Electronic Colloquium on Computational Complexity, Report No. 83 (2000)

Testing graphs for colorability properties

Eldar Fischer *

Abstract

Let P be a property of graphs. An ϵ-test for P is a randomized algorithm which, given the ability to make queries whether a desired pair of vertices of an input graph G with n vertices are adjacent or not, distinguishes, with high probability, between the case of G satisfying P and the case that it has to be modified by adding and removing more than $\epsilon\binom{n}{2}$ edges to make it satisfy P. The property P is called testable, if for every ϵ there exists an ϵ-test for P whose total number of queries is independent of the size of the input graph. Goldreich, Goldwasser and Ron [7] showed that certain graph properties, like k-colorability, admit an ϵ-test. In [2] a first step towards a logical characterization of the testable graph properties was made by proving that all first order properties of type "$\exists\forall$" are testable while there exist first order graph properties of type "$\forall\exists$" which are not testable. For proving the positive part, it was shown that all properties describable by a very general type of coloring problem are testable.

While this result is tight from the standpoint of first order expressions, further steps towards the characterization of the testable graph properties can be taken by considering the coloring problem instead. It is proven here that other classes of graph properties, describable by various generalizations of the coloring notion used in [2], are testable, showing that this approach can broaden the understanding of the nature of the testable graph properties. The proof combines some generalizations of the methods used in [2] with additional methods.

It is also observed that these testability results, as well as results from [2] about existence of non-testable properties, can be extended to certain combinatorial structures related to graphs, such as tournaments.

1 Introduction

In all places where it is not stated otherwise, all graphs considered are finite, undirected, and have neither loops nor parallel edges. We also assume that all such graphs (as well as the other

*NEC Research Institute, 4 Independence Way, Princeton NJ 08540, USA; and DIMACS. Email address: fischer@research.nj.nec.com

ISSN 1433-8092

combinatorial structures discussed in the following) have a labeled set of vertices. In what follows, we use the notation of [4] except where stated otherwise.

Let P be a property of graphs. A graph G with n vertices is called *ϵ-far from satisfying P* if no graph $\tilde{G}$ with the same vertex set, which differs from G in no more than $\epsilon\binom{n}{2}$ places (i.e. can be constructed from G by adding and removing no more than $\epsilon\binom{n}{2}$ edges), satisfies P. An *ϵ-test* for P is a randomized algorithm which, given the quantity n and the ability to make queries whether a desired pair of vertices of an input graph G with n vertices are adjacent or not, distinguishes with probability at least $\frac{2}{3}$ between the case of G satisfying P and the case of G being ϵ-far from satisfying P. The property P is called *testable*, if for every fixed $\epsilon > 0$ there exists an ϵ-test for P whose total number of queries is bounded by a function of ϵ (and is independent of the size of the input graph).

As to be expected, all properties discussed here are invariant with regard to graph isomorphisms. This allows us to assume without loss of generality that a given ϵ-test actually chooses a uniformly random constant size subset of the vertices of the input graph, queries about all vertex pairs in it, and gives the output based on the (queried upon) subgraph induced by the chosen vertex set (see e.g. [2]).

The general notion of property testing was first formulated by Rubinfeld and Sudan [11], who were motivated mainly by its connection to the study of program checking. The study of this notion for combinatorial objects, and mainly for labeled graphs, was introduced by Goldreich, Goldwasser and Ron [7], who showed that several interesting graph properties, like k-colorability, are testable. The fact that k-colorability is testable was, in fact, already proven implicitly in [10] (see also [1]), using the Regularity Lemma of Szemerédi [12], but in the context of property testing it was first studied in [7]. In [2] a first step in the direction of a logical characterization of the testable graph properties was taken, by proving the following.

Theorem 1.1 ([2]) *All first order properties of type "$\exists\forall$" are testable. On the other hand, there exists a first order property of type "$\forall\exists$" which is not testable.*

The definition of first order properties is explained in detail in [2]. Briefly, these are properties that can be formulated by first order expressions about graphs, that is, expressions that contain quantifiers (over vertices), Boolean connectives, equality and adjacency.

In order to prove the positive part of Theorem 1.1 the following notion of (vertex) colorability was defined. Note that this is a generalization of the previously investigated notion of proper k-colorability.

Definition 1 *Suppose we are given c, and a finite family (with repetitions) $\mathcal{F}$ of graphs, each of which is provided with a c-coloring (i.e. a function from its vertex set to $\{1,\ldots,c\}$, which is not necessarily a proper c-coloring in the usual sense).*

A c-coloring of a graph G is called an $\mathcal{F}$-**coloring** *if no member of $\mathcal{F}$ appears as an induced subgraph of G with an identical coloring. A graph G is called* $\mathcal{F}$-**colorable** *if it admits an $\mathcal{F}$-coloring.*

This class of properties, which captures the essence (with regards to property testing) of the first order graph properties of type "$\exists\forall$", was shown in [2] to be testable, and a technique for using the Regularity Lemma of Szemerédi for obtaining results in the context of induced subgraphs was developed for this purpose.

While Theorem 1.1 seems tight from the point of view of first order expressions, many important testable graph properties are not expressible as instances of the first order expressions which were proven there to be testable. If the point of view of colorability is considered, further steps towards the understanding of the testable graph properties can be taken by considering further generalizations of the colorability notion defined above.

In the following, graph properties belonging to more general classes are shown to be testable, by combining some of the general methods of [2] with additional methods and arguments. It is also explained in the following how the results regarding testability (positive and negative) can be generalized to some other graph-related combinatorial structures.

Section 2 presents some definitions and lemmas regarding regular pairs, including the Regularity Lemma of Szemerédi, which are relevant to the proofs here, and section 3 presents some general methods underlying the following proofs. Section 4 shows the testability of graph properties belonging to the following class, which combines colorability with counting.

Definition 2 *Given a finite family $\mathcal{F}$ of c-colored graphs and a sequence $\alpha = (\alpha_1,\ldots,\alpha_c)$ of numbers satisfying $0 \le \alpha_i \le 1$ and $\sum_{i=1}^{c}\alpha_i = 1$, we call a graph G with n vertices* $(\alpha,\mathcal{F})$-**colorable** *if G admits an $\mathcal{F}$-coloring with the additional property that for every i the number of vertices of G which are colored with i is in $\{\lfloor\alpha_i n\rfloor, \lceil\alpha_i n\rceil\}$.*

This testability of $(\alpha,\mathcal{F})$-colorability (for every fixed, given, α and $\mathcal{F}$) implies the testability of some graph properties which are not expressible by colorability notions alone. An example of such a property is that of a graph G with n vertices having a clique with $\lfloor\frac{1}{2}n\rfloor$ vertices (this particular property as well as some other properties involving vertex or edge count were actually proven already in [7] to be testable; the idea here is to find a common generalization of as many previous results as possible).

Since, unlike $\mathcal{F}$-colorability, $(\alpha, \mathcal{F})$-colorability is not a property which is inherited by induced subgraphs, special arguments which apply directly to most subgraphs of a given graph have to be used in conjunction with methods similar to those of [2].

Section 5 introduces some more definitions and lemmas, mostly dealing with a certain combinatorial structure related to graphs. These are used in Section 6 to prove the testability of all graph properties belonging to the following class.

Definition 3 *A* c-**pair-coloring** *of a graph* G *with a vertex set* V *is a function from* $[V]^2$*, the set of all (unordered) pairs of vertices of* G*, to* $\{1, \ldots, c\}$*. Given a finite family (with repetitions)* $\mathcal{F}$ *of c-pair-colored graphs, an* $\mathcal{F}$-**pair-coloring** *of* G *is a c-pair-coloring* $\mathcal{C}$ *of* G *for which no member of* $\mathcal{F}$ *appears as an induced subgraph with the colors assigned to its vertex pairs being identical to those assigned by* $\mathcal{C}$*.* G *is called* $\mathcal{F}$-**pair-colorable** *if it admits an* $\mathcal{F}$*-pair-coloring.*

This is clearly a generalization of the usual notion of edge-colorability, but every property expressible in terms of the above defined (vertex) $\mathcal{F}$-colorability is expressible in terms of this new notion as well. A vertex c-coloring is transformed into a $\binom{c+1}{2}$-pair-coloring by coloring each pair with the union of the colors assigned to its vertex members; it is not hard to ensure in this context (by forbidding a constant number of pair-colored graphs) that the allowable pair-colorings have corresponding vertex c-colorings.

Although this result applies to graphs, its proof requires additional arguments pertaining to the combinatorial structure defined in Section 5.

Section 7 deals with some generalizations of results regarding testing to tournaments, demonstrating that these type of results hold not only to graphs, but hold to related combinatorial structures as well. Since most generalizations of the positive testability results are rather straightforward, this section just highlights the main lemmas and sketches the proofs. In this context, Section 8 contains a proof that there exists a non-testable first order tournament property of type "$\forall\exists$".

Finally, Section 9 summarizes some of what is known and what is not yet known regarding testing of graphs and related structures.

It should be noted that the dependence of the constants given here on ϵ is heavy, as a result of making use of the Regularity Lemma. This could be unavoidable, as there is presently no known testability proof even for the property of being triangle free, but the one making use of this lemma.

2 Regular pairs and equipartitions

For every two nonempty disjoint vertex sets A and B of a graph G, we define $e(A, B)$ to be the number of edges of G between A and B. The *edge density* of the pair is defined by $d(A, B) = \frac{e(A,B)}{|A||B|}$. We say that the pair A, B is γ-*regular*, if for any two subsets A' of A and B' of B, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, their edge density satisfies $|d(A', B') - d(A, B)| < \gamma$.

One simple yet useful property of regularity is that it is somewhat preserved when moving to subsets, as the following trivial lemma shows.

Lemma 2.1 *If A, B is a γ-regular pair with density δ, and $A' \subset A$ and $B' \subset B$ satisfy $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$ for some $\epsilon \geq \gamma$, then A', B' is a $\max\{2, \epsilon^{-1}\}\gamma$-regular pair with density at least $\delta - \gamma$ and at most $\delta + \gamma$.* □

The following lemma shows how the existence of regular pairs implies the existence of many induced subgraphs of a fixed type. Many similar lemmas have been proven in previous works, so we omit the proof here.

Lemma 2.2 *For every $0 < \eta < 1$ and k there exist $\gamma = \gamma_{2.2}(\eta, k)$ and $\delta = \delta_{2.2}(\eta, k)$ with the following property.*

Suppose that H is a graph with vertices $v_1, \ldots, v_k$, and that $V_1, \ldots, V_k$ is a k-tuple of disjoint vertex sets of G such that for every $1 \leq i < i' \leq k$ the pair $V_i, V_{i'}$ is γ-regular, with density at least η if $v_i v_{i'}$ is an edge of H, and with density at most $1 - \eta$ if $v_i v_{i'}$ is not an edge of H. Then, at least $\delta \prod_{i=1}^{k} |V_i|$ of the k-tuples $w_1 \in V_1, \ldots, w_k \in V_k$ span (induced) copies of H where each w_i plays the role of v_i. □

$\gamma_{2.2}(\eta, k)$ and $\delta_{2.2}(\eta, k)$ may and are assumed to be monotone nondecreasing in η and monotone nonincreasing in k. We also assume similar monotonicity properties for the other functions appearing in following lemmas.

Sometimes the fact that most subpairs of a given regular pair are also regular to some extent is useful. In order to prove such a lemma, we use the following alternative definition of regularity.

Definition 4 *For a pair V, W of vertex sets with density η, a vertex $v \in V$ is said to* ϵ-**agree** *with the pair if it has between $(\eta - \epsilon)|W|$ and $(\eta + \epsilon)|W|$ neighbors in W, and in addition all but at most $\epsilon|V|$ of the vertices $v' \in |V|$ are such that the number of mutual neighbors of v and v' in W*

is between $(\eta^2 - \epsilon)|W|$ and $(\eta^2 + \epsilon)|W|$. We define in an analogous manner the instance where a vertex $w \in W$ is said to ϵ-agree with the pair.

The pair V, W is called **ϵ-locally-regular** *if all but at most $\epsilon|V|$ of the vertices of V and all but at most $\epsilon|W|$ of the vertices of W do ϵ-agree with the pair.*

It was proven in [1] that this notion of regularity is equivalent in a sense to the usual notion of regularity.

Lemma 2.3 ([1]) *For every $\epsilon > 0$ there exists $\gamma = \gamma_{2.3}(\epsilon) > 0$ such that every γ-locally-regular pair is also ϵ-regular, and every γ-regular pair is also ϵ-locally-regular.*

The following corollary shows that the regularity property is inherited by most subpairs, with any given size larger than a constant, of a given regular pair. It follows directly from lemma 2.3 by combining it with large deviation inequalities (see e.g. [3], Appendix A).

Corollary 2.4 *For every ϵ there exists $\gamma = \gamma_{2.4}(\epsilon)$, such that for every δ there exists $s = s_{2.4}(\epsilon, \delta)$ satisfying the following. If V, V' is a γ-regular pair with density η, then for every $t > s$ and $t' > s$, at least $1 - \delta$ of the possible choices of $U \subset V$ and $U' \subset V'$ with $|U| = t$ and $|U'| = t'$ satisfy that U, U' is an ϵ-regular pair whose density is between $\eta - \delta$ and $\eta + \delta$.*

Proof sketch: We set $\gamma = \gamma_{2.3}(\frac{1}{2}\gamma_{2.3}(\epsilon))$, and set s large enough to ensure that if V, V' is a $\frac{1}{2}\gamma_{2.3}(\epsilon)$-locally-regular pair with density η, then for all but at most δ of the possible choices of U and U' as above the pair U, U' is $\gamma_{2.3}(\epsilon)$-locally-regular and with density between $\eta - \delta$ and $\eta + \delta$. $\square$

A partition $\mathcal{A} = \{V_i | 1 \le i \le k\}$ of the vertex set of a graph is called an *equipartition* if $|V_i|$ and $|V_{i'}|$ differ by no more than 1 for all $1 \le i < i' \le k$ (so in particular each V_i has one of two possible sizes). The order of such an equipartition $\mathcal{A}$ is the number of its sets (k in the above notation). A *refinement* of such an equipartition $\mathcal{A}$ is an equipartition of the form $\mathcal{B} = \{V_{i,j} | 1 \le i \le k, 1 \le j \le l\}$ such that $V_{i,j}$ is a subset of V_i for every $1 \le i \le k$ and $1 \le j \le l$.

In view of the above properties of regular pairs, it is useful to our purpose to ensure the existence of regular pairs in a way that nearly classifies a given input graph G. This is best done by ensuring the existence of a certain equipartition, using the Regularity Lemma of Szemerédi [12].

Lemma 2.5 (The Regularity Lemma [12]) *For every m and $\epsilon > 0$ there exists $T = T_{2.5}(m, \epsilon)$ with the following property.*

If G is a graph with $n \geq T$ vertices, and $\mathcal{A}$ is an equipartition of the vertex set of G with an order not exceeding m, then there exists a refinement $\mathcal{B}$ of $\mathcal{A}$ with order k, where $m \leq k \leq T$, for which all pairs of sets but at most $\epsilon\binom{k}{2}$ of them are ϵ-regular.

In the following, we assume that the number of vertices of the graph G, which appears in some of the formulations, is large enough (as a function of the other parameters) even when this is not mentioned explicitly.

The following corollary, many versions of which appear in various papers applying the Regularity Lemma, is useful in what follows. It is proven by combining Lemma 2.5 with Turán's Theorem and Ramsey's Theorem (see [4] for their formulation). We omit the details.

Corollary 2.6 *For every h and γ there exists $\delta = \delta_{2.6}(h, \gamma)$ such that for every graph G with $n \geq \delta^{-1}$ vertices there exist disjoint vertex sets $W_1, \ldots, W_h$ satisfying:*

- $|W_i| \geq \delta n$.
- *All $\binom{h}{2}$ pairs are γ-regular.*
- *Either all pairs are with densities at least $\frac{1}{2}$, or all pairs are with densities less than $\frac{1}{2}$.*

□

In order to deal with non-monotonic graph properties as well as monotonic ones, the following variant of Lemma 2.5 from [2], which is suited for finding certain induced subgraphs in G, is used.

Lemma 2.7 ([2]) *For every integer m and function $0 < \mathcal{E}(r) < 1$ there exists $S = S_{2.7}(m, \mathcal{E})$ with the following property.*

If G is a graph with $n \geq S$ vertices, then there exists an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ and a refinement $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ of $\mathcal{A}$ that satisfy:

- $|\mathcal{A}| = k \geq m$ *but* $|\mathcal{B}| = kl \leq S$.
- *For all $1 \leq i < i' \leq k$ but at most $\mathcal{E}(0)\binom{k}{2}$ of them the pair $V_i, V_{i'}$ is $\mathcal{E}(0)$-regular.*
- *For all $1 \leq i < i' \leq k$, for all $1 \leq j, j' \leq l$ but at most $\mathcal{E}(k)l^2$ of them the pair $V_{i,j}, V_{i',j'}$ is $\mathcal{E}(k)$-regular.*
- *All $1 \leq i < i' \leq k$ but at most $\mathcal{E}(0)\binom{k}{2}$ of them are such that for all $1 \leq j, j' \leq l$ but at most $\mathcal{E}(0)l^2$ of them $|d(V_i, V_{i'}) - d(V_{i,j}, V_{i',j'})| < \mathcal{E}(0)$ holds.*

For many proofs regarding testability, including those in Section 6, but excluding those in Section 4, it is enough to use the following corollary.

Corollary 2.8 ([2]) *For every m and $0 < \mathcal{E}(r) < 1$ there exist $S = S_{2.8}(m, \mathcal{E})$ and $\delta = \delta_{2.8}(m, \mathcal{E})$ with the following property.*

If G is a graph with $n \geq S$ vertices then there exist an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of G and an induced subgraph G' of G, with an equipartition $\mathcal{A}' = \{V_i' | 1 \leq i \leq k\}$ of the vertices of G', that satisfy:

- $S \geq k \geq m$.
- $V_i' \subset V_i$ *for all* $i \geq 1$, *and* $|V_i'| \geq \delta n$.
- *In the equipartition $\mathcal{A}'$,* **all** *pairs are $\mathcal{E}(k)$-regular.*
- *All but at most $\mathcal{E}(0)\binom{k}{2}$ of the pairs $1 \leq i < i' \leq k$ are such that $|d(V_i, V_{i'}) - d(V_i', V_{i'}')| < \mathcal{E}(0)$.*

3 Small models for large graphs

Given a large graph with a partition of its vertex set, a possibly much smaller graph can capture many of its properties (at least with respect to worst cases) by having the regularity property. The following somewhat lengthy definition gives the properties required of this smaller graph.

Definition 5 *Given a graph $\tilde{G}$ and a partition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of its vertices, a graph M is an* (h, s, t)**-based** (η, ϵ)**-model for** $\tilde{G}$ **over** $\mathcal{A}$ *if there is a partition $\mathcal{M} = \{W_{i,j} | 1 \leq i \leq k, 1 \leq j \leq h\}$ of its vertex set satisfying the following.*

- $s \leq |W_{i,j}| \leq t$ *for all* $1 \leq i \leq k$ *and* $1 \leq j \leq h$.
- *If $h > 1$, then for every fixed $1 \leq i \leq k$, all pairs $W_{i,j}, W_{i,j'}$ for $1 \leq j < j' \leq h$ are ϵ-regular. Moreover, if $h > 1$, then either all these pairs have densities at least $\frac{1}{2}$, in which case the subgraph of $\tilde{G}$ spanned by V_i is a clique, or all these pairs have densities less than $\frac{1}{2}$, in which case the subgraph of $\tilde{G}$ spanned by V_i is an edgeless graph.*
- *For every $1 \leq i < i' \leq k$ and $1 \leq j, j' \leq h$, the pair $W_{i,j}, W_{i',j'}$ is an ϵ-regular pair. Moreover, for every fixed $1 \leq i < i' \leq k$, one of the following three cases occurs. Either the densities of all the pairs $W_{i,j}, W_{i',j'}$ (where j and j' vary over $\{1, \ldots, h\}$) are between η and $1 - \eta$; or $\tilde{G}$*

contains all possible edges between V_i and $V_{i'}$, and the densities of all these pairs are at least η; or $\tilde{G}$ contains no edge between V_i and $V_{i'}$, and the densities of all these pairs are at most $1-\eta$.

In [2] it is proven implicitly that there exist certain induced subgraphs of G, which model a graph $\tilde{G}$ which has the same vertex set as G and an edge set similar to that of G. Let us review these proofs.

Existence in G of a fairly large induced subgraph, modeling a graph $\tilde{G}$ which is similar to G, follows readily from Corollary 2.8.

Lemma 3.1 (see [2]) *For every m, h and $0 < \mathcal{E}(r) < 1$ there exist $S = S_{3.1}(m,h,\mathcal{E})$ and $\delta = \delta_{3.1}(m,h,\mathcal{E})$, such that for every graph G there exists an equipartition $\mathcal{A} = \{V_i | 1 \le i \le k\}$ with $m \le k \le S$, a graph $\tilde{G}$ with the same vertex set as G which differs from G in less than $\mathcal{E}(0)\binom{n}{2}$ places, and an $(h,\delta n,n)$-based $(\frac{1}{6}\mathcal{E}(0),\mathcal{E}(k))$-model for $\tilde{G}$ over $\mathcal{A}$ which is an induced subgraph of G.*

Proof: We may assume that $\mathcal{E}(r)$ is monotone nonincreasing with r. For convenience, set $\epsilon = \mathcal{E}(0)$. We set

$$S = S_{2.8}(\max\{m, 7\epsilon^{-1}\}, \mathcal{E}')$$

using

$$\mathcal{E}'(r) = \min\{\frac{1}{6}\epsilon, \delta_{2.6}(h,\mathcal{E}(r))\mathcal{E}(r)\},$$

and set

$$\delta = \delta_{2.6}(h,\mathcal{E}(S))\delta_{2.8}(\max\{m,7\epsilon^{-1}\},\mathcal{E}').$$

We apply Corollary 2.8 to G, to find $\mathcal{A} = \{V_i | 1 \le i \le k\}$, G' and $\mathcal{A}' = \{V_i' | 1 \le i \le k\}$, that satisfy $\max\{m, 7\epsilon^{-1}\} \le k \le S_{2.8}(\max\{m,7\epsilon^{-1}\},\mathcal{E}')$ and $|V_i'| \ge \delta_{2.8}(\max\{m,7\epsilon^{-1}\},\mathcal{E}')n$, and the regularity and density properties guaranteed by the corollary with regards to $\mathcal{E}'(r)$.

We use Corollary 2.6 on the subgraph of G induced by each V_i' to obtain $W_{i,1},\ldots,W_{i,h}$, with all pairs being $\mathcal{E}(k)$-regular, and either all of the densities being at least $\frac{1}{2}$ or all of them being less than $\frac{1}{2}$. Note that now all pairs $W_{i,j}, W_{i',j'}$ are $\mathcal{E}(k)$-regular, and that their densities do not differ from those of $V_i', V_{i'}'$ by more than $\frac{1}{6}\epsilon$.

$\tilde{G}$ is defined to be the graph obtained from G by adding and removing edges according to what follows.

- For $1 \le i < i' \le k$ such that $|d(V_i, V_{i'}) - d(V_i', V_{i'}')| > \frac{1}{6}\epsilon$, for all $v \in V_i$ and $v' \in V_{i'}$ the pair vv' becomes an edge if $d(V_i', V_{i'}') \ge \frac{1}{2}$, and becomes a non-edge otherwise. This changes less than $\frac{2}{6}\epsilon\binom{n}{2}$ edges (for n large enough) because there are no more than $\frac{1}{6}\epsilon\binom{k}{2}$ such $1 \le i < i' \le k$.

- For $1 \le i < i' \le k$ such that $d(V_i', V_{i'}') < \frac{2}{6}\epsilon$, all edges between V_i and $V_{i'}$ are removed. For all $1 \le i < i' \le k$ such that $d(V_i', V_{i'}') > 1 - \frac{2}{6}\epsilon$, all non-edges between V_i and $V_{i'}$ become edges. This changes no more than $\frac{3}{6}\epsilon\binom{n}{2}$ edges in addition to those changed by the previous condition.

- For $1 \le i < i' \le k$ such that none of the above holds, the edges of $\tilde{G}$ between V_i and $V_{i'}$ remain exactly those of G.

- If $h > 1$ we perform the following. If for a fixed i all densities of pairs from $W_{i,1}, \ldots, W_{i,h}$ are less than $\frac{1}{2}$, all edges within the vertices of V_i are removed. Otherwise, all the abovementioned densities are at least $\frac{1}{2}$, in which case all non-edges within V_i become edges. This changes less than $\frac{1}{6}\epsilon\binom{n}{2}$ edges, because $k \ge 7\epsilon^{-1}$.

G' is clearly a model graph for $\tilde{G}$ as in the formulation of this lemma, and according to the calculation above $\tilde{G}$ differs from G in less than $\epsilon\binom{n}{2}$ places, as required. □

By combining this with Lemma 2.2 the following corollary is immediate (a detailed proof of a similar one is implied in [2]).

Corollary 3.2 (see [2]) *For every m, h, $\mathcal{S}(r)$ and $0 < \mathcal{E}(r) < 1$ there exist $S = S_{3.2}(m, h, \mathcal{S}, \mathcal{E})$, $t = t_{3.2}(m, h, \mathcal{S}, \mathcal{E})$ and $\delta = \delta_{3.2}(m, h, \mathcal{S}, \mathcal{E})$, such that for every graph G with $n > S$ vertices there exists an equipartition $\mathcal{A} = \{V_i | 1 \le i \le k\}$ with $m \le k \le S$, and a graph $\tilde{G}$ with the same vertex set as G which differs from G in less than $\mathcal{E}(0)\binom{n}{2}$ places, such that at least $\delta\binom{n}{t}$ of the induced subgraphs of G with t vertices are $(h, \mathcal{S}(k), t)$-based $(\frac{1}{3}, \mathcal{E}(k))$-models for $\tilde{G}$ over $\mathcal{A}$.*

Proof sketch: We may assume that $\mathcal{S}(r)$ is monotone nondecreasing. We construct $\tilde{G}$ and $\mathcal{A}$, and then describe the required model graphs. Apply Lemma 3.1 (using appropriate m' and $\mathcal{E}'(r)$) to G, to obtain $\mathcal{A}$, $\tilde{G}$ and G' which satisfy the conditions in its formulation. We now consider any graph H with t vertices and an equipartition $\{W_{i,j}' | 1 \le i \le k, 1 \le j \le h\}$ which satisfies the following, where $t \ge (\mathcal{S}(S) + 1)S$ is chosen to be large enough for the construction below to be possible.

- For every $1 \le i \le k$ if $h > 1$, if the restriction of $\tilde{G}$ to V_i is a clique, then H contains all possible edges between $W_{i,j}'$ and $W_{i,j'}'$ for any $1 \le j < j' \le h$; otherwise the restriction of $\tilde{G}$ to V_i is edgeless, in which case H contains no edges between $W_{i,j}'$ and $W_{i,j'}'$.

- For every $1 \leq i < i' \leq k$, if $\tilde{G}$ contains all edges between V_i and $V_{i'}$, then H contains all edges between $W_{i,j}$ and $W_{i',j'}$ for any $1 \leq j, j' \leq h$; if $\tilde{G}$ contains no edges between V_i and $V_{i'}$, then H contains no edges between $W_{i,j}$ and $W_{i',j'}$; and if non of the above occurs, then $W_{i,j}, W_{i',j'}$ is some fixed $\mathcal{E}(k)$-regular pair whose density is between $\frac{1}{3}$ and $\frac{2}{3}$ (a fixed regular pair satisfying these requirements exists if t is chosen to make $\lfloor \frac{t}{k} \rfloor \geq \lfloor \frac{t}{S} \rfloor$ large enough).

It follows readily that any such H is a model graph for $\tilde{G}$, and Lemma 2.2 ensures that enough of the induced subgraphs of G' satisfy the above for the statement of the proposition to hold for an appropriately chosen δ. $\square$

The above corollary is sufficient for many proofs of testability of colorability properties, because it guarantees that most large enough random induced subgraphs of G contain an induced model graph for an appropriate $\tilde{G}$, and this model can in turn be shown to be colorable only if $\tilde{G}$ is. However, in the case where the number of appearances of each color is also important, a more refined mean of obtaining model graphs must be used.

To shorten notation, in the following we call a c-coloring an *α-coloring* if it is an $(\alpha, \mathcal{F})$-coloring for some $\mathcal{F}$, e.g. $\mathcal{F} = \emptyset$. We also use some vector notation regarding α. For example, $|\alpha - \alpha'|_\infty$ is the largest difference between a component of α and the corresponding component of α'.

Given a model graph M (for some $\tilde{G}$ over some partition $\mathcal{A}$ of its vertices), and a partition $\mathcal{M} = \{W_{i,j} | 1 \leq i \leq k, 1 \leq j \leq h\}$ of the vertices of M as in the notation of the above definition of model graphs, we say that a c-coloring of M is *α-modeling* if its restriction to $\bigcup_{j=1}^{h} W_{i,j}$ is monochromatic for every fixed $1 \leq i \leq k$, and, denoting the common color of the above set by c_i, satisfies also that $|\{i | c_i = g\}| = \alpha_g k$ for every fixed $1 \leq g \leq c$. We are now ready to formulate the lemma regarding existence of model graphs which is required for proving the testability of $(\alpha, \mathcal{F})$-colorability properties.

Lemma 3.3 *For every h, c and function $\mathcal{E}(r)$ there exist $S = S_{3.3}(h, c, \mathcal{E})$, $\delta = \delta_{3.3}(h, c, \mathcal{E})$ and $s = s_{3.3}(h, c, \mathcal{E})$, such that for every graph G with $n > s$ vertices there exists an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ with $k \leq S$, satisfying that for every $t \geq s$, at least $(1 - \mathcal{E}(0))\binom{n}{t}$ of the induced subgraphs G' of G with t vertices satisfy the following.*

For every α-coloring of G' (for any $\alpha = (\alpha_1, \ldots, \alpha_c)$ satisfying $\alpha_i \geq 0$ and $\sum_{i=1}^{c} \alpha_i = 1$), there exists an induced subgraph G'' of G', which is an $(h, \delta t, t)$-based $(\frac{1}{6}\mathcal{E}(0), \mathcal{E}(k))$-model over $\mathcal{A}$ for a graph $\tilde{G}$ which differs from G in less than $\mathcal{E}(0)\binom{n}{2}$ places, so that the restriction of the given coloring of G' to G'' is α'-modeling for some α' satisfying $|\alpha - \alpha'|_\infty \leq \mathcal{E}(0)$.

Proof: We may assume $\mathcal{E}(r) \le \mathcal{E}(0)$, and set $\epsilon = \mathcal{E}(0)$. Define $\mathcal{E}'$ by

$$\mathcal{E}'(r) = \min\left\{\gamma_{2.4}\left(\min\{\frac{1}{4cr}\delta_{2.6}\left(h, \mathcal{E}(r)\right)\mathcal{E}(r), \frac{1}{6}\epsilon\}\right), \frac{1}{192}\epsilon, \frac{1}{16}\binom{r+2}{2}^{-1}\right\},$$

set

$$S = S_{2.7}(\max\{m, 7\epsilon^{-1}\}, \mathcal{E}')$$

with m to be chosen later, and set

$$\delta = \frac{1}{6cS^2}\delta_{2.6}(h, \mathcal{E}(S)).$$

Use Lemma 2.7 on G, to find the partitions $\mathcal{A} = \{V_i | 1 \le i \le k\}$ and $\mathcal{B} = \{V_{i,j} | 1 \le i \le k, 1 \le j \le l\}$ which satisfy the density and regularity properties in its formulation; the partition $\mathcal{A}$ will also play the role of the one appearing in the formulation here.

We choose

$$s = \max\left\{2Ss_{2.4}\left(\min\{\frac{1}{4cS}\delta_{2.6}\left(h, \mathcal{E}(S)\right)\mathcal{E}(S), \frac{1}{6}\epsilon\}, \min\{\frac{1}{12}\epsilon, \frac{1}{2}\epsilon\binom{S}{2}^{-1}\}\right), s_0\right\},$$

where s_0 is chosen so that with probability at least $1 - \frac{1}{2}\epsilon$ a random subgraph G' with $t \ge s$ vertices contains between $\frac{2t}{3kl}$ and $\frac{4t}{3kl}$ vertices from each $V_{i,j}$. Thus, denoting by $V'_{i,j}$ the set of vertices of G' belonging to $V_{i,j}$, by Corollary 2.4 the following properties hold in particular with probability at least $1 - \epsilon$ (these properties are similar to those of the equipartition $\mathcal{B}$).

- $\frac{2t}{3kl} \le |V'_{i,j}| \le \frac{4t}{3kl}$
- For all $1 \le i < i' \le k$, for all $1 \le j, j' \le l$ but at most $\frac{1}{16}\binom{k}{2}^{-1} l^2$ of them the pair $V'_{i,j}, V'_{i',j'}$ is $\min\{\frac{1}{4ck}\delta_{2.6}(h, \mathcal{E}(k))\mathcal{E}(k), \frac{1}{6}\epsilon\}$-regular.
- All $1 \le i < i' \le k$ but at most $\frac{1}{192}\epsilon\binom{k}{2} < \frac{1}{12}\epsilon\binom{k}{2}$ of them are such that for all $1 \le j, j' \le l$ but at most $\frac{1}{192}\epsilon l^2$ of them $|d(V_i, V_{i'}) - d(V'_{i,j}, V'_{i',j'})| \le (\frac{1}{192} + \frac{1}{12})\epsilon < \frac{1}{6}\epsilon$ holds.

We now show that such a graph G' has the properties described in the formulation of this lemma. Given an α-coloring $\mathcal{C}$ of G', choose independently, uniformly, and randomly, a vertex $v_i \in \bigcup_{j=1}^{l} V'_{i,j}$ for every $1 \le i \le k$. Choose m to be large enough to ensure (by large deviation inequalities) that with probability more than $\frac{3}{4}$, the restriction of $\mathcal{C}$ to the subgraph induced by $v_1, \ldots, v_k$ is an α'-coloring for some α' satisfying $|\alpha - \alpha'|_\infty \le \epsilon$ (this choice depends only on ϵ). Denoting by j_i the index for which $v_i \in V'_{i,j_i}$, with probability at least $\frac{3}{4}$ the number of vertices in V'_{i,j_i} which have the same color as v_i is at least $\frac{1}{4ck}|V'_{i,j}|$ (because for every fixed $1 \leq i \leq k$ and $1 \leq g \leq c$, the probability of v_i to be colored g while the number of the other vertices of V'_{i,j_i} so colored is less than $\frac{1}{4ck}|V'_{i,j_i}|$, is bounded by $\frac{1}{4ck}$). Finally, because for every fixed i and j the probability that $j_i = j$ is no more than $\frac{2}{l}$, each of the following properties holds with probability at least $\frac{3}{4}$.

- Within the family $\{V'_{1,j_1}, \ldots, V'_{k,j_k}\}$, all pairs are $\min\{\frac{1}{4ck}\delta_{2.6}(h, \mathcal{E}(k))\mathcal{E}(k), \frac{1}{6}\epsilon\}$-regular; this happens with probability at least $\frac{3}{4}$ because for every fixed $1 \leq i < i' \leq k$ the probability of the pair $V'_{i,j_i}, V'_{i',j_{i'}}$ to be otherwise is no more than $\frac{1}{4}\binom{k}{2}^{-1}$.
- All $1 \leq i < i' \leq k$ but at most $\frac{1}{6}\epsilon\binom{k}{2}$ of them satisfy $|d(V'_{i,j_i}, V'_{i',j_{i'}}) - d(V_i, V_{i'})| < \frac{1}{6}\epsilon$. This happens with probability at least $\frac{3}{4}$ because the average number of pairs $1 \leq i < i' \leq k$ that do not satisfy $|d(V'_{i,j_i}, V'_{i',j_{i'}}) - d(V_i, V_{i'})| < \frac{1}{6}\epsilon$, but do satisfy $|d(V_i, V_{i'}) - d(V'_{i,j}, V'_{i',j'})| < \frac{1}{6}\epsilon$ for all but at most $\frac{1}{192}\epsilon l^2$ of the possible $1 \leq j, j' \leq l$, is no more than $\frac{1}{48}\epsilon\binom{k}{2}$; thus with probability at least $\frac{3}{4}$ there are no more than $\frac{1}{12}\epsilon\binom{k}{2}$ such $1 \leq i < i' \leq k$.

In particular, there exists a choice of $v_1, \ldots, v_k$, with the appropriate $j_1, \ldots, j_k$, for which all of the above properties hold. We fix one such choice. Now, for every $1 \leq i \leq k$, we apply Corollary 2.6 to the subgraph induced by the set of all vertices in V'_{i,j_i} which are colored by $\mathcal{C}$ with the same color as v_i, thus obtaining sets $W_{i,1}, \ldots, W_{i,h}$ such that all their members share the same color, all pairs are $\mathcal{E}(k)$-regular, and their densities are either all at least $\frac{1}{2}$ or all less than $\frac{1}{2}$. Note that also all the pairs $W_{i,j}, W_{i',j'}$ are $\mathcal{E}(k)$-regular.

The subgraph G'' is defined as that spanned by the union of $\{W_{i,j} | 1 \leq i \leq k, 1 \leq j \leq h\}$. By constructing $\tilde{G}$ now in a manner completely analogous to the construction in the proof of Lemma 3.1, G'' is proven to be the required subgraph. □

4 Vertex-colorability with restrictions

This section contains a proof that graph properties defined in terms of (vertex) $(\alpha, \mathcal{F})$-colorability are testable. For convenience, an h-based (η, ϵ)-model graph refers to an (h, s, t)-based (η, ϵ)-model graph for any $1 \leq s \leq t$, e.g. $s = 1$ and $t = \infty$. The main combinatorial result follows from the following lemma about the impossibility of a graph far from being colorable to have a model graph with a modeling coloring, by combining it with Lemma 3.3 about the existence of such a setting.

Lemma 4.1 *For every $\epsilon > 0$ and $\alpha = (\alpha_1, \ldots, \alpha_c)$ satisfying $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{c} \alpha_i = 1$ there exist $\beta = \beta_{4.1}(\epsilon, \alpha)$, such that for every $\eta > 0$ and family $\mathcal{F}$ of c-colored graphs with up to h vertices in each member, there exists $\gamma = \gamma_{4.1}(\eta, \mathcal{F})$ satisfying the following.*

If an h-based (η, γ)-model for $\tilde{G}$ over $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ has an α'-modeling $\mathcal{F}$-coloring for any α' satisfying $|\alpha - \alpha'|_\infty < \beta$, then there exists an $(\alpha, \mathcal{F})$-colorable graph $\hat{G}$ which differs from $\tilde{G}$ in no more than $\epsilon\binom{n}{2}$ places, where n is the number of vertices of $\tilde{G}$.

Proof: Without loss of generality, assume that α_1 is the smallest non-zero coordinate of α. We choose

$$\beta = \min\{\frac{1}{2}\alpha_1, \frac{1}{2c}\epsilon\}$$

and

$$\gamma = \gamma_{2.2}(\eta, h).$$

Given $\tilde{G}$, the model graph H with its partition $\{W_{i,j} | 1 \leq i \leq k, 1 \leq j \leq h\}$, and an α'-modeling $\mathcal{F}$-coloring $\mathcal{C}$ of H satisfying the above, we generate $\hat{G}$, its partition $\hat{\mathcal{A}} = \{\hat{V}_i | 1 \leq i \leq k\}$ and its $(\alpha, \mathcal{F})$-coloring $\mathcal{D}$ as follows.

Since $|\alpha - \alpha'|_\infty < \alpha_1$, for every $1 \leq g \leq c$ either $\alpha_g = 0$ or $\alpha'_g > 0$. In the latter case, let i_g denote one index such that the vertices of $\bigcup_{j=1}^{h} W_{i_g,j}$ are all colored g. Now color as many vertices of $\tilde{G}$ as possible, under the following restrictions. For every i, every vertex of V_i is either colored the same as $\bigcup_{j=1}^{h} W_{i,j}$, or remains uncolored; and for every $1 \leq g \leq c$, no more than $\lfloor \alpha n \rfloor$ vertices of $\tilde{G}$ are colored g. In the following construction of $\hat{\mathcal{A}}$, for every i the set $\hat{V}_i$ holds the vertices of V_i which are colored in this stage, and may hold other vertices as follows. Since $|\alpha - \alpha'|_\infty \leq \frac{1}{2c}\epsilon$, no more than $\frac{1}{2}\epsilon n$ vertices remain uncolored. To complete the construction of the coloring $\mathcal{D}$, for every uncolored vertex v, a color $1 \leq g \leq c$ is chosen such that the final coloring $\mathcal{D}$ is an α-coloring; to complete the construction of $\hat{\mathcal{A}}$ and $\hat{G}$, such a v is placed in $\hat{V}_{i_g}$, and all its incident edges are modified so that H is also a model for $\hat{G}$ over $\hat{\mathcal{A}}$

It remains to prove that $\mathcal{D}$ is an $\mathcal{F}$-coloring. Supposing otherwise, assume that $v_1, \ldots, v_h$ span a graph which appears in $\mathcal{F}$ with the same colors as $\mathcal{D}$. For every $1 \leq j \leq h$, let $\hat{\imath}_j$ denote the index for which $v_j \in \hat{V}_{\hat{\imath}_j}$. Now, considering the sets $W_{\hat{\imath}_1,1}, \ldots, W_{\hat{\imath}_h,h}$, Lemma 2.2 ensures the existence of $w_1 \in W_{\hat{\imath}_1,1}, \ldots, w_h \in W_{\hat{\imath}_h,h}$ that span a graph which appears in $\mathcal{F}$ with the same colors as $\mathcal{C}$, but this is a contradiction as $\mathcal{C}$ is an $\mathcal{F}$-coloring of H. $\square$

Theorem 4.2 *For every ϵ, $\mathcal{F}$, and $\alpha = (\alpha_1, \ldots, \alpha_c)$ satisfying $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{c} \alpha_i = 1$, there exist $s = s(\epsilon, \alpha, \mathcal{F})$ and $\beta = \beta(\epsilon, \alpha, \mathcal{F})$, such that for every graph G with n vertices which is ϵ-far* *from being $(\alpha,\mathcal{F})$-colorable, and for every $t \geq s$, at least $(1-\epsilon)\binom{n}{t}$ of the sets with t vertices of G span graphs which are not $(\alpha',\mathcal{F})$-colorable for any α' satisfying $|\alpha-\alpha'|_\infty < \beta$.*

Proof: We set $\beta = \frac{1}{2}\beta_{4.1}(\frac{1}{2}\epsilon,\alpha)$. We define $\mathcal{E}$, by setting

$$\mathcal{E}(0) = \min\{\frac{1}{2}\epsilon,\beta\},$$

and for every $r > 0$ setting

$$\mathcal{E}(r) = \gamma = \gamma_{4.1}(\frac{1}{6}\mathcal{E}(0),\mathcal{F}).$$

Finally, we set $s = s_{3.3}(h,c,\mathcal{E})$, where c is the number of colors, and h is the maximal number of vertices in a member of $\mathcal{F}$.

If G is ϵ-far from being $(\alpha,\mathcal{F})$-colorable, then more than $(1-\epsilon)\binom{n}{t}$ of the induced subgraphs of G with t vertices (for any fixed $t \geq s$) have the property given in the formulation of Lemma 3.3. If any of these is $(\alpha',\mathcal{F})$-colorable for some α' satisfying $|\alpha-\alpha'|_\infty < \beta$, then it contains a $(\frac{1}{6}\mathcal{E}(0),\gamma)$-model graph for a graph $\tilde{G}$ which differs from G in less than $\frac{1}{2}\epsilon\binom{n}{2}$ places, with an α''-modeling $\mathcal{F}$-coloring for some α'' which satisfies $|\alpha-\alpha''|_\infty < 2\beta$. But then Lemma 4.1 ensures the existence of $\hat{G}$ which differs from G in less than $\epsilon\binom{n}{2}$ places and is also $(\alpha,\mathcal{F})$-colorable, which is a contradiction. □

Corollary 4.3 *For every fixed α and $\mathcal{F}$, the property of a graph being $(\alpha,\mathcal{F})$-colorable is testable.*

Proof: Given ϵ, an ϵ-test for this property is constructed as follows. Let $\beta = \beta(\min\{\frac{1}{3},\epsilon\},\alpha,\mathcal{F})$ be as in Theorem 4.2, and let $t \geq s(\min\{\frac{1}{3},\epsilon\},\alpha,\mathcal{F})$ be large enough to ensure also that given an α-coloring of any graph G with n vertices, at least $\frac{2}{3}\binom{n}{t}$ of its restrictions to sets of t vertices of G are α'-colorings for some α' satisfying $|\alpha-\alpha'|_\infty < \beta$ (such a t independent of n can be found using large deviation inequalities).

The test is as follows: A uniformly random set of t vertices of G is chosen, and all pairs of this sets are queried. The algorithm accepts G if and only if the subgraph of G induced by this set is $(\alpha',\mathcal{F})$-colorable for any α' satisfying $|\alpha-\alpha'|_\infty < \beta$ (this can be checked by an exhaustive search of all possible c-colorings of this subgraph). If G is $(\alpha,\mathcal{F})$-colorable, the existence of such a coloring ensures, by the choice of t, that with probability at least $\frac{2}{3}$ the algorithm accepts. If G is ϵ-far from being $(\alpha,\mathcal{F})$-colorable, Theorem 4.2 ensures that with probability at least $\frac{2}{3}$ the algorithm rejects. □

5 Regularity and charts

This section deals with a combinatorial structure which is a generalization of a graph. It turns out that this structure is useful for proving testability of graph properties which are defined in terms of edge or pair colorings. In the following, recall that $[V]^2$ denotes the set of the unordered pairs of members of V.

Definition 6 *A g-**chart** is an ordered pair $C = (V, P)$ consisting of a set of **vertices** V and a function $P : [V]^2 \to \{1 \dots g\}$, which is called the **pair function** of C.*

In particular, a graph may be regarded as an instance of a 2-chart, and a c-pair-colored graph may be regarded as an instance of a $2c$-chart. The notion of a subchart induced by a subset of the vertex set of the chart is defined in an analogous manner to the notion of an induced subgraph. The *preimage graphs* $C_1, \dots, C_g$ of a g-chart C are defined as the graphs having the same vertex set as C, with the edges of each C_j being exactly the pairs of C for which its pair function gives the value j.

For $1 \le j \le g$ we define the *j-density* of a pair A, B of vertex sets of a g-chart C as its density when considered as a pair of vertex sets of the preimage graph C_j of C. We call such a pair *γ-regular* if it is γ-regular considered as a pair of vertex sets of each of the preimage graphs $C_1, \dots, C_g$ of C. Most of the results related to regularity in graphs have counterparts for charts. Since the proofs are virtually the same, and many similar lemmas have been proven in previous works, we give here only their formulations.

Lemma 5.1 *If A, B is a γ-regular pair in a chart, and $A' \subset A$ and $B' \subset B$ satisfy $|A'| \ge \epsilon|A|$ and $|B'| \ge \epsilon|B|$ for some $\epsilon \ge \gamma$, then A', B' is a $\max\{2, \epsilon^{-1}\}\gamma$-regular pair.* $\square$

Lemma 5.2 (Chart version of Lemma 2.2) *For every $\eta > 0$ and h there exist $\gamma = \gamma_{5.2}(\eta, h)$ and $\delta = \delta_{5.2}(\eta, h)$ with the following property.*

Suppose that C and $D = (\{v_1, \dots, v_h\}, P)$ are charts, and that $V_1, \dots, V_h$ are disjoint vertex sets of C such that for every $1 \le i < i' \le h$ the pair $V_i, V_{i'}$ is γ-regular and its $P(v_i v_{i'})$-density is at least η. Then, at least $\delta \prod_{i=1}^{h} |V_i|$ of the h-tuples $w_1 \in V_1, \dots, w_h \in V_h$ span copies of D where each w_i plays the role of v_i. $\square$

Lemma 5.3 (Regularity Lemma for charts) *For every g, m and $\epsilon > 0$ there exists a number $T = T_{5.3}(g, m, \epsilon)$ with the following property.*

If C is a g-chart with $n \geq T$ vertices, and $\mathcal{A}$ is an equipartition of the vertex set of C with an order not exceeding m, then there exists a refinement $\mathcal{B}$ of $\mathcal{A}$ of order k, where $m \leq k \leq T$, for which all pairs of sets but at most $\epsilon\binom{k}{2}$ of them are ϵ-regular. $\square$

By using this in conjunction with Turán's Theorem and a chart version of Ramsey's Theorem, we arrive at the following chart version of Corollary 2.6.

Corollary 5.4 (Chart version of Corollary 2.6) *For every g, h and γ there exists a quantity $\delta = \delta_{5.4}(g, h, \gamma)$ such that for every g-chart C with $n \geq \delta^{-1}$ vertices there exist disjoint vertex sets $W_1, \ldots, W_h$ satisfying:*

- $|W_i| \geq \delta n$.
- *All $\binom{h}{2}$ pairs are γ-regular.*
- *There exists some $1 \leq j \leq g$ such that the j-density of every pair is at least $\frac{1}{g}$.*

$\square$

We also need the following simple lemma about charts (it is simpler than Corollary 2.8 because nothing is claimed about the densities of the regular pairs it obtains).

Lemma 5.5 *For every g, k and γ there exists $\delta = \delta_{5.5}(g, k, \gamma)$ such that for every g-chart C with $n > \delta^{-1}$ vertices, and every equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of its vertices, there exist $V_i' \subset V_i$ such that $|V_i'| \geq \delta n$ for every $1 \leq i \leq k$ and $V_i', V_{i'}'$ is γ-regular for every $1 \leq i < i' \leq k$.*

Proof: We set $\delta = \frac{1}{2}(T_{5.3}(g, k, \min\{\gamma, \frac{1}{2}\binom{k}{2}^{-1}\}))^{-1}$. We apply Lemma 5.3 to find a refinement $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ of $\mathcal{A}$ with all but at most $\frac{1}{2}\binom{k}{2}^{-1}\binom{kl}{2}$ of its pairs being γ-regular.

Choose randomly, independently and uniformly $1 \leq j_i \leq l$ for each $1 \leq i \leq k$. Clearly, with probability at least $\frac{1}{2}$, all the pairs $V_{i,j_i}, V_{i',j_{i'}}$ are γ-regular. Fixing one such choice of $j_1, \ldots, j_k$, we set $V_i' = V_{i,j_i}$ to obtain the required result. $\square$

6 Pair-colorability

In this section we prove that graph properties defined in terms of $\mathcal{F}$-vertex-colorability are testable. The main combinatorial result follows from the following lemma about the possible colorability of a model graph, by combining it with Corollary 3.2 about the existence of such model graphs.

Lemma 6.1 *For every k and family $\mathcal{F}$ of c-pair-colored graphs there exist $\gamma = \gamma_{6.1}(k,\mathcal{F})$ and $s = s_{6.1}(k,\mathcal{F})$, such that if a $(1,s,n)$-based $(\frac{1}{3},\epsilon)$-model graph M for a graph $\tilde{G}$ over a partition $\mathcal{A} = \{V_i | 1 \le i \le k\}$ is $\mathcal{F}$-pair-colorable, then there exists a graph $\hat{G}$ which may differ from $\tilde{G}$ only on the pairs $\bigcup_{i=1}^k [V_i]^2$ and is $\mathcal{F}$-pair-colorable.*

Proof: Let h denote the maximum number of vertices in a member of $\mathcal{F}$. We choose

$$\gamma = \min\left\{\frac{1}{9}, \delta_{5.5}\left(2c, k, \min\{\frac{1}{9c}, \delta_{5.4}\left(2c, h, \gamma_{5.2}(\frac{1}{9c}, h)\right)\gamma_{5.2}(\frac{1}{9c}, h)\}\right)\right\}$$

and

$$s = (\delta_{5.5}\left(2c, k, \min\{\frac{1}{9c}, \delta_{5.4}\left(2c, h, \gamma_{5.2}(\frac{1}{9c}, h)\right)\gamma_{5.2}(\frac{1}{9c}, h)\}\right)\delta_{5.4}\left(2c, h, \gamma_{5.2}(\frac{1}{9c}, h)\right))^{-1}.$$

Assume that $\{U_i | 1 \le i \le k\}$ is the partition of M corresponding to the partition $\mathcal{A}$ of $\tilde{G}$ (so U_i plays the role of $W_{1,i}$ in the notation of the definition of model graphs). Assuming that $\mathcal{C}$ is an $\mathcal{F}$-pair-coloring of M, we show how to construct $\hat{G}$ and an $\mathcal{F}$-pair-coloring $\mathcal{D}$ of $\hat{G}$. We consider M with its coloring $\mathcal{C}$ as a $2c$-chart, and use Lemma 5.5 to find $U_i' \subset U_i$ such that all pairs of $\mathcal{A}' = \{U_i' | 1 \le i \le k\}$ are $\min\{\frac{1}{9c}, \delta_{5.4}(2c, h, \gamma_{5.2}(\frac{1}{9c}, h))\gamma_{5.2}(\frac{1}{9c}, h)\}$-regular as pairs of a $2c$-chart. To each U_i' we apply Corollary 5.4 to obtain $\{U_{i,j}' | 1 \le i \le k, 1 \le j \le h\}$, with all pairs being $\gamma_{5.2}(\frac{1}{9c}, h)$-regular. $\hat{G}$ is constructed from $\tilde{G}$ as follows.

For every fixed i, if there is a color such that the density of the edges colored with this color is at least $\frac{1}{2c} > \frac{1}{9c}$ in all the pairs of $\{U_{i,j}' | 1 \le j \le h\}$, we denote it by k_i and add to $\hat{G}$ the all possible edges in $[V_i]^2$. Otherwise there exists a color such that the density of the non-edges colored with this color is at least $\frac{1}{2c} > \frac{1}{9c}$ in all pairs of $\{U_{i,j}' | 1 \le j \le h\}$, in which case we denote this color by k_i and remove from $\hat{G}$ all edges in $[V_i]^2$.

Note that for each $i < i'$ such that $\hat{G}$ has both edges and non-edges between V_i and $V_{i'}$, the density of the pair $U_i, U_{i'}$ (in M) is between $\frac{1}{3}$ and $\frac{2}{3}$, and so the regularity of this pair ensures that the density of $U_i', U_{i'}'$ is between $\frac{2}{9}$ and $\frac{7}{9}$. Thus there exist two colors $k_{i,i'}$, $k_{i,i'}'$ such that between $U_{i,j}'$ and $U_{i',j'}'$ the density of the edges between them which are colored with $k_{i,i'}$, as well as the density of the non-edges between them which are colored with $k_{i,i'}'$, is at least $\frac{1}{9c}$ (due to the regularity of $U_i, U_{i'}$ as a pair of vertex sets of the $2c$-chart defined above, we can choose $k_{i,i'}$ and $k_{i,i'}'$ which satisfy that the density of the edges colored with $k_{i,i'}$ as well as the density of the non-edges colored with $k_{i,i'}'$ is at least $\frac{2}{9c}$ between U_i' and $U_{i'}'$). Also, if the subgraph of $\hat{G}$ between V_i and $V_{i'}$ is a complete bipartite graph then $k_{i,i'}$ as above exists (as the density of $U_i, U_{i'}$ is at least $\frac{1}{3}$ and so the density of $U'_i, U'_{i'}$ is at least $\frac{2}{9}$), and if it is an edgeless bipartite graph then $k'_{i,i'}$ as above exists.

The coloring $\mathcal{D}$ of $\hat{G}$ is constructed as follows. For a pair v, v' of vertices, we find i, i' such that $v \in V_i$ and $v' \in V_{i'}$. If $i = i'$, we color the pair by k_i. Otherwise, assume without loss of generality that $i < i'$. In this case, the pair is colored by $k_{i,i'}$ if it is an edge of $\hat{G}$, and by $k'_{i,i'}$ if it is not an edge of $\hat{G}$ (since M is a model for $\tilde{G}$, and thus also for $\hat{G}$, only the values of $k_{i,i'}$ and $k'_{i,i'}$ which were defined above are used).

If a member K of $\mathcal{F}$ with h vertices appears in $\hat{G}$ with identical colors according to $\mathcal{D}$, let $\{v_1, \ldots, v_h\}$ be the vertices of $\hat{G}$ that span it, and let $\{i_1, \ldots, i_h\}$ be such that $v_j \in V_{i_j}$ for $1 \leq j \leq h$. Lemma 5.2 now guarantees the existence of $u_j \in U'_{i_j,j}$ such that the subgraph of M induced by $\{u_1, \ldots, u_h\}$ is identical to K and also has identical colors according to $\mathcal{C}$, which contradicts the assumption that $\mathcal{C}$ is an $\mathcal{F}$-pair-coloring of M. Thus, $\mathcal{D}$ is an $\mathcal{F}$-pair-coloring of $\hat{G}$. $\square$

Theorem 6.2 *For every $\epsilon > 0$ and every family of pair-colored graphs $\mathcal{F}$ there exist $\delta = \delta(\epsilon, \mathcal{F})$ and $t = t(\epsilon, \mathcal{F})$, such that for every graph G with n vertices which is ϵ-far from being $\mathcal{F}$-pair-colorable, at least $\delta\binom{n}{t}$ of the sets of t vertices of G span induced subgraphs which are not $\mathcal{F}$-pair-colorable.*

Proof: We define $\mathcal{E}$ by $\mathcal{E}(0) = \frac{1}{2}\epsilon$ and $\mathcal{E}(r) = \gamma_{6.1}(r, \mathcal{F})$ for $r > 0$; we define a function $\mathcal{S}$ by $\mathcal{S}(r) = s_{6.1}(r, \mathcal{F})$. We set

$$\delta = \delta_{3.2}(3\epsilon^{-1}, 1, \mathcal{S}, \mathcal{E})$$

and

$$t = t_{3.2}(3\epsilon^{-1}, 1, \mathcal{S}, \mathcal{E}).$$

Given a graph G with n vertices which is ϵ-far from being $\mathcal{F}$-pair-colorable, we use Lemma 3.2 to find $\tilde{G}$ which differs from G in less than $\frac{1}{2}\epsilon\binom{n}{2}$ places, and an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ where $3\epsilon^{-1} \leq k \leq S_{3.2}(3\epsilon^{-1}, 1, \mathcal{S}, \mathcal{E})$, such that at least $\delta\binom{n}{t}$ of the induced subgraphs of G are models for $\tilde{G}$ as in the formulation of this lemma.

Since G is ϵ-far from being $\mathcal{F}$-pair-colorable, and because $k \geq 3\epsilon^{-1}$, there exist no graph $\hat{G}$ which differs from $\tilde{G}$ only on $\bigcup_{i=1}^{k}[V_i]^2$ and is $\mathcal{F}$-pair-colorable (as such a graph would differ from G in less than $\epsilon\binom{n}{2}$ places), and so Lemma 6.1 ensures that all the above induced subgraphs are not $\mathcal{F}$-pair-colorable. $\square$

Corollary 6.3 *For any fixed family $\mathcal{F}$ of c-pair-colored subgraphs, the property of a graph being $\mathcal{F}$-pair-colorable is testable.*

Proof: Let $\delta(\epsilon, \mathcal{F})$ and $t(\epsilon, \mathcal{F})$ be as in Theorem 6.2. Consider now the algorithm which queries about all pairs of a uniformly random set of t' vertices of the input graph G, with t' chosen so that $(1-\delta)^{t'/t} \le \frac{1}{3}$, and accepts if and only if the subgraph of G induced by this random set is $\mathcal{F}$-pair-colorable. It is not hard to see that an $\mathcal{F}$-pair-colorable graph is always accepted, while a graph which is ϵ-far from being $\mathcal{F}$-pair-colorable is rejected with probability at least $\frac{2}{3}$. $\square$

7 Efficient testing of large tournaments

Property testing of tournaments is defined in an analogous manner to that of graphs. Namely, an ϵ-*test* for a tournament property P is a randomized algorithm which, given the ability to make queries about pairs of vertices of an input tournament T with n vertices, distinguishes with probability at least $\frac{2}{3}$ between the case of T satisfying P and the case that it has to be modified in more than $\epsilon\binom{n}{2}$ places to make it satisfy P; the property P is called *testable* if for every ϵ there exists an ϵ-test which makes a constant number of queries.

A first order tournament property is a property defined by an expression involving variable which range over vertices, the tournament adjacency relation ("points to"), boolean connectives, and the existential and universal quantifiers. A first order property of type "$\exists\forall$" is one of the form $\exists_{u_1,\ldots,u_t}\forall_{v_1,\ldots,v_s}A(u_1,\ldots,u_t,v_1,\ldots,v_s)$ where the expression A contains no quantifiers.

In the following, we sketch a proof of the testability of first order tournament properties of type "$\exists\forall$". Since the proofs of the lemmas are almost identical to the proofs of the analogous results for graphs, we concentrate in the following on their definitions and formulations.

As with graph properties, we call two tournament properties P and Q *indistinguishable* if for every ϵ these exist only a finite number of graphs which satisfy one of these properties while being ϵ-far from satisfying the other. It is not hard to see (see [2] for the proof of the analogue for graphs) that if P and Q are indistinguishable and P is testable, then so is Q.

For a finite family $\mathcal{F}$ of c-colored tournaments, we say that a tournament T is $\mathcal{F}$-*colorable* if there exists a c-coloring of T such that no member of $\mathcal{F}$ appears in G as an (induced) subtournament with identical vertex colors as those supplied by the coloring of T. The following lemma shows that the testability of properties of type "$\exists\forall$" follows from the testability of $\mathcal{F}$-colorability properties.

Lemma 7.1 *Every first order tournament property of type "$\exists\forall$" is indistinguishable the property of being $\mathcal{F}$-colorable for some $\mathcal{F}$.*

Proof Sketch: Given the property $\exists x_1,\ldots,x_t\forall y_1,\ldots,y_s A(x_1,\ldots,x_t,y_1\ldots,y_s)$, we define $\mathcal{F}$ as follows. We assume that $A(x_1, \ldots, x_t, y_1 \ldots, y_s)$ allows us to restrict our attention to cases where $x_1, \ldots, x_t, y_1 \ldots, y_s$ are all assigned distinct values (otherwise there exists a property satisfying this which is identical to P for all graphs with at least $s+t$ vertices). We use the color set

$$\{(0,0)\} \cup \{(a,b) | 1 \leq a \leq 2^{\binom{t}{2}}, 1 \leq b \leq 2^t\}.$$

For what follows we also use an enumeration of the $2^{\binom{t}{2}}$ possible tournaments with t vertices $u_1, \ldots, u_t$, and an enumeration of the 2^t possible adjacency relations of a vertex v to t vertices $u_1, \ldots, u_t$. We impose upon the coloring of T the following restrictions; each of them is expressible by forbidding certain induced subgraphs with given colorings.

- The color $(0,0)$ appears at most t times in the coloring of G (simply disallow all possible tournaments with $t+1$ vertices which are all colored with this color).
- For $1 \leq a < a' \leq 2^{\binom{t}{2}}$ and $1 \leq b, b' \leq 2^t$, at most one of the colors (a,b) and (a', b') appears in the coloring.
- Let K be a tournament with vertices $w_1, \ldots, w_s$, which are colored with $(a, b_1), \ldots, (a, b_s)$ respectively for some $a > 0$. In order to decide if such a K is to be disallowed, we consider the tournament L with vertices $u_1, \ldots, u_t, v_1, \ldots, v_s$ and the following edges. The edges within $u_1, \ldots, u_t$ are defined in correspondence to a (using the enumeration of all tournaments with t labeled vertices), and for $1 \leq j \leq s$ the edges between v_j and $u_1, \ldots, u_t$ are defined in correspondence to b_j (using the enumeration of all possible adjacencies of a vertex to t other vertices). The subtournament of L induced by $v_1, \ldots, v_s$ is made identical to K where v_i plays the role of w_i. Having thus defined L, we disallow the coloring of K where w_i is colored by (a, b_i) for $1 \leq i \leq s$, if and only if $A(u_1, \ldots, u_t, v_1, \ldots, v_s)$ is false in relation to L.

The proof that this property is indistinguishable from the given first order property is straightforward; see [2] for details of the analogue lemma for graphs. $\square$

We now sketch the proof that the above tournament properties are testable. For every two nonempty disjoint vertex sets A and B of a tournament T, we say that the pair A, B is γ-*regular*, if it is regular considered as a pair of vertex sets of the graph defined to have the same vertices as T with its edges being all pairs u, v such that $u \in A$, $v \in B$, and u points to v in T.

The role of regularity in tournaments is virtually the same as its role in graphs. Consider for example the following lemma.

Lemma 7.2 (Tournament version of Lemma 2.2) *For every $0 < \eta < 1$ and k there exist $\gamma = \gamma_{7.2}(\eta, k)$ and $\delta = \delta_{7.2}(\eta, k)$ with the following property.*

Suppose that H is a tournament with vertices $v_1, \ldots, v_k$, and that $V_1, \ldots, V_k$ is a k-tuple of disjoint vertex sets of T such that for every $1 \leq i < i' \leq k$ the pair $V_i, V_{i'}$ is γ-regular, with the density of the edges pointing from V_i to $V_{i'}$ being at least η if v_i points to $v_{i'}$ in H, and being at most $1 - \eta$ otherwise (so if $v_{i'}$ points to v_i in H then the density of edges pointing from $V_{i'}$ to V_i in T is at least η). Then, at least $\delta \prod_{i=1}^{k} |V_i|$ of the k-tuples $w_1 \in V_1, \ldots, w_k \in V_k$ span copies of H where each w_i plays the role of v_i. $\square$

An *equipartition* $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of the vertices of a tournament is defined in a manner analogous to the definition in context of graphs, but here the ordering of the sets $V_1, \ldots, V_k$ is also important. A *refinement* $\mathcal{B}$ of $\mathcal{A}$ is required also to preserve the order of $\mathcal{A}$, in the sense that if U and U' are members of $\mathcal{B}$ which are subsets of V and V' respectively in $\mathcal{A}$, then either $V = V'$ or U and U' appear in $\mathcal{B}$ in the same order as V and V' in $\mathcal{A}$.

A tournament version of the Regularity lemma was proven in previous works, notably in [5].

Lemma 7.3 (Tournament version of the Regularity Lemma [5]) *For every m and $\epsilon > 0$ there exists $T = T_{7.3}(m, \epsilon)$ with the following property.*

If T is a tournament with $n \geq T$ vertices, and $\mathcal{A}$ is an equipartition of the vertex set of T with an order not exceeding m, then there exists a refinement $\mathcal{B}$ of $\mathcal{A}$ of order k, where $m \leq k \leq T$, for which all pairs of sets but at most $\epsilon\binom{k}{2}$ of them are ϵ-regular.

In the following $d(V, V')$ denotes the number of edges pointing from a vertex of V to a vertex of V', divided by $|V||V'|$. By using the above lemma repeatedly in a manner quite analogous to the proof of Lemma 2.7 in [2], one can arrive at the following.

Lemma 7.4 (Tournament version of Lemma 2.7) *For every m and every $0 < \mathcal{E}(r) < 1$ there exists $S = S_{7.4}(m, \mathcal{E})$ with the following property.*

If T is a tournament with $n \geq S$ vertices, then there exists an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ and a refinement $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ of $\mathcal{A}$ that satisfy:

- *$|\mathcal{A}| = k \geq m$ but $|\mathcal{B}| = kl \leq S$.*
- *For all $1 \leq i < i' \leq k$ but at most $\mathcal{E}(0)\binom{k}{2}$ of them the pair $V_i, V_{i'}$ is $\mathcal{E}(0)$-regular.*

- *For all $1 \leq i < i' \leq k$, for all $1 \leq j, j' \leq l$ but at most $\mathcal{E}(k)l^2$ of them the pair $V_{i,j}, V_{i',j'}$ is $\mathcal{E}(k)$-regular.*

- *All $1 \leq i < i' \leq k$ but at most $\mathcal{E}(0)\binom{k}{2}$ of them are such that for all $1 \leq j, j' \leq l$ but at most $\mathcal{E}(0)l^2$ of them $|d(V_i, V_{i'}) - d(V_{i,j}, V_{i',j'})| < \mathcal{E}(0)$ holds.*

$\square$

The above lemma is used to find small tournaments which capture the essence of large ones. The following definition is a tournament analogue of the definition of a model graph.

Definition 7 (Model tournaments) *Given a tournament $\tilde{T}$ and a partition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of its vertices, an (h, s, t)-***based** *(η, ϵ)-***model for** *T* **over** *$\mathcal{A}$ is a tournament M with a partition $\mathcal{M} = \{W_{i,j} | 1 \leq i \leq k, 1 \leq j \leq h\}$ of its vertex set, satisfying the following.*

- *$s \leq |W_{i,j}| \leq t$ for all $1 \leq i \leq k$ and $1 \leq j \leq h$.*

- *If $h > 1$, then for every fixed $1 \leq i \leq k$, all pairs $W_{i,j}, W_{i,j'}$ for $1 \leq j < j' \leq h$ are ϵ-regular. Moreover, if $h > 1$, then for every $1 \leq j < j' \leq h$ the density of the edges pointing from $W_{i,j}$ to $W_{i,j'}$ is at least $\frac{1}{2}$, and the subtournament of T spanned by V_i is transitive.*

- *For every $1 \leq i < i' \leq k$ and $1 \leq j, j' \leq h$, the pair $W_{i,j}, W_{i',j'}$ is an ϵ-regular pair. Moreover, for every fixed $1 \leq i < i' \leq k$, one of the following three cases occurs. Either all the pairs $W_{i,j}, W_{i',j'}$ (where j and j' vary over $\{1, \ldots, h\}$) satisfy $\eta \leq d(W_{i,j}, W_{i',j'}) \leq 1 - \eta$; or all edges of $\tilde{T}$ between V_i and $V_{i'}$ point from V_i to $V_{i'}$, and $\eta \leq d(W_{i,j}, W_{i',j'})$ for all j, j'; or all edges of $\tilde{T}$ between V_i and $V_{i'}$ point from $V_{i'}$ to V_i, and $\eta \leq d(W_{i',j'}, W_{i,j})$ for all j, j'.*

Using Lemma 7.4 (or actually a tournament version of Corollary 2.8, with the aid of a tournament version of Corollary 2.6), it is not hard to formulate a tournament version of Corollary 3.2 about the existence in T of many model tournaments for a tournament $\tilde{T}$ which differs from T in less than $\epsilon\binom{n}{2}$ places. Tournament $\mathcal{F}$-colorability (and so first order tournament properties of type "$\exists\forall$") can then be proven to be testable by using the above with a lemma that states that an appropriate model for a tournament $\tilde{T}$ can be $\mathcal{F}$-colorable only if $\tilde{T}$ is $\mathcal{F}$-colorable. As these proofs are nearly identical to the proofs in [2] about testability of first order graph properties of type "$\exists\forall$", and similar methods are used here in the previous sections, they are left to the reader.

8 A non-testable first order tournament property

This section presents a proof of the existence of a non-testable first order tournament property of type "$\forall\exists$"; as in [2] it is based on the inability to efficiently test for isomorphisms, but the formulation of the first order tournament property itself requires a modified approach.

Given a (one to one and onto) ordering $\sigma : \{1, \ldots, t\} \to V$ of the vertex set of a tournament T, we call a pair of vertices u, v of T a *forward pair* if either u points to v in T and $\sigma^{-1}(u) < \sigma^{-1}(v)$ (i.e. u appears before v in σ), or v points to u and $\sigma^{-1}(v) < \sigma^{-1}(u)$. A *forward triple* is a triple of vertices u, v, w such that each of the three pairs uv, vw, uw is a forward pair according to σ.

There exist tournaments with the property that there exist a forward triple for every possible ordering of their vertices. We let T_0 denote a tournament which satisfies this with the minimum number of vertices, which we denote by t_0. It is used in defining a property which embodies the notion of tournament isomorphism and is equivalent to a first order property.

Definition 8 *A tournament T is said to satisfy* Property I *if its vertex set is the disjoint union of three sets V_{-1}, V_0, V_1 satisfying the following.*

- *$|V_0| = t_0$ and V_0 spans a copy of T_0 in T.*
- *$|V_{-1}| = |V_1|$, and the subtournaments induced by V_{-1} and V_1 are isomorphic and contain no copy of T_0.*
- *For every $-1 \leq i < i' \leq 1$, every vertex of V_i points to every vertex of $V_{i'}$.*

Two tournaments T_1 and T_2 with n vertices each are called *ϵ-apart* if no tournament differing from T_1 in no more than $\epsilon\binom{n}{2}$ places is isomorphic to T_2. For the proof of the non-testability of the above property, we use the following simple proposition, which is very similar to an analogue proposition about graphs from [2]. Because of the similarity of its proof we give here only a sketch.

Proposition 8.1 *There exists a constant $\epsilon_{8.1}$ such that for every D there exist two tournaments $H = H_{8.1}(D)$ and $H' = H'_{8.1}(D)$, both with $r = r_{8.1}(D)$ vertices, satisfying the following.*

- *H and H' do not contain an induced copy of T_0.*
- *H and H' are $\epsilon_{8.1}$-apart.*
- *H and H' have the property that every vertex has in-degree and out-degree at least $\frac{1}{6}r$; and for every set of vertices with size between $\frac{1}{6}r$ and $\frac{1}{2}r$ there are at least $\frac{1}{72}r^2$ incoming edges and $\frac{1}{72}r^2$ outgoing edges.*

- *Every tournament with D vertices appears the same number of times in H and H'.*

Proof sketch: For every t we consider all tournaments defined on the vertex set $u_1, \ldots, u_{2t}$ such that for every $1 \leq i < i' \leq t$ and $t+1 \leq i < i' \leq 2t$ the vertex $u_{i'}$ points to u_i; we set $r = 2t$. There are $2^{t^2} = 2^{r^2/4}$ such tournaments with the given vertex labeling, and none of them contains a copy of T_0, because with the given vertex order there can be no forward triple. It is not hard to see that more than half of these tournaments (for a large enough r) satisfy the third condition in the formulation of the proposition, from standard bounds on binomial distribution (see e.g. [3], Appendix A).

There are less than $E = 2^{D^2/2}$ possible tournaments with D vertices, and every one can appear less than r^D times in a tournament with r vertices, so there are at least $\frac{1}{2}r^{-DE}2^{r^2/4}$ tournaments that satisfy the first and the third conditions in the formulation in the proposition, and in addition every two of them satisfy the last condition in the formulation above. For a given tournament T with r vertices and a given ϵ, all but less than $\binom{r^2/2}{\epsilon^2 r^2/2}2^{\epsilon^2 r^2/2} < (e\epsilon^{-2})^{\epsilon^2 r^2/2}2^{\epsilon^2 r^2/2}$ of the tournaments with the same (labeled) vertex set are ϵ-apart. Thus a proper choice of $\epsilon_{8.1}$ (independent of D) will ensure for a large enough r (depending on D) the existence of T_1 and T_2 satisfying all the above conditions. $\square$

Corollary 8.2 *Property I is not testable.*

Proof: We show that there exists no ϵ-test for I where $\epsilon = \frac{1}{9}\min\{\epsilon_{8.1}, \frac{2}{72}\}$. Assuming that there exists such a test, we may assume (see the introduction) that it queries about all pairs from a uniformly random subset of size D chosen from the vertex set of the input graph, and gives output according to the resulting induced subgraph.

Assuming that $r_{8.1}(D) \geq t_0$ (the proof of Proposition 8.1 allows it) we construct T and T' as follows. T consists of a copy of $H_{8.1}(D)$ which points to a (disjoint) copy of T_0, and both of them point to another, disjoint, copy of $H_{8.1}(D)$. T' consists of a copy of $H_{8.1}(D)$ which points to a copy of T_0, and both of them point to a copy of $H'_{8.1}(D)$. It is clear that the above test will have the same probability for accepting T and for accepting T' (both have the same appearance counts for subtournaments of size D). This is a contradiction, however, since T satisfies I while T' is ϵ-far from satisfying I. $\square$

Proposition 8.3 *There exists a property of tournaments of type "$\forall\exists$" which is indistinguishable from I, and thus is not testable.*

Proof: We define our property as the conjunction of the following statements.

- Every vertex is either contained in a copy of T_0, or points to t_0 vertices which span a copy of T_0, or is pointed to by t_0 vertices which span a copy of T_0.
- No two distinct copies of T_0 exist.
- For every vertex v which points to the copy of T_0, there exists exactly one vertex w which points to v such that the copy of T_0 points to w (this has a "$\forall\exists$" formulation). Similarly, for every w to which the copy of T_0 points, there exists exactly one vertex v which points to the copy of T_0 such that w points to v.
- If v_1 and v_2 point to the copy of T_0, while it points to w_1 and w_2, and w_1 points to v_1, and w_2 points to v_2, then v_1 points to v_2 if and only if w_1 points to w_2.

Suppose that T is a tournament with n vertices satisfying the above. Define by V_0 the vertex set of the copy of T_0, by V_{-1} the set of vertices pointing to V_0, and by V_1 the set of vertices to which V_0 points (the union of these sets is the set of vertices of T). The third condition above implies that the set of edges directed from V_1 to V_{-1} forms a perfect matching between these two sets, and the fourth condition above implies that this matching is an isomorphism between the subtournament of T induced by V_1 and the subtournament induced by V_{-1}. By reversing the direction of these $\frac{1}{2}(n - t_0)$ edges, one arrives at a tournament $\tilde{T}$ which satisfies I, where V_{-1}, V_0, V_1 here play the role of the corresponding sets appearing in the definition of I.

Suppose now that T is a tournament satisfying I, that V_{-1}, V_0, V_1 is a partition of its vertex set as in the definition of Property I, and that $f : V_{-1} \to V_1$ is an isomorphism between the subtournament induced by V_{-1} and the subtournament induced by V_1. Define $\tilde{T}$ to be the tournament which results from T by reversing the direction of $u, f(u)$ for every $u \in T_{-1}$. If it is proven at this point that the only copy of T_0 in $\tilde{T}$ is the one induced by V_0, the other conditions of the property defined above easily follow.

Suppose that A is a vertex set which spans a copy of T_0, which is distinct from V_0. By the definition of I it is not contained in V_{-1} or V_1. To show a contradiction (and thus complete the proof of the proposition) we construct an ordering of A which does not contain a forward triple. We define $A_i = A \cap V_i$ for $i \in \{-1, 0, 1\}$. By the discussion above $|A_i| < t_0$ for every i, so A_i can be ordered so the subtournament of $\tilde{T}$ induced by A_i contains no forward triple. We merge the orderings of A_{-1}, A_0 and A_1 into an ordering of A by placing all vertices of A_1 before all vertices of A_0, and all these vertices before all vertices of A_{-1}. It is not hard to see now that there is no forward triple in this ordering (the only possible forward pairs which are not contained in some A_i correspond to vertex disjoint edges from A_1 to A_{-1}). $\square$

9 Concluding remarks

Testability for other combinatorial objects

Using appropriate versions of the Regularity Lemma, testability of certain properties can be proven for other combinatorial structures related to graphs. For example, testability of first order properties of type "$\exists\forall$" can be proven in the context of digraphs and in the context of g-charts for a fixed $g > 1$ as well, using arguments similar to those found in [2] and in Section 7 here (note that in both contexts existence of non-testable properties of type "$\forall\exists$" follows directly from the graph case in [2]). Testability of other, more generalized, coloring properties for these structures can be proven as well using methods analogous to those of Section 4 and Section 6.

The definition of the model structures (as in model graphs) has to be modified in each case; as an example, consider the following definition for charts.

Definition 9 *Given a g-chart $\tilde{C}$ and a partition $\mathcal{A} = \{V_i | 1 \le i \le k\}$ of its vertices, a chart M is an* (h, s, t)**-based** (η, ϵ)**-model for** $\tilde{C}$ **over** $\mathcal{A}$ *if there is a partition $\mathcal{M} = \{W_{i,j} | 1 \le i \le k, 1 \le j \le h\}$ of its vertex set satisfying the following.*

- *$s \le |W_{i,j}| \le t$ for all $1 \le i \le k$ and $1 \le j \le h$.*
- *If $h > 1$, then for every fixed $1 \le i \le k$, all pairs $W_{i,j}, W_{i,j'}$ for $1 \le j < j' \le h$ are ϵ-regular. Moreover, if $h > 1$, then there exists $1 \le l \le g$ such that the l-densities of these pairs are at least $\frac{1}{g}$, and the restriction of the pair function of $\tilde{C}$ to $[V_i]^2$ is the constant function l.*
- *For every $1 \le i < i' \le k$ and $1 \le j, j' \le h$, the pair $W_{i,j}, W_{i',j'}$ is an ϵ-regular pair. Moreover, for every fixed $1 \le i < i' \le k$ and $1 \le l \le g$, if for any $v \in V_i$ and $v' \in V_{i'}$ the pair function of $\tilde{C}$ has value l on v, v', then the l-densities of all the pairs $W_{i,j}, W_{i',j'}$ (where j and j' vary over $\{1, \ldots, h\}$) are at least η.*

It would be interesting to investigate the testability of first order properties of type "$\exists\forall$" of vertex ordered graphs; properties which use the order relation over vertices but not the "sequel" function seem more accessible in that respect.

The situation is also far from clear in the context of first order hypergraphs properties, since an appropriate hypergraph version of the Regularity Lemma is not known at the present. This presents yet another motivation for finding such a version.

A little more about vertex coloring with restrictions

Section 4 proves testability of a rather restrictive notion of vertex coloring with restrictions. With a little more work, testability of being $(\alpha, \mathcal{F})$-colorable for any $\alpha = (\alpha_1, \ldots, \alpha_c)$ satisfying a given set of weak linear inequalities, can be proven. For certain strong inequalities, such as "$\alpha_1 > 0$", a reduction to another colorability problem (with a larger set of colors) can be performed, in a similar manner to the reduction in [2] of first order properties of type "$\exists\forall$" to colorability properties. It seems that a rather wide notion of colorability with restrictions can be proven to be testable.

Coloring pairs with restrictions and beyond

Even simple properties, such as the property of a graph G with n vertices having at least $\alpha\binom{n}{2}$ edges (for a fixed α), cannot be described in terms of coloring notions without explicitly introducing restrictions on the number of appearances of each color. It is not surprising that coloring notions by themselves do not capture the essence of all testable graph properties, because many testable properties do not have one-sided tests, as proven in [7], while colorability notions seem to have such tests, as is the case with $\mathcal{F}$-pair-colorability. Note however that the tests constructed for (vertex) $(\alpha, \mathcal{F})$-colorability are not, and cannot be, one-sided.

If we define $(\alpha, \mathcal{F})$-pair-colorings and define appropriately a notion of pair-colorability with restrictions, all graph properties proven in [7] to be testable become instances of this definition. It seems that some approximation of the testability of $(\alpha, \mathcal{F})$-pair-colorability may be provable. However, such a proof using (some extension of) the current methods would be very involved.

Since the proof in [7] gives much better bounds on the dependence of the number of queries an ϵ-test makes as a function of ϵ, a proof of the testability of pair-colorability with restrictions is significant only if it brings one closer to a full characterization of all testable graph properties. However, it is not clear whether this notion even captures the simple property of a graph G with n vertices containing at least $\alpha\binom{n}{3}$ distinct triangles (for a fixed α), which seems to be testable. In fact, even an "upper bound" proof that all testable properties are expressible by some notion of $\mathcal{F}$-(l-tuple)-colorability with restrictions would be welcome, despite the fact that testability of all such properties is currently far from certain.

A little more about non-testable properties

The non-testable first order tournament property here and the non-testable graph property described in [2] are based on the idea that existence of a global isomorphism is hard to test locally. Specificly, the property I' stating that a graph G with n vertices consists of two disjoint copies of the same graph is essentially proven in [2] to be non-testable even if $o(\sqrt{\log n})$ queries are allowed.

It is provable that also querying about all pairs of a randomly chosen set of $o(\sqrt{n})$ vertices is not enough. It would be interesting to find out the size of the random set of vertices which allows a construction of a test of I' in this extended sense, and also how many queries it takes to test for I' by an adaptive algorithm, whose queries are not restricted to the pairs of a pre-chosen random vertex set.

Finally, It would be interesting to find natural non-testable graph properties other than those based on graph isomorphism. It would also be interesting to find out more about the expressive power of all first order properties in the context of property testing and indistinguishability, e.g. which "natural" graph properties are not indistinguishable from any first order property.

References

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the Regularity Lemma, *Proceedings of the* 33^{rd} *IEEE FOCS at Pittsburgh* (1992), 473–481. Also: *Journal of Algorithms* 16 (1994), 80–109.

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Proceedings of the* 40^{th} *IEEE FOCS* (1999), 656–666.

[3] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley (1992).

[4] B. Bollobás, *Extremal Graph Theory*, Academic Press, New York (1978).

[5] A. Czygrinow, S. Poljak and V. Rödl, Constructive quasi-Ramsey numbers and tournament ranking, *SIAM Journal on Discrete Mathematics* 12 (1999), 48–63.

[6] A. M. Frieze and R. Kannan, The Regularity Lemma and approximation schemes for dense problems, *Proceedings of the* 37^{th} *Annual IEEE FOCS* (1996), 12–20.

[7] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *Proceedings of the* 37^{th} *Annual IEEE FOCS* (1996), 339–348. Also: Journal of the ACM, to appear.

[8] W. T. Gowers, Lower bounds of tower type for Szemerédi's Uniformity Lemma, *Geometric and Functional Analysis* 7 (1997), 322–337.

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

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

[11] R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM Journal of Computing* 25 (1996), 252–271.

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

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'