Connectivity in Random Graphs

CS648: Course Project

R havish 220879 Pradeep 210733

Department of Computer Science and Engineering
Indian Institute of Technology Kanpur

April 15, 2025

Overview

Problem

- **Random Graph Model** $G(n, p)$**:**
  - Start with n labeled vertices
  - For each of the $\binom{n}{2}$ possible edges,
  - independently add an edge with probability p
- **Critical threshold:** $p = \frac{1}{n}$
- **Subcritical regime (**$p = \frac{1-\epsilon}{n}$**):**
  - All connected components have size $O_\epsilon(\log n)$
- **Supercritical regime (**$p = \frac{1+\epsilon}{n}$**):**
  - Emergence of **giant component** with $\Theta(n)$ vertices

History

Erdős-Rényi (1960)

- First proof of sharp connectivity threshold in $G(n, p)$
- Used intricate combinatorial arguments

Krivelevich & Sudakov(2012)

- Simplified proof using DFS algorithm
- Same threshold result but Avoids complex combinatorics through graph exploration

Theorems

Subcritical Regime

Theorem

Let $\epsilon > 0$ be small enough and $G \sim G(n, p)$.

1. *For $p = \frac{1-\epsilon}{n}$:*
   whp *all connected components of G are of size at most $\frac{7}{\epsilon^2} \ln n$.*

- No giant component exists
- All components are logarithmic in size

Supercritical Regime

Theorem

Let $\epsilon > 0$ be small enough and $G \sim G(n, p)$.

2. *For $p = \frac{1+\epsilon}{n}$:*
   whp *G contains a path of length at least $\frac{\epsilon^2 n}{5}$.*

- Implies existence of giant component
- Path length scales linearly with n

Lemma 1

Lemma

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

Let $p = \frac{1-\epsilon}{n}$. Let $k = \frac{7}{\epsilon^2} \ln n$. 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.*

Sketch.

- Apply Chernoff bound to the sum of Bernoulli variables in any interval
- Use union bound over all possible intervals
- For any interval of length kn:

$$(N - k + 1)Pr[B(kn, p) \geq k] < n^2 \cdot e^{-\frac{\epsilon^2}{3}(1-\epsilon)k} < n^2 \cdot e^{-\frac{\epsilon^2(1-\epsilon)}{3}\frac{7}{\epsilon^2}\ln n} = o(1)$$

□

Lemma 2

Lemma

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

Let $p = \frac{1+\epsilon}{n}$. Let $N_0 = \frac{\epsilon n^2}{2}$. Then **whp** $\left|\sum_{i=1}^{N_0} X_i - \frac{\epsilon(1+\epsilon)n}{2}\right| \leq n^{2/3}$.

Sketch.

- Apply Chernoff bound directly to the sum of N_0 Bernoulli variables
- Expected value: $\mathbb{E}\left[\sum_{i=1}^{N_0} X_i\right] = N_0 \cdot p = \frac{\epsilon n^2}{2} \cdot \frac{1+\epsilon}{n} = \frac{\epsilon(1+\epsilon)n}{2}$

$\square$

DFS Setup

Vertex Partition

Three disjoint sets maintained:

- **S** (Explored): Vertices with complete neighborhood exploration
- **U** (Currently under Exploration): Vertices in LIFO stack being actively explored
- **T** (Unvisited): Remaining vertices ($U = V \setminus (S \cup T)$)

Initial Configuration

- Start with $S = U = \emptyset$, $T = V$
- Fixed vertex order σ prioritizes exploration

Key Mechanism

- Stack behavior enforces depth-first exploration
- σ determines neighbor selection from T

DFS Algorithm

1. While $U \neq \emptyset$:
   - Let v = last vertex in U
   - Query T for σ-first neighbor u of v
   - If exists: Move u from T to U
   - Else: Pop v from U to S
2. If $U = \emptyset$ and $T \neq \emptyset$:
   - Push σ-first vertex from T to U
3. Final phase ($U = T = \emptyset$):
   - Query all remaining unexplored edges in $S = V$

Some Useful Observations

Vertex Movement

- Each round moves exactly one vertex:
  - From T to U (exploration) or From U to S (backtracking)

Useful Inequalities

- $|S \cup U| > \sum_{i=1}^{t} X_i$ and $|U| \leq 1 + \sum_{i=1}^{t} X_i$

Edge Revelation

- **No S-T edges:** All potential edges between S and T have been queried and excluded

Path Structure

- U always spans a path (LIFO stack behavior)
- New $u \in U$ must neighbor last $v \in U$

Proof of Theorem 1 (Subcritical Regime)

Proof by contradiction.

- **Assume:** G has component C with $|C| > k = \frac{7}{\epsilon^2} \ln n$
- Track DFS epoch when C was discovered
- At moment of adding $(k+1)$-st vertex to U:
  - $|S \cup U| = k$ (requires exactly k edge discoveries)
  - Queried edges: $\leq \binom{k}{2} + k(n-k) < kn$

Contradiction with Lemma 1

- Found interval of length kn with $\geq k$ ones
- Lemma 1: No such intervals exist whp in $G(n, \frac{1-\epsilon}{n})$

Conclusion

All components have size $\leq \frac{7}{\epsilon^2} \ln n$ whp

Proof of Theorem 2:

Assumption for Contradiction

No path of length $\frac{\epsilon^2 n}{5}$ exists in $G(n, \frac{1+\epsilon}{n})$

Tracking DFS Progress

- Run DFS until $N_0 = \frac{\epsilon n^2}{2}$ edge queries
- By Lemma 2: $\sum X_i \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3}$
- Vertex movement: $|S \cup U| > \sum X_i \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3}$

Critical Case 1: $|S| \geq \frac{n}{3}$

- At some $t \leq N_0$, $|S| = \frac{n}{3}$
- Then $|U| < \frac{n}{3}$ (by Lemma 1 inequality) $\Rightarrow$ T has $\geq \frac{n}{3}$ vertices
- Queried S-T pairs: $\geq \frac{n^2}{9} >> N_0$ Contradiction!

Proof Contd.

Critical Case 2: $|S| < \frac{n}{3}$

- If $|U| < \frac{\epsilon^2 n}{5}$:
  - $|S| \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3}$
  - Remaining T size: $n - |S| - |U|$

Query Count Contradiction

- S-T pairs queried: $|S||T| \geq \frac{\epsilon n^2}{2}$
- But total queries $N_0 = \frac{\epsilon n^2}{2}$ already exceeded
- Impossible! (Algorithm would have stopped earlier)

Conclusion

- Must have $|U| \geq \frac{\epsilon^2 n}{5}$ at that point
- $\Rightarrow$ Path of required length exists $\Rightarrow$ Giant component emerges

Theoretical Extensions

Directed Graphs (Digraphs)

- Extended DFS analysis to random digraphs $D(n, p)$ showing linear-length directed paths/cycles

Pseudo-Random Graphs

- Applied to (n, d, λ)-graphs: Proved linear paths exist in random subgraphs when $p = \frac{1+\epsilon}{d}$

Positional Games

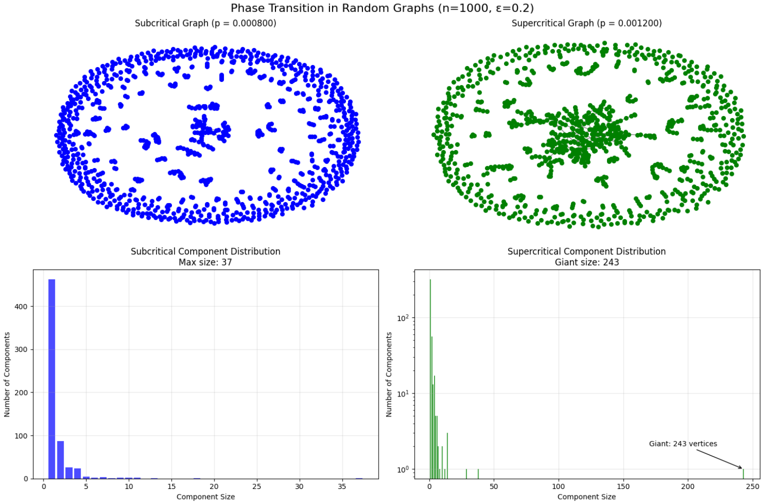
- Used in Maker-Breaker games: Strategy for creating $\Theta(\epsilon^2 n)$-length paths when $b = (1 - \epsilon)n$

Experimentation

n	ϵ	$\mathbf{p} = \frac{1-\epsilon}{n}$	$\mathbf{p} = \frac{1+\epsilon}{n}$	**Giant/2nd largest for** $\mathbf{p} = \mathbf{p} = \frac{1+\epsilon}{n}$
10	0.200	3.5 ± 1.6	4.9 ± 2.2	3.28 ± 2.55
100	0.200	12.5 ± 4.6	29.1 ± 13.8	4.16 ± 3.95
1000	0.2857	27.6 ± 14.3	367.4 ± 91.9	18.05 ± 11.76
10^4	0.1459	80.0 ± 20.6	2291.9 ± 419.1	28.21 ± 16.57
10^5	0.100	282.8 ± 68.4	17043.4 ± 1596.7	67.52 ± 24.90
10^6	0.100	550.5 ± 63.5	180686.0 ± 1377.0	351.98 ± 23.93

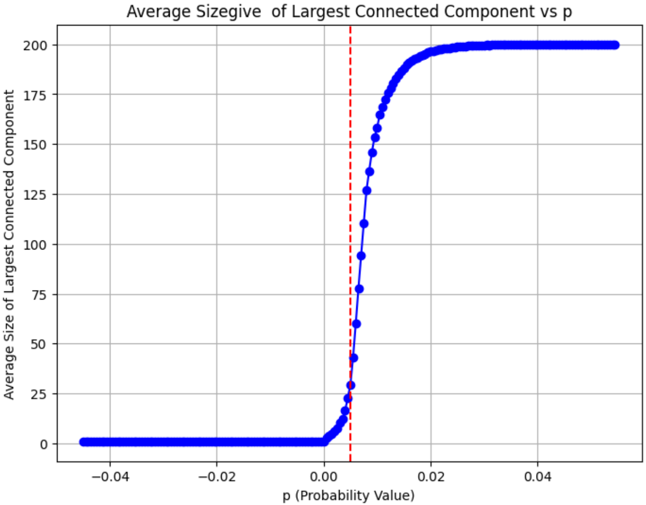
Table: Size of the Largest Component

Visualisation of Random Graphs around $p = \frac{1}{n}$



Visualization of Phase Transition in Random Graphs ($n = 1000$, $\epsilon = 0.2$)

Phase Transition in Random Graphs



Phase Transition in Random Graphs around $1/n$

Subcritical Regime: Component Size vs Theoretical Bound

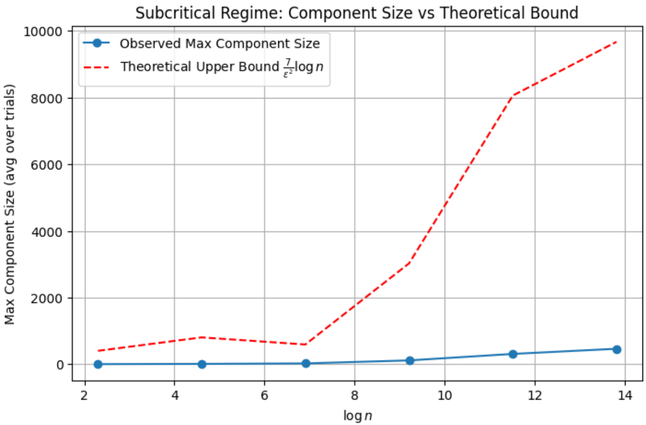


Figure: Observed vs Theoretical Component Size in Subcritical Regime.

Thank you