Applications of model theory in extremal graph combinatorics

Artem Chernikov

(IMJ-PRG, UCLA)

Logic Colloquium

Helsinki, August 4, 2015

Szemerédi regularity lemma

Theorem

[E. Szemerédi, 1975] Every large enough graph can be partitioned into boundedly many sets so that on almost all pairs of those sets the edges are approximately uniformly distributed at random.

Szemerédi regularity lemma

Theorem

[E. Szemerédi, 1975] Given $\varepsilon > 0$, *there exists* $K = K(\varepsilon)$ *such that: for any finite bipartite graph* $R \subseteq A \times B$, *there exist partitions* $A = A_1 \cup \ldots \cup A_k$ *and* $B = B_1 \cup \ldots \cup B_k$ *into non-empty sets, and a set* $\Sigma \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\}$ *of good pairs with the following properties.*

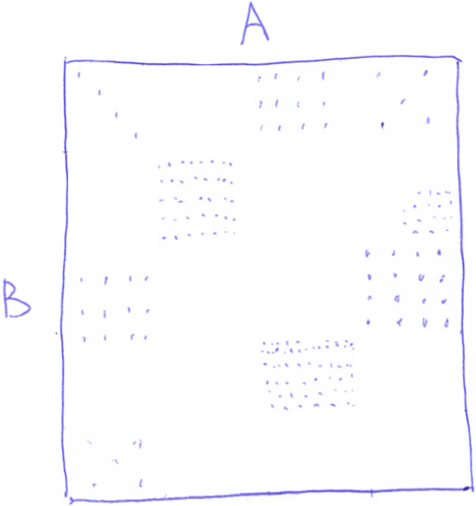
1. *(Bounded size of the partition)* $k \leq K$.
2. *(Few exceptions)* $\left| \bigcup_{(i,j) \in \Sigma} A_i \times B_j \right| \geq (1 - \varepsilon) |A| |B|$.
3. *(ε-regularity) For all* $(i, j) \in \Sigma$, *and all* $A' \subseteq A_i, B' \subseteq B_j$:

$$\left| \left| R \cap (A' \times B') \right| - d_{ij} \left| A' \right| \left| B' \right| \right| \leq \varepsilon |A| |B|,$$

where $d_{ij} = \frac{\left| R \cap (A_i \times B_j) \right|}{\left| A_i \times B_j \right|}$.

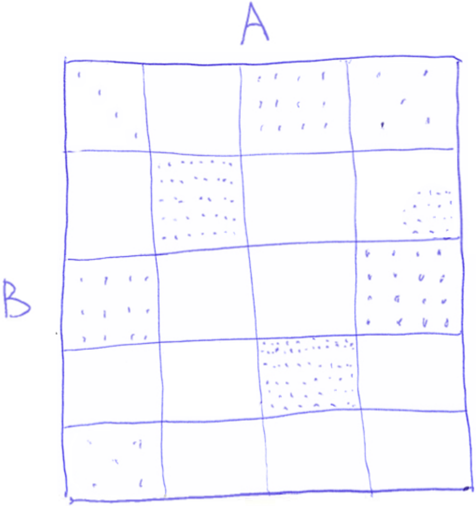
Szemerédi regularity lemma

Consider the incidence matrix of a bipartite graph (R, A, B):



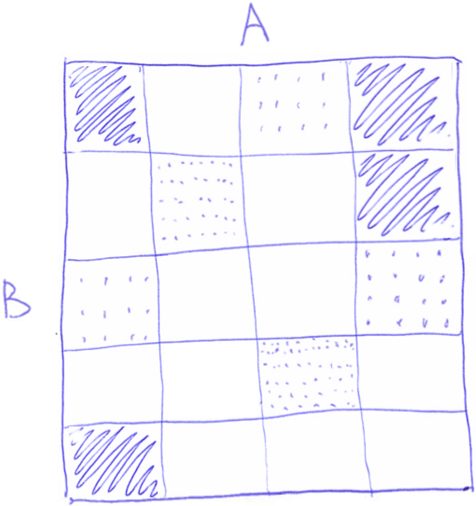
Szemerédi regularity lemma

Consider the incidence matrix of a bipartite graph (R, A, B):



Szemerédi regularity lemma

Consider the incidence matrix of a bipartite graph (R, A, B):



Szemerédi regularity lemma: bounds and applications

- Exist various versions for weaker and stronger partitions, for hypergraphs, etc.
- Increasing the error a little one may assume that the sets in the partition are of (approximately) equal size.
- Has many applications in extreme graph combinatorics, additive number theory, computer science, etc.
- [T. Gowers, 1997] The size of the partition $K(\varepsilon)$ grows as an exponential tower $2^{2^{\cdots}}$ of height $\left(1/\varepsilon^{\frac{1}{64}}\right)$.
- Can get better bounds for restricted families of graphs (e.g. coming from algebra, geometry, etc.)? Some recent positive results fit nicely into the *model-theoretic* classification picture.

Shelah's classification program

Theorem

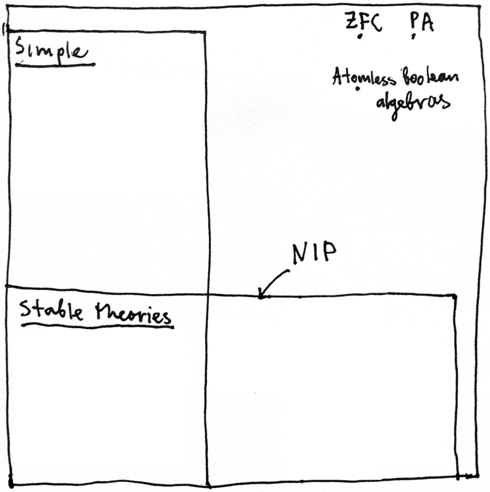
[M. Morley, 1965] Let T be a countable first-order theory. Assume T has a unique model (up to isomorphism) of size κ for some uncountable cardinal κ. Then ***for any*** *uncountable cardinal λ it has a unique model of size λ.*

- Morley's conjecture: for any T, the function

  $$f_T : \kappa \mapsto |\{M : M \models T, |M| = \kappa\}|$$

  is non-decreasing on uncountable cardinals.
- Shelah's "radical" solution: describe all possible functions (distinguished by T (not) being able to encode certain combinatorial configurations).
- Additional outcome: stability theory and its generalizations.
- Later, Zilber, Hrushovski and many others: geometric stability theory — close connections with algebraic objects interpretable in those structures.

Model-theoretic classification



- See G. Conant's ForkingAndDividing.com for an interactive map of the (first-order) universe.

Stability

- Given a theory T in a language L, a (partitioned) formula $\phi(x, y) \in L$ (x, y are tuples of variables), a model $M \models T$ and $b \in M^{|y|}$, let $\phi(M, b) = \{a \in M^{|x|} : M \models \phi(a, b)\}$.
- Let $\mathcal{F}_{\phi,M} = \{\phi(M, b) : b \in M^{|y|}\}$ be the family of ϕ-definable subsets of M. All dividing lines are expressed as certain conditions on the combinatorial complexity of the families $\mathcal{F}_{\phi,M}$ (independent of the choice of M).

Definition

1. A formula $\phi(x, y)$ is k-stable if **there are no** $M \models T$ and $(a_i, b_i : i < k)$ in M such that
$$M \models \phi(a_i, b_j) \iff i \leq j.$$
2. $\phi(x, y)$ is *stable* if it is k-stable for some $k \in \omega$.
3. A theory T is stable if it implies that all formulas are stable.

Stable examples

Example

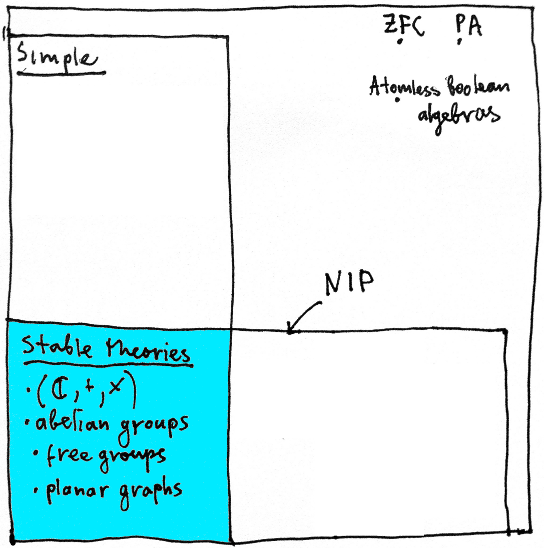
The following structures are stable:

1. abelian groups and modules,
2. $(\mathbb{C}, +, \times, 0, 1)$ (more generally, algebraically/separably/differentially closed fields),
3. [Z. Sela] free groups (in the pure group language $(\cdot, ^{-1}, 0)$),
4. planar graphs (in the language with a single binary relation).

Stability theory

- There is a rich machinery for analyzing types and models of stable theories (ranks, forking calculus, weight, indiscernible sequences, etc.).
- These results have substantial infinitary Ramsey-theoretic content (in disguise).
- Making it explicit and finitizing leads to results in combinatorics.
- The same principle applies to various generalizations of stability.

Stable regularity lemma



Recalling general regularity lemma

Theorem

[E. Szemerédi, 1975] Given $\varepsilon > 0$*, there exists* $K = K(\varepsilon)$ *such that: for any finite bipartite graph* $R \subseteq A \times B$*, there exist partitions* $A = A_1 \cup \ldots \cup A_k$ *and* $B = B_1 \cup \ldots \cup B_k$ *into non-empty sets, and a set* $\Sigma \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\}$ *of good pairs with the following properties.*

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

$$\left| \left| R \cap \left(A' \times B' \right) \right| - d_{ij} \left| A' \right| \left| B' \right| \right| \leq \varepsilon |A| |B|,$$

where $d_{ij} = \frac{\left| R \cap \left(A_i \times B_j \right) \right|}{\left| A_i \times B_j \right|}$.

Stable regularity lemma

Theorem

[M. Malliaris, S. Shelah, 2012] Given $\varepsilon > 0$ *and* k, *there exists* $K = K(\varepsilon, k)$ *such that:*
for any k-stable finite bipartite graph $R \subseteq A \times B$, *there exist partitions* $A = A_1 \cup \ldots \cup A_k$ *and* $B = B_1 \cup \ldots \cup B_k$ *into non-empty sets, and a set* $\Sigma \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\}$ *of good pairs with the following properties.*

1. *(Bounded size of the partition)* $k \leq K$.
2. *(No exceptions)* $\Sigma = \{1, \ldots, k\} \times \{1, \ldots, k\}$.
3. *(ε-regularity) For all* $(i, j) \in \Sigma$, *and all* $A' \subseteq A_i, B' \subseteq B_j$:

$$\left| \left| R \cap \left(A' \times B' \right) \right| - d_{ij} \left| A' \right| \left| B' \right| \right| \leq \varepsilon \left| A \right| \left| B \right|,$$

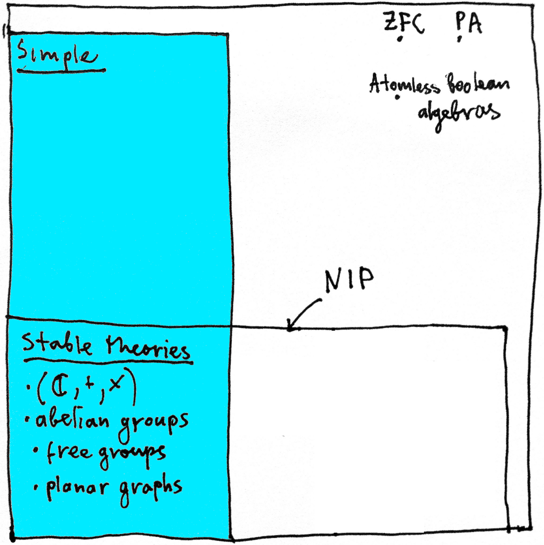
   where $d_{ij} = \frac{\left| R \cap \left(A_i \times B_j \right) \right|}{\left| A_i \times B_j \right|}$.

4. *Moreover, can take* $K \leq \left(\frac{1}{\varepsilon} \right)^c$ *for some* $c = c(k)$.

Stable regularity lemma, some remarks

- In particular this applies to finite graphs whose edge relation (up to isomorphism) is definable in a model of a stable theory.
- An easier proof is given recently by [M. Malliaris, A. Pillay, 2015] and applies also to infinite definable stable graphs, with respect to more general measures.

Simple theories



Recalling general regularity lemma

Theorem

[E. Szemerédi, 1975] Given $\varepsilon > 0$, there exists $K = K(\varepsilon)$ such that: for any finite bipartite graph $R \subseteq A \times B$, there exist partitions $A = A_1 \cup \ldots \cup A_k$ and $B = B_1 \cup \ldots \cup B_k$ into non-empty sets, and a set $\Sigma \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\}$ of good pairs with the following properties.

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

$$\left|\left|R \cap \left(A' \times B'\right)\right| - d_{ij}\left|A'\right|\left|B'\right|\right| \leq \varepsilon |A||B|,$$

where $d_{ij} = \frac{\left|R\cap\left(A_i \times B_j\right)\right|}{\left|A_i \times B_j\right|}$.

Tao's algebraic regularity lemma

Theorem

[T. Tao, 2012] If $t > 0$, there exists $K = K(t) > 0$ s. t.: whenever $\mathbf{F}$ is a finite field, $A \subseteq \mathbf{F}^n$, $B \subseteq \mathbf{F}^m$, $R \subseteq A \times B$ are definable sets in $\mathbf{F}$ of complexity at most t (i.e. $n, m \leq t$ and can be defined by some formula of length bounded by t), there exist partitions $A = A_0 \cup \ldots \cup A_k$, $B = B_0 \cup \ldots \cup B_k$ satisfying the following.

1. *(Bounded size of the partition) $k \leq K$.*
2. *(No exceptions) $\Sigma = \{1, \ldots, k\} \times \{1, \ldots, k\}$.*
3. *(Stronger regularity) For all $(i, j) \in \Sigma$, and all $A' \subseteq A_i, B' \subseteq B_j$:*

$$\left| \left| R \cap \left(A' \times B' \right) \right| - d_{ij} \left| A' \right| \left| B' \right| \right| \leq \left(c \left| \mathbf{F} \right|^{-1/4} \right) \left| A \right| \left| B \right|,$$

   where $d_{ij} = \frac{\left| R \cap \left(A_i \times B_j \right) \right|}{\left| A_i \times B_j \right|}$.

4. *Moreover, the sets $A_1, \ldots, A_k, B_1, \ldots, B_k$ are definable, of complexity at most K.*

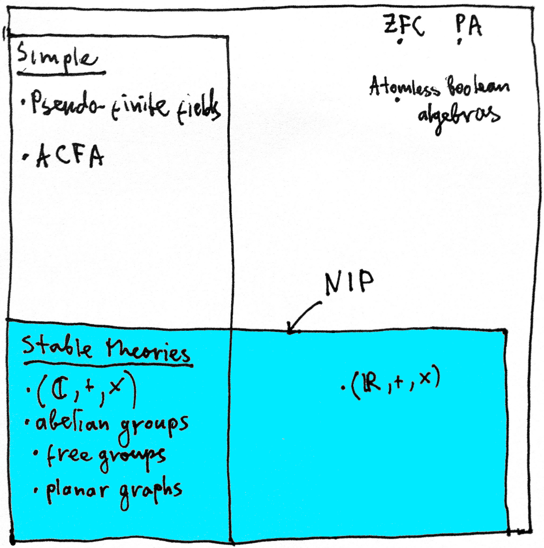
Simple theories

1. It is really a result about graphs definable in pseudofinite fields (with respect to the non-standard counting measure) — a central example of a structure with a *simple theory*.
2. A theory is *simple* if one cannot encode an infinite tree via a uniformly definable family of sets $\mathcal{F}_{\phi,M} = \left\{\phi\left(M, b\right) : b \in M^{|y|}\right\}$ in some model of T, is for any formula ϕ.
3. Some parts of stability theory, especially around forking, were generalized to the class of simple theories by Hrushovski, Kim, Pillay and others.

Simple theories and pseudo-finite fields

1. A field F is *pseudofinite* if it is elementarily equivalent to an ultraproduct of finite fields modulo a non-principal ultrafilter.
2. Model-theory of pseudofinite fields was studied extensively, starting with [J. Ax, 1968].
3. Tao's proof relied on the quantifier elimination and bounds on the size of definable subsets in pseudo-finite fields due to [Z. Chatzidakis, L. van den Dries, A. Macintyre, 1992] and some results from étale cohomology.
4. Fully model-theoretic proofs of Tao's theorem (replacing étale cohomology by some local stability and forking calculus, well-understood in the 90's) and some generalizations to larger subclasses of simple theories were given by [E. Hrushovski], [A. Pillay, S. Starchenko], [D. Garcia, D. Macpherson, C. Steinhorn].

NIP theories



Semialgebraic regularity lemma

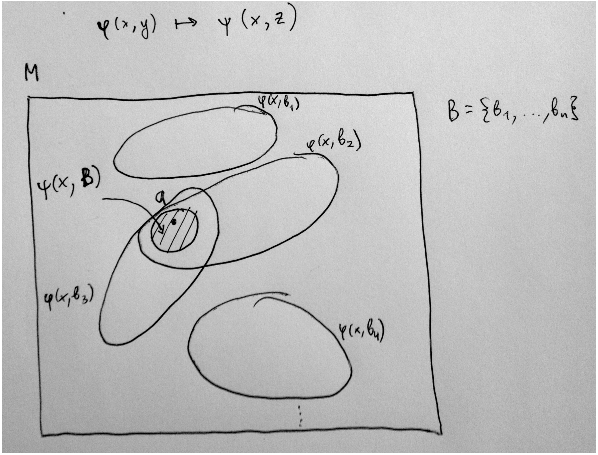
- A set $A \subseteq \mathbb{R}^d$ is *semialgebraic* if it can be defined by a finite boolean combination of polynomial equalities and inequalities.
- The *description complexity* of a semialgebraic set $A \subseteq \mathbb{R}^d$ is $\leq t$ if $d \leq t$ and A can be defined by a boolean combination involving at most t polynomial inequalities, each of degree at most t.
- Examples of semialgebraic graphs: incidence relation between points and lines on the plane, pairs of circles in $\mathbb{R}^3$ that are linked, two parametrized families of semialgebraic varieties having a non-empty intersection, etc.
- [J.Fox, M. Gromov, V. Lafforgue, A. Naor, J. Pach, 2010] + [J. Fox, J. Pach, A. Suk, 2015] Regularity lemma for semialgebraic graphs of bounded complexity.
- In a joint work with S. Starchenko we prove a generalization for graphs definable in *distal structures*, with respect to a larger class of *generically stable* measures.

Distal theories

- NIP ("No Independence Property") is an important dividing line in Shelah's classification theory generalizing the class of stable theories.
- Turned out to be closely connected to the Vapnik–Chervonenkis dimension, or VC-dimension — a notion from combinatorics introduced around the same time (central in computational learning theory).
- The class of *distal theories* was introduced and studied by [P. Simon, 2011] in order to capture the class of "purely unstable" NIP theories.
- The original definition is in terms of a certain property of indiscernible sequences.
- [C., Simon, 2012] gives a combinatorial characterization of distality:

Distal structures

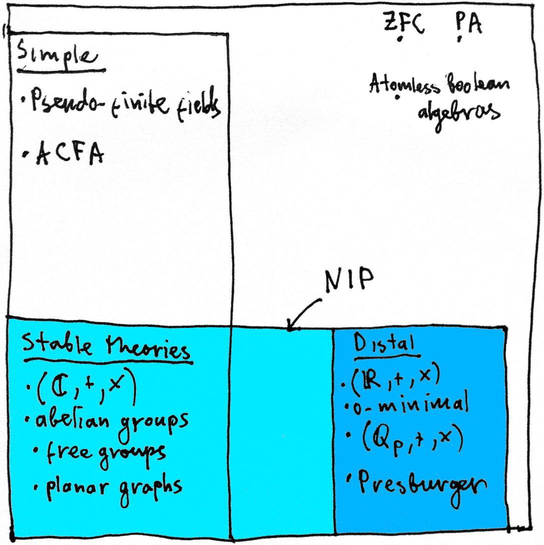
- **Theorem/Definition** An NIP structure M is *distal* if and only if for every definable family $\{\phi(x, b) : b \in M^d\}$ of subsets of M there is a definable family $\{\psi(x, c) : c \in M^{kd}\}$ such that for every $a \in M$ and every finite set $B \subset M^d$ there is some $c \in B^k$ such that $a \in \psi(x, c)$ and for every $a' \in \psi(x, c)$ we have $a' \in \phi(x, b) \Leftrightarrow a \in \phi(x, b)$, for all $b \in B$.



Examples of distal structures

- All (weakly) o-minimal structures are distal, e.g. $M = (\mathbb{R}, +, \times, e^x)$.
- Any p-minimal theory with Skolem functions is distal. E.g. $(\mathbb{Q}_p, +, \times)$ for each prime p is distal (e.g. due to the p-adic cell decomposition of Denef).
- Presburger arithmetic.

Distal theories



Distal regularity lemma

Theorem

[C., Starchenko] Let M be distal. For every definable $R(x,y)$ and every $\varepsilon > 0$ there is some $K = K(\varepsilon, R)$ such that: for any generically stable measures μ on $M^{|x|}$ and ν on $M^{|y|}$, there are $A_0, \ldots, A_k \subseteq M^{|x|}$ and $B_0, \ldots, B_k \subseteq M^{|y|}$ uniformly definable by instances of formulas depending just on R and ε , and a set $\Sigma \subseteq \{1, \ldots, k\}^2$ such that:

1. *(Bounded size of the partition) $k \leq K$,*
2. *(Few exceptions) $\omega\left(\bigcup_{(i,j)\in\Sigma} A_i \times B_j\right) \geq 1 - \varepsilon$, where ω is the product measure of μ and ν,*
3. *(The best possible regularity) for all $(i,j) \in \Sigma$, either $(A_i \times B_j) \cap R = \emptyset$ or $A_i \times B_j \subseteq R$.*
4. *Moreover, K is bounded by a polynomial in $\left(\frac{1}{\varepsilon}\right)$.*

Generically stable measures and some examples

- By a generically stable measure we mean a finitely additive probability measure on the Boolean algebra of definable subsets of M^n that is "well-approximated by frequency measures". The point is that in NIP (via VC theory) uniformly definable families of sets satisfy a uniform version of the weak law of large numbers with respect to such measures.
- Examples of generically stable measures:
  - A (normalized) counting measure concentrated on a finite set.
  - Lebesgue measure on $[0, 1]$ over reals, restricted to definable sets.
  - Haar measure on a compact ball over p-adics.
- Moreover, we show that any structure such that all graphs definable in it satisfy this strong regularity lemma is distal.

An application (in case I still have time)

- Let (G, V) be an undirected graph. 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 $v \neq v' \in V_0$.
- A class of finite graphs $\mathcal{G}$ has the *Erdős-Hajnal property* if there is $\delta > 0$ such that every $G \in \mathcal{G}$ has a homogeneous subset of size $\geq |V(G)|^{\delta}$.
- **Erdős-Hajnal conjecture.** For every finite graph H, the class of all H-free graphs has the Erdős-Hajnal property.
- **Fact.** If $\mathcal{G}$ is a class of finite graphs closed under subgraphs and $\mathcal{G}$ satisfies distal regularity lemma (without requiring definability of pieces), then $\mathcal{G}$ has the Erdős-Hajnal property.
- Thus, we obtain many new families of graphs satisfying the Erdős-Hajnal conjecture (e.g. quantifier-free definable graphs in arbitrary valued fields of characteristic 0).