Testing Hereditary Properties of Ordered Graphs and Matrices

Noga Alon
Schools of Mathematics and Computer Science
Tel-Aviv University
Tel-Aviv, Israel
Email: nogaa@tau.ac.il

Omri Ben-Eliezer
School of Computer Science
Tel-Aviv University
Tel-Aviv, Israel
Email: omrib@mail.tau.ac.il

Eldar Fischer
Department of Computer Science
Technion - Israel Institute of Technology
Haifa, Israel
Email: eldar@cs.technion.ac.il

***Abstract*—We consider properties of edge-colored *vertex-ordered graphs* – graphs with a totally ordered vertex set and a finite set of possible edge colors – showing that any hereditary property of such graphs is strongly testable, i.e., testable with a constant number of queries. We also explain how the proof can be adapted to show that any hereditary property of two-dimensional matrices over a finite alphabet (where row and column order is not ignored) is strongly testable. The first result generalizes the result of Alon and Shapira [FOCS'05; SICOMP'08], who showed that any hereditary graph property (without vertex order) is strongly testable. The second result answers and generalizes a conjecture of Alon, Fischer and Newman [SICOMP'07] concerning testing of matrix properties.**

The testability is proved by establishing a removal lemma for vertex-ordered graphs. It states that if such a graph is far enough from satisfying a certain hereditary property, then most of its induced vertex-ordered subgraphs on a certain (large enough) constant number of vertices do not satisfy the property as well.

The proof bridges the gap between techniques related to the regularity lemma, used in the long chain of papers investigating graph testing, and string testing techniques. Along the way we develop a Ramsey-type lemma for multipartite graphs with "undesirable" edges, stating that one can find a Ramsey-type structure in such a graph, in which the density of the undesirable edges is not much higher than the density of those edges in the graph.

***Index Terms*—property testing; removal lemma; ordered graphs; Ramsey; hereditary properties;**

I. Introduction

Property Testing is dedicated to finding fast algorithms for decision problems of the following type: Given a combinatorial structure S, distinguish quickly between the case where S satisfies a property $\mathcal{P}$ and the case where S is *far* from satisfying the property. Being far means that one needs to modify a significant fraction of the data in S to make it satisfy $\mathcal{P}$. Property Testing was first formally defined by Rubinfeld and Sudan [46], and the investigation in the combinatorial context was initiated by Goldreich, Goldwasser and Ron [33]. This area has been very active over the last twenty years, see, e.g. [32] for various surveys on it.

In this paper we focus on property testing of two-dimensional structures over a finite alphabet, or equivalently, two-variable functions with a fixed finite range. Specifically, we consider graphs and matrices. *Graphs* are functions $G\colon \binom{V}{2} \to \{0,1\}$ where V is the vertex set; more generally *edge-colored graphs* (with a finite color set Σ) are functions $G\colon \binom{V}{2} \to \Sigma$. *Matrices* over a finite alphabet Σ (or *images*) are functions $M\colon U \times V \to \Sigma$. In this paper we generally consider edge-colored graphs rather than standard graphs, as the added generality will prove useful later, so the term *graph* usually refers to an edge-colored graph.

For a fixed finite set Σ, a *property* of functions over Σ is simply a collection of functions whose range is Σ. Specifically, any collection of (edge-colored) graphs $G\colon \binom{V}{2} \to \Sigma$ is an *ordered graph property*. As a special case, an *unordered graph property* is an ordered graph property that is also *invariant under vertex permutations*: If $G \in \mathcal{P}$ and π is any permutation on V_G, then the graph G^π, defined by $G^\pi(\pi(u)\pi(v)) = G(uv)$ for any $u \neq v \in V_G$, satisfies $G^\pi \in \mathcal{P}$. Similarly, an *(ordered) matrix property*, or an *image property*, is a collection of functions $M\colon [m] \times [n] \to \Sigma$. For simplicity, most definitions given below are stated only for graphs, but they carry over naturally to matrices.

A graph $G : \binom{[n]}{2} \to \Sigma$ is ϵ-far from the property $\mathcal{P}$ if one needs to modify the value $G(ij)$ for at least $\epsilon\binom{n}{2}$ of the edges ij to make G satisfy $\mathcal{P}$, where ij denotes the (unordered) edge $\{i,j\} \in \binom{[n]}{2}$. A *tester* for the property $\mathcal{P}$ is a randomized algorithm that is given a parameter $\epsilon > 0$ and query access to its input graph G. The tester must distinguish, with error probability at most $1/3$, between the case where G satisfies $\mathcal{P}$ and the case where G is ϵ-far from satisfying $\mathcal{P}$. The tester is said to have *one-sided error* if it always accepts inputs from $\mathcal{P}$, and rejects inputs that are ϵ-far from $\mathcal{P}$ with probability at least $2/3$. It is desirable to obtain testers that are efficient in terms of the *query complexity* (i.e. the maximal possible number of queries made by the tester). A property $\mathcal{P}$ is *strongly testable* if there is a one-sided error tester for $\mathcal{P}$ whose query complexity is bounded by a function $Q(\mathcal{P}, \epsilon)$. In other words, the query complexity of the tester is independent of the size of the input.

From now on, we generally assume (unless it is explicitly stated that we consider unordered graphs) that the vertex set V of a graph G has a total ordering (e.g. the natural one for $V = [n]$), which we denote by $<$. The (induced) *ordered subgraph* of the graph $G : \binom{V}{2} \to \Sigma$ on $U \subseteq V$, where the

Research of the first author supported in part by a USA-Israeli BSF grant 2012/107, by an ISF grant 620/13 and by the Israeli I-Core program.

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.83

elements of U are $u_1 < \ldots < u_k$, is the graph $H\colon \binom{[k]}{2} \to \Sigma$ which satisfies $H(ij) = G(u_i u_j)$ for any $i < j \in [k]$. For a family $\mathcal{F}$ of "forbidden" graphs, the property $\mathcal{P}_{\mathcal{F}}$ of *$\mathcal{F}$-freeness* consists of all graphs G for which any ordered subgraph H of G satisfies $H \notin \mathcal{F}$. Finally, a property $\mathcal{P}$ is *hereditary* if it is closed under taking induced subgraphs. That is, for any $G \in \mathcal{P}$ and any ordered subgraph H of G, it holds that $H \in \mathcal{P}$. Note that a property $\mathcal{P}$ is hereditary if and only if $\mathcal{P} = \mathcal{P}_{\mathcal{F}}$ for some (possibly infinite) family $\mathcal{F}$ of graphs over Σ.

The analogous notions of ordered subgraphs, $\mathcal{F}$-freeness and hereditary properties for matrices are "structure preserving". Here, the *ordered submatrix* of the matrix $M : [m] \times [n] \to \Sigma$ on $A \times B$, where the elements of A and B are $a_1 < \ldots < a_k$ and $b_1 < \ldots < b_l$, is the matrix $N : [k] \times [l] \to \Sigma$ defined by $N(i,j) = M(a_i, b_j)$ for any $i \in [k]$ and $j \in [l]$.

II. Previous results on graphs and matrices

Some of the most interesting results in property testing have been those that identify large families of properties that are efficiently testable, and those that show that large families of properties cannot be tested efficiently.

One of the most widely investigated questions in property testing has been that of characterizing the efficiently testable unordered graph properties. In the seminal paper of Goldreich, Goldwasser and Ron [33] it was shown that all unordered graph properties that can be represented by a certain graph partitioning, including properties such as k-colorability and having a large clique, are (two-sided) testable using a constant number of queries. See also [35]. Alon, Fischer, Krivelevich and Szegedy [5] showed that the property of $\mathcal{F}$-freeness is strongly testable for any finite family $\mathcal{F}$ of forbidden unordered graphs (here the term *unordered graphs* refers to the usual notion of graphs with no order on the vertices). Their main technical result, now known as the *induced graph removal lemma*, is a generalization of the well-known *graph removal lemma* [4], [47].

Theorem II.1 (Induced graph removal lemma [5]). *For any finite family $\mathcal{F}$ of unordered graphs and $\epsilon > 0$ there exists $\delta = \delta(\mathcal{F}, \epsilon) > 0$, such that any graph G which is ϵ-far from $\mathcal{F}$-freeness contains at least δn^q copies of some $F \in \mathcal{F}$ with q vertices.*

The original proof of Theorem II.1 uses a strengthening of the celebrated Szemerédi graph regularity lemma [48], known as the *strong graph regularity lemma*.

It is clear that having a removal lemma for a family $\mathcal{F}$ immediately implies that $\mathcal{F}$-freeness is strongly testable: A simple tester which picks a subgraph H whose size depends only on $\mathcal{F}$ and ϵ, and checks whether H contains graphs from $\mathcal{F}$ or not, is a valid one-sided tester for $\mathcal{F}$-freeness. Hence, removal lemmas have a major role in property testing. They also have implications in different areas of mathematics, such as number theory and discrete geometry. For more details, see the survey of Conlon and Fox [22].

By proving a variant of the induced graph removal lemma that also holds for infinite families, Alon and Shapira [9] generalized the results of [5]. The infinite variant is as follows.

Theorem II.2 (Infinite graph removal lemma [9]). *For any finite or infinite family $\mathcal{F}$ of unordered graphs and $\epsilon > 0$ there exist $\delta = \delta(\mathcal{F}, \epsilon) > 0$ and $q_0 = q_0(\mathcal{F}, \epsilon)$, such that any graph G which is ϵ-far from $\mathcal{F}$-freeness contains at least δn^q copies of some $F \in \mathcal{F}$ on $q \le q_0$ vertices.*

Theorem II.2 directly implies that *any* hereditary unordered graph property is strongly testable, exhibiting the remarkable strength of property testing.

Theorem II.3 (Hereditary graph properties are strongly testable [9]). *Let Σ be a finite set with $|\Sigma| \ge 2$. Any hereditary unordered graph property over Σ is strongly testable.*

Alon, Fischer, Newman and Shapira later presented [7] a complete combinatorial characterization of the graph properties that are testable (with two-sided error) using a constant number of queries, building on results from [27], [35]. Independently, Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi [16], and later Lovász and Szegedy [41], obtained analytic characterizations of testable properties through the theory of graph limits. See also [39], [40].

An efficient finite induced removal lemma for binary matrices with no row and column order was obtained by Alon, Fischer and Newman [6]. In this case, δ^{-1} is polynomial in ϵ^{-1} (where ϵ, δ play the same roles as in the above removal lemmas). It was later shown by Fischer and Rozenberg [28] that when the alphabet is bigger than binary, the dependence of δ^{-1} on ϵ^{-1} is super-polynomial in general, and in fact testing submatrix-freeness over a non-binary alphabet is at least as hard as testing triangle-freeness in graphs, for which the dependence is also known to be super-polynomial in general [1], see also [2]. Actually, the main tool in [6] is an efficient conditional regularity lemma for ordered binary matrices, and it was conjectured there that this regularity lemma can be used to obtain a removal lemma for ordered binary matrices.

Conjecture II.4 (Ordered binary matrix removal lemma [6]). *For any finite family $\mathcal{F}$ of ordered binary matrices and any $\epsilon > 0$ there exists $\delta = \delta(\mathcal{F}, \epsilon)$ such that any $n \times n$ binary matrix which is ϵ-far from $\mathcal{F}$-freeness contains at least δn^{a+b} copies of some $a \times b$ matrix from $\mathcal{F}$.*

In contrast to the abundance of general testing results for two-dimensional structures with an inherent symmetry, such as unordered graphs and matrices, no similar results for *ordered* two-dimensional structures (i.e. structures that do not have any underlying symmetry) have been established. Even seemingly simple special cases, such as F-freeness for a single ordered graph F, or M-freeness for a single 2×2 ordered matrix M, have not been known to be strongly testable in general [2]. A good survey on the role of symmetry in property testing is given by Sudan [49], who suggests that the successful characterization of the strongly testable unordered graph properties is attributable to the underlying symmetry of these properties. See also [34].

Despite the lack of general results as above for the ordered case, property testing of multi-dimensional ordered structures has recently been an active area of research. Notable examples of properties that were investigated in the setting of ordered matrices include monotonicity (see, e.g., [19], [20] for some of the recent works in the matrix setting), extensions of monotonicity such as k-monotonicity [18] and more generally poset properties [26], visual and geometric properties of images, such as connectedness, convexity, being a half plane [44], [14] and being a Lipschitz function [11], [15], and local properties, such as consecutive pattern-freeness [13]. Ordered graphs were less investigated in the context of property testing, but are the subject of many works in Combinatorics and other areas. See, e.g., a recent work on Ramsey-type questions in the ordered setting [23], in which it is shown that Ramsey numbers of simple ordered structures might differ significantly from their unordered counterparts.

Finally, we mention a relevant result on *one-dimensional* structures. Alon, Krivelevich, Newman and Szegedy [8] showed that regular languages are strongly testable. One can combine this result with the well-known Higman's lemma in order theory [37] to show that any hereditary property of words (i.e. one dimensional functions) over a finite alphabet is strongly testable.

III. Our contributions

We prove generalizations of Theorems II.3 and II.2 to the ordered setting, as well as analogous results for matrices. The following result generalizes Theorem II.3.

Theorem III.1 (Hereditary properties of ordered graphs are strongly testable). *Fix a finite set Σ with $|\Sigma| \geq 2$. Any hereditary ordered graph property over Σ is strongly testable.*

To prove Theorem III.1, we establish an *order-preserving* induced graph removal lemma, which holds for finite and infinite families of ordered graphs. This is a generalization of Theorem II.2.

Theorem III.2 (Infinite ordered graph removal lemma). *Fix a finite set Σ with $|\Sigma| \geq 2$. For any (finite or infinite) family $\mathcal{F}$ of ordered graphs $F : \binom{[n_F]}{2} \to \Sigma$ and any $\epsilon > 0$ there exist $q_0 = q_0(\mathcal{F}, \epsilon)$ and $\delta = \delta(\mathcal{F}, \epsilon) > 0$, such that any ordered graph $G : \binom{[n]}{2} \to \Sigma$ that is ϵ-far from $\mathcal{F}$-freeness contains at least δn^q induced copies of some graph $F \in \mathcal{F}$ on $q \leq q_0$ vertices.*

An analogue of Theorem III.1 for matrices is also proved.

Theorem III.3 (Hereditary properties of ordered matrices are strongly testable). *Fix a finite set Σ with $|\Sigma| \geq 2$. Any hereditary (ordered) matrix property over Σ is strongly testable.*

As in the case of ordered graphs, to prove Theorem III.3 we establish the following ordered matrix removal lemma, which holds for finite and infinite families of matrices, and settles a generalized form of Conjecture II.4.

Theorem III.4 (Infinite ordered matrix removal lemma). *Fix a finite set Σ with $|\Sigma| \geq 2$. For any (finite or infinite) family $\mathcal{F}$ of ordered matrices over Σ and any $\epsilon > 0$ there exist $q_0 = q_0(\mathcal{F}, \epsilon) > 0$ and $\delta = \delta(\mathcal{F}, \epsilon) > 0$, such that any ordered matrix over Σ that is ϵ-far from $\mathcal{F}$-freeness contains at least $\delta n^{q+q'}$ copies of some $q \times q'$ matrix $F \in \mathcal{F}$, where $q, q' \leq q_0$.*

Actually, the proof of Theorem III.4 is almost identical to that of Theorem III.2, so we only describe what modifications are needed to make the proof of Theorem III.2 also work here, for the case of square matrices. However, all proofs can be adapted to the non-square case as well. An outline for the proof of the graph case, and of its adaptation to the matrix case, is given in Section V. The full proofs of these results are provided in the full version of this paper [3]. In section VI we state and prove a Ramsey-type lemma for multipartite graphs in the presence of undesirable edges, which is needed for the proof, and we believe it is interesting in its own right.

To the best of our knowledge, Theorems III.1 and III.3 are the first known testing results of this type for ordered two-dimensional structures, and Theorems III.2 and III.4 are the first known order-preserving removal lemmas for two-dimensional structures.

It is interesting to note that some of the properties mentioned in Section II, such as monotonicity, k-monotonicity, and forbidden-poset type properties in matrices, are hereditary (as all of them can be characterized by a finite set of forbidden submatrices), so Theorem III.3 gives a new proof that these properties, and many of their natural extensions, are strongly testable. Naturally, our general testers are much less efficient than the testers specifically tailored for each of these properties (in terms of the dependence of the underlying constants on the parameters of the problem), but the advantage of our result is its generality, that is, the fact that it applies to any hereditary property. Thus, for example, for any fixed ordered graph H and any integer k, the property that an ordered graph G admits a k-edge coloring with no monochromatic (ordered) induced copy of H is strongly testable. As mentioned above Ramsey properties of this type have been considered in the Combinatorics literature, see [23] and the references therein. Another family of examples includes properties of (integer) intervals on the line. Any interval can be encoded by an edge connecting its two endpoints, where the order on the vertices (the endpoints) is the usual order on the real line. A specific example of a hereditary property is that the given set of intervals is closed under intersection. The forbidden structure is a set of 4 vertices $i < j < k < l$ where ik and jl are edges (representing intervals) whereas jk is a non-edge.

Finally, there are various examples of unordered hereditary graph properties that have simple representations using a small *finite* forbidden family of *ordered* subgraphs, while in the unordered representation, the forbidden family is infinite. Some examples of such properties are bipartiteness, being a chordal graph, and being a strongly chordal graph [24], [17]. For such properties, when the input graph is supplied with the "right" ordering of the vertices, one can derive the strong testability using the version of Theorem III.2 for *finite* families of forbidden ordered subgraphs, instead of using the infinite unordered version, Theorem II.2.

IV. Discussion and open questions

Several possible directions for future research follow from our work.

Dependence between the parameters

Our proofs rely heavily on strong variants of the graph regularity lemma. Regularity-based proofs generally have a notoriously bad dependence between the parameters of the problem. In the notation of Theorem III.2, for a fixed finite family $\mathcal{F}$ of forbidden ordered subgraphs, δ^{-1} is generally very large in terms of ϵ^{-1}, meaning that the number of queries required for the corresponding tester for such properties is very large in terms of ϵ^{-1}. Indeed, the original Szemerédi regularity lemma imposes a tower-type dependence between these parameters [36], [30], [42], while the variant we use is at least as strong (and at least as expensive) as the strong regularity lemma [5], which is known to have a wowzer (tower of towers) type dependence between its parameters [21], [38]. Note that for infinite families $\mathcal{F}$ the dependence between the parameters may be arbitrarily bad [10].

In a breakthrough result of Fox [29], the first known proof for the (unordered) graph removal lemma that does not use the regularity lemma is given. However, the dependence between the parameters there is still of a tower type. In any case, it will be interesting to try to obtain a proof for the ordered case, that does not go through the strong regularity needed in our proof.

Better dependence for specific properties

As discussed in Section II, for ordered binary matrices there is an efficient conditional regularity lemma [6], in which the dependence of δ^{-1} on ϵ^{-1} is polynomial. It will be interesting to try to combine the ideas from our proof with this binary matrix regularity lemma, to obtain a removal lemma for finite families of ordered binary matrices with better dependence between the parameters. Ideally, one hopes for a removal lemma with polynomial dependence, but even obtaining such a lemma with, say, exponential dependence will be interesting.

More generally, it will be interesting to find large families of hereditary ordered graph or matrix properties that have more efficient testers than those obtained from our work. See, e.g., [31] for recent results of this type for unordered graph properties.

Characterization of strongly testable ordered properties

For unordered graphs, Alon and Shapira [9] showed that a property is strongly testable using an *oblivious* one-sided tester, which is a tester whose behavior is independent of the size of the input, *if and only if* the property is (almost) hereditary. It will be interesting to obtain similar characterizations in the ordered case.

More generally, in the ordered case there are other general types of properties that may be of interest. Ben-Eliezer, Korman and Reichman [13] recently raised the question of characterizing the efficiently testable *local properties*, i.e., properties that are characterized by a collection of forbidden local substructures. It will also be interesting to identify and investigate large classes of *visual* (or *geometric*) properties. Due to the lack of symmetry, obtaining a complete characterization of the efficiently testable properties of ordered graphs and matrices seems to be very difficult. In fact, considering that all properties whatsoever can be formulated as properties of ordered structures (e.g. strings), any characterization here will have to define and refer to some "graphness" of our setting, even that we do not allow the graph symmetries.

Generalization to ordered hypergraphs and hypermatrices

It will be interesting to obtain similar removal lemmas (and consequently, testing results) for the high-dimensional analogues of ordered graphs and matrices, namely ordered k-uniform hypergraphs and k-dimensional hypermatrices. Such results were proved for unordered hypergraphs [45], [43], [50].

Analytic analogues via graph limits

The theory of graph limits has provided a powerful approach for problems of this type in the unordered case [16], [40], [39]. It will be interesting to define and investigate a limit object for ordered graphs; this may also help with the characterization question above.

V. Proof outline

Here we provide a sketch of the proofs for our removal lemmas. For the full details, please refer to the full version of this paper [3].

A proof of a graph removal lemma typically goes along the following lines: First, the vertex set of the graph is partitioned into a "constant" (not depending on the input graph size itself) number of parts, and a corresponding *regularity scheme* is found. The regularity scheme essentially allows that instead of considering the original graph, one can consider a very simplified picture of a constant size structure approximately representing the graph. On one hand, the structure has to approximate the original graph in the sense that we can "clean" the graph, changing only a small fraction of the edges, so that the new graph will not contain anything not already "predicted" by the representing structure. On the other hand, the structure has to be "truthful", in the sense that everything predicted by it in fact already exists in the graph.

In the simplest case, just a regular partition given by the original Szemerédi Lemma would suffice. More complex cases, like [5] and [9], require a more elaborate regularity scheme. In our case, we provide a regularity scheme that addresses both edge configuration and vertex order, combining a graph regularity scheme with a scheme for strings.

Given a regularity scheme, we provide the graph cleaning procedure, and prove that if the cleaned graph still contains a forbidden subgraph, then the original graph already contains a structure containing many such graphs (this will consist of some vertex sets referenced in the regularity scheme). We show how to use the scheme to prove the removal lemma and the testability theorem for the case of a finite family $\mathcal{F}$ of forbidden subgraphs, and later, we describe how to extend it for the case of a possibly infinite family $\mathcal{F}$. The latter case also requires a formal definition of what it means for the regularity scheme to predict the existence of a forbidden subgraph, while for the finite case it is enough to keep it implicit.

To extract the regularity scheme we need two technical aids. One of which is just a rounding lemma that allows us to properly use integer quantities to approximate real ones. While in many works the question of dealing with issues related to the divisibility of the number of vertices is just hand-waved away, the situation here is complex enough to merit a formal explanation of how rounding works.

In Section VI we develop a Ramsey-type theorem that we believe to be interesting in its own right. The use of Ramsey-type theorems is prevalent in nearly all works dealing with regularity schemes, as a way to allow us to concentrate only on "well-behaved" structures in the scheme when we are about to clean the graph. Because of the extra complication of dealing with vertex-ordered graphs, we cannot just find Ramsey-type instances separately in different parts of the regularity scheme. Instead, we need to find the well-behaved structure "all at once", and furthermore assure that we avoid enough of the "undesirable" parts where the regularity scheme does not reflect the graph. The fraction of undesirable features, while not large, must not depend on any parameters apart from the original distance parameter ϵ (and in particular must not depend on the size of the regularity scheme), which requires us to develop the new Ramsey-type theorem.

Roughly speaking, the theorem states the following: If we have a k-partite edge-colored graph with sufficiently many vertices in each part, then we can find a subgraph where the edges between every two parts are of a single color (determined by the identity of the two parts). However, we do it in a way that satisfies another requirement: If additionally the original graph is supplied with a set of "undesirable" edges comprising an α fraction of the total number of edges, then the subgraph we find will include not more than an $(1+\eta)\alpha$ fraction of the undesirable edges, for an η as small as we would like (in our application $\eta = 1$ will suffice).

Finding a regularity scheme

To prove the removal lemma we need a *regularity scheme*, that is a sequence of vertex sets whose "interaction" with the graph edges, and in our case also the graph vertex order, allows us to carry a cleaning procedure using combinatorial lemmas.

Historically, in the case of properties like triangle-freeness in ordinary graphs, a regular equipartition served well enough as a regularity scheme. One needs then to just remove all edges that are outside the reach of regularity, such as edges between the sets that do not form regular pairs. When moving on to more general properties of graphs, this is not enough. We need a robust partition (see [27]) instead of just a regular one, and then we can find a subset in each of the partition sets so that these "representative" sets will all form regular pairs. This allows us to decide what to do with problem pairs, e.g. whether they should become complete bipartite graphs or become edgeless (we also need to decide what happens *inside* each partition set, but we skip this issue in the sketch).

For vertex ordered graphs, a single robust partition will not do. The reason is that even if we find induced subgraphs using sets of this partition, there will be no guarantees about the vertex order in these subgraphs. The reason is that the sets of the robust partition could interact in complex ways with regards to the vertex order. Ideally we would like every pair of vertex sets to appear in one of the following two possible ways: Either one is completely before the other, or the two are completely "interwoven".

To interact with the vertex order, we consider the robust partition along with a secondary interval partition. If we consider what happens between two intervals, then all vertices in one of the intervals will be before all vertices in the other one. This suggests that further dividing a robust partition according to intervals is a good idea. However, we also need that inside each interval, the relevant robust partition sets will be completely interwoven. In more explicit terms, we consider what happens when we intersect them with intervals of a *refinement* of the original interval partition. If these intersections all have the "correct" sizes in relation to the original interval (i.e., a set that intersects an interval also intersects all relevant sub-intervals with sufficient vertex counts), then we have the "every possible order" guarantee.

To obtain the formulation and existence proof of a regularity scheme suitable for ordered graphs, we first present the concept of *approximating partitions*, showing several useful properties of them. Importantly, the notion of a robust partition is somewhat preserved when moving to a partition approximating it. Then, we develop the lemma that gives us the required scheme. Roughly speaking, it follows the following steps.

- We find a base partition P of the graph G, robust enough with regards to the graph edge colors, so as to ensure that it remains robust even after refining it to make it fit into a secondary interval partition.
- We consider an interval partition J of the vertex set V of G, that is robust with respect to P. That is, if we partition each interval of J into a number of smaller intervals (thus obtaining a refinement J'), most of the smaller intervals will contain about the same ratio of members of each set of P as their corresponding bigger intervals.
- Now we consider what happens if we construct a partition resulting from taking the intersections of the members of P with members of J'. In an ideal world, if a set of P intersects an interval of J, then it would intersect "nicely" also the intervals of J' that are contained in that interval. However, this is only mostly true. Also, this "partition by intersections" will usually not be an equipartition.
- We now modify a bit both P and J, to get Q and I that behave like the ideal picture, and are close enough to P and J. Essentially we move vertices around in P to make the intersections with the intervals in J' have about the same size inside each interval of J. We also modify

the intersection set sizes (which also affects J a little) so they will all be near multiples of a common value (on the order of n). This is so we can divide them further into an equipartition that refines both the robust graph partition and the interval partition. The rounding Lemma mentioned above helps us here.

The above process generates the following scheme. Q is the modified base equipartition, and its size (i.e. number of parts) is denoted by k. I is the modified "bigger intervals" equipartition, and its size is denoted by m. We are allowed to require in advance that m will be large enough (that is, to have m bigger than a predetermined constant m_0). There is an equipartition Q' of size mt which refines both Q and I. That is, each part of Q' is fully contained in a part of Q and a part of I, and so each part of Q contains exactly mt/k parts of Q'. Moreover, there is the "smaller intervals" equipartition I' which refines I, and has size mb where $b = r(m,t)$ for a two-variable function r that we are allowed to choose in advance (r is eventually chosen according to the Ramsey-type arguments needed in the proof). Each part of I contains exactly b parts of I'. Finally, there is a "perfect" equipartition Q'' which refines Q' and I' and has size mbt, such that inside any bigger interval from I, the intersection of each part of Q' with each smaller interval from I' consists of exactly one part of Q''. Additionally, Q' can be taken to be very robust, where we are allowed to choose the robustness parameters in advance.

We are guaranteed that the numbers m and t are bounded in terms of the above function r, the robustness parameters, and m_0 for which we required that $m \geq m_0$. These bounds do not depend on the size of the input graph.

A more formal statement of the regularity scheme is given in Section VII.

Proving a finite removal lemma

Consider an ordered colored graph $G : \binom{[n]}{2} \to \Sigma$, and consider a regularity scheme consisting of equipartitions Q, I, Q', I', Q'' for G as described above.

We start by observing that if Q'' is robust enough, then there is a tuple W of "representatives" for Q'', satisfying the following conditions.

- For each part of Q'' there is exactly one representative, which is a subset of this part.
- Each representative is not too small: it is of order n (where the constants here may depend on all other parameters discussed above, but not on the input size n).
- All pairs of representatives are very regular (in the standard Szemerédi regularity sense).
- The densities of the colors from Σ between pairs of representatives are usually similar to the densities of those colors between the pairs of parts of Q'' containing them. Here the *density* of a color $\sigma \in \Sigma$ between vertex sets A and B is the fraction of σ-colored edges in $A \times B$.

Actually, the idea of using representatives, as presented above, was first developed in [5]. Note that each part of Q' contains exactly b representatives (since it contains b parts from Q'') and each small interval of I' contains exactly t representatives.

Now if Q' is robust enough then the above representatives for Q'' also represent Q' in the following sense: Densities of colors between pairs of representatives are usually similar to the densities of those colors between the pairs of parts of Q' containing them.

Consider a colored graph H whose vertices are the small intervals of I', where the "color" of the edge between two vertices (i.e. small intervals) is the $t \times t$ "density matrix" described as follows: For any pair of representatives, one from each small interval, there is an entry in the density matrix. This entry is the set of all colors from Σ that are dense enough between these two representatives, i.e., all colors whose density between these representatives is above some threshold.

An edge between two vertices of H is considered *undesirable* if the density matrix between these intervals differs significantly from a density matrix of the large intervals from I containing them. If Q' is robust enough, then most density matrices for pairs of small intervals are similar to the density matrices of the pairs of large intervals containing them. Therefore, the number of undesirables in H is small in this case.

Consider now H as an m-partite graph, where each part consists of all of the vertices (small intervals) of H that are contained in a certain large interval from I. We apply the undesirability-preserving Ramsey on H, and then a standard multicolored Ramsey within each part, to obtain an induced subgraph D of H with the following properties.

- D has exactly $d_{\mathcal{F}}$ vertices (small intervals) inside each part of H, where $d_{\mathcal{F}}$ is the maximum number of vertices in a graph from the forbidden family $\mathcal{F}$.
- For any pair of parts of H, all D-edges between these parts have the same "color", i.e. the same density matrix.
- For any part of H, all D-edges inside this part have the same "color".
- The fraction of undesirables among the edges of D is small.

Finally we wish to "clean" the original graph G as dictated by D. For any pair Q'_1, Q'_2 of (not necessarily distinct) parts from Q', let I_1, I_2 be the large intervals from I' containing them, and consider the density matrix that is common to all D-edges between I_1 and I_2. In this matrix there is an entry dedicated to the pair Q'_1, Q'_2, which we refer to as the set of colors from Σ that are "allowed" for this pair. The cleaning of G is done as follows: For every $u \in Q'_1$ and $v \in Q'_2$, if the original color of uv in G is allowed, then we do not recolor uv. Otherwise, we change the color of uv to one of the allowed colors.

It can be shown that if D does not contain many undesirables, then the cleaning does not change the colors of many edges in G. Therefore, if initially G is ϵ-far from $\mathcal{F}$-freeness, then there exists an induced copy of a graph $F \in \mathcal{F}$ in G with $l \leq d_{\mathcal{F}}$ vertices after the cleaning. Considering our cleaning method, it can then be shown that there exist representatives

$R_1, \ldots, R_l$ with the following properties. For any i, all vertices of R_i come before all vertices of R_{i+1} in the ordering of the vertices, and for any $i < j$, the color of $F(ij)$ has high density in $R_i \times R_j$. Recalling that all pairs of representatives are very regular, a well-known lemma implies that the representatives $R_1, \ldots, R_l$ span many copies of F, as desired.

From finite to infinite removal lemma

After the finite removal lemma is established, adapting the proof to the infinite case is surprisingly not difficult. The only problem of the finite proof is that we required D to have exactly $d_{\mathcal{F}}$ vertices in each large interval, where $d_{\mathcal{F}}$ is the maximal number of vertices of a graph in $\mathcal{F}$. This requirement does not make sense when $\mathcal{F}$ is infinite. Instead we show that there is a function $d_{\mathcal{F}}(m,t)$ that "plays the role" of $d_{\mathcal{F}}$ in the infinite case.

$d_{\mathcal{F}}(m,t)$ is roughly defined as follows: We consider the (finite) collection $\mathcal{C}(m,t)$ of all colored graphs with loops that have exactly m vertices, where the set of possible colors is the same as that of H (so the number of possible colors depends only on $|\Sigma|$ and t). We take $d_{\mathcal{F}}(m,t)$ to be the smallest number that guarantees the following. If a graph $C \in \mathcal{C}(m,t)$ exhibits (in some sense) a graph from $\mathcal{F}$, then C also exhibits a graph from $\mathcal{F}$ with no more than $d_{\mathcal{F}}(m,t)$ vertices.

The rest of the proof follows as in the finite case, replacing any occurrence of $d_{\mathcal{F}}$ in the proof with $d_{\mathcal{F}}(m,t)$. Here, if G contains a copy of a graph from $\mathcal{F}$ after the cleaning, then there is a set of no more than $d_{\mathcal{F}}(m,t)$ different representatives that are very regular in pairs and have the "right" densities with respect to some $F \in \mathcal{F}$ with at most $d_{\mathcal{F}}(m,t)$ vertices, so we are done as in the finite case.

From ordered graphs to ordered matrices

To prove Theorem III.4 for square matrices, we reduce the problem to a graph setting. Suppose that $M : U \times V \to \Sigma$ is a matrix, and add an additional color σ_0 to Σ. All edges between U and V will have the original colors from Σ, and edges inside U and inside V will have the new color σ_0. Note that we are not allowed to change colors to or from the color σ_0, as it actually signals "no edge". The proof now follows from the proof for graphs: We can ask the partition I into large intervals to "respect the middle", so all parts of I are either fully contained in U or in V. Moreover, colors of edges inside U or inside V are not modified during the cleaning step, and edges between U and V are not recolored to σ_0, since this color does not appear at all between the relevant representatives (and in particular, does not appear with high density).

To adapt the proof of Theorem III.4 for non-square matrices, we need the divisibility condition to be slightly different than respecting the middle. In the case that $m = o(n)$, we need to construct two separate "large intervals" equipartitions, one for the rows and one for the columns, instead of one such equipartition I as in the graph case. The rest of the proof does not change.

VI. A QUANTITATIVE RAMSEY-TYPE THEOREM

The multicolored Ramsey number $\mathrm{Ram}(s,k)$ is the minimum integer n so that in any coloring of K_n with s colors there is a monochromatic copy of K_k. It is well known that this number exists (i.e. is finite) for any s and k. For our purposes, we will also need a different Ramsey-type result, that keeps track of "undesirable" edges, as described below.

Let Σ be a finite alphabet. A *k-partite Σ-chart* $G = (V_1, \ldots, V_k, c)$ is defined by k disjoint vertex sets $V_1, \ldots, V_k$ and a function $c : E_G \to \Sigma$, where $E_G = \bigcup_{1 \le i < j \le k} V_i \times V_j$. In other words, it is an edge-colored complete k-partite graph. Given a k-partite Σ-chart, we would like to pick a given number of vertices from each partition set, so that all edges between remaining vertices in each pair of sets are of the same color. However, in our situation we also have a "quantitative" requirement: A portion of the edges is marked as "undesirable", and we require that in the chart induced on the picked vertices the ratio of undesirable edges does not increase by much.

Formally, we prove the following, which is stated as a theorem because we believe it may have uses beyond the use in this paper.

Theorem VI.1. *There exists a function $R_{VI.1} : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \times (0,1] \to \mathbb{N}$, so that if $G = (V_1, \ldots, V_k, c)$ is a k-partite Σ-chart with $n \ge R_{VI.1}(|\Sigma|, k, t, \alpha)$ vertices in each class, and $B \subseteq \bigcup_{i<j \in [k]} (V_i \times V_j)$ is a set of "undesirable edges" of size at most $\epsilon \binom{k}{2} n^2$, then G contains an induced subchart $H_{VI.1}(G, B, t, \alpha) = (W_1, \ldots, W_k, c \restriction_{\bigcup_{1 \le i < j \le k} (W_i \times W_j)})$ with the following properties.*

- *$|W_i| = t$ for every $1 \le i \le k$.*
- *$c \restriction_{W_i \times W_j}$ is a constant function for every $1 \le i < j \le k$.*
- *The size of $B \cap (\bigcup_{1 \le i < j \le k} (W_i \times W_j))$ is at most $(1 + \alpha)\epsilon \binom{k}{2} t^2$.*

In our use for the proofs of Theorem III.2 and III.4, these "vertices" would actually be themselves sets of a robust partition of the original graph, and "colors" will encode densities; an undesirable pair would have the "wrong" densities. Also, in our use case the undesirability of an edge will be determined solely by its color and the W_i that its end vertices belong to, which means that for each $1 \le i < i' \le k$ the edge set $W_i \times W_{i'}$ consist of either only desirable edges or only undesirable edges. When this happens, a later pick of smaller sets $W_i' \subset W_i$ will still preserve the ratio of undesirable edges (we will in fact perform such a pick using the original Ramsey's theorem inside each W_i). The following corollary summarizes our use of the theorem.

Definition VI.2. *Given a k-partite Σ-chart $G = (V_1, \ldots, V_k, c)$ and a set $B \subseteq \bigcup_{i<j \in [k]} (V_i \times V_j)$, we say that B is* orderly *if for every $1 \le i < j \le k$ there are no $e \in (V_i \times V_j) \cap B$ and $e' \in (V_i \times V_j) \setminus B$ for which $c(e) = c(e')$. In other words, the "position" and color of an edge fully determines whether it is in B.*

Corollary VI.3. *There exists a function* $R_{VI.3} : \mathbb{N}\times\mathbb{N}\times\mathbb{N} \to \mathbb{N}$*, so that if* $G = (\bigcup_{i=1}^{k} V_i, c)$ *is a* Σ*-colored graph with* $|V_i| = n \geq R_{VI.3}(|\Sigma|, k, t)$ *for any* $i \in [k]$ *and* $V_i \cap V_j = \emptyset$ *for any* $i \neq j \in [k]$*, and* $B \subseteq \bigcup_{i<j\in[k]}(V_i \times V_j)$ *is an orderly set of "undesirable edges" of size at most* $\epsilon\binom{k}{2}n^2$*, then* G *contains an induced subgraph* D *satisfying the following.*

- *The vertex set of* D *is* $\bigcup_{i=1}^{l} U_i$ *where* $U_i \subseteq V_i$ *and* $|U_i| = t$ *for any* $i \in [k]$*.*
- *For any* $i \in [k]$*, all edges inside* U_i *have the same color.*
- *For any* $i < j \in [k]$*, all edges in* $U_i \times U_j$ *have the same color.*
- $\sum_{i<j\in[k]} |B \cap (U_i \times U_j)| \leq 2\epsilon\binom{k}{2}t^2$*.*

Proof: Take $R_{VI.3}(s, k, t) = R_{VI.1}(s, k, \mathrm{Ram}(s, t), 1)$ (recall that $\mathrm{Ram}(s, t)$ denotes the "traditional" s-colored Ramsey function). By Theorem VI.1, there exists a chart $H = (W_1, \ldots, W_k)$ with the following properties.

- $W_i \subseteq V_i$ and $|W_i| = \mathrm{Ram}(t, |\Sigma|)$ for every $i \in [k]$.
- For any pair $i < j \in [k]$, all edges in $W_i \times W_j$ have the same color.
- $\sum_{i<j\in[k]} |B \cap (W_i \times W_j)| \leq 2\epsilon\binom{k}{2}\left(\mathrm{Ram}(t, |\Sigma|)\right)^2$.

Observe that for any pair $i < j \in [k]$, either $W_i \times W_j \subseteq B$ or $(W_i \times W_j) \cap B = \emptyset$, since all edges in $W_i \times W_j$ have the same color and B is orderly. Therefore, the number of pairs $i < j$ for which $(W_i \times W_j) \cap B \neq \emptyset$ is at most $2\epsilon\binom{k}{2}$. Now we apply the traditional Ramsey's theorem inside each W_j to obtain a set $U_j \subseteq W_j$ of size t such that all edges inside W_j have the same color. Since $\sum_{i<j} |B \cap (U_i \times U_j)| \leq \sum_{i<j:(W_i\times W_j)\cap B\neq\emptyset} |B \cap (U_i \times U_j)| \leq 2\epsilon\binom{k}{2}t^2$, the proof follows. ■

Before moving to the proof of Theorem VI.1 itself, let us quickly note that a quantitative counterpart for the traditional (not k-partite) graph case does not exist (indeed, Corollary VI.3 is a way for us to circumvent such issues).

Proposition VI.4. *For any* $\alpha > 0$*,* m*,* k*, and large enough* l*, for infinitely many* n *there is a graph* G *and a set of undesirable pairs* B*, so that* G *has* n *vertices,* B *consists of at most* $\frac{1}{mk}\binom{n}{2}$ *pairs,* G *has no independent set of size* l*, and every clique of* l *vertices in* G *holds at least* $(\frac{1}{m} - \alpha)\binom{l}{2}$ *members of* B*.*

Proof: We construct G for any n that is a multiple of mk larger than lk. The graph G will be the union of k vertex-disjoint cliques, each with n/k vertices. In particular, G contains no independent set with l vertices, and any clique with l vertices must be fully contained in one of the cliques of G.

Now B will be fully contained in the edge-set of G, and will consist of the edge-set of mk vertex-disjoint cliques with n/mk vertices each, so that each of the cliques of G contains m of them. It is now not hard to see that any clique with l vertices in G will contain at least $(\frac{1}{m} - \alpha_l)\binom{l}{2})$ members of B, where $\lim_{l\to\infty} \alpha_l = 0$. ■

Moving to the proof, the following is our main lemma. It essentially says that we can have a probability distribution over "Ramsey-configurations" in our chart that has some approximate uniformity properties.

Lemma VI.5. *There exists a function* $R_{VI.5} : \mathbb{N}\times\mathbb{N}\times\mathbb{N}\times(0, 1] \to \mathbb{N}$*, so that if* $G = (V_1, \ldots, V_k, c)$ *is a* k*-partite* Σ*-chart with* $n \geq R_{VI.5}(|\Sigma|, k, t, \delta)$ *vertices in each class, then* G *contains a randomized induced subchart* $H_{VI.5}(G, t, \delta) = (W_1, \ldots, W_k, c\restriction_{\bigcup_{1\leq i<j\leq k}(W_i\times W_j)})$ *satisfying the following properties.*

- *Either* $|W_i| = t$ *for every* $1 \leq i \leq k$*, or the chart is empty (*$W_i = \emptyset$ *for every* i*).*
- $c\restriction_{W_i\times W_j}$ *is a constant function for every* $1 \leq i < j \leq k$ *(with probability* 1*).*
- *For every* $1 \leq i \leq k$*, every* $v \in V_i$ *will be in* W_i *with probability at most* t/n*.*
- *For every* $1 \leq i < j \leq k$*, every* $v \in V_i$ *and every* $w \in V_j$*, the probability for both* $v \in W_i$ *and* $w \in W_j$ *to hold is bounded by* $(t/n)^2$*.*
- *The probability that the chart is empty is at most* δ*.*

Before we prove this lemma, we show how it implies Theorem VI.1.

Proof of Theorem VI.1: We set $R_{VI.1}(a, k, t, \alpha) = R_{VI.5}(a, k, t, \alpha/3)$. Given the k-partite Σ-chart G, we take the randomized subchart $H = H_{VI.5}(G, t, \alpha/3) = (W_1, \ldots, W_k, c\restriction_{\bigcup_{1\leq i<j\leq k}(W_i\times W_j)})$, and prove that with positive probability it is the required subchart.

Let $B' = B \cap (\bigcup_{1\leq i<j\leq k} W_i \times W_j)$ denote the set of undesirable pairs that are contained in H. By the probability bound on pair containment and by the linearity of expectation, $\mathrm{E}[|B'|] \leq (t/n)^2|B| \leq \epsilon\binom{k}{2}t^2$. Therefore, the probability for $|B'|$ to be larger than $(1+\alpha)\epsilon\binom{k}{2}t^2$ is bounded by $\frac{1}{1+\alpha} \leq 1 - \alpha/2$. Therefore, with positive probability, both $|B'|$ is not too large and H is not the empty chart. Such an H is the desired subchart. ■

To prove Lemma VI.5 we shall make good use of the following near-trivial observation.

Observation VI.6. *There exists a function* $m_{VI.6} : \mathbb{N}\times\mathbb{N}\times(0, 1] \to \mathbb{N}$*, such that if* A *is a set of size at least* $m_{VI.6}(k, t, \delta)$ *and* $\mathcal{A} = (A_1, \ldots, A_k)$ *is a partition of* A *to* k *sets, then there exists a randomized subset* $B = B_{VI.6}(\mathcal{A}, t, \delta)$ *satisfying the following properties.*

- *Either* $|B| = t$ *or* $B = \emptyset$*.*
- B *is fully contained in a single* A_i*.*
- *For every* $a \in A$*, the probability for* $a \in B$ *is at most* $t/|A|$*.*
- *The probability for* $B = \emptyset$ *is at most* δ*.*

Proof: To choose the randomized subset B, first choose a random index I where $\Pr[I = i] = |A_i|/|A|$ for all $1 \leq i \leq k$. Next, if $|A_I| < t$ then set $B = \emptyset$, and otherwise set B to be a subset of size exactly t of A_I, chosen uniformly at random from all $\binom{|A_I|}{t}$ possibilities. Setting $m_{VI.6}(k, t, \delta) = tk/\delta$, it is not hard to see that all properties for the random set B indeed hold. ■

Proof of Lemma VI.5: The proof is done by induction over k. The base case $k = 1$ is easy – set $R_{VI.5}(|\Sigma|, 1, t, \delta) = t$, and let W_1 be a uniformly random subset of size t of V_1.

For the induction step from $k - 1$ to k, we set

$R_{VI.5}(|\Sigma|, k, t, \delta) = m_{VI.6}(|\Sigma|^s, r, \frac{1}{k+1}\delta)$, where $s = m_{VI.6}(|\Sigma|^{k-1}, t, \frac{1}{k+1}\delta)$ and $r = R_{VI.5}(|\Sigma|, k-1, t, \frac{1}{k+1}\delta)$. We set $W_1, \ldots, W_k$ to be the result of the following random process.

First, we set $V_1' \subseteq V_1$ to be a uniformly random subset of size exactly s. Then, for every $2 \le i \le k$, we set $V_i' \subseteq V_i$ to be the random set $B_{VI.6}(\mathcal{V}_i, r, \frac{1}{k+1}\delta)$, where $\mathcal{V}_i$ is the partition of V_i obtained by classifying every $v \in V_i$ according to the colors $\langle c(w, v)\rangle_{w \in V_1'}$, i.e., two vertices in V_i are in the same partition set if all their pairs with vertices from V_1' have the same colors.

If any of the V_i' came out empty, we set all W_i to $\emptyset$ and terminate the algorithm (this occurs with probability at most $\frac{k-1}{k+1}\delta$), and otherwise we continue. Note now, in particular, that for every $w \in V_1$ and $v \in V_i$ the probability for both $w \in V_1'$ and $v \in V_i'$ to hold is bounded by $(s/n)(r/n)$. This is since the probability guarantees of Observation VI.6 hold for any possible value of V_1'. Also, since each V_i' was independently drawn, for $v \in V_i$ and $w \in V_j$ (where $1 < i < j \le k$) the probability for both $v \in V_i'$ and $w \in V_j'$ to hold is bounded by $(r/n)^2$.

We now let H' denote the $(k-1)$-partite Σ-chart induced by $V_2', \ldots, V_k'$, and use the induction hypothesis to (randomly) set $W_2, \ldots, W_k$ as the corresponding vertex sets of $H_{VI.5}(H', t, \frac{1}{k+1}\delta)$. As before, if we receive empty sets then we also set $W_1 = \emptyset$ and terminate. Note that for $1 < i \le k$ and $v \in V_i'$, the probability for v to be in W_i is bounded by t/r. Hence, for $v \in V_i$, the probability for $v \in W_i$ to hold is bounded by $(r/n)(t/r) = t/n$. Similarly, for $1 < i < j \le k$, for every $v \in V_i$ and $w \in V_j$ the probability for both $v \in W_i$ and $w \in W_j$ to hold is bounded by $(t/n)^2$. Also by similar considerations, for $v \in V_1$ and $w \in V_i$, the probability for both $v \in V_1'$ and $w \in W_i$ to hold is bounded by $(s/n)(t/n)$.

Finally, we set W_1 to be the random set $B_{VI.6}(\mathcal{V}', t, \frac{1}{k+1}\delta)$, where $\mathcal{V}'$ is the partition of V_1' obtained by classifying each $v \in V_1'$ by the colors $\langle c(v, w)\rangle_{w \in W_i}$. Note that $c(v, w)$ in that expression depends only on v and the index i for which $w \in W_i$, because of how we chose each V_i' above. In particular, after the choice of W_1, the function $c \restriction_{W_1 \times W_i}$ is constant for each $1 < i \le k$. Again, if we got an empty set for W_1, we set all $W_2, \ldots, W_k$ to be empty as well. By similar considerations as in the preceding steps, also here, for any $v \in V_1$ the probability of $v \in W_1$ is bounded by t/n, and for $w \in V_i$ where $1 < i \le k$, the probability of both $v \in W_1$ and $w \in W_i$ is bounded by $(t/n)^2$.

The probability of obtaining empty sets in any of the steps is bounded by δ by a union bound, and all other properties of the random sets $W_1, \ldots, W_k$ have already been proven above. ■

VII. The Regularity Scheme

Here we provide a more formal statement of the regularity scheme needed for the proof. We start with the definitions required to formally present the scheme.

Basic definitions

Recall that a Σ-*colored graph* $G = (V, c_G)$ is defined by a totally ordered set of vertices V and a function $c_G : \binom{V}{2} \to \Sigma$, and a (k, Σ)-*chart* $C = (V_1, \ldots, V_k, c_C)$ is an edge-colored complete k-partite graph with the parts $V_1, \ldots, V_k$. For C and G as above, where additionally $V = \bigcup_{i=1}^k V_i$, we say that C is a *partition* of G if $V = \bigcup_{i=1}^k V_i$ and $c_G(e) = c_C(e)$ for any edge $e \in E_C$. Moreover, C is *equitable* if $||V_i| - |V_j|| \le 1$ for any $1 \le i, j \le k$; an equitable partition is usually called an *equipartition*. C is an *interval partition* if $v < v'$ for any $v \in V_i, v' \in V_j$ with $i < j$.

For a partition C as above, a (k', Σ)-chart C' which is also a partition of G is said to be a *refinement* of C if we can write $C' = (V_{11}, \ldots, V_{1j_1}, \ldots, V_{k1}, \ldots, V_{kj_k}, c_{C'})$ where $V_i = \bigcup_{l=1}^{j_i} V_{il}$. Note that $c_G(e) = c_C(e) = c_{C'}(e)$ for any edge $e \in E_C$.

For two partitions $P = (V_1, \ldots, V_k)$ and $Q = (U_1, \ldots, U_l)$ of a colored graph G, the *least common refinement (LCR)* $P \sqcap Q$ of P and Q is the partition $(V_1 \cap U_1, \ldots, V_1 \cap U_l, \ldots, V_k \cap U_1, \ldots, V_k \cap U_l)$ (after removing empty sets from the list). Note that even if P and Q are equitable, $P \sqcap Q$ is not necessarily equitable.

Definitions: Robustness and regularity

For two disjoint sets of vertices U, W and a coloring $c \colon U \times W \to \Sigma$, we say that the *density* of $\sigma \in \Sigma$ in (U, W, c) is $d_\sigma(U, W, c) = |(U \times W) \cap c^{-1}(\sigma)|/|U||W|$. the squared density is denoted by $d_\sigma^2(U, W, c)$. The *index* of (U, W, c) is

$$\mathrm{ind}(U, W, c) = \sum_{\sigma \in \Sigma} d_\sigma^2(U, W, c).$$

Note that $0 \le \mathrm{ind}(U, V, c) \le 1$ always holds.

For a partition C as above we define the *index* of C as

$$\mathrm{ind}(C) = \sum_{1 \le i < i' \le k} \frac{|V_i||V_{i'}|}{\binom{|V|}{2}} \mathrm{ind}(V_i, V_{i'}, c\restriction_{V_i \times V_{i'}})$$

where $V = \bigcup_{i=1}^k V_i$. By the Cauchy-Schwarz inequality, for any two partitions C, C' of G where C' is a G-refinement of C we have $0 \le \mathrm{ind}(C) \le \mathrm{ind}(C') \le 1$.

For a function $f : \mathbb{N} \to \mathbb{N}$ and a constant $\gamma > 0$, we say that an equipartition C of size k is (f, γ)-*robust* if there exists no refining equipartition C' of C of size at most $f(k)$ for which $\mathrm{ind}(C') > \mathrm{ind}(C) + \gamma$. The notion of robustness is stronger, for fast enough f, than the more commonly used notion of regularity given below. This was first proved (implicitly) by Ruzsa and Szemerédi in [47]. Here, For a Σ-colored graph $G = (V, c)$ and an equipartition $P = (V_1, \ldots, V_k)$ of G, a pair (V_i, V_j) is ϵ-*regular* if $|d_\sigma(V_i, V_j) - d_\sigma(V_i', V_j')| \le \epsilon$ for any $\sigma \in \Sigma$ and $V_i' \subseteq V_i, V_j' \subseteq V_j$ that satisfy $|V_i'| \ge \epsilon|V_i|, |V_j'| \ge \epsilon|V_j|$. P is an ϵ-*regular partition* if all but at most $\epsilon\binom{k}{2}$ of the pairs (V_i, V_j) are ϵ-regular.

Statement of the regularity scheme

We are finally ready to provide the statement of the regularity scheme, used in the proof of the removal lemma.

Lemma VII.1. *For any positive integer k, real value γ, functions $r : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$, and any n-vertex ordered colored graph G (for large enough n), there exist*

- *An interval equipartition I of G into m parts, where $k \leq m \leq S(\gamma, k, f, r)$.*
- *An equipartition Q' of G into mt parts (not necessarily an interval equipartition) which refines I and is additionally (f, γ)-robust, where $t \leq T(\gamma, k, f, r)$.*
- *An interval equipartition I' into $m \cdot r(m,t)$ parts also refining I, so that the LCR $Q'' = Q' \sqcap I'$ is an equipartition into exactly $mt \cdot r(m,t)$ parts (so each set of Q' intersects "nicely" all relevant intervals in I').*

The flexibility of the statement of Lemma VII.1 has an unexpected benefit: It allows to move from a removal lemma that works for finite forbidden families to one that also works for infinite families with almost no effort, where historically [9], moving from a finite unordered graph removal lemma to an infinite one was harder.

For the existence proof of the regularity scheme, and for more details on how to use it to prove the desired removal lemma, please refer to the full version of this paper [3].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for useful comments.

REFERENCES

[1] N. Alon, *Testing subgraphs in large graphs*, Random Structures & Algorithms 21 (2002), 359–370, doi:10.1002/rsa.10056; also Proc. 42nd Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 2001, 434–441.

[2] N. Alon, O. Ben-Eliezer, *Efficient removal lemmas for matrices*, Proc. 21st RANDOM (2017), to appear.

[3] N. Alon, O. Ben-Eliezer, E. Fischer, *Testing hereditary properties of ordered graphs and matrices*, arXiv:1704.02367, 2017.

[4] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, R. Yuster, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms 16 (1994), 80–109, doi:10.1006/jagm.1994.1005; also Proc. 33th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 1992, 473–481.

[5] N. Alon, E. Fischer, M. Krivelevich, M. Szegedy, *Efficient testing of large graphs*, Combinatorica 20 (2000), 451–476, doi:10.1007/s004930070001; also Proc. 40th Annu. Symp. Foundations of Computer Science, IEEE, 1999, 656–666.

[6] N. Alon, E. Fischer, I. Newman, *Efficient testing of bipartite graphs for forbidden induced subgraphs*, SIAM J. Comput. 37 (2007), 959–976, doi:10.1137/050627915.

[7] N. Alon, E. Fischer, I. Newman, A. Shapira, *A combinatorial characterization of the testable graph properties: it's all about regularity*, SIAM J. Comput. 39 (2009), 143–167, doi:10.1137/060667177; also Proc. 38th Ann. ACM Symp. Theory of Computing (STOC), 251–260, ACM, 2006.

[8] N. Alon, M. Krivelevich, I. Newman, M. Szegedy, *Regular languages are testable with a constant number of queries*, SIAM J. Comput. 30 (2001), 1842–1862, doi:10.1137/S0097539700366528; also Proc. 40th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 1999, 645–655.

[9] N. Alon, A. Shapira, *A characterization of the (natural) graph properties testable with one-sided error*, SIAM J. Comput. 37 (2008), 17031727, doi:10.1137/06064888X; also Proc. 46th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 2005, 429–438.

[10] N. Alon, A. Shapira, *Every monotone graph property is testable* SIAM J. Comput. 38 (2008), 505–522, doi:10.1137/050633445; also Proc. 37th Ann. ACM Symp. Theory of Computing (STOC), 128–137, ACM, 2005.

[11] P. Awasthi, M. Jha, M. Molinaro, S. Raskhodnikova, *Testing Lipschitz functions on hypergrid domains*, Algorithmica 74 (2016), 1055–1081, doi:10.1007/s00453-015-9984-y.

[12] M. Axenovich, Y. Person, S. Puzynina, *A regularity lemma and twins in words*, J. Combin. Theory Ser. A 120 (2013), 733–743, doi:10.1016/j.jcta.2013.01.001.

[13] O. Ben-Eliezer, S. Korman, D. Reichman, *Deleting and testing forbidden patterns in multi-dimensional arrays*, Proc. 44th International Colloquium on Automata, Languages, and Programming (ICALP), 9:1–9.14, 2017, doi:10.4230/LIPIcs.ICALP.2017.9.

[14] P. Berman, M. Murzabulatov, S. Raskhodnikova, *Tolerant testers of image properties*, Proc. 43rd International Colloquium on Automata, Languages, and Programming (ICALP), 90:1–90:14, 2016, doi:10.4230/LIPIcs.ICALP.2016.90.

[15] E. Blais, S. Raskhodnikova, G. Yaroslavtsev, *Lower bounds for testing properties of functions over hypergrid domains*, IEEE 29th Conf. Computational Complexity (CCC), IEEE, 2014, 309–320, doi:10.1109/CCC.2014.38.

[16] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, K. Vesztergombi, *Graph limits and parameter testing*, Proc. 38th Ann. ACM Symp. Theory of Computing (STOC), 261–270, ACM, 2006, doi:10.1145/1132516.1132556.

[17] A. Brandstädt, V. B. Le, J. P. Spinrad, *Graph classes: a survey*, SIAM Monographs on Discrete Mathematics and Applications, SIAM, 1999, doi:10.1137/1.9780898719796.

[18] C. Canonne, E. Grigorescu, S. Guo, A. Kumar, K. Wimmer, *Testing k-monotonicity*, Proc. 8th Innovations in Theoretical Computer Science (2017), to appear.

[19] D. Chakrabarty, C. Seshadhri, *Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids*, Proc. 2013 ACM Symp. Theory of Computing (STOC), 419–428, ACM, 2013, doi:10.1145/2488608.2488661.

[20] D. Chakrabarty, C. Seshadhri, *An optimal lower bound for monotonicity testing over hypergrids*, Theory Comput. 10 (2014), 453–464, doi:10.4086/toc.2014.v010a017.

[21] D. Conlon, J. Fox, *Bounds for graph regularity and removal lemmas*, Geom. Funct. Anal. 22 (2012), 1191–1256, doi:10.1007/s00039-012-0171-x.

[22] D. Conlon, J. Fox, *Graph removal lemmas*, Surveys in Combinatorics 2013, 1–49, London Math. Soc. Lecture Note Ser. 409, Cambridge Univ. Press, Cambridge, 2013, doi:10.1017/CBO9781139506748.002.

[23] D. Conlon, J. Fox, C. Lee, B. Sudakov, *Ordered Ramsey numbers*, J. Combin. Theory Ser. B 122 (2017), 353–383, doi:10.1016/j.jctb.2016.06.007.

[24] P. Damaschke, *Forbidden ordered subgraphs*, Topics in combinatorics and graph theory, 219–229, Physica, 1990, doi:10.1007/978-3-642-46908-4.

[25] U. Feige, T. Koren, M. Tennenholtz, *Chasing ghosts: competing with stateful policies*, SIAM J. Comput. 46 (2017), 190–223, doi:10.1137/14100227X; also Proc. 55th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 2014, 100–109.

[26] E. Fischer, I. Newman, *Testing of matrix-poset properties*, Combinatorica 27 (2007), 293–327, doi:10.1007/s00493-007-2154-3.

[27] E. Fischer, I. Newman, *Testing versus estimation of graph properties*, SIAM J. Comput. 37 (2007), 482–501; also Proc. 37th Ann. ACM Symp. Theory of Computing (STOC), 138–146, ACM, 2005, doi:10.1145/1060590.1060612.

[28] E. Fischer, E. Rozenberg, *Lower bounds for testing forbidden induced substructures in bipartite-graph-like combinatorial objects*, Proc. 11th RANDOM, Vol. 4627 of Lecture Notes in Comput. Sci. (2007), 464-478, doi:10.1007/978-3-540-74208-1.

[29] J. Fox, *A new proof of the graph removal lemma*, Ann. of Math. 174 (2011), 561–579, doi:10.4007/annals.2011.174.1.17.

[30] J. Fox, L. M. Lovász *A tight lower bound for Szemerédi's regularity lemma*, Combinatorica (2017), doi:10.1007/s00493-016-3274-4.

[31] L. Gishboliner, A. Shapira, *Removal lemmas with polynomial bounds*, Proc. 49th Ann. ACM Symp. Theory of Computing (STOC), 510–522, ACM, 2017, doi:10.1145/3055399.3055404.

[32] O. Goldreich (ed.), *Property Testing: Current Research and Surveys*, Vol. 6390 of Lecture Notes in Comput. Sci., Springer, 2010, doi:10.1007/978-3-642-16367-8.

[33] O. Goldreich, S. Goldwasser, D. Ron, *Property testing and its connection to learning and approximation*, J. ACM 45 (1998), 653–750; also Proc. 37th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 1996, 339–348, doi:10.1145/285055.285060.

[34] O. Goldreich, T. Kaufman, *Proximity oblivious testing and the role of invariances*, in Studies in complexity and cryptography, 173–190, Vol.

6650 of Lecture Notes in Comput. Sci., Springer, 2011, doi:10.1007/978-3-642-22670-0.

[35] O. Goldreich, L. Trevisan, *Three theorems regarding testing graph properties*, Random Structures Algorithms 23 (2003), 23–57, doi:10.1002/rsa.10078; also Proc. 42th Annu. Symp. Foundations of Computer Science (FOCS), IEEE, 2001, 460–469.

[36] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Funct. Anal. 7 (1997), 322–337, doi:10.1007/PL00001621.

[37] G. Higman, *Ordering by divisibility in abstract algebras*, Proc. London Math. Soc. (3) 2 (1952), 326–336.

[38] S. Kalyanasundaram, A. Shapira, *A wowzer-type lower bound for the strong regularity lemma*, Proc. Lond. Math. Soc. (3) 106 (2013), 621–649, doi:10.1112/plms/pds045.

[39] L. Lovász, *Large networks and graph limits*, American Mathematical Society, Providence, RI, 2012.

[40] L. Lovász, B. Szegedy, *Szemerédi's Lemma for the Analyst*, Geom. Funct. Anal. 17 (2007), 252–270, doi:10.1007/s00039-007-0599-6.

[41] L. Lovász, B. Szegedy, *Testing properties of graphs and functions*, Israel. J. Math. 178 (2010), 113–156, doi:10.1007/s11856-010-0060-7.

[42] G. Moshkovitz, A. Shapira, *A short proof of Gowers' lower bound for the regularity lemma*, Combinatorica 36 (2016), 187–194, doi:10.1007/s00493-014-3166-4.

[43] B. Nagle, V. Rödl, M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures & Algorithms 28 (2006), 113–179, doi:10.1002/rsa.20117.

[44] S. Raskhodnikova, *Approximate testing of visual properties*, Approximation, randomization, and combinatorial optimization, 370–381, Lecture Notes in Comput. Sci., 2764, Springer, 2003, doi:10.1007/978-3-540-45198-3.

[45] V. Rödl, J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures & Algorithms 25 (2004), 1–42, doi:10.1002/rsa.20017.

[46] R. Rubinfeld, M. Sudan, *Robust characterizations of polynomials with applications to program testing*, Siam J. Comput. 25 (1996), 252–271, doi:10.1137/S0097539793255151.

[47] I. Z. Ruzsa, E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics, Vol. II, 939–945, Coll. Math. Soc. J. Bolyai 18, North-Holland, Amsterdam-New York, 1978.

[48] E. Szemerédi, *Regular partitions of graphs*, Proc. Colloque Inter. CNRS No. 260 (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.

[49] M. Sudan, *Invariance in property testing*, in Property Testing: Current Research and Surveys (O. Goldreich, ed.), Vol. 6390 of Lecture Notes in Computer Science, Springer, 2010, 211–227.

[50] T. Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A 113 (2006), 1257–1280, doi:10.1016/j.jcta.2005.11.006.