EE447 Homework4
Multi-hop Neighbours in ER Networks

付昊源 517021910753

May 22, 2020

Problem Statement

In an ER network, we can consider a graph $G = \mathcal{G}(n, p)$ as a random graph, i.e. each edge in G has the probability p of being present and $(1-p)$ of being absent, independently of the other edges. Besides, we further sample the edges in G for two times to construct two new graphs G_1 and G_2. In each of the new graph, each edge of G has the probability s of being preserved and naturally $(1-s)$ of being deleted, and G_1 and G_2 are sampled independently with each other. Thus, we can formulate two new graphs as $G_1 = \mathcal{G}(n, ps)$, $G_2 = \mathcal{G}(n, ps)$. Note that G_1 and G_2 are related with G, and there cannot be any new edge in G_1 or G_2. All these relationships can be formulated as

$$\begin{aligned} E(G) &\supseteq E(G_1) \cup E(G_2) \\ V(G) &= V(G_1) = V(G_2) \end{aligned} \tag{1}$$

where $E()$ means the edge set and $V()$ means the vertices set.

We define that two different nodes u and v in network G are 2-hop neighbors if and only if their shortest distance in G is exactly 2. Provided that $0 < s \leq 1$ and $p \ll 1 \ll np$, we are assigned to prove the following statements:

1. The summation of expected number of 2-hop neighbors for all nodes in network G can be approximated by n^3p^2.
2. The probability of nodes u, v being 2-hop neighbors in both sampled networks G_1 and G_2 if they are 2-hop neighbors in network G can be approximated by s^4.

Proof of statement 1

Suppose the set of vertices is $\{v_1, v_2, \cdots, v_n\}$ and we use P_{ij} to denote the probability of an edge from v_i to v_j exists. Clearly, $P_{ij} = P_{ji} = p$. We further use V_i^1 and V_i^2 to denote the set of 1-hop neighbors and the set of 2-hop neighbors for vertex v_i, respectively.

For an arbitrary vertex v_i in G, the expected number of items in V_i^1 is

$$E(|V_i^1|) = \sum_{j \neq i} 1 \cdot P_i j = (n-1) \cdot p \approx np \tag{2}$$

since $np \gg p$.

We first calculate the number of items in V_i^1 for the following two properties. One is that, if a vertex v_j is the 1-hop neighbor of v_i, then it cannot be 2-hop neighbor of v_i. The other is, if v_j is the 2-hop neighbor of v_i, then there must exist v_k, who is 1-hop neighbor of both v_i and v_j. These two properties both relate to 1-hop neighbors, so we can further calculate the expected number of 2-hop neighbors for a given vertex v_i based on $E(|V_i^1|)$:

$$E(|V_i^2|) = \sum_{j \neq i, j \notin V_i^1} \sum_{k \in V_i^1} 1 \cdot P_{jk} = (n - np - 1)np \cdot p \approx n^2p^2 \tag{3}$$

Thus, for all vertices in network G, the summation of expected number of 2-hop neighbors can be approximated by $n \cdot n^2p^2 = n^3p^2$

Proof of statement 2

According to the setting in statement 2, node u and v have already been 2-hop neighbors in graph G, which is the given condition. We can construct a set $V_{uv} \subset V(G)$, which contains all nodes satisfying

$$\forall v_i \in V_{uv}, \quad \{u, v_i\} \in E(G) \wedge \{v_i, v\} \in E(G) \tag{4}$$

If node u and v are still 2-hop neighbors in sampled graph G_1 and G_2, the intermediate node connecting u and v must also exist in V_{uv}. Without loss of generality, assume

$$\begin{aligned} v_{i_1} \in V_{uv}, \quad \{u, v_{i_1}\} \in E(G_1) \wedge \{v_{i_1}, v\} \in E(G_1) \\ v_{i_2} \in V_{uv}, \quad \{u, v_{i_2}\} \in E(G_2) \wedge \{v_{i_2}, v\} \in E(G_2) \end{aligned} \tag{5}$$

In this way, we only care about the appearance or absence of the four edges in equation 5. If u and v are 2-hop neighbors both in G_1 and G_2, the four edges above must all exist. Due to the independence of sampling, the probability is $s \cdot s \cdot s \cdot s = s^4$.