The Algorithmic Aspects of the Regularity Lemma *

N. Alon † R. A. Duke ‡ H. Lefmann § V. Rödl ¶ R. Yuster‖

Abstract

The Regularity Lemma of Szemerédi is a result that asserts that every graph can be partitioned in a certain regular way. This result has numerous applications, but its known proof is not algorithmic. Here we first demonstrate the computational difficulty of finding a regular partition; we show that deciding if a given partition of an input graph satisfies the properties guaranteed by the lemma is co-NP-complete. However, we also prove that despite this difficulty the lemma can be made constructive; we show how to obtain, for any input graph, a partition with the properties guaranteed by the lemma, efficiently. The desired partition, for an n-vertex graph, can be found in time $O(M(n))$, where $M(n) = O(n^{2.376})$ is the time needed to multiply two n by n matrices with $0,1$-entries over the integers. The algorithm can be parallelized and implemented in NC^1. Besides the curious phenomenon of exhibiting a natural problem in which the search for a solution is easy whereas the decision if a given instance is a solution is difficult (if P and NP differ), our constructive version of the Regularity Lemma supplies efficient sequential and parallel algorithms for many problems, some of which are naturally motivated by the study of various graph embedding and graph coloring problems.

*Part of this paper was presented at the 33^{rd} IEEE Symposium on Foundations of Computer Science

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Research supported in part by a United States Israel BSF Grant

‡School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

§Lehrstuhl Informatik II, Universität Dortmund, Postfach 500 500, W-4600 Dortmund 50, Germany

¶Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

‖Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

1 Introduction

One of the main reasons for the fast development of combinatorics during the recent years is the widely used application of combinatorial methods in the study and the development of efficient algorithms. It is therefore surprising that many results proved by applying some of the modern combinatorial techniques merely consist of existence proofs and do not yield efficient algorithms for the corresponding problems. Many examples of this type appear in [22] and in [2]. Some of these, such as the Lovász Local Lemma whose algorithmic aspects have been studied recently in [6] and [3], are general techniques that have various algorithmic applications; a constructive version for such a technique immediately supplies efficient procedures for many problems.

In the present paper we consider another general technique of this form. This is the Regularity Lemma of Szemerédi [28]. This lemma has numerous applications in various areas including combinatorial number theory ([27], [14]), computational complexity([19]) and, mainly, extremal graph theory ([5], [7], [8], [12], [13], [15], [16], [17], [23], [24], [25], [26]). Not all of these applications have obvious corresponding algorithmic problems, but many do. A number of these problems, as well as various similar ones, are naturally suggested by the study of graph embedding, graph coloring and graph decomposition.

The known proof of the Regularity Lemma is not algorithmic. Here we prove two main results concerning the constructive aspects of this lemma. The first result shows the computational difficulties; we show that deciding if a given partition of an input graph satisfies the regularity properties guaranteed by the lemma is co-NP-complete. The second result shows that despite these difficulties the lemma can be made constructive; we show how to obtain, for any input graph, a partition with the required properties efficiently. The desired partition, for an n-vertex graph, can be found in time $O(M(n))$, where $M(n) = O(n^{2.376})$ is the time needed to multiply two n by n matrices with $0,1$-entries over the integers (see [9]). The algorithm can be parallelized and supplies efficient sequential and parallel algorithms for many problems. Some of these follow easily from the known applications of the lemma, together with our constructive version of it, and some other similar applications are new, and require several additional ideas.

In order to describe our results more precisely, we need several technical definitions, which mostly follow the ones in [28]. If $G = (V, E)$ is a graph, and A, B are two disjoint subsets of V, let $e(A, B)$ denote the number of edges of G with an endpoint in A and an endpoint in B. If A and B are non-empty, define the *density of edges* between A and B by $d(A, B) = \frac{e(A,B)}{|A||B|}$. For $\epsilon > 0$, the pair (A, B) is called *ϵ-regular* if for every $X \subset A$ and $Y \subset B$ satisfying $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$, the inequality

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

holds.

An *equitable partition* of a set V is a partition of V into pairwise disjoint classes $C_0, C_1, \ldots, C_k$, in which all the classes C_i for $1 \leq i \leq k$ have the same cardinality. The class C_0 is called the *exceptional class* and may be empty. An equitable partition of the set of vertices V of G into the classes $C_0, C_1 \ldots, C_k$, with C_0 being the exceptional class, is called *ϵ-regular* if $|C_0| \leq \epsilon|V|$, and all but at most $\epsilon\binom{k}{2}$ $(\leq \epsilon k^2)$ of the pairs (C_i, C_j) for $1 \leq i < j \leq k$ are ϵ-regular.

The following lemma is proved in [28] and is usually called the Regularity Lemma.

Lemma 1.1 (The Regularity Lemma [28]) *For every $\epsilon > 0$ and every positive integer t there is an integer $T = T(\epsilon, t)$ such that every graph with $n > T$ vertices has an ϵ-regular partition into $k + 1$ classes, where $t \leq k \leq T$.* $\Box$

It is worth noting that in [28] the author raises the question if the assertion of the lemma holds even if we do not allow any irregular pairs in the definition of a regular partition. This, however, is false, as observed by several researchers including L. Lovász, P. Seymour, T. Trotter and ourselves. A simple example showing that irregular pairs are necessary is a bipartite graph with vertex classes $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ in which $a_i b_j$ is an edge iff $i \leq j$.

In the present paper we first prove the following two new results.

Theorem 1.2 *The following decision problem is co-NP-complete.*
Instance: An input graph G, an integer $k \geq 1$ and a parameter $\epsilon > 0$, and a partition of the set of vertices of G into $k + 1$ parts.
Problem: Decide if the given partition is ϵ-regular.

We note that the proof actually gives that the problem remains co-NP-complete even for $\epsilon = 1/2$ and for $k = 2$.

Theorem 1.3 (A constructive version of the Regularity Lemma) *For every $\epsilon > 0$ and every positive integer t there is an integer $Q = Q(\epsilon, t)$ such that every graph with $n > Q$ vertices has an ϵ-regular partition into $k + 1$ classes, where $t \leq k \leq Q$. For every fixed $\epsilon > 0$ and $t \geq 1$ such a partition can be found in $O(M(n))$ sequential time, where $M(n)$ is the time for multiplying two n by n matrices with $0, 1$ entries over the integers. It can also be found in time $O(\log n)$ on an EREW PRAM with a polynomial number of parallel processors.*

We note that the dependence of $Q(\epsilon, t)$ (as well as that of $T(\epsilon, t)$ in the original lemma) on ϵ and t, as well as the dependence of the running time of the algorithm in the last theorem on these two parameters, is rather horrible; in fact *$log^* Q$* is a polynomial (of degree about 20) in $1/\epsilon$ and in t. However, in the applications ϵ will always be fixed (though small) and t will always be fixed (though large), whereas the size n of the graph will grow. However, in section 5 we will sketch how these parameters can be improved in some cases.

In most of the known applications of the Regularity Lemma, the only non-constructive part is the lemma itself. Therefore, Theorem 1.3 supplies numerous immediate algorithmic applications. Here is a new, relatively simple application. Recall that a *topological copy* of a graph H is a graph obtained from H by replacing some of its edges by pairwise disjoint paths. Topological copies, especially those in which the paths that replace edges are relatively short, arise in the study of various graph embedding and emulation problems.

Proposition 1.4 *For any positive $\delta > 0$ there is a positive $c = c(\delta)$ such that for every m and for every graph H with m edges, every graph G with $n \geq cm$ vertices and with at least δn^2 edges contains a topological copy of H in which each edge is replaced by a path of length 4. Such a copy of H in G can be found in polynomial sequential time as well as in NC.*

In some applications, the construction of the uniform partition does not yield a straightforward $O(M(n))$ algorithm for the problem, and further refinements of the partition are needed in order to obtain this bound. We present such an application, which is the algorithmic version of the main result of [23]. This application is a certain (slightly unnatural and yet, we believe, interesting) polynomial procedure for estimating the chromatic number of a graph G in the following sense: either it supplies a short proof that G contains a small subgraph which is not $(k-1)$-colorable or it provides a proper $(k-1)$-coloring of an approximation of G. Here is the exact statement.

Proposition 1.5 *Let k be an integer, $k \geq 3$, and let c and ϵ be positive constants. Then there exist an integer $n_0 = n_0(k, \epsilon)$ and a function $f(k, \epsilon)$ such that if $G = (V, E)$ is any graph with n vertices, $n \geq n_0$, and cn^2 edges, then either*

(i) there exists a subgraph $G^* = (V^*, E^*)$ of G with $\chi(G^*) \geq k$ and $|V^*| \leq f(k, \epsilon)$, or

(ii) there exists a set $E^{**} \subseteq E$ of edges of G with $|E^{**}| \leq \epsilon \cdot n^2$, such that the subgraph $G^{**} = (V, E \setminus E^{**})$ satisfies $\chi(G^{**}) \leq k - 1$.

Furthermore there is an algorithm with input a graph $G = (V, E)$ as above and running time $O(M(n))$ which either yields a subgraph G^* as in (i) or a set of edges E^{**} as in (ii) together with a proper coloring $\Delta: V \longrightarrow \{1, 2, \ldots, k-1\}$ of the subgraph $G^{**} = (V, E \setminus E^{**})$.

The rest of the paper is organized as follows. In section 2 we present the proof of Theorem 1.2 that shows the difficulties in obtaining a constructive version of the Regularity Lemma. In Section 3 we show how to overcome these difficulties and prove Theorem 1.3. Section 4 contains various algorithmic applications, including the proofs of Propositions 1.4 and 1.5, and outlined proofs of several additional applications. Section 5 introduces a variant of the Regularity Lemma which can be used in some cases to achieve stronger results.

2 The complexity of deciding if a given partition is regular

In this section we prove the following result, which implies Theorem 1.2.

Theorem 2.1 *The following problem is co-NP-complete: Given $\epsilon > 0$ and a bipartite graph G with vertex classes A, B such that $|A| = |B| = n$, determine if G is ϵ-regular. (I.e., determine if $\{\emptyset, A, B\}$ is an ϵ-regular partition of G).*

The proof of Theorem 2.1 involves a series of two main reductions and a few minor ones. We start from the basic CLIQUE problem and reduce it to the $K_{k,k}$ Problem which is the following:
The $K_{k,k}$ Problem: Given a positive integer k and a bipartite graph G with vertex classes A, B such that $|A| = |B| = n$, determine if G contains the complete bipartite graph with k vertices in each vertex class.

The $K_{k,k}$ Problem is mentioned in [18] as NP-complete, but since the proof is not presented we describe one here.

Lemma 2.2 *The following decision problem (HALF SIZE CLIQUE) is NP-complete: Given a graph G on n vertices where n is odd, decide if G contains a subgraph isomorphic to $K_{\frac{n+1}{2}}$.*

Proof Let a graph $G = (V, E)$ and a positive integer k be the input to CLIQUE. We may assume that G has an odd number of vertices since we can add an isolated vertex. If $k \leq \frac{|V|+1}{2}$ we define $G^* = G + K_{|V|+1-2k}$ (where $G + H$ is the graph obtained from the disjoint union of G and H by joining every vertex of G to every vertex of H). Otherwise, we let G^* be the disjoint union of G and $E^{2k-|V|-1}$ where E^i is the graph on i isolated vertices. In any case, G^* has a HALF SIZE CLIQUE iff G has a subgraph isomorphic to K_k. $\Box$

Lemma 2.3 *The $K_{k,k}$ Problem is NP-Hard.*

Proof Let $G = (V, E)$ be the input to HALF SIZE CLIQUE. We reduce to $K_{k,k}$. Assume that $V = \{1, \ldots, n\}$. Define a bipartite graph $H = (A \cup B, F)$ as follows:

$$A = \{\alpha_{ij} | 1 \leq i, j \leq n\}$$

$$B = \{\beta_{ij} | 1 \leq i, j \leq n\}$$

$$F = \{(\alpha_{ij}, \beta_{kl}) | i = k \ or \ (i,k) \in E \ and \ l \neq i \ and \ j \neq k\} \ .$$

Proposition: G has a clique of size $\frac{n+1}{2}$ iff H has a subgraph isomorphic to $K_{(\frac{n+1}{2})^2, (\frac{n+1}{2})^2}$.
(a) Assume that G has a clique on W where $|W| = \frac{n+1}{2}$. Define

$$A' = \{\alpha_{ij} | i \in W, j \in V \setminus W\} \cup \{\alpha_{ii} | i \in W\}$$

$$B' = \{\beta_{ij} | i \in W, j \in V \setminus W\} \cup \{\beta_{ii} | i \in W\} \ .$$

The subgraph of H spanned by $B' \cup A'$ is complete bipartite, and

$$|A'| = |B'| = (\frac{n+1}{2})(\frac{n-1}{2}) + (\frac{n+1}{2}) = (\frac{n+1}{2})^2 \ .$$

(b) Suppose that H has a subgraph isomorphic to $K_{(\frac{n+1}{2})^2, (\frac{n+1}{2})^2}$. Let $H' = (A', B')$ be a maximal complete bipartite subgraph of H containing it. We may assume that

$$|A'| \geq |B'| \geq (\frac{n+1}{2})^2 \ . \tag{1}$$

Define $A^* = \{i | \alpha_{ij} \in A' \ \ for \ some \ \ j\}$, $s(i) = |\{j | \alpha_{ij} \in A'\}|$. Define B^* and $t(i)$ analogously. Finally, define $C^* = A^* \cap B^*$.
Claim: The following equalities hold for $s(i)$ and $t(i)$:

1. $s(i) = n - |B^*|$ for $i \in A^* \setminus C^*$
2. $s(i) = n - |B^*| + 1$ for $i \in C^*$
3. $t(i) = n - |A^*|$ for $i \in B^* \setminus C^*$
4. $t(i) = n - |A^*| + 1$ for $i \in C^*$.

Proof We begin by proving the first equality.
(i) If $i \in A^* \setminus C^*$ then for every $k \in B^*$ we have $\alpha_{ik} \notin A'$ since otherwise we have that $(\alpha_{ik}, \beta_{kj})$ is an edge for some j in contradiction to the definition of F.
(ii) If $i \in A^* \setminus C^*$ then for every $k \notin B^*$ we have $\alpha_{ik} \in A'$. To see this we note that by the definition of A^*, $\alpha_{ik'} \in A'$ for some k'. Let β_{jl} be any vertex in B'. Since $(\alpha_{ik'}, \beta_{jl}) \in F$ it follows from the definition of F that $(i,j) \in E$, $l \neq i$, $j \neq k'$. The other option is impossible because $i \neq j$ since $j \in B^*$ and $i \notin B^*$. For the same reason, $j \neq k$ and so $(\alpha_{ik}, \beta_{jl}) \in F$. Now from the maximality of H' it follows that $\alpha_{ik} \in A'$ as stated.
From (i) and (ii), the first equality (1) follows.
(iii) If $i \in C^*$ then for all $k \in B^*, k \neq i$ we have $\alpha_{ik} \notin A'$ as in (i) above.
(iv) If $i \in C^*$ then for all $k \notin B^*$ and also for $k = i$ we have $\alpha_{ik} \in A'$ as in (ii) above.

From (iii) and (iv) the second equality follows. The third and fourth equalities are proved analogously, replacing A by B and s by t. This completes the proof of the claim.

Denote $x = |A^*|$, $y = |B^*|$, $z = |C^*|$. By the last claim we obtain

$$|A'| = \sum_{i \in A^*} s(i) = x(n-y) + z \ ,$$

$$|B'| = \sum_{i \in B^*} t(i) = y(n-x) + z \ .$$

By (1) we have

$$x(n-y) + z \geq y(n-x) + z \geq (\frac{n+1}{2})^2 \ . \tag{2}$$

By (2) and trivial inclusions we have

$$n \geq x \geq y \geq z \geq 0 \ . \tag{3}$$

In order to satisfy (2) we must therefore have $y(n-x) + y \geq (\frac{n+1}{2})^2$, but the only values of x and y under the constraints (3) that satisfy the last inequality are $x = \frac{n+1}{2}$ and $y = \frac{n+1}{2}$, and for these values the inequality is actually an equality. It now follows from (2) and $z \leq y$, that $z = \frac{n+1}{2}$. Since $z = |C^*|$ and since C^* is a clique in G, the proposition follows. $\Box$

Since the reduction in Lemma 2.3 can be done in $O(n^2)$ time, it follows that the $K_{k,k}$ problem is NP-complete. $\Box$

In order to deduce the co-NP-completeness of ϵ-regularity, we need a stricter version of the $K_{k,k}$ problem.

Lemma 2.4 *The following problem is NP-complete: Given a bipartite graph $G = (A \cup B, E)$ where $|A| = |B| = n$ and $|E| = \frac{n^2}{2} - 1$ decide if G contains a subgraph isomorphic to $K_{\frac{n}{2},\frac{n}{2}}$.*

Proof: Let $G = (A \cup B, E)$ and k be the input to the $K_{k,k}$ Problem. We reduce to the restricted problem mentioned in the lemma. Our first step is similar to Lemma 2.2. We add $2k - n$ isolated vertices to each class if $k > \frac{n}{2}$ or add $n - 2k$ vertices (to each class) that are connected to every vertex of the opposite class if $k < \frac{n}{2}$. Thus, we may assume that n is even and $k = \frac{n}{2}$. However, the size of the edge set of G is not necessarily $\frac{n^2}{2} - 1$.

(a) Suppose $|E| < \frac{n^2}{2} - 1$ (note that if $|E| < \frac{n^2}{4}$ we are done, so we assume that $|E| \geq \frac{n^2}{4}$). Add to the vertex class A two sets of vertices of size $\frac{n}{2}$ each, A_1, A_2. Similarly add B_1, B_2 to B. We add edges to obtain the following complete bipartite subgraphs: (A_1, B_1), (A_2, B_2), (A_1, B), (B_1, A). Denote the modified graph by $G^* = (A^* \cup B^*, E^*)$. The size of E^* is

$$|E^*| = |E| + 2(\frac{n}{2})^2 + 2(\frac{n}{2})n < \frac{(2n)^2}{2} - 1 \ .$$

If G contained a $K_{\frac{n}{2},\frac{n}{2}}$ then by adding to it the vertices of A_1 and B_1 we get that G^* contains a $K_{n,n}$. On the other hand, if G^* contains an $H = K_{n,n}$, then no vertex of A_2 or B_2 belongs to H and by omitting from H the vertices of A_1 and B_1, we get that G contains a $K_{\frac{n}{2},\frac{n}{2}}$. We can now add edges between A_2 and B and between B_2 and A, but we must take care that the degrees of the vertices of A_2 in (A_2, B) and, respectively, the vertices of B_2 in (B_2, A) do not exceed $\frac{n}{2} - 1$, because we do not want to create any new copies of $K_{n,n}$ in G^* during the process. We add these edges until the size of E^* reaches exactly $\frac{(2n)^2}{2} - 1$. To prove that this is possible we must show that

$$|E| + 2(\frac{n}{2})^2 + 2(\frac{n}{2})n + 2\frac{n}{2}(\frac{n}{2} - 1) \geq \frac{(2n)^2}{2} - 1 \ .$$

Rearranging the last inequality we see that it is equivalent to $|E| \geq n - 1$ which is true for every $n \geq 2$ and $|E| \geq \frac{n^2}{4}$.

(b) Suppose $|E| > \frac{n^2}{2} - 1$. Extend G to G^* in a way similar to the one done in case (a) above, but with sets A_1, A_2, B_1, B_2 of size n each. We create the complete bipartite subgraphs on (A_1, B_1), (A_1, B) and (B_1, A). A simple computation yields that $|E^*| = |E| + 3n^2 < \frac{(3n)^2}{2} - 1$. A similar argument to the above shows that G^* contains a $K_{\frac{3n}{2},\frac{3n}{2}}$ iff G contains a $K_{\frac{n}{2},\frac{n}{2}}$. We can now add edges between A_2 and B_2 and between A_2 and B and between B_2 and A while maintaining the degrees of the vertices of A_2 and B_2 less than $\frac{n}{2}$ in (A_2, B) and (B_2, A) respectively. We keep doing so until the size of E^* becomes exactly $\frac{(3n)^2}{2} - 1$. To prove that this is possible we must show that

$$|E| + 3n^2 + n^2 + 2n(\frac{n}{2} - 1) \geq \frac{(3n)^2}{2} - 1 \ ,$$

but this is true for every $n \geq 2$ and $|E| > \frac{n^2}{2} - 1$. $\square$

The co-NP-completeness of ϵ-regularity is a simple consequence of Lemma 2.4. A bipartite graph G with n vertices in each class and exactly $\frac{n^2}{2} - 1$ edges contains a $K_{\frac{n}{2},\frac{n}{2}}$ iff it is not ϵ-regular for $\epsilon = \frac{1}{2}$. To see this, suppose first that it is not ϵ-regular, and let X and Y be any offending sets. We must have

$$|d(X, Y) - \frac{1}{2} + \frac{1}{n^2}| \geq \frac{1}{2} \ .$$

This is possible iff $d(X, Y) = 1$. Since $|X| \geq \frac{1}{2}n$, $|Y| \geq \frac{1}{2}n$, the subgraph of G spanned by X and Y contains a $K_{\frac{n}{2},\frac{n}{2}}$. The other direction is simpler. This completes the proof of Theorem 2.1, since verifying that two sets X and Y are offending can be done in linear time. $\square$

3 Finding a regular partition efficiently

In this section we prove Theorem 1.3. The basic idea that provides the proof (despite the assertion of Theorem 2.1) is the design of an efficient approximation algorithm. Given a graph G and an

$\epsilon > 0$ this algorithm computes some $\epsilon' < \epsilon$ (which is a function of ϵ). In case G is not ϵ-regular, the algorithm correctly reports this is the case, and provides evidence showing that G is not ϵ' $(< \epsilon)$-regular. In case G is ϵ' $(< \epsilon)$ regular the algorithm decides it is ϵ-regular. In any other case (i.e., if G is ϵ-regular but not ϵ'-regular) the algorithm behaves according to one of the above two possibilities and we have no control on the possibility it chooses.

The detailed proof of Theorem 1.3 is rather lengthy, and we split it into several lemmas. Let H be a bipartite graph with equal color classes $|A| = |B| = n$. Let d be the average degree of H. For two distinct vertices $y_1, y_2 \in B$ define the *neighbourhood deviation* of y_1 and y_2 by

$$\sigma(y_1, y_2) = |N(y_1) \cap N(y_2)| - \frac{d^2}{n}.$$

For a subset $Y \subset B$ denote the *deviation* of Y by

$$\sigma(Y) = \frac{\sum_{y_1, y_2 \in Y} \sigma(y_1, y_2)}{|Y|^2} .$$

Lemma 3.1 *Let H be a bipartite graph with equal classes $|A| = |B| = n$, and let d denote the average degree of H. Let $0 < \epsilon < \frac{1}{16}$. If there exists $Y \subset B$, $|Y| \geq \epsilon n$ such that $\sigma(Y) \geq \frac{\epsilon^3}{2} n$ then at least one of the following cases occurs.*

1. *$d < \epsilon^3 n$.*
2. *There exists in B a set of more than $\frac{1}{8}\epsilon^4 n$ vertices whose degrees deviate from d by at least $\epsilon^4 n$.*
3. *There are subsets $A' \subset A$, $B' \subset B$, $|A'| \geq \frac{\epsilon^4}{4} n$, $|B'| \geq \frac{\epsilon^4}{4} n$ and $|d(A', B') - d(A, B)| \geq \epsilon^4$.*

Moreover, there is an algorithm whose input is a graph H with a set $Y \subset B$ as above that outputs either
(i) The fact that 1 holds, or
(ii) The fact that 2 holds and a subset of more than $\frac{1}{8}\epsilon^4 n$ members of B demonstrating this fact, or
(iii) The fact that 3 holds and two subsets A' and B' as in 3 demonstrating this fact.
The algorithm runs in sequential time $O(M(n))$, where $M(n) = O(n^{2.376})$ is the time needed to multiply two n by n $0,1$ matrices over the integers, and can be parallelized and implemented in time $O(\log n)$ on an EREW PRAM with a polynomial number of parallel processors.

Proof We first assume that cases 1 and 2 do not happen and prove that case 3 must happen. Denote $Y' = \{y \in Y |\ |deg(y) - d| < \epsilon^4 n\}$. Note that Y' is not empty since case 2 does not happen. Choose a $y_0 \in Y'$ that maximizes $\sum_{y \in Y} \sigma(y_0, y)$. Let us estimate this quantity. Clearly,

$$\sum_{y' \in Y'} \sum_{y \in Y, y \neq y'} \sigma(y', y) = \sigma(Y)|Y|^2 - \sum_{y' \in Y \setminus Y'} \sum_{y \in Y, y \neq y'} \sigma(y', y) \geq \frac{\epsilon^3}{2} n|Y|^2 - \frac{\epsilon^4}{8} n|Y|n .$$

Since $|Y'| \leq |Y|$ we get

$$\sum_{y \in Y} \sigma(y_0, y) \geq \frac{\epsilon^3}{2} n|Y| - \frac{\epsilon^4}{8} n^2 \geq \frac{3}{8} \epsilon^3 n|Y| \ . \tag{4}$$

There are at least $\frac{\epsilon^4}{4} n$ vertices $y \in Y$ whose neighbourhood deviation with y_0 is greater than $2\epsilon^4 n$. To see this we note that if there were less we would get, using the fact that the neighbourhood deviation cannot exceed n,

$$\sum_{y \in Y} \sigma(y_0, y) \leq \frac{\epsilon^4}{4} n^2 + |Y| 2\epsilon^4 n \leq \frac{\epsilon^3}{4} n|Y| + 2\epsilon^4 n|Y| < \frac{3}{8} \epsilon^3 n|Y| \ ,$$

which contradicts (4).

Hence there is a set $B' \subset Y$, $|B'| \geq \frac{\epsilon^4}{4} n$, $y_0 \notin B'$, and for every vertex $b \in B'$ we have $|N(b) \cap N(y_0)| > \frac{d^2}{n} + 2\epsilon^4 n$. Define $A' = N(y_0)$. Clearly, $|A'| \geq d - \epsilon^4 n \geq \epsilon^3 n - \epsilon^4 n \geq 15\epsilon^4 n > \frac{\epsilon^4}{4} n$. We will show that $|d(A', B') - d(A, B)| \geq \epsilon^4$. Indeed,

$$e(A', B') = \sum_{b \in B'} |N(y_0) \cap N(b)| > \frac{|B'| d^2}{n} + 2\epsilon^4 n|B'| \ .$$

Therefore,

$$d(A', B') - d(A, B) > \frac{d^2}{n|A'|} + \frac{2\epsilon^4 n}{|A'|} - \frac{d}{n} \geq \frac{d^2}{n(d + \epsilon^4 n)} + 2\epsilon^4 - \frac{d}{n} = 2\epsilon^4 - \frac{d\epsilon^4}{d + \epsilon^4 n} \geq \epsilon^4 \ .$$

This completes the proof of the non-algorithmic part of the lemma.

The existence of the required sequential algorithm is simple. One can clearly check if 1 holds in time $O(n^2)$. Similarly, it is trivial to check if 2 holds in $O(n^2)$ time, and in case it holds to exhibit the required subset of B establishing this fact. If both cases above fail we continue as follows. For each $y_0 \in B$ with $|deg(y_0) - d| < \epsilon^4 n$ we find the set of vertices $B_{y_0} = \{y \in B | \ \sigma(y_0, y) \geq 2\epsilon^4 n\}$. The last proof guarantees the existence of at least one such y_0 for which $|B_{y_0}| \geq \frac{\epsilon^4}{4} n$. The subsets $B' = B_{y_0}$ and $A' = N(y_0)$ are the required ones. Since the computation of all the quantities $\sigma(y, y')$ for $y, y' \in B$ can be done by squaring the adjacency matrix of H the claimed sequential running time follows. The parallelization is obvious. $\square$

The basic idea in the proof of the next lemma resembles an idea used in [1].

Lemma 3.2 *Let H be a bipartite graph with equal classes $|A| = |B| = n$. Let $2n^{-1/4} < \epsilon < \frac{1}{16}$. Assume that at most $\frac{1}{8}\epsilon^4 n$ vertices of B deviate from the average degree of H by at least $\epsilon^4 n$. Then, if H is not ϵ-regular then there exists $Y \subset B$, $|Y| \geq \epsilon n$ such that $\sigma(Y) \geq \frac{\epsilon^3}{2} n$.*

Proof We assume that for every such Y, $\sigma(Y) \le \frac{\epsilon^3}{2}n$, and show that H is ϵ-regular. It is not too difficult to see that the pair (A, B) is ϵ-regular if and only if for every two subsets $X \subset A$, $Y \subset B$, $|X| = \lceil \epsilon |A| \rceil$, $|Y| = \lceil \epsilon |B| \rceil$, the inequality $|d(A, B) - d(X, Y)| \le \epsilon$ holds. We therefore fix two such subsets, and show that the inequality holds.

Claim:

$$\sum_{x \in X} (|N(x) \cap Y| - \frac{d|Y|}{n})^2 \le e(A, Y) + |Y|^2 \sigma(Y) + \frac{2}{5}\epsilon^5 n^3 \ . \tag{5}$$

Proof Let $M = (m_{i,j})$ be the adjacency matrix of H. Then,

$$\sum_{x \in X} (|N(x) \cap Y| - \frac{d|Y|}{n})^2 \le \sum_{x \in A} (|N(x) \cap Y| - \frac{d|Y|}{n})^2$$

$$= \sum_{x \in A} (\sum_{y \in Y} m_{x,y} - \frac{d|Y|}{n})^2 = \sum_{x \in A} (\sum_{y \in Y} {m^2}_{x,y} + \frac{d^2|Y|^2}{n^2} + \sum_{y, y' \in Y, y \ne y'} m_{x,y} m_{x,y'} - 2\frac{d|Y|}{n} \sum_{y \in Y} m_{x,y})$$

$$= e(A, Y) + \frac{d^2|Y|^2}{n} + \sum_{y, y' \in Y, y \ne y'} |N(y) \cap N(y')| - 2e(A, Y)\frac{d|Y|}{n}$$

$$= e(A, Y) + \frac{d^2|Y|^2}{n} + \sum_{y, y' \in Y, y \ne y'} (\sigma(y, y') + \frac{d^2}{n}) - 2e(A, Y)\frac{d|Y|}{n}$$

$$\le e(A, Y) + \sigma(Y)|Y|^2 + \frac{2d^2|Y|^2}{n} - 2e(A, Y)\frac{d|Y|}{n} \ .$$

It remains to show that

$$\frac{|Y|^2 d^2}{n} - e(A, Y)\frac{d|Y|}{n} \le \frac{1}{5}\epsilon^5 n^3 \ .$$

Multiplying the last inequality by $\frac{1}{d|Y|^2}$ and rearranging the terms we must show that

$$d(A, Y) \ge \frac{d}{n} - \frac{\frac{1}{5}\epsilon^5 n^3}{d|Y|^2} \ .$$

Indeed,

$$d(A, Y) = \frac{e(A, Y)}{n|Y|} \ge \frac{(d - \epsilon^4 n)(|Y| - \frac{1}{8}\epsilon^4 n)}{n|Y|} = \frac{d}{n} - \epsilon^4 - \frac{\frac{1}{8}\epsilon^4 d}{|Y|} + \frac{\frac{1}{8}\epsilon^8 n}{|Y|} \ge \frac{d}{n} - \epsilon^4 - \frac{1}{8}\epsilon^3 \ .$$

Using the bounds on ϵ which imply that $1 < \epsilon^4 n$ and since $|Y| \le 1 + \epsilon n$,

$$\frac{\frac{1}{5}\epsilon^5 n^3}{d|Y|^2} \ge \frac{\frac{1}{5}\epsilon^5 n^2}{(\epsilon n + 1)^2} \ge \frac{\frac{1}{5}\epsilon^5}{(\epsilon + \epsilon^4)^2} \ge \epsilon^4 + \frac{1}{8}\epsilon^3 \ .$$

This completes the proof of the claim.

Returning to the proof of the lemma, we note that by the Cauchy-Schwarz Inequality

$$\sum_{x \in X} (|N(x) \cap Y| - \frac{d|Y|}{n})^2 \ge \frac{1}{|X|}((\sum_{x \in X} |N(x) \cap Y|) - \frac{d|X||Y|}{n})^2 \ .$$

Therefore by the previous claim we obtain

$$((\sum_{x\in X}|N(x)\cap Y|)-\frac{d|X||Y|}{n})^2 \le |X|(e(A,Y)+|Y|^2\sigma(Y)+\frac{2}{5}\epsilon^5 n^3)\ .$$

Dividing the last inequality by $|X|^2|Y|^2$ we obtain

$$|d(X,Y)-d(A,B)|^2 \le \frac{1}{|X||Y|^2}(e(A,Y)+|Y|^2\sigma(Y)+\frac{2}{5}\epsilon^5 n^3)\ .$$

Using the fact that $e(A,Y) \le (d+\epsilon^4 n)|Y| + \frac{1}{8}\epsilon^4 n^2$ and $\sigma(Y) \le \frac{\epsilon^3}{2}n$ and $\epsilon > 2n^{-1/4}$ we estimate

$$|d(X,Y)-d(A,B)|^2 \le \frac{1}{|X||Y|^2}((d+\epsilon^4 n)|Y| + \frac{1}{8}\epsilon^4 n^2 + |Y|^2\frac{\epsilon^3}{2}n + \frac{2}{5}\epsilon^5 n^3) \le$$

$$\le \frac{n+\epsilon^4 n}{\epsilon^2 n^2} + \frac{\frac{1}{8}\epsilon^4 n^2}{\epsilon^3 n^3} + \frac{\epsilon^3 n}{2\epsilon n} + \frac{\frac{2}{5}\epsilon^5 n^3}{\epsilon^3 n^3} \le \frac{\epsilon^2(1+\epsilon^4)}{16} + \frac{\epsilon^5}{128} + \frac{9}{10}\epsilon^2 < \epsilon^2.$$

Therefore, (A,B) is an ϵ-regular pair. □

Corollary 3.3 *Let H be a bipartite graph with equal classes $|A| = |B| = n$. Let $2n^{-1/4} < \epsilon < \frac{1}{16}$. There is an $O(M(n))$ algorithm that verifies that H is ϵ-regular or finds two subsets $A' \subset A$, $B' \subset B$, $|A'| \ge \frac{\epsilon^4}{16}n$, $|B'| \ge \frac{\epsilon^4}{16}n$, such that $|d(A,B) - d(A',B')| \ge \epsilon^4$. The algorithm can be parallelized and implemented in NC^1.*

Proof We begin by computing d, the average degree of H. If $d < \epsilon^3 n$, then by a trivial computation H is ϵ-regular, and we are done.
Next, we count the number of vertices in B whose degrees deviate from d by at least $\epsilon^4 n$. If there are more than $\frac{\epsilon^4}{8}n$ such vertices, then the degrees of at least half of them deviate in the same direction and if we let B' be such a set of vertices, then $|B'| \ge \frac{\epsilon^4}{16}n$. A simple computation yields that $|d(B',A) - d(B,A)| \ge \epsilon^4$, and we are done.
By Lemma 3.2, it is now sufficient to show that if there exists $Y \subset B$ with $|Y| \ge \epsilon n$ and $\sigma(Y) \ge \frac{\epsilon^3 n}{2}$, then one can find in $O(M(n))$ time the required subsets A' and B'. But this follows from the assertion of Lemma 3.1. The parallelization is immediate. □

With every equitable partition P of the set of vertices of a graph into $k+1$ classes $C_0, C_1, \ldots, C_k$ we associate, following [28], a number called the *index* of P defined by

$$ind(P) = \frac{1}{k^2}\sum_{1\le r<s\le k} d(C_r, C_s)^2.$$

The following lemma is proved in [28].

Lemma 3.4 *Fix k and γ and let $G = (V, E)$ be a graph with n vertices. Let P be an equitable partition of V into classes $C_0, C_1, \ldots, C_k$. Assume $|C_1| > 4^{2k}$ and $4^k > 600\gamma^{-5}$. Given proofs that more than γk^2 pairs (C_r, C_s) are not γ-regular (where by proofs we mean subsets $X = X(r,s) \subset C_r$, $Y = Y(r,s) \subset C_s$ that violate the condition of γ-regularity of (C_r, C_s)), then one can find in $O(n)$ time a partition P' (which is a refinement of P) into $1 + k4^k$ classes, with an exceptional class of cardinality at most*

$$|C_0| + \frac{n}{4^k}$$

and such that

$$ind(P') \geq ind(P) + \frac{\gamma^5}{20} .$$

□

Theorem 1.3 is a simple consequence of Lemma 3.4 and Corollary 3.3. Given any $\epsilon > 0$ and a positive integer t, we define the constants $N = N(\epsilon, t)$, $T = T(\epsilon, t)$ as follows; let b be the least positive integer such that

$$4^b > 600(\frac{\epsilon^4}{16})^{-5}, \quad b \geq t .$$

Let f be the integer valued function defined inductively as

$$f(0) = b, \quad f(i+1) = f(i)4^{f(i)} .$$

Put $T = f(\lceil 10(\frac{\epsilon^4}{16})^{-5} \rceil)$ and put $N = max\{T4^{2T}, \frac{32T}{\epsilon^5}\}$. We prove the theorem with $Q = N$ $(\geq T)$. Given a graph $G = (V, E)$ with n vertices where $n \geq N$, we show how to construct in $O(M(n))$ time an ϵ-regular partition of G into $k+1$ classes where $t \leq k \leq T$ $(\leq Q)$.

The following procedure achieves this goal:

1. Arbitrarily divide the vertices of G into an equitable partition P_1 with classes $C_0, C_1, \ldots, C_b$, where $|C_1| = \lfloor n/b \rfloor$ and hence $|C_0| < b$. Denote $k_1 = b$.

2. For every pair (C_r, C_s) of P_i, verify if it is ϵ-regular or find $X \subset C_r$, $Y \subset C_s$, $|X| \geq \frac{\epsilon^4}{16}|C_1|$, $|Y| \geq \frac{\epsilon^4}{16}|C_1|$, such that $|d(X, Y) - d(C_s, C_t)| \geq \epsilon^4$.

3. If there are at most $\epsilon\binom{k_i}{2}$ pairs that are not verified as ϵ-regular, then halt. P_i is an ϵ-regular partition.

4. Apply Lemma 3.4 where $P = P_i$, $k = k_i$, $\gamma = \frac{\epsilon^4}{16}$, and obtain a partition P' with $1 + k_i4^{k_i}$ classes.

5. Let $k_{i+1} = k_i4^{k_i}$, $P_{i+1} = P'$, $i = i + 1$, and go to step 2.

Claim: The procedure above is correct and can be implemented in the asserted time.

Proof Denote $\gamma = \frac{\epsilon^4}{16}$. We prove by induction on the iteration counter i that $ind(P_i) \geq \frac{(i-1)\gamma^5}{20}$ and that the size of the exceptional class of P_i (denoted by ${C_0}^i$) is at most $\gamma n(1-\frac{1}{2^i})$. This is true for $i=1$ since the index is nonnegative and since

$$|{C_0}^1| = n - \lfloor \frac{n}{b} \rfloor b < b < T \leq \frac{n\epsilon^5}{32} < n\frac{\gamma}{2} \ .$$

Assuming it is true for i we prove it for $i+1$. By Lemma 3.4

$$ind(P_{i+1}) \geq ind(P_i) + \frac{\gamma^5}{20}$$

and so $ind(P_{i+1}) \geq \frac{i\gamma^5}{20}$. Since $\gamma 4^{k_i} \geq 2^{i+1}$ it follows that $|{C_0}^{i+1}| - |{C_0}^i| \leq \frac{n}{4^{k_i}} \leq \frac{\gamma n}{2^{i+1}}$.
Since the index cannot exceed $\frac{1}{2}$, we conclude that after at most $\lceil 10\gamma^{-5} \rceil$ iterations the algorithm must halt, producing an equitable partition $P = P_i$. Note that k_i is exactly $f(i)$ as described above, and so the number of classes in P is at most T. The size of the exceptional class of P is at most $\gamma n < \epsilon n$. We must also verify that when applying Lemma 3.4 and Corollary 3.3 the conditions in their statements are met. This is easily verified by the choices of N and b, and by the fact that during the whole process the number of classes does not exceed T. Note also that the restriction that $\epsilon < \frac{1}{16}$ in Corollary 3.3 may be dropped because we may always produce an $\frac{1}{16}$-regular partition for larger ϵ.
Finally, we note that the number of iterations is constant (does not depend on n), and that the running time of each iteration is bounded by the $O(M(c))$ bound of Corollary 3.3 where c is the size of the classes in the equitable partition, which is less than n. $\square$

4 Algorithmic applications

We start this section with the proof of Proposition 1.4. We note that although, as far as we know, this is a new result, its proof, though non-trivial, is a rather standard application of the Regularity Lemma, and is given here mainly as a relatively simple representative example of a result proved by this lemma. The existence proof, together with Theorem 1.3 and a few additional ideas yield efficient (sequential and parallel) procedures for the corresponding algorithmic problem.

Proof of Proposition 1.4. Without making any attempt to optimize our constants, we prove the assertion with

$$c = c(\delta) = \frac{70}{\delta} Q(\frac{\delta}{20}, \frac{20}{\delta}),$$

where Q is the function that appears in Theorem 1.3. To simplify the notation, we omit all floor and ceiling signs whenever these are not essential. Let H be a graph with m edges, and let G be a graph with $n \geq cm$ vertices and with at least δn^2 edges. Define $\epsilon = \frac{\delta}{20}$ and $t = \frac{20}{\delta}$ and apply Theorem 1.3 to G with these values of ϵ and t. We obtain (constructively) an ϵ-regular partition of the set of vertices of G into $k+1$ classes, where $t \leq k \leq Q(\epsilon, t)$. By the definition of c each non-exceptional class in the partition has more than $\frac{60}{\delta}m$ vertices. Since most pairs of classes are ϵ-regular, and since the total number of edges of G is at least δn^2 and the number of edges inside the classes as well as the number of these that touch the exceptional class is very small, an easy edge counting shows that there is an ϵ-regular pair on two of the classes, which we denote by A and B, so that $d(A, B) > \frac{\delta}{2}$. Let A' be the set of all vertices of A whose number of neighbors in B is at least $\frac{\delta}{3}|B|$. By ϵ-regularity $|A'| \geq (1-\epsilon)|A| > \frac{30}{\delta}m > 2m$ (since otherwise $A - A'$ and B would violate ϵ- regularity.) We claim that between any two vertices of A' there are at least $6m$ internally vertex disjoint paths of length 4, each containing two internal vertices of B and one of A. To see this, fix two vertices a_1 and a_2 in A' and suppose we have already found $s < 6m$ such disjoint paths between them. Delete all the vertices of these paths from A and B, and let B_1 and B_2 denote, respectively, the set of all neighbors of a_1 and a_2 among the remaining vertices of B. Clearly

$$|B_i| \geq \frac{\delta}{3}|B| - 12m > \epsilon|B| \ for\ i = 1, 2.$$

Therefore, by ϵ-regularity (applied to the pair A and B_1 and to the pair A and B_2) almost all vertices of A have many neighbors in B_1 and in B_2 and we can thus find an additional path of length 4 consisting of such a vertex of A that has not been used in the previous paths and of a neighbor of it in B_1 and another one in B_2. This completes the proof of the claim.

Returning to the proof of the proposition, we can now construct the copy of H on any set of $|V(H)|$ vertices of A' (since $|V(H)| \leq 2m$ we can choose that many vertices in A'). To do so, we construct the paths corresponding to the edges of H one by one. After constructing paths corresponding to some $p < m$ of the edges, suppose we need a path connecting a_1 and a_2. By the claim there are $6m$ internally vertex disjoint paths between these two endpoints. Up to now we have only used at most $2m + 3p < 5m$ vertices, so at least one of the paths above can be chosen, completing the proof of existence of the required copy of H.

It remains to check that the above proof is algorithmic. The sequential polynomial time implementation is immediate. For the parallel implementation, observe that the proof implies that any *maximal* set of internally disjoint paths of length 4 between each two members of A' is of size at least $6m$. Such a set can be easily found in NC by applying one of the many known NC algorithms for finding a maximal independent set in an appropriately defined graph (see, e.g., [20],[21], [4]). (The graph considered here is of course the one whose set of vertices are all possible paths of length 4 with the required endpoints where two are adjacent if they share an internal vertex). Once the sets of $6m$ paths between each pair are given, we can find the required copy of H as follows. Let p denote the number of vertices of H, $q \leq m$ the number of its edges and choose arbitrarily p vertices of A'. Define a new graph whose set of vertices is the $6mq$ paths consisting of the q sets of $6m$ disjoint paths each joining pairs corresponding to adjacent vertices of H. Two such paths are adjacent iff they intersect or iff they belong to the same set of $6m$ paths. It is easy to check that any maximal independent set of vertices in this new graph gives a copy of H as required, completing the proof of the proposition. □

Many additional algorithmic applications of theorem 1.3 follow from the known applications of the Regularity Lemma. Here is a list of results, whose detailed proofs are mostly omitted, where in each case the appropriate reference in which existence is proved is mentioned in brackets. The existence proof in most of these cases can be translated, with some minor additional ideas, into a sequential algorithm by Theorem 1.3. Ideas similar to ones given in the last proof usually yield a parallel implementation as well.

Proposition 4.1 (Existence proved in [5]) *For every $\epsilon > 0$ and for every integer h, there exists a positive integer $n_0 = n_0(\epsilon, h)$ such that for every graph H with h vertices and chromatic number $\chi(H)$, there exists a polynomial algorithm that given a graph $G = (V, E)$ with $n > n_0$ vertices and minimum degree $d > \frac{\chi(H)-1}{\chi(H)}n$ finds a set of $(1-\epsilon)n/h$ vertex disjoint copies of H in G.*

Proof (Outline) By a correct choice of the constants (see [5]), n_0 is determined. Note also that H is a fixed graph and let $k = \chi(H)$. Given a graph $G = (V, E)$ with $n > n_0$ vertices and minimum degree $d > \frac{k-1}{k}n$, we construct a regular partition of G with parameters γ^2, t that are also functions of ϵ and h (see [5]). Let $q+1$ be the number of classes in this partition, $t \leq q \leq Q(\gamma^2, t)$. Let L be the graph on the vertices $1, \dots, q$ where two vertices are connected if the corresponding classes in the partition are γ^2-regular and their density is at least $1/t+\gamma$. Clearly, L can be constructed in $O(|E|)$ time, given the partition. It can be shown that such a graph must contain at least $\frac{q}{k}(1-11k^2/t)$ disjoint copies of K_k. Such a set of copies may be found in constant time by exhaustive search on L. It can also be shown that every such "k-clique" of classes in the partition, contains at least $(1-\gamma)\frac{n(1-\gamma^2)}{qh}$ vertex disjoint copies of the complete k-partite graph having h vertices in each of its color classes (and, therefore, $(1-\gamma)n(1-\gamma^2)/h$ copies of H). These copies can be found in $O(n^2)$ time by simply constructing the copies one by one, and for every copy, constructing its vertices in each color class one by one, where every class in the partition contains exactly one color class. The details of this procedure can be found in Lemma 2.2 in [5]. By summing up the number of copies found in each of the k-cliques, the proposition is proved. □

Proposition 4.2 (Existence can be proved using the methods of [8]) *For any $\delta > 0$ and $d \geq 1$ there is a $c = c(\delta, d)$ such that for every bipartite graph H with m vertices and maximum degree d, any graph G with $n \geq cm$ vertices and with at least δn^2 edges contains a copy of H. Such a copy of H in G can be found in polynomial time.*

Proposition 4.3 (Existence proved in [8]) *For any $d \geq 1$ there is a $c = c(d)$ such that for every graph H with m vertices and maximum degree d, for any graph G with $n \geq cm$ vertices either G or its complement contains a copy of H. Such a copy of H in G or in its complement can be found in polynomial time.*

Proposition 4.4 (Existence (for a special case) proved in [17]) *For every positive integer k and for every $\gamma > 0$, there exists an $\epsilon = \epsilon(\gamma, k)$ such that given a graph G with n vertices and less than ϵn^k copies of K_k (the complete graph on k vertices), one can find in polynomial time a set of at most γn^2 edges of G whose deletion will leave G K_k-free.*

In some cases the construction of the regular partition is not enough in order to achieve an $O(M(n))$ running time, and additional refinements of the partition are needed in order to obtain this bound. Such an approach is demonstrated in the proof of Proposition 1.5, which is more difficult than those of the other applications mentioned in this section.

Proof of Proposition 1.5. In the proof of Proposition 1.5 we use the following result:

Lemma 4.5 (proved in [23]) *Let α, β be positive real numbers with $0 < \alpha \leq 1$ and $0 < \beta \leq \frac{1}{2}$, and let r be a positive integer.*

Then there exist a positive constant $\epsilon_d(\alpha, \beta)$ and a positive integer $l_0 = l_0(\alpha, \beta, r)$ such that for every r-partite graph $G = ((A_i)_{i=1}^r, E)$ with $|A_i| = m > l_0$ the following holds.

There are subsets $B_i \subseteq A_i$ with $|B_i| = l_0$ for $i = 1, 2, \ldots, r$ such that if the pair (A_i, A_j) is $\epsilon_d(\alpha, \beta)$-regular with density of edges $d(A_i, A_j) \geq 2\beta$, then for all subsets $C_i \subseteq B_i$ and $C_j \subseteq B_j$ with $|C_i| = |C_j| = \lceil \alpha \cdot l_0 \rceil$ there exists an edge between C_i and C_j.

Remark 4.1 *It follows from the proof in [23] that in Lemma 4.5 the values of ϵ_d and l_0 can be computed as follows:*

(i) *Let t be the least positive integer with $8 \cdot (1-\beta)^{\alpha \cdot t} < 1$.*

(ii) *Let $l_0 = l_0(\alpha, \beta, r)$ be the least positive integer with $\binom{r}{2} \cdot \left(8 \cdot (1-\beta)^{\alpha \cdot t}\right)^{l_0} < 1$ and with $\lceil \alpha \cdot l_0 \rceil \equiv 0 \bmod (t)$.*

(iii) $\epsilon_d(\alpha, \beta) = min \left\{ \beta, \frac{(1-\beta)^{t^2}}{2 \cdot t} \right\}$.

In particular, the value of $\epsilon_d(\alpha, \beta)$ does not depend on r.

In what follows we refer to the auxiliary graph as the graph whose vertices are the classes in the ϵ-regular partition and whose edges correspond to dense regular pairs.

Let $G = (V, E)$ be a graph on n vertices with $n \geq n_1(k, \epsilon)$, where $n_1(k, \epsilon)$ is large enough for the following considerations. Moreover, let G have $c \cdot n^2$ edges for some positive constant $c > 0$.

The idea of the algorithm is as follows. We construct an ϵ_1-regular partition of G, where ϵ_1 is appropriate. Then delete all "bad edges" from G, i.e. those which touch the exceptional class, or lie completely in one class of the ϵ_1-regular partition or lie in a low density pair. For the resulting graph G_1 construct the auxiliary graph $Aux(G_1)$. Now construct an ϵ_2-regular partition of the graph $Aux(G_1)$, delete the bad edges to obtain a subgraph G_2 of $Aux(G_1)$, and construct the auxiliary graph $Aux(G_2)$. If $\chi(Aux(G_2)) \geq k$, then by Lemma 4.5 we infer that there exists a small subgraph G_2^* of G_2 with $\chi(G_2^*) \geq k$, cf. [23]. Then reconstruct a copy of G_2^* in G_1, by choosing the vertices of that copy one by one. On the other hand, if $\chi(Aux(G_2)) \leq k-1$, then we obtain a good coloring for $Aux(G_2)$, which induces a good coloring of G_2, which again induces a good coloring of a subgraph G_1^* of G_1. The subgraph G_1^* differs from G in at most $\epsilon \cdot n^2$ edges, and this gives the algorithm. In the following we give the argument in detail. We remark that in our approach of the proof of Proposition 1.5 one application of the Regularity Lemma does not suffice in general, as the desired subgraph G^* of $Aux(G_1)$ (with $\chi(G^*) \geq k$) cannot be reconstructed according to our approach.

Let k be a positive integer and let $\epsilon > 0$ be a positive real number. Put

$$\begin{aligned} h_2 &= \left\lceil \frac{3}{\epsilon} \right\rceil \\ \beta_2 &= \frac{\epsilon}{6} \\ \epsilon_2 &= \min\left\{ \frac{\epsilon}{12}, \epsilon_d\left(\frac{1}{k-1}, \beta_2\right) \right\}, \end{aligned} \tag{6}$$

where $\epsilon_d(\alpha, \beta)$ is computed according to Lemma 4.5 and Remark 4.1 for the parameters $\alpha = \frac{1}{k-1}$ and $\beta = \beta_2$.

Precompute $M_2 = Q(\epsilon_2, h_2)$ according to Theorem 1.3.

Let $l_0 = l_0(\alpha, \beta, r)$ be given by Lemma 4.5 and Remark 4.1 for the parameters $\alpha = \frac{1}{k-1}$, $\beta = \beta_2$ and $r = M_2$.

Put

$$\begin{aligned} \beta_1 &= \frac{\epsilon}{3} \\ \epsilon_1 &= \min\left\{ \frac{\epsilon}{12}, \left(\frac{\beta_1}{3}\right)^{l_0 M_2} \right\} \end{aligned}$$

$$h_1 \quad = \quad \max\left\{\left\lceil\frac{3}{\epsilon}\right\rceil, Q\left(\epsilon_2, l_0\cdot(M_2+1)\right)\right\} \ , \tag{7}$$

Finally, put

$$n_1(k,\epsilon) = Q(\epsilon_1, h_1)\left(\frac{3}{\beta_1}\right)^{l_0 M_2} \tag{8}$$

These precomputations can be done in running time independent of n, hence require $O(1)$ time only.

By Theorem 1.3 we can construct for the given graph G on n vertices, $n \geq n_1(k,\epsilon)$, in running time $O(n^{2.376})$ an ϵ_1-regular partition with k_1+1 classes, where $h_1 \leq k_1 \leq M_1$, where M_1 is a constant depending on h_1 and ϵ_1 only. Let this ϵ_1-regular partition of G be given by $V = A_0 \cup A_1 \cup \ldots \cup A_{k_1}$, where A_0 is the exceptional class. In particular, $|A_0| \leq \epsilon_1 \cdot n$ and $|A_1| = |A_2| = \ldots = |A_{k_1}|$ and all - up to $\epsilon_1 \cdot k_1^2$ - pairs (A_i, A_j), $1 \leq i < j \leq k_1$, are ϵ_1-regular.

Now delete in running time at most $O(n^2)$ all edges in G, which have at least one endpoint in the class A_0, or which have both endpoints in some class A_i for some i, $1 \leq i \leq k_1$, or which have one endpoint in the class A_i and the other in A_j, where (A_i, A_j) is an ϵ_1-irregular pair or satisfies $d(A_i, A_j) < \beta_1$. The total number of deleted edges is at most

$$\begin{aligned}
& \epsilon_1 \cdot n^2 + k_1 \cdot \binom{\frac{n}{k_1}}{2} + \epsilon_1 \cdot k_1^2 \cdot \left(\frac{n}{k_1}\right)^2 + \beta_1 \cdot \binom{k_1}{2} \cdot \left(\frac{n}{k_1}\right)^2 \\
\leq \quad & 2 \cdot \epsilon_1 \cdot n^2 + \frac{n^2}{2 \cdot k_1} + \frac{\beta_1}{2} \cdot n^2 \\
\leq \quad & \frac{\epsilon}{2} \cdot n^2 \ ,
\end{aligned}$$

as $\epsilon_1 \leq \epsilon/12$, $k_1 \geq h_1 \geq \lceil 3/\epsilon \rceil$ and $\beta_1 = \epsilon/3$ by (7). Denote the resulting subgraph of G by $G_1 = (V_1, E_1)$. In particular,

$$\begin{aligned}
|E_1| \quad & \geq \quad |E| - \frac{\epsilon}{2} \cdot n^2 \\
& \geq \quad (c - \frac{\epsilon}{2}) \cdot n^2 \ .
\end{aligned} \tag{9}$$

Construct in running time at most $O(n^2)$ the auxiliary graph $Aux(G_1) = (V_1^{Aux}, E_1^{Aux})$ corresponding to G_1 with vertex set $V_1^{Aux} = \{1, 2, \ldots, k_1\}$ and edge set E_1^{Aux} given by $\{i,j\} \in E_1^{Aux}$ if there exists an edge between A_i and A_j in G_1. Then,

$$\begin{aligned}
|E_1^{Aux}| \quad & \geq \quad \frac{|E_1|}{\left(\frac{n}{k_1}\right)^2} \\
& \geq \quad k_1^2 \cdot (c - \frac{\epsilon}{2}) \qquad \text{by (9).}
\end{aligned} \tag{10}$$

Apply, as before, to the graph $Aux(G_1)$ the algorithm from Theorem 1.3 and obtain in running time $O(k_1^{2.376}) = O(1)$ an ϵ_2-regular partition of $Aux(G_1)$ with k_2+1 classes $B_0, B_1, \ldots, B_{k_2}$, where $h_2 \leq k_2 \leq M_2$. The class B_0 is the exceptional class with $|B_0| \leq \epsilon_2 \cdot k_1$ and $|B_1| = |B_2| = \ldots = |B_{k_2}|$, and at most $\epsilon_2 \cdot k_2^2$ pairs (B_i, B_j) are ϵ_2-irregular, $1 \leq i < j \leq k_2$.

As above, delete from $Aux(G_1)$ all edges, which touch the class B_0 or which are contained in some class B_i, $1 \leq i \leq k_2$, or are contained in a pair (B_i, B_j), which is ϵ_2-irregular or satisfies $d(B_i, B_j) < 2 \cdot \beta_2$. Let $G_2 = (V_2, E_2)$ be the resulting subgraph. Then

$$\begin{aligned} |E_2| &\geq |E_1^{Aux}| - \left(2 \cdot \epsilon_2 \cdot k_1^2 + \frac{k_1^2}{2 \cdot k_2} + \beta_2 \cdot k_1^2 \right) \\ &\geq |E_1^{Aux}| - \frac{\epsilon}{2} \cdot k_1^2 && \text{by (6).} \\ &\geq k_1^2 \cdot (c - \epsilon) && \text{by (10).} \end{aligned} \tag{11}$$

Construct the auxiliary graph $Aux(G_2) = (V_2^{Aux}, E_2^{Aux})$ for G_2 with vertices $V_2^{Aux} = \{1, 2, \ldots, k_2\}$ and edges $\{i, j\} \in E_2^{Aux}$, if there exists an edge between B_i and B_j in G_2. This can be done in running time $O(k_1^2) = O(1)$. Then

$$\begin{aligned} |E_2^{Aux}| &\geq \frac{|E_2|}{\left(\frac{k_1}{k_2}\right)^2} \\ &\geq k_2^2 \cdot (c - \epsilon) \; . \end{aligned} \tag{12}$$

Going backwards, we delete in running time $O(n^2)$ some more edges of the graph G_1. If some edge $\{i, j\}$ satisfies $\{i, j\} \in E_1^{Aux}$ but $\{i, j\} \notin E_2$, then delete all edges in G_1 between the vertex sets A_i and A_j. As a deleted edge of $Aux(G_1)$ corresponds to deleting at most $\frac{n^2}{k_1^2}$ edges of G_1, and as at most $\frac{\epsilon}{2} \cdot k_1^2$ edges were deleted from $Aux(G_1)$ to obtain G_2, the resulting subgraph $G_1^* = (V_1^*, E_1^*)$ of G_1 satisfies

$$\begin{aligned} |E_1^*| &\geq |E_1| - \frac{\epsilon}{2} \cdot k_1^2 \cdot \frac{n^2}{k_1^2} \\ &\geq (c - \epsilon) \cdot n^2 && \text{by (9).} \end{aligned} \tag{13}$$

Hence, G_1^* is a subgraph of the original graph G, and they differ by at most $\epsilon \cdot n^2$ edges.

Now check by any reasonable algorithm in running time that depends on k_2, but not on n, whether $\chi(Aux(G_2)) \geq k$ or not. We consider two cases according to the value of $\chi(Aux(G_2))$:

Suppose first that $\chi(Aux(G_2)) \leq k - 1$. Then determine in running time that depends only on k_2 a good coloring for $Aux(G_2)$, say $\Delta_2^{Aux} \colon V_2^{Aux} \longrightarrow \{1, 2, \ldots, k^*\}$ for some $k^* < k$. This induces in running time that depends only on k_1 a coloring $\Delta_2 \colon V_2 \longrightarrow \{1, 2, \ldots, k^*\}$ by $\Delta_2(v) = \Delta_2^{Aux}(i)$ if $v \in B_i$. If $v \in B_0$, color v by color 1. By construction Δ_2 is a good coloring for G_2, i.e.

$\chi(G_2) < k$. Now Δ_2 induces in running time $O(n)$ another coloring $\hat{\Delta_1}: V_1^* \longrightarrow \{1, 2, \ldots, k^*\}$ by $\hat{\Delta_1}(v) = \Delta_2(j)$ if $v \in A_j$. Moreover, color vertices $v \in A_0$ by color 1. (Indeed, we could easily neglect the exceptional classes, as they do not contain edges and are not touched by any edges.) We claim that $\hat{\Delta_2}$ is a good coloring for G_1^*. Namely, suppose, there are $v, w \in V_1^*$ with $\{v, w\} \in E_1^*$ and with $\hat{\Delta_1}(v) = \hat{\Delta_1}(w)$. Assume that $v \in A_i$ and $w \in A_j$. By construction, there is an edge between i and j in G_2, which implies $\Delta_2(i) \neq \Delta_2(j)$. Hence, the subgraph G_1^*, which differs from the original graph G by at most $\epsilon \cdot n^2$ edges, is colored by a good coloring using at most $k-1$ colors.

Now assume that $\chi(Aux(G_2)) \geq k$. Apply Lemma 4.5 with $\alpha = \frac{1}{k-1}$, $\beta = \beta_2$ and $r = k_2$ for the graph G_2. We may do so because by Remark 4.1 $l_0{}' = l_0(\frac{1}{k-1}, \beta_2, k_2) \leq l_0(\frac{1}{k-1}, \beta_2, M_2) = l_0$ and since by the choice of h_1 in (7), $|B_i| \geq l_0$ for $i = 1, \ldots, k_2$. We search for the l_0' element subsets $C_i \subseteq B_i$ guaranteed by Lemma 4.5 by any reasonable algorithm, in $O(1)$ time. Let $G_2^{**} = (\cup_{i=1}^{k_2} C_i, E_2^{**})$ be the subgraph of G_2, induced by $C_1 \cup C_2 \cup \ldots \cup C_{k_2}$. This subgraph has $m = l_0' \cdot k_2 \leq l_0 \cdot k_2$ vertices, where $m = m(k, \epsilon)$ is independent of n. We claim that this subgraph satisfies

$$\chi(G_2^{**}) \geq k \; . \tag{14}$$

Suppose for contradiction that this is not the case. Then there exists a good coloring $\Delta: \cup_{i=1}^{k_2} C_i \longrightarrow \{1, 2, \ldots, k-1\}$. In each set C_i one color, say, t_i occurs at least $\lceil \frac{l_0}{k-1} \rceil$ times. For $i = 1, 2, \ldots, k_2$ pick subsets $D_i \subseteq C_i$ with $D_i \subseteq \Delta^{-1}(t_i)$ and $|D_i| = \left\lceil \frac{l_0}{k-1} \right\rceil$. Induce a coloring $\Delta^{Aux}: V_2^{Aux} \longrightarrow \{1, 2, \ldots, k-1\}$ by $\Delta^{Aux}(i) = t_i$ for $i \in V_2^{Aux}$. As $\chi(Aux(G_2)) \geq k$, there exists an edge $\{i, j\} \in E_2^{Aux}$, where $\Delta^{Aux}(i) = \Delta^{Aux}(j)$. By construction, the pair (B_i, B_j) is ϵ_2-regular with density $d(B_i, B_j) \geq 2 \cdot \beta_2$. As by (6) $\epsilon_2 \leq \epsilon_d(\frac{1}{k-1}, \beta_2)$, we infer by Lemma 4.5 that there exists an edge between D_i and D_j, which proves (14). We can now reconstruct a copy of G_2^{**} in G_1^*, in a way similar to the procedure used in the proofs of Propositions 1.4 and 4.1. This is possible since

$$n_1 \geq \left(\frac{3}{\beta_1}\right)^{l_0 \cdot M_2}$$

and since

$$\epsilon_1 \leq \left(\frac{\beta_1}{3}\right)^{l_0 \cdot M_2}$$

Notice that nonedges in G_2^{**} correspond to pairs (A_i, A_j) in G_1^*, where there is no edge between A_i and A_j in G_1^*. The algorithm is now complete.

Summarizing, all operations require a total running time of $O(n^{2.376})$. $\square$

5 Variants of the Regularity Lemma

As mentioned in the introduction, the dependence of $Q(\epsilon, t)$, the upper bound on the number of partition classes, on the parameters ϵ and t in the algorithm in Theorem 1.3 is rather horrible. We can improve this situation in some cases. In order to do so we would make use of a variant of the Regularity Lemma in which we have better control over the parameters. In this section we will sketch such an alternative approach. The details will be given in [10].

Let $G = (V, E)$, $V = \cup_{1 \le i \le k} V_i$, be a k-partite graph. We will consider partitions of the set $V_1 \times V_2 \times \ldots V_k$, where each partition class is of the form $W_1 \times W_2 \times \ldots \times W_k$, $W_i \subseteq V_i$ for $i = 1, 2, \ldots, k$. We will call such sets *cylinders*. We say that the cylinder $W_1 \times W_2 \times \ldots \times W_k$ is *ϵ-regular* if the subgraph of G induced on the set $\cup_{1 \le i \le k} W_i$ is such that all $\binom{k}{2}$ of the pairs (W_i, W_j), $1 \le i < j \le k$, are ϵ-regular.

A lemma similar to the following was proved in [11].

Lemma 5.1 *Let $G = (V, E)$ be a k-partite graph with $V = \cup_{1 \le i \le k} V_i$, $|V_i| = N$, $i = 1, 2, \ldots, k$.*

Then for every $\epsilon > 0$ there exists a partition $\mathcal{K}$ of $V_1 \times V_2 \times \ldots \times V_k$ into q cylinders such that

(*i*) $q \le 4^h$, *where* $h \le \frac{\binom{k}{2}}{\epsilon^5}$, *and*

(*ii*) *all but $\epsilon \cdot N^k$ of the k-tuples $(v_1, v_2, \ldots, v_k)$, $v_i \in V_i$, $i = 1, 2, \ldots, k$, are in ϵ-regular cylinders of $\mathcal{K}$.*

The proof of this Lemma, which as in Lemma 3.4 involves refining partitions and computing an index for each of them, can be combined with Corollary 3.3, much as in the proof of Theorem 1.3 to obtain the following constructive version:

Lemma 5.2 *Let ϵ be a positive constant and let k be a fixed positive integer. Set $\gamma = \frac{\epsilon^4}{16}$ and $h = 2^8 \cdot \binom{k}{2} \cdot \epsilon^{-17}$.*

Then for each k-partite graph $G = (V, E)$, $V = \cup_{1 \le i \le k} V_i$, $|V_i| = N$, $i = 1, 2, \ldots, k$, with $\gamma^{h+1} \cdot N \ge 1$, there exists a partition $\mathcal{K}$ of $V_1 \times V_2 \times \ldots \times V_k$ into q cylinders such that

(*i*) $q \le 4^h$, *and*

(*ii*) *all but $\epsilon \cdot N^k$ of the k-tuples $(v_1, v_2, \ldots, v_k)$, $v_i \in V_i$, $i = 1, 2, \ldots, k$, are in ϵ-regular cylinders of $\mathcal{K}$.*

Moreover, there is an algorithm whose input is a graph G as above which produces such a partition in $O\left(\binom{k}{2} \cdot 4^{h+1} \cdot M(N)\right)$ sequential time (where $M(N)$ is the time for multiplying two N by N matrices with entries $0, 1$ over the integers).

While the number of cylinders in the partition produced by this algorithm is still exponential in $\frac{1}{\epsilon}$, it is very small in comparison with the upper bound on the number of classes in the ϵ-regular partition of the set of vertices which is guaranteed by the Regularity Lemma of Szemerédi.

As an application of the algorithm in Lemma 5.2 we will briefly consider a result concerning the number of subgraphs of a given graph G on n vertices which are isomorphic to a given labeled graph H on k vertices, where $k < \left(\frac{1}{17} \cdot \log\log n\right)^{\frac{1}{2}}$.

Again we need some definitions. Let $G = (V, E)$, $V = \cup_{1 \le i \le k} V_i$ be a k-partite graph. For each i and j, $1 \le i < j \le k$, we set $d_{i,j}$ equal to the density of the pair (V_i, V_j). That is, $d_{i,j} = d(V_i, V_j)$. Let $H = (W, F)$ be a graph whose k-element vertex set $W = \{w_1, w_2, \ldots, w_k\}$ is ordered by $w_1 < w_2 < \ldots < w_k$. We say that an induced subgraph H' of G is *partite-isomorphic* to H if $H' = (W', F')$, where $W' = \{w_1', w_2', \ldots, w_k'\}$, $w_i' \in V_i$, $i = 1, 2, \ldots, k$, and the mapping $w_i \longmapsto w_i'$ is an isomorphism. For each such choice of $G = (\cup_{1 \le i \le k} V_i, E)$ and $H = (W, F)$ and each i and j, $1 \le i < j \le k$, set

$$\hat{d}_{i,j} = \begin{cases} d_{i,j} & \text{if } (w_i, w_j) \in F \\ 1 - d_{i,j} & \text{otherwise.} \end{cases}$$

Let $f_<(H, G)$ denote the number of induced subgraphs of G which are partite-isomorphic to H.

Lemma 5.3 *Let $G = (V, E)$, $V = \cup_{1 \le i \le k} V_i$, be a k-partite graph and let δ be a positive constant with $\delta < \frac{1}{k}$.*

If for each i and j, $1 \le i < j \le k$, the pair (V_i, V_j) is ϵ-regular, where $\epsilon = \left(\frac{\delta}{k}\right)^2$, then we have

$$\left| \, f_<(G, H) - \prod_{1 \le i < j \le k} \hat{d}_{i,j} \cdot \prod_{1 \le i \le k} |V_i| \, \right| \le \delta \cdot \prod_{1 \le i \le k} |V_i| \; .$$

Now let $G = (V, E)$ be any labeled graph on n vertices and suppose that the vertex set of G, $V = \{v_1, v_2, \ldots, v_n\}$, is ordered by $v_1 < v_2 < \ldots < v_n$. Let the set $W = \{w_1, w_2, \ldots, w_k\}$, $k \le n$, be ordered by $w_1 < w_2 < \ldots < w_k$. We say that a graph H with vertex set W is *order-isomorphic* to an induced subgraph H' of G with vertex set W' if there exists an isomorphism $\varphi\colon W \longrightarrow W'$ with the property that for each i and j, $1 \le i < j \le k$, $w_i < w_j$ implies $\varphi(w_i) < \varphi(w_j)$. Let $H_1, H_2, \ldots, H_t$, $t = 2^{\binom{k}{2}}$, be a list of all labeled graphs on the set W. We shall use $\sigma_k(G)$ to denote the t-dimensional vector $\sigma_k(G) = (h_1, h_2, \ldots, h_t)$, in which for each i, $1 \le i \le t$, h_i is the number of induced subgraphs of G to which H_i is order-isomorphic. Lemma 5.2 and Lemma 5.3 can now be combined to yield an efficient algorithm which computes a vector $\bar{\sigma}_k(G) = (\bar{h}_1, \bar{h}_2, \ldots, \bar{h}_t)$ which approximates $\sigma_k(G)$ when k is sufficiently small relative to n. Specifically we have the following result [10]:

Theorem 5.4 *There is an algorithm whose input is a labeled graph G with n vertices, n sufficiently large, and a list $(H_1, H_2, \ldots, H_t)$, $t = 2^{\binom{k}{2}}$, of all labeled graphs on k vertices for $3 \le k < \left(\frac{1}{17} \cdot \log\log n\right)^{\frac{1}{2}}$, and whose output is an approximation $\bar{\sigma}_k(G) = (\bar{h}_1, \bar{h}_2, \ldots, \bar{h}_t)$ to the vector*

$\sigma_k(G) = (h_1, h_2, \ldots, h_t)$ with the property that

$$\sum_{1 \le i \le t} |h_i - \bar{h}_i| \le \frac{1}{t} \sum_{1 \le i \le t} h_i \ .$$

This algorithm runs in $O\left(n^2 \cdot M(n)\right)$ sequential time.

We will briefly outline a proof of this result which relies on the algorithm given in Lemma 5.2 for finding a regular partition of the type described there.

Proof (Outline) Suppose we are given a labeled graph $G = (V, E)$, $|V| = n$, with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ ordered by $v_1 < v_2 < \ldots < v_n$ and a list $(H_1, H_2, \ldots, H_t)$, $t = 2^{\binom{k}{2}}$, of all labeled graphs on k vertices, where $k < \left(\frac{1}{17} \cdot \log\log n\right)^{\frac{1}{2}}$. Set

$$\epsilon = \left(\frac{2^8 \cdot k^2 \cdot \log\log n}{\log n}\right)^{\frac{1}{17}} \ . \tag{15}$$

Partition V into $m = \left\lceil \frac{k^2}{\epsilon} \right\rceil$ parts $V_1, V_2, \ldots, V_m$, each of size $N = \frac{n}{m}$ with

$$V_j = \left\{ v_{(j-1)\cdot N+1}, v_{(j-1)\cdot N+2}, \ldots v_{j \cdot N} \right\}$$

for $j = 1, 2, \ldots, m$. By the choice of m at most $\epsilon \cdot n^k$ k-tuples of vertices from V have more than one point in a single one of the sets V_l, $1 \le l \le m$.

Choosing γ and h as in Lemma 5.2, it can be checked that for sufficiently large n, say n for which $\log\log n < (\log n)^{\frac{1}{2}}$, we have $\gamma^{h+1} \cdot N \ge 1$. This was required in Lemma 5.2 to insure that the sizes of the pairs to be checked by the algorithm in Corollary 3.3 remain sufficiently large until the algorithm halts.

For each of the $\binom{m}{k}$ choices of k-sets $V_{i_1}, V_{i_2}, \ldots, V_{i_k}$ with $i_1 < i_2 < \ldots < i_k$ we use the algorithm of Lemma 5.2 to find a partition $\mathcal{K}$ of $V_{i_1} \times V_{i_2} \times \ldots \times V_{i_k}$ into at most 4^h cylinders with the property that at most $\epsilon \cdot N^k$ of the k-tuples in $V_{i_1} \times V_{i_2} \times \ldots \times V_{i_k}$ are not in ϵ-regular cylinders of the k-partite subgraph of G induced by $\cup_{1 \le j \le k} V_{i,j}$. For each ϵ-regular cylinder of each such partition and each labeled graph H_j, $1 \le j \le t$, we use the formula of Lemma 5.3 to obtain an estimate of the number of induced subgraphs of G which have as vertices a k-tuple in that cylinder and which are partite isomorphic to H_j. For each j, $1 \le j \le t$, we add these estimates to obtain the entry $\bar{h}_j$ of $\bar{\sigma}_k(G)$. It can be shown that the error $|h_j - \bar{h}_j|$ is less than $2 \cdot k \cdot n^k \sqrt{\epsilon}$ for $j = 1, 2, \ldots, t$. For $k < \left(\frac{1}{17} \cdot \log\log n\right)^{\frac{1}{2}}$ and our choice of ϵ we have that $2 \cdot k \cdot \sqrt{\epsilon} \cdot n^k < 2^{-\binom{k}{2}} \cdot n^k$ for sufficiently large n, from which we obtain the inequality

$$\sum_{1 \le i \le t} |h_i - \bar{h}_i| \le \frac{1}{t} \cdot \sum_{1 \le i \le t} h_i \ .$$

Finally we indicate that this procedure can be implemented in the asserted time. For each of the $\binom{m}{k}$ choices of the products $V_{i_1} \times V_{i_2} \times \ldots \times V_{i_k}$ we use the algorithm of Lemma 5.2. Computing the various densities required to use Lemma 5.3 can be clearly done in running time $O(n^2)$. By our choices of m and h we have $\binom{m}{k} = O(n)$ and $4^{h+1} = O(n)$ (h according to Lemma 5.2). Summarizing, this gives the total running time $O\left(n^2 \cdot M(n)\right)$. $\square$

Our general approach to approximating the number of copies of H in G, for an ϵ-regular k-partite graph G, is quite similar to methods already described in [13], [24], and in abstract form in [26]. Indeed our algorithm which supplies such approximations could be obtained by combining these results with Theorem 1.3. We note, however, that due to the way in which the bound $Q(\epsilon, t)$ of that Theorem depends on $\frac{1}{\epsilon}$, an algorithm of this sort would only achieve the sort of accuracy obtained by Theorem 5.4 for subgraphs on k vertices with $k = \log^{(p)} n$, where $\log^{(p)}$ means the p-fold iterated logarithm, and p is in this case a polynomial in $\frac{1}{\epsilon}$.

The variant of the Regularity Lemma given in Lemma 5.1 is in some sense nearly optimal. It can be shown that there exists a bipartite graph $G = (V, E)$, $V = V_1 \cup V_2$, $|V_1| = |V_2| = N$, with the property that any partition of $V_1 \times V_2$ which is such that at most $\epsilon \cdot N^2$ pairs are not in ϵ-regular cylinders must contain at least $c \cdot 2^{\frac{1}{\sqrt{\epsilon}}}$ cylinders for some positive constant $c > \frac{1}{4}$.

References

[1] N. Alon, *Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory*, Combinatorica 6 (1986), 207-219.

[2] N. Alon, *Non-constructive proofs in combinatorics*, Proc. of the International Congress of Mathematicians, Kyoto 1990, Japan, Springer Verlag, Tokyo (1991), 1421-1429.

[3] N. Alon, *A parallel algorithmic version of the local lemma*, Random Structures and Algorithms 2 (1991), 367-378. Also: Proc. 32^{nd} Annual FOCS, IEEE (1991), 586-593.

[4] N. Alon, L. Babai and A. Itai, *A fast and simple randomized parallel algorithm for the maximal independent set problem*, J. Algorithms 7 (1986), 567-583.

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

[6] J. Beck, *An algorithmic approach to the Lovász Local Lemma I*, Random Structures and Algorithms 2 (1991), 343-365.

[7] B. Bollobás, P. Erdös, M. Simonovits and E. Szemerédi, *Extremal graphs without large forbidden subgraphs*, in: Advances in Graph Theory (Cambridge Combinatorial Conference, Trinity College, 1977), Ann. Discrete Math. 3 (1978), 29-41.

[8] V. Chvátal, V. Rödl, E. Szemerédi and T. Trotter, *The Ramsey Number of a graph with bounded maximum degree*, J. Combinatorial Theory Ser. B 34 (1983), 239-243.

[9] D. Coppersmith and S. Winograd, *Matrix multiplication via arithmetic progressions*, J. of Symbolic Computation 9 (1990), 251-280.

[10] R. A. Duke, H. Lefmann and V. Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs in a given graph*, preprint.

[11] N. Eaton and V. Rödl, *Ramsey numbers for sparse graphs*, preprint.

[12] P. Erdös, P. Frankl and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs and Combinatorics 2 (1986), 113-121.

[13] F. Franek and V. Rödl, *On Erdös conjecture for multiplicities of complete subgraphs*, Graphs and Combinatorics, to appear.

[14] P. Frankl, R. L. Graham and V. Rödl, *On subsets of Abelian groups with no 3-term arithmetic progression*, J. Combinatorial Theory Ser. A 45 (1987), 157-161.

[15] P. Frankl and Z. Füredi, *Exact solution of some Turán-type problems*, J. Combinatorial Theory Ser. A 45 (1987), 226-262.

[16] Z. Füredi, *The maximum number of edges in a minimal graph of diameter 2*, J. Graph Theory, to appear.

[17] Z. Füredi, M. Goemans and D. J. Kleitman, *On the maximum number of triangles in wheel-free graphs*, to appear.

[18] M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company. New York, 1979.

[19] A. Hajnal, W. Maass and G. Turán, *On the communication complexity of graph properties*, Proc. 20th ACM STOC, ACM Press (1988), 186-191.

[20] R. M. Karp and A. Wigderson, *A fast parallel algorithm for the maximal independent set problem*, Proc. 16th ACM STOC, ACM Press (1984), 266-272.

[21] M. Luby, *A simple parallel algorithm for the maximal independent set problem*, SIAM J. Computing 15 (1986), 1036-1053.

[22] C. H. Papadimitriou, *On graph theoretic lemmata and complexity classes*, Proc. 31st Annual IEEE FOCS, IEEE (1990), 794-801. See also: *On the complexity of the parity argument and other inefficient proofs of existence*, JCSS, to appear.

[23] V. Rödl and R. Duke, *On Graphs with small subgraphs of large chromatic number*, Graphs and Combinatorics 1 (1985), 91-96.

[24] V. Rödl, *On universality of graphs with uniformly distributed edges*, Discrete Mathematics 59 (1986), 125-134.

[25] R. Rödl and A. Rucinski, *Random Graphs with monochromatic triangles in every edge coloring*, submitted.

[26] M. Simonovits and V. T. Sós, *Szemerédi's partition and quasirandomness*, Random Structures and Algorithms 2 (1991), 1-10.

[27] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica 27 (1975), 199-245.

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