Structural Limits

Jaroslav Nešetřil Patrice Ossona de Mendez

Charles University
Praha, Czech Republic

LIA STRUCO

CAMS, CNRS/EHESS
Paris, France

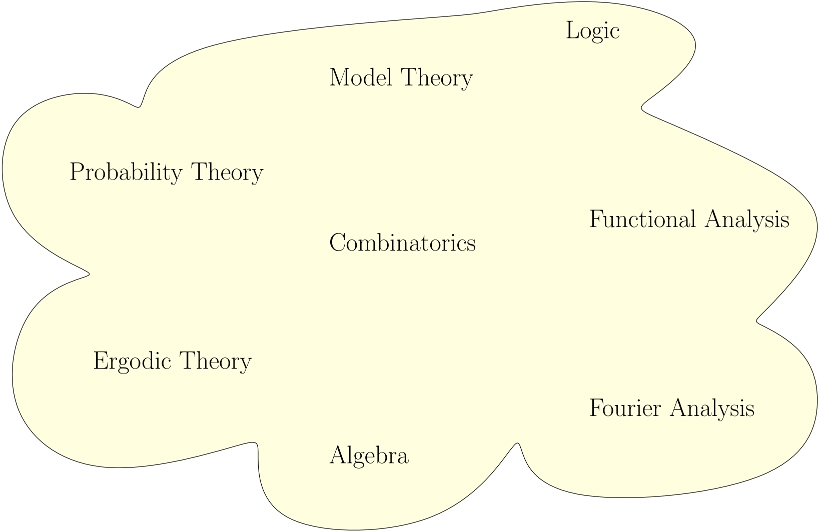
— Minicourse — Paris 2018 —

Introduction

Issues

- How to describe/approximate a network?
- How much is a network structured? How much is it random-like?
- How to check whether a network has (or is close to have) some property?
- How to compare the structures of two networks?
- How to represent limits of networks?
- Asymptotic structure of the networks in a convergent sequence?

Tools


Logic
Model Theory
Probability Theory
Functional Analysis
Combinatorics
Ergodic Theory
Fourier Analysis
Algebra

Many kinds of limits

- Scaling limit (Le Gall)
- Elementary limit (Gödel, Löwenheim–Skolem)
- Left limit (Lovász *et al.*; Aldous-Lyons)
- Local limit (Benjamini, Schramm)
- Structural limits (Nešetřil, OdM)
- Local-global limit (Bollobás, Riordan)
- Local-global structural limits (Nešetřil, OdM)

Menu

Menu

First Course

Fricassee of classical limit

Second Course

S

Third Course

Modeling limit

De

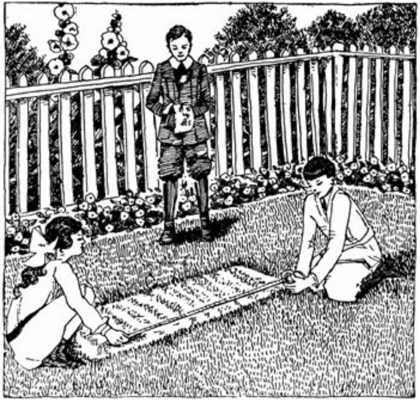
Sweet o

Classical limits

Three main types

Scaling	Non-standard	Sampling
Metric	Logic	Statistics
Gromov–Hausdorff limit	Ultralimit	Limit distribution

Metric convergence



Hausdorff metric

For a subset A of a metric space (X, d) and a positive real $r > 0$ let $A^{(r)} = \bigcup_{x \in A} B_r(x)$.

Definition

Let A, B be closed subsets of X. The *Hausdorff distance* of A and B is

$$d_H(A, B) = \inf\{r > 0 \mid B \subset A^{(r)} \text{ and } A \subset B^{(r)}\}.$$

Theorem

If (X, d) is a compact metric space, then space of all closed subsets of X with Hausdorff metric is a compact metric space.

Gromov-Hausdorff metric

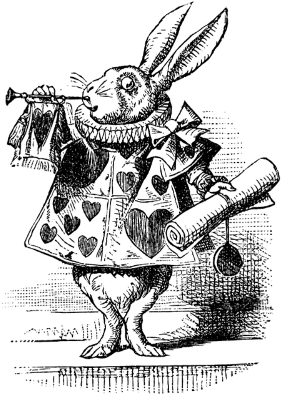
Definition

Let A, B be closed metric spaces. The *Gromov-Hausdorff distance* of A and B is

$$d_{GH}(A, B) = \inf_{f,g} d_H(f_{A \hookrightarrow X}(A), g_{B \hookrightarrow X}(B)),$$

where $f_{A \hookrightarrow X}$ (resp. $g_{B \hookrightarrow X}$) denotes an isometric embedding of A (resp. B) into some metric space X.

Logical convergence



First-Order Logic

- *atomic formulas*, *Boolean formulas*, *existential first-order formulas*, *first-order formulas*.
- The *quantifier rank* $\mathrm{qrank}(\phi)$ of ϕ is the maximum nesting of quantifiers of its sub-formulas.

For a formula $\phi[x_1, \ldots, x_n]$ with free variables $x_1, \ldots, x_n$,

$$G \models \phi[a_1, \ldots, a_n] \iff \phi \text{ is true in } G \text{ when } x_i \leftarrow a_i.$$

Elementary equivalence

Definition

Two graphs G and H are *elementarily equivalent*, noted $G \equiv H$ if they satisfy the same first-order sentences.

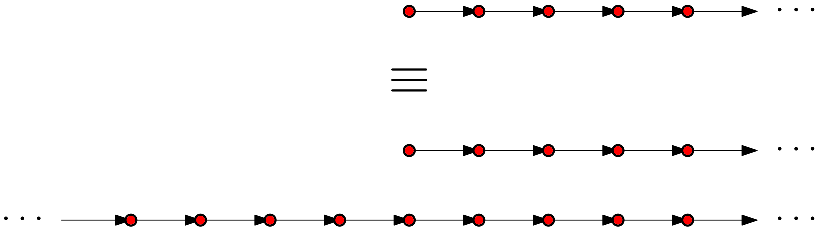
G and H are *n-elementarily equivalent*, noted $G \equiv^n H$, if they satisfy the same first order sentences of quantifier rank n.

The n-elementary equivalence (and the elementary equivalence) can be checked using Ehrenfeucht-Fraïssé games.

Elementary Equivalence

Remark

Two **finite** elementary equivalent structures are isomorphic, but it is not usually the case for infinite structures.



Compactness

Let L and $G_1, G_2, \ldots$ be graphs. Then G_n *converges elementarily* to L if for every ϕ,

$$L \models \phi \quad \Rightarrow \quad \exists n_0,\ \forall n \geq n_0 \quad G_n \models \phi.$$

Compactness

Let L and $G_1, G_2, \ldots$ be graphs. Then G_n *converges elementarily* to L if for every ϕ,

$$L \models \phi \quad \Rightarrow \quad \exists n_0,\ \forall n \geq n_0 \quad G_n \models \phi.$$

Theorem (Sequential compactness theorem)

Let $G_1, G_2, \ldots$ be a sequence of graphs. Then there exists a subsequence that converges elementarily to a limit L, which is at most countable.

Compactness

Let L and $G_1, G_2, \ldots$ be graphs. Then G_n *converges elementarily* to L if for every ϕ,

$$L \models \phi \quad \Rightarrow \quad \exists n_0,\ \forall n \geq n_0 \quad G_n \models \phi.$$

Theorem (Sequential compactness theorem)

Let $G_1, G_2, \ldots$ be a sequence of graphs. Then there exists a subsequence that converges elementarily to a limit L, which is at most countable.

Remark

The limit is unique up to elementary equivalence.

Metric space

Let $\mathcal{G}raph_\omega$ = at most countable graphs.
For $\mathfrak{G}, \mathfrak{H} \in \mathcal{G}raph_\omega / \equiv$ define

$$\operatorname{dist}(\mathfrak{G}, \mathfrak{H}) = 2^{-sup\{n, \quad G \equiv^n H,\ G \in \mathfrak{G}, H \in \mathfrak{H}\}}.$$

Then dist is an ultrametric and $(\mathcal{G}raph_\omega / \equiv, \operatorname{dist})$ is a **compact metric space**.

Metric space

Let $\mathcal{G}raph_\omega$ = at most countable graphs.
For $\mathfrak{G}, \mathfrak{H} \in \mathcal{G}raph_\omega / \equiv$ define

$$\mathrm{dist}(\mathfrak{G}, \mathfrak{H}) = 2^{-\sup\{n, \quad G \equiv^n H,\ G \in \mathfrak{G}, H \in \mathfrak{H}\}}.$$

Then dist is an ultrametric and $(\mathcal{G}raph_\omega / \equiv, \mathrm{dist})$ is a **compact metric space**.

Remark

According to Gödel completeness theorem, elements of $\mathcal{G}raph_\omega / \equiv$ correspond to the complete theories of graphs.

Limit Object

Proposition (Gödel+Löwenheim–Skolem)

Every elementarily convergent sequence of finite graphs has a limit, which is an at most countable graph.

BUT

No characterization of elementary limits of finite graphs

Trakhtenbrot's theorem states that the problem of existence of a finite model for a single first-order sentence is undecidable.

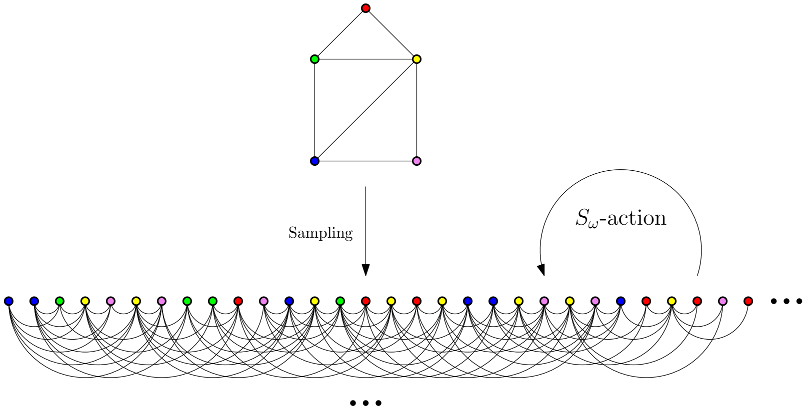
Sampling and Left convergence

Sampling and Left convergence

Problem

Consider a very large network and pick N random sample vertices. What can you learn about the large network by looking at these N vertices?

Infinite Exchangeable Graph



Left Convergence and Lovász Profile

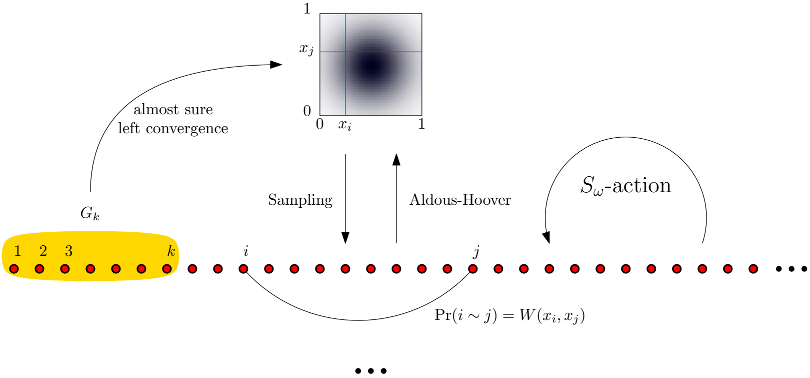
$$t(F,G) := \frac{\hom(F,G)}{|G|^{|F|}} = \Pr[f : F \to G \text{ is a homomorphism}].$$

$(G_n)_{n\in\mathbb{N}}$ is *left convergent* if $t(F,G_n)$ converges for each F.

Example (Pseudorandom graphs)

Assume $t(K_2,G_n) \to p$ and $t(C_4,G_n) \to p^4$.
Then $(G_n)_{n\in\mathbb{N}}$ is left-convergent (Chung, Graham, Wilson '89)

Infinite Exchangeable Graph



Sampling

Problem

Consider a very large network and pick N random sample vertices. What can you learn about the large network by looking at these N vertices?
Can you decide wether it is likely that the large network is close or far from having some given property P?

This is the setting of Property Testing

Example: is a large graph triangle-free?

Pick 3 random vertices and check wether they form a triangle.
If yes **reject**.

By repeating the same test $2/c$ times

- if G is triangle-free we **accept** it with probability 1,
- if G contains at least $c|G|^3$ triangles, we **reject** G with probability $> 2/3$.

Example: is a large graph triangle-free?

Pick 3 random vertices and check wether they form a triangle.
If yes **reject**.

By repeating the same test $2/c$ times

- if G is triangle-free we **accept** it with probability 1,
- if G contains at least $c|G|^3$ triangles, we **reject** G with probability $> 2/3$.

Triangle Removal Lemma (Ruzsa and Szemerédi, 1976)

$\forall \epsilon > 0\ \exists c > 0$ such that if a graph G contains at most $c|G|^3$ triangles then is possible to remove at most $\epsilon|G|^2$ edges and end up with a graph that is triangle-free.

Example: is a large graph triangle-free?

Pick 3 random vertices and check wether they form a triangle.
If yes **reject**.

By repeating the same test $f(\epsilon)$ times

- if G is triangle-free we **accept** it with probability 1,
- if at least $\epsilon\binom{|G|}{2}$ edges have to be deleted from G to make it triangle-free, we **reject** G with probability $> 2/3$.

Triangle Removal Lemma (Ruzsa and Szemerédi, 1976)

$\forall\epsilon > 0\ \exists c > 0$ such that if a graph G contains at most $c|G|^3$ triangles then is possible to remove at most $\epsilon|G|^2$ edges and end up with a graph that is triangle-free.

Example: is a large graph triangle-free?

Pick 3 random vertices and check wether they form a triangle.
If yes **reject**.

By repeating the same test times

- if G is triangle-free we **accept** it with probability 1,
- if G is ϵ-far from being triangle-free, we **reject** G with probability $> 2/3$.

Triangle Removal Lemma (Ruzsa and Szemerédi, 1976)

$\forall \epsilon > 0\ \exists c > 0$ such that if a graph G contains at most $c|G|^3$ triangles then is possible to remove at most $\epsilon|G|^2$ edges and end up with a graph that is triangle-free.

How to prove the removal lemma?

Szemerédi Regularity Lemma

Theorem (Szemerédi, 1978)

$\forall m$ and $\epsilon > 0$, $\exists T$ such that every G of order $\geq T$ has an ϵ-regular equipartition of order k, where $m \leq k \leq T$.

- A pair (X, Y) is ϵ-regular if $\forall X' \subseteq X \; \forall Y' \subseteq Y$

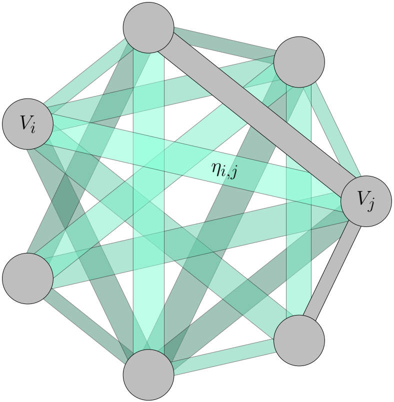
  $$|X'| \geq \epsilon|X| \text{ and } |Y'| \geq \epsilon|Y| \implies |d(X,Y) - d(X',Y')| < \epsilon.$$

- A partition $V_0, V_1, \ldots, V_k$ is an ϵ-regular equipartition if $|V_i| = |V_j|$ (for $1 \leq i < j \leq k$), $|V_0| \leq \epsilon|G|$ and if all the pairs (V_i, V_j) but an ϵ fraction are ϵ-regular.

Szemerédi Regularity Lemma

Theorem (Szemerédi, 1978)

$\forall m$ and $\epsilon > 0$, $\exists T$ such that every G of order $\geq T$ has an ϵ-regular equipartition of order k, where $m \leq k \leq T$.



The Trilogy

Counting Lemma

Removing Lemma

Generalizations of the removal lemma

Theorem (Erdős, Frankl and Rödl 1986)

For any $0 < \epsilon < 1$ there is $\delta = \delta(\epsilon) > 0$ such that every graph G that is ϵ-far from being H-free contains at least $\delta|G|^{|H|}$ copies of H.

Theorem (Alon, Fischer, Krivelevich and (M.) Szegedy 2000)

For any $0 < \epsilon < 1$, there is $\delta = \delta(\epsilon) > 0$ such that every graph G that is ϵ-far from being induced H-free contains at least $\delta|G|^{|H|}$ induced copies of H.

Generalizations of the removal lemma

Theorem (Alon, Shapira 2005)

$\forall \mathcal{F}$, $\forall \eta > 0$, $\exists c > 0, C > 0, n_0$ such that $\forall H$ of order $n \geq n_0$: if $\forall F \in \mathcal{F}$ with order $|F| \leq C$, H contains $\leq cn^{|F|}$ copies of F, then one can delete $\eta\binom{n}{2}$ edges from H to get H' which contains no copy of any member of $\mathcal{F}$.

$$\Downarrow$$

Theorem (Alon, Shapira 2005)

Every monotone graph property is testable.

Is Regularity Lemma essential?

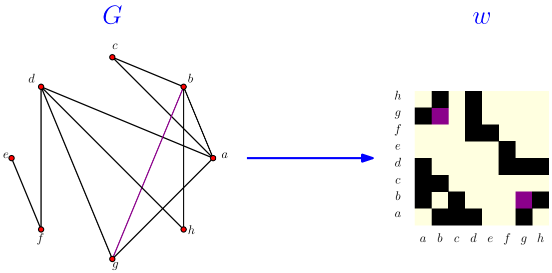
Theorem (Alon, Fischer, Newman, Shapira 2006)

A graph property is testable if and only if it is regular-reducible.

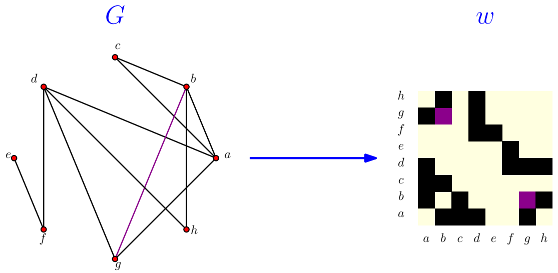
In other words (intuitively)

A graph property $\mathcal{P}$ is testable if and only if $\mathcal{P}$ is such that knowing a regular partition of a graph G is sufficient for telling whether G is ϵ-far or ϵ-close to satisfying $\mathcal{P}$.

Graphons: Intuition



Graphons: Intuition



Graphons: Definition

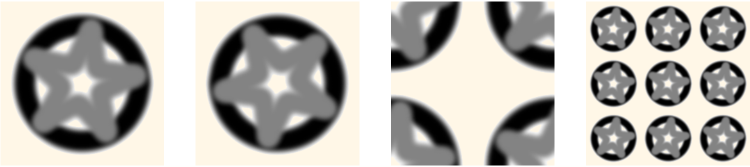
A *graphon* is a measurable symmetric $w : [0,1] \times [0,1] \to [0,1]$;

$\mathcal{W}_0$ is the set of all graphons.

Γ is the group of measure preserving bijections on $[0,1]$

$\Gamma \curvearrowright \mathcal{W}_0$: for $(w, \phi) \in \mathcal{W}_0 \times \Gamma$

$$w^{\phi}(x, y) := w(\phi(x), \phi(y)).$$



The Cut Norm

$$\|w\|_\square := \sup_{S,T\subseteq[0;1]} \left| \int_{S\times T} w(x,y)\, \mathrm{d}x\, \mathrm{d}y \right|$$
$$= \sup_{S\times T\subseteq[0;1]^2} |w(S\times T)|$$

- Fréchet '15; Littlewood '30 (Bilinear Forms)
- Clarkson, Adams '33; Morse '55 (Variation measure)
- ∗ Frieze and Kannan '99 (Matrix approximation)
- ∗ Gowers '01 (Fourier analysis)
- ∗ Borgs, Chayes, Lovász, Sós, Vesztergombi '08 (Graph Limits)

The Cut Metric and the Graphon Space

$$\delta_\square(u,v) := \inf_{\phi\in\Gamma} \|u^\phi - v\|_\square$$

Borgs, Chayes, Lovász, Sós, Vesztergombi '08

Theorem (Lovász, Szegedy)

Let $\mathcal{X}_0$ denote the image of $\mathcal{W}_0$ under the identification $u \sim w$ if $\delta_\square(u,v) = 0$. Then the metric space $(\mathcal{X}_0, \delta_\square)$ is **compact**.

We call $(\mathcal{X}_0, \delta_\square)$ the *Graphon Space*.

Left Limits and Cut Metric

For simple graph F and a graphon $w \in \mathcal{W}_0$ define

$$t(F,w) := \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} w(x_i, x_j) \, \mathrm{d}x.$$

Left Limits and Cut Metric

For simple graph F and a graphon $w \in \mathcal{W}_0$ define

$$t(F,w) := \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} w(x_i, x_j)\, \mathrm{d}x.$$

Lemma (Counting Lemma; Lovász, Szegedy 2006)

Let F be a simple graph and let $u, w \in \mathcal{W}_0$. Then

$$|t(F,u) - t(F,w)| \leq \|F\|\, \delta_\square(u,w).$$

Inverse counting lemma

Theorem (Borgs, Chayes, Lovász, Sós, Vesztergombi 2008)

For $k \geq 1$ and $u, w \in \mathcal{W}_0$, if

$$|t(F,u) - t(F,v)| \leq 2^{-k^2}$$

for every F with $|F| \leq k$ then

$$\delta_\square(u,w) \leq \frac{50}{\sqrt{\log k}}.$$

Corollary

the Lovász profile and the cut metric define the same notion of convergence.

Left convergence (to sum up)

- Approximation: regularity lemma (Szemerédi)
- Property testing (Alon–Shapira, Austin–Tao)
- Convergence of $\delta_\square$ or of Lovász profile

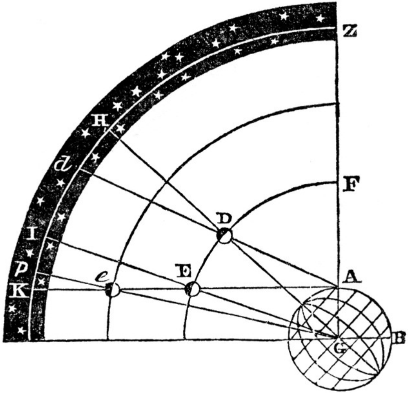
$$t(F, G_n) = \frac{\hom(F, G_n)}{|G_n|^{|F|}}.$$

- Distributional limit as an exchangeable random infinite graph (Aldous–Hoover–Kallenberg, Diaconis–Janson).
- Analytic limit as a graphon (Lovász–Szegedy)

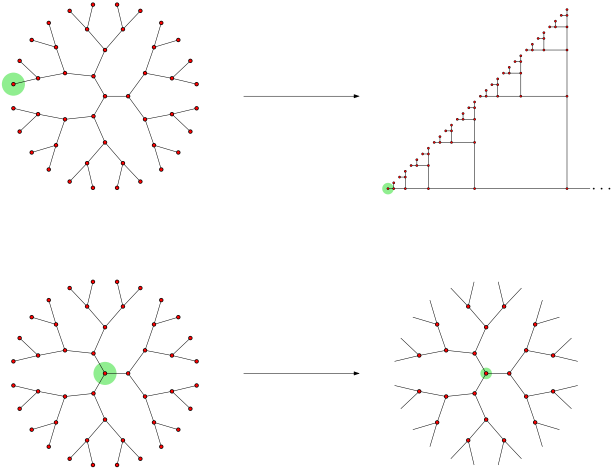
$$\text{symmetric } W : [0,1] \times [0,1] \to [0,1]$$

(up to weak-equivalence)

Sampling and Local convergence



Typical Views



Local convergence

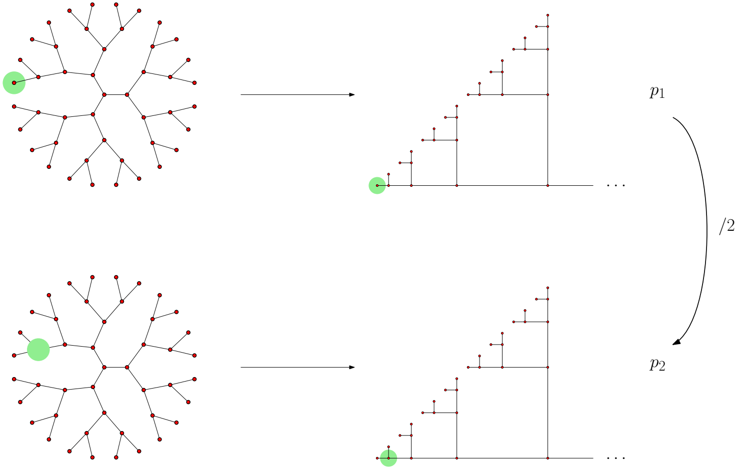
Local profile:

$$P_\rho((F,r),G) := \frac{|\{v : B_\rho(G,v) \simeq (F,r)\}|}{|G|} = \Pr[B_\rho(G,X) \simeq (F,r)].$$

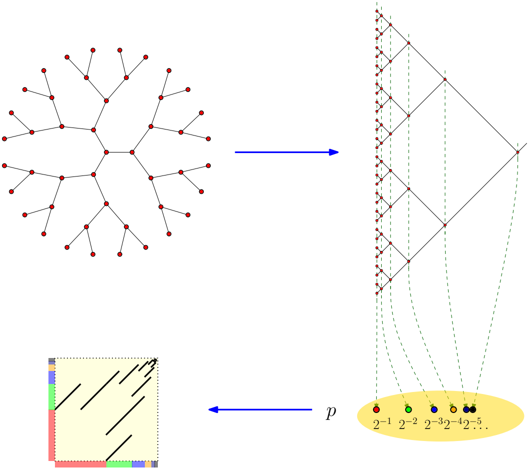
Local convergence:

Convergence of $P_\rho((F,r),G_n)$ for every ρ and (F,r).

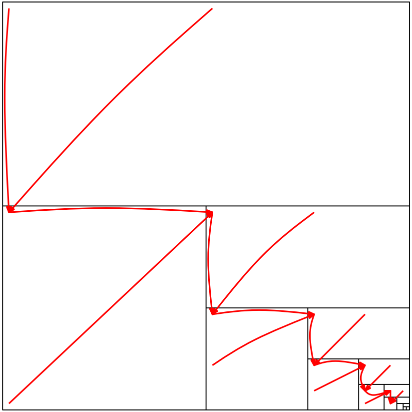
Unimodularity



Statistics on local views



Limit Object?



→ a Borel graph with the same generic vertex properties.

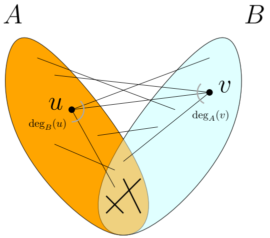
Graphings

From d measure preserving involutions $f_1, \ldots, f_d$

$$x \sim y \iff$$
$$(x \neq y) \text{ and } (\exists i)\ y = f_i(x)$$

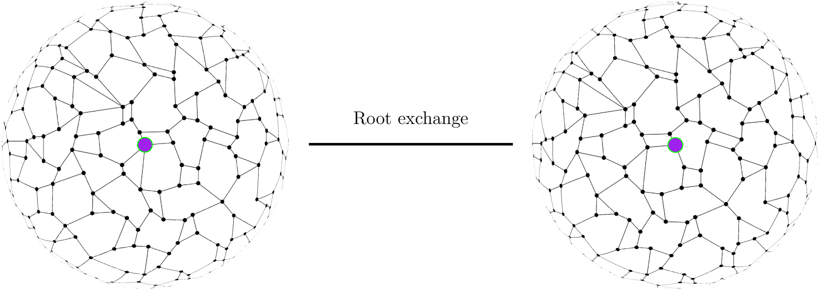
A Borel graph G satisfying the Mass Transport Principle

$$\int_A \deg_B(u)\,\mathrm{d}u = \int_B \deg_A(v)\,\mathrm{d}v.$$



$\Sigma_{u \in A} \deg_B(u) = \Sigma_{v \in B} \deg_A(v)$

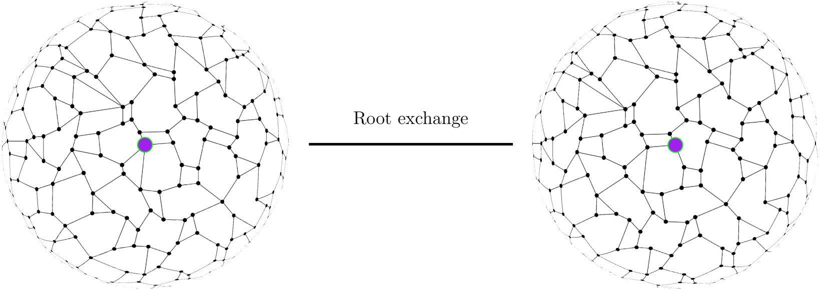
The graph of graphs



Problem

How to cope with symmetries?

The graph of graphs

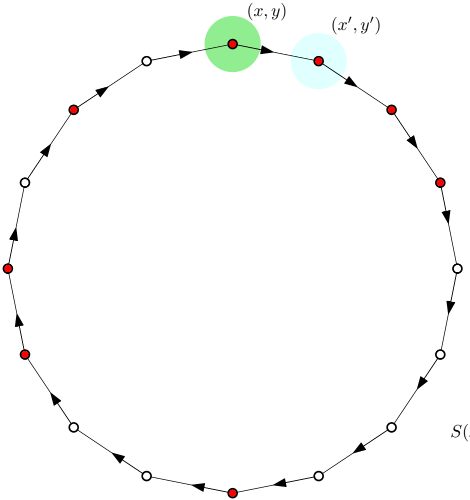


Problem

How to cope with symmetries?

Coloring!

Bernoulli graphing



$x = .1111000010011010\ldots$
$y = .0101100100001111\ldots$

$x' = .111000010011010\ldots$
$y' = .10101100100001111\ldots$

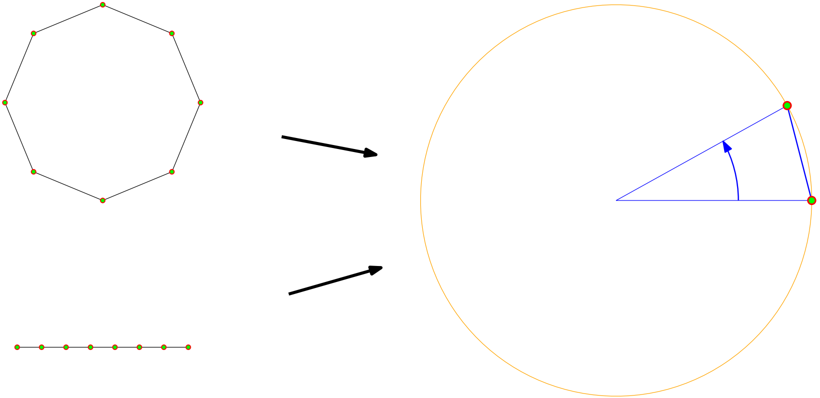
$$(x', y') = S(x, y)$$

where

$$S(x, y) = \left(2x - \lfloor 2x \rfloor, \frac{y + \lfloor 2x \rfloor}{2}\right).$$

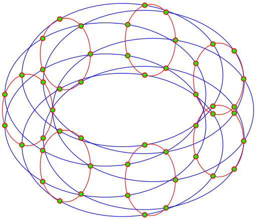
Some handmade examples

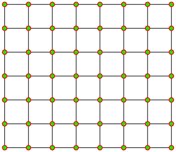
Paths and Cycles

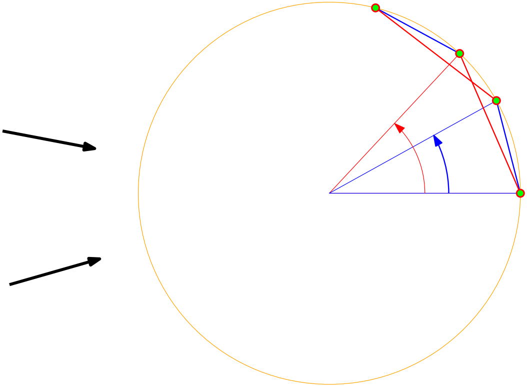


$x \in \mathbb{R}/\mathbb{Z} \qquad x \mapsto x \pm \alpha$

Grids

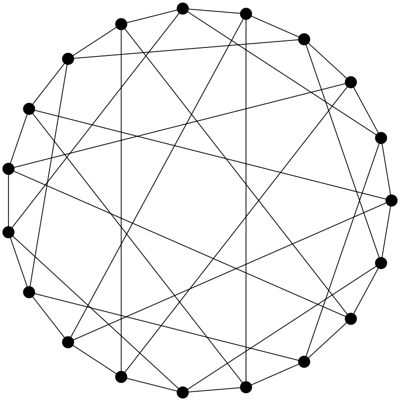






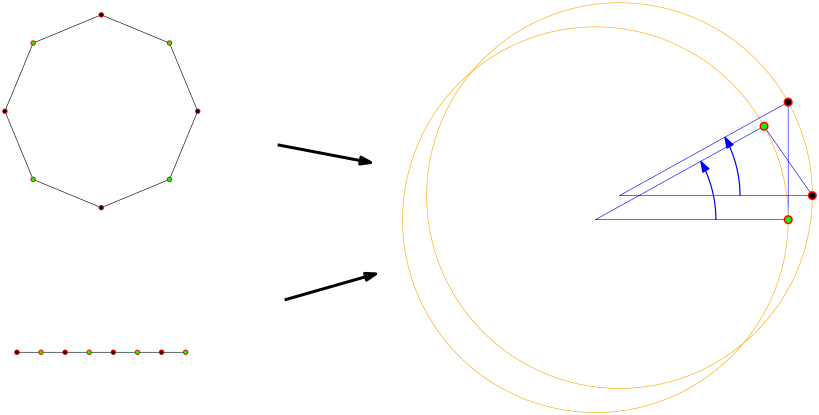
$$x \in \mathbb{R}/\mathbb{Z} \qquad x \mapsto \begin{cases} x \pm \alpha \\ x \pm \beta \end{cases}$$

High-girth Regular Graphs



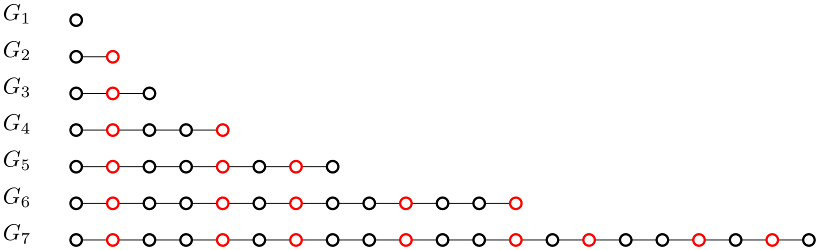
$$(x, y) \in (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}) \qquad (x, y) \mapsto \begin{cases} (x, y) \pm (\alpha, 0) \\ (x, y) \pm (\beta, \beta) \end{cases}$$

Bipartite Paths and Cycles

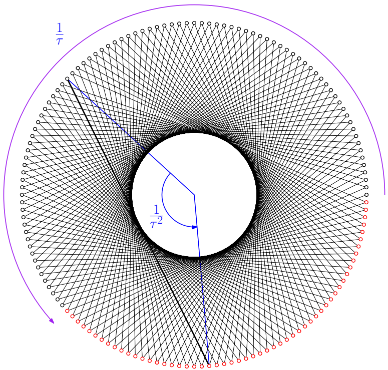


$$(x, c) \in (\mathbb{R}/\mathbb{Z}) \times \{0, 1\} \qquad (x, c) \mapsto (x \pm \alpha, 1 - c)$$

Fibonacci Sequence

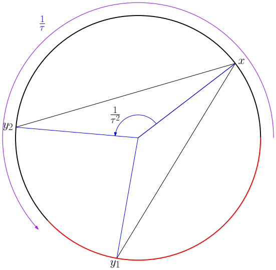


Fibonacci Sequence



G_{11}

Fibonacci Sequence Local Limit

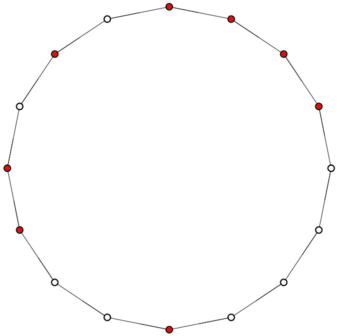


In $\mathbb{R}/\mathbb{Z}$:

$$x \sim y \iff x \equiv y \pm \frac{1}{\tau^2}$$

$$\text{Black}(x) \iff x \in \left[0, \frac{1}{\tau}\right]$$

de Bruijn Sequences



$$S(x,y) = \left(2x - \lfloor 2x \rfloor, \frac{y + \lfloor 2x \rfloor}{2}\right).$$

$x < 1/2 \rightarrow$ white $\qquad x \geq 1/2 \rightarrow$ red

The inverse problem

Conjecture (Aldous, Lyons 2006)

Every graphing is the local limit of a sequence of connected finite graphs.

Would solve…

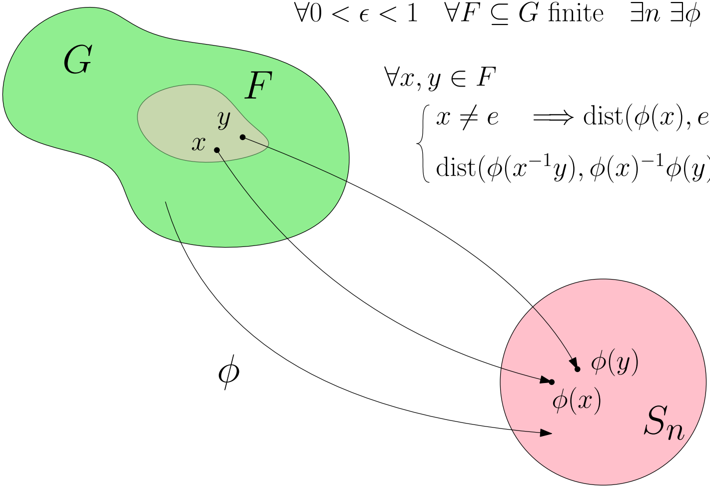
- Weiss conjecture that all finitely generated groups are sofic,
- the Direct Finiteness Conjecture of Kaplansky on group algebras,
- a conjecture of Gottschalk on surjunctive groups in topological dynamics,
- the Determinant Conjecture on Fuglede-Kadison determinants,
- Connes' Embedding Conjecture for group von Neumann algebras.

Sofic groups

$$\forall 0 < \epsilon < 1 \quad \forall F \subseteq G \text{ finite} \quad \exists n \, \exists \phi : G \to S_n$$

$$\forall x, y \in F$$

$$\begin{cases} x \neq e \quad \Longrightarrow \operatorname{dist}(\phi(x), e) > 1 - \epsilon \\ \operatorname{dist}(\phi(x^{-1}y), \phi(x)^{-1}\phi(y)) < \epsilon \end{cases}$$



Local limits (to sum up)

- Property testing (Benjamini–Schramm–Shapira)
- Convergence of

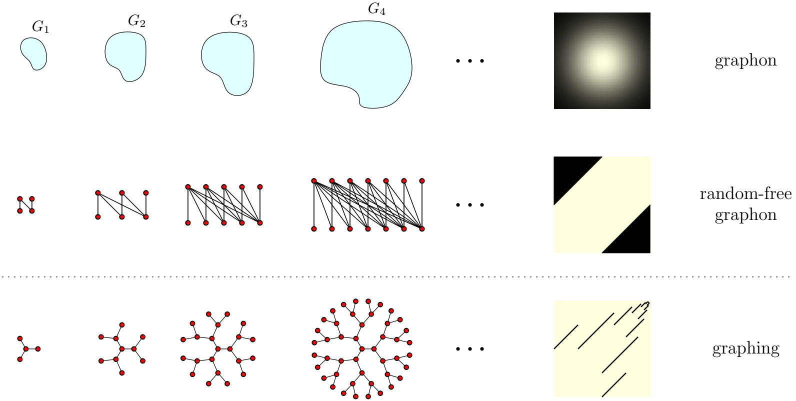
$$\frac{|\{v, B_\rho(G_n, v) \simeq (F, r)\}|}{|G_n|}.$$

- Distributional limit as a unimodular distribution on rooted connected countable graphs (Benjamini–Schramm).
- Analytic limit as a graphing = Borel graph satisfying the Mass Transport Principle (MTP) (Aldous–Lyons, Elek)

$$\forall A, B \in \Sigma \qquad \int_A \mathrm{d}_B(x)\,\mathrm{d}\nu(x) = \int_B \mathrm{d}_A(x)\,\mathrm{d}\nu(x).$$

(Converse: Aldous–Lyons conjecture)

Limit objects and Sparsity



Conclusion

A triple dichotomy

Distributional	vs	Analytic
Dense	vs	Sparse
Kernel	vs	Borel