Internet Analytics (COM-308)
Problem Set 1

Problem 1

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.

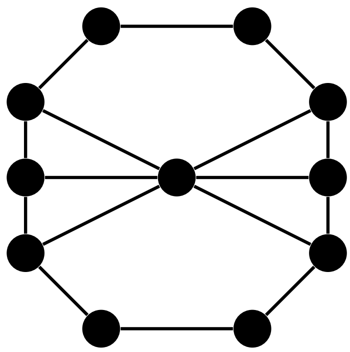
(i) Each node has $n-1$ possible neighbours, with an edge to each with probability p. This gives a total expected node degree of $(n-1)p$.

(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 $1+(n-2)p$.

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

Problem 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.



Problem 3

We saw in class that a *local bridge* is an edge (u, v) without a short alternative path, i.e., if we removed (u, v) from the graph, $d(u, v) \geq 3$. Compute the probability that some arbitrary edge (u, v) in the random graph $(V, E) = G(n, p)$ is a local bridge.

Problem 4

In the Watts-Strogatz small-world network model, we start with an n-cycle C_n, whose nodes are labeled $1, \ldots, n$. Then, we add every edge (u, v) such that $dist(u, v) \leq k$ (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)?