

# Additive Approximation of Generalized Turán Questions

Noga Alon\*

Clara Shikelman †

November 17, 2021

## Abstract

For graphs  $G$  and  $T$ , and a family of graphs  $\mathcal{F}$  let  $\text{ex}(G, T, \mathcal{F})$  denote the maximum possible number of copies of  $T$  in an  $\mathcal{F}$ -free subgraph of  $G$ . We investigate the algorithmic aspects of calculating and estimating this function. We show that for every graph  $T$ , finite family  $\mathcal{F}$  and constant  $\epsilon > 0$  there is a polynomial time algorithm that approximates  $\text{ex}(G, T, \mathcal{F})$  for an input graph  $G$  on  $n$  vertices up to an additive error of  $\epsilon n^{v(T)}$ . We also consider the possibility of a better approximation, proving several positive and negative results, and suggesting a conjecture on the exact relation between  $T$  and  $\mathcal{F}$  for which no significantly better approximation can be found in polynomial time unless  $P = NP$ .

## 1 Introduction

Many natural computational problems can be formulated as *graph modification problems*. In these we are given an input graph  $G$  and we aim to apply the smallest number of modifications and get a graph which has some predefined property. Both the allowed modifications and the desired properties vary. The most common modifications are adding, deleting, or editing edges or vertices. As for the desired properties, these are usually either graph properties coming from classical graph theory or properties motivated by real world applications such as Molecular Biology [8], [14], [15], Circuit Design [11] or Machine Learning [6].

Garey and Johnson [13] considered 18 edge and vertex modification problems. Yannakakis [22] proved that such modification problems are NP-hard for properties such as outerplanar and transitively orientable, Asano [4], and Asano and Hirata [5] established NP-hardness for several properties expressible through forbidding families of minors or topological minors, El-Mallah and Colbourn [11] proved NP-hardness for properties defined by forbidden minors and induced subgraphs. In [22] Yannakakis posed the question of proving NP-hardness not only for specific properties, but for general families of properties.

In [18] Natanzon, Shamir and Sharan studied edge modification problems for hereditary properties such as being Perfect. They showed that not only are these problems NP-hard, even finding an approximate answer, up to some constant multiplicative factor, is NP-hard. Other works have also investigated the question of approximation, see for example [16] and [7].

---

\*Department of Mathematics, Princeton University, Princeton, New Jersey, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-1855464 and BSF grant 2018267.

†Sackler School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: clarashk@mail.tau.ac.il. Research supported in part by an ISF grant.

In [1] Alon, Shapira and Sudakov investigated this question for the general family of monotone graph properties<sup>1</sup>. The only relevant edge-modification for such properties is edge deletion. Note that any monotone property can be defined as the property of a graph being  $\mathcal{F}$ -free where  $\mathcal{F}$  is an appropriate (finite or infinite) family of graphs. Thus the question becomes the following: for a graph  $G$ , let  $\text{ex}(G, K_2, \mathcal{F})$  denote the maximum number of edges in a subgraph of  $G$  that contains no copy of  $F \in \mathcal{F}$ . The following theorem shows that this value can be approximated as follows.

**Theorem 1.1** ([1]). *For any fixed  $\epsilon > 0$  and any family of graphs  $\mathcal{F}$  there is a deterministic algorithm that given a graph  $G$  on  $n$  vertices computes  $\text{ex}(G, K_2, \mathcal{F})$  up to additive error of  $\epsilon n^2$  in time  $O(n^2)$ .*

A complimentary theorem shows that excluding simple cases, a significantly better approximation is NP-hard

**Theorem 1.2** ([1]). *Let  $\mathcal{F}$  be a family of graphs and let  $G$  be a graph on  $n$  vertices. Then,*

1. *If there is a bipartite graph in  $\mathcal{F}$ , then there is a fixed  $\delta > 0$  for which it is possible to approximate  $\text{ex}(G, K_2, \mathcal{F})$  within an additive error of  $n^{2-\delta}$  in polynomial time.*
2. *On the other hand, if there are no bipartite graphs in  $\mathcal{F}$ , then for any fixed  $\delta > 0$  it is NP-hard to approximate  $\text{ex}(G, K_2, \mathcal{F})$  within an additive error of  $n^{2-\delta}$ .*

Note that if  $G = K_n$ , the complete graph on  $n$  vertices, then finding  $\text{ex}(K_n, K_2, \mathcal{F})$  is the classical Turán question. This question and its many variations are in the heart of extremal graph theory. Recently in [3] the systematic study of the following variation was initiated

**Definition 1.3.** *For graphs  $G$  and  $T$  and a family of graphs  $\mathcal{F}$  let  $\text{ex}(G, T, \mathcal{F})$  denote the maximum number of copies of  $T$  in an  $\mathcal{F}$ -free subgraph of  $G$ .*

Following the spirit of Theorems 1.1 and 1.2 and the above generalization of the classical Turán theorem, we consider the following question. Given a forbidden family of graphs  $\mathcal{F}$ , a graph  $T$ , and a constant  $\epsilon > 0$ , is there a polynomial time algorithm that given a graph  $G$  determines  $\text{ex}(G, T, \mathcal{F})$  up to an additive error of  $\epsilon n^{v(T)}$ ? If so, is the problem of finding a significantly better approximation NP-hard?

We answer the first question by proving the following

**Theorem 1.4.** *For every constant  $\epsilon > 0$ , finite family of forbidden graphs  $\mathcal{F}$  and fixed graph  $T$ , there is a deterministic polynomial time algorithm that given a graph  $G$  on  $n$  vertices approximates  $\text{ex}(G, T, \mathcal{F})$  within an additive error of  $\epsilon n^{v(T)}$ .*

The proof is based on variants of Szemerédi's regularity lemma, extending the methods in [1].

As for the question of better approximation, the natural generalization of the easy part of Theorem 1.2 holds also in our case. For a fixed graph  $T$ , an  $m$  blow-up of  $T$  is the graph obtained by replacing each vertex of  $T$  by an independent set of size  $m$  and each edge by a complete bipartite graph between the corresponding independent sets.

---

<sup>1</sup>Monotone graph properties are properties that are closed under edge and vertex deletion, for example being  $K_3$  free or being planar.

**Proposition 1.5.** *Let  $T$  be a fixed graph and let  $\mathcal{F}$  be a family of graphs such that there is a graph  $F \in \mathcal{F}$  which is a subgraph of a blowup of  $T$ . Then there is a fixed  $\epsilon := \epsilon(T, \mathcal{F}) > 0$  such that  $\text{ex}(G, T, \mathcal{F})$  can be calculated in polynomial time up to additive error of  $n^{v(T)-\epsilon}$  for any graph  $G$  on  $n$  vertices.*

The above is a straightforward application of the following simple proposition from [3].

**Proposition 1.6** ([3]). *Let  $T$  be a fixed graph. Then  $\text{ex}(n, T, F) = \Omega(n^{v(T)})$  if and only if  $F$  is not a subgraph of a blow-up of  $T$ . Otherwise,  $\text{ex}(n, T, F) \leq n^{v(T)-\epsilon}$  for some  $\epsilon := \epsilon(T, F) > 0$*

If indeed  $F$  is a subgraph of a blow up of  $T$ , then for  $\epsilon := \epsilon(T, F) > 0$ , if  $n$  is large enough we have that

$$0 \leq \text{ex}(G, T, F) \leq \text{ex}(K_n, T, F) \leq n^{v(T)-\epsilon} + 0.$$

Thus 0 is a trivial approximation of  $\text{ex}(G, T, F)$  up to additive error of  $n^{v(T)-\epsilon}$ . For any family of graphs  $\mathcal{F}$ , if  $F \in \mathcal{F}$  then  $\text{ex}(G, T, \mathcal{F}) \leq \text{ex}(G, T, F)$ , and the required result follows.

As for the extension of the second part of Theorem 1.2 we prove the following special case

**Theorem 1.7.** *Let  $k > m \geq 2$  be integers, then for every  $\epsilon > 0$  approximating  $\text{ex}(G, K_m, K_k)$  up to additive error of  $n^{m-\epsilon}$  is NP-hard for a given input graph  $G$  on  $n$  vertices.*

We believe that excluding the cases covered by Proposition 1.5 no better approximation is possible, and so we suggest the following conjecture.

**Conjecture 1.8.** *For every graph  $T$ , family of graphs  $\mathcal{F}$  such that no  $F \in \mathcal{F}$  is a subgraph of a blow-up of  $T$ , and  $\epsilon > 0$ , it is NP-hard to approximate  $\text{ex}(G, T, \mathcal{F})$  up to additive error of  $n^{v(T)-\epsilon}$  for a given input graph  $G$  on  $n$  vertices.*

The rest of the paper is organized as follows. Section 2 is dedicated to definitions and preliminary results, most of these concern variations of the regularity lemma. Section 3 is the proof of the main lemma used to establish Theorem 1.4. Sections 4 and 5 contain the proofs of Theorems 1.4 and 1.7 respectively, and finally Section 6 includes some further remarks concerning Conjecture 1.8 and open problems.

## 2 Regularity Lemmas and Auxiliary Results

From here on let  $\epsilon_1, \epsilon_2, \dots$  and  $\delta_1, \delta_2, \dots$  be positive constants depending on  $\epsilon, T$  and  $\mathcal{F}$  and tending to zero as  $\epsilon$  tends to zero. In some cases the indices refer to the lemma or theorem from which the corresponding constant arises. Additionally, let  $t = v(T)$  denote the number of vertices of  $T$ .

### 2.1 Two Versions of the Regularity Lemma

Given a graph  $G$  and two disjoint sets of vertices  $A, B \subset V(G)$ , let  $e(A, B)$  denote the number of edges with one end point in each set and let the density of edges between the sets be defined as:

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

**Definition 2.1** ( $\epsilon$ -Regular Pair). *Given a graph  $G$  and a pair of disjoint sets of vertices  $A, B \subset V(G)$ , we say that the pair  $(A, B)$  is  $\epsilon$ -regular, if for any two subsets  $A' \subseteq A$  and  $B' \subseteq B$ , such that  $|A'| \geq \epsilon|A|$  and  $|B'| \geq \epsilon|B|$ , the inequality  $|d(A', B') - d(A, B)| \leq \epsilon$  holds.*

**Definition 2.2** ( $\epsilon$ -Regular partition). *Given a graph  $G$  and  $\epsilon > 0$  a partition  $P = \{V_i\}_{i=0}^k$  of  $V(G)$  for which  $|V_i| = |V_j|$  for every  $i, j \geq 1$ ,  $|V_0| < k$ , and all but at most  $\epsilon \binom{k}{2}$  of the pairs  $(V_i, V_j)$  are  $\epsilon$ -regular is called an  $\epsilon$ -regular partition.*

The following theorem is the regularity lemma of Szemerédi [20]. It is convenient to use the following version which appears, for example, in [12], Theorem 3.7.

**Theorem 2.3** ([20],[12]). *For any  $\epsilon > 0$  and integer  $k$  there exist integers  $N_{2.3} := N(\epsilon, k)$  and  $K_{2.3} := K(\epsilon, k)$  such that the following holds. If  $|V| > N_{2.3}$  then for any partition  $V = V_0 \cup \dots \cup V_k$  with  $|V_0| < k$  and  $|V_1| = \dots = |V_k|$  and any graph  $G$  on the vertex set  $V$  there exists a partition  $V = U_0 \cup \dots \cup U_{k'}$  such that*

1.  $|U_0| < k' < K_{2.3}$ .
2. For each  $U_i$  with  $i \geq 1$  there is  $V_j$  such that  $U_i \subset V_j$ .
3.  $|U_1| = \dots = |U_{k'}|$ .
4.  $U_0 \cup \dots \cup U_{k'}$  is an  $\epsilon$ -regular partition of  $G$ .

We will also need the following algorithmic version of the regularity lemma that uses a stronger notion of regularity. For the following see [2] and [19]

**Definition 2.4** ( $f$ -Regular partition). *For a function  $f : \mathbb{N} \rightarrow (0, 1)$  and a graph  $G$  we say that a partition  $P = \{V_i\}_{i=0}^k$  of  $V(G)$  such that  $|V_0| < k$  and  $|V_i| = |V_j|$  for every  $i, j \geq 1$  is an  $f$ -regular partition if all pairs  $(V_i, V_j)$ ,  $1 \leq i < j \leq k$ , are  $f(k)$ -regular.*

**Theorem 2.5** ([2]). *For every  $l, \epsilon > 0$  and non-increasing function  $f : \mathbb{N} \rightarrow (0, 1)$ , there is an integer  $K_{2.5} := K_{2.5}(f, \epsilon, l)$  so that for a given graph  $G$  on  $n$  vertices, with  $n > K_{2.5}$ , one can add or remove at most  $\epsilon n^2$  edges of  $G$  and get a graph  $G_0$  that has an  $f$ -regular partition of order  $k$ , where  $l \leq k \leq K_{2.5}$ . Furthermore, the needed changes and the partition can be found in polynomial time.*

Note that in [2] the definition of  $f$ -regular partition is slightly different. There is no set  $V_0$  and the sizes of the sets are such that for every  $i, j$ ,  $||V_i| - |V_j|| \leq 1$ . It is easy to check that the two versions are equivalent. For simplicity we will use the version which has a set  $V_0$  and  $k$  sets of equal size throughout the paper.

## 2.2 Definitions of Weighted Graphs and Partition Graphs

We use the following definitions for weighted graphs.

**Definition 2.6.**

1. A weighted graph  $W$  is a graph on  $n$  vertices with a weight function  $w : E(K_n) \rightarrow [0, 1]$ , where we identify between  $w(e) = 0$  and  $e \notin E(W)$ .

2. For a fixed graph  $T$  let  $\mathcal{T}$  be the set of copies of  $T$  in  $W$  and let

$$\mathcal{N}(W, T) = \sum_{T \in \mathcal{T}} \prod_{(v_i, v_j) \in E(T)} w(v_i, v_j).$$

Call  $W$  a  $T$ -free graph if  $\mathcal{N}(W, T) = 0$ .

(Note that in an unweighted graph  $G$ ,  $\mathcal{N}(G, T)$  is just the number of copies of  $T$  in  $G$ .)

3. Let  $W'$  be a weighted graph on  $V(W)$  with a weight function  $w' : E(W') \rightarrow [0, 1]$ . We say that  $W'$  is a conventional subgraph of  $W$  if  $\forall e \in K_n$  either  $w'(e) = w(e)$  or  $w'(e) = 0$ .

Given a graph  $G$  and an  $\epsilon$ -regular partition of it, say  $P = \{V_i\}$ , we would like to associate a weighted graph to  $G$  and  $P$  and to relate subgraphs of the weighted graph to subgraphs of  $G$ . To do this we use the following definitions:

**Definition 2.7.**

- Given a graph  $G$  and an  $\epsilon$ -regular partition  $P = \{V_i\}_{i=0}^k$  of its vertices define the  $(\epsilon, d)$ -partition graph  $W$  to be a weighted graph on  $k$  vertices  $\{v_1, \dots, v_k\}$ . The weight function is  $w(v_i, v_j) = d(V_i, V_j)$  if  $(V_i, V_j)$  is an  $\epsilon$ -regular pair with density at least  $d$  and  $w(v_i, v_j) = 0$  otherwise.
- For a conventional subgraph of  $W$ , say  $W'$ , let  $G_{W'}$  be the following subgraph of  $G$  on the same set of vertices  $V(G) = V(G_{W'})$ . An edge  $e = \{u, u'\} \in E(G)$  is also an edge of  $G_{W'}$  if and only if  $u \in V_i, u' \in V_j$  and  $w'(v_i, v_j) > 0$ . The vertices of  $V_0$  form an independent set.

### 2.3 Embedding and Counting Copies of Fixed Graphs

For a graph  $R$  and an integer  $h$  let  $R(h)$  be the  $h$ -blowup of  $R$ , that is the graph obtained by replacing each vertex with an independent set of size  $h$ , and each edge with a complete bipartite graph between the corresponding independent sets.

**Theorem 2.8** (Embedding Lemma, see, e.g., [17]). *Given  $d > \epsilon > 0$ , a graph  $R$ , and a positive integer  $m_{2.8}$ , construct a graph  $G$  by replacing every vertex of  $R$  by  $m_{2.8}$  vertices, and by replacing each edge of  $R$  by an  $\epsilon$ -regular pair of density at least  $d$ . Let  $F$  be a subgraph of  $R(h)$  of maximum degree  $\Delta > 0$ , let  $\delta = d - \epsilon$  and  $\epsilon_0 = \delta^\Delta / (2 + \Delta)$ . If  $\epsilon \leq \epsilon_0$  and  $h - 1 \leq \epsilon_0 m_{2.8}$  then  $F \subset G$ .*

As we aim to use the embedding lemma with an  $(\epsilon, d)$  partition graph, we need to choose  $d$  to suit a family of graphs  $\mathcal{F}$ . Let  $\Delta(\mathcal{F}) = \max_{F \in \mathcal{F}} \{\Delta(F)\}$ , and let  $v(\mathcal{F}) = \max_{F \in \mathcal{F}} \{v(F)\}$ . For a fixed  $\epsilon$  and family  $\mathcal{F}$  let  $\epsilon_0 = \epsilon$  and let  $d_{2.8} := d(\epsilon, \Delta(\mathcal{F})) = \delta + \epsilon$  as they appear in Theorem 2.8. Note that  $d_{2.8}$  tends to zero as  $\epsilon$  tends to zero. Let  $m_{2.8} := m_{2.8}(\epsilon, \mathcal{F}) = v(\mathcal{F}) \frac{1}{\epsilon}$ .

**Lemma 2.9.** *Let  $W$  be a weighted graph on  $n$  vertices,  $\epsilon > 0$  be a constant and  $P = \cup_{i=1}^k V_i$  a partition of  $V(W)$  into  $k \geq \frac{t^2}{\epsilon}$  parts of equal size. Let  $\mathcal{T}'$  be the set of all copies of  $T$  in  $W$  using at least two vertices from the same  $V_i$ . Then*

$$\sum_{T \in \mathcal{T}'} \prod_{(v_i, v_j) \in E(T)} w(v_i, v_j) < \epsilon n^t.$$

*Proof.* As  $w(v_i, v_j) \leq 1$  note that  $\sum_{T \in \mathcal{T}'} \prod_{(v_i, v_j) \in E(T)} w(v_i, v_j) \leq |\mathcal{T}'|$ , and so it is enough to bound the number of copies of  $T$  in  $\mathcal{T}'$ . A copy from  $\mathcal{T}'$  uses vertices from at most  $t - 1$  sets of the partition, and in each such set there are  $\frac{n}{k}$  vertices. Thus

$$|\mathcal{T}'| < \binom{t}{2} k^{t-1} \left(\frac{n}{k}\right)^t < \frac{t^2}{k} n^t$$

as  $\epsilon \geq \frac{t^2}{k}$  the required result follows.  $\square$

Note that the above lemma can be used for unweighted graphs by choosing  $w(v_i, v_j) \in \{0, 1\}$  appropriately.

**Lemma 2.10.** *Let  $T$  be a graph on  $t$  vertices, let  $G$  be a graph on  $n$  vertices,  $P = V_0 \cup \{V_i\}_{i=1}^k$  an  $\epsilon$ -regular partition of its vertices and let  $W$  be its  $(\epsilon, d_{2.8})$  partition graph. Then there exists  $\delta_{2.10} := \delta_{2.10}(\epsilon, t)$ , that tends to zero when  $\epsilon$  and  $d_{2.8}$  tend to zero, such that*

$$|\mathcal{N}(G, T) - |V_i|^t \mathcal{N}(W, T)| < \delta_{2.10} n^t$$

*Proof.* Let  $\mathcal{N}^P(G, T)$  be the number of copies of  $T$  in  $G$  using at most one vertex from each set  $\{V_i\}_{i=1}^k$  and not using vertices from  $V_0$ , edges between sets  $(V_i, V_j)$  which are not  $\epsilon$ -regular or with density smaller than  $d_{2.8}$ . Call these copies *conventional*.

There are at most  $\epsilon t^2 \binom{k}{2} \left(\frac{n}{k}\right)^2 n^{t-2} \leq \epsilon n^t$  copies of  $T$  using an edge of an irregular pair, there are at most  $t^2 \binom{k}{2} d_{2.8} \left(\frac{n}{k}\right)^2 n^{t-2} \leq d_{2.8} n^t$  copies using edges between sets of small density and there are at most  $t k n^{t-1}$  copies using a vertex from  $V_0$ . Together with Lemma 2.9 we get that for  $\delta_0 > 0$  that tends to zero when  $\epsilon$  and  $d_{2.8}$  tend to zero

$$|\mathcal{N}(G, T) - \mathcal{N}^P(G, T)| < \delta_0 n^t. \quad (1)$$

Let  $\mathcal{T} = \{T_i\}$  be the set of all copies of  $T$  in  $W$ . For a given copy  $T_i$  let  $V(T_i) = \{v_{i_1}, \dots, v_{i_t}\}$  be the sets of vertices in  $W$  that  $T_i$  uses, let  $V_{i_1}, \dots, V_{i_t}$  be the corresponding sets in  $P$  and let  $\mathcal{N}(G, T_i)$  be the number of copies of  $T$  in  $G$  using a vertex from  $V_{i_j}$  for the role of  $v_{i_j}$ .

A simplified version of Lemma 1.6 from [9] is the following

**Lemma 2.11.** *Let  $T$  be a fixed graph on  $t$  vertices, and let  $G$  be a graph with an  $\epsilon$ -regular partition into  $t$  parts  $P = V_1, \dots, V_t$ . Let  $d_{i,j}$  be the density between the sets  $V_i$  and  $V_j$ , and assume that  $d_{i,j} = 0$  for every  $\{i, j\} \notin E(T)$ . Then for  $\delta_{2.11} := \delta_{2.11}(\epsilon, t)$  that tends to zero as  $\epsilon$  tends to zero,*

$$|\mathcal{N}(G, T) - |V_1|^t \prod_{(i,j) \in E(T)} d_{i,j}| \leq \delta_{2.11} |V_1|^t$$

In our case, when counting the number of copies of  $T_i$  we focus only on the sets  $V_{i_1}, \dots, V_{i_t}$ , and assume that the density between two sets that do not correspond to an edge of  $T$  is zero. Thus,

$$|\mathcal{N}(G, T_i) - |V_1|^t \prod_{(v_{i_k}, v_{i_j}) \in E(T_i)} w(v_{i_k}, v_{i_j})| \leq \delta_{2.11} |V_1|^t$$

Each conventional copy of  $T$  in  $G$  can be mapped to a copy  $T_i$  in  $W$  by mapping each vertex  $v$  of  $T$  to the set  $V_i$  it is contained in. Thus  $\mathcal{N}^P(G, T) = \sum_{T_i \in \mathcal{T}} \mathcal{N}(G, T_i)$ .

Finally, note that  $|V_1| \leq \frac{n}{k}$ , and together with the above we get

$$\begin{aligned}
|\mathcal{N}^P(G, T) - |V_1|^t \mathcal{N}(W, T)| &= \left| \sum_{T_i \in \mathcal{T}} (\mathcal{N}(G, T_i) - |V_1|^t \prod_{(v_i, v_j) \in E(T_i)} w(v_i, v_j)) \right| \\
&\leq \sum_{T_i \in \mathcal{T}} |\mathcal{N}(G, T_i) - |V_1|^t \prod_{(v_i, v_j) \in E(T_i)} w(v_i, v_j)| \\
&\leq |\mathcal{T}| \delta_{2.11} |V_1|^t \leq k^t \delta_{2.11} \left(\frac{n}{k}\right)^t = \delta_{2.11} n^t
\end{aligned}$$

This together with (1) gives the required result. □

## 2.4 Homomorphisms of Graphs and Homomorphism Freeness

**Definition 2.12.** Given graphs  $G$  and  $F$  we say that the function  $\varphi : V(F) \rightarrow V(G)$  is a homomorphism from  $F$  to  $G$  if it maps edges to edges.

**Definition 2.13.** Given a graph  $G$  and a family of graphs  $\mathcal{F}$ , we say that  $G$  is  $\mathcal{F}$ -homomorphism-free ( $\mathcal{F}_{hom}$ -free, for short), if there is no homomorphism from  $F$  to  $G$  for any  $F \in \mathcal{F}$ .

**Definition 2.14.** For a weighted graph  $W$ , a graph  $T$  and a family of graphs  $\mathcal{F}$  define the homomorphism extremal number to be

$$\text{ex}_{hom}(W, T, \mathcal{F}) = \max\{\mathcal{N}(W_0, T) : W_0 \text{ is a conventional subgraph of } W \text{ and is } \mathcal{F}_{hom}\text{-free}\}$$

**Lemma 2.15.** Let  $G$  be a graph,  $\mathcal{F}$  a finite family of graphs,  $\epsilon > 0$ ,  $P = \bigcup_{i=0}^k V_i$  an  $\epsilon$ -regular partition of  $G$  and  $W$  its  $(\epsilon, d_{2.8})$ -partition graph. If  $v(G) \geq m_{2.8}k$ , then the following holds:

1. If  $G$  is  $\mathcal{F}$ -free then  $W$  is  $\mathcal{F}_{hom}$ -free.
2. If  $W$  is  $\mathcal{F}_{hom}$ -free then  $G_W$  is  $\mathcal{F}$ -free.

*Proof.* For the first part, assume towards contradiction that there is a graph  $F \in \mathcal{F}$  such that there is a homomorphism of it into  $W$ . To apply Lemma 2.8 first note that this means that  $F$  is a subgraph of  $W(v(\mathcal{F}))$ , which is the graph obtained by replacing every vertex of  $W$  by an independent set of size  $v(\mathcal{F})$  and every edge with positive weight by a complete bipartite graph between the corresponding independent sets. Furthermore, as  $n \geq m_{2.8}k$  and the edges correspond to  $\epsilon$ -regular pairs in the partition with density at least  $d_{2.8}$  we get that indeed  $G$  must contain a copy of  $F$ . This is a contradiction as we assumed that  $G$  is  $\mathcal{F}$ -free.

As for the second part, assume that there is a copy of some  $F \in \mathcal{F}$  in  $G_W$ . Let  $V(F) = \{v_1, \dots, v_f\}$  be the vertices of this copy and assume that in the copy in  $G_W$  vertex  $v_j$  comes from the set  $V_{k_j}$ . If  $(v_i, v_j)$  is an edge in the copy  $F$ , this means that there is an edge between the vertices  $v_{k_i}, v_{k_j}$  in  $W$ . Thus the mapping  $v_i \rightarrow v_{k_i}$  is a homomorphism. As we assumed that  $W$  is  $\mathcal{F}_{hom}$ -free this is a contradiction. □

## 2.5 Auxiliary Lemmas

Here we prove simple lemmas to be used in the proof of Theorem 1.4.

**Lemma 2.16.** *Let  $G$  and  $G'$  be graphs on  $n$  vertices, such that one can edit (i.e. add or remove)  $\frac{\delta}{t^2}n^2$  edges of  $G$  and get  $G'$ . Then for every fixed graph  $T$  and family of graphs  $\mathcal{F}$ ,*

$$|ex(G, T, \mathcal{F}) - ex(G', T, \mathcal{F})| \leq \delta n^t.$$

*Proof.* Let  $E$  be the set of edges we need to delete from  $G$  to make it a subgraph of  $G'$ , and let  $G_0$  be the subgraph of  $G$  which is  $\mathcal{F}$ -free and has the maximum possible number of copies of  $T$ . If  $G'_0$  is the subgraph of  $G_0$  obtained by deleting from it any edges from  $E$  then  $G'_0$  is an  $\mathcal{F}$ -free subgraph of  $G'$  and the following holds

$$ex(G', T, \mathcal{F}) \geq \mathcal{N}(G'_0, T) \geq \mathcal{N}(G_0, T) - \frac{\delta}{t^2}t^2n^2n^{t-2} = ex(G, T, \mathcal{F}) - \delta n^t.$$

As this is symmetric for  $G$  and  $G'$  we get the needed result.  $\square$

**Lemma 2.17.** *For every  $\epsilon > 0$  and  $r \in \mathbb{N}$  there is  $\delta_{2.17} := \delta_{2.17}(\epsilon, r)$  that tends to zero as  $\epsilon$  tends to zero such that for every set of constants  $0 < \alpha_1, \dots, \alpha_r < 1$*

$$\prod_{i=1}^r (\alpha_i - \epsilon) > \prod_{i=1}^r \alpha_i - \delta_{2.17}$$

*Proof.* As  $\alpha_i < 1$  the following holds

$$\begin{aligned} \prod_{i=1}^r (\alpha_i - \epsilon) &= \prod_{i=1}^r \alpha_i + \sum_{i=1}^r (-\epsilon)^i \sum_{I \subset [r], |I|=r-i} \prod_{j \in I} \alpha_j \\ &\geq \prod_{i=1}^r \alpha_i - \sum_{i=1}^r \epsilon^i \sum_{I \subset [r], |I|=r-i} \prod_{j \in I} \alpha_j \\ &\geq \prod_{i=1}^r \alpha_i - \sum_{i=1}^r \epsilon^i \binom{r}{i} \\ &= \prod_{i=1}^r \alpha_i - ((1 + \epsilon)^r - 1) \end{aligned}$$

Thus taking  $\delta_{2.17} = ((1 + \epsilon)^r - 1)$  gives the needed result.  $\square$

## 3 The Main Lemma

The algorithm for Theorem 1.4 gets a graph  $G$  and uses the regularity lemma to get a partition graph  $W$ , and then solves the extremal question for  $W$ . To prove the correctness of the algorithm we need to show that indeed the answer for  $W$  gives a good estimate for the original extremal question on the graph  $G$ .

**Lemma 3.1.** *Let  $T$  be a fixed graph,  $\mathcal{F}$  a finite family of graphs,  $\epsilon > 0$  and  $k \in \mathbb{N}$ . Then there exists  $\delta_{3.1} := \delta_{3.1}(\epsilon, T, \mathcal{F})$  that tends to zero as  $\epsilon$  tends to zero, for which the following holds. Let  $G$  be a graph on  $n$  vertices, let  $P$  be an  $f(k)$ -partition of  $G$  into  $k$  parts as in Theorem 2.5 where  $f(k) = \min\{\epsilon, k \cdot K_{2.3}^{-1}\}$  and let  $W$  be the  $(f(k), d_{2.8})$  partition graph it gives. Then*

$$\text{ex}(G, T, \mathcal{F}) \leq \left(\frac{n}{k}\right)^t \text{ex}_{hom}(W, T, \mathcal{F}) + \delta_{3.1} n^t$$

*Proof of Lemma 3.1.* Let  $W_0$  be a conventional subgraph of  $W$  such that  $\mathcal{N}(W_0, T) = \text{ex}_{hom}(W, T, \mathcal{F})$ . Assume towards contradiction that  $G$  has an  $\mathcal{F}$ -free subgraph  $G_{\text{ex}}$  such that

$$\mathcal{N}(G_{\text{ex}}, T) > \mathcal{N}(W_0, T) \left(\frac{n}{k}\right)^t + \delta_{3.1} n^t. \quad (2)$$

where  $\delta_{3.1}$  is chosen in the end and tends to zero as  $\epsilon$  tends to zero.

Using Theorem 2.3 we can find a refinement of  $P = \{V_i\}_{i=1}^k$  which is an  $\epsilon$ -regular partition of  $G_{\text{ex}}$  into  $k_0 \cdot k < K_{2.3}$  parts where  $k_0 \geq \frac{1}{\epsilon}$ . Call this partition  $Q = V_0 \cup \{\{V_{i,j}\}_{j=1}^{k_0}\}_{i=1}^k$ , where  $V_{i,j} \subset V_i$ . Note that as  $|V_0| < K_{2.3}$  and all of the other sets are of equal size, we get that

$$|V_{i,j}| > K_{2.3}^{-1} \cdot n = k(K_{2.3})^{-1} \cdot \frac{n}{k} \geq kK_{2.3}^{-1} \cdot |V_i|,$$

and so as  $P$  is an  $f$ -regular partition with  $f(k) \leq kK_{2.3}^{-1}$  it holds that

$$d_G(V_i, V_j) > d_G(V_{i,l_i}, V_{j,l_j}) - \epsilon \quad (3)$$

for every choice of  $1 < l_i, l_j < k_0$  and  $1 < i, j < k$ , where  $d_G$  is the density in  $G$ . Note that here we used the strong regularity in an essential way, as  $|V_{i,l_i}|$  is far smaller than  $|V_i|$ .

Let  $W^*$  be the  $(\epsilon, d_{2.8})$  partition graph of  $Q$  and  $G_{\text{ex}}$ . Note that by Lemma 2.10 and by (2) this means that

$$\left(\frac{n}{k \cdot k_0}\right)^t \mathcal{N}(W^*, T) \geq \mathcal{N}(G_{\text{ex}}, T) - \delta_{2.10} n^t \geq \left(\frac{n}{k}\right)^t \mathcal{N}(W_0, T) + (\delta_{3.1} - \delta_{2.10}) n^t$$

and in particular

$$\left(\frac{1}{k_0}\right)^t \mathcal{N}(W^*, T) \geq \mathcal{N}(W_0, T) + (\delta_{3.1} - \delta_{2.10}) k^t. \quad (4)$$

To obtain a contradiction to the existence of  $G_{\text{ex}}$  we first prove the following lemma

**Lemma 3.2.** *There exists a choice of  $r_i \in [k_0]$  for every  $i \in [k]$  such that the following holds. Let  $W^{**}$  be the subgraph of  $W^*$  spanned by the vertices  $v_{1,r_1}, \dots, v_{k,r_k}$  then*

$$k_0^t \mathcal{N}(W^{**}, T) \geq \mathcal{N}(W^*, T) - \epsilon_{2.9} (k \cdot k_0)^t$$

*Proof.* Let  $\mathcal{T}^*$  be the set of copies of  $T$  in  $W^*$  such that for every two vertices of the copy,  $v_{i,k_i}$  and  $v_{j,k_j}$ ,  $i \neq j$ . Using Lemma 2.9 we get that

$$\sum_{T \in \mathcal{T}^*} \prod_{(v_i, v_j) \in E(T)} w^*(v_i, v_j) > \mathcal{N}(W^*, T) - \epsilon_{2.9} (k \cdot k_0)^t.$$

For every  $i$  let us choose  $r_i$  uniformly at random, and let  $W'$  be the subgraph spanned by the chosen vertices. For  $T \in \mathcal{T}^*$  let  $A(T)$  be the event that all of the vertices of  $T$  are in  $W'$ . Then

$$\begin{aligned}\mathbb{E}[k_0^t \mathcal{N}(W', T)] &= \mathbb{E}\left[k_0^t \sum_{T \in \mathcal{T}^*} \mathbf{1}_{A(T)} \prod_{v_i, v_j \in E(T)} w^*(v_i, v_j)\right] = \\ &= k_0^t \sum_{T \in \mathcal{T}^*} \frac{1}{k_0^t} \prod_{v_i, v_j \in E(T)} w^*(v_i, v_j) > \mathcal{N}(W^*, T) - \epsilon_{2.9}(k \cdot k_0)^t\end{aligned}$$

Thus, there must be some choice of  $r_1, \dots, r_k$  that gives the needed inequality.  $\square$

Given the graph  $W^{**}$ , that is the choice of  $r_i$  for every  $i$ , we can now find an  $\mathcal{F}_{hom}$ -free conventional subgraph of  $W$ , say  $W_1$ , for which  $\mathcal{N}(W_1, T) > \mathcal{N}(W_0, T)$ . Take  $(v_i, v_j)$  to be an edge of  $W_1$  if and only if  $(v_{i, r_i}, v_{j, r_j})$  is an edge of  $W^{**}$ . First note that by (3) and the fact that  $G_{ex}$  is a subgraph of  $G$  we get that

$$\begin{aligned}w_1(v_i, v_j) &= d_G(V_i, V_j) \geq d_G(V_{i, r_i}, V_{j, r_j}) - \epsilon \\ &\geq d_{G_{ex}}(V_{i, r_i}, V_{j, r_j}) - \epsilon = w^{**}(v_{i, r_i}, v_{j, r_j}) - \epsilon.\end{aligned}\tag{5}$$

Second, the mapping  $\varphi : V(W^{**}) \rightarrow V(W_1)$ ,  $\varphi(v_{i, r_i}) = v_i$  maps edges with non-negative weight to edges with non-negative weight. Thus as  $W^{**}$  is  $\mathcal{F}_{hom}$ -free then so is  $W_1$ . Finally, note that if  $\mathcal{T}(W_1)$  and  $\mathcal{T}(W^{**})$  are the copies of  $T$  in  $W_1$  and  $W^{**}$  respectively, then  $\varphi$  is a bijection between them. Thus, the following holds

$$\begin{aligned}\mathcal{N}(W_1, T) &= \sum_{T \in \mathcal{T}(W_1)} \prod_{(v_i, v_j) \in E(T)} w_1(v_i, v_j) \\ &= \sum_{T \in \mathcal{T}(W^{**})} \prod_{(v_{i, r_i}, v_{j, r_j}) \in E(T)} w_1(\varphi^{-1}(v_{i, r_i}), \varphi^{-1}(v_{j, r_j})) \\ &\stackrel{(5)}{\geq} \sum_{T \in \mathcal{T}(W^{**})} \prod_{(v_{i, r_i}, v_{j, r_j}) \in E(T)} (w^{**}(v_{i, r_i}, v_{j, r_j}) - \epsilon) \\ &\stackrel{2.17}{\geq} \mathcal{N}(W^{**}, T) - \delta_{2.17} k^t \\ &\stackrel{3.2}{>} \left(\frac{1}{k_0}\right)^t \mathcal{N}(W^*, T) - \epsilon_{2.9} k^t - \delta_{2.17} k^t \\ &\stackrel{(4)}{\geq} \mathcal{N}(W_0, T) + (\delta_{3.1} - \delta_{2.10} - \epsilon_{2.9} - \delta_{2.17}) k^t > \mathcal{N}(W_0, T)\end{aligned}$$

The last inequality holds since  $\delta_{3.1}$  can be chosen so that  $\delta_{3.1} > \delta_{2.10} + \epsilon_{2.9} + \delta_{2.17}$ .

The existence of such  $W_1$  is the needed contradiction, as we chose  $W_0$  to be a conventional  $\mathcal{F}_{hom}$ -free subgraph of  $W$  that has the maximum possible value of  $\mathcal{N}(W_0, T)$ . Thus there cannot be a graph  $G_{ex}$  which is  $\mathcal{F}$ -free and has more than  $(\frac{n}{k})^t \text{ex}_{hom}(W, T, \mathcal{F}) + \delta_{3.1} n^t$  copies of  $T$ .  $\square$

## 4 Proof of Theorem 1.4

The algorithm itself is rather straightforward. Let  $n_0 = m_{2.8} K_{2.5}$ , if  $V(G) < n_0$  we can use brute-force as there is a constant number of options to check. If  $V(G) > n_0$  then for  $\epsilon_1 > 0$  Theorem

2.5 gives us an efficient algorithm for finding  $\epsilon_1 n^2$  edges to add or remove from  $G$  and to find a  $k \cdot K_{2,3}^{-1}$ -regular partition  $P = V_0 \cup \{V_i\}_{i=1}^k$  of the edited graph, where  $K_{2,5} > k > \frac{1}{\epsilon_1}$ .

In Theorem 2.5 we change  $\epsilon_1 n^2$  edges, and as  $|V_0| < K_{2,5}$  we can remove all edges using a vertex from  $V_0$ , and change at most  $\epsilon_1 n^2 + K_{2,5}n < \epsilon_2 n^2$  edges of the original graph  $G$ . This will give us a graph that has an  $f(k)$ -partition of its vertices into  $k$  parts of equal size. Call this graph  $G^*$ .

By Lemma 2.16 if we prove that we found the approximated answer for  $G^*$ , up to an additive error of  $\epsilon_3 n^t$ , then we have also found an answer for  $G$  itself up to an additive error term of  $(\epsilon_2 + \epsilon_3)n^t$ . Thus we may focus on the graph  $G^*$ , remembering that

$$|\text{ex}(G, T, \mathcal{F}) - \text{ex}(G^*, T, \mathcal{F})| \leq (\epsilon_2 + \epsilon_3)n^t \quad (6)$$

Let  $W$  be the  $(\epsilon_1, d_{2,8})$  partition graph of  $P$  and  $G^*$ . We use brute-force to find a conventional subgraph of it, say  $W_0$ , which is  $\mathcal{F}_{hom}$ -free and maximizes  $\mathcal{N}(W_0, T)$ . As  $W$  has a constant number of vertices, we can do this in constant time (not depending on the size of  $G$ ). By Lemma 2.15  $G_{W_0}^*$  is  $\mathcal{F}$ -free, and so

$$\mathcal{N}(G_{W_0}^*, T) \leq \text{ex}(G^*, T, \mathcal{F})$$

To prove that  $G_{W_0}^*$  gives the needed approximation, it is left to show that

$$\mathcal{N}(G_{W_0}^*, T) \geq \text{ex}(G^*, T, \mathcal{F}) - \epsilon n^t.$$

Indeed Lemma 2.10 shows that

$$|\mathcal{N}(G_{W_0}^*, T) - \left(\frac{n}{k}\right)^t \mathcal{N}(W_0, T)| < \delta_{2,10} n^t. \quad (7)$$

By Lemma 3.1 there is no  $\mathcal{F}$ -free subgraph of  $G^*$  that has more than  $\left(\frac{n}{k}\right)^t \mathcal{N}(W_0, T) + \delta_{3,1} n^t$  copies of  $T$ , thus

$$|\text{ex}(G^*, T, \mathcal{F}) - \left(\frac{n}{k}\right)^t \mathcal{N}(W_0, T)| < \delta_{3,1} n^t. \quad (8)$$

Thus by using equations (6), (7), and (8) we get that

$$|\mathcal{N}(G_{W_0}^*, T) - \text{ex}(G, T, \mathcal{F})| < (\epsilon_2 + \epsilon_3 + \delta_{3,1} + \delta_{2,10}) n^t$$

and as  $\epsilon_2, \epsilon_3, \delta_{3,1}$  and  $\delta_{2,10}$  all tend to zero as  $\epsilon$  tends to zero, their sum can be made as small as needed.  $\square$

## 5 Proof of Theorem 1.7

Before we prove the theorem, we establish the following two simple lemmas.

**Lemma 5.1.** *For any integers  $2 \leq m < k$  it is NP-hard to calculate  $\text{ex}(G, K_m, K_k)$  exactly.*

*Proof.* We show that calculating  $\text{ex}(G, K_m, K_k)$  exactly is at least as hard as calculating  $\text{ex}(G, K_2, K_k)$  exactly, which by Theorem 1.2 is NP-hard.

Given a graph  $G$ , we construct a graph  $G'$  by adding  $n^{m+1}$  disjoint copies of  $K_m$  on each edge. These copies use  $m-2$  new vertices each, and a single edge from  $G$ . As there are at most  $n^m$  copies of  $K_m$  in  $G$  itself, it is easy to check that

$$\text{ex}(G', K_m, K_k) \leq \text{ex}(G, K_2, K_k) n^{m+1} + n^m \leq \text{ex}(G', K_m, K_k) + n^m.$$

Therefore

$$\frac{\text{ex}(G', K_m, K_k)}{n^{m+1}} - \frac{1}{n} \leq \text{ex}(G, K_2, K_k) \leq \frac{\text{ex}(G', K_m, K_k)}{n^{m+1}}$$

implying that for  $n > 1$   $\text{ex}(G, K_2, K_k)$  can be calculated exactly from the exact value of  $\text{ex}(G', K_m, K_k)$ .  $\square$

**Lemma 5.2.** *Let  $G$  be a graph, let  $N$  be an integer and let  $G'$  be the  $N$ -blowup of  $G$ . Then for every  $2 \leq m < k$  we have  $\text{ex}(G', K_m, K_k) = \text{ex}(G, K_m, K_k) \cdot N^m$ .*

*Proof.* Suppose without loss of generality that  $V(G) = [n]$ . For each  $i \in [n]$  let  $U_i \subseteq V(G')$  be the blowup-set corresponding to the vertex  $i \in V(G)$ . Thus  $|U_1| = \dots |U_n| = N$  and  $V(G')$  is the disjoint union  $U_1 \cup \dots \cup U_n$ . Let  $F$  be a  $K_k$ -free subgraph of  $G$  satisfying  $\mathcal{N}(F, K_m) = \text{ex}(G, K_m, K_k)$ . Then the  $N$ -blowup of  $F$  is a  $K_k$ -free subgraph of  $G'$  with  $\mathcal{N}(F, K_m) \cdot N^m = \text{ex}(G, K_m, K_k) \cdot N^m$  copies of  $K_m$ . This shows that  $\text{ex}(G', K_m, K_k) \geq \text{ex}(G, K_m, K_k) \cdot N^m$ . In the other direction, let  $F'$  be a  $K_k$ -free subgraph of  $G'$  satisfying  $\mathcal{N}(F', K_m) = \text{ex}(G', K_m, K_k)$ . For each  $1 \leq i \leq n$ , let  $u_i$  be a random vertex of  $U_i$ . Set  $F := F'[\{u_1, \dots, u_n\}]$ . By linearity of expectation, we have

$$\mathbb{E}[\mathcal{N}(F, K_m)] = \mathcal{N}(F', K_m) \cdot N^{-m}.$$

On the other hand, note that  $F$  is  $K_k$ -free (as it is a subgraph of  $F'$ ), and that  $F$  is a subgraph of  $G$ . It follows that  $\mathcal{N}(F, K_m) \leq \text{ex}(G, K_m, K_k)$  with probability 1. We conclude that

$$\text{ex}(G', K_m, K_k) \cdot N^{-m} = \mathcal{N}(F', K_m) \cdot N^{-m} = \mathbb{E}[\mathcal{N}(F, K_m)] \leq \text{ex}(G, K_m, K_k),$$

as required. This completes the proof.  $\square$

Now we are ready to prove Theorem 1.7

*Proof.* We show that the problem of approximating  $\text{ex}(G, K_m, K_k)$  up to additive error  $|V(G)|^{m-\varepsilon}$  is as hard as the problem of computing  $\text{ex}(G, K_m, K_k)$  exactly. Let  $G$  be a graph on  $n$  vertices. Set  $N = (2n)^{m/\varepsilon}$ , and let  $G'$  be the  $N$ -blowup of  $G$ . Note that  $G'$  can be constructed from  $G$  in polynomial time. Suppose  $R$  is an integer satisfying  $|R - \text{ex}(G', K_m, K_k)| \leq |V(G')|^{m-\varepsilon} = (nN)^{m-\varepsilon} < \frac{1}{2}N^m$ . By Lemma 5.2, we have  $|\frac{R}{N^m} - \text{ex}(G, K_m, K_k)| < \frac{1}{2}$ . It follows that  $\text{ex}(G, K_m, K_k)$  is the integer nearest to the number  $\frac{R}{N^m}$ . Therefore if we can approximate  $\text{ex}(G', K_m, K_k)$  up to additive error  $|V(G')|^{m-\varepsilon}$  in polynomial time, then we can also compute  $\text{ex}(G, K_m, K_k)$  exactly in polynomial time, where here we use the fact that  $|V(G')| = nN$  is polynomial in  $n$ . But this is NP-hard by Lemma 5.1, completing the proof.  $\square$

## 6 Concluding remarks and open questions

### 6.1 Improving on Theorem 1.7

In the introduction we have posed the following conjecture,

**Conjecture 1.8** *For every graph  $T$ , family of graphs  $\mathcal{F}$  such that no  $F \in \mathcal{F}$  is a subgraph of a blow-up of  $T$ , and  $\varepsilon > 0$ , it is NP-hard to approximate  $\text{ex}(G, T, \mathcal{F})$  up to additive error of  $n^{v(T)-\varepsilon}$  for a given input graph  $G$  on  $n$  vertices.*

Allowing a smaller additive error, we can get the above for a large family of graphs.

**Proposition 6.1.** *For any  $\epsilon > 0$ , 3-connected graph  $T$  on  $t$  vertices and a family of 3-connected graphs  $\mathcal{F}$  such that no  $F \in \mathcal{F}$  is a subgraph of a blow-up of  $T$ , it is NP-hard to approximate  $\text{ex}(G, T, \mathcal{F})$  up to additive error of  $n^{t-2-\epsilon}$ .*

*Proof.* We show that the question of calculating  $\text{ex}(G, K_2, \mathcal{F})$  exactly can be reduced to the question of approximating  $\text{ex}(G, T, \mathcal{F})$  up to additive error of  $n^{c(t-2)/(c+2)-\epsilon'}$  where  $c$  is a constant as large as needed and  $0 < \epsilon' < \epsilon$ . As every  $F \in \mathcal{F}$  is not a subgraph of a blow-up of  $T$ , in particular it is not bipartite, and thus the former is known to be NP-hard by Theorem 1.2. As  $c$  can be as large as we want it can be chosen so that  $c(t-2)/(c+2) - \epsilon' > t-2-\epsilon$ .

Given a graph  $G$  on  $n$  vertices, let us construct  $G^+$  as follows. Let  $T'$  be a subgraph of  $T$  obtained by removing two arbitrary vertices connected by an edge, say  $\{u, v\}$ . For each edge  $e \in G$  add  $t-2$  independent sets of size  $n^c$  to the graph. Connect the appropriate pairs of sets by complete bipartite graphs to create a blow-up of  $T'$ , and then connect sets corresponding to neighbors of  $u$  and  $v$  in  $T$  to the two endpoints of  $e$ . Call the new copies of  $T$  created *external* and the copies spanned by  $G$  *internal*.

Note that  $N := v(G^+) \leq n + (t-2)n^c n^2$  and  $N^{c(t-2)/(c+2)-\epsilon'} < \frac{1}{5}n^{c(t-2)}$ . Furthermore, in  $G^+$  every edge of  $G$  takes part in a fixed number of external copies of  $T$  in  $G^+$ , say  $X_T \geq n^{c(t-2)}$  and in at most  $O(n^{t-2})$  internal copies. In addition, no new copies of graphs from  $\mathcal{F}$  are created. This is true as  $T$  and the graphs in  $\mathcal{F}$  are 3-connected and no graph in  $\mathcal{F}$  is a subgraph of a blow-up of  $T$ .

Let  $\overline{\text{ex}}(G^+, T, \mathcal{F}) = \mathcal{N}(G^+, T) - \text{ex}(G^+, T, \mathcal{F})$  where  $\mathcal{N}(G^+, T)$  is the number of copies of  $T$  in  $G^+$ , and similarly  $\overline{\text{ex}}(G, K_2, \mathcal{F}) = |E(G)| - \text{ex}(G, K_2, \mathcal{F})$ . First note that

$$\overline{\text{ex}}(G^+, T, \mathcal{F}) \leq \overline{\text{ex}}(G, K_2, \mathcal{F})X_T + O(n^t).$$

Indeed, by deleting  $\overline{\text{ex}}(G, K_2, \mathcal{F})$  edges from  $G$  we can make  $G^+$  into an  $\mathcal{F}$ -free graph, as all of the copies of  $\mathcal{F}$  are spanned by the vertices coming from  $G$ . Removing these edges will remove all of the external copies of  $T$  using them together with some internal copies.

Furthermore, if we removed  $\overline{\text{ex}}(G^+, T, \mathcal{F})$  copies of  $T$  from  $G^+$  and made it  $\mathcal{F}$ -free, we may assume that we have done this by removing only edges from  $G$ , say  $e$  of them, and that  $G$  was made  $\mathcal{F}$ -free. Each edge of  $G$  takes part in at least  $X_T$  distinct copies of  $T$ , so

$$\overline{\text{ex}}(G^+, T, \mathcal{F}) \geq eX_T \geq \overline{\text{ex}}(G, K_2, \mathcal{F})X_T.$$

From the above

$$\frac{\overline{\text{ex}}(G^+, T, \mathcal{F}) - O(n^t)}{X_T} \leq \overline{\text{ex}}(G, K_2, \mathcal{F}) \leq \frac{\overline{\text{ex}}(G^+, T, \mathcal{F})}{X_T},$$

and as

$$\frac{N^{c(t-2)/(c+2)-\epsilon'}}{X_T} \leq \frac{1}{5} \text{ and } \frac{O(n^t)}{X_T} < \frac{1}{5},$$

if we calculated  $\overline{\text{ex}}(G^+, T, \mathcal{F})$  up to an additive error of  $N^{c(t-2)/(c+2)-\epsilon'}$  then we calculated  $\overline{\text{ex}}(G, K_2, \mathcal{F})$  up to an additive error of  $2/5$ , as this is an integer this means we have calculated it exactly and this is known to be NP-hard.  $\square$

The full assertion of Conjecture 1.8 remains open. The following questions address several special cases.

1. Is approximating  $\text{ex}(G, K_m, K_{m+1})$  up to additive error of  $n^{m-\epsilon}$  NP-hard for any integer  $m \geq 2$  and any  $\epsilon > 0$ ? The case  $m = 2$  is proved in [1] and we can also prove it for  $m = 3$ .
2. Given a family of graphs  $\mathcal{F}$  such that  $\chi(F) \geq m + 2$  for every  $F \in \mathcal{F}$ , is it NP-hard to approximate  $\text{ex}(G, K_m, \mathcal{F})$  up to additive error of  $n^{m-\epsilon}$  for every  $\epsilon > 0$ ? Theorem 1.7 is of course a special case of this.

## 6.2 Calculating $\text{ex}(G, T, F)$ exactly

Proposition 6.1 implies that for many graphs  $T$  and  $F$  there is no efficient algorithm that calculates  $\text{ex}(G, T, F)$  exactly. Nevertheless, for some special cases this calculation is possible in polynomial time. We mention two simple examples.

**Proposition 6.2.** *For a graph  $G$  on  $n$  vertices the following can be solved in polynomial time*

1.  $\text{ex}(G, kK_2, K_{1,2})$ , where  $kK_2$  is a matching of size  $k \geq 1$ .
2.  $\text{ex}(G, K_2, K_{1,t+1})$ .

*Proof.* Part 1 is trivial. It is known since Edmonds [10] that given an input graph  $G$  a matching of maximum size can be found in polynomial time. As any  $K_{1,2}$ -free subgraph of  $G$  is a matching it is clear that maximizing the largest matching also maximizes the number of copies of  $kK_2$ .

The proof of part 2 follows the idea of the proof of the f-factor theorem of Tutte [21]. Given a graph  $G$ , we may assume that it has no isolated vertices. First, replace each vertex  $v \in V(G)$  with an independent set of size  $d(v)$ , say  $V(v) = \{v_1, \dots, v_{d(v)}\}$ , and for every edge  $e = \{u, v\}$  of  $G$  choose arbitrarily vertices  $v_i, u_j$  and connect them, making sure that at the end of the processes each vertex of the new graph is of degree exactly one. Note that there is a one-to-one correspondence between edges in the new graph and edges in  $G$ .

Second, for every independent set  $V(v)$  corresponding to a vertex  $v \in V(G)$  such that  $d(v) \geq t + 1$ , add a new independent set  $U(v)$  of size  $d(v) - t$ . Connect all of the vertices in  $V(v)$  to all the vertices in  $U(v)$ . Call the new graph obtained by this processes  $G^*$ .

By [10] there is a polynomial time algorithm that finds a maximum matching in  $G^*$ , call this maximum matching  $M$ . Note that we may assume that  $M$  saturates all the vertices in the sets  $U(v)$ . Indeed, if some  $w \in U(v)$  is not saturated we may add to  $M$  an edge between  $w$  and some vertex of  $V(v)$ , say  $v_1$ , that is not adjacent to  $U(v)$  in  $M$  and if prior to this addition there was an edge in  $M$  adjacent to  $v_1$ , delete it. A vertex  $v_1$  must exist as  $V(v) > U(v)$  and the replacement of edges does not make the size of  $M$  smaller.

Each edge in  $M$  that is not adjacent to some  $U(v)$  corresponds to an edge of  $G$ . Let us keep in  $G$  only these edges and call the new graph  $G'$ . For every vertex  $v$  the set  $U(v)$  was saturated in  $M$  and so the number of edges adjacent to  $V(v)$  and a vertex not in  $U(v)$  is at most  $|V(v)| - |U(v)| = d_G(v) - (d_G(v) - t) = t$ . Thus the new graph  $G'$  is a subgraph of  $G$  with maximum degree at most  $t$ . The number of edges in  $G'$  is exactly the number of edges in  $M$  that are not adjacent to some  $U(v)$ , that is  $e(G') = |M| - \sum_{v \in V(G)} (d_G(v) - t)$ .

It is left to show that  $G'$  has the maximum possible number of edges. Indeed, assume that  $G''$  is a subgraph of  $G$  with maximum degree at most  $t$  which has more edges than  $G'$ . Looking at  $G^*$  take into a matching  $M'$  each edge corresponding to an edge in  $G''$ . This results in a matching as each  $v_i \in V(v)$  has exactly one neighbor outside of  $U(v)$ . In addition, for each  $v \in V(G)$  choose

$d_G(v) - t$  edges between  $U(v)$  and  $V(v)$  to add to  $M'$  while keeping it a matching. As the maximum degree in  $G''$  is at most  $t$  there will be  $|U(v)| = d_G(v) - t$  unsaturated vertices in  $V(v)$  to connect to  $U(v)$ . The resulting matching will be of size  $e(G'') + \sum_{v \in V} (d_G(v) - t) > |M|$ , in contradiction to the maximality of  $M$ .

Thus we get that  $\text{ex}(G, K_2, K_{1,t+1}) = e(G') = |M| - \sum_{v \in V(G)} (d_G(v) - t) = |M| - \frac{1}{2}e(G) + v(G)t$ , and we can find  $|M|, e(G)$ , and  $v(G)$  in polynomial time.  $\square$

It would be interesting to characterize all pairs of graphs  $T$  and  $F$  for which  $\text{ex}(G, T, F)$  can be calculated exactly in polynomial time, for a given input graph  $G$ .

**Acknowledgments** We thank two anonymous referees for helpful comments, and thank Lior Gishboliner for improving and simplifying the proof of Theorem 1.7.

## References

- [1] N. Alon, A. Shapira, and B. Sudakov, Additive approximation for edge-deletion problems, *Annals of Mathematics* (2009): 371-411.
- [2] N. Alon, A. Shapira, and U. Stav, Can a graph have distinct regular partitions?, *SIAM Journal on Discrete Mathematics* 23.1 (2009): 278-287.
- [3] N. Alon, and C. Shikhelman, Many T copies in H-free graphs, *Journal of Combinatorial Theory, Series B* 121 (2016): 146-172.
- [4] T. Asano, An application of duality to edge-deletion problems, *SIAM J. on Computing*, 16 (1987), 312-331.
- [5] T. Asano and T. Hirata, Edge-deletion and edge-contraction problems, *Proc. of STOC* (1982), 245-254.
- [6] N. Bansal, A. Blum and S. Chawla, Correlation clustering, *Proc. of the 43rd IEEE FOCS* (2002), 238-247
- [7] L. Cai, Fixed-parameter tractability of graph modification problems for hereditary properties, *Information Processing Letters*, 58.4 (1996): 171-176.
- [8] K. Cirino, S. Muthukrishnan, N. Narayanaswamy and H. Ramesh, graph editing to bipartite interval graphs: exact and asymptotic bounds, *Proc. of 17th FSTTCS* (1997), 37-53.
- [9] R.A. Duke, H. Lefmann, and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, *SIAM Journal on Computing* 24.3 (1995): 598-620.
- [10] J. Edmonds, Paths, trees, and flowers, *Canadian Journal of Mathematics* 17.3 (1965): 449-467.
- [11] E.S. El-Mallah, C.J. Colbourn, The complexity of some edge deletion problems, *IEEE transactions on circuits and systems* 35.3 (1988): 354-362.
- [12] P. Frankl, and V. Rödl, Extremal problems on set systems, *Random Structures & Algorithms*, 20(2), (2002) 131-164.

- [13] M.R. Garey and D.S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman and Co., San Francisco, 1979.
- [14] P. W. Goldberg, M. C. Golumbic, H. Kaplan and R. Shamir, Four strikes against physical mapping of DNA, *Journal of Computational Biology* 2 (1995), 139-152.
- [15] M. C. Golumbic, H. Kaplan and R. Shamir, On the complexity of DNA physical mapping, *Advances in Applied Mathematics*, 15 (1994), 251-261.
- [16] S. Khot, and V. Raman, Parameterized complexity of finding subgraphs with hereditary properties, *Theoretical Computer Science* 289.2 (2002): 997-1008.
- [17] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, The regularity lemma and its applications in graph theory, *Theoretical aspects of computer science*, (2002): 84-112.
- [18] A. Natanzon, R. Shamir and R. Sharan, Complexity classification of some edge modification problems, *Discrete Applied Mathematics* 113 (2001), 109-128
- [19] V. Rödl and M. Schacht, Regular partitions of hypergraphs: counting lemmas, *Combinatorics, Probability and Computing* 16.6 (2007): 887-901.
- [20] E. Szemerédi, Regular partitions of graphs, *Colloques Internationaux C.N.R.S. No 260 - Problèmes Combinatoires et Théorie des Graphes*, Orsay (1976).
- [21] W.T. Tutte, A short proof of the factor theorem for finite graphs, *Canadian Journal of mathematics*, 6, (1954): 347-352.
- [22] M. Yannakakis, Edge-deletion problems, *SIAM J. Comput.* 10 (1981), 297-309.