



Artem Chernikov · Sergei Starchenko

Regularity lemma for distal structures

Received August 24, 2015

Abstract. It is known that families of graphs with a semialgebraic edge relation of bounded complexity satisfy much stronger regularity properties than arbitrary graphs, and can be decomposed into very homogeneous semialgebraic pieces up to a small error (see e.g. [33, 2, 16, 18]). We show that similar results can be obtained for families of graphs with the edge relation uniformly definable in a structure satisfying a certain model-theoretic property called distality, with respect to a large class of generically stable measures. Moreover, distality characterizes these strong regularity properties. This applies in particular to graphs definable in arbitrary o-minimal structures and in p -adics.

Keywords. NIP, VC-dimension, distal theories, o-minimality, p -adics, Erdős–Hajnal conjecture, regularity lemma

1. Introduction

In this paper by a graph we always mean an *undirected* graph, i.e. a graph $G = (V, E)$ consists of a set of vertices V together with a *symmetric* set of edges $E \subseteq V \times V$.

As usual, we say that a subset $V_0 \subseteq V$ is *homogeneous* if either $(v, v') \in E$ for all $v \neq v' \in V_0$ or $(v, v') \notin E$ for all $v \neq v' \in V_0$, i.e. the induced graph on V_0 is either complete or empty (we ignore the diagonal).

A classical theorem of Erdős–Szekeres [15] states that every graph on n vertices contains a homogeneous subset of size at least $\frac{1}{2} \log n$ (all log's are to base 2), and this bound is tight up to a constant multiple.

Since the families of graphs with a forbidden induced subgraph have much stronger structural properties than arbitrary graphs, they have much bigger homogeneous subsets.

Theorem 1.1 (Erdős–Hajnal [13]). *For any finite graph H there is a constant $c = c(H) > 0$ such that every H -free finite graph on n vertices contains a homogeneous subset of size at least $e^{c\sqrt{\log n}}$.*

A. Chernikov: Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095-1555, USA; e-mail: chernikov@math.ucla.edu

S. Starchenko: Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556, USA; e-mail: Starchenko.1@nd.edu

Mathematics Subject Classification (2010): Primary 03C45, 03C98, 05C35, 05C69, 05D10, 05C25; Secondary 14P10, 03C64

However, the following conjecture is widely open (see e.g. [3, 10]).

Erdős–Hajnal Conjecture. For every finite graph H there is a constant $\delta = \delta(H) > 0$ such that every H -free graph on n vertices has a homogeneous subset of size at least n^δ .

In the bi-partite case one has better bounds. Let $G = (V, E)$ be a graph. We say that a pair of subsets $V_1, V_2 \subseteq V$ is *homogeneous* if either $V_1 \times V_2 \subseteq E$ or $(V_1 \times V_2) \cap E = \emptyset$.

Theorem 1.2 (Erdős, Hajnal and Pach [14]). *For any finite graph H there is a constant $\delta = \delta(H) > 0$ such that every H -free graph on n vertices has a homogeneous pair V_1, V_2 with $|V_1|, |V_2| \geq n^\delta$.*

The following definition is taken from [17].

Definition 1.3. Let \mathcal{G} be a class of finite graphs.

- (1) \mathcal{G} has the *Erdős–Hajnal property* if there is $\delta > 0$ such that every $G = (V, G) \in \mathcal{G}$ has a homogeneous subset V_0 of size $|V_0| \geq |V|^\delta$.
- (2) \mathcal{G} has the *strong Erdős–Hajnal property* if there is $\delta > 0$ such that every $G \in \mathcal{G}$ has a homogeneous pair V_1, V_2 with $|V_1|, |V_2| \geq \delta|V|$.

Remark 1.4. It is shown in [2] that if a family \mathcal{G} of finite graphs has the strong Erdős–Hajnal property and is closed under taking induced subgraphs then it has the Erdős–Hajnal property.

In this paper we consider families of graphs whose edge relations are given by a fixed definable relation in a first-order structure.

Definition 1.5. Let \mathcal{M} be a first-order structure and $R \subseteq M^k \times M^k$ be a definable relation. Consider the family \mathcal{G}_R of all finite graphs $V = (G, E)$ where $G \subseteq M^k$ is a finite subset and $E = (V \times V) \cap R$. We say that R satisfies the *(strong) Erdős–Hajnal property* if the family \mathcal{G}_R does.

We extend this notion to the bi-partite case.

Definition 1.6. Let \mathcal{M} be a first-order structure and $R \subseteq M^m \times M^n$ a definable relation.

- (1) A pair of subsets $A \subseteq M^m, B \subseteq M^n$ is called *R -homogeneous* if either $A \times B \subseteq R$ or $(A \times B) \cap R = \emptyset$.
- (2) We say that the relation R satisfies the *strong Erdős–Hajnal property* if there is a constant $\delta = \delta(R) > 0$ such that for any finite subsets $A \subseteq M^m$ and $B \subseteq M^n$ there are $A_0 \subseteq A$ and $B_0 \subseteq B$ such that $|A_0| \geq \delta|A|$ and $|B_0| \geq \delta|B|$, and the pair A_0, B_0 is R -homogeneous.

Our motivation for this work comes from the following remarkable theorem by Alon et al.

Theorem 1.7 ([2, Theorem 1.1]). *If $R \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a semialgebraic relation then R has the strong Erdős–Hajnal property.*

Remark 1.8. (1) Although this is not stated explicitly in [2], but can be easily derived from the proof, homogeneous pairs in the above theorem can be chosen to be relatively uniformly definable.

(2) The above theorem was generalized by Basu [5] to (topologically closed) relations definable in arbitrary o-minimal expansions of real closed fields.

Besides the Erdős–Hajnal property for semialgebraic graphs, the above theorem has many other applications, including those in unit distance problems [32], improved bounds in the higher dimensional semialgebraic Ramsey theorem [11], [2, Theorem 1.2], algorithmic property testing [18], and can also be used to obtain a strong Szemerédi-type regularity lemma for semialgebraic graphs [16, 18] (see also Section 5).

The aim of this article is to demonstrate that the above result by Alon et al., along with its key implications, holds at a much larger level of generality, namely for families of graphs whose edge relation is definable in a structure satisfying a certain model-theoretic property called *distality* (see Section 2.3) and with respect to the class of so-called *generically stable* measures (as opposed to just the counting ones, see Section 2.4). In particular, this applies to graphs definable in arbitrary o-minimal structures and in p -adics with analytic expansions, with respect to the Lebesgue (respectively, Haar) measure on a compact interval (respectively, compact ball).

The following is one of the key results of our paper (see Theorem 3.1).

Theorem 1.9. *Let \mathcal{M} be a distal structure and $R \subseteq M^n \times M^m$ a definable relation. Then there is a constant $\delta = \delta(R) > 0$ such that for any generically stable measures μ_1, μ_2 on M^n and M^m respectively there are definable sets $A \subseteq M^n$ and $B \subseteq M^m$ such that $\mu_1(A) \geq \delta$, $\mu_2(B) \geq \delta$, and the pair A, B is R -homogeneous.*

Remark 1.10. It is not hard to see that our Theorem 1.9 implies Theorem 1.7, by taking \mathcal{M} to be the ordered field of real numbers, and considering measures concentrated on finite sets. Thus distal structures provide a natural framework for a model-theoretic approach to Ramsey-type results in geometric combinatorics.

Remark 1.11. It is demonstrated by Malliaris–Shelah [29] (see also [9] for an alternative proof) that if \mathcal{M} is a stable structure and $R \subseteq M^k \times M^k$ is a definable relation then the family \mathcal{G}_R of finite graphs has the Erdős–Hajnal property. However, in general this family does not have the *strong* Erdős–Hajnal property (see Section 6.1).

Remark 1.12. Our proof of Theorem 1.9 (and the density version in Corollary 4.6) gives explicit bounds on δ and the number of parameters in the definitions of A and B in terms of the VC-density of the edge relation R . In particular, for o-minimal structures and for p -adics, we obtain a bound in terms of the number of variables involved in R , due to the corresponding bounds for VC-density from [4]. We were informed by Pierre Simon that after reading our paper he found another proof of this result which is faster, but does not give bounds [40].

Here is a brief summary of the paper. In Section 2 we introduce the context and the notation: first-order structures and definable sets, distality, Keisler measures and generic

stability. In Section 3 we prove a definable generalization of Theorem 1.7 for bi-partite graphs. In Section 4 we improve it to a density version, which enables us to obtain an analogue for hypergraphs and a version allowing additional parameters in the definition of the edge relation. This gives in particular a lot of new families of graphs satisfying the strong Erdős–Hajnal property (see Example 4.11). In Section 5 we obtain a strong regularity lemma for hypergraphs definable in distal structures, generalizing the result for semialgebraic hypergraphs from [16, 18].

In Section 6 we consider the converse to our results from the previous sections. First, in Section 6.1 we demonstrate a very explicit failure of the definable counterpart of Theorem 1.7 in the theory of algebraically closed fields of positive characteristic, even without requiring definability of the homogeneous subsets. It turns out in particular that every field interpretable in a distal structure is of characteristic 0. In Section 6.2 we prove that distality of a structure is in fact equivalent to the definable counterpart of Theorem 1.7.

Some further questions concerning incidence phenomena and higher-dimensional Ramsey theory in our setting will be addressed in a future paper.

2. Preliminaries

2.1. Model-theoretic notation

We assume familiarity only with the very basic notions of model theory such as first-order structures and formulas that can be found in any introductory model theory book (e.g. [30]). By a *structure* we always mean a first-order structure.

Our notations are standard. We will denote first-order structures by script letters \mathcal{M}, \mathcal{N} etc., and use M, N etc. to denote their underlying sets. Very often we will not distinguish singletons and tuples: e.g. we may use x to denote a tuple of variables (x_1, \dots, x_n) , a to denote an element of M^n , and $|x|$ to denote the length of the tuple $x = (x_1, \dots, x_n)$.

If \mathcal{M} is a structure and $\phi(x, y)$ is a formula in the language of \mathcal{M} , then for $a \in M^{|y|}$, as usual, we denote by $\phi(M, a)$ the subset of $M^{|x|}$ defined by $\phi(x, a)$, that is, $\phi(M, a) = \{b \in M^{|x|} : \mathcal{M} \models \phi(b, a)\}$.

A subset $X \subseteq M^n$ is called *definable* if there is a formula $\phi(x, y)$ and $a \in M^{|y|}$ such that $X = \phi(M, a)$; if we want to specify the set of parameters, then for $A \subseteq M$, a definable subset $X \subseteq M^n$ is called *A-definable* (or *definable over A*) if we can choose a as above in $A^{|y|}$. Also if we want to specify ϕ we say that such a set X is *ϕ -definable*.

2.2. VC-dimension and NIP

Vapnik–Chervonenkis dimension, or VC-dimension, is an important notion in combinatorics and statistical learning theory. Let X be a set, finite or infinite, and let \mathcal{F} be a family of subsets of X . Given $A \subseteq X$, we say that it is *shattered* by \mathcal{F} if for every $A' \subseteq A$ there is some $S \in \mathcal{F}$ such that $A \cap S = A'$. A family \mathcal{F} is said to be a *VC-class* if there is some $n < \omega$ such that no subset of X of size n is shattered by \mathcal{F} . In this case the *VC-dimension* of \mathcal{F} , denoted by $\text{VC}(\mathcal{F})$, is the smallest integer n such that no subset of X

of size $n + 1$ is shattered by \mathcal{F} . For a set $B \subseteq X$, let $\mathcal{F} \cap B = \{A \cap B : A \in \mathcal{F}\}$, and let $\pi_{\mathcal{F}}(n) = \max\{|\mathcal{F} \cap B| : B \subseteq X, |B| = n\}$.

Fact 2.1 (Sauer–Shelah lemma). *If $\text{VC}(\mathcal{F}) \leq d$ then $\pi_{\mathcal{F}}(n) \leq \sum_{i \leq d} \binom{n}{i} = O(n^d)$ for all $n \geq d$.*

If $S \subseteq X$ and $x_1, \dots, x_n \in X$, we let $\text{Av}(x_1, \dots, x_n; S) = \frac{1}{n} |\{i \leq n : x_i \in S\}|$ (we do not assume that the x_i are distinct).

Fact 2.2 (VC-theorem [44]; see also [21, Section 4] for a discussion). *For any $k > 0$ and $\varepsilon > 0$ there is $n = O(k(1/\varepsilon)^2 \log(1/\varepsilon))$ satisfying the following. For any finite probability space (X, μ) and any family \mathcal{F} of subsets of X of VC-dimension $\leq k$, there are $x_1, \dots, x_n \in X$ such that for any $S \in \mathcal{F}$ we have $|\mu(S) - \text{Av}(x_1, \dots, x_n; S)| \leq \varepsilon$.*

An important class of NIP theories was introduced by Shelah in his work on the classification program [37]. It has attracted a lot of attention recently, both from the point of view of pure model theory and due to some applications in algebra and geometry. We refer to [1, 39] for an introduction to that area.

As observed early on in [25], the original definition of NIP is equivalent to the following one (see [4] for a more detailed account).

Definition 2.3. Let T be a complete theory and $\phi(x, y)$ a formula in T , where x, y are tuples of variables, possibly of different length. We say that the formula $\phi(x, y)$ is *NIP* if there is a model \mathcal{M} of T such that the family $\{\phi(\mathcal{M}, a) : a \in M^{|y|}\}$ is a VC-class. In this case we define the VC-dimension of $\phi(x, y)$ to be the VC-dimension of this class. (It is easy to see that by elementary equivalence the above does not depend on the model \mathcal{M} of T .)

A theory T is *NIP* if all formulas in T are NIP.

Slightly abusing terminology we say that a structure \mathcal{M} is NIP if its complete theory $\text{Th}(\mathcal{M})$ is NIP. Restated differently, \mathcal{M} is an *NIP structure* if for every formula $\phi(x, y)$ the family $\mathcal{F}_{\phi} = \{\phi(\mathcal{M}, a) : a \in M^{|y|}\}$ of ϕ -definable sets is a VC-class.

Given a set $\Delta(x, y)$ of formulas and a set $B \subseteq M^{|y|}$, we say that $\pi(x)$ is a Δ -type over B if $\pi(x) \subseteq \bigcup_{\phi(x, y) \in \Delta, b \in B} \{\phi(x, b), \neg\phi(x, b)\}$ and there is some $\mathcal{N} \succeq \mathcal{M}$ and some $a \in N^{|x|}$ satisfying simultaneously all formulas from $\pi(x)$. By a *complete* Δ -type over B we mean a maximal Δ -type over B . We will denote by $S_{\Delta}(B)$ the collection of all complete Δ -types over B . In view of the remarks above, the following is an immediate corollary of the Sauer–Shelah lemma.

Fact 2.4. *A structure \mathcal{M} is NIP if and only if for any finite set $\Delta(x, y)$ of formulas there is some $d \in \mathbb{N}$ such that $|S_{\Delta}(B)| = O(|B|^d)$ for any finite $B \subseteq M^{|y|}$.*

2.3. Distality

The class of distal theories is defined and studied in [38], with the aim of isolating the class of purely unstable NIP theories (as opposed to the class of stable theories which

are always NIP; see also [39]). The original definition is in terms of some properties of indiscernible sequences, but the following explicit combinatorial characterization of distality given in [8] can be used as an alternative definition.

Fact 2.5. *Let T be a complete NIP theory and \mathcal{M} a model of T . The following are equivalent:*

- (1) T is distal (in the sense of the original definition, see Fact 6.4).
- (2) For every formula $\phi(x, y)$ there is a formula $\psi(x, y_1, \dots, y_n)$ with $|y_1| = \dots = |y_n| = |y|$ such that: for any finite $B \subseteq M^{|y|}$ with $|B| \geq 2$ and any $a \in M^{|x|}$, there are $b_1, \dots, b_n \in B$ such that $\mathcal{M} \models \psi(a, b_1, \dots, b_n)$ and $\psi(x, b_1, \dots, b_n) \vdash \text{tp}_\phi(a/B)$ (i.e. for any $b \in B$ either $\phi(M, b) \supseteq \psi(M, b_1, \dots, b_n)$ or $\phi(M, b) \cap \psi(M, b_1, \dots, b_n) = \emptyset$).

Remark 2.6. It is not hard to see that if \mathcal{M} satisfies (2) for all formulas $\phi(x, y)$ with $|x| = 1$ then it satisfies it for all formulas, i.e. \mathcal{M} is distal. Besides, any \mathcal{M} satisfying (2) is automatically NIP (easy to see using the equivalence from Fact 2.4), so the assumption that \mathcal{M} is NIP in Fact 2.5 is used to deduce (2) from (1).

Remark 2.7. An immediate corollary of Fact 2.5 is that in a distal structure, for any formula $\phi(x, y)$ there is a formula $\psi'(y, y_1, \dots, y_n)$ such that for any finite $B \subseteq M^{|y|}$ with $|B| \geq 2$ and $a \in M^{|x|}$, there are some $b_1, \dots, b_n \in B$ such that $\phi(a, B) = \psi'(B, b_1, \dots, b_n)$. Namely, one can take $\psi'(y, y_1, \dots, y_n) = \forall x(\psi(x, y_1, \dots, y_n) \rightarrow \phi(x, y))$. In fact, this corollary characterizes NIP (see [8] for the details).

We list some examples of distal structures (providing more details than we normally would, for the sake of a non-model-theorist reader).

2.3.1. \mathcal{O} -minimal structures. A structure $\mathcal{M} = (M, <, \dots)$ is *\mathcal{O} -minimal* if every definable subset of M is a finite union of singletons and intervals (with endpoints in $M \cup \{\pm\infty\}$). From this assumption one obtains cell decomposition for definable subsets of M^n , for all n . Moreover, a cell decomposition of a definable set is uniformly definable in terms of its definition (see [42] for a detailed treatment of \mathcal{O} -minimality, or [36, Section 3] and references there for a quick introduction). Examples of \mathcal{O} -minimal structures include $\bar{\mathbb{R}} = (\mathbb{R}, +, \times)$, $\mathbb{R}_{\text{exp}} = (\mathbb{R}, +, \times, e^x)$, $\mathbb{R}_{\text{an}} = (R, +, \times, f \upharpoonright_{[0,1]^k})$ for f ranging over all functions real-analytic on some neighborhood of $[0, 1]^k$, or the combination of both, $\mathbb{R}_{\text{an,exp}}$. It is straightforward to verify that if \mathcal{M} is \mathcal{O} -minimal then it satisfies (2) of Fact 2.5 for all formulas $\phi(x, y)$ with $|x| = 1$.

Example 2.8 (The field of reals \mathbb{R}). By Tarski's quantifier elimination, for each n the definable subsets of \mathbb{R}^n are exactly the semialgebraic sets, that is, finite Boolean combinations of sets defined by polynomial equations $p(x_1, \dots, x_n) = 0$ and inequalities $q(x_1, \dots, x_n) > 0$ for $p(\bar{x}), q(\bar{x}) \in \mathbb{R}[x_1, \dots, x_n]$. This structure is \mathcal{O} -minimal, and so distal.

2.3.2. *Ordered dp-minimal structures.* More generally, it is proved in [38] that every ordered dp-minimal structure is distal (see Fact 6.6 below). Examples of ordered dp-minimal structures include weakly o-minimal structures and quasi-o-minimal structures. An ordered structure \mathcal{M} is *weakly o-minimal* (respectively, *quasi-o-minimal*) if in every elementary extension, every definable subset is a finite union of convex subsets (respectively, a finite boolean combination of singletons, intervals and \emptyset -definable sets [6]).

Example 2.9 (\mathbb{Z} as an ordered group). By Presburger’s quantifier elimination, for each n the definable subsets of \mathbb{Z}^n are finite boolean combinations of sets of the following types:

$$\begin{aligned} S_{\bar{a}}^= &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : a_1x_1 + \dots + a_nx_n = a_0\}, \\ S_{\bar{a}}^> &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : a_1x_1 + \dots + a_nx_n > a_0\}, \\ S_{\bar{a}}^k &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : \exists y \in \mathbb{Z} \, ky = a_0 + a_1x_1 + \dots + a_nx_n\}, \end{aligned}$$

for $\bar{a} = (a_0, a_1, \dots, a_n) \in \mathbb{Z}^{n+1}$ and $k \in \mathbb{N}$. This structure is quasi-o-minimal (see [6, Example 2]).

Example 2.10 (The valued field $\mathbb{K} = \bigcup_{n \in \mathbb{N}} \mathbb{R}((T^{1/n}))$ of Puiseux power series over \mathbb{R} , in the language $L_{\text{div}} = \{0, 1, <, +, -, \times, v(x) \leq v(y)\}$). Using quantifier elimination from [12], for each n the definable subsets of \mathbb{K}^n (in the language of valued fields) are finite boolean combinations of sets of the following types:

$$\begin{aligned} S_p^= &= \{(x_1, \dots, x_n) \in \mathbb{K}^n : p(x_1, \dots, x_n) = 0\}, \\ S_p^> &= \{(x_1, \dots, x_n) \in \mathbb{K}^n : p(x_1, \dots, x_n) > 0\}, \\ S_{p,q}^v &= \{(x_1, \dots, x_n) \in \mathbb{K}^n : v(p(x_1, \dots, x_n)) \geq v(q(x_1, \dots, x_n))\}, \end{aligned}$$

for $p, q \in \mathbb{K}[x_1, \dots, x_n]$.

This structure is a model of the complete theory RCVF of real closed fields equipped with a proper convex valuation ring, and by [12] it is weakly o-minimal.

2.3.3. *P-minimal structures with definable Skolem functions*

Example 2.11. By a result of Macintyre [28] the field of p -adics \mathbb{Q}_p eliminates quantifiers in the language $L_p = \{0, 1, +, \times, v(x) \leq v(y), P_n(x)\}$, where for $n \geq 2$ we have $P_n(x) \Leftrightarrow \exists y(x = y^n)$. It follows that for each n the definable subsets of \mathbb{Q}_p^n are finite boolean combinations of sets of the following three types:

$$\begin{aligned} S_p &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : p(x_1, \dots, x_p) = 0\}, \\ S_{p,q}^v &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : v(p(x_1, \dots, x_n)) \geq v(q(x_1, \dots, x_n))\}, \\ S_p^k &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : \exists y \in \mathbb{Q}_p \, y^k = p(x_1, \dots, x_n)\}, \end{aligned}$$

for $p, q \in \mathbb{Q}_p[x_1, \dots, x_n]$ and $k \in \mathbb{N}$.

Similarly to the o-minimal case, there is a notion of minimality for expansions of \mathbb{Q}_p . Namely, a structure \mathcal{M} in a language $L \supseteq L_p$ is *p-minimal* if in every model of $\text{Th}(\mathcal{M})$, every definable subset in one variable is quantifier-free definable just using the lan-

guage L_p [20]. P -minimal structures with an additional assumption of definability of Skolem functions satisfy an analogue of the p -adic cell decomposition of Denef. A motivating example of a p -minimal theory with definable Skolem functions is the theory $p\text{CF}_{\text{an}}$ of the field of p -adic numbers \mathbb{Q}_p expanded by all subanalytic subsets of \mathbb{Z}_p [43].

By [4, Corollary 7.8], every p -minimal structure with definable Skolem functions is dp-minimal. Since $(\mathbb{Q}_p, +, \times, 0, 1)$ is distal [38] (which can also be verified using Fact 6.6), it follows by Remark 6.7 that any p -minimal theory with Skolem functions is distal.

2.4. Keisler measures

Let \mathcal{M} be a structure. Recall that a *Keisler measure* on M^n is a finitely additive probability measure on the Boolean algebra of all definable subsets of M^n , i.e. it is a function μ that assigns to every definable $X \subseteq M^n$ a number $\mu(X) \in [0, 1]$ with $\mu(\emptyset) = 0$, $\mu(M^n) = 1$ and

$$\mu(X \cup Y) = \mu(X) + \mu(Y) - \mu(X \cap Y)$$

for all definable subsets $X, Y \subseteq M^n$. Given a formula $\phi(x)$ with parameters from M and a Keisler measure μ on $M^{|x|}$, we will write $\mu(\phi(x))$ to denote $\mu(\phi(M^{|x|}))$.

In this paper we will deal mostly with the so-called generically stable and smooth Keisler measures.

Generically stable measures on M^n are defined as Keisler measures on M^n admitting a (unique) global M -invariant extension which is both finitely satisfiable in M and definable over M (see Remark 5.12). In the NIP case, according to the following fact from [22, Theorem 3.2] (see also [39, Section 7.5]), they can also be defined in terms of the structure \mathcal{M} alone without mentioning global measures.

Fact 2.12. *Let \mathcal{M} be an NIP structure and μ a Keisler measure on M^k . Then the following are equivalent:*

- (1) *The measure μ is generically stable.*
- (2) *For every formula $\phi(x, y)$ with $|x| = k$ and $\varepsilon > 0$ there are some $a_1, \dots, a_m \in M^k$ such that $|\mu(\phi(x, b)) - \text{Av}(a_1, \dots, a_m; \phi(x, b))| < \varepsilon$ for any $b \in M^{|y|}$.*

The VC-theorem implies that in NIP theories, for any Keisler measure, a uniformly definable family of sets admits an ε -approximation by types (see [21, Section 4]). Fact 2.12 implies that with respect to a generically stable measure, there are ε -approximations by *elements* of a model rather than just by types over it. We remark that the bound on the size of ε -approximations depends just on the VC-dimension of the formula (so is uniform over all generically stable measures).

Proposition 2.13. *Let \mathcal{M} be an NIP structure. Then for any $k \in \omega$ and any $\varepsilon > 0$ there is some $n = O(k(2/\varepsilon)^2 \log(2/\varepsilon))$ such that: for any formula $\phi(x, y)$ of VC-dimension at most k and any generically stable measure μ on $M^{|x|}$, there are some $a_1, \dots, a_n \in M^{|x|}$ such that for any $b \in M^{|y|}$, $|\mu(\phi(x, b)) - \text{Av}(a_1, \dots, a_n; \phi(x, b))| < \varepsilon$.*

Proof. Let $n \in \omega$ be given for k and ε by Fact 2.2. Let $\phi(x, y)$ be a formula of VC-dimension at most k and let μ be an arbitrary generically stable measure on $M^{|x|}$. By Fact 2.12 there are some $a'_1, \dots, a'_m \in M^{|x|}$ such that for any $b \in M^{|y|}$ we have $|\mu(\phi(x, b)) - \nu(\phi(x, b))| < \varepsilon/2$, where $\nu(\phi(x, b)) = \text{Av}(a'_1, \dots, a'_m; \phi(x, b))$. Now applying Fact 2.2 to ν , we find some $a_1, \dots, a_n \in M^{|x|}$ such that for all $b \in M$ we have $|\nu(\phi(x, b)) - \text{Av}(a_1, \dots, a_n; \phi(x, b))| < \varepsilon/2$. Then

$$|\mu(\phi(x, b)) - \text{Av}(a_1, \dots, a_n; \phi(x, b))| < \varepsilon$$

for all $b \in M^{|y|}$, as wanted. □

Remark 2.14. Encoding several formulas into one we can replace a single formula $\phi(x, y)$ in Proposition 2.13 by a finite set $\Delta(x, y)$ of formulas.

Recall that a Keisler measure μ on M^n is called *smooth* if there is a unique global Keisler measure extending it. The following equivalence can be used to avoid a reference to global measures in the definition of smoothness.

Fact 2.15 ([22, Section 2]). *A Keisler measure μ on M^n is smooth if and only if the following holds. For any formula $\phi(x, y)$ with $|x| = n$ and $\varepsilon > 0$ there are some formulas $\theta_i^1(x), \theta_i^2(x)$ and $\psi_i(y)$ with parameters from M , for $i = 1, \dots, m$, such that:*

- (1) *the sets $\psi_i(M^{|y|})$ partition $M^{|y|}$,*
- (2) *for all i and $b \in M^{|y|}$, if $\mathcal{M} \models \psi_i(b)$, then*

$$\mathcal{M} \models (\theta_i^1(x) \rightarrow \phi(x, b)) \ \& \ (\phi(x, b) \rightarrow \theta_i^2(x)),$$

- (3) *for each i , $\mu(\theta_i^2(x)) - \mu(\theta_i^1(x)) < \varepsilon$.*

Every smooth measure is generically stable, and there are generically stable measures which are not smooth (though every Keisler measure in an NIP theory can be extended to a smooth one, but over a larger set of parameters). However, we have the following characterization from [38].

Fact 2.16. *Let T be NIP. Then the following are equivalent:*

- (1) *T is distal.*
- (2) *For any model \mathcal{M} of T , any generically stable measure on M^n is smooth.*

Remark 2.17. Let μ be a Keisler measure on M^n and $A \subseteq M^n$ a definable subset with $\mu(A) > 0$. Then we can localize μ to A by defining $\mu_A(X) = \mu(A \cap X)/\mu(A)$. Clearly μ_A is a Keisler measure on M^n and μ_A is generically stable (smooth) provided μ is.

Let \mathcal{M} be a structure, μ_1 a Keisler measure on M^m , and μ_2 a Keisler measure on M^n . A Keisler measure μ on M^{m+n} is called a *product measure* of μ_1 and μ_2 if for any definable subsets $X \subseteq M^m$ and $Y \subseteq M^n$ we have $\mu(X \times Y) = \mu_1(X)\mu_2(Y)$. We can extend this notion to finitely many Keisler measure μ_i on $M^{|n_i|}$ in the obvious way. A product Keisler measure always exists but in general is not unique. However, for smooth measures we have the following proposition that follows from [22, Corollary 2.5].

Proposition 2.18. *Let \mathcal{M} be a structure, μ_1 a smooth Keisler measure on M^m , and μ_2 a smooth Keisler measure on M^n . Then there is a unique product measure of μ_1 and μ_2 , and this measure is also smooth.*

If μ_1 and μ_2 are smooth measures, we will denote their unique product measure as $\mu = \mu_1 \otimes \mu_2$.

Let x_1, \dots, x_n be pairwise disjoint tuples of variables, and μ a Keisler measure on $M^{|x_1|} \times \dots \times M^{|x_n|}$. Then for each $i = 1, \dots, n$, μ induces a Keisler measure μ_i on $M^{|x_i|}$ by

$$\mu_i(Y) = \mu(M^{|x_1|} \times \dots \times M^{|x_{i-1}|} \times Y \times M^{|x_{i+1}|} \times \dots \times M^{|x_n|}),$$

and we will denote μ_i by $\mu|_{x_i}$. It is easy to see that if μ is generically stable [smooth] then every $\mu|_{x_i}$ is also generically stable [smooth]. Also in this case we will call a Keisler measure μ on $M^{|x_1|} \times \dots \times M^{|x_n|}$ a *product measure* if μ is a product of $\mu|_{x_1}, \dots, \mu|_{x_n}$.

Finally, we give some examples of smooth Keisler measures.

- Fact 2.19.** (1) *Any Keisler measure concentrated on a finite set is smooth* (as the condition (2) of Fact 2.12 is clearly satisfied).
- (2) *Let λ_n be the Lebesgue measure on the unit cube $[0, 1]^n$ in \mathbb{R}^n . Let \mathcal{M} be an o-minimal structure expanding the field of real numbers. If $X \subseteq \mathbb{R}^n$ is definable in \mathcal{M} , then, by o-minimal cell decomposition, $X \cap [0, 1]^n$ is Lebesgue measurable, hence λ_n induces a Keisler measure on M^n . This measure is smooth* (by [22, Section 6]).
- (3) *Similarly to (2), for every prime p the (normalized) Haar measure on a compact ball in \mathbb{Q}_p induces a smooth Keisler measure on \mathbb{Q}_p^n* (see [22, Section 6]).
- (4) *Any definable, definably compact group G in an o-minimal structure or over the p -adics admits a unique G -invariant generically stable measure [21, 7], which is then smooth* (by distality and Fact 2.16).

3. Strong Erdős–Hajnal for definable bi-partite graphs in distal theories

In this section we prove the key result of this paper.

Theorem 3.1. *Let \mathcal{M} be a model of a distal theory, and $R \subseteq M^n \times M^m$ a definable relation. Then there is a constant $\delta = \delta(R) > 0$ and a pair of formulas $\psi_1(x, z_1), \psi_2(y, z_2)$ such that for any generically stable measures μ_1, μ_2 on M^n and M^m respectively, there are c_1, c_2 from \mathcal{M} such that $\mu_1(\psi_1(M, c_1)) \geq \delta$, $\mu_2(\psi_2(M, c_2)) \geq \delta$, and the pair $A = \psi_1(M, c_1)$, $B = \psi_2(M, c_2)$ is R -homogeneous.*

As in [2], the above theorem will follow from an asymmetric version (see Theorem 3.6 below).

Let \mathcal{M} be a distal structure and fix a formula $\phi(x, y)$. Let $\psi(x, y_1, \dots, y_l)$ be as given for $\phi(x, y)$ by Fact 2.5.

For $\vec{d} = (d_1, \dots, d_l)$ we will denote by $C_{\vec{d}}$ the subset of $M^{|x|}$ defined by $\psi(x, \vec{d})$, and call it a *chamber*. If in addition $\vec{d} \in B^l$ then we say that $C_{\vec{d}}$ is a *B -definable chamber*.

Definition 3.2. For a chamber $C = C_{\bar{d}}$ and $b \in M^{|y|}$ we say that $\phi(x, b)$ crosses C if both $C \cap \phi(M, b)$ and $C \cap \neg\phi(M, b)$ are nonempty.

For a chamber C and a set B we will denote by $C^\#(B)$ the set of all $b \in B$ such that $\phi(x, b)$ crosses C . Note that $C^\#(M)$ is a definable set (by a formula depending just on the formula ψ defining C).

Definition 3.3. For $B \subseteq M^{|y|}$, a chamber C is called B -complete if C is B -definable and $C^\#(B) = \emptyset$.

It follows from the choice of ψ that for every finite $B \subseteq M^{|y|}$ and $a \in M^{|x|}$ there is a B -complete chamber C with $a \in C$. In particular, for any finite B the union of all B -complete chambers covers $M^{|x|}$.

Definition 3.4 ($1/r$ -cutting). Let ν be a Keisler measure on $M^{|y|}$. Adopting a definition from [31], for a positive $r \in \mathbb{R}$ we say that a family \mathcal{F} of chambers is a $1/r$ -cutting with respect to ν if $M^{|x|}$ is covered by $\{C : C \in \mathcal{F}\}$ and for every $C \in \mathcal{F}$ we have $\nu(C^\#(M)) \leq 1/r$.

The following claim is an analogue of a cutting lemma from [31] (see also Exercise 10.3.4(b) there).

Claim 3.5. *There is a constant K such that the following holds. For any positive r and for any generically stable measure ν on $M^{|y|}$ there is a finite set $S \subseteq M$ such that the family of all S -complete chambers is a $1/r$ -cutting with respect to ν , and the size of S is bounded by $Kr^2 \log 2r$.*

Proof. Consider the family of sets

$$C = \{C^\#(M) : C \text{ is an } M\text{-definable chamber}\}.$$

It is a definable family, hence has a bounded VC-dimension by NIP. Applying Proposition 2.13 with $\varepsilon = 1/r$ we obtain a subset $S \subseteq M$ of size at most $Kr^2 \log 2r$, where K is a constant that depends only on the VC-dimension of \mathcal{C} , such that for every M -definable chamber C , if $\nu(C^\#(M)) > 1/r$ then $S \cap C^\#(M) \neq \emptyset$.

Since $C^\#(S) = \emptyset$ for any S -complete cell C , we are done. □

The following theorem is an analogue of a result in [2, Section 6].

Theorem 3.6. *Let \mathcal{M} be a distal structure and let $R(x, y)$ be a definable relation. Then for any $\beta \in (0, 1/2)$ there are some $\alpha \in (0, 1)$ and formulas $\psi_1(x, z_1)$, $\psi_2(y, z_2)$ depending just on R and β such that: for any Keisler measure μ on $M^{|x|}$ and any generically stable measure ν on $M^{|y|}$, there are some $c_1 \in M^{|z_1|}$ and $c_2 \in M^{|z_2|}$ such that $\mu(\psi_1(x, c_1)) > \alpha$, $\nu(\psi_2(y, c_2)) > \beta$ and the pair of sets $\psi_1(M, c_1)$, $\psi_2(M, c_2)$ is R -homogeneous.*

Proof. Let $\phi(x, y)$ be a formula defining R . Let r be a positive real number to be determined later.

By Claim 3.5, let $S \subseteq M^{|y|}$ be a set of size at most $Kr^2 \log 2r$ such that for every S -complete chamber C we have $\nu(C^\#(M)) \leq 1/r$. It is not hard to see that there is a constant K_1 and a number $l = l(\psi) \in \mathbb{N}$ such the number of S -definable chambers is at most $K_1|S|^l$. Thus the number of S -complete chambers is at most $K'r^{2l} \log^l 2r$, where K' is a constant.

As the set of S -complete chambers covers $M^{|x|}$, there is an S -complete chamber C_0 with $\mu(C_0) \geq 1/(K'r^{2l} \log^l(2r))$.

For the set $D = M^{|y|} \setminus C_0^\#(M)$, we have $\nu(D) \geq 1 - 1/r$, and for every $d \in D$, $\phi(x, d)$ does not cross C_0 . In particular all $a \in C_0$ have the same ϕ -type over D . Note that D is a disjoint union of $D_1 = \{d \in D : C_0 \subseteq \phi(M, d)\}$ and $D_2 = \{d \in D : C_0 \cap \phi(M, d) = \emptyset\}$, and both (C_0, D_1) and (C_0, D_2) are R -homogeneous. Thus either $\nu(D_1) \geq \frac{1}{2} - \frac{1}{2r}$ or $\nu(D_2) \geq \frac{1}{2} - \frac{1}{2r}$. Let $\psi_1 := \psi$ be the formula such that an instance of it defines C_0 , and let ψ_2 be the formula such that an instance of it defines either D_1 or D_2 , depending on which one has larger measure. By assumption there are only finitely many choices for both depending on the original data. So given $\beta \in (0, 1/2)$ we can find r with $\frac{1}{2} - \frac{1}{2r} = \beta$, and take any positive $\alpha < 1/(K'r^{2l} \log^l(2r))$. Then, encoding finitely many choices for ψ_1, ψ_2 into one formula, we can conclude the theorem. \square

Proof of Theorem 3.1. We can take any $\beta \in (0, 1/2)$ and let α be as in Theorem 3.6. Now take $\delta = \min\{\alpha, \beta\}$. \square

Remark 3.7. We will see in Corollary 4.6 that one can allow an extra parameter in R without affecting the uniform choice of ψ_1, ψ_2 .

4. Density version and a generalization to hypergraphs

First we prove that Theorem 3.1 can be strengthened to a density version. It seems that this implication is folklore, as it is mentioned in [2, Corollary 7.1 and the remark afterwards] without definability of the homogeneous subsets, and stated in [16]. However, the proofs in both places are very sketchy, so we give a complete proof verifying definability of the homogeneous subsets, and in addition working with Keisler measures. Our argument is an elaboration on the proof of [33, Theorem 3.3].

Proposition 4.1. *Let \mathcal{M} be a distal structure and $R(x, y)$ a definable relation. Given $\alpha > 0$ there is $\varepsilon > 0$ such that for any Keisler measure μ on $M^{|x|}$, any generically stable measure ν on $M^{|y|}$, and a product measure ω of μ and ν , if $\omega(R(x, y)) \geq \alpha$ then there are uniformly definable (in terms of α and R only) $A_0 \subseteq M^{|x|}$ and $B_0 \subseteq M^{|y|}$ with $\mu(A_0) \geq \varepsilon$, $\nu(B_0) \geq \varepsilon$, and $A_0 \times B_0 \subseteq R$.*

We fix a distal structure \mathcal{M} and a definable relation $R(x, y)$. By Theorem 3.1 there is a constant $\delta > 0$ and formulas $\psi_1(x, z_1)$ and $\psi_2(y, z_2)$ such that for any measure μ on $M^{|x|}$ and any generically stable measure ν on $M^{|y|}$ there are some $A \subseteq M^{|x|}$ and $B \subseteq M^{|y|}$ definable by an instance of ψ_1 and ψ_2 respectively, with $\mu(A), \nu(B) \geq \delta$, such that either $A \times B \subseteq R$ or $A \times B \cap R = \emptyset$.

Now we fix Keisler measures μ, ν as in the proposition and let ω be a product Keisler measure of μ, ν on $M^{|x|+|y|}$.

For definable sets $A \subseteq M^{|x|}$ and $B \subseteq M^{|y|}$ we denote by $d(A, B)$ the density of R in $A \times B$, that is,

$$d(A, B) = \frac{\omega((A \times B) \cap R)}{\mu(A)\nu(B)},$$

and we set $d(A, B) = 0$ if $\mu(A)\nu(B) = 0$. The following claim is a basic step.

Claim 4.2. *Let $A \subseteq M^{|x|}$ and $B \subseteq M^{|y|}$ be definable sets with $d(A, B) > 1 - \delta^2$. Then there are subsets $A_1 \subseteq A$ and $B_1 \subseteq B$ defined uniformly (in terms of A, B and R) such that $\mu(A_1) \geq \delta\mu(A)$, $\nu(B_1) \geq \delta\nu(B)$ and $A_1 \times B_1 \subseteq R$.*

Proof. Using Remark 2.17 we apply Theorem 3.1 to μ_A and ν_B , the localizations of μ on A and of ν on B , respectively. This gives us $A' \subseteq M^{|x|}$ and $B' \subseteq M^{|y|}$ defined by instances of ψ_1 and ψ_2 respectively such that for the sets $A_1 = A' \cap A$ and $B_1 = B' \cap B$ we have $\mu(A_1) \geq \delta\mu(A)$, $\nu(B_1) \geq \delta\nu(B)$ and either $A_1 \times B_1 \subseteq R$ or $(A_1 \times B_1) \cap R = \emptyset$.

If $(A_1 \times B_1) \cap R = \emptyset$ then

$$\omega((A \times B) \cap R) \leq \omega(A \times B) - \omega(A_1 \times B_1) \leq (1 - \delta^2)\mu(A)\nu(B),$$

contradicting the assumption $d(A, B) > 1 - \delta^2$. □

It is not hard to see that Proposition 4.1 follows from Claim 4.2 and the following claim by iterating sufficiently (but boundedly) many times and taking the conjunction of the corresponding defining formulas.

Claim 4.3. *For any $0 < \alpha < 1 - \delta^2$ there is some $h > 0$ such that for any definable A and B with $d(A, B) \geq \alpha$ there are uniformly definable (in terms of R, α, A, B) subsets $A' \subseteq A$ and $B' \subseteq B$ with $\mu(A') \geq h\mu(A)$, $\nu(B') \geq h\nu(B)$ and $d(A', B') \geq \alpha/(1 - \delta^2)$.*

Proof. We pick $d \in (0, 1)$ to be specified later.

We choose $A_0 \subseteq A$ and $B_0 \subseteq B$ homogeneous with respect to R and with $\mu(A_0) \geq \delta\mu(A)$ and $\nu(B_0) \geq \delta\nu(B)$ (applying Theorem 3.6 and Remark 2.17 to μ_A and ν_B , the localizations of μ on A and of ν on B , respectively).

If $A_0 \times B_0 \subseteq R$ then we take $A' = A_0, B' = B_0, h = \delta$ and we are done. So assume

$$(A_0 \times B_0) \cap R = \emptyset. \tag{4.1}$$

Let $a_0 = \mu(A_0)/\mu(A)$ and $b_0 = \nu(B_0)/\nu(B)$. Let $\alpha' = d(A, B)$, so $\alpha' \geq \alpha$.

From (4.1) it follows that $a_0b_0 \leq 1 - \alpha' \leq 1 - \alpha$. In particular at least one of a_0 or b_0 is at most $\sqrt{1 - \alpha}$.

We assume $a_0 \leq \sqrt{1 - \alpha}$. Let $A_1 = A \setminus A_0$ and $B_1 = B \setminus B_0$.

Case 1: $\nu(B_0) \leq d\nu(B)$. For definable $A' \subseteq A$ and $B' \subseteq B$ we write $\omega(R(A', B'))$ for $\omega((A' \times B') \cap R)$.

Since there are no R -edges between A_0 and B_0 , we have

$$\omega(R(A_0, B_1)) + \omega(R(A_1, B_0)) + \omega((A_1, B_1)) = \omega(R(A, B)) \geq \alpha\mu(A)\nu(B). \tag{4.2}$$

We also have

$$\begin{aligned} \mu(A_0)v(B_1) + \mu(A_1)v(B_0) + \mu(A_1)v(B_1) \\ = \mu(A)v(B) - \mu(A_0)v(B_0) \leq \mu(A)v(B)(1 - \delta^2). \end{aligned} \tag{4.3}$$

A very simple combinatorial statement is that if $r_1 + r_2 + r_3 \geq r$ and $s_1 + s_2 + s_3 \leq s$ then there is $i \in \{1, 2, 3\}$ with $r_i/s_i \geq r/s$.

So in this case we can choose $A' \in \{A_0, A_1\}$ and $B' \in \{B_0, B_1\}$ such that

$$d(A', B') \geq \alpha/(1 - \delta^2),$$

and $\mu(A') \geq h\mu(A)$, $v(B') \geq hv(B)$, where

$$h = \min\{\delta, 1 - d, 1 - \sqrt{1 - \alpha}\}.$$

Case 2: $v(B_0) > dv(B)$. In this case we take $A' = A_1 = A \setminus A_0$ and $B' = B$. As above, let $B_1 = B \setminus B_0$. Let $d' = 1 - d$.

The maximal possible measure of the set of R -edges between A_0 and B is

$$\omega(R(A_0, B)) = \omega(R(A_0, B_1)) \leq \mu(A_0)v(B_1) \leq \sqrt{1 - \alpha} d' \mu(A)v(B).$$

Thus for the measure of the set of R -edges between A_1 and B we obtain

$$\omega(R(A_1, B)) \geq \alpha\mu(A)v(B) - \sqrt{1 - \alpha} d' \mu(A)v(B).$$

Since $\mu(A_1)v(B) \leq (1 - \delta)\mu(A)v(B)$, we obtain

$$d(A_1, B) = \frac{\omega(R(A_1, B))}{\mu(A_1)v(B)} \geq \frac{\alpha - \sqrt{1 - \alpha} d'}{1 - \delta} = \alpha \frac{1 - \sqrt{1 - \alpha} (d'/\alpha)}{1 - \delta}.$$

As d' decreases to 0^+ , the right side goes increasingly to $\alpha/(1 - \delta)$. Since $0 < \delta < 1$, we have $1/(1 - \delta^2) < 1/(1 - \delta)$. So we can choose $d \in (0, 1)$ such that for $d' = 1 - d$ the right side is at least $\alpha/(1 - \delta^2)$.

Combining the two cases we take h as in Case 1. Uniform definability of A', B' in all the cases follows from the uniform definability of A_0, B_0 and the construction, so as always we can encode finitely many formulas into a single one. \square

Now we can use this proposition inductively to prove the analogue of Proposition 4.1 for hypergraphs, essentially following the proof of [16, Theorem 8.2].

Proposition 4.4. *Let \mathcal{M} be a distal structure and $R(x_0, \dots, x_{h-1})$ a definable relation. Given $\alpha > 0$ there is $\varepsilon > 0$ such that: given a generically stable product measure ω on $M^{|x_0|} \times M^{|x_1|} \times \dots \times M^{|x_{h-1}|}$ with $\omega(R) \geq \alpha$ there are definable sets $A_i \subseteq M^{|x_i|}$ with $\omega|_{x_i}(A_i) \geq \varepsilon$ for all $i < h$ such that $\prod_{i < h} A_i \subseteq R$. Moreover, each A_i is defined by an instance of a formula that depends only on R and α .*

Proof. Let $h \geq 2$ be given, and assume inductively that we have proved the assertion for all $i \leq h$. Let $R(x_0, \dots, x_h)$ and $\alpha > 0$ be given. Let ω be a generically stable product measure on $M^{|x_0|} \times \dots \times M^{|x_h|}$. Applying Proposition 4.1 with $h = 2$ to the binary relation $R(x_0; x_1, \dots, x_h)$ we find some $\varepsilon' > 0$, A_0 with $\omega|_{x_0}(A_0) \geq \varepsilon'$ and $A \subseteq M^{|x_1|} \times \dots \times M^{|x_h|}$ with $\omega|_{x_1, \dots, x_h}(A) \geq \varepsilon'$ such that R holds on all elements of $A_0 \times A$ (the corresponding projections of ω are clearly generically stable). Moreover, $A = R'(M^{|x_1|} \times \dots \times M^{|x_h|})$ for some uniformly definable (depending only on R and α) relation R' . We apply the inductive assumption to R' with $h - 1$, $\alpha = \varepsilon'$ and $\omega|_{x_1, \dots, x_h}$, which gives us some $\varepsilon'' > 0$ and uniformly definable sets A_i , $1 \leq i \leq h$, with $\omega|_{x_i}(A_i) \geq \varepsilon''$ and such that $A_1 \times \dots \times A_h \subseteq R'$, which implies $A_0 \times A_1 \times \dots \times A_h \subseteq R$. Take $\varepsilon = \min\{\varepsilon', \varepsilon''\}$. All the data are chosen uniformly depending only on R, α . \square

The density version implies a generalization of Theorem 3.6 to hypergraphs.

Corollary 4.5. *Let \mathcal{M} be a distal structure and $R(x_0, \dots, x_{h-1})$ a definable relation. Then there is $\delta > 0$ such that for any generically stable measures μ_i on $M^{|x_i|}$, there are A_i with $\mu_i(A_i) \geq \delta$ for all $i < h$, uniformly definable in terms of R , and such that either $\prod_{i < h} A_i \subseteq R$ or $\prod_{i < h} A_i \cap R = \emptyset$.*

Proof. Since a product of generically stable measures is generically stable, the measure $\omega = \mu_0 \otimes \dots \otimes \mu_{h-1}$ is generically stable, and either $\omega(R) \geq 1/2$ or $\omega(\neg R) \geq 1/2$. Applying Proposition 4.4 with $\alpha = 1/2$ to R and to $\neg R$ we obtain some $\varepsilon_1, \varepsilon_2$ respectively. But then $\delta = \min\{\varepsilon_1, \varepsilon_2\}$ satisfies the conclusion. \square

Moreover, the formulas defining homogeneous subsets can be chosen depending just on the formula defining the edge relation, and not on the parameters used (in the semialgebraic setting this corresponds to saying that the complexity of the homogeneous subsets is bounded in terms of the complexity of the edge relation, and does not depend on the choice of the coefficients of the polynomials involved).

Corollary 4.6. *Let \mathcal{M} be a distal structure and $\phi(x_0, \dots, x_{h-1}, y)$ a formula. Given $\alpha > 0$ there is $\varepsilon > 0$ such that: for a definable relation $R(x_0, \dots, x_{h-1}) = \phi(x_0, \dots, x_{h-1}, c)$ with some $c \in M^{|y|}$ and a generically stable product measure ω on $M^{|x_0|} \times M^{|x_1|} \times \dots \times M^{|x_{h-1}|}$ with $\omega(R) \geq \alpha$ there are definable sets $A_i \subseteq M^{|x_i|}$ with $\omega|_{x_i}(A_i) \geq \varepsilon$ for all $i < h$ and $\prod_{i < h} A_i \subseteq R$. Moreover, each A_i is defined by an instance of a formula that depends only on ϕ and α .*

Proof. This follows immediately from Proposition 4.4 applied to the relation

$$R'(x_0, \dots, x_{h-1}, y) = \phi(x_0, \dots, x_{h-1}, y)$$

and to the generically stable product measure $\omega' = \omega \otimes \delta_c$, where δ_c is a (generically stable) $\{0, 1\}$ -valued measure on $M^{|y|}$ concentrated on c . \square

Example 4.7. Let λ_n be the Lebesgue measure on \mathbb{R}^n restricted to the unit cube, i.e. $\lambda_n(X) = \Lambda_n(X \cap I_n)$ where Λ_n is the standard Lebesgue measure and I_n is the unit cube in \mathbb{R}^n .

Let \mathcal{R} be an o-minimal expansion of \mathbb{R} and $R(x_1, \dots, x_n; u)$ be a formula. Then for any $\alpha > 0$ there is some $\varepsilon > 0$ such that for any $c \in \mathbb{R}^{|u|}$ with $\lambda_n(R(\mathbb{R}^n; c)) \geq \alpha$ there are definable $A_i \subseteq \mathbb{R}$, $i = 1, \dots, n$, with $\lambda_1(A_i) \geq \varepsilon$ and $A_1 \times \dots \times A_n \subseteq R(\mathbb{R}^n, c)$.

This follows from Corollary 4.6 and Fact 2.19.

Also we get a generalization of the original semialgebraic counting version over finite sets from Theorem 1.7 with additional control on the parameters over which the homogeneous subsets are defined.

Corollary 4.8. *Let \mathcal{M} be a distal structure and let a formula $\phi(x, y, z)$ be given. Then there is some $\delta = \delta(\phi) > 0$ and formulas $\psi_1(x, z_1)$ and $\psi_2(y, z_2)$ depending just on ϕ and satisfying the following. For any definable relation $R(x, y) = \phi(x, y, c)$ for some $c \in M^{|z|}$ and finite $A \subseteq M^{|x|}$ and $B \subseteq M^{|y|}$ there are some $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \delta|A|$, $|B'| \geq \delta|B|$ and*

- (1) *the pair A', B' is R -homogeneous,*
- (2) *there are some $c_1 \in A^{|z_1|}$ and $c_2 \in B^{|z_2|}$ such that $A' = \psi_1(A, c_1)$ and $B' = \psi_2(B, c_2)$.*

Proof. Let $\psi_1(x, z_1)$, $\psi_2(x, z_2)$, ε be as given by Corollary 4.6. Then the existence of A', B' follows by defining $\mu(X)$ (resp., $\nu(X)$) to be the normalized number of points in $X \cap A$ (resp., $X \cap B$). Such Keisler measures are always generically stable by Fact 2.19.

For (2), by Remark 2.7 we can find some formulas $\psi'_1(x, z'_1)$ and $\psi'_2(x, z'_2)$ such that for any finite sets A, B and c_1, c_2 there are some $c'_1 \in A^{|z'_1|}$ and $c'_2 \in B^{|z'_2|}$ such that $\psi_1(A, c_1) = \psi'_1(A, c'_1)$ and $\psi_2(B, c_2) = \psi'_2(B, c'_2)$. \square

We will show in Section 6 that the most basic version of Corollary 4.8 characterizes distality. This is not the case however if we do not require definability of the homogeneous subsets.

Remark 4.9. (1) If every definable relation in \mathcal{M} satisfies the strong Erdős–Hajnal property and \mathcal{N} is interpretable in \mathcal{M} , then every definable relation in \mathcal{N} satisfies the strong Erdős–Hajnal property.

(2) Let \mathcal{M} and \mathcal{N} be two structures in the same language and assume that \mathcal{N} embeds into \mathcal{M} . If all quantifier-free definable relations in \mathcal{M} satisfy the strong Erdős–Hajnal property, then so do all quantifier-free definable relations in \mathcal{N} .

By the remark and Corollary 4.8 we have the following.

Corollary 4.10. *If \mathcal{M} is distal and \mathcal{N} is interpretable in \mathcal{M} (resp., embeds into \mathcal{M}), then all definable (resp., quantifier-free definable) relations in \mathcal{N} satisfy the strong Erdős–Hajnal property.*

Example 4.11. The following relations satisfy the strong Erdős–Hajnal property.

- (1) Definable relations in an arbitrary algebraically closed field of characteristic 0 (since ACF_0 is interpretable in the distal theory RCF of real closed fields).

- (2) Definable (in the language L_{div}) relations in an arbitrary nontrivially valued algebraically closed field of residue characteristic 0 (since its theory $\text{ACVF}_{0,0}$ is interpretable in the theory RCVF of real closed valued fields—see e.g. [4, Corollary 6.3]—which is distal in view of Example 2.10).
- (3) Quantifier-free definable relations in an arbitrary field of characteristic 0 (as the field can be embedded into some model of ACF_0).
- (4) Quantifier-free definable (in L_{div}) relations in an arbitrary valued field of equicharacteristic 0 (as it can always be embedded into a model of $\text{ACVF}_{0,0}$).

Thus (see Remark 1.4) we obtain many new families of graphs satisfying the Erdős–Hajnal conjecture.

Remark 4.12. (1) Every relation satisfying the strong Erdős–Hajnal property is NIP.
 (2) If all definable relations on \mathcal{M} satisfy the Erdős–Hajnal property then \mathcal{M} is NIP.

Proof. (1) If the relation $R(x, y)$ is not NIP, then for any finite bi-partite graph G there are some $A \subseteq M^{|x|}$ and $B \subseteq M^{|y|}$ such that G is isomorphic to $(A, B, R \cap (A \times B))$. By the optimality of the bound $O(\log n)$ on the size of homogeneous subsets in arbitrary bi-partite graphs it follows that R does not have the strong Erdős–Hajnal property.

(2) If the relation $R(x, y)$ is not NIP, let $R' \subseteq M^{|x|+|y|} \times M^{|x|+|y|}$ be defined by $R'(ab, cd) \Leftrightarrow R(a, d) \vee R(c, b)$. This is a symmetric relation such that for any finite graph G there is some set $A \subseteq M^{|x|+|y|}$ such that G is isomorphic to (R', A) (see e.g. [26, Lemma 2.2]). Again optimality of the logarithmic bound for arbitrary graphs implies that R' does not have the Erdős–Hajnal property. \square

5. Regularity lemma for distal hypergraphs

5.1. Regularity lemmas for restricted families of graphs

Szemerédi’s regularity lemma is a fundamental result in graph combinatorics with many versions and applications in extremal combinatorics, number theory and computer science (see [24] for a survey). In its simplest form for bi-partite graphs, it can be stated as follows.

Fact 5.1. *If $\varepsilon > 0$, then there exists $K = K(\varepsilon)$ such that: for any finite bi-partite graph $R \subseteq A \times B$, there exist partitions $A = A_1 \cup \dots \cup A_{k_1}$ and $B = B_1 \cup \dots \cup B_{k_2}$ into nonempty sets, and a set $\Sigma \subseteq \{1, \dots, k_1\} \times \{1, \dots, k_2\}$, with the following properties:*

- (1) Bounded size of the partition: $k_1, k_2 \leq K$.
- (2) Few exceptions: $|\bigcup_{(i,j) \in \Sigma} A_i \times B_j| \geq (1 - \varepsilon)|A \times B|$.
- (3) ε -regularity: for all $(i, j) \in \Sigma$, and all $A' \subseteq A_i$ and $B' \subseteq B_j$, one has

$$||R \cap (A' \times B')| - d_{ij}|A'| |B'|| \leq \varepsilon|A| |B|, \quad \text{where } d_{ij} = \frac{|R \cap (A_i \times B_j)|}{|A_i| |B_j|}.$$

In general the bound on the size of the partition K is known to grow as an exponential tower of height $1/\varepsilon$, and the result is less informative in the case of sparse graphs. Recently several improved regularity lemmas were obtained in the context of definable sets in certain structures or in restricted families of structures.

- (1) [41] Algebraic graphs of bounded complexity in large finite fields (equivalently, definable graphs in pseudofinite fields): pieces of the partition are algebraic of bounded complexity, no exceptional pairs, stronger regularity. Some generalizations and simplifications were obtained in [34, 19] and by Hrushovski (unpublished).
- (2) [27] Graphs of bounded VC-dimension: density arbitrarily close to 0 or 1, the size of the partition is bounded by a polynomial in $1/\varepsilon$.
 - (a) [29] Graphs without arbitrarily large half-graphs, corresponding to the case of stable graphs (no exceptional pairs).
 - (b) [16, 18] Semialgebraic graphs of bounded complexity.

We remark that the classes of structures in (1), (2)(a) and (2)(b) are orthogonal to each other. In the next section we give a generalization of (2)(b) to graphs definable in arbitrary distal structures. As remarked before, the stable and the distal cases present two extremal cases of general NIP structures.

5.2. Distal regularity lemma

We work in a model \mathcal{M} of a distal theory T . We have sorts S_1, \dots, S_k (i.e. definable subsets of some powers of \mathcal{M}) and a definable relation $R \subseteq S_1 \times \dots \times S_k$.

Notation 5.2. (a) Let $\vec{S} = S_1 \times \dots \times S_k$.

- (b) We call a subset $X \subseteq \vec{S}$ *rectangular* if it is of the form $X = X_1 \times \dots \times X_k$.
- (c) For $A \subseteq M$ and a finite set $\vec{\Delta} = \{\Delta_i(x_i, y_i) : i = 1, \dots, k\}$ of formulas, a rectangular subset $X = X_1 \times \dots \times X_k$ is called *Δ -definable* over A if each X_i is a finite Boolean combination of sets from $\{\Delta_i(x_i, a) : a \in A\}$. (In fact, we will need only conjunctions of Δ_i and their negations, i.e. partial $\vec{\Delta}$ -types.)
- (d) Given Keisler measures μ_i on each sort S_i , for a rectangular definable $X = X_1 \times \dots \times X_k$ we set

$$\mu(X) = \mu_1(X_1) \cdot \mu_2(X_2) \cdot \dots \cdot \mu_k(X_k).$$

- (e) By a *rectangular definable partition* of \vec{S} we mean a finite partition \mathcal{P} of \vec{S} consisting of rectangular definable sets.
- (f) For rectangular definable partitions $\mathcal{P}, \mathcal{P}_1$ of \vec{S} we write $\mathcal{P} \sqsubset \mathcal{P}_1$ if \mathcal{P} refines \mathcal{P}_1 , that is, for each $X \in \mathcal{P}$ there is $Y \in \mathcal{P}_1$ with $X \subseteq Y$.
- (g) Given Keisler measures μ_i on each sort S_i , for a rectangular definable partition \mathcal{P} of \vec{S} , we define the *defect* of \mathcal{P} to be

$$\text{def}(\mathcal{P}) := \sum_{\substack{X \in \mathcal{P} \\ X \text{ is not } R\text{-homogeneous}}} \mu(X).$$

Obviously, if $\mathcal{P}_1 \sqsubset \mathcal{P}$ then $\text{def}(\mathcal{P}_1) \leq \text{def}(\mathcal{P})$.

Proposition 5.3. *There is some constant $c = c(R)$ such that: for any $\varepsilon > 0$ and any generically stable measures μ_i on S_i , for $i = 1, \dots, k$, there is a rectangular uniformly definable (in terms of R and ε) partition \mathcal{P} of \vec{S} with $|\mathcal{P}| \leq (1/\varepsilon)^c$ and $\text{def}(\mathcal{P}) \leq \varepsilon$.*

We give a proof of the proposition in several claims, essentially following the proof of [18, Theorem 1.3] but working with Keisler measures.

Using Proposition 4.4 we find that the following holds.

Claim 5.4. *There is a constant $\delta = \delta(R)$ and formulas $\Delta_i(x_i, y_i)$, $i = 1, \dots, k$, such that for any generically stable measures μ_i on S_i there are a_i , $i = 1, \dots, k$, such that the sets $X_i \subseteq S_i$ defined by $\Delta_i(x_i, a_i)$ are R -homogeneous and $\mu_i(X_i) \geq \delta$.*

Remark 5.5. Since we are going to keep track of the parameters used in Δ_i , it is more convenient to assume that in the above claim $y_1 = \dots = y_k = y$ and $a_1 = \dots = a_k = a$. This can always be achieved by a concatenation of variables.

We fix δ from the previous claim and let $\vec{\Delta} = \{\Delta_i(x_i, y) : i = 1, \dots, k\}$, where the Δ_i are from the above claim.

Claim 5.6. *Let X be a definable rectangular subset of \vec{S} with $\mu(X) > 0$. Then there is some $a \in M^{|y|}$ and a rectangular set Y which is $\vec{\Delta}$ -definable over $\{a\}$, such that $X \cap Y$ is R -homogeneous and $\mu(X \cap Y) \geq \delta^k \mu(X)$.*

Proof. Apply Claim 5.4 to measures μ_i relativized to the sets X_i . □

Claim 5.7. *Let \mathcal{P} be a rectangular partition of \vec{S} which is $\vec{\Delta}$ -definable over a finite set A . Then there is a rectangular partition \mathcal{P}_1 which is $\vec{\Delta}$ -definable over a finite set A_1 with*

- (1) $|\mathcal{P}_1| \leq (k + 1)|\mathcal{P}|$,
- (2) $|A_1| \leq |A| + |\mathcal{P}|$,
- (3) $\text{def}(\mathcal{P}_1) \leq (1 - \delta^k) \text{def}(\mathcal{P})$.

Proof. Let $X \in \mathcal{P}$ be nonhomogeneous. We can partition it into $k + 1$ rectangular subsets, all of them $\vec{\Delta}$ -definable using one extra parameter a_X , such that one of these subsets is R -homogeneous and is of measure at least $\delta^k \mu(X)$. Namely, if $X = \prod_{i=1}^k X_i$, by Claim 5.6 there are some $Y_i \subseteq X_i$ such that $Y = \prod_{i=1}^k Y_i$ is an R -homogeneous subset of X with $\mu(Y) \geq \delta^k \mu(X)$, and we take a partition of X into $k + 1$ pieces given by the sets $Y_1 \times \dots \times Y_k$ and $X_1 \times \dots \times X_{i-1} \times (X_i \setminus Y_i) \times Y_i \times \dots \times Y_k$ for all $i = 1, \dots, k$.

Replacing each nonhomogeneous $X \in \mathcal{P}$ with such a subpartition we obtain \mathcal{P}_1 satisfying the requirements. □

Thus, by induction on n we can construct a rectangular partition \mathcal{P}_n of \vec{S} which is $\vec{\Delta}$ -definable over a finite set A_n and such that

- (1) $\text{def}(\mathcal{P}_n) \leq (1 - \delta^k)^n$,
- (2) $|\mathcal{P}_n| \leq (k + 1)^n$,
- (3) $|A_n| \leq \sum_{j < n} (k + 1)^j = \frac{(k + 1)^n - 1}{(k + 1) - 1} \leq (k + 1)^n$.

In particular, given $\varepsilon > 0$, using (1) and (2) above, after

$$N = \frac{\log \varepsilon}{\log(1 - \delta^k)} = \left(-\frac{1}{\log(1 - \delta^k)} \right) \log \frac{1}{\varepsilon}$$

steps we have $\text{def}(\mathcal{P}_N) \leq \varepsilon$ with $|\mathcal{P}_N| \leq (k + 1)^N \leq 2^{kN} \leq (1/\varepsilon)^c$, where $c = -k/\log(1 - \delta^k)$ is a positive constant depending only on R .

This finishes the proof of Proposition 5.3.

From the above k -partite version we obtain a regularity lemma for hypergraphs.

Theorem 5.8 (Distal regularity lemma). *Let $P \subseteq M^d$ be a definable set and $R(x_1, \dots, x_k)$ with $|x_i| = d$ for all $1 \leq i \leq k$ be a definable relation. Then there is some constant $c = c(R)$ such that the following holds. For any $\varepsilon > 0$ and for any generically stable measure μ on P , there is a partition $P = P_1 \cup \dots \cup P_K$ with $K = O((1/\varepsilon)^c)$ such that the P_i 's are uniformly definable (in terms of R and ε) and*

$$\sum \mu(P_{i_1}) \dots \mu(P_{i_k}) \leq \varepsilon,$$

where the sum is over all tuples (i_1, \dots, i_k) such that $(P_{i_1}, \dots, P_{i_k})$ is not R -homogeneous.

Proof. Let $\mathcal{P}_N, A_N, c, \bar{\Delta}$ be as given in the proof of Proposition 5.3 for $S_i = P$ and $\mu_i = \mu$, for all $1 \leq i \leq k$.

Using Fact 2.4, we obtain constants c_2 and c_3 depending only on R and such that the number of $\bar{\Delta}$ -types over any finite set A is bounded by $c_2|A|^{c_3}$. Finally, we partition P into realizations of complete $\bar{\Delta}$ -types over A_N , say $P = \bigcup_{i \leq K} P_i$. It follows that there will be at most $c_2(1/\varepsilon)^{c_3}$ parts. It is easy to see that this partition of P satisfies the homogeneity condition because the rectangular partition

$$\mathcal{P} := \{P_{i_1} \times \dots \times P_{i_k} : 1 \leq i_1, \dots, i_k \leq K\}$$

refines \mathcal{P}_N and $\text{def}(\mathcal{P}_N) \leq \varepsilon$. □

5.3. Finding definable equipartitions

Normally in the conclusion of a regularity lemma one is able to choose parts of (approximately) equal measure. We give a sufficient condition for this in the definable setting. To simplify some expressions, given real numbers r_1, r_2 and $\varepsilon > 0$, we write $r_1 \approx^\varepsilon r_2$ to denote that $|r_1 - r_2| < \varepsilon$.

Definition 5.9. We say that a structure \mathcal{M} *uniformly cuts finite sets* if for every formula $\phi(x, y)$ and every $\varepsilon > 0$ there is a formula $\chi(x, z)$ such that for any sufficiently large finite set $A \subseteq M^{|x|}$, any $b \in M^{|y|}$ and any $0 \leq m \leq |\phi(A, b)|$ there is some $c \in M^{|z|}$ such that

$$\frac{|\phi(A, b) \cap \chi(A, c)|}{|\phi(A, b)|} \approx^\varepsilon \frac{m}{|\phi(A, b)|}.$$

Note that this is a property of $\text{Th}(\mathcal{M})$.

Example 5.10. (1) Assume that there is a definable linear order $x < y$ on \mathcal{M} . Then clearly \mathcal{M} uniformly cuts finite sets, and χ in Definition 5.9 can be chosen independently of ϕ and ε (using lexicographic ordering for subsets of M^n for $n > 1$).

(2) Let $\mathcal{M} = (\mathbb{Q}_p, +, \times)$. Then \mathcal{M} uniformly cuts finite sets. For subsets of M this follows from the fact that every ball in \mathbb{Q}_p is a disjoint union of exactly p balls; then an argument similar to the proof that for an atomless measure, every set of positive measure contains subsets of arbitrary smaller measure, can be carried out up to ε , in a number of steps bounded in terms of ε (one can check using quantifier elimination in the p -adics that

in this case χ cannot be chosen independent of ε). To extend this to subsets of M^n for $n > 1$, note that if k is an infinite field and $A \subseteq k^n$ is a finite set then there is a uniformly definable linear map $f : k^n \rightarrow k$ that is one-to-one on A .

Proposition 5.11. *Let \mathcal{M} be a distal structure and assume that it uniformly cuts finite sets. Then for every formula $\phi(x, y)$ and $\delta > 0$ there is some $\chi(x, z)$ such that: for any generically stable measure μ on \mathcal{M} with $\mu(\{c\}) = 0$ for any singleton $c \in M^{|x|}$, if $0 \leq \alpha \leq \mu(\phi(x, a))$ then we can find some $b \in M^{|z|}$ with $\mu(\phi(x, a) \cap \chi(x, b)) \approx^\delta \alpha$.*

Proof. Fix $\varepsilon > 0$, and let $\chi(x, z)$ be an arbitrary formula. As \mathcal{M} is distal, it follows by Fact 2.16 that μ is smooth over M . Let $\theta_i^1(x), \theta_i^2(x), \psi_i(y, z), i = 1, \dots, n$, list all of the formulas over M given by Fact 2.15 for each of $\phi(x, y)$ and $\phi(x, y) \wedge \chi(x, z)$, with respect to μ and ε . Let \mathcal{B} be the finite Boolean algebra of subsets of $M^{|x|}$ generated by $\Theta = \{\theta_i^t(M) : i = 1, \dots, n, t = 1, 2\}$. Clearly the number of atoms in \mathcal{B} is at most 4^n . By assumption every definable set of positive μ -measure is infinite. Then for all sufficiently large $m \in \mathbb{N}$ we can choose a set $C \subseteq M^{|x|}$ with $|C| = m$ such that for every atom A of \mathcal{B} ,

$$\left| \frac{|C \cap A|}{m} - \mu(A) \right| < \frac{\varepsilon}{2 \cdot 4^n}.$$

It then follows that for every $i \in \{1, \dots, n\}$ and $t \in \{1, 2\}$ we have $|\theta_i^t(C)|/|C| \approx^{\varepsilon/2} \mu(\theta_i^t(M))$. But by the choice of θ_i^0, θ_i^1 this implies that for any set D from

$$\Delta(M) = \{\phi(M, a) : a \in M\} \cup \{\phi(M, a) \cap \chi(M, b) : a, b \in M\}$$

we have $|D \cap C|/|C| \approx^\varepsilon \mu(D)$.

Now let $0 < \alpha < \beta := \mu(\phi(M, a))$ be given (if $\alpha \in \{0, \beta\}$ then there is nothing to do). Let $\varepsilon := \delta/4$, and let $\chi(x, z)$ be as given by Definition 5.9 for $\phi(x, y)$ and ε . Take m sufficiently large (to be specified later). Then for C with $|C| = m$ chosen as above with respect to ε and χ we have in particular $|\phi(C, a)|/|C| \approx^\varepsilon \beta$. Let $l := |\phi(C, a)|$. We may assume that there is some $k \in \mathbb{N}$ with $k \leq l$ such that $k/l \approx^\varepsilon \alpha/\beta$ (since $\alpha > 0$, by choosing m sufficiently large we may assume that l is arbitrarily large). Then, using the fact that both $\beta, k/l \leq 1$ we obtain $\alpha \approx^\varepsilon \beta k/l \approx^\varepsilon \frac{l}{m} \frac{k}{l} = \frac{k}{m}$, so $\alpha \approx^{2\varepsilon} k/m$.

By the choice of $\chi(x, z)$, there is some $b \in M^{|z|}$ such that

$$\frac{|\phi(C, a) \cap \chi(C, b)|}{l} \approx^\varepsilon \frac{k}{l},$$

which implies $|\phi(C, a) \cap \chi(C, b)|/m \approx^\varepsilon k/m$, and so $|\phi(C, a) \cap \chi(C, b)|/m \approx^{3\varepsilon} \alpha$. By the assumption on C this implies $\mu(\phi(x, a) \wedge \chi(x, b)) \approx^{4\varepsilon} \alpha$; but $4\varepsilon = \delta$. \square

Remark 5.12. Recall that a global measure μ is *definable* over a small model M if it is $\text{Aut}(\mathbb{U}/M)$ -invariant and for every formula $\phi(x, y) \in L$ and every closed subset X of $[0, 1]$, the set $\{q \in S_{|y|}(M) : \mu(\phi(x, b)) \in X \text{ for any } b \in \mathbb{U}^{|y|} \text{ with } b \models q(y)\}$ is closed. The measure μ is *finitely satisfiable* if for every $\phi(x, b) \in L(\mathbb{U})$ with $\mu(\phi(x, b)) > 0$ there is some $a \in M^{|x|}$ such that $\models \phi(a, b)$ holds. As mentioned before, in an NIP

structure, a Keisler measure μ over \mathbb{M} is generically stable if and only if it admits a global M -invariant extension which is both definable over M and finitely satisfiable in M (see [22, Theorem 3.2]).

Corollary 5.13. *Assume that T uniformly cuts finite sets in such a way that χ in Definition 5.9 can be chosen independent of ε (e.g. if \mathcal{M} has a definable linear order). Then under the assumptions of Proposition 5.11 we can choose $b \in \mathbb{U}$ such that $\mu_1(\phi(x, a) \cap \chi(x, b)) = \alpha$, where μ_1 is the unique global Keisler measure extending μ .*

Proof. As μ_1 is generically stable over \mathcal{M} , it is in particular definable over \mathcal{M} . That is, for every $\delta > 0$ the set $\{b \in \mathbb{U} : \alpha - \delta \leq \mu_1(\phi(x, a) \wedge \chi(x, b)) \leq \alpha + \delta\}$ is type-definable over \mathcal{M} (and consistent). It then follows by compactness that we can find some $b^* \in \mathbb{U}$ with $\mu_1(\phi(x, a) \wedge \psi(x, b)) = \alpha$. \square

Corollary 5.14. *Let \mathcal{M} be a distal structure and assume that it uniformly cuts finite sets. Then in Theorem 5.8 for any μ satisfying in addition $\mu(\{c\}) = 0$ for all $c \in M^d$ and any $\delta > 0$ we can find a partition P_1, \dots, P_K with $\mu(P_i) \approx^\delta \mu(P_j)$ for all $1 \leq i, j \leq K$ (and with the parts P_i uniformly definable in terms of $R, \varepsilon, \delta, \chi$).*

Proof. We follow the standard repartition argument (see e.g. [18, proof of Theorem 1.3]).

Let (P, R) be a k -uniform hypergraph, and let $P = P_1 \cup \dots \cup P_K$ be a partition of its vertices given by Theorem 5.8 for $\varepsilon/2$, with $K \leq c_1(2/\varepsilon)^{c_2}$. Fix $\delta > 0$ and μ satisfying the assumptions; we will find a new partition $P = Q_1 \cup \dots \cup Q_{K'}$ satisfying the conclusion of the corollary for ε and δ .

Let $K' = \lceil 4(2^k/\varepsilon)K \rceil$; without loss of generality $0 < \delta < 1/K'$. Fix an arbitrary $0 < \delta' < \delta/K'$. Using Proposition 5.11 we can partition each P_i into $P_i = S_i \cup \bigcup Q_{i,j}$ with $\mu(Q_{i,j}) \approx^{\delta'} 1/K'$ for all j and with the remainder $\mu(S_i) < 1/K'$. Let now $S = \bigcup_i S_i$; again using Proposition 5.11 we can partition S into $S = T \cup \bigcup U_j$ with $\mu(U_j) \approx^{\delta'} 1/K'$ and $\mu(T) < 1/K'$. As δ' was sufficiently small compared to δ and $1/K'$, calculating the error we get $\mu(T) \approx^\delta 1/K'$.

We claim that $P = \bigcup Q_{i,j} \cup \bigcup U_j \cup T$ is the required partition. We relabel it as $P = Q_1 \cup \dots \cup Q_{K'}$. Then K' is still a polynomial in $1/\varepsilon$. Note that $\mu(S) < K/K'$, so at most K parts of the new partition are contained in S . Hence the sum $\sum \mu(Q_{i_1}) \dots \mu(Q_{i_k})$ over all tuples (i_1, \dots, i_k) for which not all of P_{i_1}, \dots, P_{i_k} are subsets of parts of the original partition is at most

$$K(K')^{k-1} \left(\frac{1}{K'} + \delta \right)^k \leq K(K')^{k-1} \frac{2^k}{(K')^k} = \frac{2^k K}{K'} = \frac{\varepsilon}{4}.$$

Together with the assumption on the original partition it then follows that $\sum \mu(Q_{i_1}) \dots \mu(Q_{i_k})$, where the sum is over all tuples (i_1, \dots, i_k) such that $(Q_{i_1}, \dots, Q_{i_k})$ is not R -homogeneous, is bounded by $\varepsilon/2 + \varepsilon/4 < \varepsilon$.

It follows from the construction that the new partition is uniformly definable in terms of the old one and δ , and thus uniformly definable in terms of R, ε and δ . \square

Remark 5.15. Answering a question from an earlier version of our article, Pierre Simon has recently demonstrated that all distal structures uniformly cut finite sets.

6. Equivalence to distality

6.1. Strong Erdős–Hajnal fails in ACF_p

Fields of positive characteristic give a standard example of the failure of the strong Szemerédi–Trotter bound on the number of incidences between points and lines. In a personal communication Terence Tao suggested that it may also be used for the failure of the strong Erdős–Hajnal property in this setting, which turned out to be the case indeed.

Let \mathbb{F} be a field. For a set $P \subseteq \mathbb{F}^2$ of points and a set L of lines in \mathbb{F}^2 we denote by $I(P, L) \subseteq P \times L$ the incidence relation, i.e. $I(P, L) = \{(p, l) \in P \times L : p \in l\}$. As remarked in Example 4.11, every field \mathbb{F} of characteristic 0 satisfies the strong Erdős–Hajnal property with respect to quantifier-free formulas. In particular we have

Proposition 6.1. *Let \mathbb{F} be a field of characteristic 0. Then there is a constant $\delta > 0$ such that for any finite (sufficiently large) set $P \subseteq \mathbb{F}^2$ of points and any finite (sufficiently large) set L of lines in \mathbb{F}^2 there are some $P_0 \subseteq P$ and $L_0 \subseteq L$ with $|P_0| \geq \delta|P|$, $|L_0| \geq \delta|L|$ and $I(P_0, L_0) = \emptyset$.*

We show that the assumption of characteristic 0 cannot be removed.

Proposition 6.2. *Fix a prime p , and let $\mathbb{F} = \mathbb{F}_p^{\text{alg}}$. Then the conclusion of Proposition 6.1 fails in \mathbb{F} .*

Proof. Assume towards a contradiction that \mathbb{F} satisfies that conclusion. Since every finite field of characteristic p can be embedded into \mathbb{F} , the following would be true:

Let \mathbb{F}_q be a finite field of characteristic p , of size q . Let P be the set of all points in \mathbb{F}_q , and let L be the set of all lines in \mathbb{F}_q^2 of the form $y = ax + b$. Then there are $P_0 \subseteq P$ and $L_0 \subseteq L$ with $|P_0| \geq \delta|P|$ and $|L_0| \geq \delta|L|$ and such that $I(P_0, L_0) = \emptyset$.

We show that this is impossible. We have $|P| = q^2$ and $|L| = q^2$. Since \mathbb{F}_q has size q , every line contains exactly q points, therefore $|I(P, L)| = |L|q = q^3$. Notice also that every point belongs to exactly q lines in L .

We fix k large enough so that $1/p^k < \delta$, and let $\delta_0 = 1/p^k$. Since $q = p^n$ for some n , $\delta_0 q$ is an integer for every $q \geq p^k$.

Hence we can choose $P_0 \subseteq P$ with $|P_0| = \delta_0|P|$ and $L_0 \subseteq L$ with $|L_0| = \delta_0|L|$ such that $I(P_0, L_0) = \emptyset$. Let $P_1 = P \setminus P_0$ and $L_1 = L \setminus L_0$. We have $|P_1| = (1 - \delta_0)q^2$ and $|L_1| = (1 - \delta_0)q^2$.

Consider $I(P_0, L)$. Since every point belongs to exactly q lines in L , we have $|I(P_0, L)| = |P_0|q = \delta_0 q^3$. Since $I(P_0, L_0) = \emptyset$, we have $I(P_0, L) = I(P_0, L_1)$, so $|I(P_0, L_1)| = \delta_0 q^3$.

On the other hand, from the Cauchy–Schwarz inequality (see e.g. [35, p. 1]) it follows that $|I(P_0, L_1)| \leq \sqrt{|L_1|} \sqrt{|I(P_0, L_1)| + |P_0|^2}$.

Thus we have

$$\delta_0 q^3 \leq \sqrt{(1 - \delta_0)q^2} \sqrt{\delta_0 q^3 + \delta_0^2 q^4} = \sqrt{(1 - \delta_0)\delta_0 q^5 + (1 - \delta_0)\delta_0^2 q^6}.$$

Since $1 - \delta_0 < 1$, the above inequality fails for large enough q —a contradiction. \square

Corollary 6.3. *Let K be an infinite field definable in a distal structure \mathcal{M} . Then $\text{char}(K) = 0$.*

Proof. By [23, Corollary 4.5] every infinite NIP field of characteristic $p > 0$ contains $\mathbb{F}_p^{\text{alg}}$. But then \mathcal{M} cannot satisfy the strong Erdős–Hajnal property by the proposition above, contradicting distality. \square

In particular the theory ACF_p admits no distal expansion. No examples of NIP theories with this property have been known until now.

6.2. Equivalence to distality

In this section we assume some familiarity with NIP theories (see e.g. [39]) and recall some facts about distal theories. We fix a theory T and a big sufficiently saturated model \mathbb{U} of T . Recall that a sequence $(a_i : i \in I)$ of elements of M^n indexed by a linear order I is *indiscernible* over a set of parameters $A \subseteq M$ if for any $i_1 < \dots < i_k$ and $j_1 < \dots < j_k$ from I we have $\text{tp}(a_{i_1}, \dots, a_{i_k}/A) = \text{tp}(a_{j_1}, \dots, a_{j_k}/A)$. Given a linear order I , by a *Dedekind cut* in I we mean a cut $I = I_1 + I_2$ (i.e., $I = I_1 \cup I_2$, $I_1 \cap I_2 = \emptyset$ and $a < b$ for all $a \in I_1$ and $b \in I_2$) such that I_1 has no maximal element and I_2 has no minimal element. We denote by I^* the reverse of the order on I .

Fact 6.4 ([8]). *Let T be a complete NIP theory. Then the following are equivalent:*

- (1) *T is distal (in the sense of Fact 2.5).*
- (2) *Every indiscernible sequence $I \subseteq M^d$ in any model of M of T is distal. That is, for any two distinct Dedekind cuts of I , if some two elements fill them separately, then they also fill them simultaneously: if $I = I_1 + I_2 + I_3$ and we have some a and b from M^d such that both $I_1 + a + I_2 + I_3$ and $I_1 + I_2 + b + I_3$ are indiscernible, then $I_1 + a + I_2 + b + I_3$ is indiscernible.*

Given an indiscernible sequence $I = (a_i : i \in [0, 1])$, one defines the *average measure* μ of I as the global Keisler measure given by $\mu(\phi(x)) = \lambda_1(\{i \in [0, 1] : a_i \models \phi(x)\})$ for all definable sets $\phi(x)$, where λ_1 is the Lebesgue measure on $[0, 1]$. It follows from NIP that this Keisler measure is well-defined, i.e. the corresponding set of indices is measurable for every $\phi(x)$ with parameters from \mathbb{U} . We say that $\mathfrak{c} = (I_1, I_2, t)$ is a *polarized cut* of I if $I = I_1 + I_2$ is a cut of I and $t \in \{1, 2\}$ specifies whether the cut is approached from the left or from the right. It follows from NIP that for a polarized Dedekind cut $\mathfrak{c} = (I_1, I_2, t)$ and a set of parameters $A \subseteq \mathbb{U}$ we have a complete *limit type* of \mathfrak{c} over A denoted by $\text{lim}(\mathfrak{c}/A)$ and defined by $\phi(x) \in \text{lim}(\mathfrak{c}/A) \Leftrightarrow$ the set $\{i \in I_t : \mathbb{U} \models \phi(a_i)\}$ is unbounded from above in case $t = 1$, or from below in case $t = 2$.

Fact 6.5. *Let T be NIP, let I be an indiscernible sequence and let μ be the average measure of I .*

- (1) *The measure μ is generically stable [22, Proposition 3.7].*
- (2) *The support of μ is exactly the set of limit types of cuts of I . That is, if for some formula $\phi(x)$ we have $\mu(\phi(x)) > 0$ then $\phi(x) \in \text{lim}(\mathfrak{c}/\mathbb{U})$ for some polarized Dedekind cut \mathfrak{c} of I [38, Lemma 2.20].*
- (3) *I is distal if and only if μ is smooth [38, Proposition 2.21].*

Recall that a sequence $(a_i : i \in I)$ is *totally indiscernible* if for any $i_1 \neq \dots \neq i_k$ and $j_1 \neq \dots \neq j_k$ from I we have $\text{tp}(a_{i_1}, \dots, a_{i_k}/A) = \text{tp}(a_{j_1}, \dots, a_{j_k}/A)$

Fact 6.6 ([38]). *Let T be dp-minimal. Then it is distal if and only if no infinite indiscernible sequence is totally indiscernible.*

Remark 6.7. It follows that if T' is a dp-minimal expansion of a distal dp-minimal theory T , then T' is distal as well.

Indeed, If T' expands T and $(a_i : i \in I)$ is an infinite L' -indiscernible sequence, then it is in particular L -indiscernible, so it is not totally- L -indiscernible by distality of T , so of course not totally- L' -indiscernible.

Fact 6.8 (Strong base change, [38, Lemma 2.8]). *Let T be NIP. Let I be an indiscernible sequence and $A \supseteq I$ a set of parameters. Let $(c_i : i < \alpha)$ be a sequence of pairwise distinct polarized Dedekind cuts in I . For each i , let d_i fill the cut c_i (i.e., if $c_i = (I_1, I_2, t)$ then $I_1 + d_i + I_2$ is indiscernible). Then there exist $(d'_i : i < \alpha)$ in \mathbb{U} such that*

- (1) $\text{tp}((d'_i)_{i < \alpha} / I) = \text{tp}((d_i)_{i < \alpha} / I)$,
- (2) for each $i < \alpha$, $\text{tp}(d'_i / A) = \lim(c_i / A)$.

Finally, we will use the following finitary version of a characteristic property of NIP theories.

Fact 6.9. *Let $\phi(x, y)$ be an NIP formula. Then there are some $k, N \in \mathbb{N}$ such that for any indiscernible sequence $I = (a_i : i < n)$ from $M^{|x|}$ with $n \geq N$ and any $b \in M^{|y|}$, the set $\phi(I, b)$ is a disjoint union of at most k intervals.*

Proof. This follows from the usual characterization of NIP via bounded alternation on indiscernible sequences (see e.g. [1, Proposition 4]) plus compactness. □

Theorem 6.10. *Let T be an NIP theory. The following are equivalent:*

- (1) T is distal.
- (2) For any definable relation $R(x, y)$ and any global generically stable measures μ_1, μ_2 there are some definable $X \subseteq \mathbb{U}^{|x|}$ and $Y \subseteq \mathbb{U}^{|y|}$ which are R -homogeneous and satisfy $\mu_1(X), \mu_2(Y) > 0$.
- (3) For any definable relation $R(x, y)$ there is some $\delta > 0$ and some formulas $\psi_1(x, z_1)$ and $\psi_2(x, z_2)$ such that for all finite $A \subseteq \mathbb{U}^{|x|}$ and $B \subseteq \mathbb{U}^{|y|}$ there are some $c_i \in \mathbb{U}^{|z_i|}$, $i = 1, 2$, such that $|\psi_1(A, c_1)| \geq \delta|A|$, $|\psi_2(B, c_2)| \geq \delta|B|$ and the pair of sets $\psi_1(A, c_1), \psi_2(B, c_2)$ is R -homogeneous.

Proof. (1) \Rightarrow (2) and (1) \Rightarrow (3) follow from Corollaries 4.5 and 4.8.

(2) \Rightarrow (1). Assume that $I = (a_i)_{i \in \mathcal{I}}$ is a nondistal indiscernible sequence, with $\mathcal{I} = [0, 1]$. This means that I can be written as $I = I_1 + I_2 + I_3$ (where $I_j = (a_i : i \in \mathcal{I}_j)$ and $\mathcal{I}_1, \mathcal{I}_2^*, \mathcal{I}_2, \mathcal{I}_3^*$ are without last elements) in such a way that there are some $c, d \in \mathbb{U}$ such that $I_1 + c + I_2 + I_3$ and $I_1 + I_2 + d + I_3$ are indiscernible, but $I_1 + c + I_2 + d + I_3$ is not.

Then there is a formula $\phi(I'_1, x, I'_2, y, I'_3)$ with some finite $I'_j \subset I_j$, say $I'_j = (a_i : i \in \mathcal{I}'_j)$, $\mathcal{I}'_j \subset \mathcal{I}_j$ for $j \in \{1, 2, 3\}$, such that $\phi(I'_1, a_j, I'_2, a_k, I'_3)$ holds for any $\mathcal{I}'_1 < j < \mathcal{I}'_2 < k < \mathcal{I}'_3$, but $\models \neg\phi(I'_1, c, I'_2, d, I'_3)$. Let $[j_1, j_2]$ be some interval of \mathcal{I} between \mathcal{I}'_1 and \mathcal{I}'_2 , and $[k_1, k_2]$ some interval between \mathcal{I}'_2 and \mathcal{I}'_3 . Let $J = (a_i : i \in [j_1, j_2])$ and $K = (a_i : i \in [k_1, k_2])$. Let μ be the average measure of J , and ν the average mea-

sure of K (we may assume that both sequences are indexed by $[0, 1]$ by taking an order preserving bijection). Then both μ and ν are generically stable by Fact 6.5(1).

Now assume $X = \xi(\mathbb{U})$ and $Y = \chi(\mathbb{U})$ are definable subsets of $\mathbb{U}^{|\mathfrak{c}|}$ with $\mu(X), \nu(Y) > 0$, where ξ, χ are formulas with parameters in some small model $M \supseteq I$. By Fact 6.5(2) there is some polarized Dedekind cut \mathfrak{c} of J such that $\xi(x) \in \lim_J(\mathfrak{c}/M)$, and some polarized Dedekind cut \mathfrak{d} of K such that $\chi(x) \in \lim_K(\mathfrak{d}/M)$.

It follows by compactness, indiscernibility of I and taking an automorphism of \mathbb{U} that there are some c' filling \mathfrak{c} and d' filling \mathfrak{d} (separately, as cuts in I) such that $\neg\phi(I'_1, c', I'_2, d', I'_3)$ holds. By Fact 6.8 we can find some c'', d'' such that $\neg\phi(I'_1, c'', I'_2, d'', I'_3)$ still holds, but moreover $c'' \models \lim(\mathfrak{c}/M), d'' \models \lim(\mathfrak{d}/M)$. In particular, $\models \xi(c'') \wedge \chi(d'')$. On the other hand, by the choice of ϕ and the definition of μ, ν there are some $j < k$ in I such that $\models \xi(a_j) \wedge \chi(a_k) \wedge \phi(I'_1, a_j, I'_2, a_k, I'_3)$. This shows that the relation $R(x, y) = \phi(I'_1, x, I'_2, y, I'_3)$ is not homogeneous on $X \times Y$. As X, Y were arbitrary definable sets of positive measure, we conclude.

(3) \Rightarrow (1). Assume that T is not distal, and we will show that (3) cannot hold. Working in \mathbb{U} we have some $I_i = (a_j^i : j \in \mathbb{Q})$ for $i \in \{1, 2, 3\}$ and a, b such that $I = I_1 + I_2 + I_3, I_1 + a + I_2 + I_3$ and $I_1 + I_2 + b + I_3$ are indiscernible, but $I_1 + a + I_2 + b + I_3$ is not. This implies in particular that there is a formula $\phi \in L$ such that $\models \neg\phi(J'_1, a, J'_2, b, J'_3)$ for some finite $J'_i \subset I_i$ with $J'_i = (a_j : j \in \mathcal{J}_i)$, but $\models \phi(J'_1, a', J'_2, b', J'_3)$ for any $a', b' \in I$ such that $\mathcal{J}'_1 < a' < \mathcal{J}'_2 < b' < \mathcal{J}'_3$.

Let now $R(x, y; c) := \phi(J'_1, x, J'_2, y, J'_3)$ with $c := J'_1 J'_2 J'_3$. Assume that there are $\psi_i(x, y), i \in \{1, 2\}$, and $\delta > 0$ as required by (3) for R . As T is NIP, it follows by Fact 6.9 that there are some $k, N \in \omega$ such that for any indiscernible sequence $K = (a_j : j < n)$ with $n \geq N$ and any $d_i \in \mathbb{U}, i \in \{1, 2\}$, the set $\psi_i(K, d_i)$ is a disjoint union of at most k intervals. Without loss of generality it then follows from (3) that there is some $k' \in \omega$ such that for any finite indiscernible sequences $A = (a_j : j < n)$ and $B = (b_j : j < n)$ with $n \geq N$ we can find intervals $A_0 \subseteq A$ with $|A_0| \geq |A|/k'$ and $B_0 \subseteq B$ with $|B_0| \geq |B|/k'$ such that (A_0, B_0) is $R(x, y; c)$ -homogeneous. We are going to show that this property fails.

Relabeling the sequence we may assume that $I_1 = I_{1,0} + I_{1,1} + \dots$ and $I_3 = \dots + I_{3,1} + I_{3,0}$, with each $I_{i,j}$ indexed by \mathbb{Q} , and that $J'_1 \subset I_{1,0}, J'_3 \subset I_{3,0}$. Let $I'_1 := I_1 \setminus I_{1,0}$ and $I'_3 := I_3 \setminus I_{3,0}$.

By indiscernibility of I , automorphism and compactness, for any Dedekind cuts \mathfrak{c} of I'_1 and \mathfrak{c}' of I'_3 we can find some a' and b' which fill those cuts (separately, viewed as cuts in I) and are such that $\models \neg\phi(J'_1, a', J'_2, b', J'_3)$.

For each $i \in \omega$, let $(c_{i,j} : j \in \omega)$ be an infinite increasing sequence of cuts of $I_{1,i}$, and let $(c'_{i,j} : j \in \omega)$ be a decreasing sequence of cuts of $I_{3,i}$. By the previous remark, let $a_{i,j}$ and $b_{i,j}$ be such that $a_{i,j}$ fills the cut $c_{i,j}$ and $b_{i,j}$ fills the cut $c'_{i,j}$, and $\models \neg\phi(J'_1, a_{i,j}, J'_2, b_{i,j}, J'_3)$ holds.

Next using Fact 6.8 and induction we can choose $a'_{i,j}, b'_{i,j}$ such that

- $\text{tp}(a'_{i,j}, b'_{i,j}/I) = \text{tp}(a_{i,j}, b_{i,j}/I)$,
- $\text{tp}(a'_{i,j}/IA_{i,j}) = \lim(c_{i,j}/IA_{i,j})$, where $A_{i,j} = \{a'_{i,j'} : j' < j\} \cup \{a'_{i',j'} : i' < i, j' \in \omega\}$,

- $\text{tp}(b'_{i,j}/IB_{i,j}) = \lim(c'_{i,j}/IB_{i,j})$, where $B_{i,j} = \{b'_{i,j'} : j' < j\} \cup \{b'_{i',j'} : i' < i, j' \in \omega\}$.

From this we have:

- (a) $\models \neg\phi(J'_1, a'_{i,j}, J'_2, b'_{i,j}, J'_3)$ holds for any $i, j \in \omega$.
- (b) I'_1 with all the $\{a'_{i,j} : i, j \in \omega\}$ added in the corresponding cuts is indiscernible.
- (c) I'_3 with all the $\{b'_{i,j} : i, j \in \omega\}$ added in the corresponding cuts is indiscernible.
- (d) $\models \phi(J_1, a', J_2, b', J_3)$ holds for any $a' \in I'_1$ and $b' \in I'_3$.

Here (a) follows from the first bullet above and the choice of $a_{i,j}, b_{i,j}$; using the second bullet it is easy to show that (b) holds, and that the sequence has the same EM-type as I (similarly for (c)); (d) was already observed above.

In view of (a)–(d) above, for any $m \in \omega$ we can choose indiscernible sequences $A = (a_j : j < 2k'm)$ and $B = (b_j : j < 2k'm)$ such that for any $l_1, l_2 < 2k'$ we have $\models \neg R(a_{l_1m+l_2}, b_{l_2m+l_1}; c)$ and $\models R(a_{l_1m+j_1}, b_{l_2m+j_2}; c)$ for any $j_1, j_2 \in (2k', m)$. It then follows that for all sufficiently large m , for any choice of an interval $A_0 \subseteq A$ with $|A_0| \geq |A|/k' \geq 2m$ and $B_0 \subseteq B$ with $|B_0| \geq |B|/k' \geq 2m$, the sets (A_0, B_0) cannot be $R(x, y; c)$ -homogeneous—contradicting the choice of k' . \square

Remark 6.11. Pierre Simon has also observed a version of the implication (2) \Rightarrow (1) in Theorem 6.10 after seeing a preliminary version of our results.

The above proof shows that an NIP theory is distal if and only if property (3) in Theorem 6.10 holds for all finite indiscernible sequences A, B . As the following proposition shows, in an arbitrary NIP theory property (3) almost holds for A, B indiscernible sequences, except for the uniform definability of one of the homogeneous subsets.

Proposition 6.12. *Let $\phi(x, y)$ be NIP. Then there is $\varepsilon > 0$ depending only on ϕ such that for any indiscernible sequences $A = (a_i : i < n)$ and $B = (b_i : i < m)$ (in fact Δ -indiscernible for some finite Δ depending just on ϕ is enough) there are $A_0 \subseteq A$ and $B_0 \subseteq B$ such that $|A_0| \geq \varepsilon|A|$ and $|B_0| \geq \varepsilon|B|$ and either $\phi(a, b)$ holds for all $a \in A_0$ and $b \in B_0$, or $\neg\phi(a, b)$ holds for all $a \in A_0$ and $b \in B_0$.*

Proof. By Fact 6.9 there is k such that $\phi(x, y)$ cannot alternate on an indiscernible sequence more than k times. We divide B into $k+1$ intervals of almost equal length. Namely, for $i < k+1$ let

$$B_i = \left\{ b_j : i \cdot \frac{m}{k+1} \leq j < (i+1) \cdot \frac{m}{k+1} \right\}.$$

Then for every $a \in A$ there is some interval B_{i_a} not containing any alternation points. It follows that for some $i' < k+1$, there are $|A|/(k+1)$ points in A which do not alternate inside $B_{i'}$, and then at least half of them satisfy ϕ or $\neg\phi$. So we can take $\varepsilon = 1/(2(k+1))$. \square

Acknowledgments. We thank Pierre Simon for pointing out a mistake in an earlier version of our results. We thank Dave Marker, Kobi Peterzil, Caroline Terry, Roland Walker and the anonymous referee for their numerous comments that have helped to improve the presentation.

The work presented in this paper began during the program on Model Theory, Arithmetic Geometry and Number Theory at MSRI, Berkeley, Spring 2014. We thank MSRI for their hospitality. The first author was partially supported by ValCoMo (ANR-13-BS01-0006), by the Fondation Sciences Mathématiques de Paris (FSMP) and by a public grant overseen by the French National Research Agency (ANR) as part of the Investissements d’Avenir program (reference: ANR-10-LABX-0098). The second author was partially supported by NSF.

References

- [1] Adler, H.: An introduction to theories without the independence property. *Arch. Math. Logic* (accepted)
- [2] Alon, N., Pach, J., Pinchasi, R., Radoičić, R., Sharir, M.: Crossing patterns of semi-algebraic sets. *J. Combin. Theory Ser. A* **111**, 310–326 (2005) [Zbl 1099.14048](#) [MR 2156215](#)
- [3] Alon, N., Pach, J., Solymosi, J.: Ramsey-type theorems with forbidden subgraphs. *Combinatorica* **21**, 155–170 (2001) [Zbl 0989.05124](#) [MR 1832443](#)
- [4] Aschenbrenner, M., Dolich, A., Haskell, D., Macpherson, D., Starchenko, S.: Vapnik–Chervonenkis density in some theories without the independence property, I. *Trans. Amer. Math. Soc.* **368**, 5889–5949 (2016) [Zbl 06551585](#) [MR 3458402](#)
- [5] Basu, S.: Combinatorial complexity in o-minimal geometry. *Proc. London Math. Soc.* **100**, 405–428 (2010) [Zbl 1186.52017](#) [MR 2595744](#)
- [6] Belegardek, O., Peterzil, Y., Wagner, F.: Quasi-o-minimal structures. *J. Symbolic Logic* **65**, 1115–1132 (2000) [Zbl 0968.03043](#) [MR 1791366](#)
- [7] Chernikov, A., Simon, P.: Definably amenable NIP groups. *J. Amer. Math. Soc.* **31**, 609–641 (2018) [Zbl 06870167](#) [MR 3787403](#)
- [8] Chernikov, A., Simon, P.: Externally definable sets and dependent pairs II. *Trans. Amer. Math. Soc.* **367**, 5217–5235 (2015) [Zbl 06429160](#) [MR 3335415](#)
- [9] Chernikov, A., Starchenko, S.: A note on the Erdős–Hajnal property for stable graphs. *Proc. Amer. Math. Soc.* **146**, 785–790 (2018) [Zbl 06816426](#) [MR 3731711](#)
- [10] Chudnovsky, M.: The Erdős–Hajnal conjecture—a survey. *J. Graph Theory* **75**, 178–190 (2014) [Zbl 1280.05086](#) [MR 3150572](#)
- [11] Conlon, D., Fox, J., Pach, J., Sudakov, B., Suk, A.: Ramsey-type results for semi-algebraic relations. *Trans. Amer. Math. Soc.* **366**, 5043–5065 (2014) [Zbl 1306.14027](#) [MR 3217709](#)
- [12] Dickmann, M. A.: Elimination of quantifiers for ordered valuation rings. *J. Symbolic Logic* **52**, 116–128 (1987) [Zbl 0624.03025](#) [MR 0877859](#)
- [13] Erdős, P., Hajnal, A.: Ramsey-type theorems. *Discrete Appl. Math.* **25**, 37–52 (1989) [Zbl 0715.05052](#) [MR 1031262](#)
- [14] Erdős, P., Hajnal, A., Pach, J.: A Ramsey-type theorem for bipartite graphs. *Geombinatorics* **10**, 64–68 (2000) [Zbl 0978.05052](#) [MR 1784373](#)
- [15] Erdős, P., Szekeres, G.: A combinatorial problem in geometry. *Compos. Math.* **2**, 463–470 (1935) [Zbl 0012.27010](#) [MR 1556929](#)
- [16] Fox, J., Gromov, M., Lafforgue, V., Naor, A., Pach, J.: Overlap properties of geometric expanders. *J. Reine Angew. Math.* **671**, 49–83 (2012) [Zbl 1306.05171](#) [MR 2983197](#)
- [17] Fox, J., Pach, J.: Erdős–Hajnal-type results on intersection patterns of geometric objects. In: *Horizons of Combinatorics*, Springer, Berlin, 79–103 (2008) [Zbl 1170.05327](#) [MR 2432528](#)
- [18] Fox, J., Pach, J., Suk, A.: A polynomial regularity lemma for semi-algebraic hypergraphs and its applications in geometry and property testing. *SIAM J. Comput.* **45**, 2199–2223 (2016) [Zbl 1353.05090](#) [MR 3585030](#)

- [19] García, D., Macpherson, D., Steinhorn, C.: Pseudofinite structures and simplicity. *J. Math. Logic* **15**, no. 1, 1550002, 41 pp. (2015) [Zbl 1337.03041](#) [MR 3367776](#)
- [20] Haskell, D., Macpherson, D.: A version of o-minimality for the p -adics. *J. Symbolic Logic* **62**, 1075–1092 (1997) [Zbl 0894.03015](#) [MR 1618009](#)
- [21] Hrushovski, E., Pillay, A.: On NIP and invariant measures. *J. Eur. Math. Soc.* **13**, 1005–1061 (2011) [Zbl 1220.03016](#) [MR 2800483](#)
- [22] Hrushovski, E., Pillay, A., Simon, P.: Generically stable and smooth measures in NIP theories. *Trans. Amer. Math. Soc.* **365**, 2341–2366 (2013) [Zbl 1294.03023](#) [MR 3020101](#)
- [23] Kaplan, I., Scanlon, T., Wagner, F. O.: Artin–Schreier extensions in NIP and simple fields. *Israel J. Math.* **185**, 141–153 (2011) [Zbl 1261.03120](#) [MR 2837131](#)
- [24] Komlós, J., Simonovits, M.: Szemerédi’s regularity lemma and its applications in graph theory. In: *Combinatorics, Paul Erdős is Eighty*, Vol. 2, János Bolyai Math. Soc., Budapest, 295–352 (1996) [Zbl 0851.05065](#) [MR 1395865](#)
- [25] Laskowski, M. C.: Vapnik–Chervonenkis classes of definable sets. *J. London Math. Soc.* (2) **45**, 377–384 (1992) [Zbl 0766.60016](#) [MR 1171563](#)
- [26] Laskowski, M. C., Shelah, S.: Karp complexity and classes with the independence property. *Ann. Pure Appl. Logic* **120**, 263–283 (2003) [Zbl 1010.03024](#) [MR 1949710](#)
- [27] Lovász, L., Szegedy, B.: Regularity partitions and the topology of graphons. In: *An Irregular Mind*, János Bolyai Math. Soc., Budapest, 415–446 (2010) [Zbl 1242.05188](#) [MR 2815610](#)
- [28] Macintyre, A.: On definable subsets of p -adic fields. *J. Symbolic Logic* **41**, 605–610 (1976) [Zbl 0362.02046](#) [MR 0485335](#)
- [29] Malliaris, M., Shelah, S.: Regularity lemmas for stable graphs. *Trans. Amer. Math. Soc.* **366**, 1551–1585 (2014) [Zbl 1283.05234](#) [MR 3145742](#)
- [30] Marker, D.: *Model Theory: an Introduction*. Springer (2002) [Zbl 1003.03034](#)
- [31] Matoušek, J.: *Lectures on Discrete Geometry*. Grad. Texts in Math. 212, Springer (2002) [Zbl 0999.52006](#) [MR 1899299](#)
- [32] Pach, J., Radoičić, R., Vondrák, J.: On the diameter of separated point sets with many nearly equal distances. *Eur. J. Combin.* **27**, 1321–1332 (2006) [Zbl 1107.52011](#) [MR 2260123](#)
- [33] Pach, J., Solymosi, J.: Crossing patterns of segments. *J. Combin. Theory Ser. A* **96**, 316–325 (2001) [Zbl 0989.05031](#) [MR 1864126](#)
- [34] Pillay, A., Starchenko, S.: Remarks on Tao’s algebraic regularity lemma. [arXiv:1310.7538](https://arxiv.org/abs/1310.7538) (2013)
- [35] Rudnev, M.: Szemerédi–Trotter theorem and applications. <http://www.maths.bris.ac.uk/~maxmr/stlectures.pdf> (2004)
- [36] Scanlon, T.: O-minimality as an approach to the André–Oort conjecture. In: *Around the Zilber–Pink Conjecture*, Panor. Synthèses 52, Soc. Math. France, Paris, 111–165 (2017) [MR 3728313](#)
- [37] Shelah, S.: *Classification Theory and the Number of Non-isomorphic Models*. 2nd ed., Stud. Logic Found. Math. 92, North-Holland, Amsterdam (1990) [Zbl 0713.03013](#) [MR 1083551](#)
- [38] Simon, P.: Distal and non-distal NIP theories. *Ann. Pure Appl. Logic* **164**, 294 – 318 (2013) [Zbl 1269.03037](#)
- [39] Simon, P.: *A Guide to NIP Theories*. Lecture Notes in Logic 44, Cambridge Univ. Press (2015) [Zbl 1332.03001](#) [MR 3560428](#)
- [40] Simon, P.: A note on “Regularity lemma for distal structures”. *Proc. Amer. Math. Soc.* **144**, 3573–3578 (2016) [Zbl 06586361](#) [MR 3503725](#)
- [41] Tao, T.: Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets. *Contrib. Discrete Math.* **10**, 22–98 (2015) [Zbl 1370.11135](#) [MR 3386249](#)

-
- [42] van den Dries, L.: Tame Topology and o-Minimal Structures. London Math. Soc. Lecture Note Ser. 248, Cambridge Univ. Press (1998) [Zbl 0953.03045](#) [MR 1633348](#)
- [43] van den Dries, L., Haskell, D., Macpherson, D.: One-dimensional p -adic subanalytic sets. J. London Math. Soc. **59**, 1–20 (1999) [Zbl 0932.03038](#) [MR 1688485](#)
- [44] Vapnik, V. N., Chervonenkis, A. Ya.: On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab. Appl. **16**, 264–280 (1971) [Zbl 0247.60005](#) [MR 3408730](#)