The scaling limit of critical random graphs

L. Addario–Berry [1]   **N. Broutin** [2]   C. Goldschmidt [3]

[1]Department of Mathematics and Statistics
McGill University

[2]Projet Algorithms
INRIA Paris-Rocquencourt

[3]Department of Statistics
University of Warwick

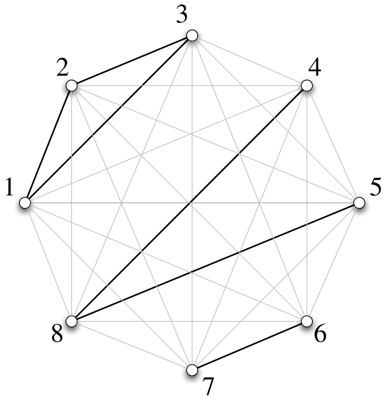
April 5, 2011

Plan

1. Random graphs
2. Exploration and branching processes
3. Large random trees
4. Comparing trees and Gromov–Hausdorff distance
5. Continuum random tree
6. Depth-first search and metric structure of large graphs

Erdős–Rényi random graphs

- n labelled vertices $\{1, 2, \ldots, n\}$
- $G_{n,p}$: each edge is present with probability $p \in [0, 1]$



Evolution of random graphs

$S_1^n \geq S_2^n \geq S_3^n \ldots$: sizes of connected components

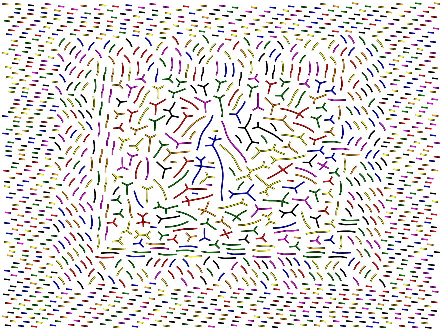
Different regimes for sizes as $n \to \infty$:

Evolution of random graphs

$S_1^n \geq S_2^n \geq S_3^n \ldots$: sizes of connected components

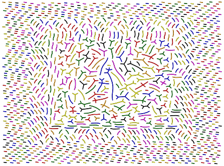
Different regimes for sizes as $n \to \infty$:

- $np = 1 - \epsilon$: $S_1^n = O(\log n)$



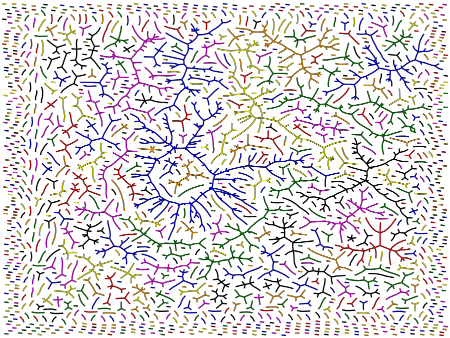
Evolution of random graphs

$S_1^n \geq S_2^n \geq S_3^n \ldots$: sizes of connected components



Different regimes for sizes as $n \to \infty$:

- $np = 1 - \epsilon$: $S_1^n = O(\log n)$
- $np = 1$: $\forall k \geq 1$ fixed $S_k^n = \Theta(n^{2/3})$

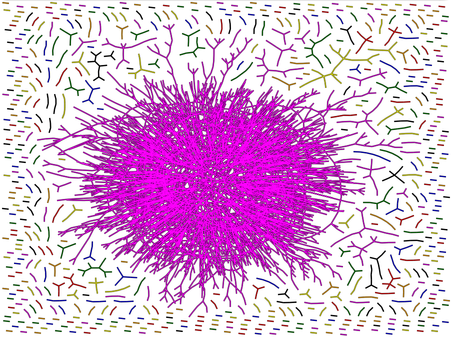


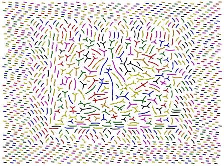
Evolution of random graphs

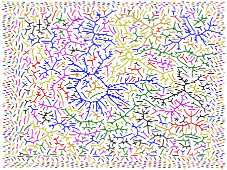
$S_1^n \geq S_2^n \geq S_3^n \ldots$: sizes of connected components

Different regimes for sizes as $n \to \infty$:

- $np = 1 - \epsilon$: $S_1^n = O(\log n)$
- $np = 1$: $\forall k \geq 1$ fixed $S_k^n = \Theta(n^{2/3})$
- $np = 1 + \epsilon$: $S_1^n = \Theta(n)$, $S_2^n = O(\log n)$





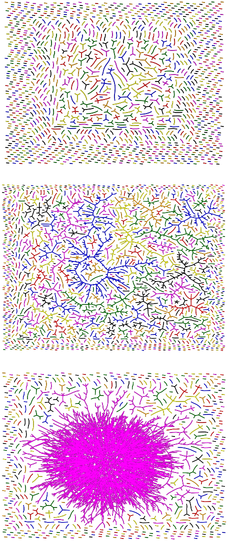


Evolution of random graphs

$S_1^n \geq S_2^n \geq S_3^n \ldots$: sizes of connected components

Different regimes for sizes as $n \to \infty$:

- $np = 1 - \epsilon$: $S_1^n = O(\log n)$
- $np = 1$: $\forall k \geq 1$ fixed $S_k^n = \Theta(n^{2/3})$
- $np = 1 + \epsilon$: $S_1^n = \Theta(n)$, $S_2^n = O(\log n)$



Component sizes and component structure

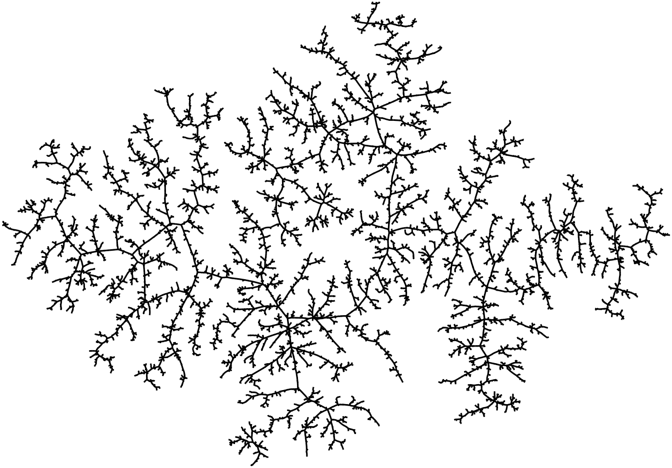
Important observation:

Given its size m, a connected component of $G(n,p)$ is distributed as $G(m,p)$ conditioned on being connected.

$\Rightarrow$ Study separately:

- sizes of connected components
- their structure given the size

What does a component look like?



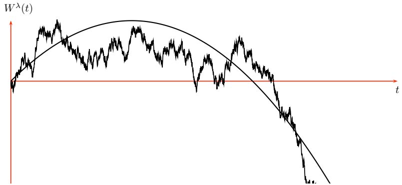
Inside the critical window: $np = 1 + \lambda n^{-1/3}$ for $\lambda \in \mathbb{R}$

Theorem (Aldous, 1997)

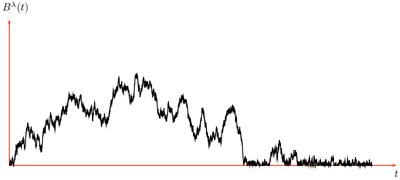
$(S_i^n; N_i^n)$ the size and surplus of the i-th largest component of $G_{n,p}$

$$((n^{-2/3} S_i^n; N_i^n) : i \geq 1) \xrightarrow[n\to\infty]{d} ((S_i; N_i) : i \geq 1)$$

as a sequence in $\ell^2_{\searrow} = \{x = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i\geq 1} x_i^2 < \infty\}$.



$W^\lambda(t) = W(t) + t\lambda - \frac{t^2}{2}$

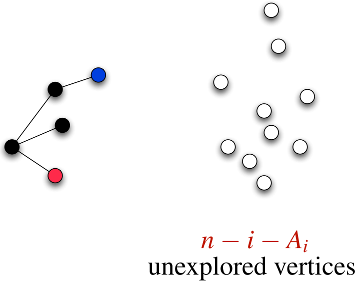


$B^\lambda(t) = W^\lambda(t) - \inf_{0<s<t} W^\lambda(s)$

Graph exploration and a branching process/random walk

Exposure vertices one after another:

- Initialization:
  one active vertex $A_0 = 1$
- At each time step:
  $A_{i+1} = A_i - 1 + Bin(n - i - A_i, p)$



The walk A_i and sizes of components

For $k \geq 0$, set $\xi_k := \inf\{i : A_i = 1 - k\}$
$\xi_k - \xi_{k-1}$ size of the k-th explored component

Asymptotics for A_i and phases in random graphs

Random walk with step distribution

$$A_{i+1} - A_i = Bin(n - i - A_i, p) - 1$$

The regimes when $np = c$

$\mathbf{E}\left[A_{i+1} - A_i\right] = \mathbf{E}Bin(n-i-A_i, p) - 1 \approx c - 1$

- $c < 1$: $A_i \to -\infty$
- $c = 1$: approx recurrent
- $c > 1$: $A_i \to \infty$

Branching process point of view

New vertices: $\mathbf{E}Bin(n - i - A_i, p) \approx c$

- $c < 1$: subcritical
- $c = 1$: critical
- $c > 1$: supercritical

The critical window

Parametrization of the probability for the critical window

$p = 1/n + \lambda n^{-4/3}$ with $\lambda \in \mathbb{R}$

$$A_{i+1} - A_i = Bin\left(n - i - A_i, \frac{1}{n} + \lambda n^{-4/3}\right) - 1$$

So for $i \sim tn^{2/3}$,

$$\mathbf{E}\left[A_{i+1} - A_i\right] \sim (\lambda - t)n^{-1/3} \quad \text{and} \quad \mathbf{Var}\left[A_{i+1} - A_i\right] \sim 1$$

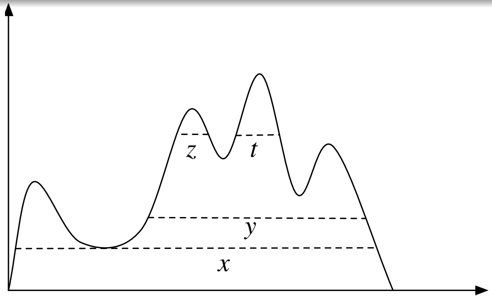
Central Limit Theorem for martingales:

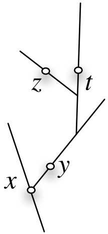
$$\left(\frac{A_{tn^{2/3}}}{n^{1/3}}\right)_{t\geq 0} \to \left(\lambda t - \frac{t^2}{2} + W(t)\right)_{t\geq 0}$$

Strategy to describe random graphs

1. Decompose into connected components
2. Extract a tree from each connected component
3. Describe the trees
4. Describe how to put back surplus edges

Real trees





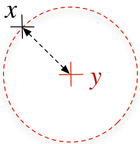
The tree in a continuous excursion

- continuous excursion $f(0) = f(1) = 0, f(s) > 0, s \in (0,1)$
- metric $d_f(x,y) = f(x) + f(y) - 2\inf_{s\in[x,y]} f(s)$.
- $x \sim_f y$ iff $d_f(x,y) = 0$
- the metric space $([0,1]/\sim_f, d_f)$ has a tree structure

Comparing subsets of a metric space: Hausdorff distance

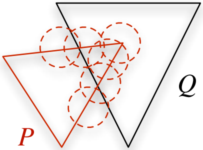
For two points $x, y \in (M, d)$

$$d(x,y) = \inf\{\epsilon > 0 : y \in B(x,\epsilon) \text{ and } x \in B(y,\epsilon)\}$$



For two subsets P, Q of a metric space (M, d): Hausdorff distance

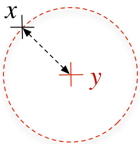
$$d_H(P,Q) = \inf\left\{\epsilon > 0 : P \subseteq \bigcup_{x \in Q} B(x,\epsilon) \text{ and } Q \subseteq \bigcup_{y \in P} B(y,\epsilon)\right\}.$$



Comparing subsets of a metric space: Hausdorff distance

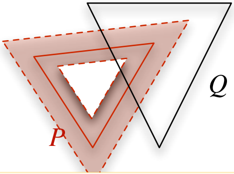
For two points $x, y \in (M, d)$

$$d(x,y) = \inf\{\epsilon > 0 : y \in B(x,\epsilon) \text{ and } x \in B(y,\epsilon)\}$$



For two subsets P, Q of a metric space (M, d): Hausdorff distance

$$d_H(P,Q) = \inf\left\{\epsilon > 0 : P \subseteq \bigcup_{x \in Q} B(x,\epsilon) \text{ and } Q \subseteq \bigcup_{y \in P} B(y,\epsilon)\right\}.$$



Comparing metric spaces: Gromov–Hausdorff distance

Using isometries

Embed both spaces isometrically into the same bigger space M:

$$d_{GH}((M_1, d_1); (M_2, d_2)) = \inf d_H(M_1, M_2),$$

where the infimum is over all metric spaces M containing both (M_1, d_1) and (M_2, d_2).

Using distortions of correspondences

Map each space inside the other using a *correspondence* $\mathcal{R} \subseteq M_1 \times M_2$

$$\mathrm{dist}(\mathcal{R}) = \sup_{(x_1,x_2),(y_1,y_2)\in\mathcal{R}} \{|d_1(x_1, y_1) - d_2(x_2, y_2)|\}$$

$$d_{GH}((M_1, d_1); (M_2, d_2)) = \frac{1}{2} \inf_{\mathcal{R}} \mathrm{dist}(\mathcal{R})$$

The Brownian continuum random tree (CRT)

Let $\mathcal{T}(2e)$ be the real tree encoded by twice a standard Brownian excursion e

Theorem (Aldous)

Let T_n be a Cayley tree of size n, seen as a metric space with the graph distance. Then,

$$n^{-1/2} T_n \xrightarrow[n\to\infty]{d} \mathcal{T}(2e)$$

with the Gromov–Hausdorff distance.

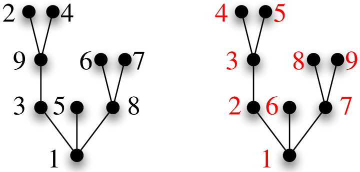
Some remarks:

- extends to all Galton–Watson trees with finite variance progeny
- the trees in random maps
- if infinite variance: Lévy trees, stable trees.

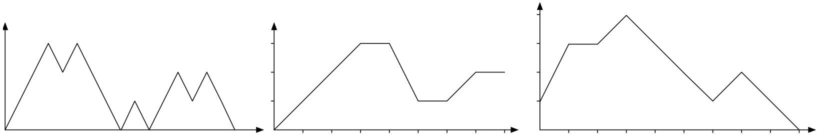
Representation of trees

A canonical order for the nodes:

- sort children by increasing label
- Depth-first order



Three different encodings of trees as nonnegative paths:



Contour process C Height process H Depth-first walk X

Walks associated with large random trees

Aldous, Le Gall, Marckert–Mokkadem

Let T_n be a Cayley tree of size n.

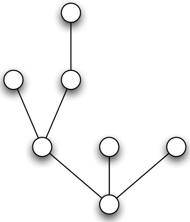
Theorem (Marckert–Mokkadem)

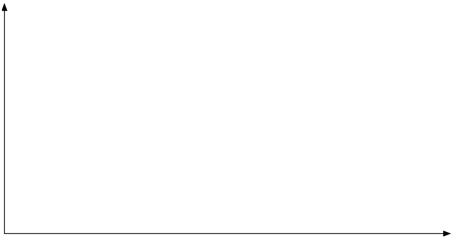
Let $e = (e(t), 0 \leq t \leq 1)$ be a standard Brownian excursion. Then,

$$\left(\frac{X(\lfloor n \cdot \rfloor)}{\sqrt{n}}; \frac{H(\lfloor n \cdot \rfloor)}{\sqrt{n}}; \frac{C(\lfloor 2n \cdot \rfloor)}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{d} (e(\,\cdot\,); 2e(\,\cdot\,); 2e(\,\cdot\,))$$

Understanding Depth-first search in graphs

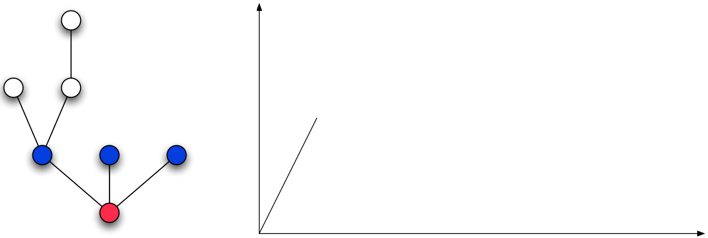
$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1





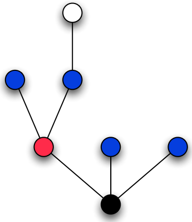
Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



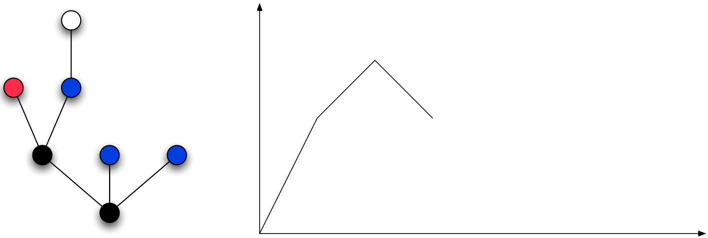
Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



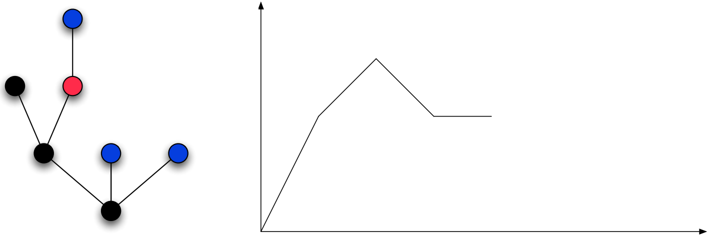
Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



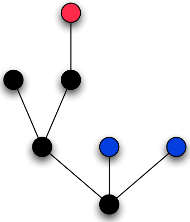
Understanding Depth-first search in graphs

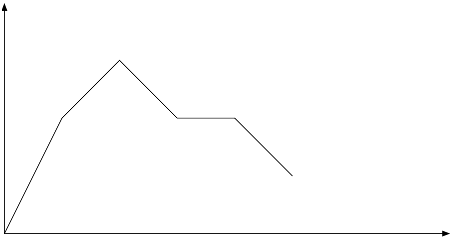
$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



Understanding Depth-first search in graphs

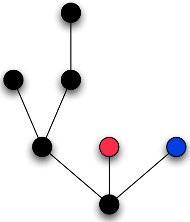
$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1

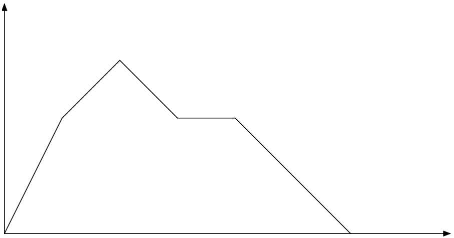




Understanding Depth-first search in graphs

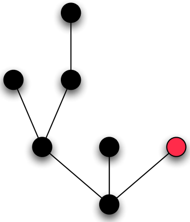
$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1

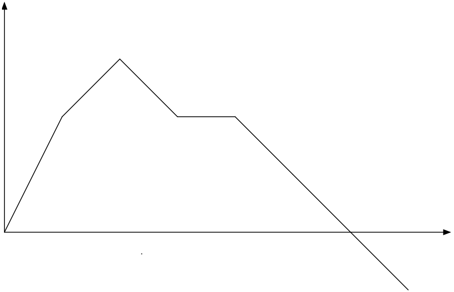




Understanding Depth-first search in graphs

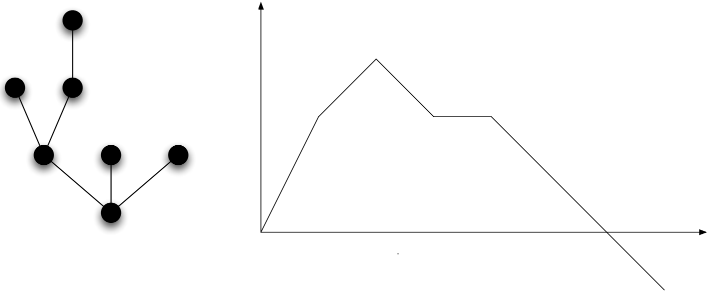
$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1





Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1

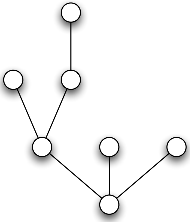


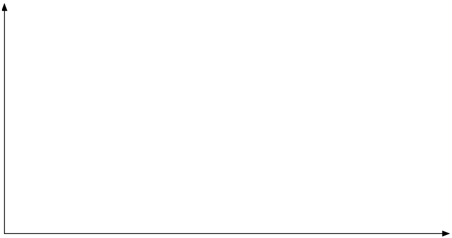
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



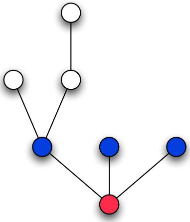


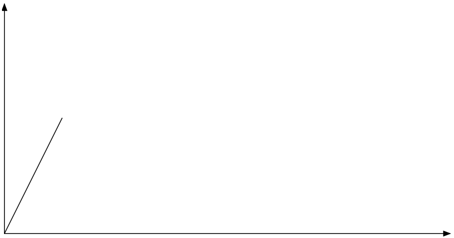
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



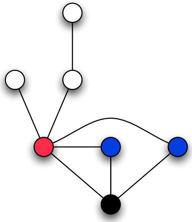


Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1

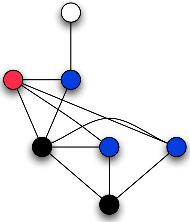


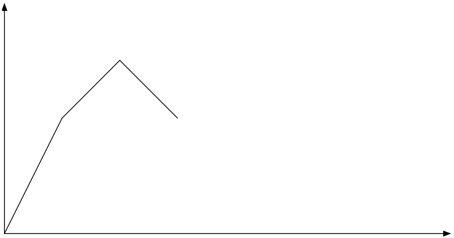
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



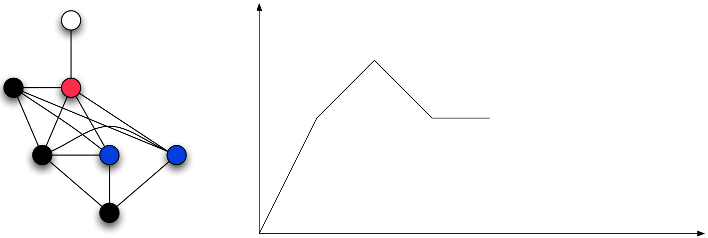


Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1

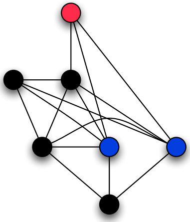


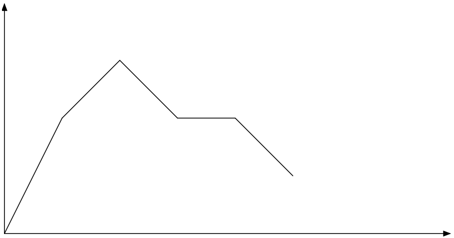
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



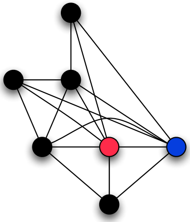


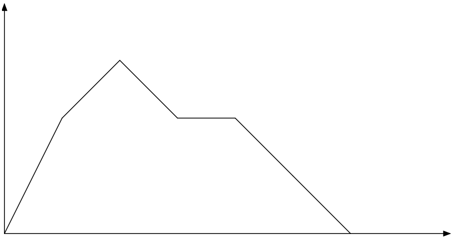
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



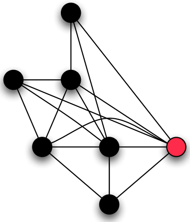


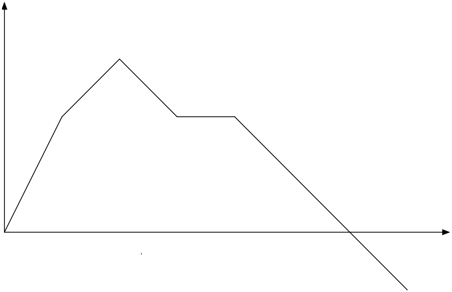
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1



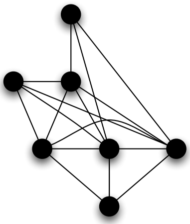


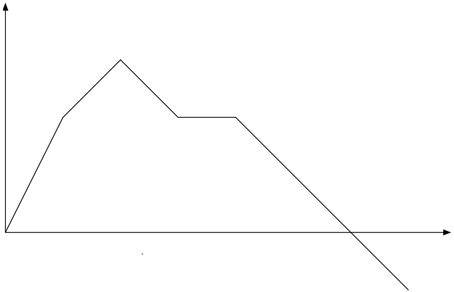
Question:

what can we change without changing the tree we obtain?

Understanding Depth-first search in graphs

$X(0) = 0$ and $X(i+1) - X(i) = \#$children of i minus 1





Question:

what can we change without changing the tree we obtain?

Answer:

Can add any of the $a(T)$ edges between two vertices active at the same time

Canonical tree and sampling random graphs

Partition the all the graphs G according to their canonical tree $T(G)$

$$\mathbb{G}_n = \bigcup_{T \in \mathbb{T}_n} \{G : T(G) = T\}$$

- Each graph G has a unique canonical tree $T(G)$
- Each canoninal tree T yields $2^{a(T)}$ graphs: $\#\{G : T(G) = T\} = 2^{a(T)}$

Uniform connected component of $G_{n,p}$ with m vertices

- Pick $\tilde{T}_m = T$ with probability $\propto (1-p)^{-a(T)}$
- Add each allowed edge with probability p.

Uniform connected connected graph with m vertices and $m-1+s$ edges

- Pick $\tilde{T}_m = T$ with probability $\propto \binom{a(T)}{s}$
- Add s random allowed edges.

Limit of canonical trees

In the critical regime: $p \sim 1/n$ and $m \sim n^{2/3}$:

$$(1-p)^{-a(T_m)} \sim \exp(m^{-3/2} a(T_m)) \sim \exp(\int_0^1 e(s)ds)$$

Definition (Tilted excursion)

Let e be a standard Brownian excursion:

$$\mathbb{P}(\tilde{e} \in \mathcal{B}) = \frac{\mathbf{E}\left[\mathbf{1}[e \in \mathcal{B}] \exp(\int_0^1 e(s)ds)\right]}{\mathbf{E}\left[\exp(\int_0^1 e(s)ds)\right]}.$$

Let $\tilde{T}_m$ be picked such that $\mathbb{P}\left(\tilde{T}_m = T\right) \propto (1-m^{-3/2})^{-a(T)}$

Theorem

$$\left(\frac{\tilde{X}^n(\lfloor n\cdot\rfloor)}{\sqrt{n}}; \frac{\tilde{C}^n(\lfloor 2n\cdot\rfloor)}{\sqrt{n}}; \frac{\tilde{H}^n(\lfloor n\cdot\rfloor)}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{d} (\tilde{e}(\cdot); 2\tilde{e}(\cdot); 2\tilde{e}(\cdot)).$$

The limit of tilted trees

Let $\mathcal{T}(\tilde{e}^{(\sigma)})$ the real tree encoded by a tilted excursion $\tilde{e}^{(\sigma)}$ of length σ.

Theorem

Let $p \sim 1/n$ and $m \sim \sigma n^{2/3}$, $\sigma > 0$. Let G_m^p be a uniform connected component of $G_{n,p}$ on m vertices.

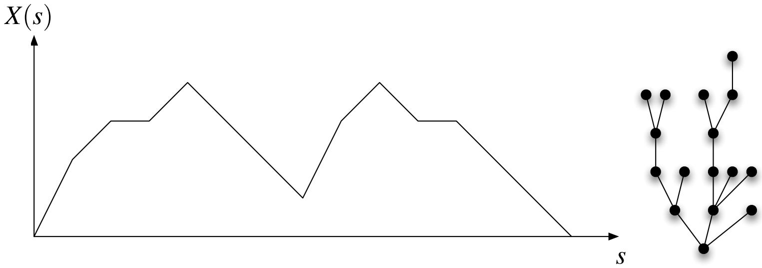
$$\frac{T(G_m^p)}{n^{1/3}} \xrightarrow[n\to\infty]{d} \mathcal{T}(2\tilde{e}^{(\sigma)}),$$

with the Gromov–Hausdorff distance.

Connected graphs as marked excursions

Bijection between:

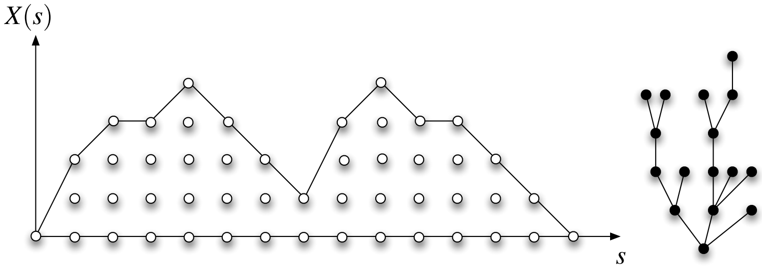
- Excursion of length m with some points under X: $(X, \mathcal{P})$.
- Connected graphs on m vertices $G^X(X, \mathcal{P})$



Connected graphs as marked excursions

Bijection between:

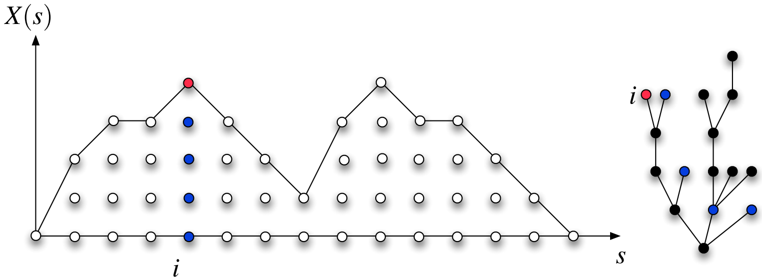
- Excursion of length m with some points under X: $(X, \mathcal{P})$.
- Connected graphs on m vertices $G^X(X, \mathcal{P})$



Connected graphs as marked excursions

Bijection between:

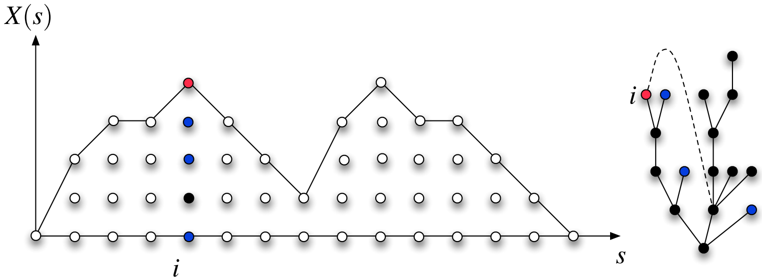
- Excursion of length m with some points under X: $(X, \mathcal{P})$.
- Connected graphs on m vertices $G^X(X, \mathcal{P})$



Connected graphs as marked excursions

Bijection between:

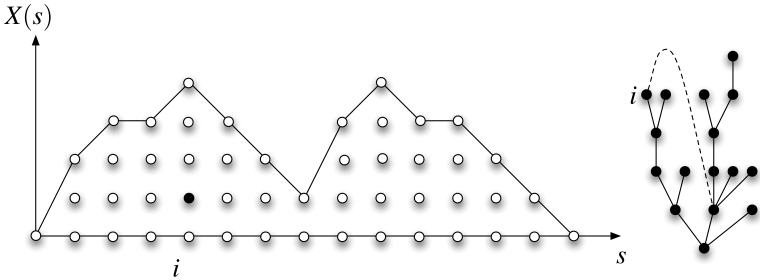
- Excursion of length m with some points under X: $(X, \mathcal{P})$.
- Connected graphs on m vertices $G^X(X, \mathcal{P})$



Connected graphs as marked excursions

Bijection between:

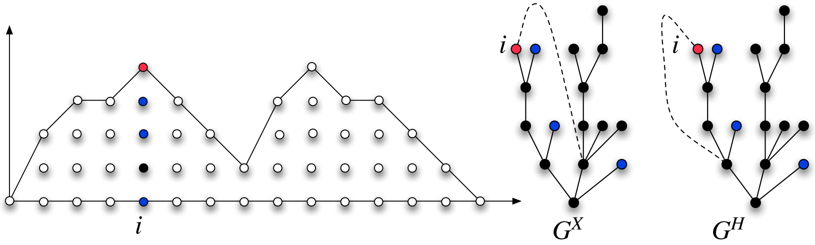
- Excursion of length m with some points under X: $(X, \mathcal{P})$.
- Connected graphs on m vertices $G^X(X, \mathcal{P})$



The limit of connected components: convergence

Need both X and H

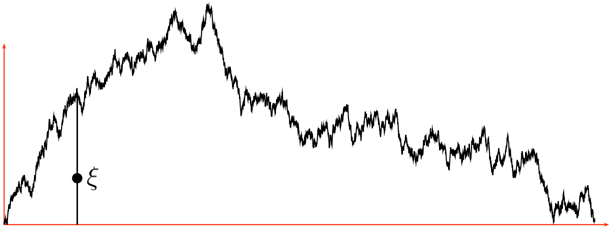
- X provides a bijection
- H provides the metric structure



Limit using the "Depth-First construction" G^X

$m^{-1/2}\tilde{X}^m(\lfloor m\cdot\rfloor) \to \tilde{e}$ and marks converge to a Poisson point process $\mathcal{P}$

Limit of connected components: construction



Metric characterization using the "height construction" G^H

$\mathcal{T}(2\tilde{e})$ where each point (ξ_x, ξ_y) of $\mathcal{P}$ identifies the leaf ξ_x with the point at distance ξ_y from the root on the path $[\![0, \xi_x]\!]$.

The limit of critical random graphs

M_i^n the i-th largest connected component of $G_{n,p}$ and S_i^n its size.
$\mathbf{M}^n = (M_1^n, M_2^n, \dots)$ as a sequence of metric spaces, with distance

$$d(\mathbf{A}, \mathbf{B}) = \left(\sum_{i \geq 1} d_{GH}(A_i, B_i)^4 \right)^{1/4}$$

$n^{-2/3}\mathbf{S}^n = (n^{-2/3} S_1^n, \dots)$ in
$\ell^2_{\searrow} = \{(x_1, x_2, \dots) : x_1 \geq 0, \dots, \sum_{i \geq 1} x_i^2 < \infty\}$.

Theorem

$$(n^{-1/3}\mathbf{M}^n, n^{-2/3}\mathbf{S}^n) \xrightarrow[n\to\infty]{d} (\mathbf{M}, \mathbf{S}) \qquad \text{where}$$

- $\mathbf{S}$ is the ordered sequence of excursion lengths of B^λ
- Given $S = (S_1, S_2, \dots)$, $(M_1, M_2, \dots)$ are independent $g(\tilde{e}^{(S_i)}, \mathcal{P}_i)$.