Lecture 4: Phase Transition in Random Graphs

Radu Balan

February 21, 2017

Analytical Results

Distributions

Today we discuss about phase transition in random graphs. Recall on the *Erdöd-Rényi class* $\mathcal{G}_{n,p}$ of random graphs, the probability mass function on $\mathcal{G}$, $P : \mathcal{G} \to [0,1]$, is obtained by assuming that, as random variables, edges are independent from one another, and each edge occurs with probability $p \in [0,1]$. Thus a graph $G \in \mathcal{G}$ with m vertices will have probability $P(G)$ given by

$$P(G) = p^m (1-p)^{\left(\begin{array}{c} n \\ 2 \end{array}\right) - m}.$$

Analytical Results
Distributions

Today we discuss about phase transition in random graphs. Recall on the *Erdöd-Rényi class* $\mathcal{G}_{n,p}$ of random graphs, the probability mass function on $\mathcal{G}$, $P : \mathcal{G} \to [0,1]$, is obtained by assuming that, as random variables, edges are independent from one another, and each edge occurs with probability $p \in [0,1]$. Thus a graph $G \in \mathcal{G}$ with m vertices will have probability $P(G)$ given by

$$P(G) = p^m (1-p)^{\left(\begin{array}{c} n \\ 2 \end{array}\right) - m}.$$

Recall the expected number of q-cliques X_q is

$$\mathbb{E}[X_q] = \left(\begin{array}{c} n \\ q \end{array}\right) p^{q(q-1)/2}$$

Analytical Results
Distributions

We shall also use *gnm* the set of all graphs on n vertices with m edges. The set $\Gamma^{n,m}$ has cardinal

$$\left(\begin{array}{c} \left(\begin{array}{c} n \\ 2 \end{array} \right) \\ m \end{array} \right).$$

In $\Gamma^{n,m}$ each graph is equally probable.

Analytical Results
Cliques

The case of 3-cliques: $\mathbb{E}[X_3] = \theta n^3 p^3$ ($\theta \sim \frac{1}{6}$).
The case of 4-cliques: $\mathbb{E}[X_4] = \theta n^4 p^6$ ($\theta \sim \frac{1}{24}$).
The first problem we consider is thesize of the largest clique of a random graph.
Note, finding the size of the largest clique (called *the clique number*) is a NP-hard problem.

Analytical Results
Cliques

The case of 3-cliques: $\mathbb{E}[X_3] = \theta n^3 p^3$ $(\theta \sim \frac{1}{6})$.
The case of 4-cliques: $\mathbb{E}[X_4] = \theta n^4 p^6$ $(\theta \sim \frac{1}{24})$.
The first problem we consider is thesize of the largest clique of a random graph.
Note, finding the size of the largest clique (called *the clique number*) is a NP-hard problem.
Idea: Analyze p so that $\mathbb{E}[X_q] \approx 1$.

- For $p > \frac{1}{n}$ and large n we expect that graphs will have a 3-clique;
- For $p > \frac{1}{n^{2/3}}$ and large n, we expect that graphs will have a 4-clique;

Analytical Results

Cliques

The case of 3-cliques: $\mathbb{E}[X_3] = \theta n^3 p^3$ $(\theta \sim \frac{1}{6})$.
The case of 4-cliques: $\mathbb{E}[X_4] = \theta n^4 p^6$ $(\theta \sim \frac{1}{24})$.
The first problem we consider is thesize of the largest clique of a random graph.
Note, finding the size of the largest clique (called *the clique number*) is a NP-hard problem.
Idea: Analyze p so that $\mathbb{E}[X_q] \approx 1$.

- For $p > \frac{1}{n}$ and large n we expect that graphs will have a 3-clique;
- For $p > \frac{1}{n^{2/3}}$ and large n, we expect that graphs will have a 4-clique;

Question: How sharp are these thresholds?

Analytical Results

3-Cliques

Theorem

Let $p = p(n)$ *be the edge probability in* $\mathcal{G}_{n,p}$.

1. *If* $p \gg \frac{1}{n}$ *(i.e.* $\lim_{n\to\infty} np = \infty$*) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 3-clique] \to 1$.
2. *If* $p \ll \frac{1}{n}$ *(i.e.* $\lim_{n\to\infty} np = 0$*) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 3-clique] \to 0$.

Analytical Results
3-Cliques

Theorem

Let $p = p(n)$ be the edge probability in $\mathcal{G}_{n,p}$.

1. *If $p \gg \frac{1}{n}$ (i.e. $\lim_{n\to\infty} np = \infty$) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 3 - clique] \to 1.$
2. *If $p \ll \frac{1}{n}$ (i.e. $\lim_{n\to\infty} np = 0$) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 3 - clique] \to 0.$

Theorem

Let $m = m(n)$ be the number of edges in $\Gamma^{n,m}$.

1. *If $m \gg n$ (i.e. $\lim_{n\to\infty} \frac{m}{n} = \infty$) then* $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}\ has\ a\ 3 - clique] \to 1.$
2. *If $m \ll n$ (i.e. $\lim_{n\to\infty} \frac{m}{n} = 0$) then* $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}\ has\ a\ 3 - clique] \to 0.$

Analytical Results
4-Cliques

Theorem

Let $p = p(n)$ *be the edge probability in* $\mathcal{G}_{n,p}$.

1. *If* $p \gg \frac{1}{n^{2/3}}$ (*i.e.* $\lim_{n\to\infty} n^{2/3} p = \infty$) *then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 4 - clique] \to 1$.
2. *If* $p \ll \frac{1}{n^{2/3}}$ (*i.e.* $\lim_{n\to\infty} n^{2/3} p = 0$) *then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 4 - clique] \to 0$.

Analytical Results

4-Cliques

Theorem

Let $p = p(n)$ *be the edge probability in* $\mathcal{G}_{n,p}$.

1. *If* $p \gg \frac{1}{n^{2/3}}$ *(i.e.* $\lim_{n\to\infty} n^{2/3}p = \infty$*) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 4-clique] \to 1$.
2. *If* $p \ll \frac{1}{n^{2/3}}$ *(i.e.* $\lim_{n\to\infty} n^{2/3}p = 0$*) then* $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ 4-clique] \to 0$.

Theorem

Let $m = m(n)$ *be the number of edges in* $\Gamma^{n,m}$.

1. *If* $m \gg n^{4/3}$ *(i.e.* $\lim_{n\to\infty} \frac{m}{n^{4/3}} = \infty$*) then* $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}\ has\ a\ 4-clique] \to 1$.
2. *If* $m \ll n^{4/3}$ *(i.e.* $\lim_{n\to\infty} \frac{m}{n^{4/3}} = 0$*) then* $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}\ has\ a\ 4-clique] \to 0$.

Analytical Results

q-Cliques

Theorem

Let $p = p(n)$ be the edge probability in $\mathcal{G}_{n,p}$. Let $q \geq 3$ be and integer.

1. *If $p \gg \frac{1}{n^{2/(q-1)}}$ (i.e. $\lim_{n\to\infty} n^{2/(q-1)} p = \infty$) then $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ q-clique] \to 1$.*
2. *If $p \ll \frac{1}{n^{2/(q-1)}}$ (i.e. $\lim_{n\to\infty} n^{2/(q-1)} p = 0$) then $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}\ has\ a\ q-clique] \to 0$.*

Analytical Results

q-Cliques

Theorem

Let $p = p(n)$ be the edge probability in $\mathcal{G}_{n,p}$. Let $q \geq 3$ be and integer.

1. *If $p \gg \frac{1}{n^{2/(q-1)}}$ (i.e. $\lim_{n\to\infty} n^{2/(q-1)} p = \infty$) then $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}$ has a $q - clique] \to 1$.*
2. *If $p \ll \frac{1}{n^{2/(q-1)}}$ (i.e. $\lim_{n\to\infty} n^{2/(q-1)} p = 0$) then $\lim_{n\to\infty} Prob[G \in \mathcal{G}_{n,p}$ has a $q - clique] \to 0$.*

Theorem

Let $m = m(n)$ be the number of edges in $\Gamma^{n,m}$. Let $q \geq 3$ be and integer.

1. *If $m \gg n^{2(q-2)/(q-1)}$ (i.e. $\lim_{n\to\infty} \frac{m}{n^{2(q-2)/(q-1)}} = \infty$) then $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}$ has a $q - clique] \to 1$.*
2. *If $m \ll n^{2(q-2)/(q-1)}$ (i.e. $\lim_{n\to\infty} \frac{m}{n^{2(q-1)/(q-1)}} = 0$) then $\lim_{n\to\infty} Prob[G \in \Gamma^{n,m}$ has a $q - clique] \to 0$.*

Analytical Results

Markov and Chebyshev Inequalities

We want to control probabilities of the random event $X_3(G) > 0$. Two important tools:

1. (Markov's Inequality) Assume X is a non-negative random variable. Then $Prob[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$.
2. (Chebyshev's Inequality) For any random variable X, $Prob[|X - E[X]| \geq t] \leq \frac{Var[X]}{t^2}$.

where $\mathbb{E}[X]$ is the mean of X, and $Var[X] = \mathbb{E}[X^2] - |\mathbb{E}[X]|^2$ is the variance of X.

Analytical Results
Markov and Chebyshev Inequalities

We want to control probabilities of the random event $X_3(G) > 0$. Two important tools:

1. (Markov's Inequality) Assume X is a non-negative random variable. Then $Prob[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$.
2. (Chebyshev's Inequality) For any random variable X, $Prob[|X - E[X]| \geq t] \leq \frac{Var[X]}{t^2}$.

where $\mathbb{E}[X]$ is the mean of X, and $Var[X] = \mathbb{E}[X^2] - |\mathbb{E}[X]|^2$ is the variance of X. *Quick Proof:*

$$Prob[X \geq t] = \int_t^\infty p_X(x)dx \leq \frac{1}{t}\int_t^\infty x p_X(x)dx \leq \frac{\mathbb{E}[X]}{t}.$$

$$Prob[|X - \mathbb{E}[X]| \geq t] = P[|X - \mathbb{E}[X]|^2 \geq t^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{t^2} = \frac{Var[X]}{t^2}.$$

Analytical Results
Proofs for the 3-clique case

For small probability: We shall use Markov's inequality to show $Prob[X_3 > 0] \to 0$ when $p \ll \frac{1}{n}$:

$$Prob[X_3 > 0] = Prob[X_3 \geq 1] \leq \frac{E[X_3]}{1} = \frac{n(n-1)(n-2)}{6} p^3 = \theta n^3 p^3 \to 0.$$

Analytical Results

Proofs for the 3-clique case

For small probability: We shall use Markov's inequality to show $Prob[X_3 > 0] \to 0$ when $p \ll \frac{1}{n}$:

$$Prob[X_3 > 0] = Prob[X_3 \geq 1] \leq \frac{E[X_3]}{1} = \frac{n(n-1)(n-2)}{6} p^3 = \theta n^3 p^3 \to 0.$$

For large probability: Since $\mathbb{E}[X_3] \to \infty$ it follows that $Prob[X_3 > 0] > 0$. We need to show that $Prob[X_3 = 0] \to 0$. By Chebyshev's inequality:

$$Prob[X_3 = 0] \leq Prob[|X_3 - \mathbb{E}[X_3]| \geq \mathbb{E}[X_3]] \leq \frac{Var[X_3]}{|\mathbb{E}[X_3]|^2}$$

Analytical Results

Proofs for the 3-clique case

For small probability: We shall use Markov's inequality to show $Prob[X_3 > 0] \to 0$ when $p \ll \frac{1}{n}$:

$$Prob[X_3 > 0] = Prob[X_3 \geq 1] \leq \frac{E[X_3]}{1} = \frac{n(n-1)(n-2)}{6} p^3 = \theta n^3 p^3 \to 0.$$

For large probability: Since $\mathbb{E}[X_3] \to \infty$ it follows that $Prob[X_3 > 0] > 0$. We need to show that $Prob[X_3 = 0] \to 0$. By Chebyshev's inequality:

$$Prob[X_3 = 0] \leq Prob[|X_3 - \mathbb{E}[X_3]| \geq \mathbb{E}[X_3]] \leq \frac{Var[X_3]}{|\mathbb{E}[X_3]|^2}$$

Need the variance of $X_3 = \sum_{(i,j,k)\in S_3} 1_{i,j,k}$,

$$X_3^2 = \sum_{(i,j,k)\in S_3} \sum_{(i',j',k')\in S_3} 1_{i,j,k} 1_{i',j',k'}.$$

Analytical Results

Proofs for the 3-clique case

$$X_3^2 = \sum_{(i,j,k)\in S_3(n)} 1_{i,j,k} + \sum_{(i,j,k)\in S_3(n)} \sum_{l\in S_1(n-3)} (1_{i,j,k}1_{i,j,l} + 1_{i,j,k}1_{j,k,l} + 1_{i,j,k}1_{k,i,l}) +$$

$$+ \sum_{(i,j,k)\in S_3(n)} \sum_{u,v\in S_2(n-3)} (1_{i,j,k}1_{i,u,v} + 1_{i,j,k}1_{j,u,v}1_{i,j,k}1_{k,u,v}) +$$

$$+ \sum_{(i,j,k)\in S_3(n)} \sum_{(i',j',k')\in S_3(n-3)} 1_{i,j,k}1_{i',j',k'}$$

Analytical Results

Proofs for the 3-clique case

$$X_3^2 = \sum_{(i,j,k)\in S_3(n)} 1_{i,j,k} + \sum_{(i,j,k)\in S_3(n)} \sum_{l\in S_1(n-3)} (1_{i,j,k}1_{i,j,l} + 1_{i,j,k}1_{j,k,l} + 1_{i,j,k}1_{k,i,l}) +$$

$$+ \sum_{(i,j,k)\in S_3(n)} \sum_{u,v\in S_2(n-3)} (1_{i,j,k}1_{i,u,v} + 1_{i,j,k}1_{j,u,v}1_{i,j,k}1_{k,u,v}) +$$

$$+ \sum_{(i,j,k)\in S_3(n)} \sum_{(i',j',k')\in S_3(n-3)} 1_{i,j,k}1_{i',j',k'}$$

$$\mathbb{E}[X_3^2] = |S_3|p^3 + 3|S_3|(n-3)p^5 + 3|S_3|\begin{pmatrix} n-3 \\ 2 \end{pmatrix} p^6 + |S_3| \begin{pmatrix} n-3 \\ 3 \end{pmatrix} p^6.$$

Thus

$$Var[X_3] = \mathbb{E}[X_3^2] - |\mathbb{E}[X_3]|^2 = \ldots = \theta(n^3p^3 + n^4p^5 + n^5p^6).$$

Analytical Results

Proofs for the 3-clique case

and:

$$Prob[X_3 = 0] \leq \frac{\theta(n^3p^3 + n^4p^5 + n^5p^6)}{\theta(n^6p^6)} \; \frac{1}{(np)^3} + \frac{1}{n} \to 0.$$

Analytical Results
Proofs for the 3-clique case

and:

$$Prob[X_3 = 0] \leq \frac{\theta(n^3p^3 + n^4p^5 + n^5p^6)}{\theta(n^6p^6)} \frac{1}{(np)^3} + \frac{1}{n} \to 0.$$

Similar proofs for the other cases (4-cliques and q-cliques).

Analytical Results

Behavior at the threshold

In general we obtain a "coarse threshold". Recall a Poisson process X with parameter λ has p.m.f. $Prob[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}$.

Theorem

In $\mathcal{G}_{n,p}$,

1. *For* $p = \frac{c}{n}$, X_3 *is asymptotically Poisson with parameter* $\lambda = c^3/6$.
2. *For* $p = \frac{c}{n^{2/3}}$, X_4 *is asymptotically Poisson with parameter* $\lambda = c^6/24$.

Analytical Results

Behavior at the threshold

In general we obtain a "coarse threshold". Recall a Poisson process X with parameter λ has p.m.f. $Prob[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}$.

Theorem

In $\mathcal{G}_{n,p}$,

1. *For $p = \frac{c}{n}$, X_3 is asymptotically Poisson with parameter $\lambda = c^3/6$.*
2. *For $p = \frac{c}{n^{2/3}}$, X_4 is asymptotically Poisson with parameter $\lambda = c^6/24$.*

Theorem

In $\Gamma^{n,m}$,

1. *For $m = cn$, X_3 is asymptotically Poisson with parameter $\lambda = 4c^3/3$.*
2. *For $m = cn^{4/3}$, X_4 is asymptotically Poisson with parameter $\lambda = 8c^6/3$.*

Analytical Results
Connected Components

$\mathcal{G}_{n,p}$ class of random graphs has a remarkable property in regards to the largest connected component. We shall express the result in the class $\Gamma^{n,m}$.

Analytical Results

Connected Components

Theorem

1. *Let* $m = m(n)$ *satisfies* $m \ll \frac{1}{2} n \log(n)$. *Then*

$$\lim_{n\to\infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 0$$

2. *Let* $m = m(n)$ *satisfies* $m \gg \frac{1}{2} n \log(n)$. *Then*

$$\lim_{n\to\infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 1$$

Analytical Results

Connected Components

Theorem

1. *Let* $m = m(n)$ *satisfies* $m \ll \frac{1}{2} n \log(n)$. *Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 0$$

2. *Let* $m = m(n)$ *satisfies* $m \gg \frac{1}{2} n \log(n)$. *Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 1$$

3. *Assume* $m = \frac{1}{2} n \log(n) + tn + o(n)$, *where* $o(n) \ll n$. *Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = e^{-e^{-2t}}$$

Analytical Results

Connected Components

Theorem

1. *Let* $m = m(n)$ *satisfies* $m \ll \frac{1}{2} n \log(n)$*. Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 0$$

2. *Let* $m = m(n)$ *satisfies* $m \gg \frac{1}{2} n \log(n)$*. Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = 1$$

3. *Assume* $m = \frac{1}{2} n \log(n) + tn + o(n)$*, where* $o(n) \ll n$*. Then*

$$\lim_{n \to \infty} Prob[G \in \Gamma^{n,m} \text{ is connected}] = e^{-e^{-2t}}$$

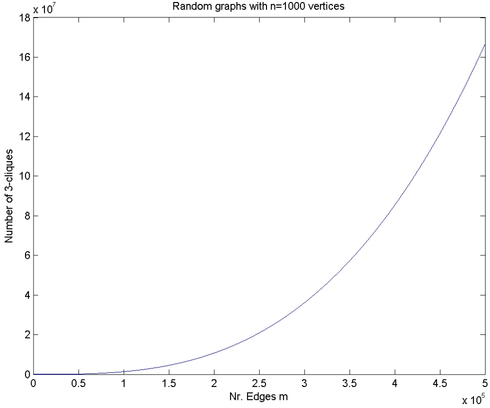
In this case $\frac{1}{2} n \log(n)$ is known as a *strong threshold*.

Numerical Results
3-cliques & Connectivity results

Results for $n = 1000$ vertices.

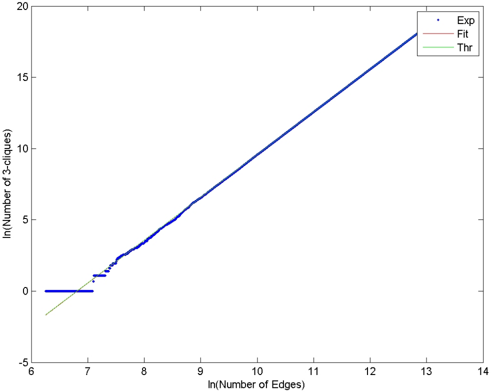
1. 3-cliques. Recall $\mathbb{E}[X_3] \sim m^3$
2. Connectivity. Recall the connectivity threshold is $\frac{1}{2} n \log(n) = 3454$.

Numerical Results
3-cliques



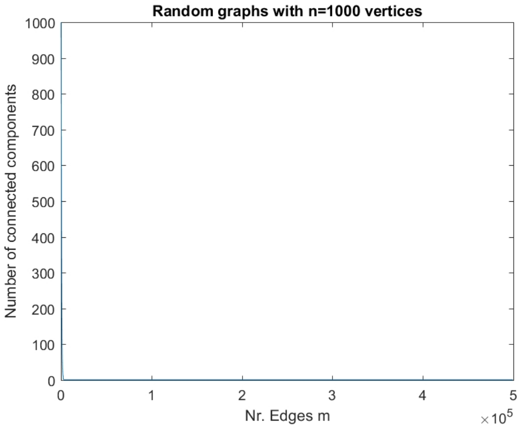
Numerical Results

3-cliques

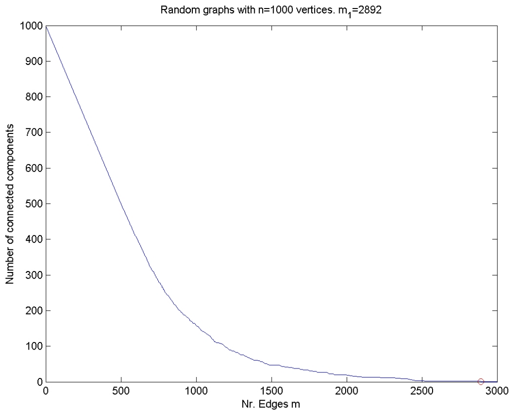


Numerical Results

Connectivity

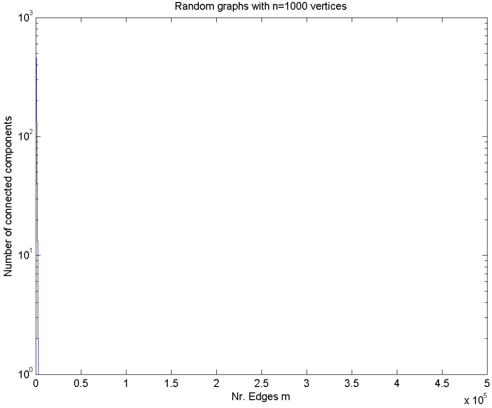


Numerical Results
Connectivity

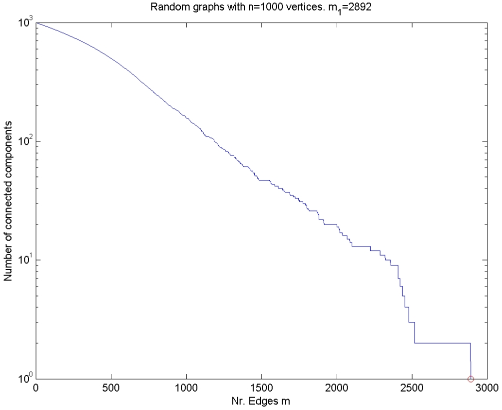


Numerical Results

Connectivity



Numerical Results
Connectivity



References

- B. Bollobás, **Graph Theory. An Introductory Course**, Springer-Verlag 1979. **99**(25), 15879–15882 (2002).
- F. Chung, **Spectral Graph Theory**, AMS 1997.
- F. Chung, L. Lu, The average distances in random graphs with given expected degrees, Proc. Nat.Acad.Sci. 2002.
- R. Diestel, **Graph Theory**, 3rd Edition, Springer-Verlag 2005.
- P. Erdös, A. Rényi, On The Evolution of Random Graphs
- G. Grimmett, **Probability on Graphs. Random Processes on Graphs and Lattices**, Cambridge Press 2010.
- J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Trans. on Knowl.Disc.Data, **1**(1) 2007.