High Order Phase Transition in Random Hypergraphs

Linyuan Lu

University of South Carolina

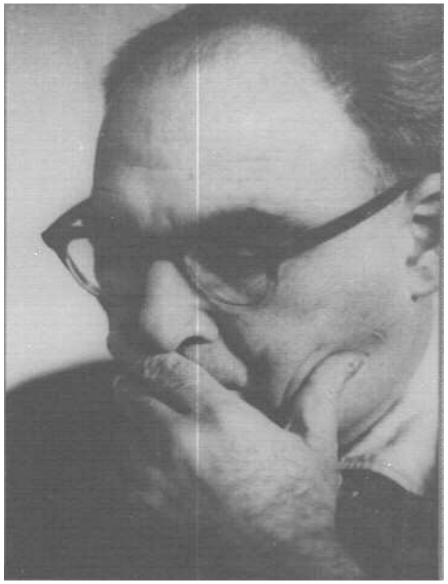
Coauthor: Xing Peng (UCSD)

Workshop on Probability and Graphs

EURANDOM, Eindhoven, Netherlands, January 6-10, 2014.

Phase transition of random graphs

Paul Erdős



Alfréd Rényi

Paul Erdős and A. Rényi, On the evolution of random graphs *Magyar Tud. Akad. Mat. Kut. Int. Kozl.* **5** (1960) 17-61.

Phase transition of random graphs

ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDÖS and A. RÉNYI
Institute of Mathematics
Hungarian Academy of Sciences, Hungary

1. Definition of a random graph

Let $E_{n,N}$ denote the set of all graphs having n given labelled vertices $V_1, V_2, \cdots, V_n$ and N edges. The graphs considered are supposed to be not oriented, without parallel edges and without slings (such graphs are sometimes called linear graphs). Thus a graph belonging to the set $E_{n,N}$ is obtained by choosing N out of the possible $\binom{n}{2}$ edges between the points $V_1, V_2, \cdots, V_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{\binom{n}{2}}{N}$. A random graph $\Gamma_{n,N}$ can be defined as an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1/\binom{\binom{n}{2}}{N}$. There is however an other slightly

Classical random graphs models

- **Uniform model $E_{n,N}$:** the set of all graphs with n vertices and N edges, equipped with uniform distribution.

Classical random graphs models

- **Uniform model $E_{n,N}$:** the set of all graphs with n vertices and N edges, equipped with uniform distribution.
- **Binomial model** $G(n,p)$: assign each pair of vertices as an edge with probability p independently. (proposed by Gilbert, is also known as Erdős-Rényi model.)

Classical random graphs models

- **Uniform model $E_{n,N}$:** the set of all graphs with n vertices and N edges, equipped with uniform distribution.
- **Binomial model $G(n,p)$**: assign each pair of vertices as an edge with probability p independently. (proposed by Gilbert, is also known as Erdős-Rényi model.)

Two models are more or less "equivalent" via $p \approx \frac{2N}{n(n-1)}$.

Classical random graphs models

- **Uniform model $E_{n,N}$:** the set of all graphs with n vertices and N edges, equipped with uniform distribution.
- **Binomial model $G(n,p)$**: assign each pair of vertices as an edge with probability p independently. (proposed by Gilbert, is also known as Erdős-Rényi model.)

Two models are more or less "equivalent" via $p \approx \frac{2N}{n(n-1)}$.

Rich literature on the phase transition of these two models: **Erdős, Rényi, Bollobás, Spencer, Janson, Knuth, Łuczak, Ruciński, Pittel, Sudakov, Krivelevich, Wormald, ...**

Evolution of $G(n, p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.

Evolution of $G(n, p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.
- $p \sim 1/n + \epsilon/n^{4/3}$, the critical window, double jump.

Evolution of $G(n, p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.
- $p \sim 1/n + \epsilon/n^{4/3}$, the critical window, double jump.
- $p \sim c/n$ for $c > 1$: Except for one "giant" component, all the other components are relatively small. The giant component has approximately $f(c)n$ vertices, where

$$f(c) = 1 - \frac{1}{c}\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}(ce^{-c})^k.$$

General question

General question on phase transition: Given a random graph model, how does the distribution of connected components evolve as the number of edges increase?

General question

General question on phase transition: Given a random graph model, how does the distribution of connected components evolve as the number of edges increase?

This is a hard problem in general. The answer depends on the model.

Chung-Lu model

Chung-Lu model $G(w_1, w_2, \ldots, w_n)$:

- $w_1, w_2, \ldots, w_n$: vertices weight/expected degrees
- $p_{ij} = w_i w_j / \sum_{i=1}^{n} w_i$: the probability of the pair (i, j) being created as an edge, independently.

Chung-Lu model

Chung-Lu model $G(w_1, w_2, \ldots, w_n)$:

- $w_1, w_2, \ldots, w_n$: vertices weight/expected degrees
- $p_{ij} = w_i w_j / \sum_{i=1}^{n} w_i$: the probability of the pair (i, j) being created as an edge, independently.

Chung and Lu (2004)
If the average degree is strictly greater than 1, then almost surely the giant component in a graph G in $G(w_1, \ldots, w_n)$ has volume $(\lambda_0 + O(\sqrt{n}\frac{\log^{3.5} n}{\mathrm{vol}(G)}))\mathrm{vol}(G)$, where λ_0 is the unique positive root of the following equation:

$$\sum_{i=1}^{n} w_i e^{-w_i \lambda} = (1 - \lambda) \sum_{i=1}^{n} w_i.$$

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

- **Kesten [1980]:** $p_c(\mathbb{Z}^2) = \frac{1}{2}$.

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

- **Kesten [1980]:** $p_c(\mathbb{Z}^2) = \frac{1}{2}$.
- **Kesten [1990]:** $p_c(\mathbb{Z}^d) \sim \frac{1}{2d}$ as $d \to \infty$.

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

- **Kesten [1980]:** $p_c(\mathbb{Z}^2) = \frac{1}{2}$.
- **Kesten [1990]:** $p_c(\mathbb{Z}^d) \sim \frac{1}{2d}$ as $d \to \infty$.
- **Alon, Benjamini, Stacey [2004]:** For random d-regular graph, $p_c = \frac{1}{d-1} + o(1)$.

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

- **Kesten [1980]:** $p_c(\mathbb{Z}^2) = \frac{1}{2}$.
- **Kesten [1990]:** $p_c(\mathbb{Z}^d) \sim \frac{1}{2d}$ as $d \to \infty$.
- **Alon, Benjamini, Stacey [2004]:** For random d-regular graph, $p_c = \frac{1}{d-1} + o(1)$.
- **Bollobás, Borgs, Chayes, and Riordan [2008]:** For desnse graph G, $p_c \approx \frac{1}{\mu}$, where μ is the largest eigenvalue of the adjacency matrix.

Graph percolation

Percolation problem: Given a host graph G, find the the threshold $p_c(G)$ so that the random subgraph G_p has a giant component if $p > p_c$ and no giant component if $p < p_c$.

- **Kesten [1980]:** $p_c(\mathbb{Z}^2) = \frac{1}{2}$.
- **Kesten [1990]:** $p_c(\mathbb{Z}^d) \sim \frac{1}{2d}$ as $d \to \infty$.
- **Alon, Benjamini, Stacey [2004]:** For random d-regular graph, $p_c = \frac{1}{d-1} + o(1)$.
- **Bollobás, Borgs, Chayes, and Riordan [2008]:** For desnse graph G, $p_c \approx \frac{1}{\mu}$, where μ is the largest eigenvalue of the adjacency matrix.
- **Chung, Lu, Horn [2009]:** $p_c \approx \frac{1}{\mu}$, for sparse graphs under some spectral conditions.

Hypergraphs

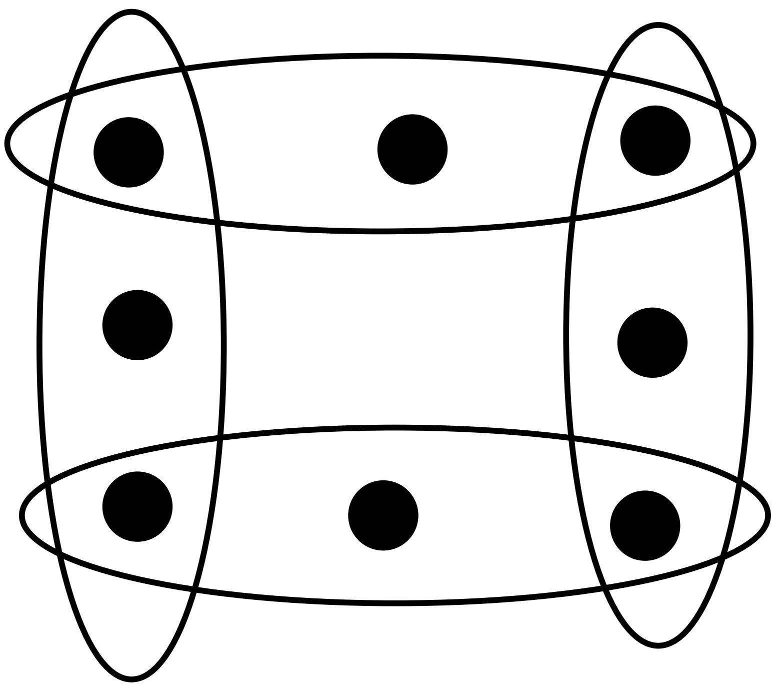
$H = (V, E)$ is an r-uniform hypergraph (r-graph, for short).

- V: the set of vertices
- E: the set of edges, each edge has carnality r.

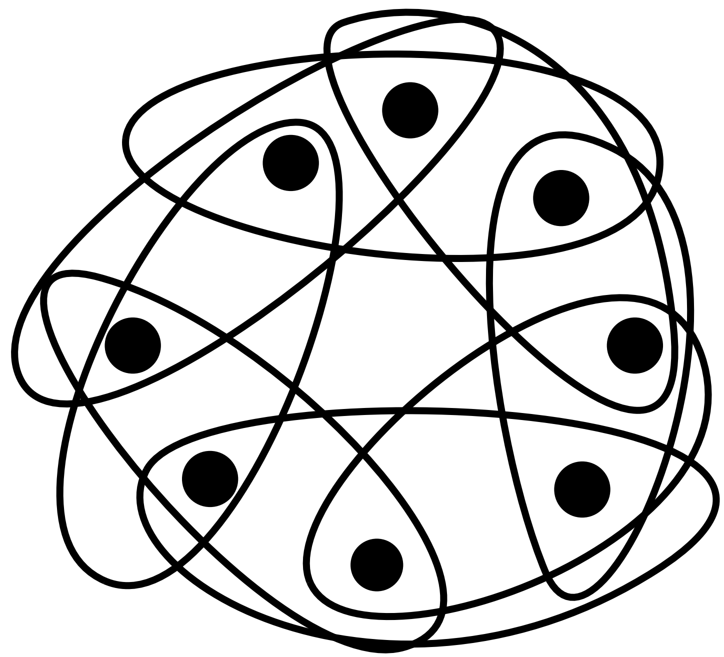
Hypergraphs

$H = (V, E)$ is an r-uniform hypergraph (r-graph, for short).

- V: the set of vertices
- E: the set of edges, each edge has carnality r.



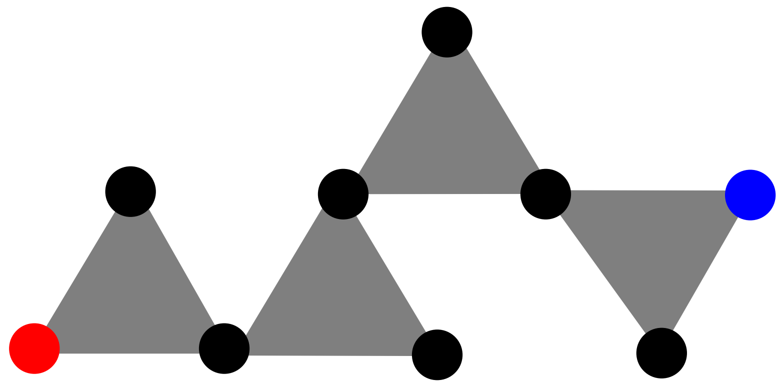
A 3-uniform loose cycle



A 3-uniform tight cycle

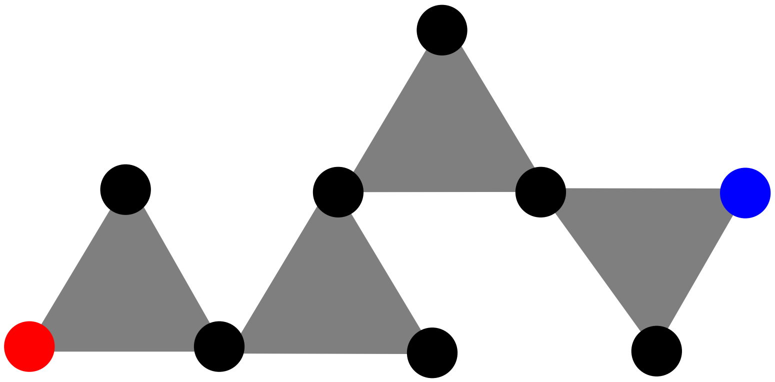
Connections in 3-graphs

- Vertex to Vertex

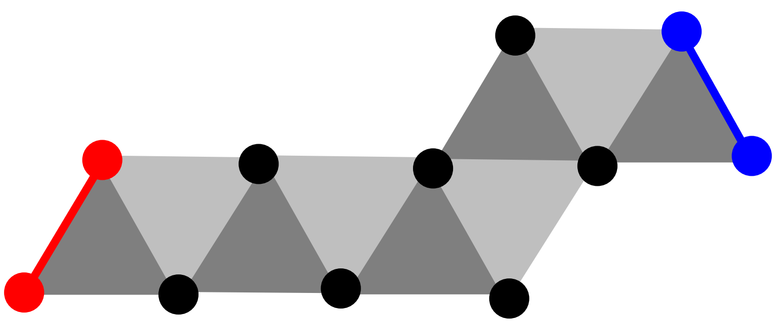


Connections in 3-graphs

- Vertex to Vertex



- Pair to Pair



High order components of H

Given an r-graph $H = (V, E)$, for $1 \leq s \leq r - 1$, define an auxiliary graph G^s:

- vertex set $\binom{V}{s}$, an s-set is called a stop.
- a pair (S, T) forms an edge in G^s if $S \cup T \subseteq F \in E$.

(There are different variations of G^s.)

High order components of H

Given an r-graph $H = (V, E)$, for $1 \le s \le r - 1$, define an auxiliary graph G^s:

- vertex set $\binom{V}{s}$, an s-set is called a stop.
- a pair (S, T) forms an edge in G^s if $S \cup T \subseteq F \in E$.

(There are different variations of G^s.)

The components of G^s are the called the s-th-order components. An s-th-order component is giant if its size is $\Theta(n^s)$.

High order components of H

Given an r-graph $H = (V, E)$, for $1 \leq s \leq r - 1$, define an auxiliary graph G^s:

- vertex set $\binom{V}{s}$, an s-set is called a stop.
- a pair (S, T) forms an edge in G^s if $S \cup T \subseteq F \in E$.

(There are different variations of G^s.)

The components of G^s are the called the s-th-order components. An s-th-order component is giant if its size is $\Theta(n^s)$.

Main question: how does the distribution of s-th order components evolve as the number of edges in random hypergraph increases?

Random hypergraphs

- **Uniform model $H^r_{n,N}$:** the set of all r-uniform hypergraphs with n vertices and N edges, equipped with uniform distribution.

Random hypergraphs

- **Uniform model $H^r_{n,N}$:** the set of all r-uniform hypergraphs with n vertices and N edges, equipped with uniform distribution.
- **Binomial model** $H_r(n,p)$: assign each r-set of vertices as an edge with probability p independently.

Random hypergraphs

- **Uniform model $H^r_{n,N}$:** the set of all r-uniform hypergraphs with n vertices and N edges, equipped with uniform distribution.
- **Binomial model** $H_r(n,p)$: assign each r-set of vertices as an edge with probability p independently.

Two models are more or less "equivalent" via $N = \binom{n}{r}p$. Here we use the binomial model.

Phase transition of s-components

$s = 1$ (vertex-to-vertex connection)

- Schmidt-Pruzan and Shamir [1985]
- Karoński and Łuczak [2002]
- Coja-Oghlan, Moore, and Sanwalani [2007]
- Kang, Behrisch and Coja-Oghlan[2010]

Phase transition of s-components

$s = 1$ (vertex-to-vertex connection)

- Schmidt-Pruzan and Shamir [1985]
- Karoński and Łuczak [2002]
- Coja-Oghlan, Moore, and Sanwalani [2007]
- Kang, Behrisch and Coja-Oghlan[2010]

$s \geq 2$

- This talk.
- Also independently by **Cooley-Kang-Person [2013+]**.

Our Results

Theorem I [Lu-Peng 2013+] For any $1 \leq s \leq r-1$, set $p = \frac{c}{\binom{n}{r-s}}$ and $m = \binom{r}{s} - 1$. Then the following statements hold for the s-th order components in $H_r(n,p)$.

- If $mc < 1 - \epsilon$, then almost surely all s-th-order connected components have size $O(\ln n)$.
- If $mc > 1 + \epsilon$, then almost surely there is a unique giant s-th-order connected component of size $(f(c) + o(1))\binom{n}{s}$, where $f(c)$ is the unique positive root of $1 - x = e^{c[(1-x)^m - 1]}$ and

$$f(c) = 1 - \sum_{k=0}^{\infty} \frac{(mk+1)^{k-1} c^k}{k! e^{(km+1)c}}.$$

Our Results

Theorem II [Lu-Peng 2013+] For any $1 \leq s \leq r-1$, set $p = \frac{c}{\binom{n}{r-s}}$ and $m = \binom{r}{s} - 1$. For any $k \geq 0$, almost surely the number of s-th-order components in $H_r(n,p)$ of exactly k-edges is $(a_k + o(1))\binom{n}{s}$, where

$$a_k = \frac{(mk+1)^{k-1}c^k}{k!e^{(km+1)c}}.$$

The majority of the small components are (r,s)-trees.

(r, s)-trees

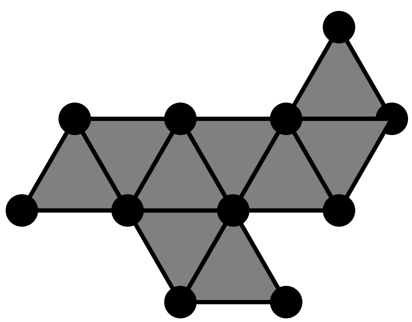
For any $1 \leq s \leq r-1$, an (r, s)-tree T_k is an s-th-order component of k edges with maximum number of vertices.

(r, s)-trees

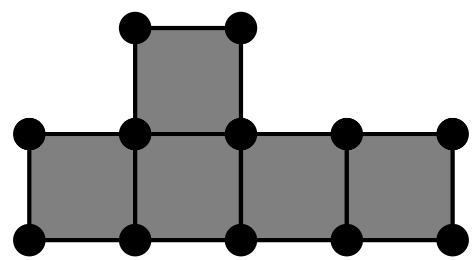
For any $1 \leq s \leq r-1$, an (r, s)-tree T_k is an s-th-order component of k edges with maximum number of vertices.

Recursive definition:

- T_0 is a single stop S_0.
- T_{k+1} can be obtained by first adding a set E_{k+1} of $r-s$ new vertices to T_k, selecting a stop $S \subset F \in E(T_k)$, then adding an new edge $S \cup E_{k+1}$.



a $(3, 2)$-tree



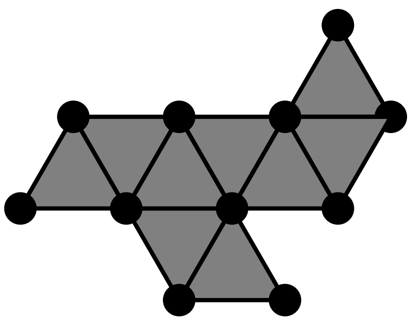
a $(4, 2)$-tree

(r, s)-trees

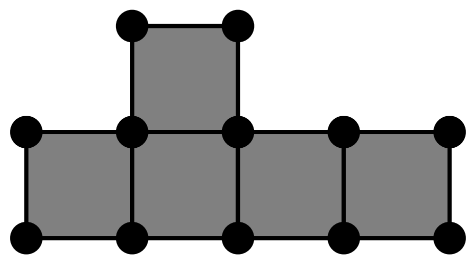
For any $1 \leq s \leq r-1$, an (r, s)-tree T_k is an s-th-order component of k edges with maximum number of vertices.

Recursive definition:

- T_0 is a single stop S_0.
- T_{k+1} can be obtained by first adding a set E_{k+1} of $r-s$ new vertices to T_k, selecting a stop $S \subset F \in E(T_k)$, then adding an new edge $S \cup E_{k+1}$.



a $(3, 2)$-tree



a $(4, 2)$-tree

Canonical partition: $S_0 \cup E_1 \cup E_2 \cup \cdots \cup E_k$

(Here $E_1, \ldots, E_k$ are indistinguishable.)

Counting (r, s)-trees

A $(2, 1)$-tree is just a tree rooted at vertex 1.

Counting (r, s)-trees

A $(2, 1)$-tree is just a tree rooted at vertex 1.

Cayley's formula [1889]: The number of trees on n vertices is n^{n-2}.

Counting (r, s)-trees

A $(2,1)$-tree is just a tree rooted at vertex 1.

Cayley's formula [1889]: The number of trees on n vertices is n^{n-2}.

Prüfer code [1918]: There is a bijection from the trees on n vertices to $[n]^{n-2}$.

Counting (r, s)-trees

A $(2, 1)$-tree is just a tree rooted at vertex 1.

Cayley's formula [1889]: The number of trees on n vertices is n^{n-2}.

Prüfer code [1918]: There is a bijection from the trees on n vertices to $[n]^{n-2}$.

Theorem [Lu-Peng 2013+]: For any $1 \leq s \leq r-1$, the number of (r, s)-trees with a fixed canonical partition $S_0 \cup E_1 \cup E_2 \cup \cdots \cup E_k$ is

$$\left(k\binom{r}{s} - k + 1\right)^{k-1}.$$

Counting (r, s)-trees

A $(2, 1)$-tree is just a tree rooted at vertex 1.

Cayley's formula [1889]: The number of trees on n vertices is n^{n-2}.

Prüfer code [1918]: There is a bijection from the trees on n vertices to $[n]^{n-2}$.

Theorem [Lu-Peng 2013+]: For any $1 \leq s \leq r - 1$, the number of (r, s)-trees with a fixed canonical partition $S_0 \cup E_1 \cup E_2 \cup \cdots \cup E_k$ is

$$\left(k\binom{r}{s} - k + 1\right)^{k-1}.$$

Remark: This theorem generalizes Cayley's formula. We also found the generalized Prüfer codes for (r, s)-trees.

Method

We couple the branching process with a special Galton-Watson process called "m-fold Poisson process" $T^{po}_{m,c}$:

- Start with one live node (the root).
- For each round, select a live node u if there is one.
- Roll a Poisson dice to produce an non-negative integer k with probability $e^{-c}c^k/k!$.
- Add mk children nodes to u, mark them live, and mark u dead.

Method

We couple the branching process with a special Galton-Watson process called "m-fold Poisson process" $T^{po}_{m,c}$:

- Start with one live node (the root).
- For each round, select a live node u if there is one.
- Roll a Poisson dice to produce an non-negative integer k with probability $e^{-c}c^k/k!$.
- Add mk children nodes to u, mark them live, and mark u dead.

Easy fact:

- If $mc < 1$, then the process terminates after finite steps.
- If $mc > 1$, then with probability $1 - x$, the process will survive forever. Here x is the solution of

$$x = e^{c(x^m - 1)}.$$

Obstacles

If $mc < 1$, coupling is easy.

Obstacles

If $mc < 1$, coupling is easy.

If $mc > 1$, there are two major obstacles:

- A new edge revealed brings in less than m new stops.
- A r-set containing a live stop may have been used before.

Obstacles

If $mc < 1$, coupling is easy.

If $mc > 1$, there are two major obstacles:

- A new edge revealed brings in less than m new stops.
- A r-set containing a live stop may have been used before.

Krivelevich-Sudakov [2012]: The Depth-First-Search (DFS) finds a path of length $\Theta(n)$ in $G(n,p)$ when $np > 1$.

Obstacles

If $mc < 1$, coupling is easy.

If $mc > 1$, there are two major obstacles:

- A new edge revealed brings in less than m new stops.
- A r-set containing a live stop may have been used before.

Krivelevich-Sudakov [2012]: The Depth-First-Search (DFS) finds a path of length $\Theta(n)$ in $G(n, p)$ when $np > 1$.

DFS helps. but we use different analysis.

Sketch

Our approaches

- We also use DFS.
- We book-keep dead stops, live stops, available edges, and potentially bad t-sets for each live stops.
- If the number of potentially bad t-sets and used edges exceeding $\epsilon\binom{n}{r-s}$ at any live stop, we restart the process.

With all efforts, we are able to find a path of length $\Omega(\frac{n^s}{log^{C_s}n})$. Then we use sprinkling to show the giant component exists and is unique.

Variations

There are different ways to define two s-sets are adjacent. For example, we define $G^{(s)}$:

- vertex set $\binom{V}{s}$.
- a pair (S, T) forms an edge in $G^{(s)}$ if $S \cup T \subseteq F \in E$ and $|S \cup T|$ as large as possible.

Variations

There are different ways to define two s-sets are adjacent. For example, we define $G^{(s)}$:

- vertex set $\binom{V}{s}$.
- a pair (S,T) forms an edge in $G^{(s)}$ if $S \cup T \subseteq F \in E$ and $|S \cup T|$ as large as possible.

The connected components in $G^{(s)}$ are the same as in G^s unless $r = 2s$. It is interesting that the threshold happens when the average degree in $G^{(s)}$ much smaller than 1.

Variations

There are different ways to define two s-sets are adjacent. For example, we define $G^{(s)}$:

- vertex set $\binom{V}{s}$.
- a pair (S, T) forms an edge in $G^{(s)}$ if $S \cup T \subseteq F \in E$ and $|S \cup T|$ as large as possible.

The connected components in $G^{(s)}$ are the same as in G^s unless $r = 2s$. It is interesting that the threshold happens when the average degree in $G^{(s)}$ much smaller than 1.
For $r = 2s$, the components in $G^{(s)}$ is similar to $G(n, p)$ and has an easier proof.

Related topics

For $1 \leq s \leq \frac{r}{2}$, we define the Laplacian eigenvalues of $G^{(s)}$ as the s-th-order Laplacian eigenvalues of the hypergraph H.

- These eigenvalues can effectively control the mixing rate of high-ordered random walks, the generalized distances/diameters, and the edge expansions.
- The Laplacian eigenvalues of $H_r(n,p)$ follows the Semi-circle Law.
- There are several directions on generalizing spectral graph theory to hypergraphs.

Related topics

For $1 \leq s \leq \frac{r}{2}$, we define the Laplacian eigenvalues of $G^{(s)}$ as the s-th-order Laplacian eigenvalues of the hypergraph H.

- These eigenvalues can effectively control the mixing rate of high-ordered random walks, the generalized distances/diameters, and the edge expansions.
- The Laplacian eigenvalues of $H_r(n, p)$ follows the Semi-circle Law.
- There are several directions on generalizing spectral graph theory to hypergraphs.

Thank you.