Approximate Multipartite Version of the Hajnal–Szemerédi Theorem

Béla Csaba*
Department of Mathematics
Western Kentucky University

Marcelo Mydlarz†
Department of Computer Science
Rutgers University

Abstract

Let q be a positive integer, and G be a q-partite simple graph on qn vertices, with n vertices in each vertex class. Let $\delta = \frac{k}{k+1}$, where $k = q + O(\log q)$. If each vertex of G is adjacent to at least δn vertices in each of the other vertex classes, q is bounded and n is large enough, then G has a K_q-factor.

1 Introduction

In this paper we will consider simple graphs. We mostly use standard notation: we denote by $V(F)$ and $E(F)$ the vertex and the edge set of the graph F, $deg_F(x)$ is the degree of the vertex $x \in V(F)$ and $\delta(F)$ is the minimum degree of F. The number of vertices of F will be denoted by $v(F)$.

Let J be a fixed graph on q vertices. If q divides $|V(F)|$ and F has a subgraph which consists of $|V(F)|/q$ vertex-disjoint copies of J, then we say that F has a J-factor.

A fundamental result in extremal graph theory is the following theorem of Hajnal and Szemerédi [5]:

Theorem 1 (Hajnal and Szemerédi) *Let G be a graph on n vertices such that $\delta(G) \geq \frac{q-1}{q}n$. If q divides n, then G contains n/q vertex-disjoint cliques of size q.*

The theorem is obvious for $q = 2$; the first non-trivial case $q = 3$ was proved by K. Corrádi and A. Hajnal [2]. A. Hajnal and E. Szemerédi proved the theorem for arbitrary q in 1970 [5]. The proof was very complicated, and did not yield an effective algorithm.

We remark that they investigated an equivalent formulation. Let us call a proper q-coloring an *equitable q-coloring* if any two color classes differ in size by at most one. Hajnal and Szemerédi showed that $\overline{G}$, the complementer of G has an equitable coloring with n/q colors. Clearly, if $\overline{G}$ has an equitable n/q-coloring, then every color class is a clique in G. Almost four decades later Mydlarz and Szemerédi [15], and independently, Kierstead and Kostochka [8] found polynomial time algorithms for finding an equitable n/q-coloring.

*Part of this work was done while the author worked at the Analysis and Stochastics Research Group at the University of Szeged. Partially supported by OTKA T049398. e-mail: bela.csaba@wku.edu

†e-mail: marcem@cs.rutgers.edu

Recently, a $O(n^3/q)$ time algorithm was published by Kierstead, Kostochka, Mydlarz and Szemerédi in [9].

We say that F is *q-partite*, if its vertex set can be divided into q classes which are independent sets. F is a *balanced* q-partite graph, if these vertex classes are of the same size. Let F be a q-partite graph with vertex classes $A_1, A_2, \ldots, A_q$. We define the *proportional minimum degree* of F by

$$\widetilde{\delta}(F) = \min_{1\leq i\leq q} \min_{v\in A_i} \left\{ \frac{deg(v, A_j)}{|A_j|} : j \neq i \right\}.$$

E. Fischer [4] considered several variants of the Hajnal-Szemerédi theorem and proposed to investigate a q-partite version of Theorem 1. He showed that if $\widetilde{\delta}(G) \geq 1 - 1/2(q-1)$ then G has a K_q-factor, and conjectured that the true bound is $\widetilde{\delta}(G) \geq 1 - 1/q$. However, this fails as the following construction by Cs. Magyar and R. Martin [14] shows. Let Γ_q be a balanced q-partite graph with vertex set $\{h_{i,j} : \ i = 1, 2, \ldots, q; \ j = 1, 2, \ldots, q\}$. The adjacency rules are as follows: $h_{i,j}h_{i',j'} \in E(\Gamma_q)$ if $i \neq i'$, $j \neq j'$ and either j or j' is in $\{1, \ldots, q-2\}$. Also, $h_{i,q}h_{i',q} \in E(\Gamma_q)$ for $i \neq i'$. No other edges exist. It is easy to see that the proportional minimum degree of Γ_q is $1 - 1/q$. If q is even then Γ_q can be covered by disjoint copies of K_qs, but it cannot if q is odd.

The conjecture below contains a small correction, an additive term that is necessary for odd values of q.

Conjecture 2 *Let G be a balanced q-partite graph on qn vertices. There exists a constant $K \geq 0$ such that if $\widetilde{\delta}(G) \cdot n \geq \frac{q-1}{q}n + K$, then G contains n vertex-disjoint cliques of size q.*

The conjecture is easily seen to hold for $q = 2$. It was shown for $q = 3$ by Cs. Magyar and R. Martin [14], and for $q = 4$ by R. Martin and E. Szemerédi [15]. The proofs of these latter cases are very involved. We remark that R. Johansson [7] proved the $q = 3$ case approximately. Also, A. Johansson, R. Johansson and K. Markström [6] considered finding a K_q-factor in balanced q-partite graphs conditioning on the usual minimum degree.

In this paper we show a relaxed version of Conjecture 2. For q being a natural number let h_q denote the qth harmonic number, that is, $h_q = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{q}$.

Theorem 3 *Let $q \geq 2$ be an integer and $k_q = q - 3/2 + h_{q-1}/2$. Then there exists an n_0 such that if $n > n_0$, G is a balanced q-partite graph on qn vertices, and $\widetilde{\delta}(G) \geq \frac{k_q}{k_q+1}$, then G has a K_q-factor.*

Notice, that $\frac{k_q}{k_q+1} - \frac{q-1}{q} = O(\log q/q^2)$, that is, the bound in the above theorem is a close estimation for the conjectured bound. In particular, for every $\varepsilon > 0$ there exists a q_0 such that if $q \geq q_0$ then $(1+\varepsilon)\frac{q-1}{q} > \frac{k_q}{k_q+1}$. We also have the following corollary of Theorem 3:

Corollary 4 *Let G be as above. Assume that H is a fixed graph such that $\chi(H) \leq q$. If $v(H)$ divides n, then G has an H-factor.*

For proving Theorem 3 and Corollary 4 our main tools will be the Regularity Lemma of E. Szemerédi [17], and the Blow-up Lemma by J. Komlós, G. Sárközy and E. Szemerédi [11, 12]. We will give a brief survey on the necessary notions in the second section.

2 Main tools for the proof

We introduce some more notation first. For any vertex v of the graph G, $deg_G(v, X)$ is the number of neighbors of v in the set X, and $e(X, Y)$ is the number of edges between the disjoint sets X and Y. $N_G(v)$ is the set of neighbors of v and $N_G(v, X)$ is the set of neighbors of v in X. For a set $S \subset V(G)$, $N(S) = \cup_{v \in S} N(v)$. We let $G|_S$ denote the subgraph of G induced by the set S.

If k is a natural number, and every vertex in the graph G has degree k, then we call the graph *k-regular*. For a real $\varepsilon \in (0, 1)$ we also will consider *ε-regular pairs*. Regular pairs play a crucial role in the Regularity Lemma of Szemerédi (more details follow later).

Let F be a multipartite graph. Given certain vertex classes $A_{i_1}, \ldots, A_{i_s}$ we will denote the s-partite subgraph of F spanned by these classes by $F(A_{i_1}, \ldots, A_{i_s})$. Throughout the paper we will apply the relation "$\ll$": $a \ll b$, if a is sufficiently smaller than b.

2.1 Factors of bipartite graphs

Let F be a bipartite graph with color classes A and B. By the well-known König–Hall theorem there is a perfect matching in F if and only if $|N(S)| \geq |S|$ for every $S \subset A$. The following, while simple, is a very useful consequence of this result, we record it here for future reference.

Lemma 5 *If F is a balanced bipartite graph on $2n$ vertices, and $deg(x) \geq n/2$ for every $x \in V(F)$, then there is a perfect matching in F.*

Notice, that Lemma 5 is precisely Conjecture 2 (and Theorem 3) in the case when $q = 2$.

If $f : V(F) \to \mathbf{N}$ is a function, then an f-factor is a subgraph F' of F such that $deg_{F'}(v) = f(v)$ for every $v \in V$. We will need special f-factors, namely when $f \equiv r$ for some $r \in \mathbf{N}$. Then F' is an r-regular subgraph of F. For $x \in (0, 1)$ we let $\rho(x) = \frac{x + \sqrt{2x-1}}{2}$. If the minimum degree of F is large enough, then one can find a sufficiently dense spanning regular subgraph (see [3] by Csaba):

Theorem 6 *Let $F(A, B)$ be a balanced bipartite graph on $2n$ vertices, and assume that $\delta = \delta(F)/n \geq 1/2$. Then F has an s–regular spanning subgraph for all $0 \leq s \leq \lfloor \rho(\delta) n \rfloor$.*

2.2 Regularity Lemma

The *density* between disjoint sets X and Y is defined as:

$$d(X, Y) = \frac{e(X, Y)}{|X||Y|}.$$

In the proof of Theorem 3, Szemerédi's Regularity Lemma [17, 13] plays a pivotal role. We will need the following definition to state the Regularity Lemma.

Definition 1 (Regularity condition) *Let $\varepsilon > 0$. A pair (A, B) of disjoint vertex-sets in G is ε-regular if for every $X \subset A$ and $Y \subset B$, satisfying*

$$|X| > \varepsilon|A|, \; |Y| > \varepsilon|B|$$

we have

$$|d(X,Y) - d(A,B)| < \varepsilon.$$

This definition implies that ε-regular pairs are highly uniform bipartite graphs; namely, the density of any reasonably large subgraph is almost the same as the density of the regular pair.

We will use the following form of the Regularity Lemma:

Lemma 7 (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 $\ell + 1$ clusters $W_0, W_1, \ldots, W_\ell$, and there is a subgraph G' of G with the following properties:*

- $\ell \leq M$,
- $|W_0| \leq \varepsilon |V|$,
- *all clusters W_i, $i \geq 1$, are of the same size m $\left(\leq \lfloor \frac{|V|}{\ell} \rfloor < \varepsilon |V|\right)$,*
- *$deg_{G'}(v) > deg_G(v) - (d+\varepsilon)|V|$ for all $v \in V$,*
- *$G'|_{W_i} = \emptyset$ (W_i is an independent set in G') for all $i \geq 1$,*
- *all pairs (W_i, W_j), $1 \leq i < j \leq \ell$, are ε-regular, each with density either 0 or greater than d in G'.*

Often we call W_0 the *exceptional cluster*. In the rest of the paper we will assume that $0 < \varepsilon \ll d \ll 1$.

Definition 2 (Reduced graph) *Apply Lemma 7 to the graph $G = (V,E)$ with parameters ε and d, and denote the clusters of the resulting partition by $W_0, W_1, \ldots, W_\ell$, W_0 being the exceptional cluster. We construct a new graph G_r, the reduced graph of G' in the following way: The non-exceptional clusters of G' are the vertices of the reduced graph G_r (hence $|V(G_r)| = \ell$). We connect two vertices of G_r by an edge if the corresponding two clusters form an ε-regular pair with density at least d.*

The following corollary is immediate:

Corollary 8 *Apply Lemma 7 with parameters ε and d to the graph $G = (V,E)$ satisfying $\delta(G) \geq \gamma n$ $(|V| = n)$ for some $\gamma > 0$. Let G_r denote the reduced graph of G'. Then $\delta(G_r) \geq (\gamma - \theta)\ell$, where $\theta = 2\varepsilon + d$.*

Next we show that the property of being balanced can be inherited by the reduced graph. Besides, one can avoid to have *mixed clusters*, that is, no cluster will contain vertices from more than one of the vertex classes of G.

Lemma 9 *Let G be a balanced q-partite graph with $\delta(G) \geq \gamma n$ for some $\gamma > 0$. Then applying the Regularity Lemma we obtain a reduced graph G_r of G which can be modified so that G_r is a balanced q-partite graph. Moreover, the modified G_r has no mixed clusters.*

Proof: First we show how to avoid having mixed clusters. Let W_i and W_j be any two clusters. If W_i has more than εm vertices from at least two different vertex classes then (W_i, W_j) cannot be an ε-regular pair by the definition. Since every cluster in G_r has some neighbors by the minimum degree condition of G, we get that in every cluster at least $(1-\varepsilon)m$ vertices belong to the same vertex class. This in turn implies that by deleting εn vertices from the clusters we arrive at a reduced graph in which every cluster is a subset of some vertex class of G, and all have the same size. We denote the common cluster size by m. Hence, the vertex classes of G naturally determine the classes of G_r.

It is still possible that G_r has two classes with different number of clusters. However, initially the exceptional cluster W_0 had at most εn vertices. Since the clusters have the same size m, we conclude that the difference between any two cluster classes can be at most $\varepsilon\ell$ clusters. Hence, after discarding a total of at most $q\varepsilon\ell$ clusters we get a balanced reduced graph G_r. □

During this procedure we may decrease the minimum degree of G_r, however, no cluster loses more than $q\varepsilon\ell$ neighbors. Therefore, we get the following lemma, the proof is implied by Lemma 8 and Lemma 9.

Lemma 10 *Applying the Regularity Lemma and Lemma 9 for a balanced q-partite graph G with $\widetilde{\delta}(G) \geq k_q/(k_q+1)$ we get a balanced q-partite reduced graph G_r with $\widetilde{\delta}(G_r) \geq k_q/(k_q+1) - (2+q)\varepsilon - d$.*

Given an ε-regular pair (A, B), we may increase A and B by adding some new vertices to both. We expect that after this procedure the new pair will be η-regular for some small η, although $\eta > \varepsilon$.

Lemma 11 *Assume that $0 < 2\varepsilon^{1/6} < 1/K$. Let (A, B) be an ε-regular pair with $m = |A| = |B|$, and add $K\varepsilon m$ vertices to A and $K\varepsilon m$ vertices to B, thereby obtaining the sets $\widetilde{A}$ and $\widetilde{B}$, respectively. Then the resulting new pair $(\widetilde{A}, \widetilde{B})$ is $2\varepsilon^{1/3}$-regular with density at least $d - 2\varepsilon^{1/3}$.*

Proof: Let $A' \subset \widetilde{A}$ and $B' \subset \widetilde{B}$ such that $|A'|, |B'| = \sqrt{\varepsilon}m$. By Convexity of Density it is sufficient to verify the regularity condition for subsets of this size (for details see [13]). Then

$$(d-\varepsilon)|A'| \cdot |B'| - 2K\varepsilon^{3/2}m^2 \leq e(A', B') \leq (d+\varepsilon)|A'| \cdot |B'| + 2K\varepsilon^{3/2}m^2$$

by the ε-regularity of the original pair. Hence,

$$d - \varepsilon - 2K\sqrt{\varepsilon} \leq d(A', B') \leq d + \varepsilon + 2K\sqrt{\varepsilon}.$$

Since $K < \varepsilon^{-1/6}/2$, this implies that

$$d - \varepsilon - \varepsilon^{1/3} \leq d(A', B') \leq d + \varepsilon + \varepsilon^{1/3}.$$

It is easy to see that $d - \varepsilon^{5/6} \leq d(\widetilde{A}, \widetilde{B}) \leq d + \varepsilon^{5/6}$, since K is not very large. Therefore

$$|d(\widetilde{A}, \widetilde{B}) - d(A', B')| \leq 2\varepsilon^{1/3},$$

which proves the lemma. □

We will need the following Slicing Lemma from [13]:

Lemma 12 (Slicing Lemma) *Let (A, B) be an ε-regular–pair with density d for some $\epsilon > 0$. We arbitrarily halve A and B, getting the sets A', A'' and B', B'', respectively. Then the following holds: (A', B') and (A'', B'') are 2ε–regular pairs with density at least $d - \varepsilon$.*

A stronger one-sided property of regular pairs is super-regularity:

Definition 3 (Super-Regularity condition) *Given a graph G and two disjoint subsets $A, B \subset V(G)$, the pair (A, B) is* (ε, δ)-super-regular*, if it is ε-regular and furthermore,*

$$deg(a) > \delta|B|, \text{ for all } a \in A,$$

and

$$deg(b) > \delta|A|, \text{ for all } b \in B.$$

Let $\varepsilon > 0$ and assume that the pair (A, B) is ε-regular with density d. Mark those vertices of A which have less than $(d - \varepsilon)|B|$ neighbors and those which have more than $(d + \varepsilon)|B|$. By the definition of ε-regularity, there can be at most $2\varepsilon|A|$ marked vertices in A. Repeat the same procedure for B so as to mark those vertices which have too many or too few neighbors in A. If we get rid of the marked vertices of A and B then we will have a $(3\varepsilon, d - 3\varepsilon)$-super-regular pair (A', B'). That is, we proved that every regular pair contains a large super-regular pair:

Lemma 13 *Let (A, B) be an ε-regular pair with density d. Then it has a $(3\varepsilon, d-3\varepsilon)$-super-regular subpair (A', B') where $A' \subset A$, $|A'| = |A| - \lceil 2\varepsilon|A| \rceil$ and $B' \subset B$, $|B'| = |B| - \lceil 2\varepsilon|B| \rceil$.*

We will repeatedly make use of the following result, which states that random sub-pairs of (ε, δ)-super-regular pairs are likely to be super-regular, with somewhat weaker parameters.

Lemma 14 *Let (A, B) be an (ε, δ)-super-regular pair with density d and k be a positive integer. Assume that $|A| = |B| = m$, and $k|m$. Divide A and B into k random subsets: $A = A_1 \cup A_2 \cup \ldots \cup A_k$ and $B = B_1 \cup B_2 \cup \ldots \cup B_k$, each having size m/k. Then with probability tending to one as m tends to infinity we have that (A_i, B_j) is an (ε', δ')-super-regular pair with density d' for every $1 \le i, j \le k$, where $\varepsilon' = 2\varepsilon^{1/5}$, $\delta - \varepsilon' \le \delta'$ and $d - \varepsilon' \le d'$.*

Proof: The proof follows from a theorem of Y. Kohayakawa and V. Rödl [10]. They showed that two *local conditions* imply η-regularity. Namely, if most of the degrees and co-degrees are close to the average in a pair, then the pair is η-regular with a small η. More precisely, let the density of the (X, Y) pair be d. Let D be the collection of all pairs $\{v, w\}$ of X such that $deg(v), deg(w) \ge (d - \eta)|Y|$ and $deg(v, N(w)) \le (d + \eta)^2|Y|$. If $|D| > (1 - 5\eta)|X|^2/2$ then (X, Y) is $(16\eta)^{1/5}$-regular. It is easy to see that if (A, B) is an ε-regular pair, then it satisfies these local conditions with $\eta = \varepsilon$. Moreover, if we split the pair randomly into sub-pairs then with probability at least $1 - 1/n$ we get that the local conditions are satisfied with $\eta = 2\varepsilon$ for all sub-pairs. This can be shown using Azuma's inequality (see Alon and Spencer's book [1]). Hence, with high probability all the sub-pairs will be $(32\varepsilon)^{1/5}$-regular. Since the individual degrees do not decrease much with high probability, we also have that $\delta' \ge \delta - \varepsilon'$ and $d' \ge d - \varepsilon'$. $\square$

Remark 1 *When we apply Lemma 14 we may have to discard at most $k-1$ vertices from a cluster in order to satisfy the divisibility condition of the lemma. Since we use Lemma 14 at most $q-2$ times during the embedding algorithm, putting the discarded vertices into W_0 will not increase the size of the exceptional cluster substantially. Hence, we will assume that the divisibility condition is satisfied whenever we apply Lemma 14.*

Let G_r be the reduced graph of the graph G such that edges in G_r represent ε-regular pairs with density at least d. Assume that $\widehat{G}_r$ is a cluster graph which we get by randomly splitting the clusters of G_r into sub-clusters of equal size. The new sub-clusters will be called *split copies* of the original cluster, and we will use "$\widehat{\ }$" to indicate that we refer to a split copy.

Two split copies will be connected if they form an ε'-regular pair with density d' where $\varepsilon' \leq 2\varepsilon^{1/5}$ and $d' \geq d - \varepsilon'$. By the previous lemma if $W_iW_j \in E(G_r)$ and $\widehat{W}_i, \widehat{W}_j$ arose from W_i and W_j by the random splitting, then $\widehat{W}_i\widehat{W}_j \in E(\widehat{G}_r)$. We will call $\widehat{G}_r$ the *refinement* of G_r. Notice that $\widetilde{\delta}(\widehat{G}_r) \geq \widetilde{\delta}(G_r)$.

2.3 Blow-up Lemma

Let H and G be two graphs on n vertices. Assume that we want to find an isomorphic copy of H in G. In order to achieve this one can apply a very powerful tool, the Blow-up Lemma of Komlós, Sárközy and Szemerédi [11, 12].

Theorem 15 (Blow-up Lemma) *Given a graph R of order r and positive integers δ, Δ, there exists a positive $\varepsilon = \varepsilon(\delta, \Delta, r)$ such that the following holds: Let $n_1, n_2, \dots, n_r$ be arbitrary positive parameters and let us replace the vertices $v_1, v_2, \dots, v_r$ of R with pairwise disjoint sets $W_1, W_2, \dots, W_r$ of sizes $n_1, n_2, \dots, n_r$ (blowing up R). We construct two graphs on the same vertex set $V = \cup_i W_i$. The first graph F is obtained by replacing each edge $v_iv_j \in E(R)$ with the complete bipartite graph between W_i and W_j. A sparser graph G is constructed by replacing each edge v_iv_j arbitrarily with an (ε, δ)-super-regular pair between W_i and W_j. If a graph H with $\Delta(H) \leq \Delta$ is embeddable into F then it is already embeddable into G.*

3 Outline of the embedding algorithm

Since a K_q-factor is a subgraph, finding such a factor will be considered an embedding problem. Let us denote the union of n vertex-disjoint copies of K_qs by H. We will show Theorem 3 by exhibiting a randomized algorithm which will embed H into G with high probability.

The algorithm will proceed in two stages. The main goal of Stage 1 is to find a K_q-factor in a reduced graph of G such that the vast majority of the vertices of G will be in some clique and the cluster sizes in the cliques are approximately equal. In Stage 2 we will achieve that all vertices of G will be in some non-exceptional cluster (that is, W_0 will be emptied) and the edges of the cliques in the factor will be super-regular. Furthermore, if C is a clique in the factor, then the clusters of C will have the same size. Then we will finish the embedding with the help of the Blow-up Lemma. We remark that the second stage is a technically somewhat challenging part, however, this stage has become routine in proofs of these types of embedding theorems.

Stage 1

- Apply the Regularity Lemma to G with appropriately chosen parameters $0 < \varepsilon \ll d \ll 1$. We obtain the balanced q-partite reduced graph G_r. The cluster classes of G_r are denoted by $A_1, A_2, \ldots, A_q$, here $|A_1| = |A_2| = \ldots = |A_q| = \ell$.
- Apply the Factor Finder Algorithm in order to construct a K_q-factor in a refinement of G_r. The cliques in the factor will contain most of the vertices of G.

Stage 2

- Put some vertices into the exceptional cluster W_0 in order to achieve that all edges in the cliques represent super-regular pairs.
- Distribute the vertices of W_0 among the non-exceptional clusters while maintaining super-regularity in the cliques.
- Move vertices between clusters in order to achieve that the clusters in cliques have equal sizes while maintaining super-regularity in the cliques.
- Apply the Blow-up Lemma in order to finish the embedding.

4 The first stage of the embedding algorithm

Given the graph G, we apply the Degree Form of the Regularity Lemma with parameters ε and d such that $0 < \varepsilon \ll d \ll 1$. Then we find the reduced graph G_r. By Lemma 9 and Lemma 10 we may assume that G_r is a balanced q-partite reduced graph on $q\ell$ vertices with $\widetilde{\delta}(G_r) \geq k_q/(k_q+1) - (2+q)\varepsilon - d \geq k_q/(k_q+1) - qd$ where $k_q = q - 3/2 + h_{q-1}/2$. It turns out that with a proportional minimum degree this large we will have room to spare if $q \geq 3$.

In what follows we will denote $k_q/(k_q+1)$ by $\widetilde{\delta}$, and the cluster classes of G_r will be denoted by $A_1, A_2, \ldots, A_q$. Recall that our goal is to show that $H \subset G$, where H is the disjoint union of n copies of K_qs.

4.1 Outline of the Factor Finder Algorithm

The Factor Finder algorithm is a recursive algorithm with base case $q = 2$. The algorithm works on G_r, and finds a K_q-factor in a refinement of G_r. Below we give a brief outline of the method, without elaborating the technical details.

If $q = 2$ it is an easy exercise to find the K_2-factor (a perfect matching) in G. Thus, we will focus on the cases when $q \geq 3$. For an easier understanding we begin with the outline of the method for the case $q = 3$. Let U be an arbitrary cluster of G_r. Observe that every edge in the neighborhood of U gives a triangle with U. This simple observation helps us finding several triangles in the reduced graph if the proportional minimum degree is larger than $1/2$. We will show that if the proportional minimum degree of G_r is at least 0.68, then it is possible to assign exactly 0.64ℓ edges to every cluster of A_1. This gives $0.64\ell^2$ triangles in G_r. Moreover, every cluster of G_r will appear in exactly 0.64ℓ triangles. Then we split every cluster randomly into 0.64ℓ sub-clusters, and these sub-clusters will be grouped into $0.64\ell^2$ disjoint triangles. Hence, we can find a triangle factor in a refinement of G_r. Notice that by Lemma 14 we may assume that the edges of the triangles will represent regular pairs with somewhat weaker parameters than the edges of G_r.

The algorithm works recursively for larger values of q. Given a cluster $U \in A_1$ we will look for a K_{q-1}-factor in the neighborhood of U. These $(q-1)$-cliques can be found using the Factor Finder Algorithm. Say, that we assign s cliques to U this way. These cliques are made of sub-clusters in $A_2 \cup \ldots \cup A_q$. In fact, these sub-clusters will have size m/s where m is the size of a cluster in G_r. The algorithm will randomly split U into s sub-clusters, and find a one-to-one mapping between the sub-clusters of U and the $(q-1)$-cliques in its neighborhood. The technical details will follow in the next subsection.

4.2 The Factor Finder Algorithm

The first case: $q = 2$

First, notice that in this case $k_q = 1$, therefore, $\widetilde{\delta} = 1/2$. It is easy to find a K_2-factor (a perfect matching) in a balanced bipartite graph G_r with a proportional minimum degree this large. This case is settled by Lemma 5. We further look at this case in order to show how to include some 'randomly chosen' edges into the perfect matching when the proportional minimum degree is larger than $1/2$.

Assume that the proportional minimum degree is $1/2 + \psi$ for some $0 < \psi < 1$. One can introduce some randomness in finding the perfect matching as follows. Pick $\psi\ell/2$ clusters randomly from the first vertex class, and find neighbors for them randomly. Then pick $\psi\ell/2$ clusters randomly from the other vertex class, and find neighbors for them randomly. This way we have found random neighbors for $\psi\ell$ clusters. In the remaining clusters the minimum degree is sufficiently large for having a perfect matching. Therefore, we can find a perfect matching in such a way that $\psi\ell$ clusters have randomly chosen neighbors.

As it turns out later on, this small extra randomness will be very helpful. When finishing the embedding of H we will need the proportional minimum degree to be a bit larger than $1/2$ in order to perform the above procedure. But that will be provided for $q \geq 3$ (recall that we use recursion).

Finding a triangle factor

As a warm-up we discuss this case in detail. First, apply Theorem 6 for the graphs $G_r(A_1, A_2)$ and $G_r(A_1, A_3)$. We get two μ-regular bipartite graphs $R(A_1, A_2)$ and $R(A_1, A_3)$, with $\mu = \rho(\widetilde{\delta}(G_r))\ell$. We define R to be the 3-partite graph on $A_1 \cup A_2 \cup A_3$ such that $E(R) = E(R(A_1, A_2)) \cup E(R(A_1, A_3))$. It is easy to see that $deg_R(W) = \mu$ for every $W \in A_2 \cup A_3$.

We are going to cut the clusters of $A_2 \cup A_3$ randomly into μ sub-clusters of equal size. The new cluster classes are denoted by $\widehat{A}_2$ and $\widehat{A}_3$. Roughly speaking, we will assign the split copies of $\widehat{A}_2 \cup \widehat{A}_3$ to the clusters of A_1 such that every cluster of A_1 will receive 2μ split copies, and every split copy will be assigned to exactly one cluster in A_1.

More formally, let us define a surjective function σ: its domain is the set of split copies, and its range is A_1. It satisfies the following requirements: whenever $U \in A_2 \cup A_3$, and $\widehat{U}$

is a split copy of U, then $\sigma(\widehat{U}) \in N_R(U)$, moreover, if $\widehat{U}$ and $\widehat{U}'$ are different split copies of U, then $\sigma(\widehat{U}) \neq \sigma(\widehat{U}')$. For every $W \in A_1$ we introduce two sets associated with it:

$$N_2(W) = \{\widehat{U} : \widehat{U} \in \widehat{A}_2,\ \sigma(\widehat{U}) = W\},$$

and

$$N_3(W) = \{\widehat{U} : \widehat{U} \in \widehat{A}_3,\ \sigma(\widehat{U}) = W\}.$$

It is easy to see that every cluster of $\widehat{A}_2 \cup \widehat{A}_3$ will participate in one of the $N_i(W)$ sets, and $|N_i(W)| = \mu$ for $i = 2, 3$ and every $W \in A_1$.

Our next goal is to show that $\widehat{G}_r(N_2(W), N_3(W))$, the induced subgraph of the refinement of G_r on $N_2(W)$ and $N_3(W)$, has a perfect matching $M(W)$ for every $W \in A_1$. Having this perfect matching in hand we can construct μ triangles for every $W \in A_1$: cut the clusters of A_1 randomly into μ sub-clusters, and assign the split copies of W to the edges of $M(W)$ bijectively. This way we construct triangles each having cluster size m/μ.

Hence, what is left: for every $W \in A_1$ find the perfect matchings in the bipartite subgraphs $\widehat{G}_r(N_2(W), N_3(W))$. We claim that the minimum degree in these bipartite graphs is in fact sufficiently large to guarantee the existence of a perfect matching. For that we will show that every cluster is adjacent to at least half of the clusters in the other class. We use a simple claim which we record here for future purposes, the proof is left for the reader.

Claim 16 *Let $F = (V, E)$ be a graph and let $S \subset V$. Then every $u \in V$ is adjacent to at least a $\frac{\delta(F)-(|V|-|S|)}{|S|}$ proportion of the vertices of S.*

We use Claim 16 in order to guarantee that every cluster of $N_2(W)$ is adjacent to at least half of the clusters of $N_3(W)$. We will also have that every cluster $N_3(W)$ is adjacent to at least half of the clusters of $N_2(W)$.

Let $\widehat{U} \in N_2(W)$ be an arbitrary cluster. By Claim 16, $\widehat{U}$ is adjacent to at least a $(\widetilde{\delta}(G_r) - (1 - \mu/\ell))\ell/\mu$ proportion of the vertices of $N_3(W)$. Similarly, every $\widehat{U} \in N_3(W)$ is adjacent to at least a $(\widetilde{\delta}(G_r) - (1 - \mu/\ell))\ell/\mu$ proportion of $N_2(W)$. Easy calculation shows that if $\widetilde{\delta}(G_r) = 0.68$, then $\mu/\ell = \rho(\widetilde{\delta}(G_r)) = 0.64$, and

$$\widetilde{\delta}(\widehat{G}_r(N_2(W), N_3(W))) \geq \frac{(\widetilde{\delta}(G_r) - (1 - \mu/\ell))\ell}{\mu} = 0.5.$$

This implies the existence of a perfect matching in $\widehat{G}_r(N_2(W), N_3(W))$ if $\widetilde{\delta}(G_r) \geq 0.68$. Notice that $0.68 < 0.69 < \frac{3-3/2+h_2/2}{3-3/2+h_2/2+1} - qd$ if d and ε are sufficiently small. That is, there exists a real number $\gamma_3 > 0$ such that if $\widetilde{\delta}(\widehat{G}_r) \geq \frac{k_3}{k_3+1} - \gamma_3$ then $\widehat{G}_r$ has a triangle factor. Hence, we can find a triangle factor in $\widehat{G}_r$ with a smaller bound that is required by Theorem 3.

This latter fact will be important for us later on. Recall the discussion of case $q = 2$ for finding the perfect matching with some randomly chosen edges. Clearly, for $k_3 = 3/2 + h_2/2$ the proportional minimum degree will be larger than $1/2$ when it comes to finding the perfect matchings in the $\widehat{G}_r(N_2(W), N_3(W))$ graphs. Hence, we can perform the randomized procedure for finding the perfect matchings.

We remark, that the cluster size in $\widehat{G}_r$ is $\frac{m}{\mu}$, and the number of clusters in each vertex class is $\mu\ell = \rho(\widetilde{\delta}(G_r))\ell^2$.

The general case

Assume now that $q > 3$. Assume further that there exists a constant $\gamma_{q-1} > 0$ such that if the proportional minimum degree in a balanced $(q-1)$-partite cluster graph F is at least $k_{q-1}/(k_{q-1}+1) - \gamma_{q-1}$, then some refinement $\widehat{F}$ has a K_{q-1}-factor. We are given G_r, a balanced q-partite graph with vertex classes $A_1, A_2, \ldots, A_q$ such that $\widetilde{\delta}(G_r) \geq k_q/(k_q+1) - qd$. This time our goal will be to find a K_q-factor in a refinement $\widehat{G}_r$.

Set $\mu = \rho(\widetilde{\delta}(G_r))\ell$. We consider the bipartite subgraphs $G_r(A_1, A_i)$ and apply Theorem 6 to get the μ-regular bipartite graphs $R(A_1, A_i)$ for every $2 \leq i \leq q$. Let R be a q-partite graph such that $V(R) = \cup_{i\geq 1} A_i$ and $E(R) = \cup_{i\geq 2} E(R(A_1, A_i))$. As before, $deg_R(U) = \mu$ where $U \in A_2 \cup \ldots \cup A_q$.

Similarly to the case $q = 3$ we randomly split every cluster in $A_2 \cup A_3 \cup \ldots \cup A_q$ into μ sub-clusters of equal size thereby getting $\widehat{A}_i$ from A_i for $2 \leq i \leq q$.

We define a surjective function σ: its domain is the set of split copies, and its range is A_1. It satisfies the following requirements: whenever $U \in A_2 \cup \ldots \cup A_q$, and $\widehat{U}$ is a split copy of U, then $\sigma(\widehat{U}) \in N_R(U)$, moreover, if $\widehat{U}$ and $\widehat{U}'$ are different split copies of U, then $\sigma(\widehat{U}) \neq \sigma(\widehat{U}')$. For every $W \in A_1$ we introduce $q-1$ sets associated with it:

$$N_i(W) = \{\widehat{U} : \widehat{U} \in \widehat{A}_i,\ \sigma(\widehat{U}) = W\}$$

for $2 \leq i \leq q$. It is easy to see, that every cluster of $\widehat{A}_2 \cup \ldots \cup \widehat{A}_q$ will participate in one of the $N_i(W)$ sets, and $|N_i(W)| = \mu$ for every $2 \leq i \leq q$ and every $W \in A_1$.

Let us consider the balanced $(q-1)$-partite graphs $\widehat{G}_r(N_2(W), \ldots, N_q(W))$ for every $W \in A_1$. As before, we can give a lower bound on the proportional minimum degree in these graphs with the help of Claim 16:

$$\widetilde{\delta}(\widehat{G}_r(N_2(W), \ldots, N_q(W))) \geq \frac{(\widetilde{\delta}(G_r) - (1 - \mu_q/\ell))\ell}{\mu_q}.$$

In case $q = 3$ we had to check whether this quantity was at least $1/2$, this time we have to check that this number is sufficiently large so as to guarantee the existence of a K_{q-1}-factor in these graphs.

Say, that we can find a K_{q-1}-factor $M(W)$ for every $W \in A_1$. Then we construct the desired K_q-factor the following way: cut the clusters of A_1 randomly into $|M(W)|$ sub-clusters, and assign the split copies of W to the $(q-1)$-cliques of $M(W)$ bijectively. This way we get $|M(W)|$ cliques of size q. In Lemma 19 we show that every cluster of $\widehat{G}_r$ will have size $m/|M(W)|$.

It is useful to introduce a new function in order to show that $\widetilde{\delta}(\widehat{G}_r(N_2(W), \ldots, N_q(W)))$ is sufficiently large for finding a K_{q-1}-factor. Let

$$\Phi(x) = \frac{x - (1 - \rho(x))}{\rho(x)}$$

for $x \in (0.5, 1)$. It is easy to see that $\Phi(x)$ is continuous in its domain.

In Lemma 17 below we take a step towards proving that $\widetilde{\delta}(G_r) \geq \frac{k_q}{k_q+1} - qd$ is sufficiently large for finding the clique factor using the $\Phi(x)$ function.

Lemma 17 *If $q \geq 3$ then*

$$\Phi\left(\frac{k_q}{k_q+1}\right) \geq \frac{k_{q-1}}{k_{q-1}+1} + \frac{1}{100k_q^2(k_q+\sqrt{k_q^2-1})}.$$

Proof: Using the definition of $\Phi(x)$ we get that

$$\Phi\left(\frac{k_q}{k_q+1}\right) = \frac{\frac{k_q}{k_q+1} - (1-\rho(\frac{k_q}{k_q+1}))}{\rho(\frac{k_q}{k_q+1})}.$$

Since

$$\rho\left(\frac{k_q}{k_q+1}\right) = \frac{\frac{k_q}{k_q+1} + \sqrt{\frac{k_q-1}{k_q+1}}}{2},$$

we get that

$$\Phi\left(\frac{k_q}{k_q+1}\right) = 2\frac{\frac{k_q-2}{2(k_q+1)} + \frac{1}{2}\sqrt{\frac{k_q-1}{k_q+1}}}{\frac{k_q}{k_q+1} + \sqrt{\frac{k_q-1}{k_q+1}}} = 1 - \frac{\frac{2}{k_q+1}}{\frac{k_q}{k_q+1} + \sqrt{\frac{k_q-1}{k_q+1}}} = 1 - \frac{2}{k_q+\sqrt{k_q^2-1}}.$$

We will show that

$$1 - \frac{2}{k_q+\sqrt{k_q^2-1}} > \frac{k_{q-1}}{k_{q-1}+1} + \frac{1}{100k_q^2(k_q+\sqrt{k_q^2-1})}$$

is a valid inequality. Equivalently, we claim that

$$\frac{1}{k_{q-1}+1} > \frac{2}{k_q+\sqrt{k_q^2-1}} + \frac{1}{100k_q^2(k_q+\sqrt{k_q^2-1})}.$$

Multiplying by $100k_q^2(k_q+\sqrt{k_q^2-1})(k_{q-1}+1)$ we get

$$100k_q^2\left(k_q+\sqrt{k_q^2-1}\right) > (200k_q^2+1)(k_{q-1}+1).$$

Using that $k_q = k_{q-1}+1+1/(2q-2)$, reordering and canceling terms, we get that

$$100k_q^2\sqrt{k_q^2-1} + \frac{100k_q^2}{q-1} > 100k_q^3 + k_q - \frac{1}{2(q-1)}.$$

We make the inequality even stronger by discarding the term '$-1/(2q-2)$' from the right hand side. Then dividing by k_q and taking the square of both sides gives

$$\left(\frac{100k_q}{q-1}\right)^2 + 2(100k_q)^2\frac{\sqrt{k_q^2-1}}{q-1} + (100k_q)^2(k_q^2-1) > (100k_q^2)^2 + 200k_q^2 + 1.$$

It is easy to see that $\sqrt{k_q^2-1} > q-1$ for $q \geq 3$, and that $(100k_q/(q-1))^2 > 0$. These imply the inequality above, which in turns proves the lemma. $\square$

Using the continuity of $\Phi(x)$ and Lemma 17, we get that for every $q \geq 3$ there is a $\gamma_q > 0$ real number such that

$$\Phi\left(\frac{k_q}{k_q+1} - \gamma_q\right) \geq \frac{k_{q-1}}{k_{q-1}+1}.$$

Choose ε and d such that $0 < \varepsilon \ll d$ and $qd \leq \gamma_q$. By Lemma 10 the proportional minimum degree in G_r will be at least $k_q/(k_q+1) - \gamma_q$. Therefore, the Factor Finder algorithm will succeed and find a K_q-factor in a refinement of $\widehat{G}_r$.

We remark that the bound of $k_q = q - 3/2 + h_{q-1}/2$ could be improved somewhat. We didn't want to optimize this bound. It already gives the correct order of magnitude for our embedding method: $k_q = q + O(\log q)$, without having tedious computations in the proof of Lemma 17.

More on the Factor Finder algorithm

Let us explore more properties of the Factor Finder algorithm, which will be useful later on. Set $s_1(q) = \ell$ for every $q \geq 3$. Given a cluster $W \in A_1$ we denote its degree in $R(A_1, A_i)$ by $s_2(q)$, that is, $s_2(q) = \rho(\widetilde{\delta}(G_r))\ell$. The recursive process guarantees that we can construct a K_{q-1}-factor in the $q-1$ neighborhoods of W, each having size $s_2(q)$. Now for finding the K_{q-1}-factor we again apply recursion, and want to find a K_{q-2} factor in $s_2(q)$ different balanced $(q-2)$-partite graphs. The size of the vertex classes of these balanced graphs will be denoted by $s_3(q)$. In general, when proceeding with the recursion, step-by-step we construct balanced $(q-i)$-partite graphs, in which we look for a K_{q-i}-factor. The number of these graphs is $s_1(q) \cdot s_2(q) \cdots s_i(q)$. The number of clusters in a class of these balanced graphs are denoted by $s_i(q)$. We stop at $i = q-2$, when we arrive at balanced bipartite graphs, in which we are looking for perfect matchings.

We can compute the number of cliques in the K_q-factor which contain some split copy of a given cluster.

Lemma 18 *Let U be an arbitrary cluster in G_r. Each split copy of U appears in $\Pi_{i=2}^{q-1} s_i(q)$ cliques in the K_q-factor of $\widehat{G}_r$.*

Proof: We want to apply induction, but to do that we have to be careful. The statement we will prove by induction is as follows:
Claim: Let F be a balanced a-partite cluster graph with cluster classes of size ℓ, and W be a cluster of F. If $\widetilde{\delta}(F) \geq k_j/(k_j+1) - \gamma_a$ where $j \geq a$, and we apply the Factor Finder algorithm then the number of a-cliques containing a split copy of W is $\Pi_{i=2}^{a-1} s_i(j)$.

It is easy to see that this statement is stronger than that of the lemma. Notice, that we have to keep track of the size of the cluster classes, too.

We will show that in the case when $a = 3$ the above statement holds. Let $j \geq 3$. First assume that $U \in A_1$. The algorithm finds the neighborhoods $N_2(U) \subset A_2$ and $N_3(U) \subset A_3$, both having size $s_2(j)$. Next we look for a perfect matching between these two sets, every edge of this matching with U will result in a triangle. Hence, the number of triangles having a split copy of U is $s_2(j)$.

Suppose, that $U \in A_2$, and let $W \in N_R(U, A_1)$ be arbitrary. Then there will be a triangle which contains a split copy of W and a split copy of U. Since this holds for every cluster of $N_R(U, A_1)$, and this set has $s_2(j)$ clusters, there are $s_2(j)$ triangles which contain a split copy of U.

Assume now that $a > 3$ and that the induction hypothesis holds up to $a-1$. Let $j \geq a$. As above, we begin with the case $U \in A_1$. The algorithm first finds an $(a-1)$-partite cluster graph in which every cluster class has size $s_2(j)$, and U is adjacent to every cluster of this graph. We want to find a K_{a-1}-factor in some refinement of it by the Factor Finder algorithm. Let W be an arbitrary cluster from the "first" cluster class of the $a-1$ classes. We have $s_2(j)$ possible choices for W. The following is easy to see from the definition of the $s_f(g)$ numbers: for $1 \leq i \leq a-2$ the cluster classes of the $(a-i)$-partite graphs constructed by the Factor Finder algorithm will be of size $s_{i+1}(j)$. Hence, applying the induction hypothesis, there are $\Pi_{i=2}^{a-2} s_{i+1}(j)$ cliques on $a-1$ clusters which contain a split copy of W. We have $s_2(j)$ choices for W, therefore, the number of a-cliques containing a split copy of U is $s_2(j)\Pi_{i=2}^{a-2} s_{i+1}(j) = \Pi_{i=2}^{a-1} s_i(j)$.

Finally, we consider the case $a > 3$ when $U \in A_t$ for $t > 1$. In the first step there are $s_2(j)$ clusters of A_1 such that these are adjacent to U in $R(A_1, A_t)$. Let W be any of these clusters. Consider the $(a-1)$-partite cluster graph which is constructed for W by the algorithm. This cluster graph has classes of size $s_2(j)$. As above, we can apply induction, and get that the algorithm finds $\Pi_{i=2}^{a-2} s_{i+1}(j)$ cliques on $a-1$ clusters which contain a split copy of U. We repeat this for every cluster in $N_R(U, A_1)$, that results in $s_2(j)$ different $(a-1)$-partite graphs. In each of these we find $\Pi_{i=2}^{a-2} s_{i+1}(j)$ cliques on $a-1$ clusters containing a split copy of U. Overall, split copies of U appear in $s_2(j)\Pi_{i=2}^{a-2} s_{i+1}(j) = \Pi_{i=2}^{a-1} s_i(j)$ cliques on a clusters. □

Obviously, $s_1(q) > s_2(q) > s_3(q) > \ldots > s_{q-1}(q) > \frac{2\ell}{k_q+1}$ for $q \geq 3$. The last inequality follows from Claim 16 and the fact that the proportional minimum degree in the last graph is $\geq 1/2 + \psi_q$ for some positive constant ψ_q depending only on q. (Recall that $k_3/(k_3 + 1) - 0.68 > 0.01$, hence, $\psi_3 > 0.01$, and because of Lemma 17 the property of ψ_q being positive is inherited for larger values of q.) Observe, that the overall number of cliques in the K_q-factor is $\Pi_{i=1}^{q-1} s_i(q) = \nu_q \ell^{q-1}$, where ν_q is a constant. We have proved the following.

Lemma 19 *Every cluster in the refinement $\widehat{G}_r$ has size $\widehat{m} = m/(\nu_q \ell^{q-2})$, and the number of clusters in each vertex class of $\widehat{G}_r$ is $\widehat{\ell} = \Pi_{i=1}^{q-1} s_i(q) = \nu_q \ell^{q-1}$.*

5 The second stage – Finishing the proof of Theorem 3

In this section we discuss how to finish the embedding of H into G. Observe, that by applying Lemma 13, Proposition 14 and the Blow-up Lemma we are able to embed most of H into G: The edges in the cliques of the K_q-factor of $\widehat{G}_r$ represent ε'-regular pairs, which by Lemma 13 can be made super regular. Applying the Blow-up Lemma we get that most of H can be embedded into G, at most $3\varepsilon' n + |W_0| \leq 4\varepsilon' n$ vertices are left out, here $\varepsilon' \leq 2^{q-2}\varepsilon^{1/5^{q-2}}$. This follows from the repeated applications of Lemma 14.

Our main goal in this section is to embed the *whole* of H with the help of the Blow-up Lemma. For that we will try to find a K_q-factor in such a way that every edge in the cliques will represent $(\varepsilon', d-\varepsilon')$-super-regular pairs. Moreover, every vertex of G will sit in a cluster of some clique, every cluster of a clique will have the same size, and the size of any two clusters will differ by at most 1. We will achieve this goal in three steps.

- First, we discard those vertices from the cliques which do not have many neighbors in other clusters of the cliques, and put them into W_0, the exceptional cluster. At the end of this step every edge in a clique will represent a super-regular pair with a sufficiently large density.
- Second, we will distribute the vertices of W_0 among the non-exceptional clusters while maintaining the super-regularity of the pairs in the cliques. This step may create cliques with clusters having unequal size.
- Third, we move vertices between clusters so as to get equal size clusters in the cliques, but keep super-regularity, we call this the *balancing step.*

These steps prepare us to apply the Blow-up Lemma, that will finish the embedding.

We need an important lemma, which will be crucial for making the cluster sizes equal in every clique. In order to state it, let us define q directed graphs: $L_1, L_2, \ldots, L_q$. Here $V(L_i) = \widehat{A}_i$, the class containing the split copies of the clusters of the ith class. Let $\widehat{U}_1, \widehat{U}_2 \in \widehat{A}_i$, we will have the directed edge $(\widehat{U}_1, \widehat{U}_2) \in E(L_i)$, if $\widehat{U}_1$ is adjacent to all the clusters of the q-clique which contains $\widehat{U}_2$ except $\widehat{U}_2$ itself. That is, if $\widehat{W}$ is a cluster of this clique, then the $(\widehat{U}_1, \widehat{W})$ pair is ε'-regular. We will also say that $\widehat{U}_1$ is adjacent to the clique of $\widehat{U}_2$. We will show the following:

Lemma 20 *Let $U_1, U_2 \in A_i$ for some $1 \leq i \leq q$, and let $\widehat{U}_1$ be any split copy of U_1 in $\widehat{G}_r$. Then with probability at least $1 - 1/(2q\widehat{\ell})^2$ there are more than $\frac{1}{8}\psi_q^2 s_{q-1}(q)\Pi_{i=3}^{q-1} s_i(q)$ split copies of U_2 such that $\widehat{U}_1$ is adjacent to its clique.*

The main message of Lemma 20 is that out of the $\Pi_{i=2}^{q-1} s_i(q)$ cliques in the factor which contain some split copy of U_2 a constant proportion is adjacent to some split copy of U_1, independently of the choice of U_1 and U_2.

Proof: We begin with the case $q = 3$. First assume that $U_1, U_2 \in A_1$. Every triangle that contains split copy of U_2 has a split copy of one cluster from $N_R(U_2, A_2)$ and a split copy of one cluster from $N_R(U_2, A_3)$. Set $s = deg_R(U_2, A_2)$. Recall that s is so large that by Claim 16 every cluster in A_1 is adjacent to at least $(1/2 + \psi_3)s$ clusters in $N_R(U_2, A_2)$ and $N_R(U_2, A_3)$. Hence, the perfect matching between $N_R(U_2, A_2)$ and $N_R(U_2, A_3)$ has at least $2\psi_3 s$ edges which are adjacent to U_1. This proves the lemma in this special case. Notice that this part of the proof is not probabilistic.

Let us assume now that $U_1, U_2 \in A_i$ for $i = 2, 3$. While above we didn't need the randomly chosen edges in the perfect matching, this time they play a crucial role. Say that $U_1, U_2 \in A_2$, and set $s = deg_R(U_2, A_1)$. Let $W \in A_1$ be any cluster that is adjacent to U_1 and U_2. There are at least $(1/2 + \psi_3)s$ such clusters in A_1. In the R graph W has exactly s neighbors in A_2. Out of those neighbors at least $(1/2 + \psi_3)s$ are adjacent to U_1. Since U_2 has at least $(1/2 + \psi_3)s$ neighbors in $N_R(W, A_2)$ as well, the common neighborhood of U_1 and U_2 in $N_R(W, A_2)$ has at least $2\psi_3 s$ clusters. At this point we will use the randomly chosen edges.

The probability of randomly choosing an edge that contains U_2 such that the other endpoint is adjacent to U_1 is at least ψ_3^2, since U_2 will be chosen with probability $\psi_3/2$, and the chance that U_2 will be matched to a neighbor of U_1 is at least $2\psi_3$. Summing up for every $W \in A_1$ we get that the expected number of split copies of U_2 that are in such a triangle that U_1 is adjacent to the other two clusters in that triangle is at least $\psi_3^2 s_2(3)/2$. Using Azuma's inequality we get that with probability at least $1 - 1/(6\widehat{\ell})^2$ there are more than $\psi_3^2 s_2(3)/8$ split copies of U_2 such that $\widehat{U}_1$ is adjacent to its clique.

It is useful to look at this question from a different point of view that is easier to generalize for larger values of q. Let $1 \leq i \leq 3$, and fix $U_1 \in A_i$. Consider the 4-partite graph that we obtain from G_r when 'pulling out' U_1 from A_i: the new vertex classes are $U_1, A_i - U_1$, the other two classes remain intact. We also add new edges apart from the edges of G_r: every cluster of A_i will be adjacent to U_1. Then one looks for 4-cliques in this new graph using straightforward modification of the Factor Finder Algorithm. Every 4-clique we find will correspond to a triangle in the triangle factor of G_r that is adjacent to U_1.

Assume now that $q \geq 4$. Fix U_1 and construct the new $(q+1)$-partite graph, similarly to the previous case. We will follow the line of arguments of the proof of Lemma 18. When computing the number of cliques having a split copy $\widehat{U}_2$, at every step we have to take into account whether the clusters are in the neighborhood of U_1, that is, we act like there were $q+1$ vertex classes. This shrinks the sizes: if the cluster class size in question is $s_i(q)$ for q vertex classes, then out of this many clusters at least $s_{i+1}(q)$ are adjacent to U_1.

This estimation works smoothly until at the end we have to find a perfect matching in a bipartite graph having cluster classes of size $s_{q-1}(q)$ each. Then U_1 is adjacent to at least $(1/2 + \psi_q)s_{q-1}(q)$ clusters in both classes. At this point we repeat the argument of the case $q = 3$ and get that the expected number of cliques containing some split copy $\widehat{U}_2$ that are adjacent to U_1 is at least $\frac{1}{2}\psi_q^2 s_{q-1}(q)\Pi_{i=3}^{q-1} s_i(q)$. Here we applied the bound of Lemma 18. Using Azuma's inequality we obtain that U_1 will be adjacent to at least $\frac{1}{8}\psi_q^2 s_{q-1}(q)\Pi_{i=3}^{q-1} s_i(q)$ split copies of U_2 with probability at least $1 - 1/(2q\widehat{\ell})^2$. □

Observe, that if $\widehat{U}, \widehat{U}'$ are split copies of U, then $\widehat{U}$ is adjacent to the clique of $\widehat{U}'$. Together with the so called union bound in probability theory this implies the following:

Corollary 21 *With positive probability there are at least $\frac{1}{8}\psi_q^2 s_{q-1}(q)\Pi_{i=3}^{q-1} s_i(q)$ vertex disjoint directed paths of length at most two between any two clusters in L_i, for every $1 \leq i \leq q$.*

5.1 The final steps of the embedding

We have acquired the knowledge to achieve our main goal, in the rest of the section we discuss how to finish the embedding step by step.

Making super-regular pairs

In the first step we make every edge in the cliques of the factor super-regular by applying Lemma 13, the discarded vertices will be put into W_0. Then the extremal cluster W_0 has increased in size, but will still remain reasonably small: $|W_0| \leq \varepsilon' n$, where $\varepsilon' \leq 2\varepsilon^{5^{2-q}}$. Observe that $\varepsilon' \ll d' = d - \varepsilon'$ if ε is sufficiently small.

Distributing the vertices of W_0

In the second step we will distribute the vertices of W_0 among the $\widehat{\ell}$ clusters of $\widehat{G}_r$. Let $v \in W_0$ and $\widehat{U}$ be a cluster. We say that v is adjacent to the clique of $\widehat{U}$ if v has at least $(d - \varepsilon')\widehat{m}$ neighbors in every cluster in the clique of $\widehat{U}$, except in $\widehat{U}$ itself. Notice, that the proof of Lemma 20 shows that for every $v \in W_0$ there are at least $\frac{1}{8}\psi_q^2 \ell s_{q-1}(q)\Pi_{i=3}^{q-1} s_i(q)$

clusters such that v is adjacent to their cliques. Since the number of cliques is $\widehat{\ell} = \Pi_{i=1}^{q-1} s_i(q)$, every vertex is adjacent to $c_q\widehat{\ell}$ cliques, where $c_q = s_{q-1}(q)\psi_q^2/(8s_2(q))$.

When distributing the vertices of W_0 we are allowed to put a vertex v into a cluster $\widehat{U}$ if v is adjacent to the clique of $\widehat{U}$. We pay attention to distribute the vertices *evenly*, that is, at the end no cluster will get more than $2|W_0|/(c_q\widehat{\ell})$ new vertices from W_0. Since every vertex is adjacent to many cliques, this can be achieved. After this step every edge of every clique in the K_q-factor will represent super-regular pairs.

Balancing

It is possible that the clusters have different sizes in a clique, hence, we have to perform the balancing algorithm. For that we assign a number ν_i to the ith clique for every i such that $\sum_i \nu_i = n$ and $|\nu_i - \nu_j| \leq 1$ for every i, j. Notice that since $n/\hat{\ell}$ is not necessarily an integer, the ν_i numbers may differ by 1. We partition the clusters of $\widehat{G}_r$ into three sets: $S_<, S_=$ and $S_>$. A cluster from the ith clique will belong to $S_<$ if it has less than ν_i vertices. We put a cluster of the ith clique into $S_>$ if the cluster has more than ν_i vertices. Finally, $S_=$ will contain the rest with equality. We will apply Corollary 21 in order to find directed paths in the L_i graphs from clusters of $S_>$ to clusters in $S_<$.

Say, that $\widehat{U}_1 \in S_>, \widehat{U}_2 \in S_<$ and there is a path of length one between them, that is, $\widehat{U}_1\widehat{U}_2 \in E(L_i)$ for some $1 \leq i \leq q$. Then the vast majority of the vertices of $\widehat{U}_1$ are adjacent to the clique of $\widehat{U}_2$. Pick as many as needed (and possible) among these and place them to $\widehat{U}_2$. If the path is of length two, then choose a cluster $\widehat{U}_3$ such that $\widehat{U}_1\widehat{U}_3$ and $\widehat{U}_3\widehat{U}_2$ belong to $E(L_i)$. Again, the vast majority of the vertices in $\widehat{U}_1$ are adjacent to the clique of $\widehat{U}_3$ and the vast majority of the vertices of $\widehat{U}_3$ are adjacent to the clique of $\widehat{U}_2$. Hence, by placing vertices from $\widehat{U}_1$ to $\widehat{U}_3$ and the same number of vertices from $\widehat{U}_3$ to $\widehat{U}_2$ we decrease the discrepancy of $\widehat{U}_1$ and $\widehat{U}_2$ such that we keep the edges super-regular in all the cliques in question. Observe, that we perform the balancing algorithm such that we do not take out more than $2|W_0|/(c_q\widehat{\ell})$ vertices from any of the clusters, and do not put in more than $2|W_0|/(c_q\widehat{\ell})$ vertices to any of the cluster.

We can apply Lemma 11, and get that the edges of the cliques represent $(\widehat{\varepsilon}, \widehat{d})$-super-regular pairs, where $\widehat{\varepsilon} \leq C(\varepsilon')^{1/3}$ and $\widehat{d} \geq d' - \widehat{\varepsilon}$, and C is a constant.

Using the Blow-up Lemma

At this point we recognize that with positive probability all conditions of the Blow-up Lemma are satisfied. That is, the cluster sizes in each clique of the K_q-factor in $\widehat{G}_r$ are equal, and the edges in these cliques represent super-regular pairs. Hence, given an arbitrary clique in $\widehat{G}_r$ having clusters with t vertices, we can find t vertex disjoint copies of K_q in it using the Blow-up Lemma. Since every vertex of G sits in some clique of the K_q-factor of $\widehat{G}_r$, we proved Theorem 3. □

6 Finding an H-factor

Next we show how to find an H-factor, if H is a fixed q-colorable graph.

Proof of Corollary 4: We embed vertex disjoint copies of H as follows. First, we find a K_q-factor in the reduced graph of G rather than G itself. Then we find an equitable q-coloring of $\widetilde{H}$, where $\widetilde{H}$ is the vertex disjoint union of q copies of H. The coloring goes as follows. Assume that the color classes of H have sizes $c_1, c_2, \ldots, c_q$ for some proper

q-coloration. Then the coloring of $\widetilde{H}$ will follow a 'rotation scheme:' the jth color class of the ith copy of H will be colored $i + j - 1 \bmod q$. This way every color class of $\widetilde{H}$ will have size $c_1 + \ldots + c_q = v(H)$.

After finding the K_q-factor in G_r we will make the pairs in the cliques super-regular. Then the distribution of the vertices of W_0 can be performed the same way as above. Only balancing will be slightly different. This time for every i we assign a number ν_i to the ith clique such that $v(H)$ divides ν_i, $|\nu_i - \nu_j| \leq v(H)$ and $\sum_i \nu_i = n$. Then we use Corollary 21 for the balancing step, and get that for all i the clusters in the ith clique will have ν_i vertices.

It is easy to see that G has an $\widetilde{H}$-factor: we can embed the copies of $\widetilde{H}$ in the cliques of the K_q-factor with the help of the Blow-up Lemma. Since $\widetilde{H}$ is the vertex disjoint union of q copies of H, this way we have found an H-factor in G. $\square$

Acknowledgment The authors would like to thank Péter Hajnal and Endre Szemerédi for the helpful conversations and the anonymous referees for their helpful remarks that improved the presentation of the paper.

References

[1] N. Alon, J. Spencer, The Probabilistic Method, John Wiley & Sons, 3rd edition, 2008.

[2] H. Corrádi and A. Hajnal, On the Maximal Number of Independent Circuits in a Graph, Acta Math. Hung., **14** (1963) 423-439.

[3] B. Csaba, Regular Spanning Subgraphs of Bipartite Graphs of High Minimum Degree, The Electronic Journal of Combinatorics, #N21 (2007).

[4] E. Fischer, Variants of the Hajnal-Szemerédi Theorem, Journal of Graph Theory **31** (1999) 275-282.

[5] A. Hajnal and E. Szemerédi, Proof of a Conjecture of Erdős, in "Combinatorial Theory and Its Applications, II" (P. Erdős, and V. T. Sós, Eds.), Colloquia Mathematica Societatis János Bolyai, North-Holland, Amsterdam/London, 1970.

[6] A. Johansson, R. Johansson, K. Markström, Factors of r-partite graphs and bounds for the strong chromatic number, Ars Combinatoria (2010) to appear.

[7] R. Johansson, Triangle-factors in a balanced blown-up triangle, Discrete Mathematics **211** (2000) 249-254.

[8] H. Kierstead, A. Kostochka, A short proof of the Hajnal-Szemerédi Theorem on equitable coloring, Combinatorics, Probability and Computing **17** (2008) 265-270.

[9] H. Kierstead, A. Kostochka, M. Mydlarz, E. Szemerédi, A fast algorithm for equitable coloring, Combinatorica **30** (2010) 217-224.

[10] Y. Kohayakawa, V. Rödl, Szemerédi's Regularity Lemma and Quasi-randomness, Recent Advances in Algorithms and Combinatorics, CMS Books in Mathematics (2003) 289-351.

[11] J. Komlós, G.N. Sárközy and E. Szemerédi, Blow-up Lemma, Combinatorica, **17** (1997) 109-123.

[12] J. Komlós, G.N. Sárközy and E. Szemerédi, An Algorithmic Version of the Blow-up Lemma, Random Struct. Alg., **12** (1998) 297-312.

[13] J. Komlós, M. Simonovits, Szemerédi's Regularity Lemma and its Applications in Graph Theory, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 295-352.

[14] Cs. Magyar, R. Martin, Tripartite version of the Corrádi-Hajnal theorem, Discrete Mathematics **254** (2002) 289-308.

[15] R. Martin, E. Szemerédi, Quadripartite version of the Hajnal-Szemerédi theorem, Discrete Mathematics **308** (2008) 4337-4360.

[16] M. Mydlarz, E. Szemerédi, Algorithmic Brooks' Theorem (2007) manuscript.

[17] E. Szemerédi, Regular Partitions of Graphs, Colloques Internationaux C.N.R.S N$^{\underline{\text{o}}}$ 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay (1976) 399-401.