Announcement

- Homework 1 is out, due in two weeks from now.
- Exercises:
  - Probabilistic inequalities
  - Erdős-Rényi graphs
  - Empirical properties of networks
- You need to do 100 out of 150 points.
- You all have to do Problems 1.2(b) and 1.3.
- If you decide to do everything in the homework, the extra points count as bonus.
Announcement

- You can find a list of suggested papers for your project in the class Web page. Topics of interest include:
  - Stochastic graph models
  - Estimating models from network data
  - Strategic graph models
  - Diffusion
  - Social learning
  - Subgraphs
  - Learning
  - Cuckoo hashing
  - Kidney exchange
  - Financial networks
  - ...

- You can still propose your own project.
- Reminder: programming projects in groups of at most 2 persons.
Overview

Figure: In the last lecture we proved that the threshold for connectivity in $G(n, p)$ is $\frac{\log n}{n}$. Today, we will see the phase transition of the giant component of $G(n, p)$ where $p = \frac{1}{n}$.

We will go over two different proofs which are based on different tools.

- Branching processes (Lecture notes from Stanford available on the Web site)
- Depth first search (Readings: Krivelevich-Sudakov paper)
Phase transition

Michael Krivelevich  Benny Sudakov
the phase transition in random graphs — a simple proof

The Erdős-Rényi paper, which launched the modern theory of random graphs, has had enormous influence on the development of the field and is generally considered to be a single most important paper in Probabilistic Combinatorics, if not in all of Combinatorics
Phase transition — proof sketch

[Krivelevich and Sudakov, 2013] give a simple proof for the transition based on running depth first search (DFS) on $G$

- $S$ : vertices whose exploration is complete
- $T$ : unvisited vertices
- $U = V - (S \cup T)$ : vertices in stack

observation:

- the set $U$ always spans a path
  - when a vertex $u$ is added in $U$, it happens because $u$ is a neighbor of the last vertex $v$ in $U$; thus, $u$ augments the path spanned by $U$, of which $v$ is the last vertex
- epoch is the period of time between two consecutive emptyings of $U$
  - each epoch corresponds to a connected component
Phase transition — proof sketch

Lemma

Let $\epsilon > 0$ be a small enough constant and let $N = \binom{n}{2}$
Consider the sequence $\bar{X} = (X_i)_{i=1}^{N}$ of i.i.d. Bernoulli random variables with parameter $p$

1. let $p = \frac{1-\epsilon}{n}$ and $k = \frac{7}{\epsilon^2} \ln n$
then whp there is no interval of length $kn$ in $[N]$, in which at least $k$ of the random variables $X_i$ take value 1

2. let $p = \frac{1+\epsilon}{n}$ and $N_0 = \frac{\epsilon n^2}{2}$
then whp $\left| \sum_{i=1}^{N_0} X_i - \frac{\epsilon (1+\epsilon)n}{2} \right| \leq n^{2/3}$
Phase transition — useful tools

**Lemma (Union bound)**

For any events $A_1, \ldots, A_n$, $\Pr [A_1 \cup \ldots A_n] \leq \sum_{i=1}^{n} \Pr [A_i]$

**Lemma (Chebyshev’s inequality)**

Let $X$ be a random variable with finite expectation $E[X]$ and finite non-zero variance $\text{Var}[X]$. Then for any $t > 0$, 

$$\Pr [|X - E[X]| \geq t] \leq \frac{\text{Var}[X]}{t^2}$$

**Lemma (Chernoff bound, upper tail)**

Let $0 < \epsilon \leq 1$. Then, 

$$\Pr [\text{Bin}(n, p) \geq (1 + \epsilon)np] \leq e^{-\frac{\epsilon^2}{3}np}$$
Proof.

- fix interval $I$ of length $kn$ in $[N]$, $N = \binom{n}{2}$ then $\sum_{i \in I} X_i \sim \text{Bin}(kn, p)$
  1. apply Chernoff bound to the upper tail of $B(kn, p)$.
  2. apply union bound on all $(N - k + 1)$ possible intervals of length $kn$
     - upper bound the probability of the existence of a violating interval
       $$(N - k + 1)\Pr[B(kn, p) \geq k] < n^2 \cdot e^{-\frac{\epsilon^2}{3}(1-\epsilon)k} = o(1)$$

- sum $\sum_{i=1}^{N_0} X_i$ distributed binomially (params $N_0$ and $p$)
  - expectation: $N_0 p = \frac{\epsilon n^2 p}{2} = \frac{\epsilon (1+\epsilon)n}{2}$
  - standard deviation of order $n$
  - applying Chebyshev’s inequality gives the estimate
Phase transition — proof sketch

Proof.

- We run the DFS on a random input $G \sim G(n, p)$, fixing the order $\sigma$ on $V(G) = [n]$ to be the identity permutation.
- The DFS algorithm is given a sequence of i.i.d. Bernoulli($p$) random variables $\bar{X} = (X_i)_{i=1}^N$.
- The DFS algorithm gets its $i$-th query answered positively if $X_i = 1$, and answered negatively otherwise.
- The obtained graph is clearly distributed according to $G(n, p)$. 
Proof.

CASE I: \( p = \frac{1-\epsilon}{n} \)

- assume to the contrary that \( G \) contains a connected component \( C \) with more than \( k = \frac{7}{\epsilon^2} \ln n \) vertices
- consider the moment inside this epoch when the algorithm has found the \((k + 1)\)-st vertex of \( C \) and is about to move it to \( U \)
- denote \( \Delta S = S \cap C \) at that moment then \( |\Delta S \cup U| = k \), and thus the algorithm got exactly \( k \) positive answers to its queries to random variables \( X_i \) during the epoch, with each positive answer being responsible for revealing a new vertex of \( C \), after the first vertex of \( C \) was put into \( U \) in the beginning of the epoch.
Phase transition — proof sketch

Proof.

• at that moment during the epoch only pairs of edges touching $\Delta S \cup U$ have been queried, and the number of such pairs is therefore at most $\binom{k}{2} + k(n - k) < kn$

- it thus follows that the sequence $\bar{X}$ contains an interval of length at most $kn$ with at least $k$ 1’s inside — a contradiction to Property 1 of our Lemma.
Phase transition — proof sketch

Proof.

CASE II: $p = \frac{1+\epsilon}{n}$

- Assume that the sequence $\bar{X}$ satisfies Property 2 of our Lemma.

- **Claim:** After the first $N_0 = \frac{\epsilon n^2}{2}$ queries of the DFS algorithm, the set $U$ contains at least $\frac{\epsilon^2 n}{5}$ vertices. This means:
  - the giant component contains $O(f(\epsilon)n)$ vertices. The function $f(\epsilon) = \frac{\epsilon^2}{5}$ can be further improved by tightening the analysis of the probabilistic lemma. Check [Krivelevich and Sudakov, 2013], page 6.
  - the longest path is $O(n)$ since $U$ forms a path.

- The fact that we have performed $N_0$ queries implies an upper bound on $|S|$. Let’s see why.
Proof.

CASE II: \( p = \frac{1+\epsilon}{n} \)

- Assume for the sake of contradiction \( |S| \geq \frac{n}{3} \).
- Always \( |U| \leq 1 + \sum_{i=1}^{t} X_i \). Hence now \( |U| < \frac{n}{3} \).
- Combining the above and the fact that \( S, T, U \) are disjoint sets, we get \( |T| > \frac{n}{3} \).
- Contradiction! Why?
- Hence \( |S| < \frac{n}{3} \). Let’s assume now that \( |U| < \frac{\epsilon^2 n}{5} \) for the sake of contradiction. Clearly, \( T \neq \emptyset \).
Proof.

**CASE II: \( p = \frac{1+\epsilon}{n} \)**

- Since \( T \neq \emptyset \) the algorithm is still revealing the connected components of \( G \).
- Each positive answer it got resulted in moving a vertex from \( T \) to \( U \).
- By property 2 of the lemma, the number of positive answers is at least \( \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} \).
- These positive answers correspond to \( S, U \), namely \( |S \cup U| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} \).
- Since \( |U| \leq \frac{\epsilon^2 n}{5} \), then \( |S| \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3} \).
Phase transition — proof sketch

Proof.

**CASE II: \( p = \frac{1+\epsilon}{n} \)**

- All \(|S||T|\) pairs between \( S, T \) have been queried.
- However \(|S||T| > N_0\), contradiction!

\[
\frac{\epsilon n^2}{2} \geq N_0 \geq |S| \left( n - |S| - \frac{\epsilon^2 n}{5} \right)
\geq \left( \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3} \right) \left( n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{2} + n^{2/3} \right)
= \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{20} - O(\epsilon^3) n^2 > \frac{\epsilon n^2}{2}
\]

The phase transition in random graphs - a simple proof.