqslant k$ we have the same number of vertices left and this number is divisible by h. Lemma 5 finds an H^k-factor and thus an H-factor.

In case we have bad vertices satisfying (12) the main idea is the same but we have to handle the bad vertices first. More precisely, again we have to eliminate a special type of bad vertices; for a vertex $v \in A_i$ we say that it is *j-exceptional* ($j \neq i$), if

$$\deg(v, A_j) \leqslant \frac{\alpha^{1/3}}{2}|A_j|.$$

Note that if a vertex $v \in A_i$ is j-exceptional for some $j \neq i$ then it is i-bad.

First we have to eliminate the i-exceptional vertices for every $1 \leqslant i \leqslant k$. The other bad vertices are not causing any further complications.

Again we may assume that for every $1 \leqslant i \leqslant k$ we have either no i-bad vertices, or there are no i-exceptional vertices in the other A_j-s ($j \neq i$). Indeed, otherwise we could exchange an i-bad vertex in A_i with an i-exceptional vertex in A_j and this way would have decreased the number of i-bad vertices. By iterating this procedure we can achieve that either we have no more i-bad vertices, or there are no more i-exceptional vertices left.

Next we eliminate the i-exceptional vertices for every $1 \leqslant i \leqslant k$. Consider an $1 \leqslant i \leqslant k$. By the above remark if there exist i-exceptional vertices in other A_j-s (say we have x_i of them), then we do not have i-bad vertices, and thus $\Delta(G|_{A_i}) \leqslant \alpha^{1/3}|A_i|$. Using this fact, (1) and Lemma 6.a we can find a set S_i' of x_i vertex disjoint h_1-stars in $G|_{A_i}$. Here we used $c(H) \geqslant h_1 - 1$. We repeat this procedure for every $1 \leqslant i \leqslant k$. We exchange the roots of the stars in S_i' with the i-exceptional vertices.

The construction of the stars $S_i, 1 \leqslant i \leqslant m$ is similar as above. If $S_i' \neq \emptyset$, so in particular we had no i-bad vertices in A_i, then by using Lemma 6.a clearly we can find the s_i h_1-stars in $G|_{A_i}$ which are vertex disjoint from each other and the other stars in S_i'. In case $S_i' = \emptyset$ (so we may have i-bad vertices) the situation is somewhat more complicated. We still find the s_i vertex disjoint h_1-stars in $G|_{A_i}$ by using Lemma 6.a. However, these stars now may contain i-bad vertices, which is a problem. First we make sure that all stars contain at most one i-bad vertex. For this purpose, if in a star we have at least two i-bad vertices (where one of them is denoted by v) then we replace this star with another h_1-star whose root is v and the leaves are not i-bad and disjoint from the other stars. (12) makes this possible. If a star has no i-bad vertices or the one i-bad vertex is not j-exceptional for any $j \neq i$, then the star is put in S_i. Otherwise the only possibility is that the root v of this star is also a root of another h_1-star in some S_j', $j \neq i$. In this case if A_j is the set where we are planning to add v, then we just remove v from A_i and add it to A_j. Otherwise, we pick an h_k-star for S_i whose root is v and this star together with the star in S_j' form a $K(1, h_1, h_k)$. As above, for the stars in $S_i, 1 \leqslant i \leqslant m$ we remove the roots and add it to other A_i-s to achieve

$$|A_i| = \left\lfloor \frac{n}{k} \right\rfloor h + h_i^r \quad \text{for every } 1 \leqslant i \leqslant k.$$

We remove copies of H^k such that each copy contains exactly one root of a star in $\bigcup_{i=1}^k (S_i \cup S_i')$ (this root might be the same for a star in S_i and a star in S_j' for some i, $j \neq i$). Again the stars make this possible. Then we remove a copy of H^r containing h_i^r vertices for $A_i, 1 \leqslant i \leqslant k$. In the leftover in each $A_i, 1 \leqslant i \leqslant k$ we have the same number of vertices left, this number is divisible by h and we have no i-exceptional vertices left for any $1 \leqslant i \leqslant k$. Lemma 5 finds an H^k-factor and thus an H-factor.

In the general case in EC2 first we have an $A_1 \subset V(G)$ with $|A_1| = \lfloor hn/k \rfloor$ and $d(A_1) < \alpha$. If possible, we take an $A_2 \subset V(G) \backslash A_1$ in the leftover with $|A_2| = \lfloor hn/k \rfloor$ and $d(A_2) < \alpha$. We continue this process until we can, or there is no $A_{l+1} \subset V(G) \backslash (A_1 \cup \cdots \cup A_l)$ with $|A_{l+1}| = \lfloor hn/k \rfloor$ and $d(A_{l+1}) < \alpha$. Put $B = V(G) \backslash (A_1 \cup \cdots \cup A_l)$. If $l = k-1$ we get back the special extremal case that we just discussed (with somewhat worse α). Assume first that either $l \leqslant k-3$ or in case $l = k-2$, $G|_B$ does not satisfy EC1. We define i-bad vertices in $A_i, 1 \leqslant i \leqslant l$ just as in (12). In B the bad vertices are vertices v

with

$$\deg(v, A_1 \cup \cdots \cup A_l) \leqslant (1 - \alpha^{1/3})|A_1 \cup \cdots \cup A_l|. \tag{13}$$

Again let us assume first that there are no bad vertices. Since $G|_B$ does not satisfy the extremal conditions EC1 and EC2 for $k-l$, if H' is a graph with $\chi(H') = k - l$, then the method described in the previous sections succeeds in finding an H'-factor in $G|_B$. But before this, again if H does not have equal color-classes we have to adjust the cardinalities. Thus again as above for each $A_i, 1 \leqslant i \leqslant m' = \min(l, m)$ we find a set S_i of $s_i = |A_i| - \lfloor n/k \rfloor h - h_i^r$ vertex disjoint h_1-stars in $G|_{A_i}$. We remove the roots of these stars from the corresponding A_i-s and we add them to the A_i-s with $m < i \leqslant l$ (if $m < l$) and B such that now we have $|A_i| = \lfloor n/k \rfloor h + h_i^r$ for every $1 \leqslant i \leqslant l$. We remove $\sum_{i=1}^{m'} rs_i$ copies of H^k (the first l color-classes come from $A_i, 1 \leqslant i \leqslant l$, the others from B) such that each H^k copy contains exactly one root of a star in S_i. Then we remove a copy of H^r containing h_i^r vertices from $A_i, 1 \leqslant i \leqslant l$. Denote the resulting sets by $A_i', 1 \leqslant i \leqslant l, B'$. We have $h/|A_i'|$, $h/|B'|$, $|A_1'| = |A_2'| = \cdots = |A_l'|$ and $|B'| = (k-l)|A_1'|$. We define H' as the last $k-l$ color-classes of H^k. As mentioned above, the non-extremal method described in the previous sections finds an H'-factor in $G|_{B'}$ (note that for the non-extremal case a weaker degree condition is sufficient than (1)). We define a new set $B'' = \{v_1, v_2, \ldots, v_{|B'|/(k-l)h}\}$, where each vertex v_i corresponds to a copy of H' in the H'-factor of $G|_{B'}$. We also define G' on $A_1' \cup A_2' \cup \cdots A_l' \cup B''$ as $G|_{A_1' \cup \cdots \cup A_l'}$ and every $v_i \in B''$ is adjacent to all the common neighbors of all the vertices in the corresponding copy of H'. Finally we define H'' as the first l color-classes of H^k and one extra vertex that is adjacent to all other vertices. Then the Blow-up Lemma (Lemma 5) finds an H''-factor in G'. This implies an H^k-factor in $G|_{A_1' \cup \cdots \cup A_l' \cup B'}$ and finally an H-factor in G.

The handling of the bad vertices is very similar to the above special case and the details are left to the reader.

Finally let $l = k-2$. We may also assume that EC1 holds so there is a partition $B = B_1 \cup B_2$ with $|B_1| = \lfloor hn/k \rfloor$ and $d(B_1, B_2) < \alpha$. Again for simplicity we assume that there are no bad vertices. We follow the same procedure as above. Thus again for each $A_i, 1 \leqslant i \leqslant m' = \min(l, m)$ we find a set S_i of $s_i = |A_i| - \lfloor n/k \rfloor h - h_i^r$ vertex disjoint h_1-stars in $G|_{A_i}$. We remove the roots of these stars from the corresponding A_i-s and we add them to the A_i-s with $m < i \leqslant l$ (if $m < l$) and B such that now we have $|A_i| = \lfloor n/k \rfloor h + h_i^r$ for every $1 \leqslant i \leqslant l$. When we add a root to B we add it to the B_i where it has more neighbors. However, before removing the copies of H^k containing the roots and the copy of H^r, we do the following in $G|_B$. Our argument in EC1 implies that in $G|_{B_1 \times B_2}$ we can either find a set S_1' of $h_{k-1} + h_k$ h_1-stars with roots in B_1, or a set S_2' of $h_{k-1} + h_k$ h_1-stars with roots in B_2, or a set S_1' of $\lfloor h_{k-1} + h_k/2 \rfloor$ h_1-stars with roots in B_1 *and* a set S_2' of $\lfloor h_{k-1} + h_k/2 \rfloor$ h_1-stars with roots in B_2. Again as in EC1 these roots will be used to adjust the sizes of B_1 and B_2. Now we remove $\sum_{i=1}^{m'} rs_i$ copies of H^k each containing exactly one root of a star in S_i where the first $l = k-2$ color-classes come from $A_i, 1 \leqslant i \leqslant l$ and the last two classes come from the B_i where the root was added. We also remove a copy of H^r containing H_i^r vertices from $A_i, 1 \leqslant i \leqslant l$. These copies of H^k and H^r are vertex disjoint from the stars in S_1' and S_2'. At this point as in EC1, we might have to move the exceptional vertices from one of the sets B_1, B_2 to the other. Denote the resulting sets by B_1' and B_2'. Note that $2h/|B_1'| + |B_2'|$.

Here H' is the vertex disjoint union of k bipartite graphs H_i' where H_i' has color-classes of sizes h_{i-1} and h_i ($h_0 = h_k$). Let $|B_1'| \equiv x_1 \pmod{2h}$ and $|B_2'| \equiv x_2 \pmod{2h}$. Thus $x_1 + x_2 = 0$ or $2h$. Let $0 \leqslant j < k$ be the largest integer for which

$$\sum_{i=1}^{j} (h_{i-1} + h_i) \leqslant x_1.$$

We have either

$$|S_1'| \geqslant x_1 - \sum_{i=1}^{j} (h_{i-1} + h_i),$$

or

$$|S_2'| \geqslant \sum_{i=1}^{j+1} (h_{i-1} + h_i) - x_1$$

(or may be both). We may assume that we have the first possibility. Then from B_1 we remove $x_1 - \sum_{i=1}^{j}(h_{i-1} + h_i)$ roots of stars in S_1' and we add them to B_2. Now it is not hard to see that we have an H'-factor in $G|_{B'}$ and then we can find the H-factor as above with the Blow-up Lemma. This finishes the extremal cases and the proof of Theorem 1.

7. Uncited reference

[14]

Acknowledgements

We thank Jiři Fiala and an anonymous referee for their very detailed and useful criticism.

References

[1] N. Alon, R. Yuster, Almost H-factors in dense graphs, Graphs Combin. 8 (1992) 95–102.
[2] N. Alon, R. Yuster, H-factors in dense graphs, J. Combin. Theory B 66 (1996) 269–282.
[3] B. Bollobás, Extremal Graph Theory, Academic Press, London, 1978.
[4] K. Corrádi, A. Hajnal, On the maximal number of independent circuits in a graph, Acta Math. Acad. Sci. Hungar 14 (1963) 423–439.
[5] G.A. Dirac, Some theorems on abstract graphs, Proc. London Math. Soc. 2 (1952) 68–81.
[6] P. Erdős, Some recent combinatorial problems, November 1990, preprint.

[7] A. Hajnal, E. Szemerédi, Proof of a conjecture of Erdős, in: P. Erdős, A. Rényi, V.T. Sós (Eds.), Combinatorial Theory and its Applications, Vol. II, Colloq. Math. Soc. J. Bolyai, Vol. 4, North-Holland, Amsterdam, 1970, pp. 601–623.
[8] J. Komlós, G.N. Sárközy, E. Szemerédi, Proof of a packing conjecture of Bollobás, Combin. Probab. Comput. 4 (1995) 241–255.
[9] J. Komlós, G.N. Sárközy, E. Szemerédi, On the square of a Hamiltonian cycle in dense graphs, Random Struct. Algorithms 9 (1996) 193–211.
[10] J. Komlós, G.N. Sárközy, E. Szemerédi, Blow-up Lemma, Combinatorica 17 (1997) 109–123.
[11] J. Komlós, G.N. Sárközy, E. Szemerédi, On the Pósa-Seymour conjecture, J. Graph Theory 29 (1998).
[12] J. Komlós, G.N. Sárközy, E. Szemerédi, An algorithmic version of the Blow-up Lemma, Random Struct. Algorithms 12 (1998) 297–312.
[13] J. Komlós, G.N. Sárközy, E. Szemerédi, Proof of the Seymour conjecture for large graphs, Ann. Combin. 2 (1998) 43–60.
[14] E. Szemerédi, Regular partitions of graphs, Colloques Internationaux C.N.R.S. No. 260 — Problèmes Combinatoires et Théorie des Graphes, Orsay, 1976, pp. 399–401.