Poisson approximation of subgraphs in random intersection graphs

Dudley Stark

Queen Mary, University of London

Durham, 24 July 2013

- Joint work with Katarzyna Rybarczyk, Adam Mickiewicz University
- Thanks for the slides for the first part of the talk, Katarzyna!

- Joint work with Katarzyna Rybarczyk, Adam Mickiewicz University
- Thanks for the slides for the first part of the talk, Katarzyna!

Outline of this talk

1. Erdős–Rényi random graphs
2. Random intersection graphs
3. Equivalence and differences between Erdős-Rényi and random intersection graphs
4. Distribution of subgraph counts in Erdős-Rényi graphs
5. Clique covers
6. Subgraph counts in random intersection graphs

First articles

On random graphs I.

Dedicated to O. Varga, at the occasion of his 50th birthday.

By P. ERDŐS and A. RÉNYI (Budapest).

Let us consider a "random graph" $\Gamma_{n,N}$ having n possible (labelled) vertices and N edges; in other words, let us choose at random (with equal probabilities) one of the $\binom{\binom{n}{2}}{N}$ possible graphs which can be formed from the n (labelled) vertices $P_1, P_2, \ldots, P_n$ by selecting N edges from the $\binom{n}{2}$ possible edges $\widehat{P_iP_j}$ ($1 \leqq i < j \leqq n$). Thus the effective number of vertices of $\Gamma_{n,N}$ may be less than n, as some points P_i may be not connected in $\Gamma_{n,N}$ with any other point P_j; we shall call such points P_i *isolated* points. We consider the isolated points also as belonging to $\Gamma_{n,N}$. $\Gamma_{n,N}$ is called completely connected if it effectively contains all points $P_1, P_2, \ldots, P_n$ (i. e. if it has no isolated points) and is connected in the ordinary sense. In the present paper we consider asymptotic statistical properties of random graphs for $n \to +\infty$. We shall deal with the following questions:

1. What is the probability of $\Gamma_{n,N}$ being completely connected?
2. What is the probability that the greatest connected component (sub-graph) of $\Gamma_{n,N}$ should have effectively $n-k$ points? ($k=0, 1, \ldots$).
3. What is the probability that $\Gamma_{n,N}$ should consist of exactly $k+1$ connected components? ($k=0, 1, \ldots$).
4. If the edges of a graph with n vertices are chosen successively so that after each step every edge which has not yet been chosen has the same probability to be chosen as the next, and if we continue this process until the graph becomes completely connected, what is the probability that the number of necessary steps ν will be equal to a given number l?

As (partial) answers to the above questions we prove the following four theorems. In Theorems 1, 2, and 3 we use the notation

$$N_c = \left[\frac{1}{2} n \log n + c\,n\right] \qquad (1)$$

where c is an arbitrary fixed real number ($[x]$ denotes the integer part of x).

RANDOM GRAPHS

By E. N. Gilbert

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

1. Introduction. Let N points, numbered $1, 2, \cdots, N$, be given. There are $N(N-1)/2$ lines which can be drawn joining pairs of these points. Choosing a subset of these lines to draw, one obtains a graph; there are $2^{N(N-1)/2}$ possible graphs in total. Pick one of these graphs by the following random process. For all pairs of points make random choices, independent of each other, whether or not to join the points of the pair by a line. Let the common probability of joining be p. Equivalently, one may erase lines, with common probability $q = 1 - p$ from the complete graph.

In the random graph so constructed one says that *point i is connected to point j* if some of the lines of the graph form a path from i to j. If i is connected to j for every pair i, j, then the graph is said to be *connected*. The probability P_N that the graph is connected, and also the probability R_N that two specific points, say 1 and 2, are connected, will both be found.

As an application, imagine the N points to be N telephone central offices and suppose that each pair of offices has the same probability p that there is an idle direct line between them. Suppose further that a new call between two offices can be routed via other offices if necessary. Then R_N is the probability that there is some way of routing a new call from office 1 to office 2 and P_N is the probability that each office can call every other office.

Exact expressions for P_N and R_N are given in Section 2. These results are unwieldy for large N. Bounds on P_N and R_N derived in Section 3 show that

$$P_N \sim 1 - Nq^{N-1} \qquad (1)$$

and

$$R_N \sim 1 - 2q^{N-1} \qquad (2)$$

asymptotically as $N \to \infty$.

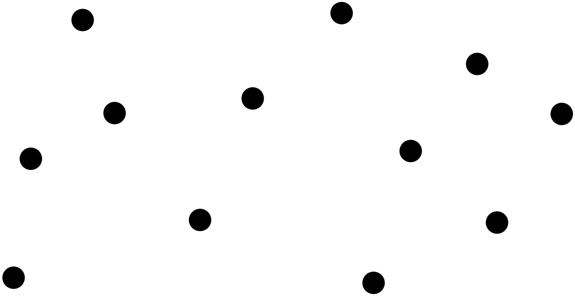
Other related results appear in a paper by Austin, Fagen, Penney, and Riordan [1]. These authors use a different random process to pick a graph and they find a generating function for the distribution of the number of connected pieces in the random graph.

2. Exact results. P_N may be expressed in terms of the number $C_{N,L}$ of connected graphs having N labeled points and L lines. Since each such graph has probability $p^L q^{-L+N(N-1)/2}$ of being the chosen graph, it follows that

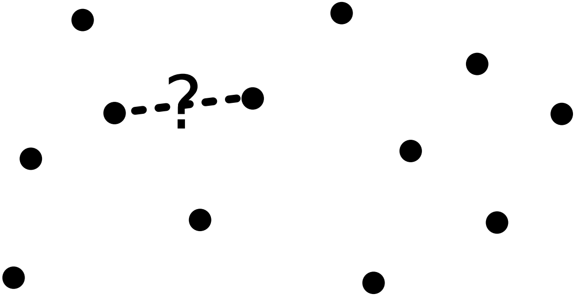
$$P_N = \sum_L C_{N,L}\, p^L q^{-L+N(N-1)/2}.$$

Received January 26, 1959; revised July 29, 1959.

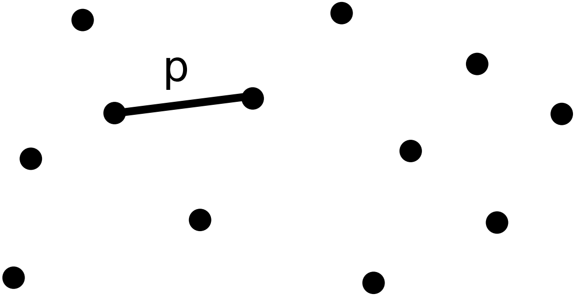
Erdős–Rényi random graph $G(n, p)$



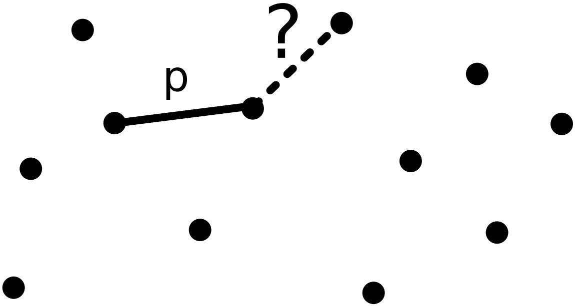
Erdős–Rényi random graph $G(n, p)$



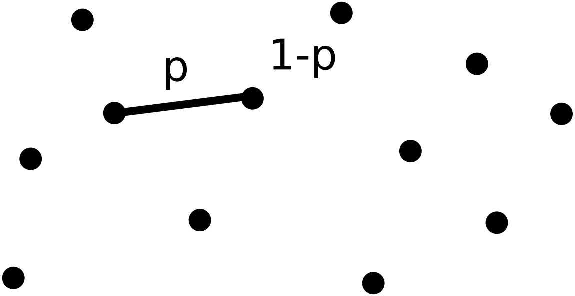
Erdős–Rényi random graph $G(n, p)$



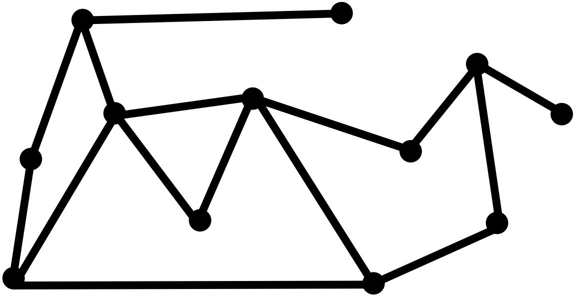
Erdős–Rényi random graph $G(n, p)$



Erdős–Rényi random graph $G(n, p)$



Erdős–Rényi random graph $G(n, p)$



Erdős–Rényi random graph $G(n, \hat{p})$

- Ω_n - a set of all graphs on n vertices ($|\Omega_n| = 2^{\binom{n}{2}}$);
- $\sigma = \mathcal{P}(\Omega_n)$;
- each edge appears independently with probability $\hat{p}$;
- A given graph $G \in \Omega_n$ on m vertices is picked with probability

$$\hat{p}^m(1-\hat{p})^{\binom{n}{2}-m}.$$

Erdős–Rényi random graph $G(n, \hat{p})$

- Ω_n - a set of all graphs on n vertices ($|\Omega_n| = 2^{\binom{n}{2}}$);
- $\sigma = \mathcal{P}(\Omega_n)$;
- each edge appears independently with probability $\hat{p}$;
- A given graph $G \in \Omega_n$ on m vertices is picked with probability

$$\hat{p}^m (1 - \hat{p})^{\binom{n}{2} - m}.$$

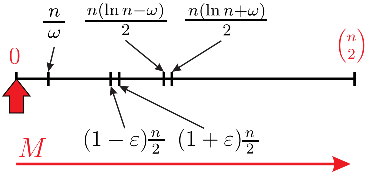
Evolution of a random graph

ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDŐS and A. RÉNYI

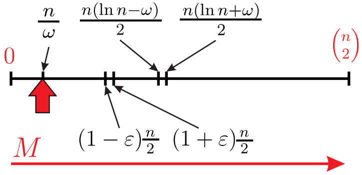
Dedicated to Professor P. Turán at his 50th birthday.



Evarage number of edges $M = \binom{n}{2} p$

As p is growing (M is growing) we expect a graph $G(n, p)$ to change. With probability $1 - o(1)$:

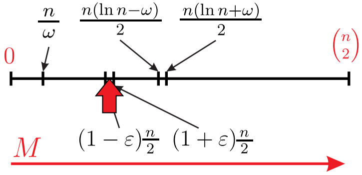
- is an empty graph;
- has no cycles (all components are trees);
- has all components: either trees or unicyclic;
- has exactly one component containing a constant fraction of vertices;
- is disconnected;
- is connected.



Evarage number of edges $M = \binom{n}{2}p$

As p is growing (M is growing) we expect a graph $G(n, p)$ to change. With probability $1 - o(1)$:

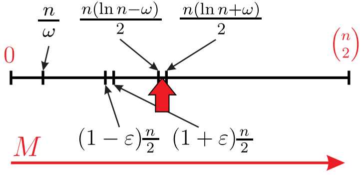
- is an empty graph;
- has no cycles (all components are trees);
- has all components: either trees or unicyclic;
- has exactly one component containing a constant fraction of vertices;
- is disconnected;
- is connected.



Evarage number of edges $M = \binom{n}{2} p$

As p is growing (M is growing) we expect a graph $G(n, p)$ to change. With probability $1 - o(1)$:

- is an empty graph;
- has no cycles (all components are trees);
- has all components: either trees or unicyclic;
- has exactly one component containing a constant fraction of vertices;
- is disconnected;
- is connected.

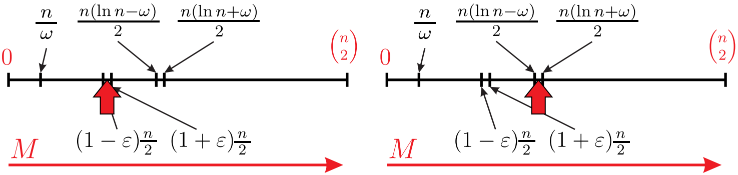


Evarage number of edges $M = \binom{n}{2}p$

As p is growing (M is growing) we expect a graph $G(n,p)$ to change. With probability $1 - o(1)$:

- is an empty graph;
- has no cycles (all components are trees);
- has all components: either trees or unicyclic;
- has exactly one component containing a constant fraction of vertices;
- is disconnected;
- is connected.

Threshold function



The study of the evolution of graphs leads to rather surprising results. For a number of fundamental structural properties A there exists a function $A(n)$ tending monotonically to $+\infty$ for $n \to +\infty$ such that

$$\text{(1)} \qquad \lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = \begin{cases} 0 & \text{if } \lim_{n\to+\infty} \frac{N(n)}{A(n)} = 0 \\ 1 & \text{if } \lim_{n\to+\infty} \frac{N(n)}{A(n)} = +\infty . \end{cases}$$

Threshold function in $G(n, \hat{p})$

A=$\{G(n,\hat{p})$ contains a $K_3\}$

Let $p_0 = \frac{1}{n}$, then

$$\lim_{n\to\infty} \Pr\{A\} = \begin{cases} 0 & \text{as } \frac{p}{p_0} \to 0; \\ 1 & \text{as } \frac{p}{p_0} \to \infty. \end{cases}$$

A=$\{G(n,\hat{p})$ is connected$\}$

Let $\hat{p}_0 = \frac{\ln n + \omega}{n}$, then

$$\lim_{n\to\infty} \Pr\{A\} = \begin{cases} 0 & \text{as } \omega \to -\infty; \\ 1 & \text{as } \omega \to \infty. \end{cases}$$

Every monotone property has a threshold function in $G(n, \hat{p})$.

w.h.p.

with high probability, i.e. with probability tending to 1 as $n \to \infty$.

Threshold function in $G(n, \hat{p})$

A=$\{G(n, \hat{p})$ contains a $K_3\}$

Let $p_0 = \frac{1}{n}$, then

$$\lim_{n\to\infty} \Pr\{A\} = \begin{cases} 0 & \text{as } \frac{p}{p_0} \to 0; \\ 1 & \text{as } \frac{p}{p_0} \to \infty. \end{cases}$$

A=$\{G(n, \hat{p})$ is connected$\}$

Let $\hat{p}_0 = \frac{\ln n + \omega}{n}$, then

$$\lim_{n\to\infty} \Pr\{A\} = \begin{cases} 0 & \text{as } \omega \to -\infty; \\ 1 & \text{as } \omega \to \infty. \end{cases}$$

Every monotone property has a threshold function in $G(n, \hat{p})$.

w.h.p.

with high probability, i.e. with probability tending to 1 as $n \to \infty$.

Motivation for further research

Why Erdős–Rényi model do not fit into many real life settings?

Example – complex network

- Relations between objects are not independent.
- Degree distribution is not Poisson.
- There are more cliques (hubs).
- There is some hidden relations which determine the structure of connections – it seems that none of the classical models capture them.

Motivation for further research

Why Erdős–Rényi model do not fit into many real life settings?

Example – complex network

- Relations between objects are not independent.
- Degree distribution is not Poisson.
- There are more cliques (hubs).
- There is some hidden relations which determine the structure of connections – it seems that none of the classical models capture them.

Motivation for further research

Why Erdős–Rényi model do not fit into many real life settings?

Example – complex network

- Relations between objects are not independent.
- Degree distribution is not Poisson.
- There are more cliques (hubs).
- There is some hidden relations which determine the structure of connections – it seems that none of the classical models capture them.

Motivation for further research

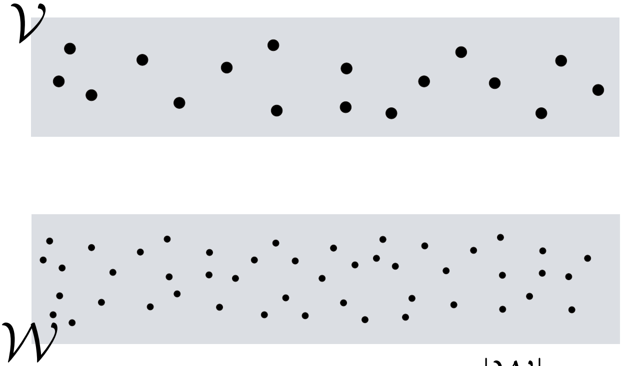
Why Erdős–Rényi model do not fit into many real life settings?

Example – complex network

- Relations between objects are not independent.
- Degree distribution is not Poisson.
- There are more cliques (hubs).
- There is some hidden relations which determine the structure of connections – it seems that none of the classical models capture them.

Random intersection graph - definition

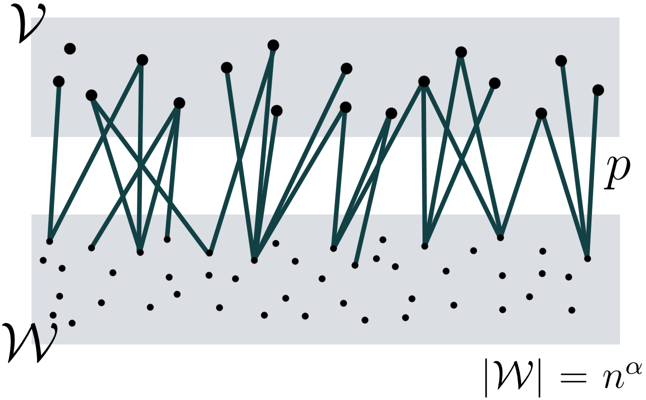
Karoński, Scheinerman, Singer-Cohen (1999)



$$|\mathcal{W}| = n^{\alpha}$$

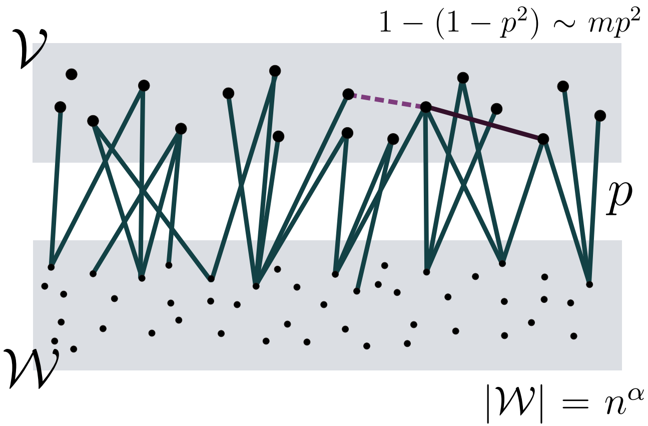
Random intersection graph - definition

Karoński, Scheinerman, Singer-Cohen (1999)



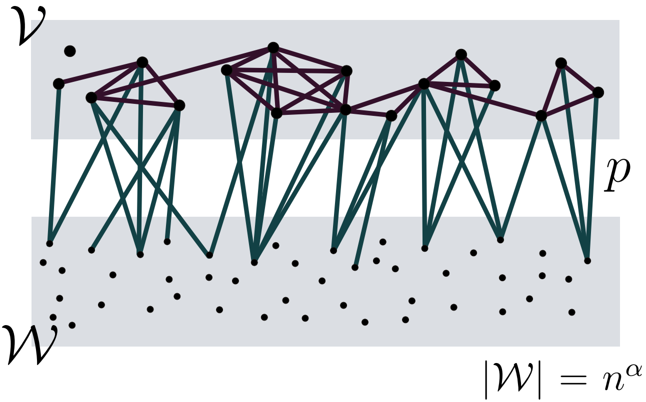
Random intersection graph - definition

Karoński, Scheinerman, Singer-Cohen (1999)



Random intersection graph - definition

Karoński, Scheinerman, Singer-Cohen (1999)



Random intersection graph - definition

$\mathcal{G}(n, m, p)$

- Set of vertices - $\mathcal{V}$, $|\mathcal{V}| = n$;
- Set of features - $\mathcal{W}$, $|\mathcal{W}| = m = n^{\alpha}$, α - constant;
- Random bipartite graph with bipartition $(\mathcal{V}, \mathcal{W})$

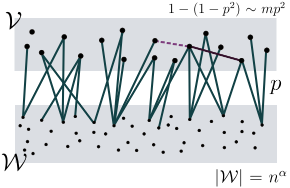
$$v_i v_j \in E(\mathcal{G}(n, m, p)) \Updownarrow W(v_i) \cap W(v_j) \neq \emptyset$$

edge probability in $\mathcal{G}(n, m, p)$

$$1 - (1 - p^2)^m \sim 1 - \exp(-mp^2)$$
$$\sim mp^2$$

(as $mp^2 \to 0$)

$$\hat{p} = 1 - \exp(-mp^2) \sim mp^2?$$

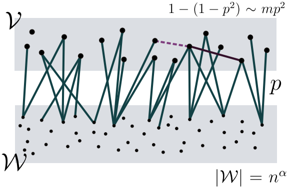


Comparison with $G(n, \hat{p})$

edge probability in $\mathcal{G}(n, m, p)$

$$1 - (1 - p^2)^m \sim 1 - \exp(-mp^2)$$
$$\sim mp^2$$

(as $mp^2 \to 0$)



$$\hat{p} = 1 - \exp(-mp^2) \sim mp^2?$$

J. Fill, E. Scheinerman, K. Singer-Cohen, 2000 – comparison

$m = n^\alpha$, where $\alpha > 0$ is a constant

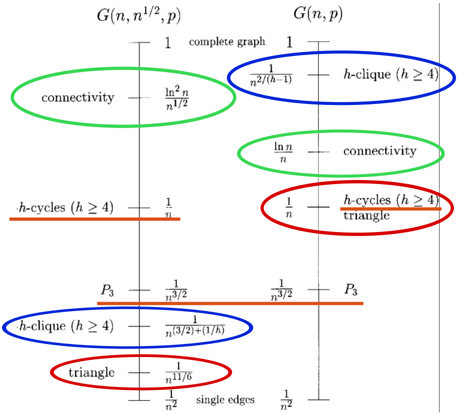


Fig. 2. A comparison of (asymptotic) edge probability at various thresholds for $G(n, n^{1/2}, p)$ and $G(n, p)$.

Threshold functions for K_3 appearance

$\mathcal{G}(n, m, p)$ (Karoński, Scheinerman, Singer-Cohen)

$$\tau = \begin{cases} \frac{1}{n\sqrt[3]{m}} & \text{for } \alpha \leq 3; \\ \frac{1}{\sqrt{nm}} & \text{for } \alpha \geq 3. \end{cases}$$

mp^2

$$m\tau^2 = \begin{cases} \frac{\sqrt[3]{m}}{n^2} & \text{for } \alpha \leq 3; \\ \frac{1}{n} & \text{for } \alpha \geq 3. \end{cases}$$

$G(n, \hat{p})$

$$\hat{\tau} = \frac{1}{n}$$

Question

Equivalence?

If $\mathcal{G}(n, m, p)$ and $G(n, \hat{p})$ for

$$\hat{p} \sim \Pr\{(v_1, v_2) \text{ is an edge in } \mathcal{G}(n, m, p)\}$$

are asymptoticly equivalent.
i.e. for any graph property $\mathcal{A}$ and any $a \in [0; 1]$

$$\Pr\{G(n, \hat{p}) \in \mathcal{A}\} \to a \quad \text{iff} \quad \Pr\{\mathcal{G}(n, m, p) \in \mathcal{A}\} \to a$$

Equivalence

Theorem (J. Fill, E. Scheinerman, K. Singer-Cohen, 2000)

Let $m = n^{\alpha}$ and $\alpha > 6$. Let $p = p(n)$ be such that

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2\ln n - \omega}{m}}$$

for some $\omega \to \infty$ (i.e. w.h.p $\mathcal{G}(n, m, p)$ is not edgeless and not complete). Let

$$\hat{p} = 1 - (1 - p_e)^m$$

($p_e \sim p^2$) and $\mathcal{A}$ be any graph property. Then for any $a \in [0; 1]$

$$\Pr\{\mathcal{G}(n, m, p) \in \mathcal{A}\} \to a \Leftrightarrow \Pr\{G(n, \hat{p}) \in \mathcal{A}\} \to a$$

Equivalence

Theorem (J. Fill, E. Scheinerman, K. Singer-Cohen, 2000)

Let $m = n^\alpha$ and $\alpha > 6$. Let $p = p(n)$ be such that

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2\ln n - \omega}{m}}$$

for some $\omega \to \infty$ (i.e. w.h.p $\mathcal{G}(n, m, p)$ is not edgeless and not complete). Let

$$\hat{p} = 1 - (1 - p_e)^m$$

$(p_e \sim p^2)$ and $\mathcal{A}$ be any graph property. Then for any $a \in [0; 1]$

$$\Pr\{\mathcal{G}(n, m, p) \in \mathcal{A}\} \to a \Leftrightarrow \Pr\{G(n, \hat{p}) \in \mathcal{A}\} \to a$$

Equivalence

Theorem (J. Fill, E. Scheinerman, K. Singer-Cohen, 2000)

Let $m = n^\alpha$ and $\alpha > 6$. Let $p = p(n)$ be such that

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2\ln n - \omega}{m}}$$

for some $\omega \to \infty$ (i.e. w.h.p $\mathcal{G}(n, m, p)$ is not edgeless and not complete). Let

$$\hat{p} = 1 - (1 - p_e)^m$$

($p_e \sim p^2$) and $\mathcal{A}$ be any graph property. Then for any $a \in [0; 1]$

$$\Pr\{\mathcal{G}(n, m, p) \in \mathcal{A}\} \to a \Leftrightarrow \Pr\{G(n, \hat{p}) \in \mathcal{A}\} \to a$$

Equivalence

Theorem (J. Fill, E. Scheinerman, K. Singer-Cohen, 2000)

Let $m = n^{\alpha}$ and $\alpha > 6$. Let $p = p(n)$ be such that

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2\ln n - \omega}{m}}$$

for some $\omega \to \infty$ (i.e. w.h.p $\mathcal{G}(n, m, p)$ is not edgeless and not complete). Let

$$\hat{p} = 1 - (1 - p_e)^m$$

($p_e \sim p^2$) and $\mathcal{A}$ be any graph property. Then for any $a \in [0; 1]$

$$\Pr\{\mathcal{G}(n, m, p) \in \mathcal{A}\} \to a \Leftrightarrow \Pr\{G(n, \hat{p}) \in \mathcal{A}\} \to a$$

Triangle counts

Given sequences a_n, b_n, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

In $G(n, \hat{p})$, what are the expected number of K_3's?

- Let X be the number of K_3's.
- $\mathbb{E}(X) = \binom{n}{3}\hat{p}^3 \sim \frac{1}{6}n^3\hat{p}^3$ and so ...
- When $\hat{p} \sim c/n$ for a constant $c > 0$, $\mathbb{E}(X) \sim c^3/6$.

Triangle counts

Given sequences a_n, b_n, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

In $G(n, \hat{p})$, what are the expected number of K_3's?

- Let X be the number of K_3's.
- $\mathbb{E}(X) = \binom{n}{3}\hat{p}^3 \sim \frac{1}{6}n^3\hat{p}^3$ and so ...
- When $\hat{p} \sim c/n$ for a constant $c > 0$, $\mathbb{E}(X) \sim c^3/6$.

Triangle counts

Given sequences a_n, b_n, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

In $G(n, \hat{p})$, what are the expected number of K_3's?

- Let X be the number of K_3's.
- $\mathbb{E}(X) = \binom{n}{3}\hat{p}^3 \sim \frac{1}{6}n^3\hat{p}^3$ and so ...
- When $\hat{p} \sim c/n$ for a constant $c > 0$, $\mathbb{E}(X) \sim c^3/6$.

Triangle counts

Given sequences a_n, b_n, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

In $G(n, \hat{p})$, what are the expected number of K_3's?

- Let X be the number of K_3's.
- $\mathbb{E}(X) = \binom{n}{3}\hat{p}^3 \sim \frac{1}{6}n^3\hat{p}^3$ and so ...
- When $\hat{p} \sim c/n$ for a constant $c > 0$, $\mathbb{E}(X) \sim c^3/6$.

Triangle counts

Given sequences a_n, b_n, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

In $G(n, \hat{p})$, what are the expected number of K_3's?

- Let X be the number of K_3's.
- $\mathbb{E}(X) = \binom{n}{3}\hat{p}^3 \sim \frac{1}{6}n^3\hat{p}^3$ and so ...
- When $\hat{p} \sim c/n$ for a constant $c > 0$, $\mathbb{E}(X) \sim c^3/6$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Triangle counts

If $\hat{p} \sim c/n$, then what is the limiting distribution of X?

- Let Γ index all possible triangles.
- Let $X = \sum_{\alpha \in \Gamma} I_\alpha$, where I_α is the indicator random variable that α is present in $G(n, \hat{p})$.
- If $\alpha, \beta \in \Gamma$ share no vertices, then I_α and I_β are independent.
- The expected number of pairs of triangles sharing vertices goes to 0.
- Therefore, $X \Rightarrow \mathrm{Po}(c^3/6)$.

Subgraph counts in $G(n, \hat{p})$

- let H_0 be a subgraph of K_n with e edges and v vertices.
- Let X be the number of H_0's.
- $\mathbb{E}(X) = \frac{(n)_v}{|\mathrm{Aut}(H_0)|}\hat{p}^e \sim \frac{1}{|\mathrm{Aut}(H_0)|}n^v\hat{p}^e$ and so ...
- When $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$, $\mathbb{E}(X) \sim \frac{c^e}{|\mathrm{Aut}(H_0)|}$.

Subgraph counts in $G(n, \hat{p})$

- let H_0 be a subgraph of K_n with e edges and v vertices.
- Let X be the number of H_0's.
- $\mathbb{E}(X) = \frac{(n)_v}{|\mathrm{Aut}(H_0)|}\hat{p}^e \sim \frac{1}{|\mathrm{Aut}(H_0)|}n^v\hat{p}^e$ and so ...
- When $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$, $\mathbb{E}(X) \sim \frac{c^e}{|\mathrm{Aut}(H_0)|}$.

Subgraph counts in $G(n, \hat{p})$

- let H_0 be a subgraph of K_n with e edges and v vertices.
- Let X be the number of H_0's.
- $\mathbb{E}(X) = \frac{(n)_v}{|\mathrm{Aut}(H_0)|}\hat{p}^e \sim \frac{1}{|\mathrm{Aut}(H_0)|} n^v \hat{p}^e$ and so ...
- When $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$, $\mathbb{E}(X) \sim \frac{c^e}{|\mathrm{Aut}(H_0)|}$.

Subgraph counts in $G(n, \hat{p})$

- let H_0 be a subgraph of K_n with e edges and v vertices.
- Let X be the number of H_0's.
- $\mathbb{E}(X) = \frac{(n)_v}{|\mathrm{Aut}(H_0)|}\hat{p}^e \sim \frac{1}{|\mathrm{Aut}(H_0)|} n^v \hat{p}^e$ and so ...
- When $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$, $\mathbb{E}(X) \sim \frac{c^e}{|\mathrm{Aut}(H_0)|}$.

Subgraph counts in $G(n, \hat{p})$

Bollobas, Rucinski,...

Definition of strictly balanced

A subgraph H_0 of K_n is strictly balanced if

$$\max_{\emptyset \subsetneq S \subsetneq V(H_0)} \frac{|E(S)|}{|S|} < \frac{e}{h},$$

where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

If $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ then, X has a limiting Poisson distribution if H_0 is strictly balanced.

Subgraph counts in $G(n, \hat{p})$

Bollobas, Rucinski,...

Definition of strictly balanced

A subgraph H_0 of K_n is strictly balanced if

$$\max_{\emptyset \subsetneq S \subsetneq V(H_0)} \frac{|E(S)|}{|S|} < \frac{e}{h},$$

where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

If $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ then, X has a limiting Poisson distribution if H_0 is strictly balanced.

Subgraph counts in $G(n, \hat{p})$

Definition of balanced

A subgraph H_0 of K_n is balanced if

$$\max_{\emptyset \subsetneq S \subsetneq V(H_0)} \frac{|E(S)|}{|S|} \leq \frac{e}{h},$$

where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

$\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ and H_0 is strictly balanced, but not strictly balanced, then $\lim_{n\to\infty} \mathrm{Var}(X) > \mathbb{E}(X)$ and X does not have limiting Poisson distribution.

Subgraph counts in $G(n, \hat{p})$

Definition of balanced

A subgraph H_0 of K_n is balanced if

$$\max_{\emptyset \subsetneq S \subsetneq V(H_0)} \frac{|E(S)|}{|S|} \leq \frac{e}{h},$$

where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

$\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ and H_0 is strictly balanced, but not strictly balanced, then $\lim_{n\to\infty} \mathrm{Var}(X) > \mathbb{E}(X)$ and X does not have limiting Poisson distribution.

Subgraph counts in $G(n, \hat{p})$

Definition of balanced

A subgraph H_0 of K_n is unbalanced if

$$\frac{|E(S)|}{|S|} > \frac{e}{h},$$

for some $\emptyset \subsetneq S \subsetneq V(H_0)$, where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

If $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ and H_0 is unbalanced, then $X \to 0$ in probability.

Subgraph counts in $G(n, \hat{p})$

Definition of balanced

A subgraph H_0 of K_n is unbalanced if

$$\frac{|E(S)|}{|S|} > \frac{e}{h},$$

for some $\emptyset \subsetneq S \subsetneq V(H_0)$, where $E(S)$ is the number of edges of H_0 contained in S.

Theorem

If $\hat{p} \sim cn^{-v/e}$ for a constant $c > 0$ and H_0 is unbalanced, then $X \to 0$ in probability.

Subgraph problem

Rybarzyk, S.(2010)

- $m = n^{\alpha}$, α – constant;
- X_n – number of K_h, $h \geq 3$, in $\mathcal{G}(n, m, p)$

$$p \sim \begin{cases} cn^{-1}m^{-\frac{1}{h}} & \text{for } 0 < \alpha < 2h/(h-1); \\ cn^{-\frac{h+1}{h-1}} & \text{for } \alpha = 2h/(h-1); \\ cn^{-\frac{1}{h-1}}m^{-\frac{1}{2}} & \text{for } \alpha > 2h/(h-1). \end{cases}$$

Subgraph problem

Rybarzyk, S.(2010)

- $m = n^{\alpha}$, α – constant;
- X_n – number of K_h, $h \geq 3$, in $\mathcal{G}(n, m, p)$

$$p \sim \begin{cases} cn^{-1}m^{-\frac{1}{h}} & \text{for } 0 < \alpha < 2h/(h-1); \\ cn^{-\frac{h+1}{h-1}} & \text{for } \alpha = 2h/(h-1); \\ cn^{-\frac{1}{h-1}}m^{-\frac{1}{2}} & \text{for } \alpha > 2h/(h-1). \end{cases}$$

Subgraph problem

Rybarzyk, S.(2010)

- $m = n^{\alpha}$, α – constant;
- X_n – number of K_h, $h \geq 3$, in $\mathcal{G}(n, m, p)$

$$p \sim \begin{cases} cn^{-1} m^{-\frac{1}{h}} & \text{for } 0 < \alpha < 2h/(h-1); \\ cn^{-\frac{h+1}{h-1}} & \text{for } \alpha = 2h/(h-1); \\ cn^{-\frac{1}{h-1}} m^{-\frac{1}{2}} & \text{for } \alpha > 2h/(h-1). \end{cases}$$

Subgraph problem

Rybarzyk, S.(2010)

(i) If $0 < \alpha < 2h/(h-1)$, then $\mathbb{E}X_n \sim c^h/h!$ and
$d_{TV}(X_n, \mathrm{Po}(\mathbb{E}X_n)) = O\left(n^{-\frac{\alpha}{h}}\right)$;

(ii) If $\alpha = 2h/(h-1)$ then $\mathbb{E}X_n \sim \left(c^h + c^{h(h-1)}\right)/h!$ and
$d_{TV}(X_n, \mathrm{Po}(\mathbb{E}X_n)) = O\left(n^{-\frac{2}{h-1}}\right)$;

(iii) If $\alpha > 2h/(h-1)$ then $\mathbb{E}X_n \sim c^{h(h-1)}/h!$ and
$d_{TV}(X_n, \mathrm{Po}(\mathbb{E}X_n)) = O\left(n^{\left(h-\frac{\alpha(h-1)}{2}-\frac{2}{h-1}\right)} + n^{-1}\right)$.

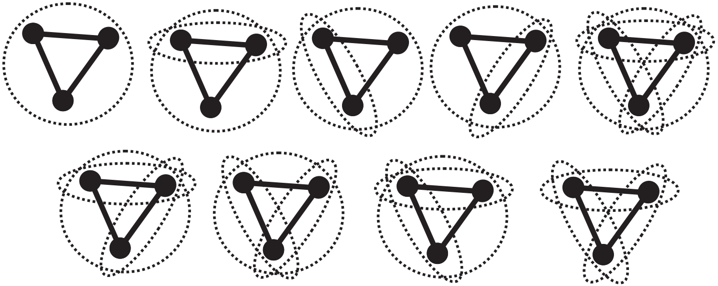
Clique covers

Definition of the clique covers of H_0

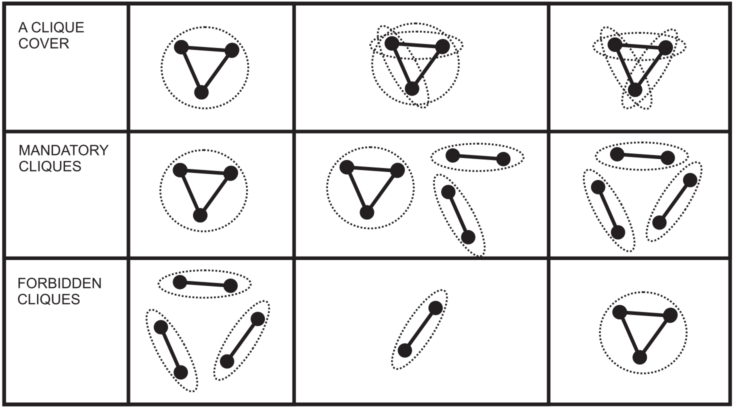
Given a fixed subgraph H_0 of K_n, define $V(H_0)$ and $E(H_0)$ to be the vertex and edge sets of H_0, respectively. A clique cover $\mathbf{C} = \{C_1, \ldots, C_t\}$ of H_0 is a set of subsets of $V(H_0)$ such that

- $|C_i| \geq 2$ for all $C_i \in \mathbf{C}$
- Each $C_i \in \mathbf{C}$ induces a clique in H_0
- For any $\{v_1, v_2\} \in E(H_0)$ there exists $C_i \in \mathbf{C}$ such that $v_1, v_2 \in C$.

Clique covers of K_3



Clique covers of K_3



Clique cover induction

Definition of induction by a clique cover

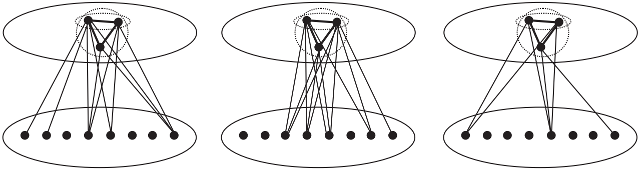
We say that H_0 is induced by clique cover $\mathbf{C} = \{C_1, \ldots, C_t\}$ if there is a family of disjoint non-empty subsets $\{W_1, \ldots, W_t\}$ of $\mathcal{W}$, such that,

- For all $1 \leq i \leq t$, each element of W_i is an object assigned to all the vertices of C_i and no other vertices from $V(H_0)$.
- Each $w \in \mathcal{W} \setminus \bigcup_{i=1}^{t} W_i$ is chosen by at most one vertex from $V(H_0)$.

Clearly, if H_0 is an induced subgraph of $\mathcal{G}(n, m, p)$, then it is induced by exactly one clique cover in $\mathcal{G}(n, m, p)$.

Clique covers of K_3

A clique cover can be induced in many different ways.



Restricted clique covers

Given $\emptyset \subsetneq S \subsetneq V(H)$, we define two different types of *restricted clique covers*, which are multisets defined by

$$\mathbf{C}[S] := \{C_i \cap S : |C_i \cap S| \geq 1,\ i \in [t]\}$$

and

$$\mathbf{C}'[S] := \{C_i \cap S : |C_i \cap S| \geq 2,\ i \in [t]\}.$$

We define

$$\sum \mathbf{C}[S] = \sum_{\substack{i \in [t] \\ |C_i \cap S| \geq 1}} |C_i \cap S|$$

and

$$\sum \mathbf{C}'[S] = \sum_{\substack{i \in [t] \\ |C_i \cap S| \geq 2}} |C_i \cap S|.$$

Let $H_0[S]$ the induced subgraph of H_0 on vertices S. Clearly, it is impossible for $\mathbf{C}$ to induce H_0 unless $\mathbf{C}[S]$ and $\mathbf{C}'[S]$ induce $H_0[S]$.

The expected number of copies of $H_0[S]$ in $\mathcal{G}(n, m, p)$ induced by $\mathbf{C}[S]$ and $\mathbf{C}'[S]$ are asymptotically of order

$$\psi(H_0, \mathbf{C}, S) := \min\left\{ n^{|S|+\alpha|\mathbf{C}[S]|} p^{\sum \mathbf{C}[S]}, n^{|S|+\alpha|\mathbf{C}'[S]|} p^{\sum \mathbf{C}'[S]} \right\}$$

Thresholds for induction by a clique cover

Consider clique cover C

Define

$$\eta_2(H_0, \mathbf{C}, S) := \begin{cases} \frac{|S| + \alpha |\mathbf{C}[S]|}{\sum \mathbf{C}[S]} & \text{if either } \alpha < \frac{|S|}{\sum \mathbf{C}[S] - |\mathbf{C}[S]|} \\ & \text{or } \sum \mathbf{C}[S] = |\mathbf{C}[S]|; \\ \frac{|S| + \alpha |\mathbf{C}'[S]|}{\sum \mathbf{C}'[S]} & \text{otherwise,} \end{cases}$$

and define

$$\eta_1(H_0, \mathbf{C}) := \min_{\emptyset \subsetneq S \subseteq V(H_0)} \eta_2(H_0, \mathbf{C}, S).$$

Then $p = n^{-\eta_1(H_0, \mathbf{C})}$ should be a threshold for **C** inducing copies of H_0.

Thresholds for induction by a clique cover

Thus, if we define

$$\eta_0 = \eta_0(H_0) := \max_{\mathbf{C}\in\mathcal{C}} \eta_1(H_0, \mathbf{C}),$$

then it is natural to expect that under suitable conditions $p = n^{-\eta_0(H_0)}$ should be the threshold for the appearance of H_0 in $\mathcal{G}(n, m, p)$.

The clique covers which will induce copies of H_0 at threshold $p = n^{-\eta_0(H_0)}$ are those for which $\eta_1(H_0, \mathbf{C}) = \eta_0$. We call them critical.

We call a clique cover $\mathbf{C} \in \mathcal{C}(H_0)$ strictly α-balanced if $\eta_2(H_0, \mathbf{C}, S) > \eta_2(H_0, \mathbf{C}, V(H_0))$ for all $\emptyset \subsetneq S \subsetneq V(H_0)$. The idea of strictly α-balanced was introduced by Jaworski and Karonski for random bipartite graphs.

Thresholds for induction by a clique cover

Thus, if we define

$$\eta_0 = \eta_0(H_0) := \max_{\mathbf{C}\in\mathcal{C}} \eta_1(H_0, \mathbf{C}),$$

then it is natural to expect that under suitable conditions $p = n^{-\eta_0(H_0)}$ should be the threshold for the appearance of H_0 in $\mathcal{G}(n, m, p)$.

The clique covers which will induce copies of H_0 at threshold $p = n^{-\eta_0(H_0)}$ are those for which $\eta_1(H_0, \mathbf{C}) = \eta_0$. We call them critical.

We call a clique cover $\mathbf{C} \in \mathcal{C}(H_0)$ strictly α-balanced if $\eta_2(H_0, \mathbf{C}, S) > \eta_2(H_0, \mathbf{C}, V(H_0))$ for all $\emptyset \subsetneq S \subsetneq V(H_0)$. The idea of strictly α-balanced was introduced by Jaworski and Karonski for random bipartite graphs.

Thresholds for induction by a clique cover

Thus, if we define

$$\eta_0 = \eta_0(H_0) := \max_{\mathbf{C}\in\mathcal{C}} \eta_1(H_0, \mathbf{C}),$$

then it is natural to expect that under suitable conditions $p = n^{-\eta_0(H_0)}$ should be the threshold for the appearance of H_0 in $\mathcal{G}(n, m, p)$.

The clique covers which will induce copies of H_0 at threshold $p = n^{-\eta_0(H_0)}$ are those for which $\eta_1(H_0, \mathbf{C}) = \eta_0$. We call them critical.

We call a clique cover $\mathbf{C} \in \mathcal{C}(H_0)$ strictly α-balanced if $\eta_2(H_0, \mathbf{C}, S) > \eta_2(H_0, \mathbf{C}, V(H_0))$ for all $\emptyset \subsetneq S \subsetneq V(H_0)$. The idea of strictly α-balanced was introduced by Jaworski and Karonski for random bipartite graphs.

Balanced and unbalanced clique covers

We call a clique cover $\mathbf{C} \in \mathcal{C}(H_0)$ α-balanced if $\eta_2(H_0, \mathbf{C}, S) = \eta_2(H_0, \mathbf{C}, V(H_0))$ for all $\emptyset \subsetneq S \subsetneq V(H_0)$.

We call a clique cover $\mathbf{C} \in \mathcal{C}(H_0)$ strictly α-unbalanced if $\eta_2(H_0, \mathbf{C}, S) < \eta_2(H_0, \mathbf{C}, V(H_0))$ for some $\emptyset \subsetneq S \subsetneq V(H_0)$.

Work in progress

≈Theorem

Suppose at least one of the critical clique covers of H_0 is strictly α-balanced and all of them are either strictly α-balanced or α-unbalanced. If $p \sim cn^{-\eta_o(H_0)}$, then X converges to a Poisson distributed random variable with parameter λ. If at least one of the critical clique covers is α-unbalanced, then $\lambda < \lim_{n\to\infty} \mathbb{E}(X)$. If all of the critical clique covers are α-balanced, then $\lambda = \lim_{n\to\infty} \mathbb{E}(X)$

What we think is true, but might not be able to show yet

If any of the critical clique covers are α-balanced, but not strictly α-balanced, then we should not have Poisson convergence.

Work in progress

$\approx$Theorem

Suppose at least one of the critical clique covers of H_0 is strictly α-balanced and all of them are either strictly α-balanced or α-unbalanced. If $p \sim cn^{-\eta_o(H_0)}$, then X converges to a Poisson distributed random variable with parameter λ. If at least one of the critical clique covers is α-unbalanced, then $\lambda < \lim_{n\to\infty} \mathbb{E}(X)$. If all of the critical clique covers are α-balanced, then $\lambda = \lim_{n\to\infty} \mathbb{E}(X)$

What we think is true, but might not be able to show yet

If any of the critical clique covers are α-balanced, but not strictly α-balanced, then we should not have Poisson convergence.

Stein's Method

We use the following well known theorem for approximating distributions by Poisson distributions.
Define the total variation distance between two integer-valued random variables by

$$d_{\mathrm{TV}}(X, Y) = \frac{1}{2} \sum_{i=0}^{\infty} |\Pr\{X = i\} - \Pr\{Y = i\}|.$$

Given a set of indices Γ and indicator variables I_α, $\alpha \in \Gamma$, let $X = \sum_{\alpha \in \Gamma} I_\alpha$ and $Y \sim \mathrm{Poisson}(\lambda)$, where $\lambda = \mathbb{E}(X)$.
A dependency graph L with vertices Γ is such that if $\{\alpha, \beta\}$ is *not* an edge in L, then I_α and I_β are independent.

Stein's Method

Then,

Theorem (Janson, Rucinski, T. Łuczak)

$$d_{TV}(X, Y) \leq \min(\lambda^{-1}, 1) \left(\sum_{\alpha \in V(L)} \pi_\alpha^2 + \sum_{\{\alpha,\beta\} \in E(L)} (\pi_\alpha \pi_\beta + \mathbb{E}(I_\alpha I_\beta) \right),$$

Our index set is $\Gamma = \{(H_i, \mathbf{C})\}$ where H_i is an isomorphic copy of H_0 and $\mathbf{C}$ is a clique cover of H_i. There is an edge in the dependency graph between $(H_i, \mathbf{C})$ and $(H_j, \mathbf{C}^*)$ if and only if H_i and H_j have vertices in common. $I_{(H_i,\mathbf{C})}$ is the indicator random variable that $\mathbf{C}$ induces H_i.

Application of Stein's Method

For $\{\alpha, \beta\} \in E(L)$, it is enough to show that

$$\mathbb{E}(I_{(H_i,\mathbf{C})} I_{(H_j,\mathbf{C}^*)}) = O(1)\mathbb{E}(I_{(H_i,\mathbf{C})})\mathbb{E}(I_{(H_j,\mathbf{C}^*)}) \frac{n^{|V(H_i)\cap V(H_j)|}}{\omega(H_i,\mathbf{C})},$$

where

$$\omega(H_i,\mathbf{C}) = \min_{\emptyset \subsetneq S \subsetneq V(H_i)} \psi(H_i,\mathbf{C}).$$

THANK YOU