The Removal Lemma for Tournaments

Jacob Fox * Lior Gishboliner † Asaf Shapira ‡ Raphael Yuster §

November 13, 2018

Abstract

Suppose one needs to change the direction of at least ϵn^2 edges of an n-vertex tournament T, in order to make it H-free. A standard application of the regularity method shows that in this case T contains at least $f_H^*(\epsilon)n^h$ copies of H, where f_H^* is some tower-type function. It has long been observed that many graph/digraph problems become easier when assuming that the host graph is a tournament. It is thus natural to ask if the removal lemma becomes easier if we assume that the digraph G is a tournament.

Our main result here is a precise characterization of the tournaments H for which $f_H^*(\epsilon)$ is polynomial in ϵ, stating that such a bound is attainable if and only if H's vertex set can be partitioned into two sets, each spanning an acyclic directed graph. The proof of this characterization relies, among other things, on a novel application of a regularity lemma for matrices due to Alon, Fischer and Newman, and on probabilistic variants of Ruzsa-Szemerédi graphs.

We finally show that even when restricted to tournaments, deciding if H satisfies the condition of our characterization is an NP-hard problem.

1 Introduction

1.1 Background and motivation

Suppose an n-vertex graph G contains cn^h copies of an h-vertex graph H. It is clear that in this case one should remove at least $c'n^2$ edges in order to turn G into an H-free graph. The celebrated removal lemma of Ruzsa and Szemerédi [17] states that (at least qualitatively) this sufficient condition is in fact necessary. More precisely, it states that there is a function $f_H(\epsilon)$ so that if one needs to remove at least ϵn^2 edges from an n-vertex graph G in order to make it H-free, then G contains at least $f_H(\epsilon)n^h$ copies of H. Besides its intrinsic interest, the removal lemma was extensively studied also due to its many applications. See [10] for more background on the lemma and its many variants.

All proofs of the removal lemma apply some version of Szemerédi's regularity lemma [19], and thus can only bound $f_H(\epsilon)$ by tower-type functions of ϵ. It is a major open problem in extremal graph theory to decide if a non-tower-type bound can be obtained even for the special case of $H = K_3$. Given the above, it is thus natural to ask for which graphs H one can obtain very efficient bounds for

*Stanford University, Stanford, CA. Supported by a Packard Fellowship, by NSF CAREER award DMS 1352121, and by an Alfred P. Sloan Fellowship. Email: jacobfox@stanford.edu.

†School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: liorgis1@post.tau.ac.il.

‡School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: asafico@tau.ac.il. Supported in part by ISF Grant 1028/16 and ERC-Starting Grant 633509.

§Department of Mathematics, University of Haifa, Haifa 31905, Israel. Email: raphy@math.haifa.ac.il. Supported in part by ISF Grant 1028/16.

$f_H(\epsilon)$. The first result of this type was obtained by Alon [1] who proved that $f_H(\epsilon)$ is polynomial in ϵ if and only if H is a bipartite graph. Alon and Shapira [4] considered the analogous question in the setting of directed graphs and proved that $f_H(\epsilon)$ is polynomial if and only if H has a homomorphism into an oriented tree or a 2-cycle, where an *oriented graph* $H = (V, E)$ is an orientation of an undirected graph, that is, a directed graph in which, for every pair of distinct vertices $x, y \in V$, there is at most one edge between x and y.

Our focus in this paper is in studying analogous questions in the setting of tournaments. The precise definition is the following: Suppose H is a fixed oriented graph. We say that an n-vertex tournament T is ϵ-far from being H-free if one should change the direction of at least ϵn^2 edges in order to turn T into an H-free tournament. It is not hard to apply the regularity method, in a way similar to [2], and show that if T is ϵ-far from being H free then T contains $f^*_H(\epsilon)n^h$ copies of H, where again $f^*_H(\epsilon)$ is a tower-type function. The question we are interested in is then for which H can $f^*_H(\epsilon)$ be bounded by a polynomial in ϵ?

As is well known, tournaments possess many properties not shared by general oriented graphs. As a result, many problems that are hard to resolve in general oriented graphs, become easier for tournaments. It is thus natural to ask if the removal lemma is easier for tournaments? Let us mention that it is known that there is at least one H for which the removal lemma is known to be easier for tournaments. Indeed, it follows from the result of [4] that $f_{C_3}(\epsilon)$ is not polynomial [1] in ϵ, while Fox and Sudakov [12] proved that $f^*_{C_3}(\epsilon)$ is polynomial in ϵ.

Let us conclude by mentioning that a further motivation for this paper was the work of Berger et al. [8] on tournaments they called *Heroes*. See [18] for more details. The work of [8] is another nice example of a phenomenon that holds in tournaments but fails to hold for general digraphs. One of the notions studied in [8] is the chromatic number of a tournament T defined as the smallest number of transitive tournaments which cover $V(T)$. As Theorem 1.2 shows, this notion is also relevant in our setting.

1.2 Our main results

Our main result in this paper gives a precise characterization of the oriented graphs H for which one can prove a removal lemma in tournaments with a polynomial bound. Let us say that an oriented graph H is *easy* if there is a constant $c = c(H)$ satisfying $f^*_H(\epsilon) \geq \epsilon^c$ for every sufficiently small $\varepsilon > 0$. If H is not easy then it is *hard*.

Theorem 1.1. *H is easy if and only if $V(H)$ can be partitioned into 2 vertex sets, each spanning an acyclic directed graph.*

It was shown in [4] that an oriented graph H satisfies $f_H(\epsilon) \geq \epsilon^c$ in general digraphs, only if H has a homomorphism into an oriented tree. Observe that Theorem 1.1 shows that $f^*_H(\epsilon)$ is polynomial for a much wider class of oriented graphs, that is, there is an entire family of oriented graphs for which the removal lemma is easier for tournaments.

We believe that the proofs of both directions of Theorem 1.1 are of independent interest. We note that the analogous "if" parts in the characterizations given in [1, 4] followed from simple density/Turán type arguments. For example, the fact that a bipartite H is easy in undirected graphs, follows from the simple reason that a graph with ϵn^2 edges contains at least $\epsilon^{h^2} n^h$ copies of H. In contrast, our proof requires a much more elaborate argument: we first show that for every oriented graph H as in the theorem there is an oriented complete bipartite graph, so that

[1] This special case of the result of [4] is actually implicit already in [17]

no matter how one completes this bipartite digraph into a tournament, one always ends up with a tournament containing a copy of H (see Lemma 3.4). We then combine this with a novel application of an efficient "conditional regularity lemma" for matrices of Alon, Fischer and Newman [3] (see also [11] and [14] for related results) in order to complete the proof. As to the "only if" part, as in previous lower bounds for removal lemmas, we also make use of variants of Ruzsa-Szemerédi [17] graphs. Our construction however, requires several additional twists such as the notion of ordered homomorphisms defines in Section 4, and the probabilistic construction from Section 2.

Let us conclude by describing our final result. It is natural to ask if the characterization given in Theorem 1.1 is "efficient", that is, how hard is it to tell if an oriented graph H is easy. It follows from the work of Bokal et al. [9] that this task is in fact NP-hard. Continuing with the theme of studying whether problems become easier when restricted to tournaments, it is natural to ask if one can at least recognize tournaments whose vertex set can be partitioned into 2 sets, each spanning an acyclic directed graph, i.e. into two transitive tournaments. The following theorem strengthens the result of Bokal et al [9] by showing that the problem is hard even for tournaments.

Theorem 1.2. *For every $k \geq 2$, the problem of deciding if a tournament is k-colorable is NP-hard.*

1.3 Organization

In Section 2 we describe a probabilistic construction that will be crucial in the proofs of both directions of Theorem 1.1. The proof of the first direction of Theorem 1.1 is given in Section 3, while the second is given in Section 4. The proof of Theorem 1.2 is given in Section 5.

2 A Preliminary Lemma

A *proper k-coloring* (or simply *k-coloring*) of an oriented graph H is a partition of $V(H)$ into k sets, each inducing an acyclic digraph. We say that H is *k-colorable* if it has a proper k-coloring. Notice that if H is a tournament then this definition coincides with the definition of a k-colorable tournament.

In this section we prove Lemma 2.1, stated below, which will be a key ingredient in the proof of both directions of Theorem 1.1. We start with some notation which we will also use in later sections. For a pair of vertices $x, y \in V$, we write $x \to y$ to mean that $(x, y) \in E$. For a pair of disjoint subsets $X, Y \subseteq V$, we use the notation $X \to Y$ to mean that there is no $(x, y) \in X \times Y$ for which $y \to x$. In other words, $X \to Y$ means that for every $(x, y) \in X \times Y$, either $x \to y$ or there are no edges between x and y.[2] Evidently, if the digraph is a tournament then $X \to Y$ is equivalent to saying that $x \to y$ for every $(x, y) \in X \times Y$. For a digraph G and a set $X \subseteq V(G)$, we use $G[X]$ to denote the subdigraph of G induced by X.

A *k-partite tournament* is an orientation of a complete k-partite graph. Notice that a bipartite tournament (i.e., a k-partite tournament for $k = 2$) is not the same as a 2-colorable tournament. A *completion* of a k-partite tournament $F = (V_1 \cup V_2 \cdots \cup V_k, E)$ is any tournament on $V(F)$ that agrees with F on the edges between the sets $V_1, ..., V_k$, i.e. any tournament obtained from F by adding k arbitrary tournaments on the sets $V_1, \ldots, V_k$.

Lemma 2.1. *For every $h \geq 2$ there are $m_0 = m_0(h)$ and $\gamma = \gamma(h) > 0$ with the following property. Let H be an oriented graph on h vertices and let D be an oriented graph on $[k]$, where $2 \leq k \leq h$.*

[2] This definition might seem strange as, for example, $x \to y$ is not the same as $\{x\} \to \{y\}$. Nevertheless, this definition is useful and should not cause confusion.

Suppose that H has a proper k-coloring, $V(H) = H_1 \cup \cdots \cup H_k$, such that $H_i \to H_j$ for every $(i,j) \in E(D)$. Then for every $m \geq m_0$ there is a k-partite tournament $F = (V_1 \cup \cdots \cup V_k, E(F))$ such that:

1. *$|V_i| = m$ for every $i = 1, ..., k$.*
2. *$V_i \to V_j$ for every $(i,j) \in E(D)$.*
3. *Every completion of F contains a collection of at least γm^2 copies of H with the property that every edge $e \in E(F)$ is contained in at most one of these copies.*

In the proof of Lemma 2.1 we use the following three claims. Denote by $\mathrm{Bin}(N,p)$ the binomial distribution with parameters N and p. We will need the following standard Chernoff-type bound.

Claim 2.2 ([6])**.** $\Pr\left[Bin(N,p) < (1-\alpha)Np\right] \leq e^{-Np\alpha^2/2}$.

The following claim is a well-known fact from Ramsey Theory.

Claim 2.3 ([16])**.** *Every tournament on 2^{k-1} vertices contains a transitive subtournament on k vertices.*

Claim 2.4. *Let $t, k \geq 1$ be integers. Then there is a collection $\mathcal{S} \subseteq [t]^k$ of size at least t^2/k^2 such that every pair of distinct k-tuples in $\mathcal{S}$ have at most one identical entry.*

Proof. We construct the collection $\mathcal{S}$ greedily: we start with an empty collection, add an arbitrary k-tuple to it, discard all k-tuples that coincide in more than one entry with the k-tuple we added and repeat. At the beginning we have all t^k of the k-tuples in $[t]^k$. At each step we discard at most $\binom{k}{2}t^{k-2}$ tuples. Therefore, at the end of the process we have a collection of size at least

$$\frac{t^k}{1+\binom{k}{2}t^{k-2}} \geq \frac{t^k}{k^2 t^{k-2}} = \frac{t^2}{k^2},$$

as required. ■

Proof of Lemma 2.1. For every $i = 1, ..., k$ put $h_i = |H_i|$. Fix an integer $m > m_0(h)$, where $m_0(h)$ will be chosen later. For convenience of presentation, we assume that m is divisible by $2h$ and by $2h_i$ for every i. Let $V_1, ..., V_k$ be pairwise-disjoint vertex sets of size m each. The edges between the sets $V_1, ..., V_k$ are oriented as follows: for every $(i,j) \in E(D)$ we direct all edges from V_i to V_j. For every $1 \leq i < j \leq k$ for which $(i,j), (j,i) \notin E(D)$, orient the edges between V_i and V_j randomly and independently with probability $1/2$. We will show that with positive probability, the resulting k-partite tournament, F, satisfies the assertion of Item 3 in the lemma, thus finishing the proof.
An *H-partition* is a tuple $(\mathcal{P}_{i,j}, \mathcal{T}_{i,j})_{i,j}$, where $1 \leq i \leq k$ and $1 \leq j \leq \frac{m}{2h_i}$, with the following properties.

- For each $1 \leq i \leq k$, $\mathcal{P}_{i,1}, ..., \mathcal{P}_{i,\frac{m}{2h_i}}$ are pairwise-disjoint subsets of V_i, each of size $h_i = |H_i|$.
- For each $1 \leq i \leq k$ and $1 \leq j \leq \frac{m}{2h_i}$, $T_{i,j}$ is a labeled transitive tournament on the set $\mathcal{P}_{i,j}$.

Note that $\bigcup_{j=1}^{m/2h_i} \mathcal{P}_{i,j}$ is a subset of V_i of size exactly $\frac{m}{2}$. The number of ways to choose an H-partition is exactly

$$\prod_{i=1}^{k} \frac{m!}{(m/2)!} \leq m^{km}. \tag{1}$$

By Claim 2.4 with parameter $t = \frac{m}{2h}$, there is a collection $\mathcal{S} \subseteq \left[\frac{m}{2h}\right]^k \subseteq \left[\frac{m}{2h_1}\right] \times \cdots \times \left[\frac{m}{2h_k}\right]$ such that $|\mathcal{S}| \geq \left(\frac{m}{2hk}\right)^2 \geq \frac{m^2}{4h^4}$, and

$$\forall s = (s_1, ..., s_k), s' = (s'_1, ..., s'_k) \in \mathcal{S}, \quad \#\left\{1 \leq i \leq k : s_i = s'_i\right\} \leq 1. \tag{2}$$

For each $i = 1, \ldots, k$ we fix a linear ordering of the vertices of H_i in which all edges point forward, that is, if $u, v \in H_i$ and $u \to v$ then u precedes v in the ordering. Such an ordering exists since H_i is acyclic. Fix an H-partition $\mathcal{Q} = (\mathcal{P}_{i,j}, \mathcal{T}_{i,j})_{i,j}$ and let $s = (s_1, ..., s_k) \in \mathcal{S}$. Since T_{i,s_i} is transitive and H_i is acyclic, H_i can be embedded into T_{i,s_i}. In what follows, when we say that T_{i,s_i} plays the role of H_i we mean that H_i is embedded in T_{i,s_i} in an order-preserving way with respect to our fixed linear ordering of H_i and the unique linear ordering of T_{i,s_i}. Let $A_{\mathcal{Q}}(s)$ be the event that $T_{1,s_1} \cup \cdots \cup T_{k,s_k}$, together with the edges of F connecting the sets $\mathcal{P}_{1,s_1}, \ldots, \mathcal{P}_{k,s_k}$, contains a copy of H with T_{i,s_i} playing the role of H_i. Then $\mathbb{P}\left[A_{\mathcal{Q}}(s)\right] \geq 2^{-\sum h_i h_j} \geq 2^{-h^2}$. Observe that by (2), the events $\{A_{\mathcal{Q}}(s) : s \in \mathcal{S}\}$ are independent. Since $|\mathcal{S}| \geq \frac{m^2}{4h^4}$, the random variable

$$X_{\mathcal{Q}} = \sum_{s \in \mathcal{S}} \mathbb{1}_{A_{\mathcal{Q}}(s)}$$

stochastically dominates a random variable with distribution $\mathrm{Bin}\left(\frac{m^2}{4h^4}, 2^{-h^2}\right)$. By Claim 2.2 with parameter $\alpha = \frac{1}{2}$ we have:

$$\mathbb{P}\left[X_{\mathcal{Q}} < \frac{2^{-h^2} m^2}{8h^4}\right] \leq \mathbb{P}\left[\mathrm{Bin}\left(\frac{m^2}{4h^4}, 2^{-h^2}\right) < \frac{2^{-h^2} m^2}{8h^4}\right] \leq \exp\left\{-\frac{m^2}{32h^4 2^{h^2}}\right\} < m^{-hm} \leq m^{-km}.$$

The strict inequality above holds if m is large enough. We choose $m_0(h)$ to be large enough so that this inequality holds for every $m \geq m_0(h)$. Set $\gamma = \gamma(h) = \frac{2^{-h^2}}{8h^4}$. By (1), there are at most m^{km} ways to choose an H-partition $\mathcal{Q}$. By the union bound over all H-partitions we get that the following event has positive probability: for every H-partition $\mathcal{Q}$, the number of $s \in \mathcal{S}$ for which $A_{\mathcal{Q}}(s)$ happened is at least γm^2. We now show that if this event happens then F satisfies the assertion of Item 3 in the lemma.

Let T be a completion of F. For every $1 \leq i \leq k$, we use Claim 2.3 to extract from V_i a collection $\mathcal{P}_{i,1}, ..., \mathcal{P}_{i,\frac{m}{2h_i}}$ of pairwise-disjoint sets, each of size h_i, such that $T[\mathcal{P}_{i,j}]$ is transitive for every $1 \leq j \leq \frac{m}{2h_i}$. We extract these sets one by one and stop when there are $\frac{m}{2}$ remaining vertices. By Claim 2.3, we can do this as long as there are at least $2^{h_i - 1}$ remaining vertices. By choosing $m_0(h)$ to be large enough we guarantee that $\frac{m}{2} \geq 2^{h_i - 1}$.

For every $1 \leq i \leq k$ and $1 \leq j \leq \frac{m}{2h_i}$, set $T_{i,j} = T[\mathcal{P}_{i,j}]$. Consider this H-partition $\mathcal{Q} = (\mathcal{P}_{i,j}, \mathcal{T}_{i,j})_{i,j}$. By our assumption, the event $A_{\mathcal{Q}}(s)$ happened for at least γm^2 of the elements $s \in \mathcal{S}$. By the definition of the event $A_{\mathcal{Q}}(s)$, if this event happened then the vertex-set $\mathcal{P}_{1,s_1} \cup \cdots \cup \mathcal{P}_{k,s_k}$ contains a copy of H (in the tournament T) with $\mathcal{T}_{i,s_i}$ playing the role of H_i. By (2), every pair of these copies can share vertices in no more than one of the clusters $V_1, \ldots, V_k$. Therefore, every edge $e \in E(F)$ (that is, an edge that connects vertices in two distinct clusters V_i, V_j) is contained in at most one of these copies. Thus, Item 3 in the lemma holds, completing the proof. ■

3 Easy Tournaments

In this section we prove the first part of Theorem 1.1. For convenience, we restate as follows.

Theorem 3.1. *For every h there are $\varepsilon_0 = \varepsilon_0(h) > 0$ and $d = d(h)$ with the following property. For every 2-colorable oriented graph H on h vertices and for every positive $\varepsilon < \varepsilon_0$, if a tournament T on $n \geq n_0(\varepsilon)$ vertices is ε-far from being H-free then T contains at least $\varepsilon^d n^h$ copies of H.*

Throughout this section, we implicitly assume that n is large enough. To make the presentation cleaner, we also implicitly assume that n is divisible by various quantities which depend on the other parameters, H and ε. It is easy to see that in order to establish Theorem 3.1, it is enough to prove it for values of n which satisfy such divisibility conditions.

We start by introducing some definitions and lemmas that we use in the proof of Theorem 3.1. Let T be a tournament on $[n]$. The *adjacency matrix* of T, denoted $A = A(T)$, is the $n \times n$ matrix in which, for every $1 \leq i \neq j \leq n$, $A_{i,j} = 1$ if $(i,j) \in E(T)$ and $A_{i,j} = 0$ if $(j,i) \in E(T)$. The main diagonal of A is set to be 0. For a pair of disjoint sets $X, Y \subseteq V(T)$ define

$$e(X,Y) = |\{(x,y) \in X \times Y : (x,y) \in E(T)\}|$$

and $d(X,Y) = \frac{e(X,Y)}{|X||Y|}$. Note that $d(X,Y) + d(Y,X) = 1$, as T is a tournament. We have $X \to Y$ if and only if $d(X,Y) = 1$, and $Y \to X$ if and only if $d(X,Y) = 0$. For a constant $\delta < \frac{1}{2}$, we say that (X,Y) is δ-homogeneous if either $d(X,Y) \geq 1-\delta$ or $d(X,Y) \leq \delta$. We say that the *dominant direction* of (X,Y) is $X \to Y$ if $d(X,Y) \geq \frac{1}{2}$ and is $Y \to X$ if $d(X,Y) < \frac{1}{2}$. The *weight* of the pair (X,Y) is $\frac{|X||Y|}{n^2}$. Let $\mathcal{P} = \{V_1, ..., V_r\}$ be a vertex-partition of T, namely suppose that $V(T) = V_1 \uplus \cdots \uplus V_r$. We say that $\mathcal{P}$ is *δ-homogeneous* if the total weight of non-δ-homogeneous pairs (V_i, V_j), $1 \leq i \neq j \leq r$, is at most δ. We say that $\mathcal{P}$ is an *equipartition* if $\big||V_i| - |V_j|\big| \leq 1$ for every $1 \leq i, j \leq r$.

Recall the definition of a bipartite tournament from Section 2. A copy of a bipartite tournament $F = (M \cup N, E)$ in a tournament T is an injection $f : V(F) \to V(T)$ such that for every $x \in M, y \in N$ it holds that $(x,y) \in E(F)$ if and only if $(f(x), f(y)) \in E(T)$. The first ingredient in the proof of Theorem 3.1 is the following lemma.

Lemma 3.2. *There is $C > 0$ such that the following holds for every integer $k \geq 1$ and every $\delta \in (0, 1/2)$. Let $F = (M \cup N, E)$ be a bipartite tournament with $|M| = |N| = k$. Then every tournament T on $n \geq n_0(k, \delta)$ vertices either contains at least $(\delta/k)^{Ck^3} n^{2k}$ copies of F or satisfies the following: there is an equipartition $\mathcal{Q} = \{Q_1, ..., Q_q\}$ of $V(T)$, where $q \geq \frac{1}{\delta}$, and there are subsets $W_i \subseteq Q_i$, such that the following hold.*

1. *For all but at most δq^2 of the pairs $1 \leq i < j \leq q$, it holds that (Q_i, Q_j) is δ-homogeneous and the dominant direction of (W_i, W_j) is the same as that of (Q_i, Q_j).*
2. *(W_i, W_j) is δ-homogeneous for every $1 \leq i < j \leq q$.*
3. *$|W_i| \geq (\delta/k)^{Ck^2} n$ for every $1 \leq i \leq q$.*

Throughout this section, C denotes the constant from Lemma 3.2. Another ingredient in the proof of Theorem 3.1 is the following simple counting lemma.

Lemma 3.3. *For every h there are $\eta = \eta(h)$ and $\alpha = \alpha(h) > 0$ such that the following holds for every oriented graph H on h vertices. Let $H_1, ..., H_\ell$ be a partition of H such that $H_1, ..., H_\ell$ induce acyclic digraphs, and for every $1 \leq i < j \leq \ell$, either $H_i \to H_j$ or $H_j \to H_i$. Let $W_1, ..., W_\ell$ be pairwise-disjoint vertex sets in a tournament T having the following properties:*

1. $|W_i| \geq 2^{h-1}$ *for every* $1 \leq i \leq \ell$.

2. *For every* $1 \leq i \neq j \leq \ell$, *if* $H_i \to H_j$ *then* $d(W_i, W_j) \geq 1 - \eta$.

Then T *contains at least* $\alpha \cdot \prod_{i=1}^{\ell} |W_i|^{h_i}$ *copies of* H, *where* $h_i = |H_i|$.

Recall the definition of a completion of a bipartite tournament from Section 2. We say that a bipartite tournament F *forces* an oriented graph H if every completion of F contains a copy of H. The following lemma is the last ingredient we need for the proof of Theorem 3.1.

Lemma 3.4. *For every 2-colorable oriented graph* H *there is a bipartite tournament that forces* H.

Proof of Theorem 3.1. Let H be a 2-colorable oriented graph on h vertices. Apply Lemma 3.4 to get a bipartite tournament $F = (M \cup N, E)$ that forces H. Note that we can clearly assume that $|M| = |N| \geq h$ (by adding additional vertices if necessary). Put $k := |M| = |N|$.
We will prove the theorem with

$$\varepsilon_0 = \varepsilon_0(h) = \min\left(\frac{1}{3k}, 3\eta(h)\right)$$

and

$$d = d(h) = 2Ck^3 + \alpha^{-1},$$

where $\eta(h)$ and $\alpha = \alpha(h)$ are from Lemma 3.3, and C is the constant from Lemma 3.2.

Let $\varepsilon < \varepsilon_0$, and let T be any tournament on n vertices which is ε-far from being H-free. Assume first that T contains at least $(\varepsilon/3k)^{Ck^3} n^{2k}$ copies of F. Since F forces H, every copy of F (in a tournament) contains a copy of H. Every copy of H is contained in at most n^{2k-h} copies of F. Recalling that $\varepsilon < \frac{1}{3k}$, we conclude that T contains at least

$$n^{-(2k-h)} (\varepsilon/3k)^{Ck^3} n^{2k} = (\varepsilon/3k)^{Ck^3} n^h \geq \varepsilon^{2Ck^3} n^h \geq \varepsilon^d n^h$$

copies of H, giving the desired result in this case.

Suppose from now on that T contains less than $(\varepsilon/3k)^{Ck^3} n^{2k}$ copies of F. We apply Lemma 3.2 to T with approximation parameter $\frac{\varepsilon}{3}$ to get an equipartition $\mathcal{Q} = \{Q_1, ..., Q_q\}$ and subsets $W_i \subseteq Q_i$ with the properties stated in the lemma. Define $\mathcal{N}$ to be the set of pairs $1 \leq i < j \leq q$ for which either (a) (Q_i, Q_j) is not $\frac{\varepsilon}{3}$-homogeneous, or (b) the dominant direction of (W_i, W_j) is not the same as that of (Q_i, Q_j). By Lemma 3.2 we have $|\mathcal{N}| \leq \frac{\varepsilon}{3} q^2$. This implies that

$$\sum_{(i,j) \in \mathcal{N}} |Q_i||Q_j| \leq \frac{\varepsilon}{3} q^2 \left(\frac{n}{q}\right)^2 = \frac{\varepsilon}{3} n^2. \tag{3}$$

Let T' be the tournament obtained from T by making the following changes.

1. Make Q_i transitive for every $i = 1, \ldots, q$.

2. For every $1 \leq i < j \leq q$, if $d(W_i, W_j) \geq 1 - \frac{\varepsilon}{3}$ then set $Q_i \to Q_j$ and if $d(W_i, W_j) \leq \frac{\varepsilon}{3}$ then set $Q_j \to Q_i$.

By Lemma 3.2, (W_i, W_j) is $\frac{\varepsilon}{3}$-homogeneous for every $1 \le i < j \le q$, so Item 2 covers all options. The number of edge-reversals made in Item 1 is at most $q\binom{n/q}{2} < \frac{n^2}{q} \le \frac{\varepsilon}{3}n^2$. Here we use the inequality $q \ge \frac{3}{\varepsilon}$, given by Lemma 3.2. In Item 2, if $(i,j) \notin \mathcal{N}$ then the number of reversals of edges between Q_i and Q_j is at most $\frac{\varepsilon}{3}|Q_i||Q_j|$. Using these facts and (3) we get that the total number of edge-reversals made in Items 1 and 2 is less than $\frac{\varepsilon}{3}n^2 + \sum_{i<j}\frac{\varepsilon}{3}|Q_i||Q_j| + \frac{\varepsilon}{3}n^2 \le \frac{\varepsilon}{3}n^2 + \frac{\varepsilon}{3}n^2 + \frac{\varepsilon}{3}n^2 = \varepsilon n^2$.

Since T is ε-far from being H-free and since T' is obtained from T by reversing less than εn^2 edges, T' must contain a copy of H. Let $Q_{i_1}, ..., Q_{i_\ell}$ be the parts of $\mathcal{Q}$ which intersect this copy. For $j = 1, ..., \ell$ define $H_{i_j} = H \cap Q_{i_j}$ and $h_{i_j} = |H_{i_j}|$. From the way we constructed T' from T in Items 1 and 2, it follows that the sets $H_{i_1}, ..., H_{i_\ell}$ are acyclic, and that for every $1 \le s < t \le \ell$ we have either $H_{i_s} \to H_{i_t}$ or $H_{i_t} \to H_{i_s}$. Moreover, for every $1 \le s \ne t \le \ell$, if $H_{i_s} \to H_{i_t}$ then $Q_{i_s} \to Q_{i_t}$ in T', implying $d(W_{i_s}, W_{i_t}) \ge 1 - \frac{\varepsilon}{3} \ge 1 - \eta(h)$ in T (see our choice of ε_0). Finally, by Lemma 3.2 we have

$$|W_{i_j}| \ge (\varepsilon/3k)^{Ck^2} n \tag{4}$$

for every $j = 1, \ldots, \ell$. So if n is large enough then $|W_{i_j}| \ge 2^{h-1}$. We conclude that $W_{i_1}, \ldots, W_{i_\ell}$ satisfy the conditions of Lemma 3.3 *in the tournament* T with respect to the partition $V(H) = H_{i_1} \cup \cdots \cup H_{i_\ell}$. By applying Lemma 3.3 and using the inequalities (4) and $\varepsilon < \frac{1}{3k}$, we get that T contains at least

$$\alpha \cdot \prod_{j=1}^{\ell} |W_{i_j}|^{h_{i_j}} \ge \alpha\,(\varepsilon/3k)^{Chk^2} n^h \ge \varepsilon^{1/\alpha}\,(\varepsilon/3k)^{Ck^3} n^h \ge \varepsilon^{2Ck^3+1/\alpha} n^h = \varepsilon^d n^h$$

copies of H. This completes the proof of the theorem. ■

Having proven Theorem 3.1, we proceed to prove Lemmas 3.2, 3.3 and 3.4.

Proof of Lemma 3.4. This lemma follows easily from Lemma 2.1. Let $V(H) = H_1 \cup H_2$ be a proper 2-coloring of H. Apply Lemma 2.1 with parameter $h = |V(H)|$ and with D being the empty graph on 2 vertices. Lemma 2.1 implies that there is a bipartite tournament $F = (V_1 \cup V_2, E(F))$, where $|V_1| = |V_2| = m := m_0(h)$, such that every completion of F contains at least $\gamma(h)m^2$ (and in particular at least one) copies of H. ■

Proof of Lemma 3.3. Set $m = 2^{h-1}$. For each $i = 1, ..., \ell$ we choose a subset $X_i \subseteq W_i$ of size m uniformly at random. For $1 \le i < j \le \ell$, let us say that (X_i, X_j) *agrees* with (H_i, H_j) if $X_i \to X_j$ whenever $H_i \to H_j$ and $X_j \to X_i$ whenever $H_j \to H_i$. By the assumption on the pairs (W_i, W_j), the probability that (X_i, X_j) does not agree with (H_i, H_j) is at most ηm^2. By the union bound, the probability that there is a pair $1 \le i < j \le \ell$ for which (X_i, X_j) does not agree with (H_i, H_j) is at most $\eta m^2\binom{\ell}{2} \le \eta m^2 h^2$. By setting $\eta(h) = \frac{1}{2}(hm)^{-2} = \frac{1}{2}h^{-2}2^{-2(h-1)}$ we get that this probability is at most $\frac{1}{2}$.

By Claim 2.3 and the choice of m we get that X_i contains a transitive subset Y_i of size $h_i = |H_i|$. Therefore, if (X_i, X_j) agrees with (H_i, H_j) for every $1 \le i < j \le \ell$ then $X = \bigcup_{i=1}^{\ell} X_i$ contains a copy of H with Y_i playing the role of H_i. Every such copy of H is contained in at most $\prod_{i=1}^{\ell}\binom{|W_i|-h_i}{m-h_i}$ such sets X. Therefore there are at least

$$\frac{\frac{1}{2}\prod_{i=1}^{\ell}\binom{|W_i|}{m}}{\prod_{i=1}^{\ell}\binom{|W_i|-h_i}{m-h_i}} = \frac{\frac{1}{2}\prod_{i=1}^{\ell}\binom{|W_i|}{h_i}}{\prod_{i=1}^{\ell}\binom{m}{h_i}} \ge \frac{1}{2}\cdot\prod_{i=1}^{\ell}\left(\frac{|W_i|}{m}\right)^{h_i} = \frac{1}{2}\cdot m^{-h}\prod_{i=1}^{\ell}|W_i|^{h_i}$$

copies of H. We choose the constant $\alpha = \alpha(h)$ to be $\alpha = \frac{1}{2}m^{-h} = 2^{-h(h-1)-1}$. ■

We now turn to the proof of Lemma 3.2. This lemma is proved using a "conditional" regularity lemma for binary matrices, proved by Alon, Fischer and Newman in [3]. Let A be an $n \times n$ matrix with 0/1 entries whose rows and columns are indexed by $1, ..., n$. For a pair of sets $R, C \subseteq [n]$, we denote by $R \times C$ the submatrix of A with rows from R and columns from C and we call it a *block*. The *dominant value* of a block is the value, 0 or 1, that appears in at least half of the entries. For a constant $\delta < \frac{1}{2}$, we say that a block is δ-homogeneous if its dominant value appears in at least a $(1-\delta)$-fraction of the entries. Clearly, if A is the adjacency matrix of a tournament T and $X, Y \subseteq V(T)$ are disjoint vertex sets, then (X, Y) is δ-homogeneous (in the tournament sense) if and only if the block $X \times Y$ is δ-homogeneous (in the matrix sense). Moreover, the dominant direction of (X, Y) is $X \to Y$ if and only if the dominant value of $X \times Y$ is 1.

The weight of a block $R \times C$ is defined as $\frac{|R||C|}{n^2}$. Let $\mathcal{R} = \{R_1, ..., R_s\}$ and $\mathcal{C} = \{C_1, ..., C_t\}$ be partitions of $[n]$. We say that $(\mathcal{R}, \mathcal{C})$ is a *δ-homogeneous* partition of A if the total weight of non-δ-homogeneous blocks $R_i \times C_j$ is at most δ. Note that if A is the adjacency matrix of a tournament T on $V(T) = [n]$, and if $\mathcal{P}$ is a partition of $[n]$ such that $(\mathcal{P}, \mathcal{P})$ is a δ-homogeneous partition of A, then $\mathcal{P}$ is a δ-homogeneous partition of T (as defined in the beginning of Section 3)[3].

Let B be a 0/1-valued $k \times k$ matrix. A *copy* of B in A is a sequence of rows $r_1 < r_2 < \cdots < r_k$ and a sequence of columns $c_1 < c_2 < \cdots < c_k$ such that $A_{r_i, c_j} = B_{i,j}$ for every $1 \leq i, j \leq k$. We are now ready to state the Alon-Fischer-Newman Regularity Lemma.

Lemma 3.5 (Alon-Fischer-Newman [3]). *There is $c > 0$ such that the following holds for every integer $k \geq 1$ and every $\delta > 0$. For every 0/1 matrix A of size $n \times n$ with $n > (k/\delta)^{ck}$, either A has a δ-homogeneous partition $(\mathcal{R}, \mathcal{C})$ with $|\mathcal{R}|, |\mathcal{C}| \leq (k/\delta)^{ck}$, or for every 0/1-valued $k \times k$ matrix B, A contains at least $(\delta/k)^{ck^2} n^{2k}$ copies of B.*

Throughout this section c denotes the constant from Lemma 3.5. Without loss of generality, we always assume that $c \geq 1$. The following lemma is an application of Lemma 3.5 to adjacency matrices of tournaments.

Lemma 3.6. *Let $F = (M \cup N, E)$ be a bipartite tournament with $|M| = |N| = k$, and let $\delta \in (0, 1/2)$. Let T be a tournament on $n \geq n_0(k, \delta)$ vertices and let $\mathcal{P}$ be an equipartition of $V(T)$. Then either T contains at least $(\delta/3k)^{2ck^2} n^{2k}$ copies of F, or T admits a δ-homogeneous equipartition that refines $\mathcal{P}$, and has at least δ^{-1} and at most $|\mathcal{P}| \cdot (3k/\delta)^{5ck}$ parts.*

As the proof of Lemma 3.6 is rather technical, we leave it to the end of this section and first show how to deduce Lemma 3.2 by two applications of Lemma 3.6.

Proof of Lemma 3.2. Let C be large enough so that

$$(\delta/k)^C \leq (\delta/15k)^{110c^2},$$

where $c \geq 1$ is the constant from Lemma 3.5. Note that C does not depend on δ or k, as δ is assumed to be less than $\frac{1}{2}$.

We assume that T contains less than $(\delta/k)^{Ck^3} n^{2k}$ copies of F and prove that the other alternative in the statement of the lemma holds. Our choice of C implies that T contains less than $(\delta/15k)^{2ck^2} n^{2k}$ copies of F. By applying Lemma 3.6 with approximation parameter $\frac{\delta}{5}$ and $\mathcal{P} = \{V(T)\}$, we get that T admits a $\frac{\delta}{5}$-homogeneous equipartition $\mathcal{Q} = \{Q_1, ..., Q_q\}$ with

$$\delta^{-1} \leq q \leq (15k/\delta)^{5ck}. \tag{5}$$

[3]The converse is not necessarily true. The fact that $\mathcal{P}$ is a δ-homogeneous partition of T does not take into account "diagonal" blocks, i.e. blocks of the form $V_i \times V_i$, $V_i \in \mathcal{P}$.

Set $\gamma = \frac{1}{2q^4}$, and note that $\gamma \geq \frac{1}{2}(\delta/15k)^{20ck}$, and hence

$$(\gamma/3k)^{2ck^2} \geq \left((1/6k)\cdot(\delta/15k)^{20ck}\right)^{2ck^2} \geq \left((\delta/15k)^{21ck}\right)^{2ck^2} \geq (\delta/15k)^{42c^2k^3}.$$

Our assumption in the beginning of the proof and our choice of C imply that T contains less than $(\gamma/3k)^{2ck^2} n^{2k}$ copies of F. Apply Lemma 3.6 to T again, now with approximation parameter γ and $\mathcal{P} = \mathcal{Q}$, to obtain a γ-homogeneous equipartition $\mathcal{W}$ which refines $\mathcal{Q}$ and satisfies

$$|\mathcal{W}| \leq |\mathcal{Q}|\cdot(3k/\gamma)^{5ck} \leq (15k/\delta)^{5ck}\cdot\left(6k\cdot(15k/\delta)^{20ck}\right)^{5ck} \leq (15k/\delta)^{110c^2k^2}. \tag{6}$$

For each $1 \leq i \leq q$ define $\mathcal{W}_i = \{W \in \mathcal{W} : W \subseteq Q_i\}$. Sample a vertex $w_i \in Q_i$ uniformly at random and let $W_i \in \mathcal{W}_i$ be such that $w_i \in W_i$. By (6) and our choice of C, we have $|\mathcal{W}| \leq (k/\delta)^{Ck^2}$, which implies that $|W_i| \geq (\delta/k)^{Ck^2} n$ for every $1 \leq i \leq q$, as required. To complete the proof, we show that with positive probability, $W_1, ..., W_q$ satisfy the assertions of Items 1 and 2 of the lemma.

Let A_1 be the event that (W_i, W_j) is δ-homogeneous for every $1 \leq i < j \leq q$. Fixing $1 \leq i < j \leq q$, the probability that (W_i, W_j) is not δ-homogeneous is $\sum \frac{|W||W'|}{|Q_i||Q_j|} = \left(\frac{q}{n}\right)^2 \sum |W||W'|$, where the sum is over all non-δ-homogeneous pairs $(W, W') \in \mathcal{W}_i \times \mathcal{W}_j$. This sum is not larger than $\gamma q^2 = \frac{1}{2q^2}$ because $\mathcal{W}$ is γ-homogeneous and by our choice of γ. By the union bound over all pairs $1 \leq i < j \leq q$, we get that $\mathbb{P}[A_1] \geq \frac{1}{2}$.

Let $1 \leq i < j \leq q$ be such that (Q_i, Q_j) is a $\frac{\delta}{5}$-homogeneous pair. We say that (Q_i, Q_j) is *bad* if $d(Q_i, Q_j) \geq 1 - \frac{\delta}{5}$ but $d(W_i, W_j) \leq \delta$, or $d(Q_i, Q_j) \leq \frac{\delta}{5}$ but $d(W_i, W_j) \geq 1 - \delta$. Otherwise (Q_i, Q_j) is *good*. Assume without loss of generality that $d(Q_i, Q_j) \geq 1 - \frac{\delta}{5}$. Then the probability that $d(W_i, W_j) \leq \delta$ is at most $\frac{\delta/5}{1-\delta} < \frac{2\delta}{5}$ (here we use $\delta < \frac{1}{2}$). We conclude that the probability that a given pair (Q_i, Q_j) is bad is less than $\frac{2\delta}{5}$. Let Z be the number of bad pairs (Q_i, Q_j). Let A_2 be the event that $Z \leq \frac{4\delta}{5}q^2$. We have $\mathbb{E}[Z] < \frac{2\delta}{5}q^2$. By Markov's inequality, we have $\mathbb{P}[Z > \frac{4\delta}{5}q^2] < \frac{1}{2}$, implying that $\mathbb{P}[A_2] > \frac{1}{2}$.

So far we showed that with positive probability, both A_1 and A_2 happen. We now show that if A_1 and A_2 happen then Items 1 and 2 in the lemma hold. Item 2 holds because A_1 happened. For Item 1, notice that if (Q_i, Q_j) is $\frac{\delta}{5}$-homogeneous and good, and if (W_i, W_j) is δ-homogeneous, then (W_i, W_j) has the same dominant direction as (Q_i, Q_j). Thus, if a pair $1 \leq i < j \leq q$ violates Item 1, then either (Q_i, Q_j) is not $\frac{\delta}{5}$-homogeneous or (Q_i, Q_j) is bad. Since $\mathcal{Q}$ is a $\frac{\delta}{5}$-homogeneous equipartition, the number of non-$\frac{\delta}{5}$-homogeneous pairs (Q_i, Q_j) is at most $\frac{\delta}{5}q^2$. Since A_2 happened, the number of bad pairs (Q_i, Q_j) is at most $\frac{4\delta}{5}q^2$. Thus, Item 1 holds for all but at most δq^2 of the pairs (Q_i, Q_j), as required. ■

Proof of Lemma 3.6. Let us assume that T contains less than $(\delta/3k)^{2ck^2} n^{2k}$ copies of F. Our goal is to show that T admits a δ-homogeneous equipartition which refines $\mathcal{P}$, and has at least δ^{-1} and at most $|\mathcal{P}|\cdot(3k/\delta)^{5ck}$ parts.

Define B to be the *bipartite adjacency matrix* of F; that is, B is a $k \times k$ matrix, indexed by $M \times N$, in which $B_{x,y} = 1$ if $(x, y) \in E(F)$ and $B_{x,y} = 0$ if $(y, x) \in E(F)$. We claim that $A = A(T)$, the adjacency matrix of T, contains less than $\left(\delta^2/3k\right)^{ck^2} m^{2k}$ copies of B. Assume otherwise. A copy of B which does not intersect the main diagonal of A corresponds to a copy of F in T. There can be no more than $O(n^{2k-1})$ copies of B which intersect the main diagonal of A. Recalling that $\delta < \frac{1}{2}$ and assuming n to be large enough, we conclude that T contains at least $\left(\delta^2/3k\right)^{ck^2} n^{2k} - O(n^{2k-1}) \geq (\delta/3k)^{2ck^2} n^{2k}$ copies of F, in contradiction to our assumption in the beginning of the proof.

Thus, $A = A(T)$ contains less than $\left(\delta^2/3k\right)^{ck^2} n^{2k}$ copies of B. By Lemma 3.5, applied with approximation parameter $\frac{\delta^2}{3}$, A has a $\frac{\delta^2}{3}$-homogeneous partition $(\mathcal{R},\mathcal{C})$ with $|\mathcal{R}|,|\mathcal{C}| \leq \left(3k/\delta^2\right)^{ck}$.

Write $\mathcal{P} = \{P_1, ..., P_p\}$. For every $i = 1, \ldots, p$, let $\mathcal{U}_i$ be the common refinement of the set P_i and the partitions $\mathcal{R}$, $\mathcal{C}$, that is $\mathcal{U}_i = \{P_i \cap R \cap C : R \in \mathcal{R}, C \in \mathcal{C}\}$. Set $q = \frac{6p|\mathcal{R}||\mathcal{C}|}{\delta}$, and note that $\delta^{-1} \leq q \leq \frac{6p}{\delta} \cdot \left(3k/\delta^2\right)^{2ck} \leq p \cdot (3k/\delta)^{5ck}$. For each $U \in \mathcal{U}_i$, partition U into parts of size $\frac{n}{q}$ and an additional part $Z_{i,U}$ of size less than $\frac{n}{q}$. Let Z_i be the union of all additional parts $Z_{i,U}$, $U \in \mathcal{U}_i$. Note that we have $|Z_i| < |\mathcal{R}| \cdot |\mathcal{C}| \cdot \frac{n}{q} \leq \frac{\delta n}{6p}$. Partition Z_i arbitrarily into parts of size $\frac{n}{q}$. Denote by $\mathcal{Q}_i$ the resulting equipartition of P_i. Then $\mathcal{Q} := \bigcup_{i=1}^{q} \mathcal{Q}_i$ is an equipartition of $V(T)$ which has q parts and refines $\mathcal{P}$.

To finish the proof, we show that $\mathcal{Q}$ is δ-homogeneous. To this end, define $\mathcal{N}$ to be the set of all non-δ-homogeneous pairs $(X, Y) \in \mathcal{Q} \times \mathcal{Q}$ with $X \neq Y$. Set $Z := \bigcup_{i=1}^{p} Z_i$ and note that $|Z| = \sum_{i=1}^{p} |Z_i| \leq \frac{\delta}{6} n$. By the definition of $\mathcal{Q}$, if $X \in \mathcal{Q}$ is not contained in Z then there are $R \in \mathcal{R}$ and $C \in \mathcal{C}$ such that $X \subseteq R \cap C$. Thus, a block $X \times Y$ for which $X, Y \not\subseteq Z$ is contained in a block $R \times C$, where $R \in \mathcal{R}$ and $C \in \mathcal{C}$. Let $\mathcal{N}_1$ be the set of pairs $(X, Y) \in \mathcal{N}$ such that either X or Y is contained in Z; let $\mathcal{N}_2$ be the set of pairs $(X, Y) \in \mathcal{N}$ such that $X, Y \not\subseteq Z$, and the block $R \times C$ containing $X \times Y$ is not $\frac{\delta^2}{3}$-homogeneous; let $\mathcal{N}_3 = \mathcal{N} \setminus (\mathcal{N}_1 \cup \mathcal{N}_2)$. Since $|Z| \leq \frac{\delta}{6} n$, we have $\sum_{(X,Y)\in\mathcal{N}_1} |X||Y| \leq 2n \cdot |Z| \leq \frac{\delta}{3} n^2$. Furthermore, as $(\mathcal{R},\mathcal{C})$ is $\frac{\delta^2}{3}$-homogeneous, we have $\sum_{(X,Y)\in\mathcal{N}_2} |X||Y| \leq \frac{\delta^2}{3} n^2$. Thus, in order to prove that $\mathcal{Q}$ is δ-homogeneous, it is enough to show that $\sum_{(X,Y)\in\mathcal{N}_3} |X||Y| \leq \frac{\delta}{3} n^2$.

By the definition of $\mathcal{N}_3$, for every $(X, Y) \in \mathcal{N}_3$ there are $R \in \mathcal{R}$ and $C \in \mathcal{C}$ such that the block $R \times C$ is $\frac{\delta^2}{3}$-homogeneous and contains the block $X \times Y$. Let $R \times C$ be a $\frac{\delta^2}{3}$-homogeneous block, and assume without loss of generality that the dominant value of $R \times C$ is 1. Let $X_1, ..., X_a$ be the parts of $\mathcal{Q}$ that are contained in R, and let $Y_1, ..., Y_b$ be the parts of $\mathcal{Q}$ that are contained in C. Define $\gamma(i, j)$ to be the fraction of pairs $(x, y) \in X_i \times Y_j$ for which $A_{x,y} = 0$. Obviously, if $\gamma(i, j) \leq \delta$ then the block $X_i \times Y_j$ is δ-homogeneous. We have

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \frac{|X_i||Y_j|}{|R||C|} \gamma(i, j) \leq \frac{\delta^2}{3},$$

because the sum on the left hand side is a lower bound for the fraction of pairs $(x, y) \in R \times C$ for which $A_{x,y} = 0$. By Markov's inequality we have

$$\sum_{(i,j):\, \gamma(i,j)>\delta} \frac{|X_i||Y_j|}{|R||C|} \leq \frac{\delta}{3}.$$

Thus, the sum of $|X||Y|$ over all pairs $(X, Y) \in \mathcal{N}_3$ for which $X \times Y \subseteq R \times C$, is at most $\frac{\delta}{3}|R||C|$. By summing over all $\frac{\delta^2}{3}$-homogeneous blocks $R \times C$ of the partition $(\mathcal{R},\mathcal{C})$, we get

$$\sum_{(X,Y)\in\mathcal{N}_3} |X||Y| \leq \sum_{R\in\mathcal{R}, C\in\mathcal{C}} \frac{\delta}{3}|R||C| \leq \frac{\delta}{3} n^2,$$

as required. ■

4 Hard Tournaments

In this section we prove the second direction of Theorem 1.1. For convenience we restate it as follows.

Theorem 4.1. *For every h there are $\varepsilon_0 = \varepsilon_0(h) > 0$ and $\alpha = \alpha(h) > 0$ with the following property. For every non-2-colorable oriented graph H on h vertices, for every positive $\varepsilon < \varepsilon_0$ and for every $n \geq n_0(\varepsilon)$ there is a tournament T on n vertices which is ε-far from being H-free but contains at most $\varepsilon^{\alpha \log(1/\varepsilon)} n^h$ copies of H.*

In this Section we prove Theorem 4.1. Throughout this section, the vertex sets of all graphs and digraphs are assumed to be subsets of $\mathbb{N}$. The reason for this assumption is that sometimes we want to have a linear ordering of the vertices. Before getting to the actual proof of Theorem 4.1, we first study some properties of homomorphisms between graphs which take into account an order of their vertex sets.

4.1 Order-preserving Homomorphisms

Definition 4.2. *(Order-Preserving Homomorphism) Let G, G' be (undirected) graphs. An* order-preserving homomorphism *from G to G' is a function $f : V(G) \to V(G')$ satisfying the following two conditions.*

1. *f is order preserving: for every $i, j \in V(G)$, if $i \leq j$ then $f(i) \leq f(j)$.*
2. *f is a graph homomorphism: for every $\{i, j\} \in E(G)$ we have $\{f(i), f(j)\} \in E(G')$.*

We write $G' \leq_{\text{hom}} G$ if there is an order-preserving homomorphism from G to G'. Notice that the relation $\leq_{\text{hom}}$ is transitive (the composition of order-preserving homomorphisms is also an order-preserving homomorphism). An *order-preserving isomorphism* is an order-preserving homomorphism which is a graph isomorphism. We write $G \cong G'$ if there is an order-preserving isomorphism between G and G'. [4]

A subgraph of any graph G is always assumed to inherit the same vertex-labeling as it had in G. The *ordered core* of G is a smallest (with respect to number of vertices) subgraph of G to which there is an order-preserving homomorphism from G. The ordered core of G is assumed to inherit the same vertex-labeling as it had in G. Notice that by definition, there is no order-preserving homomorphism from the ordered core of G to a proper induced subgraph of it. We say that a graph is an *ordered core* if it is the ordered core of itself.

Proposition 4.3. *Let G_1, G_2 be a pair of ordered cores. If $G_2 \leq_{\text{hom}} G_1$ and $G_1 \leq_{\text{hom}} G_2$ then $G_1 \cong G_2$.*

Proof. By assumption there exist order-preserving homomorphisms $f : G_1 \to G_2$ and $g : G_2 \to G_1$. Then $g \circ f$ is an order-preserving homomorphism from G_1 to itself. Since G_1 is a core, g must be surjective. The same argument shows that f is surjective. So f, g are bijections and since f, g are order-preserving we have $g = f^{-1}$. Therefore f, g are order-preserving graph isomorphisms, as required. ■

[4] Notice that two isomorphic labeled graphs may not have an order-preserving isomorphism between them. Moreover, if two graphs have an order-preserving isomorphism between them then it is unique, assuming that the vertices in each graph have different labels, which we always do in our setting.

Proposition 4.3 shows that the ordered core of a graph is unique up to order-preserving isomorphism.

Proposition 4.4. *Let G_1, G_2 be a pair of ordered cores and suppose that $G_1 \cong G_2$. Then every order-preserving homomorphism $f : G_1 \to G_2$ is an order-preserving isomorphism.*

Proof. By definition, there is an order-preserving isomorphism $g : G_2 \to G_1$. Then $f \circ g$ is an order-preserving homomorphism from G_2 to itself. By the definition of an ordered core, $f \circ g$ is a bijection. Since f, g are order-preserving we have that $f \circ g$ is the identity map and hence $f = g^{-1}$. So f is an isomorphism, as required. ■

Let H be an oriented graph. We say that an edge $(i, j) \in E(H)$ is a *forward-edge* if $i < j$ and *backward-edge* (or *backedge*) otherwise. The *backedge graph* of H is the (undirected) graph on $V(H)$ in which $\{i, j\}$ is an edge if and only if $i < j$ and $j \to i$. Note that the backedge graph depends on the labeling of the vertices of H. If we relabel the vertices of H then we may get a different backedge graph. We will need the following characterization of oriented graphs with chromatic number at most k. Although this characterization will be crucial in our proof of Theorem 4.1, it is computationally inefficient even for $k = 2$, as shown in Section 5.

Proposition 4.5. *An oriented graph H is k-colorable if and only if there is a labeling of the vertices of H for which the corresponding backedge graph is k-colorable (as a graph).*

Proof. Assume first that there is a labeling of $V(H)$ such that the corresponding backedge graph, G, has a proper (graph) k-coloring $U_1 \cup \cdots \cup U_k$. Then for every i, the set U_i is acyclic (in H) because all the edges inside it are forward-edges.

Now assume that H has a proper (acyclic) k-coloring $U_1 \cup \cdots \cup U_k$. For every $i = 1, ..., k$, label the vertices of U_i such that there are no backedges inside U_i (this is clearly possible because U_i is acyclic). Then U_i is an independent set in the backedge graph corresponding to this labeling. ■

For an oriented graph H we define a family of graphs $\mathcal{C} = \mathcal{C}(H)$, all labeled with $[h]$, as follows. We go through all $h!$ vertex-labelings of H using the labels $1, ..., h$ (where $h = v(H)$), and for each labeling we take the ordered core of the corresponding backedge graph. Let $\mathcal{C}$ be the set of all these ordered cores. Proposition 4.3 implies that $(\mathcal{C}, \leq_{\text{hom}})$ is a poset in the following sense: for every $C_1, C_2 \in \mathcal{C}$, if $C_2 \leq_{\text{hom}} C_1$ and $C_1 \leq_{\text{hom}} C_2$ then $C_1 \cong C_2$. In other words, $\leq_{\text{hom}}$ is a partial order on the set of equivalence classes of $\mathcal{C}$ under the equivalence relation $\cong$. Finally, let $K(H)$ be a maximal element of the poset $(\mathcal{C}, \leq_{\text{hom}})$, i.e. $K(H)$ is an (arbitrary) element of a maximal equivalence class. The maximality of $K(H)$ implies that for every $C \in \mathcal{C}$, if there is an order-preserving homomorphism from C to $K(H)$ (namely if $K(H) \leq_{\text{hom}} C$) then $C \cong K(H)$.

Proposition 4.6. *Let H be an oriented graph. Consider any vertex-labeling of H and let G be the corresponding backedge graph. For every order-preserving homomorphism $f : G \to K(H)$ there is a set $X \subseteq V(H)$ such that $f|_X$ is a (graph) isomorphism onto K.*

Proof. Let C be the ordered core of G. Then $f|_{V(C)}$ is an order-preserving homomorphism from C to $K(H)$. By the maximality of $K(H)$ we have $C \cong K(H)$. Then $f|_{V(C)}$ is an order-preserving isomorphism by Proposition 4.4, implying the assertion with $X = V(C)$. ■

Corollary 4.7. *Let H be a non-2-colorable oriented graph. Then the graph $K(H)$ contains a cycle $c_1 c_2 \ldots c_\ell c_1$ of length $\ell \geq 3$ with the following property. Consider any vertex-labeling of H and let G be the corresponding backedge graph. Then for every order-preserving homomorphism $f : G \to K(H)$ there are vertices $u_1 \in f^{-1}(c_1), \ldots, u_\ell \in f^{-1}(c_\ell)$ such that $u_1 u_2 \ldots u_\ell u_1$ is a cycle in G.*

Proof. By the definition of $K(H)$ there is a vertex-labeling of H such that $K(H)$ is the ordered core of the corresponding backedge graph, G_0. By Proposition 4.5, G_0 is not 2-colorable and therefore contains an odd cycle. It is easy to see that the homomorphic image of an odd cycle must contain an odd cycle. Therefore $K(H)$ contains an odd cycle, whose length is obviously at least 3. The other assertion of the corollary follows directly from Proposition 4.6. ■

4.2 Proof of Theorem 4.1

The main ingredient in the proof of Theorem 4.1 is the following construction (see [17] and [1]).

Theorem 4.8. *For every $k \geq 3$ there are $\delta_0 = \delta_0(k)$ and $c = c(k)$ such that for every $\delta < \delta_0$, for every $3 \leq \ell \leq k$ and for every sequence of distinct indices $1 \leq i_1, i_2, \ldots, i_\ell \leq k$, there is a graph $R = R(k, \delta; i_1, \ldots, i_\ell)$ with the following properties:*

1. *$V(R) = X_1 \uplus ... \uplus X_k$ and X_i is an independent set for every i.*
2. *$|V(R)| \geq \left(\frac{1}{\delta}\right)^{c \log(1/\delta)}$.*
3. *$E(R)$ is the union of at least $\delta |V(R)|^2$ pairwise edge-disjoint k-cliques, each of the form $\{x_1, \ldots, x_k\}$ with $x_i \in X_i$.*
4. *R contains at most $|V(R)|^2$ cycles $x_{i_1} x_{i_2} ... x_{i_\ell} x_{i_1}$ with $x_{i_j} \in X_{i_j}$ for $j = 1, \ldots, \ell$.*

The proof of Theorem 4.8 uses (simple variants of) Behrend's construction of a large set of integers without a 3-term arithmetic progression (see [7]). As it is similar to related constructions proved in previous papers (see, e.g., [1]) it is omitted.
We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. Let H be a non-2-colorable oriented graph on h vertices. Consider the graph $K = K(H)$ defined in Subsection 4.1. Put $k = v(K)$ and write $V(K) = \{a_1, ..., a_k\}$, where $1 \leq a_1 < a_2 < \cdots < a_k \leq h$.[5] By Corollary 4.7, K contains a cycle $(a_{i_1} a_{i_2} ... a_{i_\ell} a_{i_1})$ of length $\ell \geq 3$. Let $m_0 = m_0(h)$ and $\gamma = \gamma(h)$ be from Lemma 2.1. Set $\varepsilon_0 = \varepsilon_0(h)$ to be small enough so that every $\varepsilon < \varepsilon_0$ will satisfy the inequalities

$$(\varepsilon \gamma^{-1})^{\log(\gamma/\varepsilon)} \leq \varepsilon^{0.5 \log(1/\varepsilon)}, \quad \varepsilon < \gamma \delta_0(k), \tag{7}$$

where $\delta_0(k)$ is from Theorem 4.8. Let $\varepsilon < \varepsilon_0$. Let $R = R(k, \delta; i_1, \ldots, i_\ell)$ be the graph obtained by applying Theorem 4.8 with parameters k and

$$\delta = \varepsilon \gamma^{-1},$$

and with $i_1, \ldots, i_\ell$ being the indices of the cycle in K as above. Our choice of ε_0 guarantees that we can apply Theorem 4.8 with the above δ. Put $r = |V(R)|$ and let n be an integer which we assume, for simplicity of presentation, to be divisible by r. We will also assume that n is large enough where needed.

By the definition of K, there is a vertex-labeling of H such that K is the ordered core of the corresponding backedge graph, G_0. Hence there is an order-preserving homomorphism $g : G_0 \to K$. Denote $H_i = g^{-1}(a_i)$ for $i = 1, ..., k$. We claim that $H_1, \ldots, H_k$ have the following two properties.

[5] Recall that K inherits its vertex-labeling from the backedge graph of H whose ordered core is K and whose vertex-labels are $1, \ldots, h$.

(a) H_i is acyclic for every $i = 1, ..., k$.

(b) For every $1 \leq i < j \leq k$, if $\{a_i, a_j\} \notin E(K)$ then $H_i \to H_j$.

Property (a) follows from the definition of a backedge graph and the fact that g is a graph homomorphism. For property (b) we also need to use the fact that g is order-preserving. Define an oriented graph D on $[k]$ as follows. For every $1 \leq i < j \leq k$, if $\{a_i, a_j\} \notin E(K)$ then $(i, j) \in D$ (that is, there is a directed edge from i to j) and otherwise $(i, j), (j, i) \notin E(D)$. Then for every $(i, j) \in E(D)$ we have $H_i \to H_j$. So H satisfies the conditions of Lemma 2.1 with respect to the k-coloring $V(H) = H_1 \cup \cdots \cup H_k$ and the oriented graph D. Apply Lemma 2.1 to get a k-partite tournament $F = (V_1 \cup \cdots \cup V_k, E(F))$ such that $|V_i| = \frac{n}{r}$ (here we assume that n is large enough so that $\frac{n}{r} \geq m_0(h)$), and

$$\forall 1 \leq i < j \leq k, \quad \{a_i, a_j\} \notin E(K) \Longrightarrow V_i \to V_j. \tag{8}$$

Let $\mathcal{K}$ be the collection of k-cliques given by Item 3 in Theorem 4.8. Note that $|\mathcal{K}| \geq \delta r^2$; that the k-cliques in $\mathcal{K}$ are pairwise edge-disjoint and of the form $\{x_1, \ldots, x_k\}$ with $x_i \in X_i$; and that every edge in R is contained in (exactly) one of these k-cliques. These properties will be important in what follows. We define a tournament T on an $\frac{n}{r}$-blowup of $V(R)$, that is, each vertex $x \in V(R)$ is replaced by a vertex-set $B(x)$ of size $\frac{n}{r}$. Put $B(X_i) = \bigcup_{x \in X_i} B(x)$. The edges of T are oriented as follows.

1. $B(X_i)$ is transitive for every $i = 1, ..., k$.
2. For every $1 \leq i < j \leq k$ and for every $x_i \in X_i, x_j \in X_j$, if $\{x_i, x_j\} \notin E(R)$ then set $B(x_i) \to B(x_j)$.
3. For every $\{x_1, ..., x_k\} \in \mathcal{K}$, put a copy of F on $B(x_1) \cup \cdots \cup B(x_k)$ with $B(x_i)$ playing the role of V_i for every $i = 1, ..., k$.

Since every edge of R is contained in one of the cliques in $\mathcal{K}$, Items 2 and 3 together define the orientation of edges between $B(y)$ and $B(z)$ for every pair of vertices $y, z \in V(R)$ which belong to different clusters $X_1, \ldots, X_k$. Therefore, Items 1-3 indeed define a tournament. There is no contradiction in Item 3 because the cliques in $\mathcal{K}$ are pairwise edge-disjoint.

We now show that T satisfies our requirements, that is, T is ε-far from being H-free yet contains at most $\varepsilon^{\alpha \log(1/\varepsilon)} n^h$ copies of H (for a constant $\alpha = \alpha(h)$ that we choose later). We start with the following two observations that play a central role in the proof. First, notice that by Item 2 and by the combination of Item 3 and (8) we have the following:

$$\forall 1 \leq i < j \leq k, \quad \{a_i, a_j\} \notin E(K) \Longrightarrow B(X_i) \to B(X_j). \tag{9}$$

Secondly, let $\mathcal{C}$ be the set of all ℓ-tuples $(v_{i_1}, \ldots, v_{i_\ell}) \in B(X_{i_1}) \times \cdots \times B(X_{i_\ell})$ such that for every $j = 1, \ldots, \ell$, if $i_j < i_{j+1}$ then $v_{i_{j+1}} \to v_{i_j}$ and if $i_j > i_{j+1}$ then $v_{i_j} \to v_{i_{j+1}}$, with indices taken modulo ℓ. We will need the inequality

$$|\mathcal{C}| \leq \frac{n^\ell}{r}, \tag{10}$$

which we now prove. Given $(v_{i_1}, \ldots, v_{i_\ell}) \in \mathcal{C}$, let $x_{i_j} \in X_{i_j}$ be such that $v_{i_j} \in B(x_{i_j})$. We claim that $x_{i_1} x_{i_2} \ldots x_{i_\ell} x_{i_1}$ is a cycle in R. Let $1 \leq j \leq \ell$ and let us first handle the case that $i_j < i_{j+1}$. By the definition of $\mathcal{C}$ we have $v_{i_{j+1}} \to v_{i_j}$. By Item 2 in the construction of T above, if $\{x_{i_j}, x_{i_{j+1}}\} \notin E(R)$ then $B(x_{i_j}) \to B(x_{i_{j+1}})$, in contradiction to $v_{i_{j+1}} \to v_{i_j}$. Therefore $\{x_{i_j}, x_{i_{j+1}}\} \in E(R)$ in this case.

Similarly, if $i_j > i_{j+1}$ then by the definition of $\mathcal{C}$ we have $v_{i_j} \to v_{i_{j+1}}$. By Item 2 in the construction of T, if $\{x_{i_j}, x_{i_{j+1}}\} \notin E(R)$ then $B(x_{i_{j+1}}) \to B(x_{i_j})$, in contradiction to $v_{i_j} \to v_{i_{j+1}}$. Therefore $\{x_{i_j}, x_{i_{j+1}}\} \in E(R)$ in this case as well, proving our assertion that $x_{i_1}x_{i_2}\dots x_{i_\ell}x_{i_1}$ is a cycle in R. By Item 4 in Theorem 4.8, there are at most r^2 such cycles in R. Since T is an $\frac{n}{r}$-blowup of R and $\ell \geq 3$, we get that $|\mathcal{C}| \leq r^2 \left(\frac{n}{r}\right)^\ell \leq \frac{n^\ell}{r}$, establishing (10).

Let us prove that T contains at most $\varepsilon^{\alpha \log(1/\varepsilon)} n^h$ copies of H, where $\alpha = \alpha(h) = 0.5c$ and $c = c(k)$ is from Theorem 4.8. We will show that every copy of H in T contains vertices $v_{i_1}, \dots, v_{i_\ell}$ such that $(v_{i_1}, \dots, v_{i_\ell}) \in \mathcal{C}$ (recall the definition of $\mathcal{C}$ above). This will imply that T contains at most $|\mathcal{C}| \cdot n^{h-\ell}$ copies of H. By (10) we have $|\mathcal{C}| \leq \frac{n^\ell}{r}$, and by Item 2 in Theorem 4.8 we have $r \geq (1/\delta)^{c \log(1/\delta)}$. By using our choice of δ and the first inequality in (7), we will conclude that T contains at most

$$\frac{n^\ell}{r} \cdot n^{h-\ell} = \frac{n^h}{r} \leq \delta^{c \log(1/\delta)} n^h = \left(\varepsilon \gamma^{-1}\right)^{c \log(\gamma/\varepsilon)} n^h \leq \varepsilon^{0.5c \log(1/\varepsilon)} n^h = \varepsilon^{\alpha \log(1/\varepsilon)} n^h$$

copies of H, as required. Hence, in order to complete the proof of the theorem we only need to show that every copy of H in T contains vertices $v_{i_1}, \dots, v_{i_\ell}$ such that $(v_{i_1}, \dots, v_{i_\ell}) \in \mathcal{C}$.

Consider an embedding $\varphi : H \to T$; that is, Imφ is a copy of H in T with $\varphi(v)$ playing the role of v for every $v \in V(H)$. For $i = 1, ..., k$ define $U_i = \varphi^{-1}(B(X_i))$. Then U_i is acyclic by Item 1 in the construction of T above. Consider the vertex-labeling of H with labels $1, ..., h$ in which: (a) for every $1 \leq i < j \leq k$, the labels given to the vertices of U_i are smaller than the labels given to the vertices of U_j, and (b) for every $i = 1, ..., k$, the vertices in U_i are labeled in such a way that all edges are forward-edges, that is, for every $u, v \in U_i$ we have $u \to v$ only if $u < v$ (such a vertex-labeling of U_i exists since U_i is acyclic). Let G be the backedge graph of H with respect to this vertex-labeling. Notice that if $\{u, v\} \in E(G)$ and $u < v$ then there are $1 \leq i < j \leq k$ such that $u \in U_i$ and $v \in U_j$, as U_i is an independent set in G for every $i = 1, \dots, k$.

We claim that the function $f : V(H) \to V(K)$ which maps U_i to a_i is an order-preserving homomorphism from G to K. The fact that f is order-preserving is immediate from the definition of the labeling. To see that f is a graph homomorphism, consider any edge $\{u, v\} \in E(G)$ and assume without loss of generality that $u < v$. By the definition of a backedge graph we have $v \to u$ in H, implying that $\varphi(v) \to \varphi(u)$ in T. As mentioned before, there are $1 \leq i < j \leq k$ such that $u \in U_i$ and $v \in U_j$. Assume for the sake of contradiction, that we have $\{f(u), f(v)\} = \{a_i, a_j\} \notin E(K)$. By (9), this implies that $B(X_i) \to B(X_j)$. Since $\varphi(u) \in B(X_i)$ and $\varphi(v) \in B(X_j)$, we get a contradiction to $\varphi(v) \to \varphi(u)$. Therefore $\{f(u), f(v)\} = \{a_i, a_j\} \in E(K)$, showing that f is a homomorphism.

Having shown that f is an order-preserving homomorphism, we use Corollary 4.7 to infer that there are $u_{i_j} \in f^{-1}(a_{i_j})$, $1 \leq j \leq \ell$, such that $u_{i_1}u_{i_2}\dots u_{i_\ell}u_{i_1}$ is a cycle in G. Denote $v_{i_j} = \varphi(u_{i_j})$ and observe that by the definition of f we have $v_{i_j} \in B(X_{i_j})$. We now show that $(v_{i_1}, \dots, v_{i_\ell}) \in \mathcal{C}$. For every $1 \leq j \leq \ell$ we have $\{u_{i_j}, u_{i_{j+1}}\} \in E(G)$ (with indices taken modulo ℓ). Fix any $1 \leq j \leq \ell$ and assume first that $i_j < i_{j+1}$. Then $u_{i_{j+1}} \to u_{i_j}$ in H by the definition of the backedge graph. Therefore $v_{i_{j+1}} = \varphi(u_{i_{j+1}}) \to \varphi(u_{i_j}) = v_{i_j}$, as φ is an embedding. Similarly, if $i_j > i_{j+1}$ then $u_{i_j} \to u_{i_{j+1}}$ in H by the definition of the backedge graph, implying that $v_{i_j} = \varphi(u_{i_j}) \to \varphi(u_{i_{j+1}}) = v_{i_{j+1}}$. This shows that $(v_{i_1}, \dots, v_{i_\ell}) \in \mathcal{C}$, as required.

Having shown that T contains at most $\varepsilon^{\alpha \log(1/\varepsilon)} n^h$ copies of H, we now prove that T is ε-far from being H-free. We say that an edge e is a *cluster-edge* if it is contained in $B(X_i)$ for some $i = 1, ..., k$, and is a *cut-edge* otherwise. Let T' be any tournament obtained from T by reversing less than εn^2 edges. Our goal is to show that T' contains a copy of H. Let T'' be the tournament that agrees with T on all cut-edges and agrees with T' on all cluster-edges. Then T'' and T' disagree on less than εn^2 edges, and the same is true for T'' and T.

For every $Y = \{y_1, ..., y_k\} \in \mathcal{K}$, the tournament $T''[B(y_1) \cup \cdots \cup B(y_k)]$ is a completion of F (by Item 3 in the construction of T, and because T'' agrees with T on cut-edges). By the choice of F via Lemma 2.1, $T''[B(y_1) \cup \cdots \cup B(y_k)]$ contains a collection $\mathcal{H}(Y)$ of at least $\gamma \left(\frac{n}{r}\right)^2$ copies of H, any two of which do not share cut-edges. Let $Y = \{y_1, \ldots, y_k\}$ and $Z = \{z_1, \ldots, z_k\}$ be distinct cliques in $\mathcal{K}$. Since Y and Z are edge-disjoint, $T''[B(y_1) \cup \cdots \cup B(y_k)]$ and $T''[B(z_1) \cup \cdots \cup B(z_k)]$ do not share cut-edges. Therefore, copies of H from $\mathcal{H}(Y)$ do not share cut-edges with copies of H from $\mathcal{H}(Z)$. Put $\mathcal{H} := \bigcup_{Y \in \mathcal{K}} \mathcal{H}(Y)$. Then $\mathcal{H}$ is a collection of copies of H in T'', any two of which do not share cut-edges. By $|\mathcal{K}| \geq \delta r^2$ and our choice of δ, we have $|\mathcal{H}| \geq \delta r^2 \gamma \left(\frac{n}{r}\right)^2 = \varepsilon n^2$. Since the copies of H in $\mathcal{H}$ do not share cut-edges, one must reverse at least $|\mathcal{H}| \geq \varepsilon n^2$ cut-edges in order to destroy all copies of H in T''. Recall that T' and T'' agree on cluster-edges, and disagree on less than εn^2 edges. Therefore, one of the copies of H in T'' is also present in T'. This completes the proof of the theorem. ■

5 The Hardness of Deciding Tournament Colorability

In this section we prove Theorem 1.2. The main challenge in proving Theorem 1.2 is the case $k = 2$.

Theorem 5.1. *Deciding if a tournament is 2-colorable is NP-hard.*

After proving Theorem 5.1 we show how to deduce Theorem 1.2 from a simple reduction from the k-Colorability problem to the $(k-1)$-Colorability problem for every $k \geq 3$. Theorem 5.1 is proved by showing a reduction from a known NP-hard problem: the *Triangle-Free Cut Problem*, to the Tournament 2-Colorability problem.

Definition 5.2 (**Triangle-Free Cut**). *For an (undirected) graph G, a* triangle-free cut *of G is a 2-coloring of $V(G)$ with no monochromatic triangle.*

It is known that the problem of deciding if a given graph has a triangle-free cut is NP-hard (see [15]).

For a vertex v in a tournament we denote $N^+(v) = \{u : v \to u\}$ and $N^-(v) = \{u : u \to v\}$. If a pair of vertices u, v in a tournament satisfy $u \to v$ then we say that u *dominates* v and that v is *dominated* by u. For the proof of Theorem 5.1 we need the following proposition regarding the gadget H depicted in Figure 1.

Proposition 5.3. *H has the following properties.*

1. *H has a proper 2-coloring in which u and v have the same color and all the vertices in the set $N^-(u) \cup N^+(v)$ have the other color.*

2. *In every proper 2-coloring of H, the colors of u and v are the same.*

Proof. For Item 1, color u, v, w with one color and a, b, c, d with the other color. We now prove Item 2. Consider a 2-coloring of $V(H)$ in which u and v have different colors, say u is colored red and v is colored blue. If there is a color, red or blue, that appears in both $\{a, b\}$ and $\{c, d\}$, then the coloring is not proper, as we get a monochromatic cyclic triangle by joining either u or v. Therefore, we may assume that either a, b are colored with red and c, d are colored with blue, or vice versa. But in both cases there is no color for w as $\{a, b, w\}$ and $\{c, d, w\}$ are cyclic triangles. ■

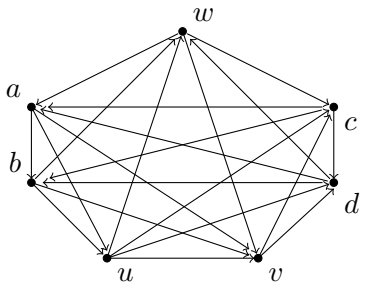

Figure 1: the gadget H

Proof of Theorem 5.1. Given a graph G with vertices $V(G) = \{x_1, ..., x_n\}$, we construct a tournament $T = T(G)$ and prove that G has a triangle-free cut if and only if T is 2-colorable. T is defined as follows. First, we put in T vertices $y_1, ..., y_n$ and set $y_i \to y_j$ for every $i < j$. We think of y_i as corresponding to the vertex x_i of G. Denote $Y = \{y_1, ..., y_n\}$. Let $C_1, ..., C_m$ be an enumeration of all triangles in G. Fix $1 \leq t \leq m$ and suppose that C_t contains the vertices $x_i, x_j, x_k \in V(G)$, where $i < j < k$. We add to T three new vertices, z_t^i, z_t^j, z_t^k, and set $z_t^i \to z_t^j \to z_t^k \to z_t^i$. So $Z_t := \left\{z_t^i, z_t^j, z_t^k\right\}$ spans a cyclic triangle. Set $Z_s \to Z_t$ for each $1 \leq s < t \leq m$. Denote $Z = \bigcup_{t=1}^m Z_t$ and set $Y \to Z$.

Let $1 \leq t \leq m$, suppose that $Z_t = \left\{z_t^i, z_t^j, z_t^k\right\}$, where $i < j < k$, and fix any $\ell \in \{i, j, k\}$. We add a copy of H (see Figure 1), denoted by H_t^ℓ, in which y_ℓ plays the role of u, z_t^ℓ plays the role of v and all other five vertices are new. Notice that this does not contradict $Y \to Z$, as we have $u \to v$ in H. Let K_t^ℓ be the subtournament of H_t^ℓ spanned by the five "new" vertices, that is $V(K_t^\ell) = V(H_t^\ell) \setminus (Y \cup Z)$. Set $K_t^i \to K_t^j \to K_t^k$ and $K_t^i \to K_t^k$. Denote $K_t = K_t^i \cup K_t^j \cup K_t^k$ and for each $1 \leq s < t \leq m$ set $K_s \to K_t$.

Define $K = \bigcup_{t=1}^m K_t$ and note that we have $|Y| = n$, $|Z| = 3m$ and $|K| = 15m$. The vertex set of the tournament $T(G)$ is $Y \uplus Z \uplus K$. So far we defined the edges of $T(G)$ inside Y, Z and K and we set $Y \to Z$. We also already put some edges between Y and K and between K and Z, namely the edges which are contained in H_t^ℓ for some $1 \leq t \leq m$ and $1 \leq \ell \leq n$. We direct all other edges from Y to K and from K to Z; that is, if a pair $(p, q) \in Y \times K$ is not contained in any H_t^ℓ then we set $p \to q$, and similarly for K and Z. In what follows we use the fact that an edge going from K to Y or from Z to K is contained in H_t^ℓ for some $1 \leq t \leq m$ and $1 \leq \ell \leq n$. This completes the definition of the tournament $T = T(G)$.

It remains to show that G has a triangle-free cut if and only if T is 2-colorable. Assume first that T admits a proper 2-coloring, $c : V(T) \to \{\text{red, blue}\}$. For each $i = 1, ..., n$ set $\phi(x_i) = c(y_i)$. We claim that ϕ is a triangle-free cut of G, that is, for every $1 \leq t \leq m$, the triangle C_t in G is not monochromatic. Fix $1 \leq t \leq m$ and suppose that C_t contains the vertices x_i, x_j, x_k. By Item 2 in Proposition 5.3, it must be the case that $c\left(z_t^i\right) = c(y_i)$, $c\left(z_t^j\right) = c(y_j)$ and $c\left(z_t^k\right) = c(y_k)$. Since the set $Z_t = \{z_t^i, z_t^j, z_t^k\} \subseteq V(T)$ spans a cyclic triangle, we deduce that $c(y_i), c(y_j), c(y_k)$ are not all identical. Our choice of ϕ guarantees that C_t is not monochromatic.

Now assume that G admits a triangle-free cut, $\phi : V(G) \to \{\text{red, blue}\}$. We define a 2-coloring c of $V(T)$ as follows. First, set $c(y_i) = \phi(x_i)$ for every $i = 1, \ldots, n$. Next, let $1 \leq t \leq m$ and suppose that $Z_t = \left\{z_t^i, z_t^j, z_t^k\right\}$. For each $\ell \in \{i, j, k\}$ set $c\left(z_t^\ell\right) = c(y_\ell)$. Recall that H_t^ℓ is a copy of H in

which y_ℓ plays the role of u and z_t^ℓ plays the role of v. Extend the coloring of $\{y_\ell, z_t^\ell\}$ to a coloring of H_t^ℓ as in Item 1 of Proposition 5.3, that is, H_t^ℓ is colored properly and any vertex that dominates y_ℓ or that is dominated by z_t^ℓ has a different color from that of y_ℓ, z_t^ℓ. This guarantees that H_t^ℓ does not contain monochromatic edges going from K to Y or from Z to K. As mentioned before, any edge in T going from K to Y or from Z to K is contained in H_t^ℓ for some $1 \leq t \leq m$ and $1 \leq \ell \leq n$. We conclude that T does not contain monochromatic edges going from K to Y or from Z to K.

It remains to show that the 2-coloring c of $V(T) = Y \cup Z \cup K$, defined in the previous paragraph, is proper. Let S be a cyclic triangle in T. We show by case analysis that S is not monochromatic. First we consider the cases (a) $S \subseteq Y \cup K$ and S intersects both Y and K, (b) $S \subseteq K \cup Z$ and S intersects both K and Z, (c) S has one vertex in each of the sets Y, Z, K. Case (a) implies that S contains an edge going from K to Y. Similarly, case (b) implies that S contains an edge that goes from Z to K. Case (c) also implies that S contains an edge from Z to K because $Y \to Z$. As proven in the previous paragraph, T does not contain any monochromatic edge going from K to Y or from Z to K. Therefore, S is not monochromatic in each of the cases (a), (b) and (c).

Given the previous paragraph, the only remaining cases to consider are $S \subseteq Y \cup Z$ and $S \subseteq K$. First, notice that the only cyclic triangles which are contained in Z are $Z_1, \ldots, Z_m$. Let $1 \leq t \leq m$ and suppose that $Z_t = \left\{z_t^i, z_t^j, z_t^k\right\}$.

By the definition of the coloring c we have $c\left(z_t^\ell\right) = c(y_\ell) = \phi(x_\ell)$ for every $\ell \in \{i, j, k\}$. The vertices of the triangle C_t (in G) are x_i, x_j, x_k. Since ϕ is a triangle-free cut, it follows that $\phi(x_i), \phi(x_j), \phi(x_k)$ are not all identical. Therefore $c\left(z_t^i\right), c\left(z_t^j\right), c\left(z_t^k\right)$ are not all identical, namely Z_t is not monochromatic.

Recall that Y is transitive and we have $Y \to Z$. Therefore $Y \cup Z$ does not contain any monochromatic cyclic triangle. Finally, every cyclic triangle inside K is contained in some K_t^ℓ. These triangles are not monochromatic because each K_t^ℓ is colored properly. This finishes the case analysis, showing that T does not contain a monochromatic cyclic triangle and completing the proof of the theorem. ■

Proof of Theorem 1.2. We will show that for every $k \geq 3$ there is a simple reduction from the k-Colorability problem to the $(k-1)$-Colorability problem. Given this reduction, we can prove the theorem by induction on k, with the base case $k = 2$ already settled by Theorem 5.1.

Let T be a tournament. We define a tournament T' as follows. The vertex-set of T' consists of two vertex-disjoint copies of T, denoted T_1 and T_2, and an additional vertex z. We set $T_1 \to T_2 \to z \to T_1$. We now show that T is $(k-1)$-colorable if and only if T' is k-colorable. First, if T is $(k-1)$-colorable then clearly T' is k-colorable: we color T_1 and T_2 according to a proper $(k-1)$-coloring of T, using the same $k-1$ colors for both T_1 and T_2, and then color z with the remaining k'th color. It is easy to see that this k-coloring of T' is proper. In the other direction, suppose that there is a proper coloring $c : V(T') \to [k]$ and assume without loss of generality that $c(z) = k$. Then it cannot be the case that both T_1 and T_2 contain a vertex with color k, as that will imply that there is a cyclic triangle in this color. Therefore, there is $i = 1, 2$ such that T_i is colored with $[k-1]$, implying that T is $(k-1)$-colorable. This completes the proof of the theorem. ■

References

[1] N. Alon, Testing subgraphs in large graphs, Random Structures and Algorithms 21 (2002), 359-370.

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Combinatorica 20 (2000), 451-476.

[3] N. Alon, E. Fischer and I. Newman, Testing of bipartite graph properties, SIAM J. Comput. 37 (2007), 959-976.

[4] N. Alon and A. Shapira, Testing subgraphs in directed graphs, JCSS 69 (2004), 354-382.

[5] N. Alon and A. Shapira, A characterization of easily testable induced subgraphs, Combin. Probab. Comput. 15 (2006), 791-805.

[6] N. Alon and J. H. Spencer, **The probabilistic method**, 3rd ed., Wiley, 2008.

[7] F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression, Proc. National Academy of Sciences USA 32 (1946), 331-332.

[8] E. Berger, K. Choromanski, M. Chudnovsky, J. Fox, M. Loebl, A. Scott, P. Seymour and S. Thomasse, Tournaments and colouring, J. Combin. Theory Ser. B 103 (2013), 1-20.

[9] D. Bokal, G. Fijavž, M. Juvan, P. M. Kayll and B. Mohar, The circular chromatic number of a digraph, J. Graph Theory 46 (2004), 227-240.

[10] D. Conlon and J. Fox, Graph removal lemmas, Surveys in Combinatorics, Cambridge University Press, 2013, 1-50.

[11] J. Fox, J. Pach and A. Suk, Erdős-Hajnal Conjecture for Graphs with Bounded VC-Dimension, in LIPIcs-Leibniz International Proceedings in Informatics (Vol. 77), 2017, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[12] J. Fox and B. Sudakov, Unavoidable patterns, J. Combin. Theory Ser. A 115 (2008), 1561-1569.

[13] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, J. ACM, 45 (1998), 653-750.

[14] L. Lovász and B. Szegedy, Regularity partitions and the topology of graphons, in An irregular mind, 415-445, Bolyai Soc. Math. Stud. 21, János Bolyai Math. Soc., Budapest, 2010.

[15] C. McDiarmid, A random bit-flipping method for seeking agreement, Random Structures and Algorithms, 1996, 121-130.

[16] J. W. Moon, **Topics on tournaments**, Holt, Rinehart and Winston, New York, 1968.

[17] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939-945.

[18] A. Shapira and R. Yuster, Unavoidable tournaments, J. Combin. Theory Ser. B 116 (2016), 191-207.

[19] E. Szemerédi, Regular partitions of graphs, in *Proc. Colloque Inter. CNRS*, 1978, 399–401.