Component Size in Erdős-Rényi Model

Lin Chen¹

¹Department of Electrical Engineering
Yale University
lin.chen@yale.edu

Theoretical Challenges in Network Science
Outline

1 Properties of DFS

2 Lemma

3 Theorem
Properties of DFS

The DFS algorithm has the following properties.

1. At each round, one vertex moves, either from $T$ to $U$, or from $U$ to $S$.
2. At any stage, it has been revealed already that $G$ has no edges between $S$ and $T$.
3. The set $U$ always spans a path. (indeed, when a vertex $u$ is added to $U$, 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.)
More Properties

1. As long as $T \neq \emptyset$, every positive answer results in moving a vertex from $T$ to $U$. Thus after $t$ queries and if $T \neq \emptyset$, we have
   \[ |S \cup U| \geq \sum_{i=1}^{t} X_i. \]
   This inequality is actually strict. Why?

2. The addition of every vertex to $U$, except the first one in a connected component, is caused by a positive answer, we have at time $t$:
   \[ |U| \leq 1 + \sum_{i=1}^{t} X_i. \]
Let \( \epsilon > 0 \) be a small enough constant. Consider a sequence \( \bar{X} = (X_i)_{i=1}^N \) of i.i.d. Bernoulli random variables with parameter \( p \).

- Let \( p = \frac{1-\epsilon}{n} \) and \( k = \frac{7}{\epsilon^2} \log n \). Then \textbf{whp} there is no interval of length \( kn \), in which at least \( k \) of the random variables \( X_i \) takes value 1.

- Let \( p = \frac{1+\epsilon}{n} \) and \( N_0 = \frac{\epsilon n^2}{2} \). Then \textbf{whp}
  \[
  \left| \sum_{i=1}^{N_0} X_i - \frac{\epsilon (1+\epsilon) n}{2} \right| \leq n^2/3.
  \]
Proof of Property 2

- \( \mathbb{E} \left[ \sum_{i=1}^{N_0} X_i \right] = N_0 p = \frac{\epsilon (1+\epsilon) n}{2} \)

- \( \text{Var} \left[ \sum_{i=1}^{N_0} X_i \right] = N_0 p (1 - p) \leq N_0 p = O(n) \)

Use Chebyshev’s inequality (\( \Pr \left[ |X - \mathbb{E}[X]| \geq k \right] \leq \frac{\text{Var}[X]}{k^2} \)), we have

\[
\Pr \left[ \left| \sum_{i=1}^{N_0} X_i - \frac{\epsilon (1+\epsilon) n}{2} \right| \geq \frac{n^{2/3}}{3} \right] \leq \frac{O(n)}{n^{4/3}} \rightarrow 0,
\]

as \( n \rightarrow \infty \).
Let $\epsilon > 0$ be a small enough constant. Let $G \sim G(n, p)$.

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

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

- Assume that $G$ contains a connected component $C$ with more than $k = \frac{7}{\epsilon^2} \log n$ vertices. Look at the epoch of $C$. Consider the moment inside this epoch when the $(k+1)$-st vertex of $C$ has been found and is about to be moved to $U$. Let $\Delta S = S \cap C$ at that moment. Then $|\Delta S \cup U| = k$. We know that the algorithm got exactly $k$ positive answers to $X_i$’s during this epoch, after the first vertex of $C$ was put into $U$ at the beginning of the epoch.

- At that moment, only pairs of edges touching $\Delta S \cup U$ have been queried. The number is at most $\binom{k}{2} + k(n - k) < kn$. We observe an interval of length at most $kn$ with at least $k$ 1’s. A contradiction.
Assume that the sequence $\bar{X}$ satisfies Property 2 of Lemma. Let $N_0 = \frac{\epsilon n^2}{2}$. Claim that $|S| < \frac{n}{3}$ at time $N_0$. If $|S| \geq \frac{n}{3}$, look at the moment $t$ where $|S| = \frac{n}{3}$ (why such a moment exists?). At that moment $|U| \leq 1 + \sum_{i=1}^{t} X_i < \frac{n}{3}$ by Property 2 of Lemma (why? hint: $\epsilon$ is small). Then $|T| = n - |S| - |U| \geq \frac{n}{3}$. The algorithm has examined all $|S| \cdot |T| \geq \frac{n^2}{9} > N_0$ pairs between $S$ and $T$ and found them to be non-edges. Contradiction.
Claim that after the first $N_0 = \frac{\epsilon n^2}{2}$ queries, the set $U$ contains at least $\frac{\epsilon n^2}{5}$ vertices. Consider time $N_0$. If $|S| < \frac{n}{3}$ and $|U| < \frac{\epsilon^2 n}{5}$, we have $T \neq \emptyset$. This means that the algorithm is still revealing the components of $G$ and each positive answer results in moving a vertex from $T$ to $U$. By Property 2 of Lemma, the number of positive answers at this time is at least $\frac{\epsilon(1+\epsilon)n}{2} - n^2/3$. Hence $|S \cup U| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^2/3$. If $|U| \leq \frac{\epsilon^2 n}{5}$, then $|S| \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^2/3$. All pairs between $S$ and $T$ have been probed (and answered negatively).
However, we have

\[ \frac{\epsilon n^2}{2} = N_0 \geq |S| \left( n - |S| - \frac{\epsilon^2 n}{5} \right) \]

\[ \geq \left( \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3} \right) \left( n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{2} + n^{2/3} \right) \]

\[ = \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{20} - O(\epsilon^3) n^2 > \frac{\epsilon n^2}{2}. \]

A contradiction. Note that we used the fact \(|S| < \frac{n}{3}\) here.