Connectivity in Random Graphs

Havish 220879 Pradeep 210733

April 2025

1 Introduction

In their seminal 1960 paper , Paul Erdős and Alfréd Rényi revolutionized graph theory by discovering a remarkable phase transition phenomenon in the Erdős-Rényi random graph model $G(n,p)$. Their fundamental result can be summarized as follows:

- For $p = \frac{1-\epsilon}{n}$ with constant $\epsilon > 0$, all connected components of $G(n,p)$ are *with high probability (whp)* of size at most $O_\epsilon(\log n)$
- For $p = \frac{1+\epsilon}{n}$, the graph undergoes a dramatic transformation: whp a *giant component* of linear size $\Theta(n)$ emerges

The original proof by Erdős and Rényi used intricate combinatorial arguments in the $G(n,p)$ model by using techniques such as branching processes.

The primary contribution of Krivelevich and Sudakov's work is a *simplified proof* of this fundamental result using the Depth-First Search (DFS) algorithm as an analytical tool. Their approach provides several key advantages:

- Avoids complex combinatorial calculations
- Reveals direct connections between component structure and edge probabilities
- Extends to show existence of long paths (linear in n) in supercritical regime

This report will reconstruct the complete proof using the DFS framework. The DFS-based methodology proves particularly powerful for analyzing connectivity thresholds, and we will discuss its applications to several extensions including:

- Random digraphs with linear-length directed paths
- Random subgraphs of pseudo-random graphs
- Positional games on complete graphs

2 DFS Algorithm Setting

The proof utilizes a **Depth-First Search (DFS)** algorithm adapted for random graph analysis. Let us formalize the setup:

Definition 2.1 (DFS Components)**.** *For a graph $G = (V,E)$ with $|V| = n$:*

- *S: Set of **completely explored** vertices (closed set)*
- *U: Stack containing the **currently explored path** (LIFO structure)*
- *T: Set of **unexplored** vertices $T = V \setminus (S \cup U)$*

Algorithm 1 DFS Exploration Process

1. Initialize $S = U = \emptyset$, $T = V$
2. While $U \cup T \neq \emptyset$:
   - If $U \neq \emptyset$, let $v =$ last vertex in U
     - Query edges (v, u) for $u \in T$ in fixed order σ
     - On first positive answer: move u from T to U
     - If no edges found: move v from U to S
   - If $U = \emptyset$: move first $u \in T$ to U

Example 2.2 (n=5 Vertex Case)**.** *Consider $V = \{1, 2, 3, 4, 5\}$ with identity order σ:*

1. *Initial state: $S = U = \emptyset$, $T = \{1, 2, 3, 4, 5\}$*
2. *Move 1 to U: $U = [1]$, $T = \{2, 3, 4, 5\}$*
3. *Query edges from 1 to T:*
   - *If edge (1,3) exists: $U = [1, 3]$, $T = \{2, 4, 5\}$*
   - *Then query edges from 3 to T*
4. *If no edges from 3: pop to S, $U = [1]$*
5. *Continue until U empties, then start new component*

Crucial Properties:

- **Path Maintenance**: U always spans a path (new vertices added at end)
- **Edge Isolation**: No edges exist between S and T at any stage
- **Bernoulli Edge Sampling**: For each edge query (v, u) between current vertex $v \in U$ and $u \in T$:
- **Querying Edges**:Whenever we want to query for an edge during dfs, We perform an independent Bernoulli trial with success probability p
- Trial success ($X_i = 1$) corresponds to edge existence: u moves from T to U
- Trial failure ($X_i = 0$): Continue querying next $u \in T$ per σ
- **Injective Querying**: Each potential edge is queried *exactly once* during the entire DFS process
- **Progress Measure**: At each algorithm step, exactly one of:
  - Vertex moves $T \to U$ (requires ≥ 1 edge query)
  - Vertex moves $U \to S$ (requires $|T|$ negative queries)

2.1 Component Growth Dynamics

Theorem 2.3 (Component Size Bounds)**.** *During DFS exploration of $G(n, p)$:*

1. *After t queries with $T \neq \emptyset$:*
$$|S \cup U| > \sum_{i=1}^{t} X_i$$
2. *Stack size bound:*
$$|U| \leq 1 + \sum_{i=1}^{t} X_i$$

Proof. The inequalities arise from the exploration mechanics:

1. Each positive answer $X_i = 1$ corresponds to moving a vertex from T to U, directly increasing $|S \cup U|$ as it may later move to S after completion of exploration. The inequality becomes strict when new components start with a "free" vertex moved from T to U without needing a positive query.

2. The +1 accounts for the initial vertex of each component. For a component C with k vertices:

$$|U| \leq 1 + (k-1) = 1 + \sum_{\text{component } C} X_i$$

since $k-1$ edges are needed to connect k vertices.

□

Example 2.4 (Component with 3 Vertices)**.** *Consider a component $C = \{v_1, v_2, v_3\}$ discovered through queries:*

- *v_1 enters U freely (no query)*
- *Edge (v_1, v_2) found ($X_i = 1$)*
- *Edge (v_2, v_3) found ($X_j = 1$)*

Here $|U| \leq 1 + (X_i + X_j) = 3$ while $|S \cup U| = 3 > X_i + X_j = 2$.

2.2 Proof of Lemma 1

Lemma 2.5 (Concentration of Bernoulli Sequences)**.** *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.*

1. *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.*

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

Proof of Part 1. :

Step 1: Analysis on a Single Interval. For any interval I of length kn in $[N]$, the sum $\sum_{i \in I} X_i$ follows a binomial distribution $B(kn, p)$.

The expected number of 1's in this interval is:

$$\mu = kn \cdot p = kn \cdot \frac{1-\epsilon}{n} = k(1-\epsilon) \tag{1}$$

Step 2: Applying Chernoff Bound. We apply the standard Chernoff bound for the upper tail of a binomial random variable:

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2 \mu/3} \tag{2}$$

For our case, we need to estimate:

$$\Pr\left[\sum_{i \in I} X_i \geq k\right] = \Pr\left[\sum_{i \in I} X_i \geq \frac{k}{k(1-\epsilon)} k(1-\epsilon)\right] \tag{3}$$

Setting $(1+\delta)\mu = k$, we get $\delta = \frac{k - k(1-\epsilon)}{k(1-\epsilon)} = \frac{\epsilon}{1-\epsilon}$.
Therefore:

$$\Pr\left[\sum_{i \in I} X_i \geq k\right] \leq e^{-\frac{\epsilon^2}{(1-\epsilon)^2} \cdot \frac{k(1-\epsilon)}{3}} \tag{4}$$

$$= e^{-\frac{\epsilon^2}{3(1-\epsilon)} k} \tag{5}$$

Step 3: Applying the Union Bound. The Union Bound states that for events $A_1, A_2, \ldots, A_m$:

$$\Pr[A_1 \cup A_2 \cup \ldots \cup A_m] \leq \sum_{i=1}^{m} \Pr[A_i] \tag{6}$$

We apply this to all possible intervals of length kn in $[N]$. There are at most $(N - kn + 1)$ such intervals, which is less than n^2 for sufficiently large n.

Therefore:

$$\Pr[\exists \text{ interval } I \text{ with } \sum_{i \in I} X_i \geq k] \leq (N - kn + 1) \cdot \Pr[B(kn, p) \geq k] \tag{7}$$

$$< n^2 \cdot e^{-\frac{\epsilon^2}{3(1-\epsilon)} k} \tag{8}$$

Step 4: Substituting the Value of k. With $k = \frac{7}{\epsilon^2} \ln n$, we get:

$$n^2 \cdot e^{-\frac{\epsilon^2}{3(1-\epsilon)} k} = n^2 \cdot e^{-\frac{\epsilon^2}{3(1-\epsilon)} \cdot \frac{7}{\epsilon^2} \ln n} \tag{9}$$

$$= n^2 \cdot e^{-\frac{7}{3(1-\epsilon)} \ln n} \tag{10}$$

$$= n^2 \cdot n^{-\frac{7}{3(1-\epsilon)}} \tag{11}$$

$$= n^{2 - \frac{7}{3(1-\epsilon)}} \tag{12}$$

For small enough $\epsilon > 0$, we have $\frac{7}{3(1-\epsilon)} > 2$. Thus, $n^{2 - \frac{7}{3(1-\epsilon)}} = o(1)$ as $n \to \infty$, which means this probability approaches zero as n grows.

This completes the proof that whp there is no interval of length kn containing at least k successes. $\square$

Proof of Part 2. :

For the second part, we need to show that when $p = \frac{1+\epsilon}{n}$ and $N_0 = \frac{\epsilon n^2}{2}$, the sum $\sum_{i=1}^{N_0} X_i$ concentrates around its expectation $\frac{\epsilon(1+\epsilon)n}{2}$ with high probability.

Step 1: Expected Value. The expected value of our sum is:

$$\mathbb{E}\left[\sum_{i=1}^{N_0} X_i\right] = N_0 \cdot p \tag{13}$$

$$= \frac{\epsilon n^2}{2} \cdot \frac{1+\epsilon}{n} \tag{14}$$

$$= \frac{\epsilon(1+\epsilon)n}{2} \tag{15}$$

Step 2: Using Chebyshev's Inequality. The variance of our sum is:

$$\mathrm{Var}\left[\sum_{i=1}^{N_0} X_i\right] = N_0 \cdot p \cdot (1-p) \tag{16}$$

$$= \frac{\epsilon n^2}{2} \cdot \frac{1+\epsilon}{n} \cdot \left(1 - \frac{1+\epsilon}{n}\right) \tag{17}$$

For large n, this simplifies to:

$$\mathrm{Var}\left[\sum_{i=1}^{N_0} X_i\right] \approx \frac{\epsilon(1+\epsilon)n}{2} \tag{18}$$

By Chebyshev's inequality:

$$\Pr\left(\left|\sum_{i=1}^{N_0} X_i - \frac{\epsilon(1+\epsilon)n}{2}\right| \geq n^{2/3}\right) \leq \frac{\frac{\epsilon(1+\epsilon)n}{2}}{n^{4/3}} = \frac{\epsilon(1+\epsilon)}{2} \cdot n^{-1/3} \tag{19}$$

This bound approaches 0 as $n \to \infty$, proving the concentration with polynomial rate of decay.

Step 3: Stronger Proof Using Chernoff Bounds. We now apply the Chernoff bound for sharper concentration:

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\delta^2\mu/3} \tag{20}$$

Setting $n^{2/3} = \delta\mu = \delta \cdot \frac{\epsilon(1+\epsilon)n}{2}$, we solve for δ:

$$\delta = \frac{2n^{2/3}}{\epsilon(1+\epsilon)n} = \frac{2}{\epsilon(1+\epsilon)} \cdot n^{-1/3} \tag{21}$$

For large n, $\delta < 1$, so we can apply the bound:

$$\Pr\left(\left|\sum_{i=1}^{N_0} X_i - \frac{\epsilon(1+\epsilon)n}{2}\right| \geq n^{2/3}\right) \leq 2e^{-\frac{\delta^2\mu}{3}} \tag{22}$$

$$= 2e^{-\frac{(2/(\epsilon(1+\epsilon))\cdot n^{-1/3})^2 \cdot \epsilon(1+\epsilon)n/2}{3}} \tag{23}$$

$$= 2e^{-\frac{4n^{1/3}}{3\epsilon(1+\epsilon)}} \tag{24}$$

This exponential decay rate is much faster than the polynomial rate from Chebyshev, giving us an exponentially strong concentration result.

For sufficiently large n, this probability becomes $o(1)$, which proves that with high probability:

$$\left|\sum_{i=1}^{N_0} X_i - \frac{\epsilon(1+\epsilon)n}{2}\right| \leq n^{2/3} \tag{25}$$

□

Theorem 2.6 (Phase Transition in Random Graphs)**.** *Let $\epsilon > 0$ be a small enough constant. Let $G \sim G(n,p)$.*

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

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

In both cases, we run the DFS algorithm on $G \sim G(n,p)$, and assume that the sequence $X = (X_i)_{i=1}^N$ of random variables defining the random graph satisfies the corresponding part of Lemma 1.

Proof of Part 1. We use proof by contradiction. Assume that G contains a connected component C with more than $k = \frac{7}{\epsilon^2}\ln n$ vertices.

Step 1: Consider the exact moment during the DFS algorithm when:

- The algorithm is exploring component C
- It has already discovered k vertices of C (these vertices are in $S \cup U$)
- It just discovered the $(k+1)$-st vertex of C

Let $\hat{S} = S \cap C$ be the vertices of C that have been completely explored at this moment. At this specific point, we know that $|\hat{S} \cup U| = k$.

Step 2: Count the Number of Edge Queries Made. The algorithm must have received exactly k positive answers to its edge queries during this exploration (each responsible for revealing a new vertex after the first vertex was added to U at the beginning).

At this moment, the algorithm has only queried edges where at least one endpoint is in $\hat{S} \cup U$. The number of such potential edges is at most:

$$\binom{k}{2} + k(n-k) = \frac{k(k-1)}{2} + k(n-k) \tag{26}$$

$$= \frac{k(k-1)}{2} + kn - k^2 \tag{27}$$

$$= kn - \frac{k(k+1)}{2} \tag{28}$$

$$< kn \tag{29}$$

Step 3: Contradiction. We have shown that within fewer than kn edge queries, the DFS algorithm has encountered k positive responses.

This means that the sequence $X = (X_i)_{i=1}^{N}$ of Bernoulli random variables contains an interval of length less than kn with at least k 1's inside.

However, this directly contradicts Property 1 of Lemma 1, which states that whp there is no interval of length kn in $[N]$ containing at least k 1's when $p = \frac{1-\epsilon}{n}$.

Therefore, our assumption must be false, and all connected components in G must have size at most $k = \frac{7}{\epsilon^2} \ln n$, completing the proof. □

Proof of Part 2. Assume that the sequence X satisfies Property 2 of Lemma 1. We claim that 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, with the contents of U forming a path of the desired length at that moment. We stop at any moment if we encounter that the set U contains $\frac{\epsilon^2 n}{5}$ vertices; i.e., $|U| \leq \frac{\epsilon^2 n}{5}$. □

Step 1: Prove that $|S| < \frac{n}{3}$ at time N_0. We establish this bound by contradiction. Suppose $|S| \geq \frac{n}{3}$ at time N_0. Then there must exist a specific moment $t \leq N_0$ when $|S|$ first reached exactly $\frac{n}{3}$ (since vertices enter S one at a time). At that moment:

- $|S| = \frac{n}{3}$ by definition of moment t
- $|U| \leq \frac{\epsilon^2 n}{5} < \frac{n}{3}$
- Therefore, $|T| = n - |S| - |U| \geq n - \frac{n}{3} - \frac{n}{3} = \frac{n}{3}$

The importance of this step: Establishing that $|S| < \frac{n}{3}$ ensures that $T \neq \emptyset$, which is necessary for the stack size bound $|U| \leq 1 + \sum_{i=1}^{t} X_i$ to be applicable. This condition guarantees that we're still in the component discovery phase.

Now, the DFS algorithm examines all pairs between S and T, which number at least:

$$|S||T| \geq \frac{n}{3} \cdot \frac{n}{3} = \frac{n^2}{9} > \frac{\epsilon n^2}{2} = N_0 \tag{30}$$

This contradicts our assumption that we've only made N_0 queries so far.

Step 2: Analyze at time N_0 when $|S| < \frac{n}{3}$. Assume for contradiction that $|U| < \frac{\epsilon^2 n}{5}$ at time N_0. By Property 2 of Lemma 1, the number of positive answers at that point is at least:

$$\sum_{i=1}^{N_0} X_i \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} \tag{31}$$

Since each positive answer results in moving a vertex from T to U, we have:

$$|S \cup U| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} \tag{32}$$

If $|U| \leq \frac{\epsilon^2 n}{5}$, then:

$$|S| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} - \frac{\epsilon^2 n}{5} \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3} \tag{33}$$

Step 3:Deriving the contradiction. All pairs between S and T have been probed by the algorithm (and returned negative). The number of such pairs is:

$$\begin{aligned} |S||T| &= |S| \cdot (n - |S| - |U|) && (34) \\ &\geq \left(\frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3}\right) \cdot \left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5}\right) && (35) \end{aligned}$$

For small enough ϵ, this simplifies to approximately:

$$\begin{aligned} |S||T| &\geq \left(\frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3}\right) \cdot \left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5}\right) && (36) \\ &= \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{20} - O(\epsilon^3 n^2) - O(n^{5/3}) && (37) \\ &> \frac{\epsilon n^2}{2} = N_0 && (38) \end{aligned}$$

This contradicts our assumption that we've only made N_0 queries. Therefore, we must have $|U| \geq \frac{\epsilon^2 n}{5}$ at time N_0, which means G contains a path of length at least $\frac{\epsilon^2 n}{5}$.

2.3 Cycle Formation Through Sprinkling

Theorem 2.7 (Path to Cycle Conversion)**.** *Let $G_1 \sim G(n, p_1)$ with $p_1 = \frac{1+\epsilon}{n}$ contain a path P of length $\ell \geq \frac{\epsilon^2 n}{5}$ whp. Let $G_2 \sim G(n, p_2)$ with $p_2 = \frac{1}{n^{7/6}}$ be independent of G_1. Then $G = G_1 \cup G_2$ contains a cycle of length at least $\frac{\epsilon^2 n}{5} - o(n)$ whp.*

Proof. We proceed through these steps:

Step 1: Identify Path Endpoints. Let $P = (v_1, v_2, ..., v_\ell)$ be the long path in G_1. Define:

- $A = \{v_1, v_2, ..., v_{n^{2/3}}\}$ (first $n^{2/3}$ vertices)
- $B = \{v_{\ell - n^{2/3}+1}, ..., v_\ell\}$ (last $n^{2/3}$ vertices)

Step 2: Analyze Sprinkling Probability. The number of potential edges between A and B is:

$$|A||B| = n^{2/3} \cdot n^{2/3} = n^{4/3}$$

The probability that G_2 contains at least one A-B edge is:

$$\begin{aligned} \Pr[\exists ab \in G_2] &= 1 - (1 - p_2)^{n^{4/3}} && (39) \\ &= 1 - \exp\left(-n^{4/3} \cdot \frac{1}{n^{7/6}} + O\left(\frac{n^{8/3}}{n^{7/3}}\right)\right) && (40) \\ &= 1 - \exp\left(-n^{1/6} + o(1)\right) && (41) \\ &\geq 1 - e^{-\sqrt[6]{n}} \to 1 \quad \text{as } n \to \infty && (42) \end{aligned}$$

Step 3: Cycle Formation. Any A-B edge $ab \in G_2$ creates a cycle through:

- Path segment $v_1 \rightsquigarrow a$ in P
- Edge ab from G_2
- Path segment $b \rightsquigarrow v_\ell$ in P

The cycle length is at least:

$$\ell - 2n^{2/3} \geq \frac{\epsilon^2 n}{5} - o(n)$$

Step 4: Preservation of Regime. The combined edge probability remains in the supercritical regime:

$$p = 1 - (1 - p_1)(1 - p_2) = \frac{1 + \epsilon}{n} + \frac{1}{n^{7/6}} - O\left(\frac{1}{n^{13/6}}\right)$$

which still satisfies $p = \frac{1+\epsilon'}{n}$ with $\epsilon' = \epsilon + o(1)$.

Thus, G contains a cycle of length at least $\frac{\epsilon^2 n}{5} - o(n)$ whp. $\square$

3 Proof of Giant Component Existence

Theorem 3.1 (Giant Component in Supercritical Regime)**.** *Let $p = \frac{1+\epsilon}{n}$ for $\epsilon > 0$ sufficiently small. Let $G \sim G(n, p)$. Then whp G contains a connected component with at least $\frac{\epsilon n}{2}$ vertices.*

Proof. We follow the DFS framework from Theorem 1 with enhanced analysis:

Step 1: Key Properties Define $N_0 = \frac{\epsilon n^2}{2}$. The sequence $X = (X_i)_{i=1}^N$ of Bernoulli(p) variables satisfies:

1. $\sum_{i=1}^{n^{7/4}} X_i \leq n^{5/6}$ (sparse initial phase)
2. $\forall n^{7/4} \leq t \leq N_0 : \left|\sum_{i=1}^{t} X_i - \frac{(1+\epsilon)t}{n}\right| \leq n^{2/3}$ (concentration)

Proof

Property 1: For $t = n^{7/4}$, $\sum_{i=1}^{n^{7/4}} X_i \leq n^{5/6}$

$$\mu = \frac{(1+\epsilon)n^{7/4}}{n} = (1+\epsilon)n^{3/4}$$
$$\delta = \frac{n^{5/6} - \mu}{\mu} \approx \frac{n^{5/6}}{n^{3/4}} - 1 = n^{1/12} - 1$$
$$\Pr\left[\sum X_i \geq n^{5/6}\right] \leq \exp\left(-\frac{\delta^2 \mu}{3}\right) \leq \exp\left(-\Omega(n^{1/12})\right) = o(1)$$

Property 2: For $n^{7/4} \leq t \leq N_0$, $\left|\sum_{i=1}^{t} X_i - \frac{(1+\epsilon)t}{n}\right| \leq n^{2/3}$

$$\mu_t = \frac{(1+\epsilon)t}{n}$$
$$\delta = \frac{n^{2/3}}{\mu_t} \leq \frac{n^{2/3}}{(1+\epsilon)n^{7/4}/n} = \frac{n^{2/3}}{n^{3/4}} = n^{-1/12}$$
$$\Pr\left[\left|\sum X_i - \mu_t\right| \geq n^{2/3}\right] \leq 2\exp\left(-\frac{\delta^2 \mu_t}{3}\right) \leq 2\exp\left(-\Omega(n^{1/6})\right)$$

Union bound over $O(n^2)$ possible t-values gives total failure probability $\leq n^2 e^{-\Omega(n^{1/6})} = o(1)$.

Step 2: Component Growth Analysis At time N_0, we establish:

- $|S| < \frac{n}{3}$ (via same contradiction argument as Theorem 1)
- $T \neq \emptyset$ (since $|S \cup U| \leq \frac{n}{3} + o(n) < n$)

For $t \in [n^{7/4}, N_0]$, the component size satisfies:

$$|S \cup U| \geq \frac{(1+\epsilon)t}{n} - n^{2/3}$$

Step 3: Non-Empty Stack Argument Assume for contradiction that $U = \emptyset$ at some $t \in [n^{7/4}, N_0]$. Then:

$$\begin{aligned}
t &\geq |S|(n - |S|) \quad \text{(all } S\text{-}T \text{ pairs queried)} && (43)\\
&\geq \left(\frac{(1+\epsilon)t}{n} - n^{2/3}\right)\left(n - \frac{(1+\epsilon)t}{n} + n^{2/3}\right) && (44)\\
&= (1+\epsilon)t\left(1 - \frac{(1+\epsilon)t}{n^2}\right) - O(n^{5/3}) && (45)\\
&\geq (1+\epsilon)t\left(1 - \frac{(1+\epsilon)\epsilon}{2}\right) - O(n^{5/3}) \quad (\text{using } t \leq N_0 = \frac{\epsilon n^2}{2}) && (46)\\
&> t \quad \text{for small enough } \epsilon > 0 && (47)
\end{aligned}$$

This contradiction implies $U \neq \emptyset$ throughout $[n^{7/4}, N_0]$, meaning all vertices added in this interval belong to the same component.

Step 4: Component Size Calculation The number of vertices added during $[n^{7/4}, N_0]$ is:

$$\begin{aligned}
\sum_{i=n^{7/4}}^{N_0} X_i &\geq \frac{(1+\epsilon)N_0}{n} - n^{2/3} - n^{5/6} && (48)\\
&= \frac{(1+\epsilon)\epsilon n}{2} - O(n^{5/6}) && (49)\\
&\geq \frac{\epsilon n}{2} \quad \text{for sufficiently large } n && (50)
\end{aligned}$$

Step 5: Conclusion These $\geq \frac{\epsilon n}{2}$ vertices form a single connected component, completing the proof. ☐

4 Extension to Directed Graphs

Definition 4.1 (Random Directed Graph Model)**.** *The random directed graph $D(n,p)$ consists of:*

- *Vertex set $[n] = \{1, 2, \ldots, n\}$*
- *Each ordered pair (i,j) with $i \neq j$ is included as a directed edge independently with probability p*
- *Total possible edges: $n(n-1)$*

Theorem 4.2 (Long Paths in Directed Graphs)**.** *Let $p = \frac{1+\epsilon}{n}$ for constant $\epsilon > 0$. Then $D(n,p)$ contains whp:*

- *A directed path of length $\Theta(\epsilon^2)n$*
- *A directed cycle of length $\Theta(\epsilon^2)n$*

Proof Sketch. The DFS-based proof for undirected graphs extends to directed graphs with these observations:

Key Similarities:

- The DFS algorithm remains valid for directed graphs by following outgoing edges
- Edge queries now check ordered pairs (v,u) instead of unordered pairs
- The sequence X becomes $n(n-1)$ i.i.d. Bernoulli(p) variables
- All concentration inequalities (Chernoff bounds) still apply as edges are independent

Modified Components:

- Total possible edges: $N = n(n-1)$ instead of $\binom{n}{2}$
- In the undirected proof, replace $\binom{n}{2}$ with $n(n-1)$ where applicable
- Path discovery follows outgoing edges only

Why the Proof Still Holds:

- The inequality $|U| \leq 1 + \sum X_i$ remains valid as it counts discovered vertices
- Component growth bounds depend only on edge count, not direction
- The contradiction argument for $|S||T| > N_0$ scales similarly:

$$|S||T| \geq \frac{n}{3} \cdot \frac{n}{3} = \frac{n^2}{9} > \frac{\epsilon n^2}{2} = N_0$$

  for sufficiently small ϵ

Cycle Formation: For directed cycles, the sprinkling argument requires:

- A directed edge from later in the path to earlier
- Probability analysis remains identical as directionality doesn't affect edge existence probabilities

Thus, the entire proof structure carries through with directed edges, yielding the same asymptotic results for path and cycle lengths. □

Theorem 4.3 (Long Paths in General Graphs)**.** *Let G be a finite graph with minimum degree at least n. Let $p = \frac{1+\epsilon}{n}$, for $\epsilon > 0$ constant. Form a random subgraph G_p of G by including every edge of G into G_p independently and with probability p. Then whp G_p has a path of length at least $\frac{\epsilon^2 n}{5}$.*

Proof. The proof follows the same DFS framework established in Theorem 1. We run the DFS process on G_p with a sequence $X = (X_i)_{i=1}^{N}$ of i.i.d. Bernoulli(p) random variables, where $N = |E(G)|$.

Key Modification: Counting Edges Between S and T

For any vertex $v \in S$, let's analyze its edge connections:

- v has at least $\delta(G) \geq n$ total edges in the original graph G
- These edges can only connect to:
  - Other vertices in S: at most $|S| - 1$ such edges
  - Vertices in U: at most $|U|$ such edges
  - Vertices in T: the remaining edges
- Therefore, v has at least $\delta(G) - (|S| - 1) - |U| \geq \delta(G) - |S| - |U|$ neighbors in T
- Summing over all $v \in S$, the total number of edges between S and T is at least:

$$|S|(\delta(G) - |S| - |U|) \geq |S|(n - |S| - |U|)$$

The remainder of the proof follows exactly as in Theorem 1, with this new lower bound.

Assume for contradiction that at time $N_0 = \frac{\epsilon n^2}{2}$, we have $|S| < \frac{n}{3}$ and $|U| < \frac{\epsilon^2 n}{5}$. Then:

$$
\begin{aligned}
|S||T| &\geq |S|(n - |S| - |U|) && (51)\\
&\geq \left(\frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3}\right) \cdot \left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5} - \frac{n}{3}\right) && (52)\\
&\geq \frac{\epsilon n}{2} \cdot \frac{2n}{3} - o(n^2) && (53)\\
&= \frac{\epsilon n^2}{3} - o(n^2) && (54)\\
&> \frac{\epsilon n^2}{2} = N_0 && (55)
\end{aligned}
$$

for sufficiently large n, which contradicts having only made N_0 queries.

Why Sprinkling Fails for Cycles

In the original $G(n,p)$ model, we could apply the sprinkling technique to create cycles from paths. Here, three critical issues arise:

1. **Limited Connectivity:** In $G(n,p)$, any two sets of $\Theta(n^{2/3})$ vertices have $\Theta(n^{4/3})$ potential edges between them. In contrast, G might have no edges at all between the endpoints of a long path.
2. **Sprinkling Ineffectiveness:** If we attempt to sprinkle additional edges with probability $p' = o(\frac{1}{n})$, we would need specific edges to exist in the original graph G. If those edges don't exist in G, sprinkling cannot create them in G_p.

$\square$

5 Long Paths in Random Subgraphs of Pseudorandom Graphs

Definition 5.1 ((n,d,λ)-Graph)**.** *A graph G is an (n,d,λ)-graph if:*

- *G is d-regular on n vertices*
- *All eigenvalues of its adjacency matrix, except the largest, satisfy $|\lambda_i| \leq \lambda$*

Theorem 5.2 (Long Paths in Random Subgraphs)**.** *Let G be an (n,d,λ)-graph with $\lambda = o(d)$. For $p = \frac{1+\epsilon}{d}$ with constant $\epsilon > 0$, the random subgraph G_p contains whp a path of length $\frac{\epsilon^2 n}{5}$.*

Proof. **Step 1: DFS Setup and Concentration**

Run DFS on G_p with coupling to Bernoulli(p) sequence X. After $t = \frac{\epsilon d n}{2}$ queries:

$$\sum_{i=1}^{t} X_i \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} \quad \text{whp (Lemma 1)}$$

Thus $|S \cup U| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3}$.

Step 2: Contradiction Setup

Assume $|U| < \frac{\epsilon^2 n}{5}$. Then:

$$|S| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{2/3} - \frac{\epsilon^2 n}{5} \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - o(n)$$

$$|T| = n - |S| - |U| \geq n\left(1 - \frac{\epsilon}{2} - \frac{3\epsilon^2}{10}\right) - o(n)$$

Step 3: Edge Counting

Using the expander mixing lemma:

$$e_G(S,T) \geq \frac{d}{n}|S||T| + \lambda\sqrt{|S||T|}$$

Substitute $|S| = \frac{\epsilon n}{2}(1 + \frac{3\epsilon}{5}) - o(n)$ and $|T| = n(1 - \frac{\epsilon}{2} - \frac{3\epsilon^2}{10}) - o(n)$:

$$\frac{d}{n}|S||T| = d\left[\frac{\epsilon}{2}\left(1 + \frac{3\epsilon}{5}\right)\left(1 - \frac{\epsilon}{2} - \frac{3\epsilon^2}{10}\right)\right] n - o(dn)$$

For small $\epsilon > 0$:

$$\frac{d}{n}|S||T| \geq \frac{\epsilon dn}{2}\left(1 + \frac{\epsilon}{5}\right) - o(dn)$$

Step 4: Contradiction Argument
The error term satisfies:

$$\lambda\sqrt{|S||T|} \leq \lambda n = o(dn) \quad (\text{since } \lambda = o(d))$$

Thus:

$$e_G(S,T) \geq \frac{\epsilon dn}{2}\left(1 + \frac{\epsilon}{5}\right) > \frac{\epsilon dn}{2} = t$$

But DFS queried only t edges between S and T. Contradiction! Therefore, $|U| \geq \frac{\epsilon^2 n}{5}$.

Step 5: Sprinkling for Cycles
For endpoint sets $A, B \subseteq U$ with $|A|, |B| = \frac{\epsilon^2 n}{10}$:

$$e_G(A,B) \geq \frac{d}{n}|A||B| - \lambda\sqrt{|A||B|} = \Theta(\epsilon^4 dn) - o(dn)$$

Add $G_{p'}$ with $p' = \frac{c}{d}$:

$$\mathbb{E}[e_{G_{p'}}(A,B)] = \frac{c}{d} \cdot \Theta(\epsilon^4 dn) = \Theta(\epsilon^4 cn)$$

Choosing $c = \Theta(1/\epsilon^4)$ ensures ≥ 1 edge whp. □

6 Positional Games on Complete Graphs

Definition 6.1 (Maker-Breaker Connectivity Game $L(n,b)$)**.** *The game $L(n,b)$ is played on the complete graph K_n by two players:*

- ***Maker****: On each turn, claims 1 unclaimed edge*
- ***Breaker****: On each turn, claims b unclaimed edges*
- *The game ends when all $\binom{n}{2}$ edges are claimed*

Objectives*:*

- *Maker aims to maximize the size of the largest connected component in her graph*
- *Breaker aims to minimize this size*

Theorem 6.2 (Path Creation in L(n,b))**.** *Let $\epsilon > 0$. In the game L(n,b) with $b = (1-\epsilon)n$, Maker has a strategy to create a path of length $\Theta(\epsilon^2)n$.*

Proof. Maker employs the following strategy:
Initialization:

- Start with $S = \emptyset$, $U = \{v_0\}$ for arbitrary $v_0 \in [n]$, and $T = [n] \setminus \{v_0\}$
- Sets S, U, and T partition $[n]$ throughout the game

Strategy Invariant: Maker maintains that U spans a path using only Maker's edges.
Maker's Turn:

1. Find the last vertex v along the path in U for which there exists an unclaimed edge (v,u) with $u \in T$
2. If such a vertex exists:

- Move all vertices in U that come after v in the path to S
- Claim the edge (v, u)
- Move u from T to U (extending the path)

3. If no such vertex exists:

- Move all vertices from U to S
- Select an arbitrary vertex $u \in T$ and set $U = \{u\}$
- Continue as in case 2 (if possible)

Key Property: At any point in the game, all edges between S and T have been claimed by Breaker.

Proof of Key Property: A vertex v moves from U to S precisely when no unclaimed edges (v, w) with $w \in T$ exist. This means either:

- Such edges never existed (impossible in K_n)
- Or all such edges have been claimed by Breaker

Therefore, every edge between S and T must have been claimed by Breaker.

Analysis after $\frac{\epsilon n}{2}$ rounds: At this point:

- Maker has claimed $\frac{\epsilon n}{2}$ edges
- $|S \cup U| \geq \frac{\epsilon n}{2}$ (each Maker edge connects to at least one new vertex)

Assume for contradiction that $|U| \leq \frac{\epsilon^2 n}{5}$. Then:

$$|S| \geq \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5} \tag{56}$$

$$|T| = n - |S| - |U| \geq n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5} \tag{57}$$

The number of edges between S and T is at least:

$$|S| \cdot |T| \geq \left(\frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5}\right) \cdot \left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{5}\right) \tag{58}$$

$$\geq \frac{\epsilon n}{2} \cdot (1 - \epsilon) n \quad \text{for small } \epsilon > 0 \tag{59}$$

However, Breaker has claimed only $\frac{\epsilon n}{2} \cdot (1 - \epsilon) n$ edges in total after $\frac{\epsilon n}{2}$ rounds, which is insufficient to claim all edges between S and T - a contradiction.

Therefore, $|U| > \frac{\epsilon^2 n}{5}$, meaning Maker has created a path of length at least $\frac{\epsilon^2 n}{5} = \Theta(\epsilon^2) n$. □

7 Experimentation

7.1 Analysis of size of the Largest Component

n	ϵ	$\mathbf{p} = \frac{1-\epsilon}{n}$	$\mathbf{p} = \frac{1+\epsilon}{n}$	**Giant/2nd**	**Trials**
10	0.200	3.5 ± 1.6	4.9 ± 2.2	3.28 ± 2.55	100
100	0.200	12.5 ± 4.6	29.1 ± 13.8	4.16 ± 3.95	50
1000	0.2857	27.6 ± 14.3	367.4 ± 91.9	18.05 ± 11.76	20
10^4	0.1459	80.0 ± 20.6	2291.9 ± 419.1	28.21 ± 16.57	10
10^5	0.100	282.8 ± 68.4	17043.4 ± 1596.7	67.52 ± 24.90	10
10^6	0.100	550.5 ± 63.5	180686.0 ± 1377.0	351.98 ± 23.93	2

Table 1: Size of the Largest Component

Transistion of Random Graphs Around $p = \frac{1}{n}$

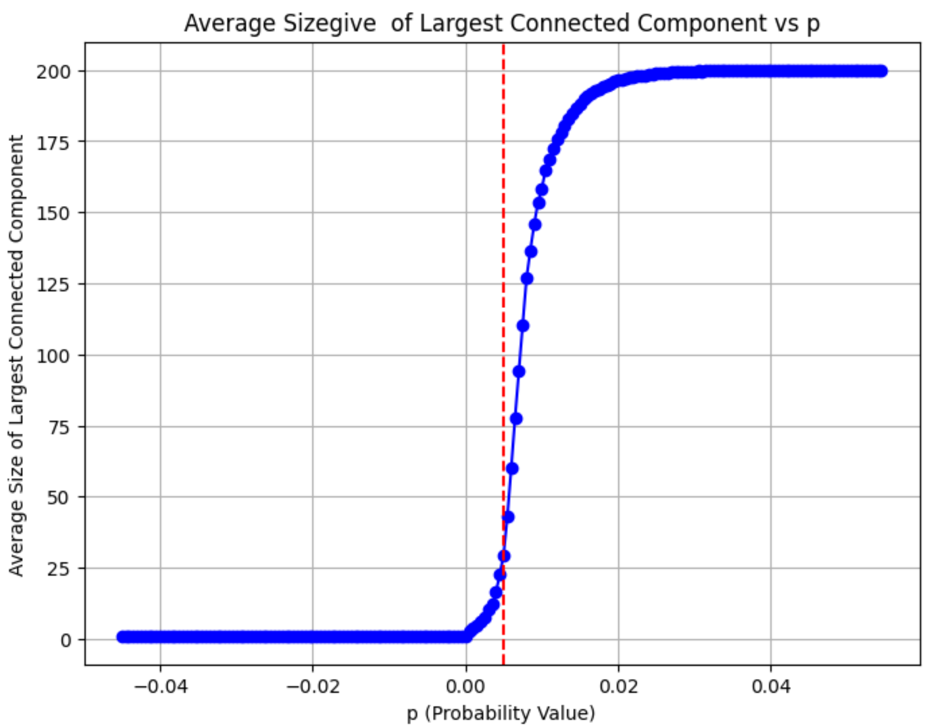

Figure 1: Visualization of random graphs near the threshold $p = \frac{1}{n}$.

Visualization of Random Graphs Around $p = \frac{1}{n}$

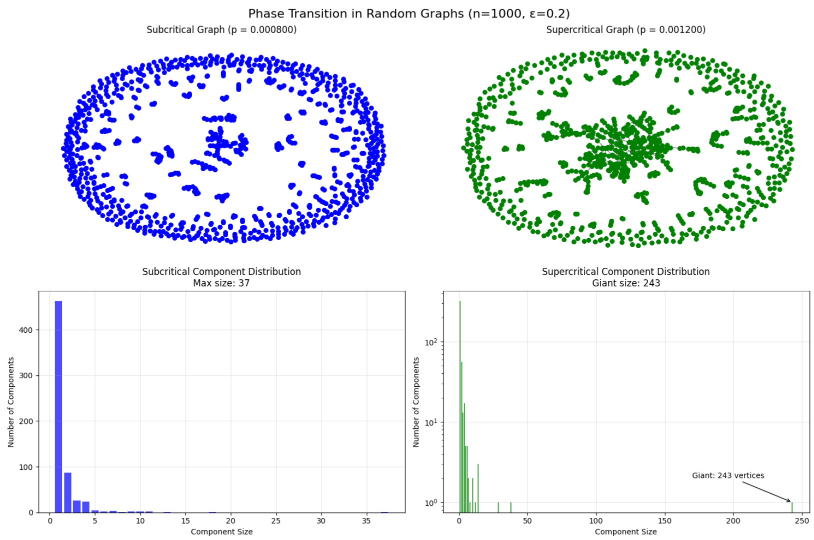

Figure 2: Visualization of random graphs near the threshold $p = \frac{1}{n}$.

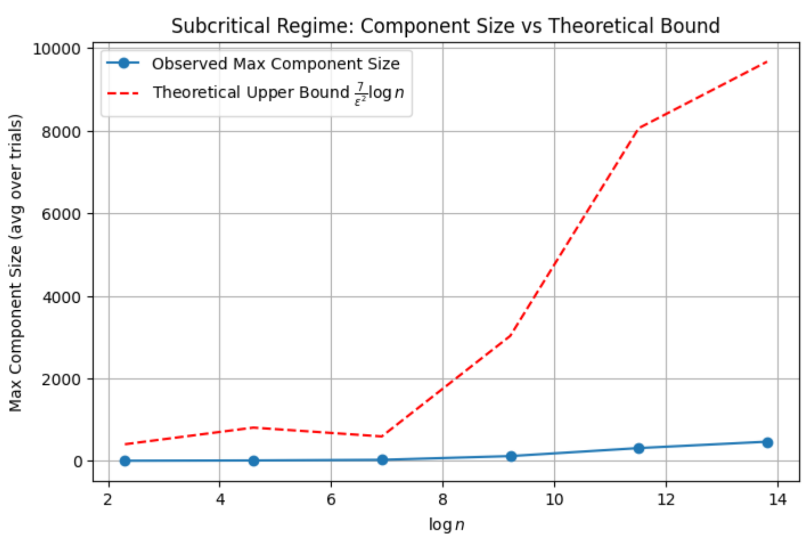

Figure 3: Subcritical Regime: Component Size vs Theoretical Bound. The observed maximum component size matches the theoretical upper bound of $\frac{7}{\epsilon^2} \log n$.

References

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), 1–12.

[2] N. Alon and J. H. Spencer, *The probabilistic method*, 3rd Ed., Wiley, New York, 2008.

[3] M. Bednarska and T. Luczak, *Biased positional games and the phase transition*, Random Struct. Alg. **18** (2001), 141–152.

[4] M. Bednarska and O. Pikhurko, *Biased positional games on matroids*, Eur. J. Combin. **26** (2005), 271–285.

[5] I. Benjamini and O. Schramm, *Percolation beyond* $\mathbb{Z}^d$*, many questions and a few answers*, Electron. Comm. Probab. **1** (1996), 71–82.

[6] I. Ben-Eliezer, M. Krivelevich and B. Sudakov, *Long cycles in subgraphs of (pseudo)random directed graphs*, J. Graph Th. **70** (2012), 284–296.

[7] B. Bollobás, *Random graphs*, 2nd ed., Cambridge University Press, Cambridge, 2001.

[8] P. Erdős and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5** (1960), 17–61.

[9] A. Frieze, M. Krivelevich and R. Martin, *The emergence of a giant component in random subgraphs of pseudo-random graphs*, Random Struct. Alg. **24** (2004), 42–50.

[10] S. Janson, T. Luczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[11] R. Karp, *The transitive closure of random digraph*, Random Struct. Alg. **1** (1990), 73–93.

[12] M. Krivelevich and B. Sudakov, *Pseudorandom graphs.* In: *More sets, graphs and numbers*, E. Györi, G. O. H. Katona and L. Lovász, Eds., Bolyai Society Mathematical Studies Vol. 15, 2006, 199–262.