Exercice 1: linear path in Erdős Rényi graphs

Recall that the Depth First Search (DFS) is a graph search algorithm that visits all vertices of a graph \( G = (V, E) \) as follows. It maintains three sets of vertices, letting \( S \) be the set of vertices whose exploration is complete, \( T \) be the set of unvisited vertices, and \( U = V \setminus (S \cup T) \), where the vertices of \( U \) are kept in a stack (the last in, first out data structure). It is also assumed that some order \( \sigma \) on the vertices of \( G \) is fixed. The algorithm starts with \( S = U = \emptyset \) and \( T = V \), and runs until \( U \cup T = \emptyset \).

At each round of the algorithm, if the set \( U \) is non-empty, the algorithm queries \( T \) for neighbors of the last vertex \( v \) that has been added to \( U \), scanning \( T \) according to \( \sigma \). If \( v \) has a neighbor \( u \) in \( T \), the algorithm deletes \( u \) from \( T \) and inserts it into \( U \). If \( v \) does not have a neighbor in \( T \), then \( v \) is popped out of \( U \) and is moved to \( S \). If \( U \) is empty, the algorithm chooses the first vertex of \( T \) according to \( \sigma \), deletes it from \( T \) and pushes it into \( U \). In order to complete the exploration of the graph, whenever the sets \( U \) and \( T \) have both become empty (at this stage the connected component structure of \( G \) has already been revealed), we make the algorithm query all remaining pairs of vertices in \( S = V \), not queried before.

(a) Show that at any stage of the algorithm, it has been revealed already that the graph \( G \) has no edges between the current set \( S \) and the current set \( T \).

(b) Show that the set \( U \) always spans a path.

We will 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. Let \( N = \binom{n}{2} \).

(c) When the DFS algorithm is fed with a sequence of i.i.d. Bernoulli \( p \) random variables \( \bar{X} = (X_i)_{i=1}^{N} \), so that is gets its \( i \)-th query answered positively if \( X_i = 1 \) and answered negatively otherwise, show that the so obtained graph is distributed according to \( G(n, p) \).

Thus, studying the component structure of \( G \) can be reduced to studying the properties of the random sequence \( \bar{X} \).

(d) Show that after \( t \) queries and assuming that \( T \) never emptied, we have \( |S \cup U| \geq \sum_{i=1}^{t} X_i \) and \( |U| \leq 1 + \sum_{i=1}^{t} X_i \).

The probabilistic part of our argument is provided by the following result. Let \( \epsilon > 0 \) be a small enough constant. Consider the sequence \( \bar{X} = (X_i)_{i=1}^{N} \) of i.i.d. Bernoulli random variables with parameter \( p = \frac{1+\epsilon}{n} \).

(e) Let \( N_0 = \frac{c n^2}{2} \). Show that with high probability \( \left| \sum_{i=1}^{N_0} X_i - \frac{c (1+\epsilon)n}{2} \right| \leq n^{2/3} \).
We run the DFS algorithm on $G \sim G(n, p)$, and assume that the sequence $\bar{X} = (X_i)_{i=1}^N$ of random variables, defining the random graph $G \sim G(n, p)$ and guiding the DFS algorithm, satisfies the above property.

We will show that after the first $N_0 = \frac{ep^2}{8}$ queries of the DFS algorithm, the set $U$ contains at least $\frac{e^2n}{5}$ vertices, which implies that with high probability $G$ contains a path of length at least $e^2n/5$.

(f) Show that if at time $t \leq N_0$, $|S| = n/3$, then $|T| \geq n/3$. Show that the algorithm examined at least $n^2/9$ pairs which leads to a contradiction. Conclude that $|S| < n/3$ at time $N_0$.

(g) Assume that $|U| < \frac{e^2n}{5}$. Show that $|S \cup U| \geq \frac{c(1+e)n}{2} - n^2/3$ and conclude similarly as above.

**Exercice 2: counting isolated subgraphs**

We wish to count the isolated copies of a connected graph $H$ in $G(n, p)$, i.e. the connected components of $G(n, p)$ which are isomorphic to $H$. Let $N_H$ count the isolated copies of $H$. Let $v_H$ and $e_H$ be the number of vertices and edges of $H$.

(a) Show that $\mathbb{E}[N_H] = O(v_Hp^{e_H}e^{-v_Gnp})$. Show that with high probability $N_H = 0$ if $e_H > v_H$ or if $e_H = v_H$ and $p \ll 1/n$ or $p \gg 1/n$.

We now consider the case $np \to c > 0$. We first consider the case where $H$ is a triangle and denote by $N_3$ the number of isolated triangles: $N_3 = \sum_G I_G$ where the sum is over all triangles $G$ of $K_n$ and $I_G$ is the indicator that $G$ is an isolated triangle.

(b) Compute $\mathbb{E}[I_G]$ and $\mathbb{E}[N_3]$. (c) For $G' = \{u', v', w'\}$ and $G = \{u, v, w\}$, define the indicator $J_{G', G}$ that $G$ is an isolated subgraph of $G(n, p)$ where all edges incident with $u', v'$ and $w'$ have been removed and the edges of the triangle $G'$ have been added. Show that the random variables $I_G$ and $J_{G', G}$ satisfy the condition of the Stein-Chen method.

(d) Conclude that $N_3$ converges in distribution to a Poisson distribution.

(e) Extend previous analysis to unicyclic connected components.

**Exercice 3: improved second moment method**

(a) Using Tchebichev’s inequality show that

$$P(X = 0) \leq \frac{\text{Var}(X)}{\mathbb{E}[X]^2}.$$ 

(b) Using the Cauchy-Schwarz inequality applied to $X = X_1(X \neq 0)$ show that

$$P(X = 0) \leq \frac{\text{Var}(X)}{(\mathbb{E}[X])^2 + \text{Var}(X)}.$$