Bundle decompositions of graphs with algorithmic applications

Béla Csaba*

Abstract

We use a recent edge decomposition method of the author for finding good approximation algorithms for special cases of the 2-EC, 2-VC and TSP(1, 2) problems.

1 Introduction

The Regularity lemma is among the most powerful tools in graph theory. The lemma appeared in [22] in 1978, a weaker version was already used earlier by Szemerédi in [21] to prove the Erdős-Turán conjecture [11].

In order to state the lemma we need some definitions. Let $G = (V, E)$ be a (simple[1]) graph, and A, B two disjoint subsets of its vertices. Then the *bipartite density* $d_G(A, B)$ or, if G is clear from the context, $d(A, B)$, is defined as follows: $d(A, B) = \frac{|E(A,B)|}{|A| \cdot |B|}$, here $E(A, B)$ stands for the set of edges with one endpoint in A and the other in B.

Given a number $\varepsilon \in (0, 1)$ we say that the (A, B) pair is ε-*regular* if for every $A' \subset A$, $|A'| \geq \varepsilon|A|$ and $B' \subset B$, $|B'| \geq \varepsilon|B|$ we have $|d(A, B) - d(A', B')| \leq \varepsilon$. An equitable partition of a graph G on n vertices is a partition of V such that $V(G) = V_1 \cup \ldots \cup V_k$ and $||V_i| - |V_j|| \leq 1$ for every $1 \leq i, j \leq k$. The Regularity lemma states that for every $\varepsilon > 0$ there is a threshold $N = N(\varepsilon)$ such that every graph G has an equitable partition into $k \leq N$ parts, the clusters, with the additional property that all but at most εk^2 pairs (V_i, V_j) are ε-regular for $1 \leq i, j \leq k$.

The threshold N for the number of parts is a tower of twos of height $O(\varepsilon^{-5})$. Unfortunately, as Gowers proved in [15], this is unavoidable in general, more precisely, there are graphs for which N has to be at least a tower of twos of height $\Omega(\varepsilon^{-1/16})$. This implies that in order to get meaningful results when using the Regularity lemma, one has to work with "astronomically large" graphs, thereby ruling out practical applications of the lemma. We remark that if the number of edges between vertex sets A and B is $o(|A| \cdot |B|)$, then the (A, B)-pair is ε-regular by definition, unless ε is very small compared to $|A|$ and $|B|$. Hence, the lemma is useful only when the density of the graph is essentially a positive constant.

There are several attempts to try to find good alternatives for the Regularity lemma. Frieze and Kannan proved a so called weak regularity lemma [12], Eaton and Rödl proved a cylindrical regularity lemma [10, 9], there are versions for C_4-free graphs [4], for graphs having bounded VC-dimension [2], for low threshold rank graphs, or for so called cut pseudorandom graphs [3]. All these results provide better bounds, but there is a drawback: they cannot be applied in so diverse areas as the original Regularity lemma.

*Bolyai Institute, University of Szeged, Szeged, Hungary. This research was partially supported by NKFIH grant KH 129597 and by the grant NKFIH-1279-2/2020 of the Ministry for Innovation and Technology, Hungary.

[1]We only consider simple graphs in this paper.

One of the main area of applications of the Regularity lemma is extremal graph theory. The Regularity lemma–Blow-up lemma method proved to be especially useful in embedding problems. In this paper we present a method, based on a paper by the author [5], which in certain cases can replace it, applicable for real-life sized graphs, and the density can be vanishing. We demonstrate the use of our method via the solution of some algorithmic questions.

We made no attempts to optimize on the constants in the paper, and will not be concerned with floor signs and divisibility in the proofs. This makes the notation simpler, easier to follow. Throughout the paper $\log x$ will denote the *natural logarithm* of x.

2 Preparations

2.1 Notations, definitions

Given a graph $G = (V, E)$ we use the notations $v(G) = |V|$ and $e(G) = |E|$. For disjoint subsets $X, Y \subset V$ we let $G[X, Y]$ denote the bipartite subgraph of G with parts X and Y that contains all the edges of G with one endpoint in X and the other endpoint in Y.

The *density* of G is defined to be $d_G = e(G) \cdot \binom{v(G)}{2}^{-1}$. Recall, that the *bipartite* density of bipartite subgraphs of G with parts A and B is defined to be $d_G(A, B) = \frac{e(G[A,B])}{|A| \cdot |B|}$. Sometimes the subscript may be omitted when there is no confusion. Similarly, when a graph in question is bipartite, density will mean bipartite density, unless stated otherwise.

Recall, that a bipartite graph $H = (A \cup B; E)$ is ε-regular for a real number $\varepsilon > 0$, if

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

whenever $X \subset A$ and $Y \subset B$ such that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$. For $\varepsilon, \delta > 0$ we call H an (ε, δ)-*super-regular*[2] pair, if it is ε-regular, and every $v \in A$ has at least $(\delta - \varepsilon)|B|$ neighbors and similarly, every $u \in B$ has at least $(\delta - \varepsilon)|A|$ neighbors.

Let $\mathcal{B}$ denote the class of balanced bipartite graphs, that is, bipartite graphs having equal-sized parts, and let $\mathcal{B}_m$ denote the class of balanced bipartite graphs having m vertices in both parts. We call a graph $H \in \mathcal{B}_m$ an (ε, δ)-*bundle,* if it is an ε-regular pair, and for every vertex v of H we have that $(\delta - \varepsilon)m \leq deg(v) \leq (\delta + \varepsilon)m$. We have the following.

Claim 2.1. *Let $0 < \varepsilon < 1/4$ and $3\varepsilon \leq d < 1$, and assume that $F \in \mathcal{B}_m$ is an ε-regular pair with density d. (i) Then F has a subgraph $H_1 \in \mathcal{B}_{m_1}$ such that H_1 is a $(2\varepsilon, d)$-super regular pair with density $d \pm \varepsilon$, and $m_1 \geq (1 - \varepsilon)m$. (ii) There is a $(3\varepsilon, d)$-super-regular pair $H_2 \subset F$, $H_2 \in \mathcal{B}_{m_2}$ with density $d \pm \varepsilon$ such that $m_2 \geq (1 - 2\varepsilon)m$.*

Proof: First we consider part (i). By ε-regularity, F may have at most εm vertices in both its vertex parts which have degree less than $(d - \varepsilon)m$. Discard these vertices from both vertex classes. Next discard some vertices from the larger part in order both parts have the same cardinality. The resulting subgraph is H_1. After these steps every remaining vertex will have at least $(d - 2\varepsilon)m$ neighbors in the opposite part, the parts will have size $m' \geq (1 - \varepsilon)m$. Moreover, since $2\varepsilon m' > \varepsilon m$,

[2]We remark that in some earlier papers, eg., in [20] super-regularity meant a somewhat different notion: ε-regularity was replaced by the condition that between any two sets $X \subset A$, $Y \subset B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $d(X, Y) \geq \delta$. This was also used by Fox [13] for proving his breakthrough result on the removal lemma. In order to avoid confusion, we call this notion (ε, δ)-super-lower-regularity.

between any two subsets of size at least $2\varepsilon m'$ each, the density of the subgraph is $d\pm\varepsilon$, by ε-regularity of the original pair.

For part (ii) we first discard the vertices that have less than $(d-\varepsilon)m$ neighbors, and then those, which have more than $(d+\varepsilon)m$ neighbors. We discarded at most $2\varepsilon m$ vertices from both parts of F. Then, as above, we discard some further vertices from the larger part to make the subgraph balanced. This way we have obtained H_2, which has $m_2 \geq (1-2\varepsilon)m$ vertices in both parts. Clearly, if we take arbitrary subsets of sizes at least $3\varepsilon m_2 \geq \varepsilon m$ from both parts, the density of the subgraph is $d \pm \varepsilon$ by ε-regularity of F. □

The following is Theorem 2.8 in [16], we will refer to it as Chernoff's bound.

Theorem 2.2. *Assume that X is the sum of n independent indicator random variables: $X = X_1 + \ldots + X_n$. If $0 \leq \lambda \leq 3/2$, then*

$$P(|X - \mathbb{E}[X]| \geq \lambda\mathbb{E}[X]) \leq 2e^{-\frac{\lambda^2}{3}\mathbb{E}[X]}.$$

Similar bound holds in the case of hypergeometric distribution. For that let S be an n element set, $Q \subset S$ an m element set, and $R \subset S$ be a random subset with cardinality k, where $k, m \leq n$. Theorem 2.10 in [16], which we will refer to as Hoeffding's bound, states the following.

Theorem 2.3. *Let S, Q and R as above, and let $X = |R \cap Q|$. We have that $\mathbb{E}[X] = km/n$, and if $0 \leq \lambda \leq 3/2$, then*

$$P(|X - \mathbb{E}[X]| \geq \lambda\mathbb{E}[X]) \leq 2e^{-\frac{\lambda^2}{3}\mathbb{E}[X]}.$$

Theorem 2.2 and 2.3 will be used in cases when we will have that $\mathbb{E}[X] = \Omega(n/\log n)$ and $\lambda = \Omega(1/\log n)$. Hence, the probability bounds we obtain will be less than $1/n^{10}$ always. As the number of bad events will be at most n^3, the total probability of bad events will tend to zero as n tends to infinity.

2.2 An algorithmic edge decomposition theorem

The following result was proved by the author [5].

Theorem 2.4. *Let G be a balanced bipartite graph on $2n$ vertices with density δ_G, and let $0 < d \leq 1$ and $0 < \varepsilon < 1/4$ be real numbers such that the following is satisfied: $e^{k/4}2n^{-1/4} < \varepsilon \ll d \leq 1$, where $k = (h/\varepsilon^{12})\log(h/\varepsilon^{12})$ and $h = 3^{12} \cdot 16^3$. Then the* edge set *of G can be decomposed as follows: $E(G)$ can be written as the edge-disjoint union of the (ε, d)-bundles $F_1, \ldots, F_K$, and another balanced bipartite graph F_0, where $K = K(\varepsilon, \delta_G, d) \leq \delta_G e^k/d$. For $i \geq 1$ each F_i has at least $m = m(\varepsilon) \geq e^{-k}n$ vertices and density at least d, while F_0 has density less than d.*

The main difference of the above result and the Regularity lemma is that in Theorem 2.4 we partition the edge set of the graph, not the vertex set. This hardens the use of the theorem. The main result of the present paper, discussed in the next section, is that in some cases partition of the edge set can be turned into partition of the vertex set so that quasirandomness remains in certain cluster pairs.

Let us remark, that in [5] another decomposition theorem is also proved, in which the exponent of ε is smaller in the expression of r. However, the proof of that result is existential, therefore, we use the above theorem in order to be able to find effective approximation algorithms.

3 Imitating the Regularity lemma in an important special case

While bundle decomposition can be obtained for graphs having sublinear average degree, and the number of vertices does not have to be astronomically large, it seems that the Regularity lemma of Szemerédi is much more powerful. The problem with bundle decompositions is that the clusters (vertex parts) of different bundles may intersect, a vertex usually appears in several bundles of the decomposition. Unlike the Regularity lemma, bundle decomposition does not produce vertex partitioning – if one wants to arrange the majority of edges in bundles then this is of course not possible in general by the Gowers bound [15].

However, in many graph embedding problems it is enough to find some substructure of the host graph, e.g., a large matching, in which the vertices are in regular pairs. Let us give a general overview of the so called Regularity Method, which is the combined application of the Regularity lemma and a counting or an embedding lemma, e.g., the Blow-up lemma [18, 19, 17]. Assume that G, F are graphs, and we want to show that $F \subset G$. When F is spanning or almost spanning, the Regularity Method works roughly as follows.

1. Apply the Regularity lemma for G with appropriate parameters $0 < \varepsilon \ll d \ll 1$, these may depend on F.
2. Construct the *reduced graph:* its vertices are the non-exceptional clusters, and two clusters are adjacent if we have an ε-regular pair between them having density at least d.
3. Find a special substructure in the reduced graph – this can be a large matching, a triangle-factor, etc., depending on F.
4. Distribute the vertices of F among the clusters of the reduced graph such that no cluster receives more vertices than its size, furthermore, adjacent vertices of F are assigned to adjacent clusters. Here the above special substructure plays an important role.
5. Use the Blow-up lemma [19] for finishing the embedding procedure.

There are other details we haven't touched upon, but the big picture is essentially the above. The method was developed by Komlós, Sárközy and Szemerédi in [**?**, **?**, **?**], see also [20, 19].

Our goal in this section is to find a subgraph of G, using Theorem 2.4, that contains *vertex disjoint* bundles, so that we will be able to imitate the Regularity Method for embedding bipartite graphs.

The host graph G, for which we apply bundle decomposition, cannot be arbitrary. We assume that G is an r-regular n-vertex graph, where $r = r(n)$ is a natural number. This is not a significant restriction in many applications, see e.g. [8]. In Section 4 we will show some algorithmic applications of this method, in which the Regularity lemma is replaced by Theorem 2.4 and another statement, a corollary of it, which will be discussed in the next section.

3.1 An easy-to-apply form for r-regular graphs

Theorem 3.1. *Assume that $0 < d < 1$ is a real number, $\kappa \geq 1$ is an integer, and G is a balanced bipartite graph on $2n$ vertices with parts U and V such that for every vertex $v \in U \cup V$ we have $deg(v) = r$ for some $r = r(n) \in \mathbb{N}$ with $r/n \gg d$. Set $\varepsilon = d^{4\kappa+2}$. Assume further that $e^{M/4}2n^{-1/4} < \varepsilon$, where $M = (h/\varepsilon^{12})\log(h/\varepsilon^{12})$ and $h = 3^{12} \cdot 16^3$. Then there exist partitions $U = U_0 \cup U_1 \cup \ldots \cup U_K$*

and $V = V_0 \cup V_1 \ldots \cup V_K$ *such that* $|U_0 \cup V_0| \leq 10d^{1/4}n$*, and for every* $1 \leq i \leq K$ *we have that* $|U_i| = |V_i| \geq \varepsilon^2 e^{-M} n/d$ *and the* (U_i, V_i) *pair is a* d^{κ}*-regular pair with density at least* $d/2 - 10\varepsilon/d^2$*. The number of pairs in the decomposition is at most* $K \leq \frac{14\eta d}{\varepsilon^2}\frac{r}{n}e^M$*.*

Roughly speaking, Theorem 3.1 asserts that if every degree is the same in a bipartite graph, then one can partition the vast majority of the vertex set so that opposite clusters having the same index give a θ-regular pair for some sufficiently small θ. Note that in several proofs in extremal graph theory the Szemerédi Regularity lemma is used for finding the same substructure what is provided by Theorem 3.1, that is, a perfect (or almost perfect) matching in the so called reduced graph. Given this matching one can embed various kind of bipartite graphs (spanning trees, collection of even cycles, etc.) into the host graph with the help of the Blow-up lemma.

In order to prove Theorem 3.1 we use a randomized algorithm. The rough sketch of it is as follows. (1) Apply Theorem 2.4 with appropriate parameters, (2) randomly partition the edge sets of the bundles so that we obtain bundles with the same density, (3) using that G is r-regular and every bundle has the same density, we are able to distribute the vast majority of vertices into vertex-disjoint regular pairs.

It is important to note that the requirement $d \ll r/n$ is essential, otherwise the whole edge set of G would be lost when applying Theorem 2.4. Next we give further details of the above method, and prove its correctness.

3.1.1 Step 1: Decomposing G into bundles

Let $0 < d \ll r/n \leq 1$ be numbers that satisfy the conditions of Theorem 3.1, and let κ be a positive integer. Set $\varepsilon = d^{4\kappa+2}$, $\nu = d^{4\kappa}$, and $\eta = d^{1/4}$. Apply Theorem 2.4 with parameters ε and d obtaining the bundles $H_1, \ldots, H_K$ and the sparse subgraph H_0. Here for every $i \geq 1$ we have that H_i is an (ε, δ_i)-bundle having $m_i \geq m_b$ vertices in both parts, and $\delta_i \geq d$ while H_0 has density less than d and edge set $E(H_0) = E(G) - \cup_{i=1}^{K} E(H_i)$.

For later purposes we introduce the initially empty set W_0 of so called *exceptional* vertices into which we will put some vertices during the algorithm, but it will never be larger than $10\eta n$.

3.1.2 Step 2: Achieving that every bundle has the same density

After Step 1 we may have very dense and sparse bundles as well, the density could be as low as $d - \varepsilon$ and as high as 1. We randomly partition the edge sets of the bundles so that the resulting bundles will have density precisely $d' = d/2$.

This goes as follows. Let H be a bundle with parts Q and S, having $e_H = k \cdot d' m_H^2 + \rho m_H^2$ edges, where $|Q| = |S| = m_H$, k is a positive integer and $0 \leq \rho < d'$.

Choose a random subset of $E(H)$ of size ρm_H^2, and put it into H_0. The rest of the edges of $E(H)$ are randomly partitioned into k equal sized edge sets, so each has exactly $d' m_H^2$ edges. These edge sets are the edge sets of the new bundles $H^{(1)}, \ldots, H^{(k)}$ which we obtain from H, all of them have parts Q and S. It is easy to see that $k \leq 1/d' = 2/d$. Using Hoeffding's inequality we get that with high probability we obtained $(2\varepsilon, d')$-bundles this way. The above procedure is repeated for every bundle in the decomposition.

After renaming the bundles we obtain the following decomposition: $E(G) = E(H_0) \cup E(H_1) \cup \ldots \cup E(H_{K'})$, where $K' \leq 2K/d$, for every $1 \leq i \leq K'$ we have that H_i is a $(2\varepsilon, d')$-bundle, and the density of H_0 is at most $3d/2$.

3.1.3 Step 3: Adjusting the sizes of the parts

Let H be a bundle which we obtained in the previous step, and denote its parts by Q and S. The density of H is precisely d', and $|Q| = |S| = m_H \geq m_b$. Assume that $m_H = k\varepsilon m_b + \rho m_b$, where $0 \leq \rho < \varepsilon$ and k is a positive integer.

Choose random subsets $Q_0 \subset Q$ and $S_0 \subset S$ such that $|Q_0| = |S_0| = \rho m_b$. Let $Q' = Q - Q_0$ and $S' = S - S_0$. Hoeffding's inequality shows[3] that with high probabiliy $H[Q', S']$ is a $(3\varepsilon, d')$-bundle, and the density of $H[Q', S']$ is in the interval $[d' - \varepsilon, d' + \varepsilon]$.

We add every edge in $H[Q_0, S_0] \cup H[Q_0, S'] \cup H[Q', S_0]$ to H_0. Since $H[Q, S]$ was a $(2\varepsilon, d')$-bundle, we add at most $2\varepsilon(d' + 2\varepsilon)m_H^2 < 3\varepsilon e(H)$ edges to H_0 this way.

The above procedure is repeated for every bundle in the decomposition. At the end we have that the density of H_0 is less than $3\varepsilon + 3d/2 < 2d$, since from every bundle less than a 3ε proportion of edges is added to H_0.

3.1.4 Step 4: Randomly distributing the vertices

At this point for every bundle $H_i[Q_i, S_i]$, where $1 \leq i \leq K'$, we have precisely $k_i \varepsilon m_b$ vertices in both parts where the k_i numbers are positive integers. For every $1 \leq i \leq K'$ we define k_i pairs of initially "empty boxes" A_j^i, B_j^i, where $j = 1, 2, \ldots, k_i$. These boxes will be "filled up" by randomly chosen vertices from the bundles.

Next we define an auxiliary bipartite graph $\mathcal{D}$. One of its parts is $U \cup V$, the set of vertices of G. The vertices of the other part are identified by the empty boxes defined above. Hence, one part of $\mathcal{D}$ has n vertices, the other part has $2\sum_{i=1}^{K'} k_i$ vertices. Given a $u \in U$ we join it to every A_j^i boxes if $v \in Q_i$, and similarly, we join $v \in V$ to every B_j^i boxes if $v \in S_i$, here $1 \leq j \leq k_i$. Hence, for every $w \in U \cup V$ we have

$$deg_{\mathcal{D}}(w) = \sum_{i: w \in Q_i \cup S_i} k_i.$$

Note that $\mathcal{D}$ is the disjoint union of two bipartite graphs, since G is bipartite.

Next for every $w \in U \cup V$ we randomly, independently of other choices, choose a neighboring box in $\mathcal{D}$, and insert w into this box. This way for every i we will obtain the bipartite subgraphs $H_i[A_j^i, B_j^i]$ from the bundle H_i.

3.1.5 Step 5: Preparations and discarding some vertices to W_0

First we need a definition for describing this step. We call a vertex $w \in U \cup V$ *bad,* if, after Step 3, w is incident to at least ηr edges of H_0. Denote $\mathcal{R}$ the set of bad vertices. The set $V - \mathcal{R}$ includes the *good* vertices. We call a bundle *bad*, if after Step 3 at least an η proportion of its vertices belong to $\mathcal{R}$. If a bundle is not bad, we call it *good.* Finally, all boxes of bad bundles are called *bad boxes.*

We begin this step with putting every bad vertex into W_0. Next we also put the vertices of bad boxes into W_0, even though some of these vertices could be good ones. At this point the box pairs (A_j^i, B_j^i) are γ-regular for some γ, which is perhaps too large for us. This is due to the fact that the degrees in $\mathcal{D}$ are not necessarily sufficiently close to the average degree of $\mathcal{D}$, hence, we will need one further step, and before that, another definition.

[3] In fact 3ε-regularity follows easily from the definition of regular pairs, we need Hoeffding's inequality only for bounding the degrees.

Recall, that we set $\nu = d^{4k}$. Let

$$\delta_{\mathcal{D}} = \min\{deg_{\mathcal{D}}(v) : v \in U \cup V \text{ is good}\}$$

and

$$\Delta_{\mathcal{D}} = \max\{deg_{\mathcal{D}}(v) : v \in U \cup V \text{ is good}\}.$$

We distribute the good vertices of $U \cup V$ into the classes $\mathcal{C}_1, \dots, \mathcal{C}_h$, where

$$\mathcal{C}_g = \{v \in U \cup V : (1+\nu)^{g-1} \leq \frac{deg_{\mathcal{D}}(v)}{\delta_{\mathcal{D}}} < (1+\nu)^g,\ v \text{ is good}\}$$

for $1 \leq g \leq h$, later we will show that $h \leq 7\eta/\nu$. Note that $(1+\nu)^{h-1}\delta_{\mathcal{D}} \leq \Delta_{\mathcal{D}} < (1+\nu)^h \delta_{\mathcal{D}}$.

3.1.6 Step 6: Finding the final partition of U and V

Let us fix a pair of good boxes X and Y, where $X = A^i_j$ and $Y = B^i_j$ for some i and j. Without loss of generality assume that $|X| \geq |Y|$. The box Y is divided into the sub-boxes Y_g, $g = 1, \dots, h$, where $Y_g = Y \cap \mathcal{C}_g$.

Call a sub-box Y_g small, if $|Y_g| \leq d\nu|Y|$, otherwise the sub-box is called large. We discard the vertices of small sub-boxes into W_0. Denote the remaining set of sub-boxes by $Y'_1, \dots, Y'_l$. Next for every k we randomly choose a subset $X_k \subset X$ without replacement such that $|X_k| = |Y'_k|$ for $1 \leq k \leq l$. We discard the vertices of $X - \cup_1^l X_k$ into W_0. The d^{κ}-regular pairs we are looking for are the $H[X_k, Y'_k]$ subgraphs.

After repeating the above procedure for every $X = A^i_j, Y = B^i_j$ pair of good boxes ($1 \leq i \leq K'$ and $1 \leq j \leq k_i$) we obtain the desired decomposition of G, as we will prove next.

3.2 Proof of Theorem 3.1 – an analysis of the above algorithm

The proof of Theorem 3.1 is divided into two parts. In the first part we assume that we are after Step 3, right before Step 4, while in the second one we analyze the effect of the final three steps of the algorithm.

3.2.1 The first part of the proof

As is indicated above, in this subsection we assume that the first three steps have been executed, but Step 4 have not started yet. Recall, that $\mathcal{R}$ denotes the set of bad vertices, and its complement, $V - \mathcal{R}$ includes the *good* vertices, here a vertex $w \in U \cup V$ is *bad,* if it is incident to at least ηr edges of H_0. Moreover, a bundle is called *bad,* if an η proportion of its vertices belong to $\mathcal{R}$, otherwise it is called *good.*

Claim 3.2. *We have* $|\mathcal{R}| \leq 4dn^2/(\eta r)$.

Proof: The number of edges in H_0 after Step 3 is less than $2dn^2$. Hence, we must have $|\mathcal{R}|\eta r \leq 2e(H_0) < 4dn^2$. Rearranging gives the claim. □

Lemma 3.3. *The total number of edges in bad bundles is less than* $15|\mathcal{R}|r/4\eta \leq 15dn^2/\eta^2$.

Proof: Let H be an arbitrary bad bundle having m vertices in both parts. Since it is bad, it has at least $\eta m/2$ bad vertices in one of its parts. The number of edges incident to these bad vertices is at least $\eta(d'-3\varepsilon)m^2/2 \geq e(H)\eta/3$, since $e(H) \leq (d'+3\varepsilon)m^2$. Therefore at least an $\eta/3$ proportion of edges must be incident to bad vertices in the bad bundles.

The total number of edges in the bundles is at least $4nr/5$. Denote the proportion of edges in bad bundles by x. Then we have the following inequality:

$$x\frac{\eta}{3} \cdot \frac{4nr}{5} \leq |\mathcal{R}|r.$$

From this we obtain that

$$x \leq \frac{15|\mathcal{R}|}{4\eta n}.$$

Using that the total number of edges in bundles is at most rn, we get the claimed bound for the total number of edges in bad bundles. □

We also define a subset of good vertices. We say that a good vertex w is *η-good,* if the total degree of w in bad bundles is at most $30dn/\eta^3 = 60d'n/\eta^3$, otherwise the good vertex is called *η-bad.* Using Lemma 3.3 one easily obtains the following:

Claim 3.4. *The number of η-bad vertices of G is less than ηn.*

Proof: Let b denote the number of η-bad vertices. Then the total degree of η-bad vertices in bad bundles is at most $b\frac{30dn}{\eta^3}$, which must be at most $2 \cdot 15dn^2/\eta^2$ by Lemma 3.3. Comparing the two expressions and rearranging gives the claim. □

Lemma 3.5. *Assume that $w \in U \cup V$ is a good vertex. Then we have*

$$\frac{(1-\eta)r}{(d'+3\varepsilon)\varepsilon m_b} \leq deg_{\mathcal{D}}(w) \leq \frac{r}{(d'-3\varepsilon)\varepsilon m_b}.$$

Proof: Since the bundles are edge-disjoint and w is a good vertex, we get the following

$$(1-\eta)r \leq \sum_{i:w\in V(H_i)} deg_{H_i}(w) \leq r.$$

The degree of w in bundle H_i is between $(d'-3\varepsilon)k_i\varepsilon m_b$ and $(d'+3\varepsilon)k_i\varepsilon m_b$, hence,

$$(1-\eta)r \leq \sum_{i:w\in V(H_i)} (d'+3\varepsilon)k_i\varepsilon m_b$$

and

$$\sum_{i:w\in V(H_i)} (d'-3\varepsilon)k_i\varepsilon m_b \leq r.$$

These imply that

$$\frac{(1-\eta)r}{(d'+3\varepsilon)\varepsilon m_b} \leq \sum_{i:w\in V(H_i)} k_i \leq \frac{r}{(d'-3\varepsilon)\varepsilon m_b}.$$

Since, as we noted earlier, $deg_{\mathcal{D}}(w) = \sum_{i:w\in V(H_i)} k_i$, this finishes the proof of the lemma. □

Later we have to estimate the number of η-good vertices which choose a bad box in Step 4.

Lemma 3.6. *Let us assume that w is an η-good vertex. Then the probability that we choose a bad box for w is less than*

$$100\frac{d'n}{\eta^3 r}.$$

Proof: The degree of w is $\mathcal{D}$ in at least

$$\frac{(1-\eta)r}{(d'+3\varepsilon)\varepsilon m_b}$$

by Lemma 3.5. By the definition of η-good vertices, w is incident to at most $60d'n/\eta^3$ edges in bad bundles. Let us consider a bad bundle having $m = k\varepsilon m_b$ vertices in its parts, out of which one contains w. Then w has at least $(d'-3\varepsilon)m$ neighbors in this bad bundle, and in the auxiliary graph $\mathcal{D}$ it is adjacent to all the k bad boxes of this bundle. That is, for k edges in $\mathcal{D}$ we "use up" at least $(d'-3\varepsilon)k\varepsilon m_b$ from the edges incident to w.

This implies that overall w is adjacent to at most

$$60\frac{d'n}{(d'-3\varepsilon)\eta^3\varepsilon m_b} < 70\frac{n}{\eta^3\varepsilon m_b}$$

bad boxes in $\mathcal{D}$. Hence, the probability of choosing a bad box is at most this latter expression divided by the degree in $\mathcal{D}$. This is at most

$$70\frac{n}{\eta^3\varepsilon m_b}\frac{(d'+3\varepsilon)\varepsilon m_b}{(1-\eta)r} < 100\frac{d'n}{\eta^3 r}.$$

□

We need that the probability bound in Lemma 3.6 is much smaller than 1, as the vertices that choose a bad box are put into the "wastebasket" W_0. Hence, we require that $d'n \ll \eta^3 r$, which is easily seen to be satisfied by our choice for η.

3.2.2 The second part of the proof – analysis of Step 4, Step 5 and Step 6

As we proved above in Lemma 3.5, the degrees of good vertices in graph $\mathcal{D}$ are almost the same. In Step 4. the vertices of G will choose uniformly, independently at random a neighboring box. Clearly, the probability that a vertex v picks a given neighboring box is the reciprocal of its degree in $\mathcal{D}$. We need the following estimation on the degree distribution of $\mathcal{D}$.

Claim 3.7. *If $d' \geq 7\varepsilon/\eta$, then*

$$1 \leq \frac{\Delta_{\mathcal{D}}}{\delta_{\mathcal{D}}} \leq (1+3\eta).$$

Proof: By Lemma 3.5 we have that

$$\frac{(1-\eta)r}{(d'+3\varepsilon)\varepsilon m_b} \leq \delta_{\mathcal{D}}$$

and

$$\Delta_{\mathcal{D}} \leq \frac{r}{(d'-3\varepsilon)\varepsilon m_b}.$$

Hence

$$\frac{\Delta_{\mathcal{D}}}{\delta_{\mathcal{D}}} \le \frac{d' + 3\varepsilon}{(1-\eta)(d' - 3\varepsilon)} = \frac{1}{1-\eta} + \frac{6\varepsilon}{(1-\eta)(d'-3\varepsilon)} \le 1 + 3\eta,$$

where the last inequality follows from the condition of the claim. □

Next we give a bound for the number h of $\mathcal{C}_g$ classes.

Claim 3.8. *We have $h \le 7\eta/\nu$.*

Proof: In Claim 3.7 we proved that $\Delta_{\mathcal{D}} \le (1+3\eta)\delta_{\mathcal{D}}$, and we also noted that $\Delta_{\mathcal{D}} \ge (1+\nu)^{h-1}\delta_{\mathcal{D}}$. Hence $(1+\nu)^{h-1} \le (1+3\eta)$. Elementary calculus tells that $(1+\nu) \ge e^{\nu/2}$ and $(1+3\eta) \le e^{3\eta}$. These imply that

$$e^{(h-1)\nu/2} \le e^{3\eta},$$

from which we get that $(h-1)\nu/2 \le 3\eta$, hence, $h - 1 \le 6\eta/\nu$. Using that $\eta \gg \varepsilon$, we get what was desired. □

Let us consider an arbitrary good bundle $H_i[Q_i, S_i]$. Right before Step 4 it is a $(3\varepsilon, d')$-bundle with density $d' \pm \varepsilon$. Denote Q' and S' the sets of good vertices in the two parts. By the definition of good bundles we have $|Q'|, |S'| \ge (1-\eta)m_i$, where $|Q_i| = |S_i| = m_i$.

Claim 3.9. *If $\eta \le 1/10$ then $H[Q', S']$ is a 4ε-regular pair with density $d' \pm 4\varepsilon$.*

Proof: After deleting at most one-tenth of the vertices from both parts, we have $4\varepsilon|Q'| \ge 3\varepsilon|Q_i|$ and $4\varepsilon|S'| \ge 3\varepsilon|S_i|$. Since $H[Q_i, S_i]$ is 3ε-regular, the density of $H[Q', S']$ is $(d' \pm \varepsilon) \pm 3\varepsilon = d' \pm 4\varepsilon$, and if $X \subset Q'$, $|X| \ge 4\varepsilon|Q'|$ and $Y \subset S'$, $|Y| > 4\varepsilon|S'|$,hence, the density of $H[X, Y]$ must be $d' \pm 4\varepsilon$. This finishes the proof of the claim. □

Let $[X_g, Y_g]$ be a pair of large sub-boxes of the box pair $[X, Y]$, where we obtained $[X, Y]$ by the random distribution of Step 4 from the pair $H[Q', S']$. By Claim 3.9 we have that $H[Q', S']$ is (4ε)-regular with density $d' \pm 4\varepsilon$.

In order to prove that $[X_g, Y_g]$ is an $\sqrt[4]{\nu}$-regular pair, we will use a lemma from [1]. First we need a definition. Let F be a bipartite graph with vertex parts A and B. Let a denote the average degree of the vertices of A. Given two (not necessarily distinct) vertices, $x_1, x_2 \in A$ we let

$$\sigma(x_1, x_2) = |N(x_1) \cap N(x_2)| - \frac{a^2}{|B|}.$$

For a subset $A' \subset A$ the *deviation* of A' is

$$\sigma(A') = \frac{\sum_{x_1, x_2 \in A'} \sigma(x_1, x_2)}{|A'|^2}.$$

The following is Lemma 3.2 in [1].

Lemma 3.10. *Let F be a bipartite graph with vertex parts A and B such that $m = |A| = |B|$. Let $2m^{-1/4} < \vartheta < 1/16$. Assume that at most $\vartheta^4 m/8$ vertices of A deviate from the average degree of F by at least $\vartheta^4 m$. Then if F is not ϑ-regular then there exists $A' \subset A$, $|A'| \ge \vartheta m$ such that $\sigma(A') \ge \vartheta^3 m/2$.*

For being able to use the above lemma for $[X_g, Y_G]$, we go through a few steps. First we estimate the $\sigma(A')$ deviations in the subgraph $H[Q', S' \cap \mathcal{C}_g]$, secondly we compute the effect of Step 4 (random distribution) in $H[Q', Y_g]$, then we consider the deviation in the subgraph $H[X, Y_g]$, and finally, we give an upper bound for the pair $[X_g, Y_g]$.

For all these computations we have to estimate the average degrees and the $\sigma(X)$ values in the box pairs of good bundles. Let S_g denote the set $S' \cap \mathcal{C}_g$, and set $s_g = |S_g|$.

Claim 3.11. *The average degree of $H[Q', S_g]$ is $a = (d' \pm 8\varepsilon)s_g$. Moreover, at most $8\varepsilon|Q'|$ vertices of Q' has degrees outside of the interval $a \pm 4\varepsilon|S'|$.*

Proof: In the proof we use that Y_g is a large sub-box, hence, $s_g > d\nu|S'|/2$ with high probability. The second part of the claim also follows from the (4ε)-regularity of $H[Q', S']$. □

Lemma 3.12. *The deviation of $\Gamma \subset Q'$ in $H[\Gamma, S_g] \subset H[Q', S_g]$ is less than $8\sqrt{\varepsilon}s_g$ if $|\Gamma| \geq \sqrt{\varepsilon}|Q'|$.*

Proof: By (4ε)-regularity there at most $4\varepsilon|Q'|$ vertices in Γ that have degree less than $(d' - 4\varepsilon)s_g$ into S_g, and similarly, at most $4\varepsilon|Q'|$ such vertices, that have more than $(d' + 4\varepsilon)s_g$ neighbors in S_g.

Let $x_1 \in \Gamma$ be a non-exceptional vertex. Since $(d' - 4\varepsilon)s_g > \varepsilon|S'|$, there are at least $(1 - 8\varepsilon)|Q'|$ vertices in Γ, which have $(d' \pm 4\varepsilon)deg(x_1, S_G)$ neighbors in S_g. Hence, we have the following bound:

$$(|\Gamma|^2 - 8|\Gamma|\varepsilon|Q'|)(d' - 4\varepsilon)^2 s_g \leq \sum_{x_1, x_2 \in \Gamma} |N(x_1) \cap N(x_2) \cap S_g| \leq 7|\Gamma| \cdot |Q'|\varepsilon s_g + (|\Gamma|^2 - 7|\Gamma|\varepsilon|Q'|)(d' + 4\varepsilon)^2 s_g,$$

where we used that $|\Gamma|^2 - 8\varepsilon|\Gamma| \cdot |Q'| \leq (|\Gamma| - 4\varepsilon|Q'|)^2 \leq |\Gamma|^2 - 7\varepsilon|\Gamma|$.

We only need to work with the upper bound for estimating the deviation of Γ. Below we use that $|\Gamma| \geq \sqrt{\varepsilon}|Q'|$:

$$\frac{\sum_{x_1, x_2 \in \Gamma} |N(x_1) \cap N(x_2) \cap S_g|}{|\Gamma|^2} - \frac{a^2}{s_g} \leq 7\sqrt{\varepsilon}s_g + 9\varepsilon s_g - 7\sqrt{\varepsilon}(d')^2 s_g - 63\varepsilon\sqrt{\varepsilon}s_g + 3\varepsilon s_g < 8\sqrt{\varepsilon}s_g.$$

□

Lemma 3.13. *The $[X_g, Y_g]$ pair of sub-boxes is $(3\sqrt{4}\nu)$-regular with high probability.*

Proof: As we mentioned before, the crucial step was in Lemma 3.12. Using Chernoff's bound we get that the average degree in $H[Q', Y_g]$ is $a(1 \pm 2\nu)$ with high probability, and that for every $x_1, x_2 \in Q'$ their common neighborhood in Y_g is

$$\frac{|N(x_1) \cap N(x_2) \cap Y_g|}{|Y_g|} = (1 \pm 2\nu)\frac{|N(x_1) \cap N(x_2) \cap S_g|}{s_g}.$$

Hence, the deviation of any sufficiently large $\Gamma \subset Q'$ is at most $(1 + 2\nu)\sigma(\Gamma)$.

The deviation grows at most at a similar pace when we turn from Q' to the random subset X chosen in Step 4. This may not be as obvious first, since we require that for every subset $\Gamma \subset X$ we have a small deviation if Γ is large enough. However, it is enough to observe that what we need is the "right" proportion of vertices in X that have about the average degree into Y_g, and the "right" proportion of pairs of vertices in X for which their common neighborhood is close enough to the expectation. That is, we only have to verify a polynomial number of events, hence, by Chernoff's bound, the deviation is still at most $(8\sqrt{\varepsilon}+4\nu)|Y_g|$. Finally, after randomly choosing $X_g \subset X$ with $|X_g| = |Y_g|$, we again lose a little by Hoeffding's inequality with high probability. At the end there will be $6\nu|X_g|$ vertices with degrees that deviate from the average by more than $6\nu|Y_g|$, and for every $\Gamma \subset X_g$ we have that $\sigma(\Gamma) \leq 6\nu|X_g|$.

If we let $\vartheta = 3\sqrt[4]{\nu}$, then by Lemma 3.10 we get that the sub-box pair $[X_g, Y_g]$ is -regular whp, proving what was desired. □

The above lemma finishes the proof of Theorem 3.1: the large sub-box pairs are the d^κ-regular (U_i, V_i) pairs $(1 \leq i \leq M)$, and $W_0 = U_0 \cup V_0$ has at most $10d^{1/4}n$ vertices. □

4 Approximation algorithms for dense intances of the 2-EC, 2-VC and TSP(1,2) problems

In this section we consider applications of Theorem 3.1. At certain points we use ideas of [7]. The main difference is that instead of the Regularity lemma of Szemerédi, Theorem 3.1 is used. This enables us to find effective approximation algorithms for relatively sparse graphs of reasonable size.

4.1 Definition of the problems

In the 2-EC problem we are given a 2-edge-connected graph G, and the goal is to find a 2-edge-connected spanning subgraph of G with minimum number of edges. The approximation ratio of an algorithm for this problem is the number of edges of the subgraph the algorithm finds divided by the optimal value.

The 2-VC problem is very similar. In this we are given a 2-vertex-connected graph G, and we want to find a 2-vertex-connected spanning subgraph of it with the minimum number of edges. The approximation ratio is defined analogously to the 2-EC case.

Finally, in the TSP(1, 2) problem we are given a complete graph with weights 1 or 2 on every edge. The goal is to find a Hamilton cycle with minimum total weight. Hence, this problem is a special instance of the travelling salesman problem. The approximation ratio of an algorithm for TSP(1, 2) is the total weight of the Hamilton cycle which is found by the algorithm divided by the total weight of the optimal solution. Note that the TSP(1, 2) problem is determined by the subgraph containing the edges with weight 1.

Each of the above optimization problems are NP-complete. Furthermore, they are NP-complete even in the restricted case when the underlying graph is r-regular for some $r < n/2$ (here, as before, n denotes the number of vertices).

4.2 The case of r-regular graphs

Let us assume that we are given a r-regular graph bipartite graph $G = (U \cup V)$ on $2n$ vertices, and a integer $\kappa = 5$ and real number $0 < d < 1$, sand assume further that $n \geq n_0$, where $n_0 = n_0(d)$ is the threshold given by Theorem 3.1. Apply Theorem 3.1 for G. We obtain the d^5-regular pairs (U_i, V_i) for $1 \leq i \leq K$, and the "wastebasket" W_0 with $|W_0| \leq 10d^{1/4}n$.

We need the algorithmic version of the Blow-up lemma of Komlós, Sárközy and Szemerédi [19].

Theorem 4.1. *[Blow-up Lemma] Given a graph R of order r and positive integers δ and Δ there exists a positive $\varepsilon = (\delta, \Delta, r)$ such that the following holds: Let $n_1, n_2, \ldots, n_r$ be arbitrary positive parameters and let us replace the vertices $v_1, v_2, \ldots, v_r$ of R with pairwise disjoint sets $W_1, W_2, \ldots, W_r$*

of sizes $n_1, n_2, \ldots, n_r$ (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 maximum degree $\Delta(H) \leq \Delta$ is embeddable into F then it is already embeddable into G.

We are going to use the Blow-up lemma for the regular pairs, that is, in case when R contains two vertices ($r = 2$) and the edge that connects them. We need a more explicit relation of ε and δ than what is given in the above formulation. In the proof it is implicit that for the $r = 2$ case it is sufficient if $\varepsilon \leq \delta^{5\Delta} \leq 1$.

Using the Blow-up lemma one can find a cycle in each regular pair that contains every vertex of the pair. That is, using a randomized polynomial time algorithm we can find M vertex disjoint cycles in G such that at most $10d^{1/4}n$ vertices are not included in them.

Now, for the 2-EC problem, we shrink these cycles into M distinct auxiliary vertices. We construct a new, auxiliary graph: whenever there is an edge between two vertices that belong to different pairs, we include an edge that connects the corresponding auxiliary vertices, we include the edges that connect W_0 vertices, and if there is an edge between a vertex $w \in W_0$ and a vertex from some regular pair, we connect the corresponding auxiliary vertex with w. Next we delete the parallel edges.

By the 2-edge-connectedness of G this auxiliary graph must also be 2-edge-connected. It has at most $M + 10d^{1/4}n$ vertices, and even a poor deterministic algorithm (e.g., use ear decomposition) can find a 2-edge-connected subgraph of it having at most $2(M + 10d^{1/4}n)$ edges. Hence, the total number of edges in the 2-edge-connected spanning subgraph we find this way is at most $n + 2(M + 10d^{1/4}n) = n + o(n)$. That is, the approximation ratio of this algorithm is $1 + o(1)$.

The case of the 2-VC problem is very similar. The only difference is that after finding the spanning cycles for the regular pairs and constructing the new graph we use a slightly different algorithm for finding a 2-VC spanning subgraph in the auxliary graph. We obtain the same approximation ratio.

For the TSP(1, 2) problem we again find the same spanning cycles in the regular pairs. Then we connect these cycles into one big cycle, using weight-2 edges if necessary. Finally, we include the vertices of W_0, forming a Hamilton cycle, using weight-2 edges if necessary. Clearly, the total weight of the Hamilton cycle we find this way is at most $n + M + 2|W_0| = n + o(n)$. Hence, the algorithm has approximation ratio $1 + o(1)$.

If G is not bipartite, but r-regular, we can randomly divide its vertex set into two (almost) equal parts. This way with high probability we obtain a bipartite graph in which every degree is in the range $r \pm O(\sqrt{n \log n})$. As Theorem 3.1 is not sensitive for a such a small deviation from the average degree, it can be applied. Let us summarize these results:

Theorem 4.2. *Let G be an n-vertex, r-regular graph. Let $0 < d \ll r/n$. Set $\varepsilon = d^{42}$. Assume further that $e^{M/4}2n^{-1/4} < \varepsilon$, where $M = (h/\varepsilon^{12})\log(h/\varepsilon^{12})$ and $h = 3^{12} \cdot 16^3$. Then there is a randomized polynomial time $(1 + o(1))$-approximation algorithm for the 2-EC, 2-VC and TSP(1, 2) problems.*

Note that we have $\kappa = 10$ in Theorem 3.1, since we need d^{10}-regular pairs for applying the Blow-up lemma.

References

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

[2] N. Alon, E. Fischer, and I. Newman, Efficient testing of bipartite graphs for forbidden induced subgraphs, SIAM J. Comput. **37** (2007), 959-976.

[3] G. Bodwin, S. Vempala, A Unified View of Graph Regularity via Matrix Decompositions, https://arxiv.org/pdf/1911.11868.pdf

[4] D. Conlon, J. Fox, B. Sudakov, Y. Zhao, The regularity method for graphs with few 4-cycles, arXiv preprint arXiv:2004.10180.

[5] B. Csaba, Regular decomposition of the edge set of graphs with applications, submitted, https://arxiv.org/pdf/2109.12394.pdf

[6] B. Csaba, A new graph decomposition method for bipartite graphs, Proceedings of MATCOS 2019, 11-14.

[7] B. Csaba, M. Karpinski, P. Krysta, Approximability of dense and sparse instances of minimum 2-connectivity, TSP and path problems, SODA 2001.

[8] B. Csaba, Regular spanning subgraphs of bipartite graphs of high minimum degree, The Electronic Journal of Combinatorics **14** (2007): 1.

[9] N. Eaton, Ramsey numbers for sparse graphs, Discrete Mathematics **185** (1998) 63-75.

[10] N. Eaton, V. Rödl, A Regularity Lemma, Tech. Report, Mathematics Department, Emory University, 1992.

[11] P. Erdős, P. Turán, On some sequences of integers, J. London Math. Soc. **11** (1936), 261-264.

[12] A. Frieze, R. Kannan, Quick approximations to matrices and applications, Combinatorica **19** (1999), 175-220.

[13] J. Fox, A new proof of the graph removal lemma, Annals of Mathematics (2) 174 (2011), 561-579.

[14] S. Gharan, L. Trevisan, A new regularity lemma and faster approximation algorithms for low threshold rank graphs. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. Springer, Berlin, Heidelberg, 2013. 303-316.

[15] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, GAFA **7** (1997), 322-337.

[16] S. Janson, T. Luczak, A. Ruczinski,

[17] J. Komlós, The blow-up lemma, Combinatorics, Probability and Computing **8** (1999): 161-176.

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

[19] J. Komlós, G.N. Sárközy and E. Szemerédi, An algorithmic version of the blow-up lemma, Random Structures & Algorithms **12** (1998): 297-312.

[20] J. Komlós, M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory. In: Combinatorics, Paul Erdős is Eighty, Vol II (D. Miklós, V. T. Sós, T. Szőnyi eds.), János Bolyai Math. Soc., Budapest (1996), 295-352.

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

[22] E. Szemerédi, Regular partitions of graphs, Colloques Internationaux C.N.R.S. No 260 - Problémes Combinatoires et Théorie des Graphes, Orsay (1976), 399-401.