Testing of Matrix Properties

Eldar Fischer[*] Ilan Newman[†]

May 7, 2001

Abstract

Combinatorial property testing, initiated by Rubinfeld and Sudan [15] and formally defined by Goldreich, Goldwasser and Ron in [12], deals with the following relaxation of decision problems: Given a fixed property P and an input f, distinguish between the case that f satisfies P, and the case that no input that differs from f in less than some fixed fraction of the places satisfies P. An (ϵ, q)-test for P is a randomized algorithm that queries at most q places of an input x and distinguishes with probability $2/3$ between the case that f has the property and the case that at least an ϵ-fraction of the places of f need to be changed in order for it to have the property.

Here we concentrate on labeled, d-dimensional grids, where the grid is viewed as a partially ordered set (poset) in the standard way (i.e as a product order of total orders). The main result here presents an $(\epsilon, poly(1/\epsilon))$-test for every property of 0/1 labeled, d-dimensional grids that is characterized by a finite collection of forbidden induced posets. Such properties include the 'monotonicity' property studied in [7, 6], other more complicated forbidden chain patterns, and general forbidden poset patterns. We also present a (less efficient) test for such properties of labeled grids with larger fixed size alphabets. All the above tests have in addition a 1-sided error probability. Another result is a test for a collection of bipartite graph properties which uses less queries than the previously known algorithms for some of them.

Both collections above are variants of properties that are defined by certain first order formulae with no quantifier alternation over the syntax containing the grid order relations (and some additional relations for the bipartite graph properties). We also show that with one quantifier alternation, a certain property can be defined, for which no test with query complexity of $O(n^{1/10})$ (for a small enough fixed ϵ) exists. The above results identify new classes of properties that are defined by means of restricted logics, and that are efficiently testable. They also lay out a platform that bridges some previous results.

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

[†]Department of Computer Science, University of Haifa, Haifa, Israel, and NEC Research Institute. E-mail: ilan@cs.haifa.ac.il, ilan@research.nj.nec.com

1 Introduction

Combinatorial property testing deals with the following relaxation of decision problems: Given a fixed property and an input f, one wants to decide whether f has the property or is 'far' from having the property. The general notion of property testing was first formulated by Rubinfeld and Sudan [15], 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 [12]. A property in this respect, is a collection of Boolean functions from a set (usually with some structure) to $\{0,1\}$[1]. Being far is measured by the *hamming distance*, namely, in how many places should the function be changed so as to have the property. An input function here is identified with its table, namely its $0/1$ value for each of the points of the domain. A property is said to be (ϵ, q)-testable if there is a randomized algorithm that for every input function f queries the values of f on on at most q chosen points of the domain, and with probability $\frac{2}{3}$ distinguishes between the case that f has the property and the case that f is ϵn-far from having the property, where n is the size of the domain. When a property P is (ϵ, q)-testable with $q = q(\epsilon)$ (i.e. q is a function of ϵ only, and is independent of n) then we say that P is ϵ-testable; we say that P is testable if it is ϵ-testable for every $\epsilon > 0$.

Property testing has recently become quite an active research area, see e.g. [12, 13, 6, 4, 1, 2, 14, 9] for an incomplete list. Apart from its theoretical appeal, and the many questions it involves, it emerges in the context of PAC learning, program checking [10, 5, 15], probabilistically checkable proofs [3] and approximation algorithms [12]. The advantage of the 'property testing' relaxation is that many properties have a randomized test that reads a very short piece of the input and also runs very fast (in sublinear time).

One of the main tasks that emerged in the field, following [12], is to identify natural collections of properties that are efficiently testable (in terms of the number of queries). Goldreich et al. [12] studied some classes of properties (mainly graph properties) and identified many properties that are testable. Alon et al. [2] considered properties of functions $f : \{1, \ldots, n\} \longrightarrow \{0,1\}$, namely, where each function is a binary string of length n. They suggested that properties that are defined by restricted logics might be testable. They proved that every regular language is testable, which is equivalent to saying that every property that is expressed by a certain second order monadic logic over ordered sequences is testable. Additional work in this line was done by [14], generalizing the above, and by [1, 9] on graph properties. In [8, 7, 6], the specific property of 'monotonicity' is studied.

Here we make another step in the direction established above: We present a logical model (including several variants) such that all properties that can be expressed in it are testable. Our structure is the d-dimensional grid $\{1, \ldots, n\}^d$, equipped with the natural product order. The language we use is that of first order formulae with the order relation. Our main positive result is that for every fixed d, every such formula using no quantifier alternation (i.e. using only one quantifier on a fixed sequence of points) is testable. Such properties have another equivalent formulation: Every such property is characterized by a finite collection of forbidden induced sub-posets (and the inverse also holds). Such properties include 'monotonicity' (stating that there are no two points x, y $\;x \leq y$ that are labeled $1, 0$ respectively). The model also includes more

[1]Sometimes, a larger range is considered.

complicated properties of chains, e.g: "all 2-dim, 0/1-labeled grids that when going towards the north-east direction in any possible way, contain no sequence of 4 points A,B,C,D labeled 0,1,0,1". More complicated examples are, e.g: "All 2-dim, 0/1-labeled grids that contain no four points A,B,C,D labeled by, say, 1,1,0,1 respectively and such that A is south-west (SW) of all, D is NE (north-east) of all, and neither B is SW of C, nor C is SW of B". We present an $(\epsilon, poly(1/\epsilon))$ 1-sided error test for such properties for every fixed dimension d. We also consider matrices over an alphabet (i.e. set of possible labels) which is not 0/1, but any fixed finite set. For these we obtain a less efficient 1-sided test, which makes $exp(exp(poly(1/\epsilon)))$ many queries.

Another variant of this model dealing with forbidden matrices rather then posets leads to the following result: We show that every bipartite graph property that can be characterized by a finite collection of forbidden induced subgraphs is efficiently testable. The query complexity that we achieve for this case is bounded by $exp(exp(poly(1/\epsilon)))$ (for a 2-sided error test; for 1-sided testing, the query complexity is triply exponential in $1/\epsilon$, and will be described only in the journal version). Such bipartite graph properties were known to be testable by [1], but the previously known query complexity was $tower(tower(poly(1/\epsilon)))$, obtained by the bound on general graph properties using the regularity lemma. Our result here, first breaks the bound of using the regularity lemma for some collection of graph properties based on families of forbidden induced subgraphs[2]. Our main technical tool is a conditional version of the Regularity Lemma that may be interesting in its own right.

On the negative side, we show that there exists a poset (matrix) property expressed by using only one "$\forall\exists$" quantifier alternation which is not ϵ-testable for some ϵ. This puts a bound on the complexity of the relevant logical model that still guarantees testability. We also discuss several other variants of this model, some of which are efficiently testable, while in some, efficient testability remains open.

To put the results in perspective, we note that our model for the one dimensional case includes only regular properties, hence the result for this case is part of the results of [2]. A special instance of such properties (for all d) is the extensively studied property of 'monotonicity', namely the property: $\forall x_1, x_2\, ((x_1 \leq x_2) \rightarrow (M(x_2) \vee \neg M(x_1)))$, [7, 6] ('$M(x)$' means 'the value of the input matrix at x is 1'). On the other hand, one variant of the model (for $d = 2$) is tightly connected to bipartite graph properties. Hence, the model we suggest, apart from identifying a large class of efficiently testable properties, generalizes some previous results and bridges some others, allowing for a broader perspective.

The rest of the paper is organized as follows: In Section 2 we define the basic model and prove some basic preliminaries, in Section 3 we construct as a warm-up a simple to prove, 2-sided error, ϵ-test, and then the 1-sided error test for the basic model of properties (one quantifier). In Section 4 we construct the test for properties of matrices over any fixed, non-binary, alphabet. In Section 5 we construct a property that can be expressed as a formula with one quantifier alternation and which is not testable by starting with a property that belongs to a more general model of matrix properties and deriving our property from it. In Section 6 we prove that bipartite graph properties that are characterized by a finite collection of forbidden induced subgraphs are efficiently testable.

[2]Bipartite graph properties that are characterized by forbidden subgraphs, rather then forbidden induced subgraphs, are trivially testable, see the end of Section 6. Unlike the case for properties defined by forbidden induced subgraphs, testing for partition-based graph properties, such as proper k-colorability, has polynomial bounds by [12].

Finally, in Section 7 we discuss variants of the poset-models, and pose some open problems and future directions.

2 Preliminaries and Notations

In the following, we omit all floor and ceiling signs whenever the implicit assumption that a quantity is an integer makes no essential difference. We make no attempt to optimize the coefficients involved, just the function types (e.g. polynomial versus exponential).

Let $[n] = \{0, \ldots, n-1\}$ with the natural order "$\leq$". The n-length d-dimensional grid is the poset (partially ordered set) $G(n,d) = [n]^d$ with the product order, namely $(\alpha_1, \ldots, \alpha_d) \leq (\beta_1, \ldots, \beta_d)$ if $\alpha_i \leq \beta_i$ for all $i = 1, \ldots, d$.

For $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $\beta = (\beta_1, \ldots, \beta_d)$, if $\alpha_j = \beta_j$ for some $1 \leq j \leq d$, then we say that α and β *share a coordinate.*

A Boolean function $f : G(n,d) \longrightarrow \{0,1\}$ is identified with $G(n,d)$ with 0/1-labels on its points. Such a function will be called a 0/1 (n,d)-matrix, or just an (n,d)-matrix. Hence a property of functions on the structure $G(n,d)$ is just a set of (n,d)-matrices. For two (n,d)-matrices M and R their distance is $dist(M,R) = |\{x| \quad M(x) \neq R(x)\}|$. For a matrix M and a property $\mathcal{P}$ we define $dist(M,\mathcal{P}) = min_{R\in\mathcal{P}} dist(M,R)$.

In the basic logical model that we study, the variables range over $G(n,d)$. The syntax includes the poset (binary) relation and the function unary relation $M(\cdot)$ ("being labeled 1"). Given a fixed set of variables, $x_1, \ldots, x_k$, a Boolean formula $\phi(x_1, \ldots, x_k)$ using the above relations specifies an allowed set of 0/1-labeled posets that are the truth assignment to it. The basic model of properties of (n,d)-matrices contains these properties that can be expressed as $\forall x_1, \ldots, x_k \phi(x_1, \ldots, x_k)$ where ϕ is a formula as above and k is a fixed constant (independent of n). For example, the well studied property 'monotonicity' is such: $\forall x_1, x_2\, ((x_1 \leq x_2) \rightarrow (M(x_2) \vee \neg M(x_1)))$. We call this model the $\forall$-poset model. Similarly the $\exists$-poset, the $\forall\exists$-poset models, etc. are defined.

Clearly if a property is expressible as an $\exists\phi(x_1, \ldots, x_k)$ formula then either every matrix is k-close to it, or the property is empty. Hence, such properties are trivially ϵ-testable. Things are starting to be interesting for $\forall$ properties. Our main result is that any such property is testable. We then show that there exists a $\forall\exists$-poset property that is not testable even for dimension $d = 2$.

We first note that testing a $\forall$-poset property is equivalent to testing that there is no forbidden fixed substructure. To make this explicit we need some definitions and observations: We say that a poset P of size k is a *subgrid poset of dimension d* if there is a set of k points in $G(n,d)$ (for some n) on which the induced order is isomorphic to P. We say that P is a 0/1-labeled poset if every point of P is labeled by 0 or 1. An (n,d)-matrix M contains a labeled poset P of size k if there are k points in M on which the order relation is isomorphic to P and the point labels are consistent with the labeling of P.

Definition 2.1 *Let F be a set of 0/1-labeled posets. The following property of (n,d) matrices is defined: $\mathcal{M}_F = \{M|\ M$ does not contain any member of F as an induced labeled poset $\}$.*

Observation 2.2 *Let $\mathcal{P}$ be a $\forall x_1, \ldots, x_k \phi(x_1, \ldots, x_k)$ type property of (n, d)-matrices. Then there is a set F of labeled k size posets, $|F| \le 2^{k^2}$, such that $\mathcal{P} = \mathcal{M}_F$.*

Proof: We can rewrite the negation of $\mathcal{P}$ as $\exists x_1, \ldots, x_k \neg\phi(x_1, \ldots, x_k)$. In turn, we can write an equivalent DNF formula $\vee m_i(x_1, \ldots, x_k)$ for $\neg\phi$, where each $m_i(x_1, \ldots, x_k)$ represents a labeled poset F_i on at most k elements. Hence, a matrix satisfies $\mathcal{P}$ if and only if it has no F_i as a labeled-subgrid-poset. ■

We also note that every property $\mathcal{M}_F$ is a $\forall x_1, \ldots, x_k \phi(x_1, \ldots, x_k)$ type property, therefore the above reduction is an equivalence. Along the sequel, our strategy for testing M for the property $\mathcal{M}_F$ will be to query some points in M and try to locate a member of F as an induced labeled poset within the queried points.

3 Algorithms for testing poset properties

We present here a 1-sided error test for $\mathcal{M}_F$ whose query complexity is polynomial in ϵ^{-1}. We start with some preliminaries, presented with a simple to prove, 2-sided error, algorithm that illustrates some of the basic ideas used throughout the following.

3.1 Preliminaries and a simple 2-sided error test

Let F be a finite collection of k-size posets. We present here a basic approach to matrix property testing, along with a 2-sided error ϵ-test for $\mathcal{M}_F$, whose number of queries is polynomial in ϵ^{-1} for any fixed dimension d.

Proposition 3.1 *For every fixed $\forall$ poset property of matrices, there exists a 2-sided ϵ-test which makes a number of queries which is polynomial in ϵ^{-1} and is independent of the size of the input.*

The proof of the proposition is implied by the algorithm below and the following Lemmata.

Let M be an (n, d)-matrix which we want to ϵ-test for $\mathcal{M}_F$. Let $m = (\frac{2^{k+1}k}{\epsilon^{k+1}})^{d-1}+1$. We divide M into m^d blocks of size $(n/m)^d$, by dividing $[n]$ into m, equal-size, intervals and taking Cartesian products. We query $\frac{8d \ln m}{\epsilon}$ uniformly random queries independently in each block of M. We tag each block as being 1, 0 or X according to the queries made: If all points that are queried in a block are labeled by '1' we tag it as '1', similarly if all points are labeled by '0' the block is taged by '0'. Otherwise, we tag it as X. Hence we get a corresponding (m, d)-matrix, M_B, in which each entry represents a block of M and is labeled by 0,1 or X.

There are two major cases: The first one is the case where there is at least an $(\epsilon/2)$-fraction of the blocks that are tagged X. In this case we answer 'No'. It will be proven below that if this happens then there exists an actual member of F within the entries of M that were queried.

The second case is when there are less than a fraction of $\epsilon/2$ of the blocks that are tagged X. In this case our intention is to check whether there is a matrix that is consistent with our knowledge of M as represented by M_B, and has the property. If we find such a matrix we answer 'Yes', if not we answer 'No'. Formally, Let M_Q be the following (n,d) matrix: For each 0-block of M, all corresponding entries of M_Q are '0', for each 1-block of M all corresponding entries in M_Q are '1'. The entries of M_Q that correspond to an X-block of M remain undefined. Now, each possibility of assigning 0/1 values to the undefined entries of M_Q and each possible choice of flipping the values in at most an $\epsilon/4$ fraction of the entries in every other block results in a 0/1-labeled matrix; we denote the set of all such matrices by $\mathcal{M}_{Q,\epsilon}$. We check if any of the members of $\mathcal{M}_{Q,\epsilon}$ has the property. If there is such a member, the algorithm answers 'Yes'. Otherwise, if every member of $\mathcal{M}_{Q,\epsilon}$ contains a member of F, the answer is 'No'. Note, this last phase of the algorithm involves no additional queries and is just a computation phase.

Clearly the query complexity of this algorithm is $poly(1/\epsilon)$ and is independent of n. We now show that it is correct with high probability.

We first show that if the algorithm answers 'No' due to the fact that at least an $\epsilon/2$ fraction of the blocks are X-tagged, then it is correct with probability 1:

Claim 3.2 *If the fraction of* X*-blocks is at least $\epsilon/2$, then the queried locations in the* X*-blocks already contain a counter-example to the property.*

Proof: Let M_B be the (m,d)-matrix defined above. We use the following d-dimensional Zarankiewicz (see [17]) type theorem that locates a (k,d)-submatrix of a given label inside any large enough matrix with enough entries of this label.

Lemma 3.3 ([11]) *For $\delta < 1$ let M be an (m,d)-matrix in which at least δm^d of the entries are marked by 'X'. If $m > (\frac{k}{\delta^{k+1}})^{d-1}$ then there is a (k,d) submatrix all of whose entries are 'X'.*

Now, M_B satisfies Lemma 3.3 with $\delta = \frac{\epsilon}{2}$, thus there is a (k,d)-submatrix, W, of M_B which is all X. Our intent is to look back at the blocks of M that correspond to the entries of W. Each such block being tagged by X contains both a '0' and a '1' entry. Hence we want to argue that any labeled poset with k elements may be found within these blocks by choosing the right labeled entry. The only difficulty is that for two entries that share a coordinate in W and are comparable, the corresponding sampled entries in M might not be comparable. For this we use the following:

Lemma 3.4 *Any d-dimensional grid poset P with $|P| = k$ can be embedded in $G(k,d)$ with no two points sharing a coordinate.*

Proof: It is enough to prove that P can be embedded in any d-dimensional grid with no two points sharing a coordinate, since we can then take the minimum subgrid containing it and it will clearly be a (k,d) grid.

Suppose that $f : P \to \{0,\ldots,m-1\}^d$ is any embedding of P into $G(m,d)$ for some m. Then we can think of f as a sequence of d functions $f_i : P \to \{0,\ldots,m-1\}$, $i = 1,\ldots,d$, so that $f_i(p)$

is the i-th coordinate of the point p in $G(m,d)$. We define a new embedding $f' : P \to G(m^{d+1}, d)$ by the formula (here $f(p)$ is taken as a d-dimensional row vector and $f(p')$ is defined as a linear combination of such vectors).

$$f'(p) = m^d \cdot f(p) + \sum_{i=1}^{d} f_i(p) m^{i-1} \cdot (1, \ldots, 1)$$

It is not hard to see that this is an embedding of the poset too, and that no two members share a coordinate (as the location of every member modulo m^d is unique in all coordinates). ■

To end the proof of Claim 3.2, let P be an arbitrary member of F. By Lemma 3.4, P can be embedded in W with no two points sharing a coordinate. For each point of W that corresponds to a point of P in this embedding, let us choose an entry in the corresponding block of M that has the same value (such a point exists as the points of W correspond to X-tagged blocks). These entries are an embedding of P in M. ■

We now claim that with high probability, the tagging of the blocks of M essentially represents the true situation.

Definition 3.5 *For a set Q of queries, we denote by G the following event: "Every 1-tagged block contains at most an $\epsilon/4$ fraction of '0' entries, and every 0-tagged block contains at most an $\epsilon/4$ fraction of '1' entries".*

Claim 3.6 *With probability at least $1 - 1/m^d$ the event G happens.* ■

To following now ends the proof of the correctness of the algorithm:

Claim 3.7 *If there are less than an $\epsilon/2$ fraction of the blocks that are tagged by* X *, then the algorithm errs with probability bounded by $1/m^d$.*

Proof: Assume that there are less than an $\epsilon/2$ fraction of the blocks that are tagged by X. Also, let us assume that the event G has happened (as asserted by Claim 3.6, this happens with probability at least $1 - 1/m^d$). If there exists a member of $\mathcal{M}_{Q,\epsilon}$ that has the property then we claim that M is close to having the property. Indeed, assuming that G happened, M is clearly at most $(\epsilon n^d/4)$-far from M_Q (disregarding the entries in the X-tagged blocks), while M_Q is at most $(3\epsilon n^d/4)$-far from any member of $\mathcal{M}_{Q,\epsilon}$ (counting *all* the entries in the X-tagged blocks), so M is at most (ϵn^d)-far from the member of $\mathcal{M}_{Q,\epsilon}$ having the property. Hence, in this case the algorithm may be wrong only if G did not happen, which occurs with a probability bounded by $1/m^d$.

If, on the other hand, no member of $\mathcal{M}_{Q,\epsilon}$ has the property, then clearly M does not have the property since given that G happened, M is in particular a member of $\mathcal{M}_{Q,\epsilon}$. Hence again the error probability is bounded by the probability that G does not happen. This completes the proof of Claim 3.7 and the correctness of the algorithms. ■

We note that the running time of the algorithm (as opposed to its query complexity) involves checking (or preparing answers in advance to) all possible $\mathcal{M}_{Q,\epsilon}$ resulting from M_Q. Although this seems to be exponential in n, it can be reduced considerably. However, since a better and faster algorithm is presented in the following, we omit the details.

We end up noting that the above is a scheme that is used throughout this paper: We try to obtain an approximation of the input matrix which consists mostly of "simple" blocks. In the above case – blocks which are known to be almost monochromatic (all '0' or all '1').

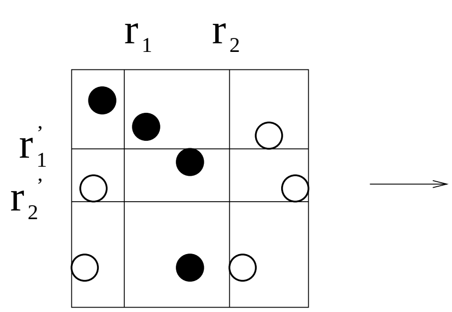
3.2 More preliminaries – pseudomatrices and a Ramsey-like lemma

In the above algorithm use was made of the following fact: Given the (k,d) 'matrix of blocks' which are all tagged X, one could choose any entry (with the right label) from each block to find the counter-example. Such 'loose formations' of entries are formalized below, with the definition of pseudogrids and pseudomatrices.

Definition 3.8 *A* (k,d) pseudogrid *is a subset* $(x_i)_{i\in[k]^d}$ *of the grid* $G(n,d)$*, such that if* i *and* j *do not share a coordinate, then* $x_i \le x_j$ *if and only if* $i \le j$ *(in the ordering of* $G(k,d)$*). Alternatively,* $(x_i)_{i\in[k]^d}$ *is a pseudogrid if and only if there exist* $\{r_{a,l} | 1 \le a \le d, 0 \le l \le k\}$ *(one may assume* $r_{a,0} = 0$ *and* $r_{a,k} = n$*) such that for every* a *and* l *the* a*'th coordinate of* x_l *is at least* r_{a,l_a} *but less than* r_{a,l_a+1}*. Namely, there exists a partition of* $G(n,d)$ *into* k^d *blocks, that is defined by the above set og intervals, such that* $(x_i)_{i\in[k]^d}$ *consists of exactly one member from each block. See Figure 1 for an example of a* $(3,2)$*-pseudo grid.*

A pseudomatrix *in an* (n,d) *matrix* A *consists of a* (k,d) *pseudogrid in* $G(n,d)$ *and its labeling according to* A*. Given a* (k,d) *pseudomatrix in* A *defined by the pseudogrid* $(x_i)_{i\in[k]^d}$*, its* label matrix *is the* (k,d) *matrix* B *(which is not in itself a submatrix of* A*) consisting of the entries of the given pseudomatrix in the corresponding locations, that is, the one satisfying* $B(i) = A(x_i)$ *for all* $i \in [k]^d$*.*

See Figure 1 for an example of a $(3,2)$-pseudo matrix with its label 3×3 matrix.



$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Figure 1: The 9 points in the grid on the left form a $(3,2)$-pseudo grid. Along with their labels (bold dots are 1's, white are 0's) they form a $(3,2)$-pseudo matrix. On the right is the 3×3 label matrix of this pseudo matrix

Pseudomatrices will be used to show the existence of counter-examples within the queried values in 1-sided tests. Their main property is established in the following:

Fact 3.9 *Let A be a (k,d)-labeled pseudo matrix of M with a label matrix W. Assume that W contains a k size labeled poset F such that no two entries share a coordinate. Then M also contains F.*

■

For 1-sided algorithms, in order to guarantee the existence of a counter example we will have to find a 'monochromatic' pseudo matrix in each block. This is done by the following 'Ramsey-type' lemma.

Lemma 3.10 *For every d and k there exists $c = c(d,k)$ such that for any $l \geq cm^c$, every 0/1-labeled (ml,d)-matrix which is partitioned into m^d blocks of size l^d, contains an (mk,d) pseudomatrix, so that its intersection with each of these blocks is a (k,d) pseudomatrix whose entries are all identical.*

In order to prove the above lemma, we first need the following one, which deals with a 1-dimensional context. In the sequel the locations in strings of length l are numbered by $0,\ldots,l-1$.

Lemma 3.11 *For every k there exists $a = a(k)$ such that for any set of m binary strings of length $l \geq a \cdot m^a$, there exist $0 = r_0 < r_1 < \ldots < r_k = l$ such that each string contains a monochromatic substring (i.e. a substring all of whose entries are of identical values) with exactly one entry in the interval $[r_i,\ldots,r_{i+1}-1]$ for every $i = 0,\ldots,k-1$.*

Proof: It is enough to prove this for odd values of k. The proof is by induction. The basis $k=1$ is trivial (it holds even for m strings of length 1).

Given that the theorem was proven for $k-2$, and $a(k-2)$ is known, we now prove it for k and find $a(k)$. We assume that none of the strings is monochromatic, as a monochromatic string will have a corresponding monochromatic substring for any $0 = r_0 < r_1 < \ldots < r_k = l$.

We first set the values of r_1 and r_{k-1} as follows: Let $l' = a(k-2)m^{a(k-2)}$, choose r_1 in the interval $[1,\ldots,l-1-l']$ uniformly at random, and set $r_{k-1} = r_1 + l'$.

For every fixed string s Let L_0 be the location of its first ('leftmost') '0' entry, L_1 the location of its first '1' entry, and R_0 and R_1 the locations of its last '0' and '1' entries respectively. We are interested in the event that none of $\{L_0,L_1,R_0,R_1\}$ fall into the interval $I = [r_1,\ldots,r_{k-1}-1]$. The significance of this event is that if it occurs for a string s, then for any entry v in the interval $[r_1,...,r_{k-1}-1]$ there are two entries of s with an identical label as v, one in the interval $[0,...,r_1-1]$, and one in the interval $[r_{k-1},...,l-1]$.

Since one of L_0 and L_1 is equal to 0 and one of R_0 and R_1 is equal to $l-1$, at most two of the four numbers above may fall into the interval I. Calculating the probability that any of L_0, L_1, R_0 and R_1 is in the interval I, it is clearly bounded by $2\frac{l'}{l-1-l'}$. It is now not hard to find $a = a(k)$

such that this probability is less than m^{-1} if $l \geq a \cdot m^a$. In particular, for such l there exist r_1 and a respective r_{k-1} for which the above event does not happen for any of the m strings. This means that for every string from the set and any of its entries in the interval $[r_1, \ldots, r_{k-1} - 1]$, there exist identical entries in the interval $[0 = r_0, \ldots, r_1 - 1]$ and in the interval $[r_{k-1}, \ldots, r_k - 1 = l - 1]$.

We now consider for each string its substring of length $l' = a(k-2)m^{a(k-2)}$ consisting of the entries in the interval $[r_1, \ldots, r_{k-1} - 1]$. To these we apply the induction hypothesis, and find $r_2, \ldots, r_{k-2}$ so that each string contains a monochromatic substring with one entry in the interval $[r_i, \ldots, r_{i+1} - 1]$ for every $1 \leq i < k-1$. By the above discussion this substring can be extended to a monochromatic substring of the original string by finding identical labeled entries, one in the interval $[r_0, \ldots, r_1 - 1]$ and one in the interval $[r_{k-1}, \ldots, r_k - 1]$. ■

We note here that it is not hard to come up with an example that shows that a string length of $l = \frac{1}{4}m^2$ is insufficient for $k = 4$. We are now ready to prove the main lemma.

Proof of Lemma 3.10: We shall use here the definition of pseudomatrices regarding the existence of a partition given by $\{r_{d',j} | 1 \leq d' \leq d, 0 \leq j \leq mk\}$ such that $(x_i)_{i \in [mk]^d}$ consists of one member from each block. In our case, the partition also has to satisfy $r_{d',k \cdot j} = l \cdot j$ for every $0 \leq j \leq m$ to ensure that the intersection of the (mk, d) pseudomatrix with each of the m^d blocks of A is a (k, d) pseudomatrix.

The proof is by induction on d. We shall actually prove the (seemingly stronger) statement that given p such matrices with $l \geq c(mp)^c$, there exists a corresponding pseudomatrix for each of them, all with the same partition $\{r_{d',j} | 1 \leq a \leq d, 0 \leq j \leq mk\}$ of $[n]^d$. The basis, $d = 1$, follows directly from Lemma 3.11, by cutting each string into m equal length substrings and applying Lemma 3.11 to these $p \cdot m$ new strings.

Given that the lemma is known for the $(d-1)$-th dimensional case, we now show it for the d dimensional case. Suppose that we are given p matrices $A_0, \ldots, A_{p-1}$ over $[ml]^d$ for which we want to find the partition. Note that each matrix can actually be viewed as being a collection of m independent matrices where in the d-th dimension there is just one block. As we require that the partition will be the same for all matrices, these different m 'slices' will fit together to a partition of the original matrix, see Figure 2.

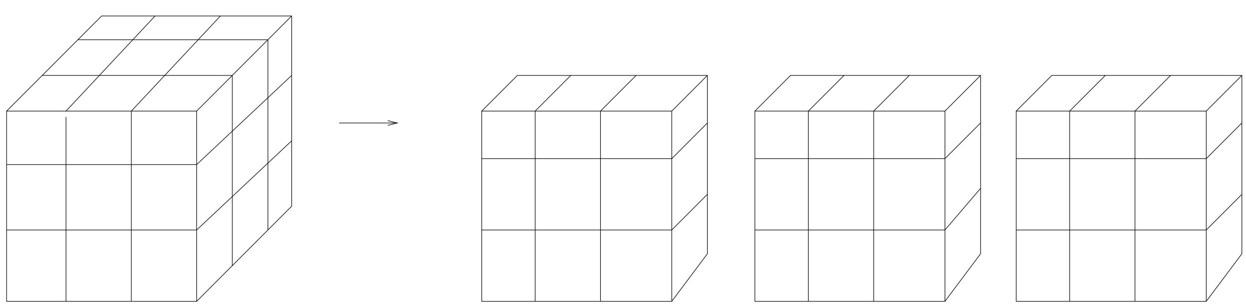


Figure 2: On the left a $(3l,\ 3)$ matrix of 3^3 blocks. It is viewed as 3 matrices, each of 3^2 blocks, by 'slicing' the matrix along the 3rd dimension.

Hence, the top level idea is the following: We look at the p matrices as $m \cdot p$ matrices that in the d dimension all contain just one block, and try to find the partition $\{r_{d,j}\}$ for the d-th dimension. To do this we just look at each l length 'string' along the d-th dimension and find the collection of $r_{d,j}$'s so that each string has a k-size monochromatic substring with exactly one point in each interval $[r_{d,j-1}, ..., r_{d,j} - 1]$. To do this we just use Lemma 3.11 (for suitable l and m). Once this is done, we may replace each l-length string by the single value that is identical to the value of the monochromatic substring that is found. This reduces each of the $m \cdot p$, d-dimensional matrices to a $(d-1)$-dimensional matrix, and as explained, finding a common partition for these matrices will end the proof. However, by the induction hypothesis we may assume that such a partition for the $(d-1)$-dimensional case exists.

Formally this is done as follows: First, for every $0 \leq s < p$, $i = (i_1, ..., i_{d-1}) \in [ml]^{d-1}$ and $0 \leq q' < m$, we consider the string of length l consisting of the entries of A_s in locations $\{(i_1, \ldots, i_{d-1}, q'l + q) | 0 \leq q < l\}$. This is a collection of $pl^{d-1}m^d$ strings of size l corresponding to the $(md)^{d-1}$ lines along the d-dimension in each of the $p \cdot m$ matrices obtained after we have 'sliced' each of the p original matrices into m matrices. However, this collection is of a size too big to apply Lemma 3.11 according to our plan. Hence we just restrict ourselves to a subset of them: For some $l' < l$, we consider only strings i for which every coordinate modulo m is between 0 and $l' - 1$. This will yield a collection of $pl'^{d-1}m^d$ strings. We do this for $l' = c(k, d-1)(pm \cdot m)^{c(k,d-1)}$, so in particular to ensure that $l' \leq l$, we require that $c(k, d-1)(pm \cdot m)^{c(k,d-1)} \leq l$.

We now want for each fixed $0 \leq q' < m$, to find $r_0, \ldots, r_k$ that will partition all $p(l'm)^{d-1}$ strings consisting of the entries of A_s in $\{(i_1, \ldots, i_{d-1}, q'l + q) | 0 \leq q < l\}$ (for any $0 \leq s < p$ and the values for $i_1, \ldots, i_{d-1}$ considered above) according to Lemma 3.11. This requires that $l \geq a(p(l'm)^{d-1})^a$ where a is the constant depending on k from Lemma 3.11.

For the above q' we then define $r_{d,q'k+j} = q'l + r_j$ for $0 \leq j < k$ (remember that $r_0 = 0$ and $r_k = l$) which defines $\{r_{d,j}\}$ of the pseudo matrix partition along the d'th dimension.

For each $0 \leq s < p$ and $i = (i_1, \ldots, i_{d-1})$ we take the common label of the substring found by Lemma 3.11 to be the entry of a new $(ml', d-1)$ matrix $A'_{ms+q'}$, at location $l'\lfloor i/l \rfloor + (i \bmod l)$ (we "delete" in the definition of the new matrices the strings which were not considered when finding $r_0, \ldots, r_k$). On these $(ml', d-1)$ matrices, $A'_0, \ldots, A'_{pm-1}$, we invoke the induction hypothesis to find the values $\{r'_{d',j} | 1 \leq d' \leq d-1, 0 \leq j \leq mk\}$, and the corresponding $(mk, d-1)$ pseudomatrices (note that $r_{d,j}$ were already set earlier). We finalize by setting $r_{d',j} = l\lfloor r'_{d',j}/l' \rfloor + (r'_{d',j} \bmod l')$ for $1 \leq d' \leq d-1$, and noting that in order to find the (mk, d) pseudomatrix in A_s for any $0 \leq s < p$, we can consider the pseudomatrices found in $A'_{ms}, \ldots, A'_{ms+m-1}$, and substitute each of their entries with the corresponding string of length k found earlier in A_s that was used in their construction. The only thing that remains to be done is to ensure that the two conditions that we have set on the relation between l, m and l' are met, which is not hard for an appropriate choice of c. ■

3.3 A 1-sided test

Here we present the 1-sided test for $\mathcal{M}_F$. We essentially follow the 2-sided error algorithm depicted above. The problem is that in the case that there are less than an $\epsilon/2$ fraction of the blocks that are tagged by X, the above algorithm may err in both ways as its correctness is based on the assumption that G happens. To overcome this we will further query more queries in each block to assure us that if the approximated matrix is far from satisfying the property, then a counter-example to the property exists within the points already queried. In turn, this also avoids the necessity of checking all (n, d) matrices approximable by the queried values, as we need only to check the queries themselves. This makes also the running time of the algorithm polynomial in ϵ^{-1}.

Theorem 3.12 *For every fixed $\forall$ poset property of matrices, there exists a 1-sided error ϵ-test which makes poly($\frac{1}{\epsilon}$) many queries, independently of the size of the input.*

We shall use the Ramsey-like lemma presented in the previous section. In order for us to be able to work with it, we need to consider only embeddings in which no two entries share a coordinate, so pseudomatrices can be considered. The following lemma in flavor of Lemma 3.4 grants this; the proof is also very similar to that of Lemma 3.4 and is therefore omitted.

Lemma 3.13 *Let $G(mk, d)$ and $G(ml, d)$ be considered as copies of $G(m, d)$ in which every entry is a (k, d) block and an (l, d) block respectively. Suppose that a k-size poset P embeds into $G(ml, d)$, then there is an embedding of P into $G(mk, d)$ so that the two embeddings map each point of P to the same respective block in $G(ml, d)$ and $G(mk, d)$, the mappings are isomorphic on corresponding blocks (in terms of the poset relation), and moreover no two points share a coordinate in $G(mk, d)$.*

Proof of Theorem 3.12: We present the 1-sided error test. We start exactly as in the algorithm of Subsection 3.1 by dividing M into m^d blocks of equal size, for $m = (\frac{2^{k+1}k}{\epsilon^{k+1}})^{d-1}+1$. We query $\frac{8d \ln m}{\epsilon}$ uniformly random queries independently in each block of M and tag each block as being 1, 0 or X according to the queries made: '1'/ '0' if all values that were queried inside the block are 1/0 respectively, and X otherwise. If there are at least an $\epsilon/2$ fraction of the blocks that are X-tagged then we answer 'No'.

If there are less than an $\epsilon/2$ fraction of the blocks that are X-tagged then we further divide each block into $l^d = (c \cdot m^c)^d$ subblocks (where $c = c(d, k)$ is the constant of Lemma 3.10) and query one arbitrary query in each subblock. We now set the outcome of the test as follows: We answer 'No' if there is a counter example among the queried points and answer 'Yes' otherwise (meaning that M is close to have the property).

Clearly the overall query complexity of this algorithm is polynomial in ϵ^{-1} (for a fixed d and k) and independent of n. We now prove its correctness.

Claim 3.14 *The algorithm is a 1-sided error algorithm with an error probability bounded by $1/m^d$.*

Proof: To show correctness, let us analyze the various cases in which the algorithm may end. If after the first round of queries there are at least an $(\epsilon/2)$-fraction of the blocks that are X-tagged, then, exactly as in the proof of the 2-sided test, a counter example to the property is guaranteed to exist already within the queried locations (with probability 1). Hence in this case, the algorithm answers 'No' and is correct with probability 1.

Assume then that there are less than an $\epsilon/2$ fraction of the blocks that are X-tagged. We then have the second phase of queries in each subblock after which we re-tag each block according to the union of the old and new queries in it. Namely, some old 1-blocks and/or 0-blocks may become X-blocks. Again we may assume that there are less than a fraction of $\epsilon/2$ of the blocks that are X, as otherwise we are back in the first case. Define now the (ml, d)-matrix, M_Q in which every entry corresponds to a subblock of M that is labeled by 0/1 as determined by the value of the query in this subblock. By Lemma 3.10 there is an (mk, d) pseudomatrix W, containing a monochromatic (k, d) pseudomatrix in each of the m^d blocks of M_Q. Note that in the natural correspondence between blocks of W and blocks of M, the label of every 0/1 block of M is identical to the label of the corresponding block of W (on X-blocks of M the label of the corresponding block of W may be either '0' or '1'). We claim that even when we restrict ourselves to the points queried in the subblocks corresponding to the points of W, and have the decision of the algorithm according to whether they contain a counter example, it is correct with very high probability (and is 1-sided). Indeed, if a counter example is found among these points then certainly the algorithm rejects correctly.

Assume now that W does not contain a counter example. We claim that M is ϵn^d-close to have the property (with high probability). Indeed, assuming that the event G happened (as in Definition 3.5), we show how to obtain a matrix, M_W from M that has no counter example by changing at most ϵn^d entries of M: First we change the entries in each 0/1 blocks of M to have the label of their block. As in Section 3.1, we may assume that this will incur a change in at most $\frac{\epsilon}{4} n^d$ of the entries. We then change every entry in an X-block to have the label as in the corresponding intersection with W. As there are at most an $(\epsilon/2)$ fraction of X-blocks, this may incur in at most $\frac{\epsilon}{2}$ additional changed entries. Hence we get a matrix M_W, that is at most $\frac{3\epsilon}{4} n^d$-far from M. We claim that M_W has no counter example. Indeed, assume that M_W contains a counter example P, then by Lemma 3.13 (looking at M_W as an (ml, d)-matrix with $l = n/m$) there is a counter example in an (mk, d)-matrix in which its m^d-size blocks are labeled as the blocks of M_W. Moreover, no two points of this counter example share a coordinate, and so its existence in a label matrix of some pseudomatrix implies its existence in the pseudomatrix itself. Now the label matrix of W is such an (mk, d) matrix, and W in itself is a pseudomatrix contained in M_Q. Hence M_Q contains P with no two points sharing a coordinate. However, each entry of M_Q corresponds to a subblock of the actual input M that contains at least one queried point that is labeled by the same label as the corresponding entry of M_Q. Therefore we conclude that there is a counter example in the queried points of M in the subblocks that correspond to the points of W contrary to our assumption. ■

An additional remark is due here: The algorithm was described as if it is adaptive, however as queries at the second stage do not depend on the answers at the first stage, the algorithm is in fact non-adaptive.

4 Testing of non-binary matrices

In this section we extend the result of the previous section to include forbidden poset properties for matrices which are not 0/1, but have entries from a fixed finite alphabet Σ. Such properties are natural extensions of 0/1-matrix properties.

The main result here is:

Theorem 4.1 *For every fixed $\forall$ poset property of matrices over a fixed finite alphabet, there exists a 1-sided ϵ-test whose number of queries and running time are doubly exponential in poly(ϵ^{-1}), and are independent of the size of the input.*

The proof uses some general ideas from the previous section, and in particular the idea of arriving to a partition into blocks for which most blocks have little 'internal features' (like being nearly monochromatic in the 0/1 case). For this we will need some additional definitions and machinery.

4.1 Homogeneity in partitions and a strengthening of Lemma 3.10

Given a partition of the input matrix into blocks, and a set of queries from each block, unlike in the 0/1 case, it cannot be guaranteed that if more than one type of label was found in many of the blocks then a counter-example to a given property $\mathcal{M}_F$ exists. It could be for example that a 0/1/2 matrix property is defined by an F all of whose members contain the label '2', while the partition of the input matrix into blocks may contain many blocks with both 0's and 1's. Therefore, a notion of a block being monochromatic is replaced with a more general notion of being 'featureless'.

In the new framework, we consider two partitions of a matrix at a time. We consider as before a partition P of M into m^d blocks, and a refinement achieved by repartitioning each block of the first partition into l^d subblocks, thereby obtaining a partition Q of M into $(ml)^d$ blocks. We label each block and each subblock with a subset of the alphabet Σ. In the construction of the tester this subset corresponds to the set of all labels found while querying entries from this block. Note that in particular a subblock is always labeled with a subset of the set used to label the whole block.

We will treat a block of P as featureless if all its subblocks have the same label. In that case we can assume that this block has the worst possible assignment (with respect to satisfying a given property) that makes use only of the members of its label set. However, in some cases it is only possible to find blocks which are close to satisfying this, which motivates the following definition.

Definition 4.2 *A labeled set is called* ϵ-homogeneous *if all but at most an ϵ fraction of its labels are identical. For such a set we call this common label the* $(1-\epsilon)$-dominant label *of the set.*

Given a partition P of the matrix M and a refinement Q of P, and given a labeling as above of these partitions, we call a block of P ϵ-homogeneous *if its label in P is the $(1-\epsilon)$-dominant label for the set of the corresponding subblocks in Q with their labels.*

Just as Lemma 3.10 was used in Section 3 to find a matrix which is both close to M and simple to check for $\mathcal{M}_F$, we shall use a similar lemma here. The pseudomatrix which we will find will be 'monochromatic' in each block, in the sense that all corresponding subblocks will have the same label (which is a subset of Σ). However, here we also need to ensure that for most blocks which are homogeneous enough, the subblocks will have the label of the block, so we will not need to modify many of the entries of M inside these blocks to arrive at the simplified matrix. We thus need the following strengthening of Lemma 3.10.

Lemma 4.3 *For every h, d and k there exists $c = c(h,d,k)$ such that for any $l \geq cm^c$, every (ml,d)-matrix labeled with a set of h labels which is partitioned into m^d blocks of size l^d, contains an (mk,d) pseudomatrix, so that its intersection with each of these blocks is a (k,d) pseudomatrix whose entries are all identical, and moreover for all but at most an ϵ fraction of the $\frac{1}{c}\epsilon^c$-homogeneous blocks (for any ϵ), the common label of the intersection with them will be identical to the corresponding dominant label.*

The proof here follows closely the outline of the proof of Lemma 3.10, with additional care taken for the homogeneous blocks (the generalization to non-binary alphabets is not hard by itself). We shall give here only a sketch. In the proof of this lemma, we also need a corresponding strengthening of Lemma 3.11, as formulated in the following.

Lemma 4.4 *For every h and k there exists $a = a(h,k)$ such that for any set of m strings of length $l \geq am^a$ over an alphabet of size h, there exist $0 = r_0 < r_1 < \ldots < r_k = l$ such that any string contains a monochromatic substring (i.e. a substring all of whose entries are identical) with exactly one entry in the location range $r_i, \ldots, r_{i+1} - 1$ for every $i = 0, \ldots, l-1$. Moreover, for all but at most an ϵ fraction of the strings which have $(1 - \frac{1}{a}\epsilon)$-dominant labels, these will also be respectively the labels of the monochromatic substrings thus found.*

Proof sketch: The proof here is by induction, and almost identical to the proof of Lemma 3.11. The basis $k = 1$ is also trivial here.

Given that the case $k-2$ was proven and $a(h,k-2)$ is known, the proof for $a(k)$ is done as follows. Similarly to the proof of Lemma 3.11, let $l' = a(k-2)m^{a(k-2)}$, let r_1 be randomly chosen in the range $1, \ldots, l-1-l'$ and set $r_{k-1} = r_1 + l'$. For every string, we set L_x and R_x for every x which appears as a label in this string (there are no more than h such x), as the locations of its first occurrence and its last occurrence in the string. As before, a proper choice of $a(h,k)$ will ensure that with probability more than $\frac{1}{2}$ none of the above values lies in $I = [r_1, r_1+1, \ldots, r_{k-1}-1]$ for any of the strings, which makes it possible to use the induction step for setting $r_2, \ldots, r_{k-2}$.

However, we must also make sure that for most of the strings which have a $(1 - \frac{1}{a(h,k)}\epsilon)$-dominant label, this label will be identical to the label of the monochromatic substring we find. By choosing a possibly higher value of $a(h,k)$, it is indeed possible to ensure that for such a string, with probability at least $1 - 2^{-k-1}\epsilon$ there will be occurrences of this label in both $[0 = r_0, \ldots, r_1 - 1]$ and $[r_{k-1}, \ldots, r_k - 1 = l - 1]$, as well as that it will be a $(1 - \frac{1}{a(h,k-2)}\epsilon)$-dominant label for the substring consisting of the entries in I.

In particular, the above means, for a proper choice of $a(h,k)$, that there exists a choice of r_1 (and r_{k-1}) for which the condition regarding L_x and R_x occurs for *all* strings, and the condition regarding dominant labels occurs for all but at most a $2^{-k}\epsilon$ fraction of the strings with such a label. This allows for the induction hypothesis to follow (summing up the fractions of the strings which failed the dominant label condition in all stages, we bound it by $\sum_{k=1}^{\infty} 2^{-k}\epsilon = \epsilon$). ■

Proof sketch of Lemma 4.3: This proof is similar to the proof of Lemma 3.10, and uses induction on d. Similarly to Lemma 3.10, we actually prove the existence of a common partition for a set of p such matrices $A_{,} \ldots, A_{p-1}$, given that $l \geq c(pm)^c$, so the partition has the corresponding pseudomatrix in each matrix, with the additional property that the fraction of blocks with a dominant label whose intersection with the pseudomatrices has a different label is less than ϵ (summing over all p matrices; this is not necessarily true for each matrix separately). The basis $d = 1$ is follows easily from Lemma 4.4.

Assume that the lemma is proven for $d-1$ and $c(h, d-1, k)$ is known. Let $\delta = \frac{1}{c}\epsilon^c$, where $c = c(h, d, k)$ will be defined latter. As in the proof of Lemma 3.10, for each A_s, for $0 \leq s < p$, we consider only part of the $pm \cdot (ml)^{d-1}$, l-size strings that are defined by 'slicing' each A_s along the d-th dimension, into m matrices, each of size $l \cdot (ml)^{d-1}$ (see Figure 2). For Lemma 3.10, this partial set was arbitrary picked by looking at some fixed residues of the coordinates. Here our goal is to find a subset (not too large) of the strings so that for all but a $\delta^{1/2}$ fraction of them the following will hold: If a string is coming from a block which contains a δ-dominant label, then this label is also a $\delta^{1/2}$-dominant for the string. To construct such a set of strings of size $pm(ml')^{d-1}$ let $L_i \in [l]$ be a random set of size l' for each $1 \leq i \leq d$. We consider only the lines for which for every $i \leq d-1$, its i-th coordinate is in L_i. With positive probability the fraction of such strings which come from blocks which contain a δ-dominant label, but for which the same label is not $\delta^{1/2}$-dominant, is less than $\delta^{1/2}$, so we fix an $L_1, \ldots, L_d$ for which this holds.

We now want to apply Lemma 4.4, to get a partition of $[l]$ into k parts so that there is a k-monochromatic substring in each of the lines with exactly one entry in each part. To apply Lemma 4.4 we need $l \geq a(pm(ml')^{d-1})^a$, where a is the constant from Lemma 4.4. This partition of $[l]$ defines the partition for the d'th coordinate. We now construct matrices $A'_0, \ldots, A'_{pm-1}$ using the labels of the monochromatic substrings found exactly as in the proof of Lemma 3.10. The fraction of the strings not conforming to a dominant label of their block is bounded by $a\delta^{1/2}$. Hence, the fraction of the blocks which correspond to blocks with δ-dominant labels in the original matrix, but do not have the same labels as a $\delta^{1/4}$-dominant one, is bounded by $a\delta^{1/4}$.

On this set of pm matrices we invoke the induction hypothesis to find partitions along the other $d-1$ dimensions, $\{r'_{d',i} | 1 \leq d' \leq d-1, 0 \leq i \leq mk\}$, exactly as in the proof of Lemma 3.10. For this we choose $l' = c(h, k, d-1)(pm \cdot m)^{c(h,k,d-1)}$.

Let $c_1 = 2c(h, k, d-1)$, we note that up to an $a\delta^{1/4} + c_1^{(1/c_1)}\delta^{1/4c_1}$ fraction of the blocks that had a δ-dominant label in $A_0, \ldots, A_p$ will not have the same label for the restriction of the appropriate pseudomatrix. Thus, there exists a choice of a large enough $c(h, k, d)$ that ensures that this is less than ϵ (remember that $\delta = \frac{1}{c}\epsilon^c$). This together with the previous two restrictions on l' vs. l and m sets an appropriate $c = c(h, k, d)$. ∎

4.2 The test

Proof of Theorem 4.1: Let F be a collection of forbidden Σ-labeled posets (for some fixed alphabet Σ), each of size at most k. In order to construct a 1-sided error ϵ-test for the property $\mathcal{M}_F$ we want to arrive at a situation similar to the 1-sided test for 0/1 properties: We want to have two partitions of the matrix into blocks P, Q, where Q is a refinement of P with the following properties: Most blocks of P are δ-homogeneous with respect to the subblocks as defined by Q

(for some small constant δ). In this case we will be able to conclude the test in a way similar to the 0/1 test. However, in order to arrive at such a situation, we cannot rely anymore on the assertion that if there are many non-homogeneous blocks then we have a counter example (as in the 0/1 case). Instead, we construct a sequence of partitions $P_0, \ldots, P_q$ of M, where each P_i is a refinement of the previous one. We will prove in the following that for such sequence, there are two members for which the above holds. The proof is by an iterative argument reminiscent of the proof of Szemerédi's Regularity Lemma. However, the dependency of the number of queries on ϵ will not be as severe as might be expected from such an argument; for every fixed property it will be doubly exponential in a polynomial in ϵ, rather than a tower.

Formally, we choose $m_0 = 1$, and let $m_i = c \cdot m_{i-1}^{c+1}$, where $c = c(h, d, k)$ is the constant provided by Lemma 4.3 and h is set to $2^{|\Sigma|}$. Let P_0 have just one block and let $P_1, ..., P_q$ be a sequence of partitions, each P_i being a partition of M into m_i^d blocks which is a refinement of the previous one. We choose $q = \frac{|\Sigma| c(h,d,k)}{(\frac{1}{3}\epsilon)^{1+c(h,d,k)}}$, so $m_q = exp(exp(poly(1/\epsilon)))$.

The algorithm proceeds as follows. First, $\frac{6}{\epsilon} \ln((m_q)^d \cdot |\Sigma|)$ uniformly random queries are made independently in each block of P_q. If a counter example is found among the query points the algorithm rejects, otherwise it accepts.

It is clear from the formulation that the algorithm has a 1-sided error. We show in the sequel that it has bounded error for all inputs.

Let us assume that there is no counter example within the set of the queried points. We will show that M is ϵn^d-close to having the property with high probability. To see this, we label each block of P_q with the set of all labels known to appear in it as a result of the queries made. We also label all the blocks of the other partitions (remember that P_q is a refinement of all of them) with the set of labels found while making queries within them. Denoting by G the event that no block of P_q (and hence no block of any $P_j,\ j < q$) contains more than a $\frac{1}{3}\epsilon$ fraction of entries of any label which did not appear in its queries, we note that with high probability G happens. The next step is to choose some $0 \leq p < q$, using the following claim.

Claim 4.5 *There exists $0 \leq p < q$ for which all but at most a $\frac{1}{3}\epsilon$ fraction of the blocks of P_p are $\frac{1}{c(h,d,k)}(\frac{1}{3}\epsilon)^{c(h,d,k)}$-homogeneous with respect to P_{p+1}.*

Proof: For every $\sigma \in \Sigma$, $0 \leq l < q$, and a set of blocks A of P_l, we denote by $\chi_{\sigma,l}(A)$ the fraction of blocks of P_l in A whose label contains the color σ. Let $\chi_l(A) = \Sigma_\sigma \chi_{\sigma,l}(A)$ and let $\psi_l = \chi_l(P_l)$. By a slight abuse of notation we let $\chi_{\sigma,l+1}(A)$ and $\chi_{l+1}(A)$ denote the respective quantities where the set of the subblocks of A from P_{l+1} is used. We claim that for any $\delta, \gamma > 0$, if P_l has more than a δ-fraction of blocks which are not γ-homogeneous, then $\psi_{l+1} \leq \psi_l - \delta \cdot \gamma$.

Indeed, we observe that for any $p' \leq l$ and any set of blocks A of p' $\chi_{p'}(A) \geq \chi_l(A)$. We also observe that $\psi_l = Pr(A)\chi_l(A) + (1 - Pr(A))\chi_l(P_l - A)$ (where $Pr(A)$ is the relative size of A to the whole matrix).

Now assume that P_l has more than a δ-fraction of blocks that are not γ-homogeneous, and let A be the set of all blocks of P_l that are not γ-homogeneous. Then $Pr(A) \geq \delta$. Also, by the above observation it is enough to show that $Pr(A)(\chi_l(A) - \chi_{l+1}(A)) \geq \delta \cdot \gamma$. By definition, for any $p \geq l$, $\chi_p(A) = \Sigma Pr(B)\chi_p(B))$ where B is a block in A and $Pr(B)$ is its relative size. However, for each block B, assuming that its label in P_l is $S, |S| = r \leq h$, then $\chi_l(B) = r$, while as B is not γ-homogeneous then $\chi_{l+1}(B) \leq Pr(S) \cdot r + (1 - Pr(S)) \cdot (r-1) \leq r - 1 + Pr(S) \leq r - \gamma$ (here $Pr(S)$ is the fraction of blocks of P_{l+1} in B that have label S). Hence $Pr(A)(\chi_l(A) - \chi_{l+1}(A)) \geq \delta \cdot \gamma$ as claimed.

To end the proof of the claim, note that since $\chi_0 \leq |\Sigma|$, if the claim does not hold for every p, it would imply that $\chi_q < 1$, which is a contradiction (every block label must contain at least one member of Σ). Hence there exists a p as in the formulation of the claim. ∎

We now return to the proof of the theorem: Choosing a p as in the formulation of the claim, we proceed as follows. We consider the (m_{p+1}, d) matrix consisting of the labels of the blocks of P_{p+1}, and its partition into $(m_p)^d$ blocks corresponding to the blocks of P_p. Over these we apply Lemma 4.3 to find a (km_p, d) pseudomatrix so that its intersection with every block from P_p is a monochromatic (k, d) pseudomatrix. By Lemma 4.3 for all but at most a $\frac{1}{3}\epsilon$ fraction of the homogeneous blocks in P_p, the restrictions of the pseudomatrix found to these blocks will have labels identical to the blocks. We let W be the label matrix corresponding to the above pseudomatrix.

We now consider the following (n, d) matrix M_W: For every entry of M whose label is in the set which is the common label of the intersection of the above pseudomatrix with the block from P_p, this will also be the entry of M_W in the respective location. Every other entry shall be replaced with an arbitrary member of the set labeling the intersection of the pseudomatrix with the entry's block.

We now note that, as in the 0/1 case, if G happens then M_W differs from M in less than ϵn^d places, because the differences can only be in entries which did not conform to the corresponding block label in P_{p+1}, or entries which were in non-homogeneous blocks of P_p, or entries which were in homogeneous blocks of P_p whose label was not the common label of the intersection of the pseudomatrix found by Lemma 4.3 within this block. Since each of these three categories contains at most $\frac{1}{3}\epsilon n^d$ of the entries, M is ϵn^d-close to M_W.

We now claim that if the algorithm accepts then M_W has the property and hence we are done. Indeed, assume on the contrary that M_W contains a counter example, then we claim that there already exists a counter-example within the locations of the matrix queried by the algorithm (regardless of G): Given the existence of a counter example, F_0, in M_W, then, by Lemma 3.13, there exists an embedding of the poset F_0 (disregarding the labels for now, they will be dealt with later) in M_W, considered as $G(m_p \cdot k, d)$, where each point in this embedding is in a different subblock, and with no two points sharing a coordinate. As W is an $(m_p \cdot k, d)$ matrix, we look at this embedding in W. By our construction, the label of each entry in a P_p block of M_W is a member of the common block label of this block in M_W. Hence, in the embedding F_0 in W, each entry is labeled by a set that contains the label of the original entry of F_0. However, as W is a label matrix of a pseudomatrix of P_{p+1} (where each block is considered as a point), an entry of M can be chosen from each block of P_{p+1} whose label is any desired member of the corresponding set-label of the point in W. In particular, we can choose for any entry the label which leads to an embedding of F_0 in M. ∎

5 A $\forall\exists$ property that is not testable

In this section we construct a $\forall\exists\phi(x_1, \ldots, x_k)$ property that is not ϵ-testable for some fixed ϵ. The construction is similar in spirit to the non-testable graph property constructed in [1]. We first construct such a property for a model which is stronger than the poset model, and then based on it derive a poset-only $\forall\exists$ property which is not testable. For the rest of this subsection we consider only 2-dimensional matrices, and use the notion of rows and columns in the usual matrix sense.

5.1 A non-testable property concerning submatrices

The property that we present here uses the alphabet $\{0, 1, 2\}$ and is in a slightly more general model than $\forall\exists$-poset: We say that a matrix satisfies the property 'permutation' if it is a row/column permutation of a symmetric matrix with all 2's on its primary diagonal, and no 2's anywhere else. It is not hard to see that this is equivalent to the matrix satisfying the following three conditions:

1. For every matrix entry which is not '2' there is an entry on the same row and an entry on the same column which are both '2'.
2. For every '2' entry there is no other entry on the same row or column which is also '2'.
3. The matrix contains none of the following 2×2 matrices as a submatrix (to ensure that the original matrix was symmetric).

$$\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}; \quad \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}; \quad \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix}; \quad \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$$

Using the above equivalent formulation, the property 'permutation' can be expressed as a $\forall\exists\phi$ type property where ϕ uses, apart from the order relation and the value relations ("the entry at location x is 0/1/2"), the two additional relations: "x_1, x_2 are on the same row" and "x_1, x_2 are on the same column" (this model is discussed further in Section 6). We claim that property 'permutation' is not testable:

Proposition 5.1 *Property 'permutation' is not $\frac{1}{10}$-testable even by a two sided error algorithm.*

Proof: Using Yao's principle [16], we define a probability distribution on inputs and show that any fixed deterministic algorithm that queries $d = O(1)$ queries has an average error (according to the distribution on inputs) of more than $\frac{1}{3}$.

Let us first define a distribution P on positive inputs, a distribution N on negative inputs, and then the distribution D will be to choose with probability $\frac{1}{2}$ a member according to P and with probability $\frac{1}{2}$ a member according to N. P is defined by first choosing randomly and uniformly a symmetric matrix B with 2's on the primary diagonal and 0's and 1's everywhere else. The input matrix A is constructed from B by permuting its rows according to a permutation which is chosen uniformly in random.

N is defined by just letting A be a uniformly random 0/1 matrix. N is a distribution also on inputs which are close to satisfying 'permutation', but a matrix selected according to N will almost surely be $\frac{1}{10}$-far from satisfying it. This can be proven by using a large deviation inequality to show that the probability of any (row) permutation of A to be close to being symmetric is much smaller than the inverse of the number of permutations. We omit further details on this point.

Now, let $\mathcal{A}$ be a deterministic algorithm for testing the above property that queries $d = O(1)$ queries. As $d = O(1)$ we may assume that $\mathcal{A}$ is non-adaptive and hence all queries are made in advance (otherwise we can go over the decision tree of the algorithm to get a nonadaptive algorithm which makes $2^d = O(1)$ queries). We further assume (by making additional $O(1)$ queries if necessary) that $\mathcal{A}$ actually queries all locations $\{x_{ij} | i, j \in I\}$ where $I = \{i_1, \ldots, i_d\}$ is some fixed set of size d.

We claim that for both distributions P and N, the distribution of the restriction of $\mathcal{A}$ to the submatrix that is queried by $\mathcal{A}$ is close to the uniform distribution on 0/1 matrices of size $d \times d$. This implies that when we choose one of P and N randomly and then choose a matrix according to this distribution, the algorithm will err with probability close to $\frac{1}{2}$.

We need the following easy claim.

Claim 5.2 *Let $\epsilon > 0$ and d be fixed, and n be large enough (in relation to ϵ and d). For a fixed set $I \subset \{1, \ldots, n\}$ of size d, let σ be a uniformly chosen random permutation over $\{1, \ldots, n\}$. Then, with probability at least $1 - \epsilon$ there are no $i, j \in I$ for which $\sigma(i) = j$.* ■

For a random matrix chosen according to N, the submatrix queried by the algorithm is a uniformly random 0/1 matrix. To end the proof, we note that for a random matrix chosen according to P, if the random permutation used to permute the rows of the intermediate matrix B satisfies the above claim, than the submatrix queried by the algorithm is also a uniformly random 0/1 matrix – it will contain no '2' entries and moreover it will contain no pair of entries of the intermediate matrix B which are correlated in the distribution P by its symmetry. ■

As a final remark, we note that it can be shown (by further analysis of the above proof) that even $n^{1/10}$ queries are not enough to $\frac{1}{10}$-test for this property (the seemingly exponential gap between the adaptive and nonadaptive algorithms can be avoided for this particular case).

5.2 Submatrices, tight submatrices and witnesses

In order to construct a non-testable property of 0/1 matrices that is strictly in the $\forall\exists$-poset model we need some machinery that will be developed here.

Relations like "x and y are not on the same row or column" can be expressed in the poset model using additional variables and quantifiers. For example, it can be seen that if $x \leq y$ (in the product ordering of the 2-dimensional matrix), then they do not share a row or a column if and only if there exist w and z which are incomparable and moreover satisfy $x \leq w \leq y$ and $x \leq z \leq y$. We call such a pair variables a *witness* for x and y not sharing a coordinate.

We can extend this further: The locations $(x_{i,j})_{i,j\in[k]}$ represent a subgrid (and their labels a submatrix) if and only if they have between them the order relations that a subgrid has, and furthermore there are no witnesses for any $x_{i,j}$ and $x_{i,j'}$ not sharing a coordinate, as well as for any $x_{i,j}$ and $x_{i',j}$ not sharing a coordinate. Note however that we cannot distinguish this way between a submatrix and its transpose.

We can also use witnesses to express other notions: Two points $x \leq y$ that are not equal but do share a coordinate, reside on consecutive values of the other coordinate if and only if there exist no z different from x and y for which $x \leq z \leq y$. This allows us to express the following definition within a first order poset property. $(x_{i,j})_{i,j\in[k]}$ are said to represent a *tight subgrid* if there exist i_0 and j_0 so that $x_{i,j} = (i_0 + i, j_0 + j)$ for every i and j. Similarly we define the notion of a *tight submatrix* using their labels. We note now that $(x_{i,j})_{i,j\in[k]}$ represent a tight subgrid (or its transpose) if and only if they satisfy the appropriate order relations, and furthermore there exist no witnesses showing that $(x_{i,j})_{i,j\in[k]}$ is not a subgrid and no witnesses showing that $x_{i,j}$ and $x_{i,j+1}$ or $x_{i,j}$ and $x_{i+1,j}$ (for any i and j) do not reside consecutively on the non-shared coordinate.

We are now ready to define the non-testable poset property.

5.3 A non-testable $\forall\exists$-poset property

In our definition we shall use the following matrices, which we shall call the three *guide matrices*. The idea would be to encode the $\{0, 1, 2\}$ labeling in the property *permutation* defined in Subsection 5.1 by copies of these matrices.

$$G_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad G_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad G_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

The property 'permutation-p' that is defined below describes matrices that have the property permutation, where each label $i \in \{0, 1, 2\}$ is replaced with the guide matrix G_i. Formally, it is defined as the input matrix satisfying the following conditions (actually each statement should be replaced by a counterpart which is invariant with respect to taking a transpose of the input matrix, but this has no essential effect on the following analysis as the guide matrices are not symmetric).

1. For every $(x_{i,j})_{i\in[5],j\in[10]}$ there either exist a witness showing that they do not form a tight subgrid, or the labels of the input matrix M in these locations are such that $(M(x_{i,j}))_{i\in[5],j\in[5]}$ is a guide matrix if and only if $(M(x_{i,j}))_{i\in[5],j\in[10]-[5]}$ is a guide matrix.
2. Similarly to the above, for every $(x_{i,j})_{i\in[10],j\in[5]}$ that form a tight subgrid, $(M(x_{i,j}))_{i\in[5],j\in[5]}$ is a guide matrix if and only if $(M(x_{i,j}))_{i\in[10]-[5],j\in[5]}$ is.
3. If $(x_{i,j})_{i\in[5],j\in[5]}$ form a tight subgrid and their labels form a guide matrix other than G_2, then the input matrix M contains a "1" on the same row as $x_{1,1}$ and a "1" on the same column

as $x_{1,1}$ (this can be formulated in terms of the order relations of the locations of the 1's with respect to $x_{1,0}$, $x_{0,1}$, $x_{2,1}$ and $x_{1,2}$).

4. If $(x_{i,j})_{i\in[5],j\in[10]}$ form a subgrid, and moreover $(x_{i,j})_{i\in[5],j\in[5]}$ and $(x_{i,j})_{i\in[5],j\in[10]-[5]}$ are both tight, then $(M(x_{i,j}))_{i\in[5],j\in[5]}$ and $(M(x_{i,j}))_{i\in[5],j\in[10]-[5]}$ cannot be both G_2.

5. Similarly if $(x_{i,j})_{i\in[10],j\in[5]}$ form a subgrid, and $(x_{i,j})_{i\in[5],j\in[5]}$ and $(x_{i,j})_{i\in[10]-[5],j\in[5]}$ are tight, then $(M(x_{i,j}))_{i\in[5],j\in[5]}$ and $(M(x_{i,j}))_{i\in[10]-[5],j\in[5]}$ cannot be both G_2.

6. If $(x_{i,j})_{i\in[10],j\in[10]}$ form a subgrid for which $(x_{i,j})_{i\in[5],j\in[5]}$, $(x_{i,j})_{i\in[5],j\in[10]-[5]}$, $(x_{i,j})_{i\in[10]-[5],j\in[5]}$, and $(x_{i,j})_{i\in[10]-[5],j\in[10]-[5]}$ are tight, then $(M(x_{i,j}))_{i\in[10],j\in[10]}$ is not any of the following formations.
$$\left(\begin{array}{c|c} G_2 & G_0 \\ \hline G_1 & G_2 \end{array}\right); \quad \left(\begin{array}{c|c} G_2 & G_1 \\ \hline G_0 & G_2 \end{array}\right); \quad \left(\begin{array}{c|c} G_0 & G_2 \\ \hline G_2 & G_1 \end{array}\right); \quad \left(\begin{array}{c|c} G_1 & G_2 \\ \hline G_2 & G_0 \end{array}\right)$$

We first establish the connection between the properties 'permutation' and 'permutation-p'.

Given an $n \times n$ matrix A labeled by $\{0,1,2\}$, we say that a $5n \times 5n$ 0/1 matrix M is the *tiling* of A, if for every $0 \le i,j \le n-1$ the tight submatrix $(M(5i+i', 5j+j'))_{0\le i',j'\le 4}$ is equal to $G_{A(i,j)}$. In other words, M is formed by replacing each entry of A with the appropriate 5×5 guide matrix.

Claim 5.3 *If M is a tiling of a matrix A which satisfies 'permutation', then M satisfies the property 'permutation-p'.*

Proof: It is not hard to see that a tiling of any 0/1/2 matrix does not contain any tight submatrices which are equal to a guide matrix (or its transpose) apart from those of the form $(M(5i+i', 5j+j'))_{0\le i',j'\le 4}$. Thus, the first two conditions in the definition of 'permutation-p' are satisfied in M. The third condition now follows from the first condition in the definition of 'permutation' (note that G_2 is the only guide matrix with a '1' entry on its second row, as well as the only one to have a '1' on its second column). The fourth and fifth conditions in 'permutation-p' follow from the second condition in 'permutation', and the sixth condition in 'permutation-p' follows from the third condition in 'permutation'. This completes the claim. ■

Claim 5.4 *If M is a $5n \times 5n$ matrix which satisfies 'permutation-p' and has any tight submatrix of the form $(M(5i+i', 5j+j'))_{0\le i',j'\le 4}$ which is equal to a guide, then M is a tiling of some $n \times n$ matrix A which satisfies 'permutation'.*

Proof: The first two conditions in the definition of 'permutation-p' imply that if M satisfies it and has any tight submatrix of the form $(M(5i+i', 5j+j'))_{0\le i',j'\le 4}$ equal to a guide, then M is a tiling of some $n \times n$ matrix A (these conditions say in essence that a tight submatrix adjacent to a tight guide submatrix is a guide itself). The last four conditions in 'permutation-p' guarantee in turn that A satisfies the three conditions in the definition of 'permutation'. ■

Theorem 5.5 *Property 'permutation-p' is not $\frac{1}{250}$-testable even by a two sided error algorithm.*

Proof: By Claim 5.3 and Claim 5.4, given an ϵ-test of a 0/1 $5n \times 5n$ matrix M for 'permutation-p', we construct a test of a 0/1/2 $n \times n$ matrix A for 'permutation' by querying the location $(\lfloor\frac{i}{5}\rfloor, \lfloor\frac{i}{5}\rfloor)$ of A whenever the location (i, j) of M is queried, and assigning to (i, j) the entry of $G_{A(\lfloor\frac{i}{5}\rfloor,\lfloor\frac{i}{5}\rfloor)}$ at $(i \bmod 5, j \bmod 5)$. In other words, the test of A is constructed by simulating the test of M on the tiling of A.

We now claim that the new test is a 25ϵ-test for the property 'permutation'. In particular, this means that the existence of a $\frac{1}{250}$-test for 'permutation-p' implies the existence of a $\frac{1}{10}$-test for 'permutation'. This would contradict Proposition 5.1, so the theorem follows.

Indeed, if A has 'permutation' then clearly its tiling M has 'permutation-p'. On the other hand let us assume that A is 25ϵ-far from satisfying 'permutation', and $\epsilon < \frac{1}{25}$. Claim 5.4 implies that any matrix M' which satisfies 'permutation-p' and is not ϵ-far from M is a tiling of some matrix A' which satisfies 'permutation', because one has to change M in at least $\frac{1}{25}(5n)^2$ places to remove all tight guide submatrices of the form $(M(5i+i', 5j+j'))_{0\leq i',j'\leq 4}$. Since A' is at least 25ϵ-far from A (by the assumption that A is far from any matrix which satisfies 'permutation'), the tiling M' of A' is at least ϵ-far from the tiling M of A. ∎

6 Bipartite graph properties

An important and natural variant of the poset-model is the model that allows to address properties that are characterized by a finite collection of forbidden submatrices. We restrict ourselves along the sequel to the 2-dimensional case. To put it in the framework of our study, we allow in our language, apart of the poset relation, the relations $row(x_1, x_2)$ which state that x_1 is on the same row as x_2, and similarly $col(x_1, x_2)$ for columns (the same model was used in the intermediate result in Section 5). The $\forall$-properties would then correspond to exactly the properties that are characterized by a finite collection of a forbidden submatrices (in a similar manner to Observation 2.2). We call this model the 'submatrix model'. As was explained in Section 5, the submatrix model is closely related to a submodel of the $\forall\exists$-poset.

The model 'submatrix' includes some interesting properties. In particular, the permutation-invariant properties in it are tightly connected to bipartite graph properties that are characterized by a collection of forbidden induced subgraphs:

Definition 6.1 *For a finite collection F of 0/1 matrices, we denote by $\mathcal{S}_F$ all 0/1-matrices that do not contain any row and/or column permutation of a member of F.*

Observation 6.2 *Every bipartite graph property (viewed as a collection of matrices) that is characterized by a finite collection of forbidden induced subgraphs is equivalent to a property $\mathcal{S}_F$ for some finite set F of matrices. In addition, every $\mathcal{S}_F$-property in the 'submatrix' model is a bipartite graph property as above.* ∎

Bipartite graph properties that are characterized by a finite set of forbidden induced subgraphs are testable: In [1], it was shown how to test $\forall$-graph properties for general graphs (not necessarily bipartite). However, the dependence on ϵ in these tests is $tower(tower(poly(1/\epsilon)))$, resulting from the use of the Regularity Lemma. Here we prove that such bipartite graph properties are testable (2-sided) with query complexity $exp(exp(poly(1/\epsilon)))$, hence, breaking the regularity lemma bounds in this case. Our main tool is a conditional (but efficient) version of the regularity lemma as stated in Lemma 6.6. The main result here is:

Theorem 6.3 *Let F be a fixed collection of 2-dimensional matrices. Then $\mathcal{S}_F$ for the 2-dimensional case is $(\epsilon, exp(exp(poly(\frac{1}{\epsilon}))))$-testable, by a two sided error algorithm.*

Using a scheme involving an appropriate Ramsey-like lemma, it is also possible to arrive at a one-sided test for these properties at the price of increasing the query complexity.

Theorem 6.4 *Let F be a fixed collection of 2-dimensional matrices. Then $\mathcal{S}_F$ for the 2-dimensional case is $(\epsilon, exp(exp(exp(poly(\frac{1}{\epsilon})))))$-testable, by a one sided error algorithm.*

The proof of Theorem 6.4 is outside the scope of this abstract, and will appear in a future full journal version thereof. To present the test proving Theorem 6.3, we will need some machinery:

Let M be a 0/1-labeled, $n \times m$ matrix. Let r be an integer, then an r-partition of M is a partition of the rows into $r' < r$ parts $\{R_1, \ldots, R_{r'}\}$ and a partition of the columns into $r'' < r$ parts $\{C_1, \ldots, C_{r''}\}$. Each submatrix of the form $R_i \times C_j$ will be called a block (note that these are not necessarily 'tight submatrices' as per the definition of the previous section). The weight of the (i,j) block is defined by $\frac{1}{nm}|R_i||C_j|$. We also define similar weights for the R_i's and C_j's, e.g. $w(R_i) = \frac{1}{n}|R_i|$.

Definition 6.5 *Let $\mathcal{P} = \{R_1, \ldots R_{r'}\} \times \{C_1, \ldots C_{r''}\}$ be an r-partition, and let $\delta > 0$. We say that $\mathcal{P}$ is a (δ, r)*-partition *if the total weight of the δ-homogeneous blocks is at least $1 - \delta$.*

Unlike the poset model, in this model we cannot easily reject in the case that a given partition is not a (δ, r)-partition. However, it is still true that we can reject the input unless it admits *some* (δ, r)-partition, as the following lemma shows.

Lemma 6.6 *Let k, δ be fixed, then for every $n \times n$, 0/1-matrix M, with n large enough, either M has a (δ, r)-partition for $r = r(\delta, k) \leq exp(exp(poly(1/\delta)))$, or for every 0/1-labeled $k \times k$ matrix F, a $g(\delta, k) > exp(-poly(1/\delta))$ fraction of the $k \times k$ submatrices of M are permutations of F.*

We note that this lemma is a 'conditional' version of Szemerédi's Regularity Lemma, as a (δ, r)-partition is in particular a regular partition in the sense of Szemerédi of the corresponding bipartite graph. The improvement over using directly the Regularity Lemma is achieved because of this conditioning. The proof of the lemma will be presented in Subsection 6.2. It allows us to reduce the testing problem to matrices that admit a (δ, r)-partition for certain δ, r, as for matrices that do not admit such partitions the lemma asserts that querying a random submatrix will find a counter example with positive probability.

Our next goal is then to construct a test for such matrices. This test will be very similar to the test in Subsection 3.1, the main problem will be to actually 'know' enough of the partition by sampling. For this we need some more machinery. It is described in the following subsection, along with the framework of the proof of Theorem 6.3.

6.1 Partition, signatures and the general framework

Assume that M has a (δ, r)-partition. We have no hope, of course, to find it by $O(1)$ many queries as we cannot even sample a single point from each row. Hence, we will need here some 'high level feature' of such partitions that can be detected by sampling. This is given in the following:

Let M be a matrix with a (δ, r) partition $\mathcal{P}$ defined by the row partition $\{R_1, \ldots, R_s\}$ and the column partition $\{C_1, \ldots, C_t\}$, $s, t \leq r$. Then $\mathcal{P}$ naturally defines a $\{0/1/\mathrm{X}\}$-labeled, $s \times t$ matrix P in the following way. The rows correspond to the sets $R_1, \ldots, R_s$ and the columns to the sets $C_1, \ldots, C_t$ so that each entry $P_{i,j}$ corresponds to the submatrix that is defined by R_i, C_j and is labeled by its $(1-\delta)$-dominant label if it is δ-homogeneous, and otherwise by X. We say that P is the *pattern* of the partition $\mathcal{P}$ of M. As the block sizes of a (δ, r)-partition need not be fixed, we will also need information about the weights of R_i, C_i, $(i,j) \in [s] \times [t]$: Let M be an $n \times n$ matrix with a (δ, r)-partition $\mathcal{P}$ defined by the row partition $\{R_1, \ldots, R_s\}$ and the column partition $\{C_1, \ldots, C_t\}$. Then a γ-*signature* of $\mathcal{P}$ is an $s \times t$, $\{0/1/\mathrm{X}\}$-labeled matrix P and two sequences $\{\alpha_i\}_1^s, \{\beta_i\}_1^t$, where P is a pattern of $\mathcal{P}$ and $|\frac{|R_i| \cdot |C_j|}{n^2} - \alpha_i \cdot \beta_j| \leq \frac{\gamma}{st}$ for every $(i,j) \in [s] \times [t]$.

Note that the signature of a partition is 'closed' under permutation of rows and columns, namely, any row/column permutation of P with the respective permutations of $\{\alpha_i\}_1^s$ and $\{\beta_i\}_1^t$ is also a signature of the same matrix. Moreover, a signature of M is also a signature of all permutations of M.

The signature of a partition has sufficient properties for constructing a test. Namely, it can be detected by sampling and it carries enough information in order to allow for a test as asserted by the following:

Lemma 6.7 *Assume that an $n \times n$, 0/1 matrix M has a (δ, r)-partition with $r \geq \frac{100}{\delta^{10}}$, then by sampling $q = poly(r)$ many queries, a γ-signature of a $(12(\delta+\gamma)^{1/4}, r^4)$-partition can be found for any $\gamma \geq \frac{1}{r^{1/2}}$, with probability $1 - exp(-\gamma r)$.*

We note that a test for such partition, as guaranteed by the lemma, can also be deduced from [12], with exponentially worse running time (and essentially the same type query complexity). The proof of Lemma 6.7 is given in Subsection 6.3. We end this discussion by showing that it indeed implies a 2-sided error test.

Proof of Theorem 6.3: Assume that we want to ϵ-test M for a permutation invariant collection of forbidden induced $k \times k$ submatrices. Our 2-sided test follows the general idea of the test in

Subsection 3.1. Blocks will now correspond to partition-blocks: Let $\delta_1 = (\frac{\epsilon}{2^7})^4$ and let $g = g(\delta_1, k)$, $r = r(\delta_1, k)$ of Lemma 6.6. For $\frac{1}{g^2}$ times, independently, we choose k random rows and k random columns of M and query all k^2 points in the $k \times k$ matrix that is defined. If we find a counter example in the queried points we answer 'No'. Otherwise, by Lemma 6.6 we may assume that with probability $(1 - exp(-1/g))$ M has a (δ_1, r)-partition.

Let $\gamma = (\frac{\epsilon}{2^7})^4$ and let $\delta_2 = 12(\delta_1 + \gamma)^{1/4}$. Given that M has a (δ_1, r)-partition, using Lemma 6.7, we can find a γ-signature of a (δ_2, r^4)-partition by sampling $poly(r)$ queries. Let P and $\{\alpha_i\}_1^s, \{\beta_i\}_1^t$ be a signature of such partition. We form, as in Subsection 3.1, an $n \times n$ matrix M_Q that represents our knowledge of M: We partition the rows of M_Q into s parts of weights $\{\alpha_i\}_1^s$ and the columns into t parts of weights $\{\beta_i\}_1^t$. For each block of P, we set every entry of the corresponding block of M_Q to have the same label as in P. Now, exactly as in Subsection 3.1, each possibility of assigning 0/1 values to the X entries of M_Q and each possible choice of flipping the values in at most an $\epsilon/4$ fraction of the entries in every 0/1 block, results in a 0/1-labeled matrix; we denote the set of all such matrices by $\mathcal{M}_{Q,\epsilon}$. We check if any of the members of $\mathcal{M}_{Q,\epsilon}$ has the property. If there is such a member, the algorithm answers 'Yes'. Otherwise, if every member $\mathcal{M}_{Q,\epsilon}$ contains a permutation of a forbidden submatrix the answer is 'No'. Note, this last phase of the algorithm involves no additional queries and is just a computation phase.

To see that the algorithm is correct we first note that if a counter example is found in the first phase of the algorithm, then the input M does not have the property with probability 1. Hence the algorithm can err only in the second phase.

We claim that, with high probability, a row/column permutation of M is ϵn^2 close to each member of $\mathcal{M}_{Q,\epsilon}$. Indeed, assume that the signature that has been found is a γ signature of a (δ_2, r^4)-partition of M. Then, by a simple calculation, it follows that some row/column permutation of M is $(2\delta_2 + 2\gamma)n^2$-close to M_Q. Furthermore, M_Q is $\frac{\epsilon}{2}n^2$-close to any member of $\mathcal{M}_{Q,\epsilon}$, and hence M is $(\frac{\epsilon}{2} + 2\delta_2 + 2\gamma)$-close to any member of $\mathcal{M}_{Q,\epsilon}$. However, by our choice of $\delta_1, \delta_2, \gamma$, M is indeed ϵn^2-close to all members of $\mathcal{M}_{Q,\epsilon}$. The failure probability here is the sum of the failure of Lemma 6.6 and Lemma 6.7, which is exponentially small.

We conclude that, given that M is ϵn^2-close to all members of $\mathcal{M}_{Q,\epsilon}$, then if it has the property then certainly one member of $\mathcal{M}_{Q,\epsilon}$ will have the property (as M itself is such member). On the other hand if M is more than ϵn^2 far from having the property then certainly every member of $\mathcal{M}_{Q,\epsilon}$ cannot have the property.

Clearly the query complexity of the test is $poly(r, 1/\epsilon)$ and by our expression for r it is $exp(exp(poly(1/\epsilon)))$, which concludes the proof. ∎

Remark: In all the above we discussed forbidden *induced* subgraphs. Not having a forbidden subgraph (rather then induced subgraph) is a monotone decreasing property. In this case, the test is trivial, by density: For a large enough density, Lemma 3.3 asserts that the answer 'No' is correct (as the graph will have a large enough complete bipartite graph), while if the density is low the answer is trivially 'Yes', as the graph is close to the empty (edge-less) one.

We now turn back to the proofs of Lemma 6.7 and 6.6 for which we need some more machinery.

6.2 Universal sets and the proof of Lemma 6.6

We begin with the following simple observation.

Observation 6.8 *An r'-partition which is a refinement of a (δ, r)-partition is in particular a $(\sqrt{\delta}, r')$-partition.* ■

The proof of Lemma 6.6 will follow from the definition and lemma below.

Let M be an $n \times m$, $0/1$ matrix. Let $C = \{c_{i_1}, \ldots, c_{i_k}\}$ be a sequence of k columns of M and $\phi \in \{0,1\}^k$. Then ϕ defines the set of rows R_ϕ of M which consists of all rows whose intersection with the columns of C (in this order) is ϕ.

Definition 6.9 *Let $0 < \alpha < 1$ and let M be a 0/1, $n \times m$ matrix. We say that an ordered set of columns $C = \{c_{i_1}, \ldots, c_{i_k}\}$ is (α, k)-universal if for every $\phi \in \{0,1\}^k$, $|R_\phi| \geq \alpha n$.*

Lemma 6.10 *Let k be a fixed integer, $\delta > 0$, then for every $n \times m$, 0/1-matrix M, either M has a (δ, r)-partition for $r \leq exp(exp(poly(k, 1/\delta)))$ or at least $g(\delta, k)$-fraction of the ordered k-column sets are $(g(\delta, k), k)$-universal, where g is a function satisfying $g \geq exp(-poly(\delta, k)))$.*

Before we proof the lemma we prove the following easier Claim:

Claim 6.11 *For every k,δ every $0/1$ matrix either has a 2^{k+1}-partition for which the total weight of the δ-homogeneous blocks is at least $(1-\delta)\delta^{k-1}$, or at least a δ^k-fraction of its ordered sets of k-columns are δ^k-universal.*

Proof: We prove the claim by induction on k. For $k = 1$, we may assume that M is not δ-homogeneous for otherwise it has a $(\delta, 1)$-partition (which is stronger than what we need). Let C_M be the set of columns of M and let $C = \{c \in C_M | c \text{ is not } \delta\text{-homogeneous}\}$. If $|C| < \delta m$ then we have a partition of the columns into three parts: the set C, the set of columns that have at most a δ-fraction of 0's, and the set of columns with at most a δ-fraction of 1's. However, since in the above partition only C is not δ-homogeneous, its small size implies that this is a $(\delta, 3)$-partition (which is stronger than what we have claimed). Hence, we now assume that $|C| \geq \delta m$. However, each $c \in C$ is $(\delta, 1)$-universal and therefore we have δm many $(\delta, 1)$-universal sets in M. This completes the proof for the case $k = 1$. Note that this actually may serve as a proof of Lemma 6.10 for the case $k = 1$.

Assuming now that the claim has been proven for all $j < k$, we prove it for k: We may assume that M has a δ^{k-1}-fraction of δ^{k-1}-universal sequences, as otherwise, by the induction hypothesis M has a 2^k partition for which the total weight of the δ-homogeneous submatrices is at least $(1-\delta)\delta^{k-2}$ which is even stronger than what we need.

Let $C_{k-1} = \{c_1, \ldots, c_{k-1}\}$ be such a fixed sequence. C_{k-1} partitions the rows of M into 2^{k-1} sets: $\{R_\phi | \ \phi \in \{0,1\}^{k-1}\}$. We use this to define a partition $\{C\} \cup \{C_{\phi,0} | \phi \in \{0,1\}^{k-1}\} \cup \{C_{\phi,1} | \phi \in$

$\{0,1\}^{k-1}\}$ of the columns into (up to) 2^k+1 sets as follows. C is the set of columns that are not δ-homogeneous for all the R_ϕ (i.e. the set of columns whose intersection with every R_ϕ is not δ-homogeneous). For every other column, we pick arbitrarily one $\phi \in \{0,1\}^{k-1}$ such that the column is δ-homogeneous for R_ϕ, and put it in either $C_{\phi,0}$ or $C_{\phi,1}$ according to which is the $(1-\delta)$-dominant label in the intersection.

By the definition of C, for any $c_k \in C$ the sequence $\{c_1, \ldots, c_k\}$ is $(\delta \cdot \delta^{k-1})$-universal. Hence, if C contains at least a δ fraction of the columns for all possible universal sequences C_{k-1} then we get at least a $\delta \cdot \delta^{k-1}$-fraction of the k ordered sets to be δ^k-universal sets, and we are done. Let us assume therefore that C_{k-1} is a sequence for which $|C| < \delta \cdot m$. But then, each of the other sets of columns is δ-homogeneous for at least one ϕ, and that ϕ defines at least a δ^{k-1}-fraction of the rows. Thus, at least a δ^{k-1}-fraction of the total area in these columns is δ-homogeneous. As the columns not in C are at least a $1-\delta$ fraction of all columns, it follows that the total weight of the δ-homogeneous blocks in M is at least $(1-\delta)\delta^{k-1}$ as needed. ∎

The following is a trivial observation:

Observation 6.12 *Assume that for a submatrix of M that is defined by a subset of rows R and a subset of columns C, an α-fraction of the ordered k column sets are β-universal. Furthermore, assume that the relative weights of R and C are w_R, w_C respectively, then at least a $w_C \cdot \alpha$ fraction of the k ordered column sets of M are $(w_R \cdot \beta, k)$-universal.*

Proof: Immediate from the definition.

Proof of Lemma 6.10 Let $\delta_1 = \delta^2$. By Claim 6.11 we may assume that there is a 2^{k+1}-partition of M for which the total weight of the δ_1-homogeneous blocks is at least $\alpha \geq (1-\delta_1)\delta_1^{k-1}$, as otherwise the claim guarantees a stronger result on k-universal sets than what we need. Our aim, now, is to amplify the fraction of area of δ_1-homogeneous submatrices by applying recursively the claim to each non δ_1-homogeneous submatrix, or to show that the required fraction of k-sets are δ^k-universal.

Formally let $a = (1-\delta_1)\delta_1^{k-1}$. We are going to do $i_0 = \frac{\log(1/(1-\delta_1))}{a}$ iterations as follows: After the i'th iteration, $i \leq i_0$, we will have a partition of M (which at this point is not necessarily a cross product of a partition of the rows and a partition of the columns) into $r_i \leq 2^{2(i+1)(k+1)}$ blocks for which the total weight of the δ_1-homogeneous blocks is at least $1-(1-a)^i + \alpha(1-a)^i - i\delta_1^{2k-1}$. Then in the $(i+1)$'th iteration, either we find that an $exp(-poly(1/\delta_1))$-fraction of the ordered k-column sets are $(exp(-poly(1/\delta_1)), k)$-universal, or we exhibit a new partition of M into $r_{i+1} \leq 2^{2(i+2)(k+1)}$ many blocks for which the total weight of δ_1-homogeneous blocks is at least $1-(1-a)^{i+1} + \alpha(1-a)^{i+1} - (i+1)\delta_1^{2k-1}$. Indeed, for $i=0$ this is what is guaranteed after one application of Claim 6.11 on M. Assume that this is shown for i, let us show it for $i+1$: We assume that after i iterations we have a partition of M into at most r_i blocks such that the total weight of the δ_1-homogeneous subblocks is at least $1-(1-a)^i + \alpha(1-a)^i - i\delta_1^{2k-1}$. We then take each block of the above partition, that is defined by a subset R of rows and a subset C of columns, of weights $w_R, w_C \geq \frac{\delta_1^k}{r_{i-1}}$ respectively, and that is not δ_1-homogeneous, and apply on it Claim 6.11. As a result, either we get for one of these blocks that at least a δ_1^k-fractions of its ordered k-column sets are (δ_1^k, k)-universal, or for every one of these blocks we get a 2^{k+1}-subpartition into subblocks of which at least an $a = (1-\delta_1)\delta_1^{k-1}$-fraction of them (in term of the total relative area) are δ_1-homogeneous.

In the former case, by Observation 6.12 we have that at least a $\frac{\delta_1^k}{r_i} \cdot \delta_1^k$-fraction of the ordered k-column sets are $(\frac{\delta_1^k}{r_i} \cdot \delta_1^k, k)$-universal, which is what we need. Hence we may assume that the latter case happens: Namely every non δ_1-homogeneous block with w_R and w_C as above is subdivided into at most $2^{2(k+1)}$ subblocks of which the total weight of δ_1-homogeneous subblocks is at least $a = (1-\delta_1)\delta_1^{k-1}$ of the weight of the block. Let x be the total weight of the δ_1-homogeneous blocks after i iterations, then the total weight at this point is at least $x + (1-x-\delta_1^k)a \geq a + x(1-a) - \delta_1^{2k-1}$. By our inductive hypothesis $x \geq 1-(1-a)^i + \alpha(1-a)^i - i\delta_1^{2k-1}$, plugging this into the above we get that this is at least $1-(1-a)^{i+1} + \alpha(1-a)^{i+1} - (i+1)\delta_1^{2k-1}$.

Having proven the above for every i we now let $i = i_0$ for which this weight is at least $1-\delta_1$. For i_0 we have $r = r_{i_0} = exp(poly(1/\delta_1))$. Hence we have nearly reached our goal. The only problem is that the resulting partition of M into blocks after the i_0'th iteration is not necessarily a cross product of a row-partition and a column-partition. To get a refinement of this partition which is a cross product, we partition the rows according to the coarsest common refinement of the row partition in each block (in other words, we partition according to the atoms of the algebra generated by the partitions of the blocks), and we do the same for the columns. Taking the cross product of these defines a legitimate partition of size $2^r = exp(exp(poly(1/\delta)))$ with at least a $1-\delta$ fraction of them being δ-homogeneous (by Observation 6.8). ∎

Proof of Lemma 6.6: Let M be as in the formulation of the Lemma, then either it has a (δ, r)-partition for $r = exp(exp(poly(1/\delta)))$ and then we are done, or, as guaranteed by Lemma 6.10, at least a $g(\delta, k)$-fraction of the ordered k-column sets of M are $(g(\delta, k), k)$-universal, for $g(\delta, k) = exp(-poly(1/\delta))$. Let $C = \{c_1, \ldots, c_k\}$ be any $(g(\delta, k), k)$-universal column set. It is quite easy to see that a row permutation of any $0/1$ $k \times k$ matrix can be found using the columns of C and choosing rows appropriately from the corresponding ϕ-sets. Hence, for such a fixed C there are at least $\binom{gn}{k} \geq (g/ek)^k n^k$ appearances of a permutation of any such $k \times k$ matrix, where $g = g(\delta, k)$. Therefore we get that there are at least a $(g/ek)^k exp(-poly(1/\delta)) = exp(-poly(1/\delta))$-fraction of the possible $k \times k$ matrices in M that are permutations of the fixed $k \times k$ matrix. ∎

6.3 (δ, r)-partitions, similarity between rows and the proof of Lemma 6.7

Our goal here is to show that a sampling of not too many entries in M can detect the signature of a (δ, r)-partition if such a partition exists. For this we need a representation of a partition by a more 'local' way, which is asserted by the following Claim 6.14 and Claim 6.16. To do this, we relate the notion of a (δ, r)-partition to relative distances between rows and columns. For the rest of this subsection we assume that $r \geq \frac{100}{\delta^{10}}$.

For two vectors $u, v \in \{0,1\}^m$ let $\mu(u, v) = \frac{1}{m}|\{i | \ u_i \neq v_i\}| = \frac{1}{m} hamming(u, v)$, namely $\mu(u, v)$ is just the normalized hamming distance.

Definition 6.13 *Let M be a $n \times n$ matrix, then $E^R(\mu(r_i, r_j))$ is the expected value of $\mu(r_i, r_j)$*

where r_i, r_j are two rows of M chosen at random. Similarly let $E^C(\mu(c_i, c_j))$ denote the respective quantity where c_i, c_j are two columns chosen at random.

One property of matrices having a (δ, r)-partition is the existence of a clustering of the rows in the following sense.

Claim 6.14 *Let M be a 0/1, $m \times m$ matrix and assume that M has a (δ, r) partition, then there is a set of r rows $R = \{r_1, \ldots, r_s\}$, $s \leq r$, such that for at least a $(1 - 2\delta^{1/4})$-fraction of the rows u of M the following holds: $\exists v \in R$, $\mu(u, v) \leq \delta^{1/4}$.*

The converse of the above also holds in the following sense:

Definition 6.15 *Let A be a 0/1, $m \times m$ matrix, $R = \{r_1, \ldots, r_s\}$ a set of rows of A, $C = \{c_1, \ldots, c_t\}$ a set of columns of A and $\delta < 1$.*

1. *R, C define a natural $(r+1)$-partition of A, where $r = max(s, t)$ in the following way: Let $R_0, \ldots, R_s$ where R_0 contains all rows u for which there is no r_i satisfying $\mu(u, r_i) \leq \delta$ and $R_i = \{u|$ such that $\mu(u, r_i) \leq \mu(u, r_j), j \neq i\}$ (if u has the same minimal distance from two different r_i's then we put it in an arbitrary one). Let $C_0, \ldots, C_t$ be the analogous column sets. We refer to this partition as the partition defined by R, C for δ.*

2. *For R as above, we say that R is δ-cover of the rows of M if for at least $(1 - \delta)$-fraction of the rows, u, of M there exists a $r_i \in R$ for which $\mu(u, r_i) \leq \delta$. Analogously C is said to be δ-cover (for the columns).*

Claim 6.16 *Let M be a 0/1, $m \times m$ matrix, $R = \{r_1, \ldots, r_s\}$ and $C = \{c_1, \ldots, c_t\}$ sets of rows, columns respectively. Then if R and C are both δ^2-covers (for rows and columns respectively) then the partition defined by R, C for δ^2 (Definition 6.15) is a $(6\delta, r+1)$-partition of M for $r = max(s, t)$.*

Moreover, let $R_0, \ldots, R_s$, $C_0, \ldots, C_t$ be as in Definition 6.15, then a 0*-signature for it is given by the following $s \times t$ matrix P and the sequences $\alpha_i = w(R_i), i = 0, ..., s$, $\beta_i = w(C_i), i = 0, ..., t$ and the (i, j) entry of P corresponds to the block $R_i \times C_j$ and its label is the $(1 - 4\delta)$-dominant label of this block if there is one or* X *otherwise.*

Before we prove the claims we need some observations:

Observation 6.17 *Let A be a 0/1 matrix, then if A is δ-homogeneous then $E^R(\mu(r_i, r_j)) \leq 2\delta$.*

Proof: As A is δ-homogeneous, we may assume w.l.o.g that A contains less then a δ fraction of 0's. Hence, choosing two rows at random and picking a random place i in both, the probability that they are not both '1' at this place is at most 2δ. Hence the expectation of the fraction of the number of places where they differ is bounded by 2δ; but this expectation is exactly $E(\mu(r_i, r_j))$. ■

Observation 6.18 *Let A be a 0/1 matrix such that $E^R(\mu(r_i, r_j)) < \delta$ and $E^C(\mu(c_i, c_j)) < \delta$, then A must be 4δ-homogeneous.*

Proof: Assume on the contrary that A is not 4δ-homogeneous. This implies that when choosing two points from A independently and uniformly at random, with probability at least 4δ they will not have the same label. This is also a lower bound on the fraction of the 2×2 submatrices that contain both 0's and 1's, as any two points with differing labels can be extended to such a submatrix. On the other hand, if $E^R(\mu(r_i, r_j)) < \delta$, then with probability more than $1 - 2\delta$ both rows of a uniformly random 2×2 submatrix are identical, as this matrix can be expressed as choosing two random places from two random rows. By the same token, if $E^R(\mu(c_i, c_j)) < \delta$ then with probability more than $1 - 2\delta$ the two columns of a random 2×2 matrix are identical. Together these would have implied that less than a 4δ fraction of the 2×2 submatrices have both 0's and 1's, a contradiction. ■

Proof of Claim 6.14: Assume that M has a (δ, r)-partition defined by the row partition $R_1, \ldots, R_s$ and the column partition $C_1, \ldots, C_t$, $s, t \leq r$. For a partition block B and a row u that intersect B, let $u|_B$ be the restriction of u to the columns in B. Assume that B is a homogeneous block that contains the rows of R_i. Then by Observation 6.17 $E(u_B, v_B) \leq 2\delta$ for two rows chosen at random from R_i. For a non δ-homogeneous block, this expression may be 1. Let $p_i = |R_i|/m$, $i = 1, \ldots, s$, and let $E_i(\mu(u, v))$ be the expectation of $\mu(u, v)$ where u, v are two rows chosen uniformly at random from R_i. Then the above implies that $\Sigma_{i=1}^{r} p_i E_i(\mu(u, v)) \leq (1 - \delta)2\delta + \delta \cdot 1 \leq 3\delta$, as this sum goes over all blocks and there are at least a $(1 - \delta)$ fraction of 0/1-blocks contributing at most 2δ each.

Now this implies that the total weight of R_i's for which $E_i(\mu(u, v)) \geq \delta^{1/2}$ is at most $3\delta^{1/2}$. Let R^* be the union of all these R_i's. Let $R_1, \ldots, R_{r'}$ be all other R_i's possibly with renumerating them. For each $i = 1, \ldots, r'$, by our assumption, $E_i(\mu(u, v)) < \delta^{1/2}$ for randomly chosen u, v, hence there is an $r_i \in R_i$ for which for at least a $(1 - \delta^{1/4})$ fraction of the v's in R_i, $\mu(r_i, v) < \delta^{1/4}$. Hence for the set $R = \{r_1, \ldots, r_{r'}\}$, $Prob(\exists v \in R,\ \mu(u, v) \leq \delta^{1/4}) \geq 1 - 3\delta^{1/2} - \delta^{1/4} \geq 1 - 2\delta^{1/4}$ (for small enough δ), as u's for which there is no such v either belongs to R^*, or are at most a $\delta^{1/4}$ fraction of $\cup_{i \geq 1} R_i$. ■

Proof of Claim 6.16: Let $R = \{r_1, \ldots, r_s\}$ be as above, let R_0 contain all rows u for which there is no r_i satisfying $\mu(u, r_i) \leq \delta^2$ and let $R_i = \{u \notin R_0 |\ such\ that\ \mu(u, r_i) \leq \mu(u, r_j), j \neq i\}$ (if u has the same minimal distance from two different r_i's then we put it in a arbitrary one). Let $C_0, \ldots, C_t$ be the analogous column sets.

By the assumptions of the claim, $|R_0| < \delta^2$. Also, for any $i \geq 1$ and any two rows $u, v \in R_i$, $\mu(u, v) \leq 2\delta^2$, as $\mu(\cdot, \cdot)$ is a distance function. Thus for $i = 1, \ldots, s$ $E_i(\mu(u, v)) \leq 2\delta^2$ where E_i is the expectation when u, v are chosen at random from R_i. Hence for the above partition into rows, $\Sigma_{i=0}^{s} \frac{|R_i|}{m} E_i(\mu(u, v)) \leq 3\delta^2$ (as for each $i > 1$ the corresponding term in this average is at most $2\delta^2$ and for $i = 0$ the weight of the term is at most δ^2). Similarly we get the analogous thing for columns. Let $\mathcal{P}$ be the partition of M into blocks that is defined by the cross product of the two partitions above.

Recall that $\frac{|R_i|}{m}, \frac{|C_i|}{m}$ are the weights $w(R_i), w(C_i)$ of the corresponding sets. Also, for a block B let $E_R(\mu(u|_B, v|_B)), E_C(\mu(u|_B, v|_B))$ be the expectation of $\mu(\cdot,\cdot)$ for two rows u, v, respectively columns, chosen at random from B. By linearity of expectation, $\Sigma_{i=0}^{s} w(R_i) \cdot E_i(\mu(u,v)) = E_B(E_R(\mu(u|_B, v|_B)))$ where in the right hand side the outer expectation is on blocks of $\mathcal{P}$ chosen according to their weights, and the inner expectation is on rows chosen at random in the block. Hence, the fact that $\Sigma_{i=0}^{s} w(R_i) E_i(\mu(u,v)) \leq 3\delta^2$ implies that the total weight of all blocks B for which $E_R(\mu(u|_B, v|_B)) > \delta$, is bounded by 3δ. By the same argument, for at most a 3δ fraction of the blocks $E_C(\mu(u|_B, v|_B)) > \delta$. Hence, for at least a $1-6\delta$ fraction of the blocks (weighted by the block weights) both $E_R(\mu(u|_B, v|_B)) \leq \delta$ and $E_C(\mu(u|_B, v|_B)) \leq \delta$. However, by the Observation 6.18 above, each such block is 4δ-homogeneous, hence at most a 6δ fraction of the blocks (measured by weights) are 4δ homogeneous. This implies that $\mathcal{P}$ is an $(6\delta, r+1)$-partition. Also, by definition, the pattern of this partition is, for each block, the $(1-4\delta)$-dominant label of this block if there is one or X otherwise. Moreover, as α_i, β_i are the exact weights of the parts in the partition, we get a 0-signature of it. ∎

We are now ready to prove Lemma 6.7.

Proof: We first start with the following partial task: Let $R' = \{r'_1, ..., r'_s\}, C' = \{c'_1, ..., c'_t\}$ be sets of rows and columns, respectively. Let $\delta_1 < 1$ and $r \geq max(s', t')$ so that $r > \frac{\gamma}{\delta_1^{10}}$. We want to check whether R', C' are δ_1-covers (for rows and columns respectively). Here is a test that approximates this with high probability in the following sense: If either R' or C' is not a $3\delta_1$-cover it will reject with high probability. If on the other hand, both R', C' are δ_1-covers, the test will accept with hight probability. Furthermore, in the latter case it will also find a γ-signature of a $(12\delta_1^{1/2}, r+1)$-partition of M, for an arbitrary given constant $\frac{1}{r^{1/2}} \leq \gamma \leq 1$:

Test 1

1. Choose randomly and independently r^4 rows and r^4 columns of M, naming the resulting sets R'' and C'' respectively. For each row compute the relative weight $w(R''_0) = \frac{|R''_0|}{|R''|}$ of the set $R''_0 = \{u \in R'' | \; \forall i \; \mu(r'_i, u) \geq 2\delta_1\}$. If $w(R''_0) \geq 2\delta_1$ then reject the test. An exact computation of $\mu(r_i, u)$ actually requires many queries, but we shall use instead an approximation which is outlined below.

   Similarly, C''_0 is computed (or approximated) for columns and the test here rejects if $w(C''_0) \geq 2\delta_1$

2. If the input is not rejected by the previous item, we compute for every i the set $R''_i = \{u \in R'' - R''_0 | \; \textit{such that} \; \mu(u, r'_i) \leq \mu(u, r'_j), j \neq i\}$, and similarly C''_i. For each R''_i (C''_i) if $w(R''_i) < \frac{\gamma}{50r}$ we discard it and put all its members into R''_0 (C''_0). Hence we get a sequence $w(R''_0), w(R''_1), ..., w(R''_s)$ (and similar for columns) so that for each $i \geq 1$, $w(R''_i) \geq \frac{\gamma}{50r}$.

   Let $(P)_{i,j}$, $(i,j) \in [s] \times [t]$, be the following matrix: To set $P_{i,j}$, for each row in R''_i we choose a random coordinate by choosing a random member of C''_j. We then set $P_{i,j}$ to be the $(1-4\delta_1^{1/2})$-dominant label of this set if there is one, or X if this set has no dominant label.

We still need to specify how the approximation of $\mu(\cdot,\cdot)$ is done in the test: To compute it for two rows r_i and $u \in R''$ we just choose at random r^2 positions, $\{p_1, ..., p_{r^2}\}$, and let $\tilde{\mu}(r_i, u) =$

$\frac{1}{r^2}\sum_{k=1}^{r^2}|r_i(p_k)-u(p_k)|$ be our estimate for $\mu(r_i,u)$, namely, we look only on the difference in the coordinates that corresponds to the columns in C''.

First let us note that for any r_i,u, $|\mu(r_i,u)-\tilde{\mu}(r_i,u)|\le\frac{\gamma}{100r}$ with probability $1-exp(-\gamma r)$ for any $\gamma<1$ (by Chernoff type bound). Assume that $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$ are δ_1-covers (for rows and columns respectively). Namely, the set R'_0, that contains all rows u of M for which $\mu(r_i,u)\ge\delta_1$ for all r_i, has $w(R'_0)\le\delta_1$. Let R^*_0 be the set of all rows u of M for which $\tilde{\mu}(r_i,u)\ge\delta_1+\frac{\gamma}{100r}$ for all r_i, then by the above $w(R^*_0)\le\delta_1$ with probability $1-exp(-\gamma r)$. Now, $R''_0=R''\cap R^*_0$, hence with probability $1-exp(-\gamma r)$, $\frac{|R''\cap R^*_0|}{r^2}=w(R''_0)\le\delta_1+\frac{\gamma}{100r}$. Thus with probability $1-exp(-\gamma r)$ Test 1 will not reject $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$. On the other hand, if one of R',C' is not a $3\delta_1$-cover then by the same reasoning Test 1 will reject with very high probability.

Now it remains to be seen that the signature is computed correctly. Indeed, under the assumption that $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$ are both δ_1-covers, Claim 6.16 asserts that $\{w(R'_0),...,w(R'_s)\}$ and $\{w(C'_0),...,w(C'_t)\}$ are the weight sequences of a 0-signature of the partition associated with $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$ and δ_1 (where R'_i,C'_i are as in Definition 6.15). Then similarly to the paragraph above, as $\tilde{\mu}$ approximates μ and a random sampling approximates densities, it follows that with probability $1-exp(-\gamma r)$ for every i $|w(R''_i)-w(R_i)|\le\frac{\gamma}{50}$ and similarly for the weights of the column sets.

Furthermore, by Claim 6.16, the pattern of the partition defined by $R'_i,\ i=0,...,s$ and $C'_j,\ j=0,...,t$ has $P_{i,j}$ equal to the $(1-2\delta_1^{1/2})$-dominant label of the submatrix defined by R'_i,C'_j if there is one, or X otherwise. Again, by the same reasoning as above, the density of 1's in such a block is well approximated (within an additive arbitrary small constant) by randomly sampling r^2 points, as done by Test 1, in the approximated block (hence allowing for some 'slack' in the 'dominance' requirement). Without giving further details here, We conclude that pattern that we get from Test 1 is a γ signature of a $(12\delta_1^{1/2},r+1)$-partition.

Clearly Test 1 uses $poly(r)$ queries and if $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$ are δ_1-covers then with high probability Test 1 does not reject. Furthermore, the signature computed is a γ-signature of a $(12\delta_1^{1/2},r+1)$-partition of M. On the other hand if one of $\{r'_1,...,r'_s\},\{c'_1,...,c'_t\}$ is not a $4\delta_1$-cover, it rejects with high probability.

We now use Test 1 to end the proof of the lemma: Assume that M has a (δ,r)-partition, then by Claim 6.14 there is a set of $s\le r$ rows $R=\{r_1,\dots,r_s\}$, and a set of $t\le r$ columns, $C=\{c_1,...,c_t\}$ that is $\delta_1=2\delta^{1/4}$-cover. By Claim 6.16 (with δ_1) the above implies that the partition that is defined by R,C is an $(6\delta_1^{1/2},r+1)$-partition. Let $R_0,...,R_s$ and $C_0,...,C_t$, be this partition (Definition 6.15).

Our aim now is to find the sets $\{r_1,...,r_s\},\{c_1,...,c_t\}$ and to find the corresponding signature. Recall that all we know is r and δ.

We choose at random r^4 rows and r^4 columns of M. Then by a Chernoff type bound, we know that with high probability we will have at least one representative from each R_i and C_j that has a weight larger than $\frac{\gamma}{20r}$. Hence these two sets certainly are $(\delta_1+\gamma)$-covers (this is true because μ is a distance function). We then apply Test 1 above for these sets. As argued, with very high probability Test 1 will assert that the chosen rows and columns are $(\delta_1+\gamma)$-covers and find a γ-signature of a $(12(\delta_1+\gamma)^{1/2},r^4)$-partition of M. ■

7 Concluding remarks

We have seen that $\forall$-poset properties are testable, for $0/1$ matrices as well as matrices over any fixed finite alphabet, while some $\forall\exists$-poset properties are not testable. In addition we have seen that bipartite graph properties that are characterized by a finite set of forbidden induced subgraph are efficiently testable. This last model is a restricted case of the model of 'submatrix' which lies in between the $\forall$-poset and the $\forall\exists$-poset models. There are some major interesting problems that remain open regarding these models:

The situation with the interesting model of 'submatrix' is not yet understood. We showed that an relatively efficient test (with dependence on $1/\epsilon$ that is better then a tower) exists for the permutation invariant case using a 'conditional' Regularity Lemma (Lemma 6.6). While we can prove a conditional Regularity Lemma, similar to Lemma 6.6 for 2 dimensional matrices in general, we don't know yet how to construct a tester for this case. Another open question is to generalize the results for the binary case to fixed, non binary alphabet. The most important question concerning this model is to obtain results for dimension higher than two, as this would related to testing of hypergraphs.

Back to the $\forall$-poset model: It would be nice to make the tests more efficient, especially in the case of non-binary alphabet. Another open problem is to better understand the model $\exists\forall$-poset. This latter model is related to some colorability problem in the spirit of [9], and currently the question whether properties in this model are testable is open.

Finally, other relations, apart of row, col (for 2-dim.) and the order may be used e.g: $succ_R(x_1, x_2)$ stating that x_2 is on the same row as x_1 and directly at the right of x_1, and similarly $succ_C(x_1, x_2)$ for columns. The relations $row(x_1, x_2) \vee col(x_1, x_2)$ and $succ_R(x_1, x_2) \vee succ_C(x_1, x_2)$ are both expressible by "$\forall$" formulae using the basic poset-model relations, Hence "$\forall$" properties using them are all $\forall\exists$-poset properties (but not necessarily $\forall$-poset properties). The model that contains $succ_R(\cdot,\cdot), succ_C(\cdot,\cdot)$ formulae with one $\forall$ quantifier generalizes the model 'submatrices' in the spirit of regular languages, namely, it can specify a submatrix with additional requirements that some columns/rows are adjacent in the whole matrix. We currently have no results for this model.

References

[1] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs. *Combinatorica* 20:451–476, 2000.

[2] N. Alon, M. Krivelevich, I. Newman and M. Szegedy, Regular Languages are Testable with a Constant Number of Queries, In 40^{th} *IEEE FOCS Conference Proceedings*, pages 645–655, 1999.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy, Proof verification and the hardness of approximation problems, *JACM*, 45(3):501–555, 1998.

[4] T. Batu, R. Rubinfeld, P. White, Fast Approximate PCPs for Multidimensional Bin-Packing Problems, In *RANDOM-APPROX99 Conference Proceedings*, pages 245–256, 1999.

[5] M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems. *JCSS*, 47:549–595, 1994.

[6] Y. Dodis, O. Goldreich, E. Lehman, S. Raskhodnikova, D. Ron and A. Samorodnitsky, Improved testing algorithms for monotonicity, In *3rd RANDOM Conference Proceedings*, pages 97–108, Springer-Verlag, Berkeley CA, 1999.

[7] F. Ergün, S. Kannan, S. R. Kumar, R. Rubinfeld and M. Viswanathan, Spot checkers, *JCSS*, 60(3):717–751, 2000.

[8] E. Fischer, On the strength of comparisons in property testing, manuscript.

[9] E. Fischer, Testing graphs for colorability properties, In 12^{th} *SODA Conference Proceedings*, pages 873–882, 2001.

[10] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan and A. Wigderson, Self-testing/correcting for polynomials and for approximate functions, In 23^{th} *ACM STOC Conference Proceedings*, pages 32–42, 1991.

[11] T. Kövary, V.T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.*, 3:50–57, 1954.

[12] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connections to learning and approximation. *JACM*, 45(4):653–750, 1998.

[13] O. Goldreich and D. Ron, Property testing in bounded degree graphs. In 29^{th} *ACM STOC Conference Proceedings*, pages 406–415, 1997.

[14] I. Newman, Testing of Function that have small width Branching Programs, In 41^{st} *IEEE FOCS Conference Proceedings*, pages 251– 258, 2000.

[15] R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing. *SIAM J. of Computing*, 25(2):252–271, 1996.

[16] A. C. Yao, Probabilistic computation, towards a unified measure of complexity. In 18^{th} *IEEE FOCS Conference Proceedings*, pages 222–227, 1977.

[17] K. Zarankiewicz, Problem P 101. *Colloq. Math.*, 2:116–131, 1951.