First Order Logic of Sparse Random Hyper-Graphs

Lázaro Alberto Larrauri Borroto

July 9, 2019

Supervisor: Marc Noy Serrano

The first order language of graphs.

- Variables $x_1, \ldots, x_n, \ldots$
- Connectives $\wedge, \vee$, equality symbol $=$, and negation symbol $\neg$.
- Quantifiers $\forall, \exists$.
- A binary relation symbol R

- Vertices.
- "And", "or", "equals", "not".
- "For all", "there exists".
- Edges $x \sim y$.

$$\forall x_1, x_2\, R(x_1, x_2) \implies \exists x_3 (\neg(x_3 = x_1) \wedge \neg(x_3 = x_2) \wedge R(x_1, x_3))$$

The binomial model

The binomial model of random graphs $G(n, p)$ is a discrete probability space where we assign to each graph $G = ([n], E)$ the probability

$$\Pr(G) = p^{|E|} \cdot (1-p)^{\binom{n}{2}-|E|}.$$

Lynch's theorem

Theorem (Lynch, 1992)

Let φ *be a sentence in the F.O. language of graphs. Then the map* $\mathrm{F}_\varphi : [0,\infty) \to \mathbb{R}$ *given by*

$$\mathrm{F}_\varphi(\beta) = \lim_{n\to\infty} \Pr(\, G(n, \beta/n) \models \varphi\,)$$

is well defined and is analytic with respect to β*.*

Overview of the proof

Some properties of $G(n, \beta/n)$:

- The number of cycles of length $3, 4 \dots, r$ are asymptotically distributed like independent Poisson variables.
- Small cycles are a.a.s. far away.
- Fixed vertices are a.a.s. far away.
- The ball of a given radius centered in fixed vertex is a.a.s. a tree. Any tree occurs with a positive probability.

Overview of the proof

For each fixed quantifier rank k:

(1) It is given a finite classification of "small" uni-cycles.

(2) It is shown that the rank k type of random graph G in $G(n, \beta/n)$ a.a.s. depends exclusively on the number of "small" uni-cycles belonging to each class.

(3) The asymptotic distribution of those quantities is obtained.

Edge sets

Definition

The **total edge set** $\mathcal{H}_{(a,\Phi,A)}(n)$ of **size** a, **symmetry group** Φ and **restrictions** A, on n **elements** is the set

$$\mathcal{H}_{(a,\Phi,A)}(n) = ([n]^a/\Phi) \ \setminus R,$$

where

$$R = \{\, [x_1, \dots, x_a] \in [n]^a/\Phi \mid x_i = x_j \text{ for some } (i,j) \in A\}$$

Graphs

Definition

An **(hyper)-graph** $([n], H_1, \ldots, H_c)$ with edge colors $1, \ldots, c$, sizes $a_1, \ldots, a_c$, symmetry groups $\Phi_1, \ldots, \Phi_c$ and restrictions $A_1, \ldots, A_c$ consists of

- The **vertex set** $[n]$ for some natural number n.
- For $i = 1, \ldots, c$, a "colored" **edge set** $H_i \subseteq \mathcal{H}_{(a_i, \Phi_i, A_i)}(n)$ whose elements have color i.

The first order language

Consider the first order purely relational language $\mathcal{L}$ with relation symbols $R_1, \ldots, R_c$ with arities $a_1, \ldots, a_c$.

A graph $G = ([n], H_1, \ldots, H_c)$ is a $\mathcal{L}$-structure in the following way:

- The universe of G is its vertex set, $[n]$.
- For each $1 \leq i \leq c$,

$$(x_1, \ldots, x_{a_i}) \in R_i^G \iff [x_1, \ldots, x_{a_i}] \in H_i.$$

The first order language

A graph $G = ([n], H_1, \dots, H_n)$ satisfies, for each $1 \le i \le c$:

- **Symmetry formulas**:

$$S_g := \big(R_i(x_1 \dots, x_{a_i}) \iff R_i(x_{g(1)} \dots, x_{g(a_i)})\big),$$

  where g is an element from Φ_i.

- **Anti-reflexivity formulas**:

$$AR_{i,(j,l)} := \big(R_i(x_1 \dots, x_{a_i}) \implies \neg(x_j = x_l)\big),$$

  where $(j, l) \in A_i$.

The random model

The random model $HG(n, p_1, \dots, p_c)$ is a discrete probability space where for each graph $G = ([n], H_1, \dots, H_c)$,

$$\Pr(G) = \prod_{i=1}^{c} p_i^{|H_i|} \cdot (1 - p_i)^{|\mathcal{H}_{(a_i, \Phi_i, A_i)}(n)| - |H_i|}.$$

We consider the **sparse** case where for each $1 \le i \le c$, $p_i(n) = \beta_i / n^{a_i - 1}$.

The theorem

We want to prove the following

Theorem

Let φ be a first order sentence in $\mathcal{L}$. Then the map $\mathrm{F} : [0, \infty)^c \to \mathbb{R}$ given by

$$\mathrm{F}(\beta_1, \dots, \beta_c) = \lim_{n\to\infty} \mathrm{Pr}(HG(n, p(n, \beta)) \models \varphi),$$

where

$$p(n, \beta) := (\beta_1/n^{a_1-1}, \dots, \beta_c/n^{a_c-1})$$

is well defined and is analytic with respect to $\beta = (\beta_1, \dots, \beta_c)$.

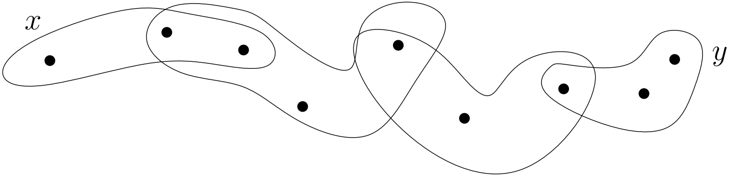
Distance and paths

Given any graph G, we define the following distance over its vertex-set:

$$d(x,y) = \min_{\substack{H \leq G, x,y \in V(H) \\ H \text{ connected}}} \big(|V(H)| - 1\big), \quad \text{or } \infty \text{ if } x,y \text{ are not connected.}$$

Definition

A path between two vertices x, y in a graph G is a minimal graph among the connected subgraphs $H \leq G$ containing both x, y.



Likelihood, trees, cycles and clusters

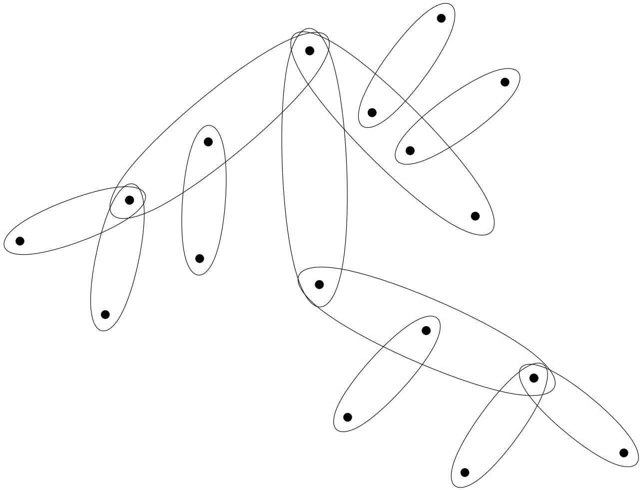
Definition

The likelihood $L(G)$ of a graph $G = (V, H_1, \ldots, H_c)$ is the number

$$|V(G)| - \sum_{i=1}^{c} |H_i|(a_i - 1).$$

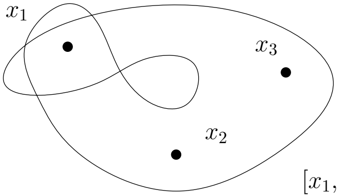
Likelihood, trees, cycles and clusters

- A tree is a connected graph with likelihood 1.

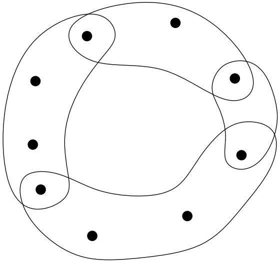


Likelihood, trees, cycles and clusters

- An unicycle is a connected graph with likelihood 0. A cycle is a minimal unicycle.

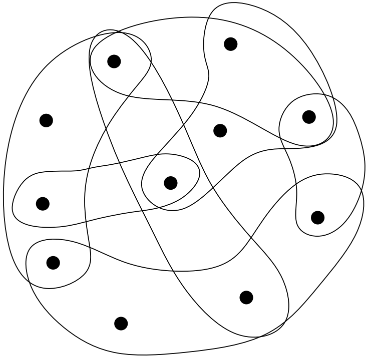


$[x_1, x_2, x_3, x_1]$



Likelihood, trees, cycles and clusters

- A cluster is a graph G with $L(G) \leq 0$ such that $L(H) > L(G)$ for any subgraph $H \leq G$.



The k-morphism relation over trees.

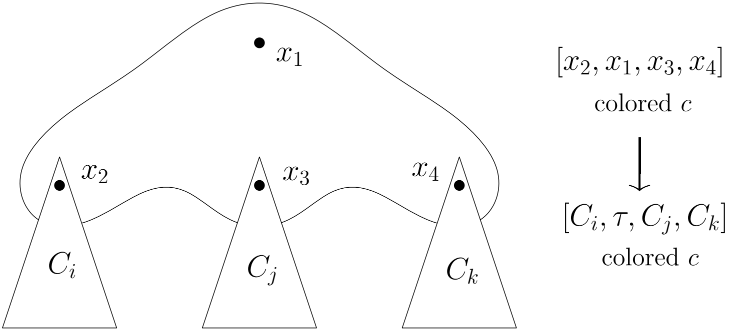
Fix $k \in \mathbb{N}$.
We define the k-morphism relation over rooted trees of the same radius inductively as follows:

- If $r(T_1) = r(T_2) = 0$ then $T_1 \overset{k}{\simeq} T_2$.

The k-morphism relation over trees.

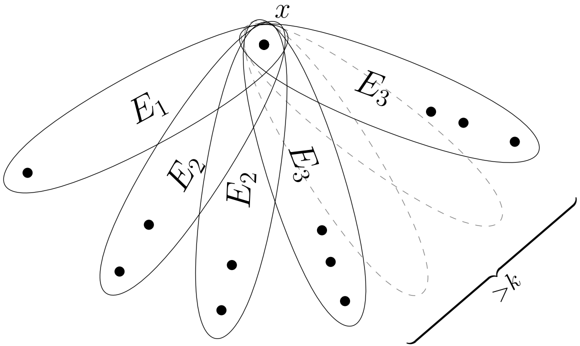
If $r(T_1) = r(T_2) > 0$:

- First we define the k-type of a an initial edge:



The k-morphism relation over trees.

- We say that $T_1 \overset{k}{\simeq} T_2$ if for any defined edge k-type E either:
  - the number of initial edges in T_1 and T_2 of k-type E is the same, or
  - both T_1 and T_2 contain no less than $k+1$ initial edges of k-type E.



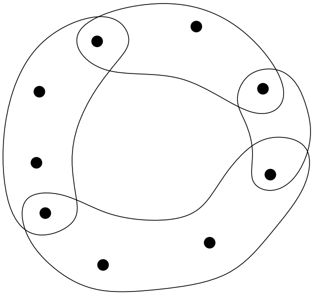
The k-morphism relation for uni-cycles

Definition

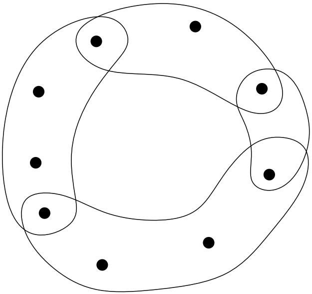
Let G_1, G_2 be uni-cycles. Let O_1, O_2 be their cycles respectively. $G_1 \overset{k}{\simeq} G_2$, if for some f:

- $O_1 \overset{f}{\simeq} O_2$, and
- $Tree(x, G_1) \overset{k}{\simeq} Tree(f(x), G_2)$ for all $x \in V(O_1)$.

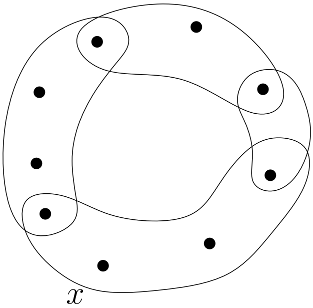
The k-morphism relation for uni-cycles



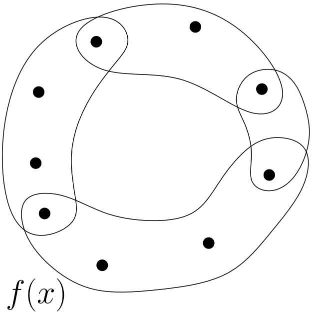
$$\overset{f}{\simeq}$$



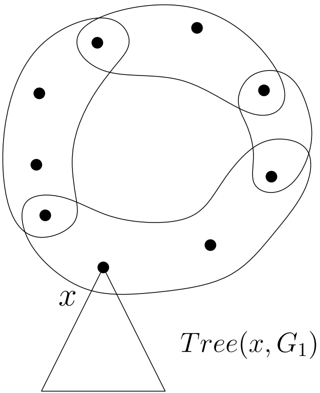
The *k*-morphism relation for uni-cycles

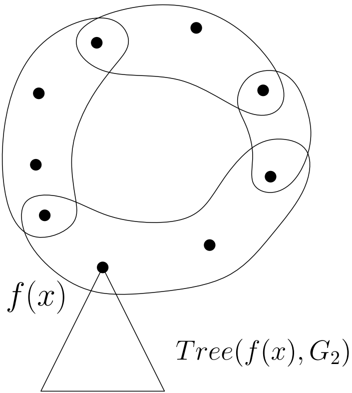


$\underset{\sim}{f}$



The k-morphism relation for uni-cycles





The k-morphism relation for cycles

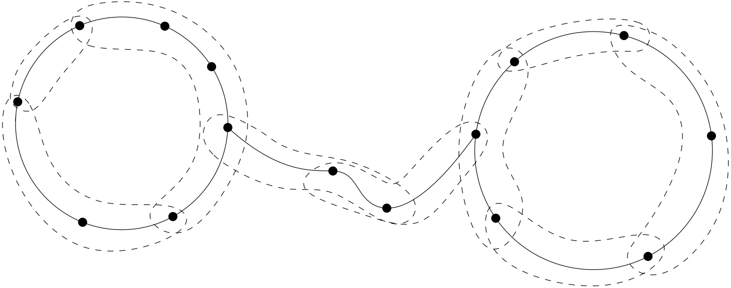
Uni-cyclic k-morphic graphs have the same first order rank k-type.

The tool for proving this are **Ehrenfeucht Fraisse games**.

The k-morphism relation gives, for any r, a **finite classification** of the cycles of diameter at most r with trees of radii at most r hanging from their vertices.

The landscape of $HG(n, p(n, \beta))$

- Clusters with diameter at most d, for any fixed d, a.a.s. will not appear. In particular, small cycles are a.a.s. far away.



The landscape of $HG(n, p(n, \beta))$

- Given any fixed cycles $O_1, \ldots, O_l$, their quantities converge in distribution to independent Poisson variables.
- Fixed vertices are a.a.s. far away.
- The ball of a given radius centered in fixed vertex is a.a.s. a tree. Any tree occurs with a positive probability.

In consequence, F.O. properties of a fixed quantifier rank k only depend on the small neighbourhoods of the small cycles in a graph.

Probabilities of uni-cycles

The main tool for computing probabilities is the following:

Theorem

(Multivatiate Brun's Sieve) For $1 \leq i \leq l$ *let* $\{X_i(n)\}_{n \in \mathbb{N}}$*, be successions of random variables s.t the* $X_i(n)$*'s are sums of random indicator variables. If for any natural numbers* $a_1, \dots, a_l$

$$\lim_{n \to \infty} E\left[\prod_{i=1}^{l} \binom{X_i}{a_i}\right] = \prod_{i=1}^{l} \frac{\lambda_i^{a_i}}{a_i!},$$

then

$$\forall x_1, \dots, x_l \in \mathbb{N}: \quad \lim_{n \to \infty} \Pr(\wedge_{i=1}^{l} X_i = x_i) = \prod_{i=1}^{l} \mathrm{Poiss}_{\lambda_i}(x_i).$$

Ending summary

- $HG(n, p(\beta, n))$ is very similar to $G(n, \beta/n)$. The ideas behind Lynch's proof can be used here after finding suitable definitions.
- The main increase in complexity is due to the fact that trees are more complicated in $HG(n, p(\beta, n))$: the types of the edges have to be considered now.
- Using Brun's Sieve simplifies the harder combinatorial proofs.

Ending summary

- $HG(n, p(\beta, n))$ is very similar to $G(n, \beta/n)$. The ideas behind Lynch's proof can be used here after finding suitable definitions.

- The main increase in complexity is due to the fact that trees are more complicated in $HG(n, p(\beta, n))$: the types of the edges have to be considered now.

- Using Brun's Sieve simplifies the harder combinatorial proofs.

Thank you for attention!