

# Internet Analytics (COM-308)

## Homework Set 1 - Solutions

### Exercise 1

We had studied the  $G(n, p)$  random graph model. Here are two ways to compute its expected node degree. State which one is correct, and argue what is wrong with the other.

1. (i) Each node has  $n - 1$  possible neighbours, with an edge to each with probability  $p$ . This gives a total expected node degree of  $(n - 1)p$ .

(ii) In the component discovery process, each node is discovered through an incoming edge, and has  $(n - 2)$  possible outgoing edges. Therefore, the expected node degree is  $1 + (n - 2)p$ .

(i) is correct. (ii) only takes into account nodes with degree at least one. Therefore, this is essentially the conditional expectation  $E[D|D > 0]$ . Also the first node does not have an incoming edge thus having different expected degree.

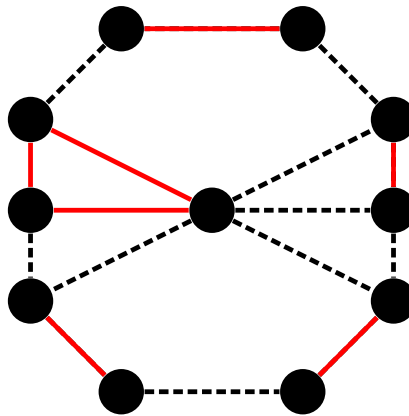
2. In a complete graph  $K_n$  of order  $n$  we delete each edge with probability  $q$  independently. Compute the clustering coefficient of this network as a function of  $n$  and  $q$ .

This problem is equivalent to the case where each edge exists independently with probability  $1 - q$ , i.e., graph  $K_n$  after edge deletion process is equivalent to a  $G(n, 1 - q)$  graph. Therefore, the clustering coefficient is  $1 - q$ .

### Exercise 2

In the following graph, all nodes must satisfy the strong triadic closure (STC) property. To ensure this, you need to mark some edges as strong, while all others are weak. Identify the largest possible set of strong edges such that the STC is not violated.

Strong edges are represented by red color and all the other edges (dotted edges) are weak.



### Exercise 3

In the Watts-Strogatz small-world network model, we start with an  $n$ -cycle  $C_n$ , whose nodes are labeled  $1, \dots, n$ . Then, we add every edge  $(u, v)$  such that  $\text{dist}(u - v) \leq k$  (i.e., every pair of nodes at distance less or equal to  $k$  on the cycle).

Compute the clustering coefficient of this network as a function of  $n$  and  $k$ .

*Would this clustering coefficient be higher than that of a  $G(n, p)$  network of same order and size (number of nodes and edges)?*

By symmetry, the clustering coefficient of the graph is identical to that of a node. Without loss of generality, we look at node 0, which has neighbors  $-k, \dots, +k$  (we assume  $n \gg k$ ).

Consider a node  $0 < i \leq k$ , and let  $n_i$  denote the number of neighbors of  $i$  within the window  $[-k, k] \setminus 0$ . We find  $n_i = (k - i) + (k - 1) = 2k - i - 1$ , where the first term accounts for neighbors to the left of  $i$ , and the second for those to the right. Everything is symmetric for  $i > 0$ . Note that we count every edge twice in this calculation. Summing over all  $i$ , we obtain that the number of edges among neighbors (triangles) is  $(3/2)k(k - 1)$ .

$$c_G = c_0 = \frac{3k(k - 1)}{2k(2k - 1)}. \quad (1)$$

The asymptotic limit with  $k \rightarrow \infty$  for the clustering coefficient is  $3/4$ .