Random Graphs

and

The Parity Quantifier

Phokion G. Kolaitis

UC Santa Cruz
&
IBM Research-Almaden

Swastik Kopparty

MIT
&
Institute for Advanced Study

What is finite model theory?

It is the study of logics on classes of finite structures.

Logics:

First-order logic FO and various extensions of FO:

- Fragments of second-order logic SO.
- Logics with fixed-point operators.
- Finite-variable infinitary logics.
- Logics with generalized quantifiers.

Main Themes in Finite Model Theory

- **Classical model theory in the finite:**
  Do the classical results of model theory hold in the finite?

- **Expressive power of logics in the finite:**
  What **can** and what **cannot** be expressed in various logics on classes of finite structures.

- **Descriptive complexity:**
  computational complexity vs. uniform definability
  (logic-based characterizations of complexity classes).

- **Logic and asymptotic probabilities on finite structures**
  0-1 laws and convergence laws.

Classical Model Theory in the Finite

- Preservation under substructures

Theorem: Tait – 1959

The Łoś -Tarski Theorem fails in the finite.

(rediscovered by Gurevich and Shelah in the 1980s)

- Preservation under homomorphisms

Theorem: Rossman – 2005

If a FO-sentence ψ is preserved under homomorphisms on all finite structures, then there is an existential positive FO-sentence ψ^* that is equivalent to ψ on all finite structures.

Descriptive Complexity

- Characterizing NP

Theorem: Fagin 1974

On the class ***G*** of all finite graphs G=(V,E),

NP = ESO (existential second-order logic).

- Characterizing P

Theorem: Immerman 1982, Vardi 1982

On the class ***O*** of all ordered finite graphs G= (V,<,E),

P = LFP (least fixed-point logic), where

LFP = FO + Least fixed-points of positive FO-formulas.

Logic and Asymptotic Probabilities

- **Notation:**
  - Q: Property (Boolean query) on the class **F** of all finite structures
  - $\mathbf{F}_n$: Class of finite structures with n in their universe
  - μ_n: Probability measure on $\mathbf{F}_n$, $n \geq 1$
  - $\mu_n(Q)$ = Probability of Q on $\mathbf{F}_n$ with respect to μ_n, $n \geq 1$.
- **Definition:** Asymptotic probability of property Q

  $\mu(Q) = \lim \mu_{n \to \infty}(Q)$ (provided the limit exists)
- **Examples:** For the uniform measure μ on finite graphs **G**:
  - μ(**G** contains a triangle) = 1.
  - μ(**G** is connected) = 1.
  - μ(**G** is 3-colorable) = 0.
  - μ(**G** is Hamiltonian) = 1.

0-1 Laws and Convergence Laws

Question: Is there a connection between the definability of a property Q in some logic L and its asymptotic probability?

Definition: Let L be a logic

- The 0-1 law holds for L w.r.t. to a measure μ_n, n$\geq$ 1, if

  $$\mu(\psi) = 0 \quad \text{or} \quad \mu(\psi) = 1,$$

  for every L-sentence ψ.

- The convergence law holds for L w.r.t. to a measure μ_n, n$\geq$ 1, if $\mu(\psi)$ exists, for every L-sentence ψ.

0-1 Law for First-Order Logic

Theorem: Glebskii et al. – 1969, Fagin – 1972
The 0-1 law holds for FO w.r.t. to the uniform measure on the class of all finite graphs.

Proof Techniques:

- Glebskii et al.
  Quantifier Elimination + Counting
- Fagin
  Transfer Theorem:
  There is a unique countable graph **R** such that for every FO-sentence ψ, we have that
  $$\mu(\psi) = 1 \text{ if and only if } \mathbf{R} \models \psi.$$
  Note:
  - **R** is Rado's graph: Unique countable, homogeneous, universal graph; it is characterized by a set of first-order extension axioms.
  - Each extension axiom has asymptotic probability equal to 1.

FO Truth vs. FO Almost Sure Truth

Everywhere true (valid)
Somewhere true & Somewhere false
Everywhere false (contradiction)

Almost surely true
Almost surely false

- **First-Order Truth**
  Testing if a FO-sentence is **true** on all finite graphs is an **undecidable** problem (Trakhtenbrot - 1950)

- **Almost Sure First-Order Truth**
  Testing if a FO-sentence is **almost surely true** on all finite graphs is a **decidable** problem; in fact, it is PSPACE-complete (Grandjean - 1985).

Three Directions of Research on 0-1 Laws

- 0-1 laws for FO on restricted classes of finite structures
  - Partial Orders, Triangle-Free Graphs, …

- 0-1 laws on graphs under variable probability measures.
  - G(n,p) with $p \neq 1/2$ (e.g., $p(n) = n^{-(1/e)}$)

- 0-1 laws for extensions of FO w.r.t. the uniform measure.

Restricted Classes and Variable Measures

- Restricted classes of finite structures

Theorem: Compton - 1986

The 0-1 law hods for the class of all finite partial orders

  - Proof uses results of Kleitman and Rothschild – 1975 about the asymptotic structure of partial orders.

- Variable probability measures

Theorem: Shelah and Spencer – 1987

Random finite graphs under the G(n,p) model with $p = n^{-\alpha}$

- If α is irrational, then the 0-1 law **holds** for FO.
- If α is rational, then the 0-1 law **fails** for FO.

0-1 Laws for Extensions of First-Order Logic

Many generalizations of the original 0-1 law, including:

- Blass, Gurevich, Kozen – 1985

  0-1 Law for Least Fixed-Point Logic LFP

  - Captures Connectivity, Acyclicity, 2-Colorability, …

- K … and Vardi – 1990

  0-1 Law for Finite-Variable Infinitary Logics $L^k_{\infty\omega}$, $k \geq 2$

  - Proper extension of LFP

- K… and Vardi – 1987, 1988

  0-1 Laws for fragments of Existential Second-Order Logic

  - Capture 3-Colorability, 3-Satisfiability, …

Logics with Generalized Quantifiers

- Dawar and Grädel – 1995
  - 0-1 Law for FO[Rig], i.e., FO augmented with the rigidity quantifier.
  - Sufficient condition for the 0-1 Law to hold for FO[**Q**], where **Q** is a collection of generalized quantifiers.
- Kaila – 2001, 2003
  - Sufficient condition for the 0-1 Law to hold for $L^k_{\infty\omega}$ [**Q**], $k \geq 2$, where is a collection of simple numerical quantifiers.
  - Convergence Law for $L^k_{\infty\omega}$ [**Q**], $k \geq 2$, where is a collection of certain special quantifiers on very sparse random finite structures.
- Jarmo Kontinen – 2010
  - Necessary and sufficient condition for the 0-1 law to hold for $L^k_{\infty\omega}[\exists^{s/t}]$, $k \geq 2$.

A Barrier to 0-1 Laws

All generalizations of the original 0-1 law are obstructed by

THE PARITY PROBLEM

The Parity Problem

- Consider the property

  Parity = "there is an odd number of vertices"

- For n odd, μ_n(Parity) = 1
- For n even, μ_n(Parity) = 0

- Hence, μ(Parity) does **not** exist.
- Thus, if a logic L can express Parity, then even the convergence law fails for L.

First-Order Logic + The Parity Quantifier

Goal of this work:

- Turn the parity barrier into a feature.
- Investigate the asymptotic probabilities of properties of finite graphs expressible in FO[$\oplus$], that is, in first-order logic augmented with the parity quantifier $\oplus$.

FO[$\oplus$]: FO + The Parity Quantifier $\oplus$

- Syntax of FO[$\oplus$]: If $\varphi(v)$ is a formula, then so is $\oplus\, v\, \varphi(v)$.
- Semantics of $\oplus\, v\, \varphi(v)$:
  - "the number of v's for which $\varphi(v)$ is true is odd"
- Examples of FO[$\oplus$]-sentences on finite graphs:
  - $\oplus\, v\, \exists\, w\, E(v, w)$
    - The number of vertices of positive degree is odd.
  - $\neg\, \exists\, v\, \oplus\, w\, E(v, w)$
    - There is **no** vertex of odd degree, i.e.,
    - The graph is Eulerian.

Vectorized FO[$\oplus$]

- Syntax: If $\varphi(v_1,\ldots,v_t)$ is a formula, then so is

$$\oplus(v_1,\ldots,v_t)\ \varphi(v_1,\ldots,v_t)$$

- Semantics of $\oplus(v_1,\ldots,v_t)\ \varphi(v_1,\ldots,v_{\mathbf{t}})$:
  - "there is an odd number of tuples $(v_1,\ldots,v_t)$ for which $\varphi(v_1,\ldots,v_t)$ is true"

- **Fact:**

$$\oplus(v_1,\ldots,v_t)\ \varphi(v_1,\ldots,v_t) \quad \text{iff} \quad \oplus\, v_1 \oplus v_2 \cdots \oplus v_t\ \varphi(v_1,\ldots,v_t).$$

  - Thus, FO[$\oplus$] is powerful enough to express its vectorized version.

The Uniform Measure on Finite Graphs

Let $\mathbf{G}_n$ be the collection of all finite graphs with n vertices

- The uniform measure on $G_{n:}$
  - If $G \in \mathbf{G}_n$, then $pr_n(G) = 1/ 2^{n(n-1)/2}$
  - If Q is a property of graphs, then
    $pr_n(Q)$ = fraction of graphs in $\mathbf{G}_n$ that satisfy Q.

An equivalent formulation

- The G(n, 1/2)-model:
  - Random graph with n vertices
  - Each edge appears with probability ½ and independently of all other edges

FO[$\oplus$] and Asymptotic Probabilities

Question: Let ψ be a FO[$\oplus$]-sentence.

What can we say about the asymptotic behavior of the sequence

$$pr_n(\psi), \quad n \geq 1 \ ?$$

Asymptotic Probabilities of FO[$\oplus$]-Sentences

Fact: The 0-1 Law **fails** for FO[$\oplus$]

Reason 1 (a blatant reason):

Let ψ be the FO[$\oplus$]-sentence $\oplus$ v (v = v)

Then

- $pr_{2n}(\psi) = 0$
- $pr_{2n+1}(\psi) = 1$.

Hence,

- $\lim_{n \to \infty} pr_n(\psi)$ does **not** exist.

Asymptotic Probabilities of FO[$\oplus$]-Sentences

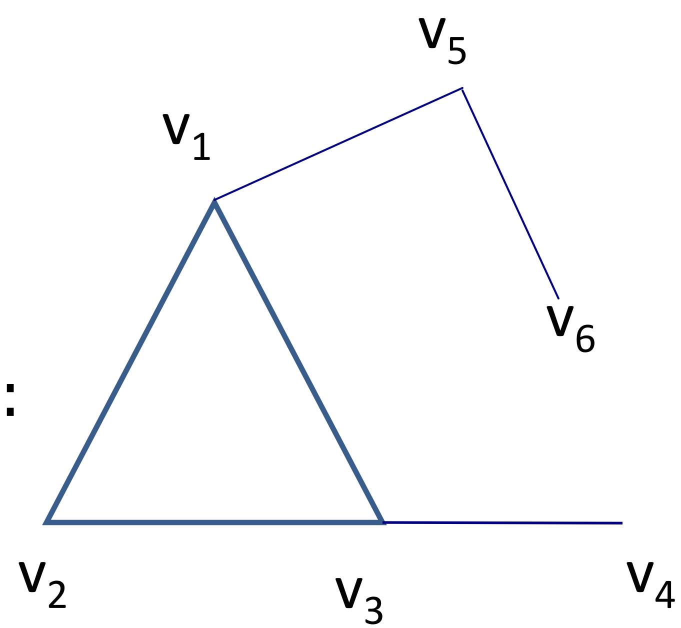
Reason 2 (a more subtle reason):

- Let φ be the FO-sentence

  $\oplus\ v_1, v_2, \ldots, v_6$

- **Fact** (intuitive, but needs proof):

  $\lim_{n \to \infty} pr_n(\varphi) = 1/2$



Modular Convergence Law for FO[$\oplus$]

Main Theorem: For every FO[$\oplus$]-sentence φ, there exist two effectively computable rational numbers a_0, a_1 such that

- $\lim_{n \to \infty} pr_{2n}(\varphi) = a_0$
- $\lim_{n \to \infty} pr_{2n+1}(\varphi) = a_1$.

Moreover,

- a_0, a_1 are of the form $s/2^t$, where s and t are positive integers.
- For every such a_0, a_1, there is a FO[$\oplus$]-sentence φ such that $\lim_{n \to \infty} pr_{2n}(\varphi) = a_0$ and $\lim_{n \to \infty} pr_{2n+1}(\varphi) = a_1$.

In Contrast

- Hella, K …, Luosto - 1996
  LFP[$\oplus$] is *almost-everywhere-equivalent* to PTIME.
  Hence, the modular convergence law **fails** for LFP[$\oplus$].

- Kaufmann and Shelah - 1985
  For every rational number r with 0 < r < 1, there is
  a sentence ψ of monadic second-order logic such that
  $\lim_{n \to \infty} pr_n(\psi) = r$.

Modular Convergence Law

Main Theorem: For every FO[$\oplus$]-sentece φ, there exist two effectively computable rational numbers a_0, a_1 of the form $s/2^t$ such that

- $\lim_{n \to \infty} pr_{2n}(\varphi) = a_0$
- $\lim_{n \to \infty} pr_{2n+1}(\varphi) = a_1$.

Proof Ingredients:

- Elimination of quantifiers.
- Counting results obtained via algebraic methods used in the study of pseudorandomness in computational complexity.
  - Functions that are uncorrelated with low-degree multivariate polynomials over finite fields.

Counting Results – Warm-up

Notation: Let H be a fixed connected graph.

- #H(G) = the number of "copies" of H as a subgraph of G
  = |Inj.Hom(H,G)| / |Aut(H)|.

Basic Question:

What is pr(#H(G) is odd), for a random graph G?

Lemma: If H is a fixed connected graph, then for all large n,
pr_n(#H(G) is odd) = $1/2 + 1/2^n$.

Proof uses results of Babai, Nisan, Szegedy – 1989.

Counting Results – Subgraph Frequencies

Definition: Let m be a positive integer and let $H_1,\ldots,H_t$ be an enumeration of all distinct connected graphs that have at most m vertices.

- The m-subgraph frequency vector of a graph G is the vector

$$freq(m,G) = (\#H_1(G) \bmod 2, \ldots, \#H_t(G) \bmod 2)$$

Theorem A: For every m, the distribution of freq(m,G) in G(n,1/2) is $1/2^n$-close to the uniform distribution over $\{0,1\}^t$, except for $\#K_1 = n \bmod 2$, where K_1 is  .

Quantifier Elimination

Theorem B: The asymptotic probabilities of FO[$\oplus$]-sentences are "determined" by subgraph frequency vectors.

More precisely:

For every FO[$\oplus$]-sentence φ, there are a positive integer m and a function g: $\{0,1\}^t \to \{0,1\}$ such that for all large n,

$$pr_n(G \vDash \varphi \;\Leftrightarrow\; g(freq(m,G))=1) \;=\; 1-1/2^n.$$

Quantifier Elimination

Theorem B: The asymptotic probabilities of FO[$\oplus$]-sentences are "determined" by subgraph frequency vectors.

Proof: By quantifier elimination.
However, one needs to prove a stronger result about formulas with free variables ("induction loading device").

Roughly speaking, the stronger result asserts that:

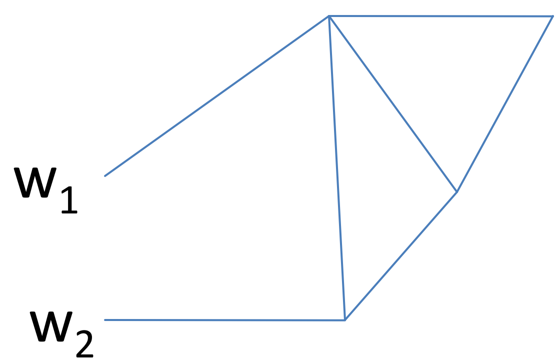
The asymptotic probability of every FO[$\oplus$]-formula $\varphi(w_1, \ldots, w_k)$ is determined by:

- Subgraph induced by $w_1, \ldots, w_k$.
- Subgraph frequency vectors of graphs anchored at $w_1, \ldots, w_k$.

Quantifier Elimination

Notation:

- $\text{type}_G\,(w_1, \ldots, w_k)$ = Subgraph of G induced by $w_1,\ldots,w_{\underline{k}}$
- Types(k) = Set of all k-types
- $\text{freq}(m,G,w_1, \ldots, w_k)$ = Subgraph frequencies of graphs anchored at $w_1, \ldots, w_k$



Quantifier Elimination

Theorem B': For every FO[$\oplus$]-formula $\varphi(w_1, \ldots, w_k)$, there are a positive integer m and a function h: Types(k) x $\{0,1\}^t \rightarrow \{0,1\}$ such that for all large n,

$pr_n(\forall \mathbf{w}\ (G \models \varphi(\mathbf{w}) \Leftrightarrow\ h(type_G(\mathbf{w})_,\ freq(m,G\ \mathbf{w}))=1)) = 1\text{-}1/2^n$.

Note:

- k = 0 is **Theorem B**.
- $\varphi(w_1, \ldots, w_k)$ quantifier-free is trivial: determined by type.

Modular Convergence Law for FO[$\oplus$]

Theorem A: For every m, the distribution of freq(m,G) in G(n,1/2) is $1/2^n$-close to the uniform distribution over $\{0,1\}^t$, except for $\#K_1 = n \bmod 2$, where K_1 is  .

Theorem B: For every FO[$\oplus$]-sentence φ, there are a positive integer m and a function g: $\{0,1\}^t \to \{0,1\}$ such that for all large n, $pr_n(G \vDash \varphi \;\Leftrightarrow\; g(freq(m,G))=1) \;=\; 1-1/2^n$.

Main Theorem: For every FO[$\oplus$]-sentence φ, there exist effectively computable rational numbers a_0, a_1 of the form $s/2^t$ such that

- $\lim_{n \to \infty} pr_{2n}(\varphi) = a_0$
- $\lim_{n \to \infty} pr_{2n+1}(\varphi) = a_1$.

Realizing All Possible Limits of Subsequences

- For every a_0, a_1 of the form $s/2^t$, there is a FO[$\oplus$]-sentence φ such that $\lim_{n \to \infty} pr_{2n}(\varphi) = a_0$ and $\lim_{n \to \infty} pr_{2n+1}(\varphi) = a_1$.

- **Example**: Take two rigid graphs H and J
  Let φ be the FO[$\oplus$]-sentence asserting
  "(G has an even number of vertices, an odd number of copies of H, and an odd number of copies of J) or
  (G has an odd number of vertices and odd number of copies of H)"
  Then
  - $\lim_{n \to \infty} pr_{2n}(\varphi) = 1/4$
  - $\lim_{n \to \infty} pr_{2n+1}(\varphi) = 1/2.$

Modular Convergence Law for FO[Mod_q]

Theorem: Let q be a prime number.
For every FO[Mod_q]-sentece φ, there exist effectively computable rational numbers $a_0, a_1, \ldots, a_{q-1}$ of the form s/q^t such that for every i with $0 \leq i \leq q-1$,

$$\lim_{n \equiv i \bmod q,\ n \to \infty} pr_n(\varphi) = a_i .$$

Open Problems

- What is the complexity of computing the limiting probabilities of FO[$\oplus$]-sentences?
  - PSPACE-hard problem;
  - In Time($2^{2^{\cdots}}$).

- Is there a modular convergence law for FO[Mod_6]?
  More broadly,
  - Understand FO[Mod_6] on random graphs.
  - May help understanding $AC^0[Mod_6]$ better.

- Modular Convergence Laws for FO[$\oplus$] on $G(n, n^{-a})$?