Topic Course on Probabilistic Methods (Week 11) Random Graphs (I)

Linyuan Lu

University of South Carolina
Selected topics

- Linearity of Expectation (2 weeks)
- Alterations (1 week)
- The second moment method (1 week)
- The Local Lemma (1-2 weeks)
- Correlation Inequalities (1 week)
- Large deviation inequalities (3 weeks)
- Poisson Paradigm (1 week)
- Random graphs (2 weeks)
- Discrepancy (1 week)
- Entropy (1 week)
Subtopics

Random graphs

- Erdős-Rényi model
- Evolution of $G(n, p)$
- Galton-Watson process
- Graph branching process
- Barely subcritical regimes
$G(n, p)$: Erdős-Rényi random graphs
- n nodes
Erdős-Rényi model

\( G(n, p) \): Erdős-Rényi random graphs

- \( n \) nodes
- For each pair of vertices, create an edge independently with probability \( p \).
Erdős-Rényi model

\[ G(n, p) \]: Erdős-Rényi random graphs

- \( n \) nodes
- For each pair of vertices, create an edge independently with probability \( p \).

An example \( G(3, \frac{1}{2}) \):
The birth of random graph theory

Paul Erdős and A. Rényi, On the evolution 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 $V_1, V_2, \ldots, V_n$ 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 $V_1, V_2, \ldots, V_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{n}{2}$. 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 $1/\binom{n}{2}$. There is however an other slightly
Evolution of $G(n, p)$

- $p = 0$: the empty graph.
- $p = \frac{c}{n}$: disjoint union of trees.
- $p = \frac{1}{n}$: cycles of any size.
- $p = \frac{c'}{n}$: the double jumps.
- $p = \frac{\log n}{n}$: one giant component, others are trees.
- $p = \Omega\left(\frac{\log n}{n}\right)$: $G(n, p)$ is connected.
- $p = \Omega(n^{\epsilon - 1})$: connected and almost regular.
- $p = \Theta(1)$: finite diameter.
- $p = 1$: dense graphs, diameter is 2.
- $p$ approaching 1: the complete graph.
Range I \( p = o(1/n) \)

The random graph \( G_{n,p} \) is the disjoint union of trees. In fact, trees on \( k \) vertices, for \( k = 3, 4, \ldots \) only appear when \( p \) is of the order \( n^{-k/(k-1)} \).
Range I \( p = o(1/n) \)
The random graph \( G_{n,p} \) is the disjoint union of trees. In fact, trees on \( k \) vertices, for \( k = 3, 4, \ldots \) only appear when \( p \) is of the order \( n^{-k/(k-1)} \).

Furthermore, for \( p = cn^{-k/(k-1)} \) and \( c > 0 \), let \( \tau_k(G) \) denote the number of connected components of \( G \) formed by trees on \( k \) vertices and \( \lambda = c^{k-1}k^{k-2}/k! \). Then,

\[
\Pr(\tau_k(G_{n,p}) = j) \rightarrow \frac{\lambda^j e^{-\lambda}}{j!}
\]

for \( j = 0, 1, \ldots \) as \( n \rightarrow \infty \).
Range II \( p \sim c/n \) for \( 0 < c < 1 \)

- In this range of \( p \), \( G_{n,p} \) contains cycles of any given size with probability tending to a positive limit.
Range II \( p \sim c/n \) for \( 0 < c < 1 \)

- In this range of \( p \), \( G_{n,p} \) contains cycles of any given size with probability tending to a positive limit.
- All connected components of \( G_{n,p} \) are either trees or unicyclic components. Almost all (i.e., \( n - o(n) \)) vertices are in components which are trees.
Range II \hspace{1em} p \sim c/n \text{ for } 0 < c < 1

- In this range of $p$, $G_{n,p}$ contains cycles of any given size with probability tending to a positive limit.
- All connected components of $G_{n,p}$ are either trees or unicyclic components. Almost all (i.e., $n - o(n)$) vertices are in components which are trees.
- The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha} \left( \log n - \frac{5}{2} \log \log n \right)$ vertices, where $\alpha = c - 1 - \log c$. 
Evolution of $G(n, p)$

Range III $p \sim 1/n + \mu/n$, the double jump

- If $\mu < 0$, the largest component has size
  $$(\mu - \log(1 + \mu))^{-1} \log n + O(\log \log n).$$
Range III \( p \sim 1/n + \mu/n \), the double jump

- If \( \mu < 0 \), the largest component has size
  \[ (\mu - \log(1 + \mu))^{-1} \log n + O(\log \log n). \]
- If \( \mu = 0 \), the largest component has size of order \( n^{2/3} \).
Range III \( p \sim 1/n + \mu/n \), the double jump

- If \( \mu < 0 \), the largest component has size \((\mu - \log(1 + \mu))^{-1} \log n + O(\log \log n)\).
- If \( \mu = 0 \), the largest component has size of order \( n^{2/3} \).
- If \( \mu > 0 \), there is a unique giant component of size \( \alpha n \) where \( \mu = -\alpha^{-1} \log(1 - \alpha) - 1 \).
Range III  \( p \sim 1/n + \mu/n \), the double jump

- If \( \mu < 0 \), the largest component has size
  \[
  (\mu - \log(1 + \mu))^{-1} \log n + O(\log \log n).
  \]

- If \( \mu = 0 \), the largest component has size of order \( n^{2/3} \).

- If \( \mu > 0 \), there is a unique giant component of size \( \alpha n \)
  where \( \mu = -\alpha^{-1} \log(1 - \alpha) - 1 \).

- Bollobás showed that a component of size at least \( n^{2/3} \) in \( G_{n,p} \) is almost always unique if \( p \) exceeds
  \[
  1/n + 4(\log n)^{1/2} n^{-4/3}.
  \]
Evolution of $G(n, p)$

Range IV \( p \sim c/n \) for \( c > 1 \)

- Except for one “giant” component, all the other components are relatively small, and most of them are trees.
Range IV \( p \sim c/n \) for \( c > 1 \)

- Except for one “giant” component, all the other components are relatively small, and most of them are trees.
- The total number of vertices in components which are trees is approximately \( n - f(c)n + o(n) \).
Evolution of $G(n, p)$

Range IV $p \sim c/n$ for $c > 1$

- Except for one “giant” component, all the other components are relatively small, and most of them are trees.
- The total number of vertices in components which are trees is approximately $n - f(c)n + o(n)$.
- The largest connected component of $G_{n,p}$ has approximately $f(c)n$ vertices, where

$$f(c) = 1 - \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k.$$
Range V  \( p = \frac{c \log n}{n} \) with \( c \geq 1 \)

- The graph \( G_{n,p} \) almost surely becomes connected.
Evolution of $G(n, p)$

Range V  
$p = c \log n / n$ with $c \geq 1$

- The graph $G_{n, p}$ almost surely becomes connected.
- If

$$p = \frac{\log n}{kn} + \frac{(k - 1) \log \log n}{kn} + \frac{y}{n} + o\left(\frac{1}{n}\right),$$

then there are only trees of size at most $k$ except for the giant component. The distribution of the number of trees of $k$ vertices again has a Poisson distribution with mean value $\frac{e^{-ky}}{k \cdot k!}$.
Range VI \( p \sim \omega(n) \log n/n \) where \( \omega(n) \rightarrow \infty \).

In this range, \( G_{n,p} \) is not only almost surely connected, but the degrees of almost all vertices are asymptotically equal.
Galton-Watson branching process: Let $Z$ be a distribution over the non-negative integers. Starting with a single node, it gives $Z$ children nodes. Each of children nodes have $Z$ children independently. The process continues, each new offspring having an independent number $A$ of children.
Galton-Watson branching process: Let $Z$ be a distribution over the non-negative integers. Starting with a single node, it gives $Z$ children nodes. Each of children nodes have $Z$ children independently. The process continues, each new offspring having an independent number $A$ of children.

- $Z_1, Z_2, \ldots, Z_t, \ldots$: a countable sequence of independent identically distributed variables, each have distribution $Z$.

- $Y_t$: the number of living children at time $t$.

\[
Y_0 = 1 \\
Y_t = Y_{t-1} + Z_t - 1.
\]
Let $T$ be the total number of nodes in Galton-Watson process. There are two essentially different cases.

- $Y_t > 0$ for all $t \geq 0$. In this case the Galton-Watson process goes on forever and $T = \infty$. 

Let $T$ be the total number of nodes in Galton-Watson process. There are two essentially different cases.

- $Y_t > 0$ for all $t \geq 0$. In this case the Galton-Watson process goes on forever and $T = \infty$.
- $Y_t = 0$ for some $t \geq 0$. In this case, $T$ is the least integer for which $Y_T = 0$. The Galton-Watson process stops with $T$ nodes.
Let $Z$ be the Poisson distribution with the expectation $c$. Write $T = T_c^{po}$.

**Theorem:** If $c \leq 1$, then $T$ is finite with probability one. If $c > 1$, then $T$ is infinite with probability $y = y(c)$, where $y$ is the unique positive real satisfying

$$e^{-cy} = 1 - y.$$
Let $Z$ be the Poisson distribution with the expectation $c$. Write $T = T^p_c$.

**Theorem:** If $c \leq 1$, then $T$ is finite with probability one. If $c > 1$, then $T$ is infinite with probability $y = y(c)$, where $y$ is the unique positive real satisfying

$$e^{-cy} = 1 - y.$$

**Proof:** Suppose $c < 1$. \[ \Pr(T > t) \leq \Pr\left(\sum_{i=1}^{t} Z_i \geq t\right) < e^{-kt}, \]

for some constant $k$. \( \lim_{t \to \infty} \Pr(T > t) = 0. \)
Suppose \( c \geq 1 \). Let \( z = 1 - y = \Pr(T < \infty) \). Then

\[
z = \sum_{i=0}^{\infty} \Pr(Z_1 = i) z^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i}{z^i} i! = e^{c(z-1)}.
\]
Suppose $c \geq 1$. Let $z = 1 - y = \Pr(T < \infty)$. Then

$$z = \sum_{i=0}^{\infty} \Pr(Z_1 = i)z^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i}{z^i} i! = e^{c(z-1)}.$$

Hence $1 - y = e^{-cy}$. When $c = 1$, this equation has a unique solution $y = 0$. When $c > 1$, there are two solutions $1$ and $y(c)$. 
Suppose $c \geq 1$. Let $z = 1 - y = \Pr(T < \infty)$. Then

$$z = \sum_{i=0}^{\infty} \Pr(Z_1 = i) z^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i}{z^i} i! = e^{c(z-1)}.$$ 

Hence $1 - y = e^{-cy}$. When $c = 1$, this equation has a unique solution $y = 0$. When $c > 1$, there are two solutions $1$ and $y(c)$. By Chernoff’s equality, for any $t$

$$\Pr(\sum_{i=1}^{t} Z_i \leq t) < e^{-\frac{(c-1)^2 t}{2c}}.$$ 

There is a $t_0$ so that $\sum_{t\geq t_0} e^{-\frac{(c-1)^2 t}{2c}} < 1$. Thus, $y > \Pr(T = \infty \mid T \geq t_0) \Pr(T \geq t_0) > 0$. \qed
Graph branching process
Let $C(v)$ denote the component of $G(n, p)$, containing a vertex $v$. Explore $C(v)$ using Breadth First Search (BFS). In this procedure all vertices will be live, dead, or neutral. The live vertices will be contained in a queue $Q$. 
Let $C(v)$ denote the component of $G(n, p)$, containing a vertex $v$. Explore $C(v)$ using Breadth First Search (BFS). In this procedure all vertices will be live, dead, or neutral. The live vertices will be contained in a queue $Q$.

**Algorithm for computing $C(v)$:**

Push $v$ into $Q$. Mark all vertices but $v$ neutral.

```plaintext
while (Q is not empty) {
    Pop Q and get $w$, mark $w$ dead
    foreach ($w'$ neutral) {
        if ($ww'$ is an edge of $G(n, p)$) {
            mark $w'$ live and push it into Q
        }
    }
}
```

Return the set of all dead vertices.
In the graph branching process, let $Y_t$ be the size of the queue at time $t$ and $N_t$ be the set of neutral vertices. Let $N_t$ be the set of neutral vertices.

\[ Z_t \sim B(N_{t-1}, p). \]

\[ N_t \sim B(n - 1, (1 - p)^t). \]

If $T = t$ it is necessary that $N_t = n - t$. We have

\[ \Pr(|C(v)| = t) \leq \Pr(B(n - 1, (1 - p)^t) = n - t). \]

Or, equivalently,

\[ \Pr(|C(v)| = t) \leq \Pr(B(n - 1, 1 - (1 - p)^t) = t - 1). \]
**Theorem:** For any positive real $c$ and any fixed integer $k$

$$\lim_{n \to \infty} \Pr(|C(v)| = k \text{ in } G(n, \frac{c}{n})) = \Pr(T^{po}_c = k).$$
**Theorem:** For any positive real $c$ and any fixed integer $k$

$$\lim_{n \to \infty} \Pr(|C(v)| = k \text{ in } G(n, \frac{c}{n})) = \Pr(T^p_c = k).$$

**Proof:** Let $\Gamma$ be the set of $k$-tuples $\vec{z} = (z_1, z_2, \ldots, z_k)$ of nonnegative integers such that the recursion $y_0 = 1$, $y_t = y_{t-1} + z_t - 1$ has $y_t > 0$ for $t < k$ and $y_k = 0$.

$$\Pr(T^{gr} = k) = \sum \Pr(Z^{gr}_i = z_i, 1 \leq i \leq k)$$
$$\Pr(T^{po} = k) = \sum \Pr(Z^{po}_i = z_i, 1 \leq i \leq k).$$

Here both sums are over $\vec{z} \in \Gamma$. 
Since \( Z_{i-1} = n - O(1) \) and \( B(Z_i, p) \) approaches the Poisson distribution, we have

\[
\lim_{n \to \infty} \Pr(B(N^{\text{gr}}_{i-1}, p) = z_i) = \Pr(Z^\text{po}_i = z_i).
\]
Since \( Z_{i-1} = n - O(1) \) and \( B(Z_i, p) \) approaches the Poisson distribution, we have

\[
\lim_{n \to \infty} \Pr\left( B\left( N_{i-1}^{gr}, p \right) = z_i \right) = \Pr\left( Z_i^{po} = z_i \right).
\]

\[
\Pr(T^{gr} = k) = \Pr\left( Z_i^{gr} = z_i, 1 \leq i \leq k \right)
\]

\[
= \prod_{i=1}^{k} \Pr\left( B\left( N_{i-1}^{gr}, p \right) = z_i \right)
\]

\[
\rightarrow \prod_{i=1}^{k} \Pr\left( B\left( Z_i^{po} \right) = z_i \right)
\]

\[
= \Pr(T^{po} = k).
\]

\( \square \)
**Theorem:** For any positive real $c$ and any integer $k$, 

$$
\text{Pr}(T_c^{po} = k) = e^{-ck} \frac{(ck)^{k-1}}{k!}.
$$
Theorem: For any positive real $c$ and any integer $k$,

$$
\Pr(T^p_c = k) = e^{-ck} \frac{(ck)^{k-1}}{k!}.
$$

Proof: We have $\Pr(T^p_c = k) = \lim_{n \to \infty} \Pr(|C(v)| = k)$ in $G(n, p)$ with $p = c/n$. 

$$
\Pr(C(v) = k) \approx \binom{n}{k-1} k^{k-2} p^{k-1} (1 - p)^{k(n-k)}
$$

$$
\to e^{-ck} \frac{(ck)^{k-1}}{k!}.
$$
\[ p = \frac{c}{n}, \ 0 \leq c \leq 1 \]

With Poisson approximation,

\[
\Pr(|C(v)| \geq u) \leq (1 + o(1)) \Pr(T_{c}^{po} \geq u) \approx \sum_{k=u}^{\infty} e^{-ck} \frac{(ck)^{k-1}}{k!}.
\]
\[ p = \frac{c}{n}, \quad 0 \leq c \leq 1 \]

With Poisson approximation,

\[
\Pr(|C(v)| \geq u) \leq (1 + o(1)) \Pr(T^p_o \geq u) \approx \sum_{k=u}^{\infty} e^{-ck} \frac{(ck)^{k-1}}{k!}.
\]

Setting \( u = (c - 1 - \ln c)^{-1} \ln n + C \ln \ln n \), we have

\[
\Pr(|C(v)| \geq u) \leq o\left(\frac{1}{n \ln n}\right).
\]
\[ p = \frac{c}{n}, \ 0 \leq c \leq 1 \]

With Poisson approximation,

\[
\Pr(|C(v)| \geq u) \leq (1 + o(1))\Pr(T_c^{po} \geq u) \approx \sum_{k=u}^{\infty} e^{-ck} \frac{(ck)^{k-1}}{k!}.
\]

Setting \( u = (c - 1 - \ln c)^{-1} \ln n + C \ln \ln n \), we have

\[
\Pr(|C(v)| \geq u) \leq o\left(\frac{1}{n \ln n}\right).
\]

Thus, the size of largest component in \( G(n, p) \) is at most \((c - 1 - \ln c)^{-1} \ln n + O(\ln \ln n)\).
\[ p = \frac{c}{n}, \quad 0 \leq c \leq 1 \]

With Poisson approximation,

\[
\Pr(|C(v)| \geq u) \leq (1+o(1))\Pr(T_{c}^{po} \geq u) \approx \sum_{k=u}^{\infty} e^{-ck} \frac{(ck)^{k-1}}{k!}.
\]

Setting \( u = (c - 1 - \ln c)^{-1} \ln n + C \ln \ln n \), we have

\[
\Pr(|C(v)| \geq u) \leq o\left(\frac{1}{n \ln n}\right).
\]

Thus, the size of largest component in \( G(n, p) \) is at most 

\[
(c - 1 - \ln c)^{-1} \ln n + O(\ln \ln n).
\]

Most of them are trees. Then number of trees of size \( k \) is

\[
(1 + o(1))e^{-ck} \frac{(ck)^{k-1}}{k!} n.
\]
Barely subcritical regimes

Let $p = (1 - \epsilon)/n$ with $\epsilon = \lambda n^{-1/3}$. 
Barely subcritical regimes

Let $p = (1 - \epsilon)/n$ with $\epsilon = \lambda n^{-1/3}$.

\[
(c - 1 - \ln c)^{-1} = (-\epsilon - \ln(1 - \epsilon))^{-1} \\
\approx \frac{2}{\epsilon^2} \\
= 2n^{2/3} \lambda^{-2}.
\]
Barely subcritical regimes

Let \( p = \frac{(1 - \epsilon)}{n} \) with \( \epsilon = \lambda n^{-1/3} \).

\[
\left( c - 1 - \ln c \right)^{-1} = \left( -\epsilon - \ln(1 - \epsilon) \right)^{-1} \\
\approx \frac{2}{\epsilon^2} \\
= 2n^{2/3} \lambda^{-2}.
\]

The size of the largest component approaches \( Kn^{2/3} \lambda^{-2} \ln n \).