Internet Analytics (COM-308)
Problem Set 1 - Solutions

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 $1 + (n-2)p$.

(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.

---

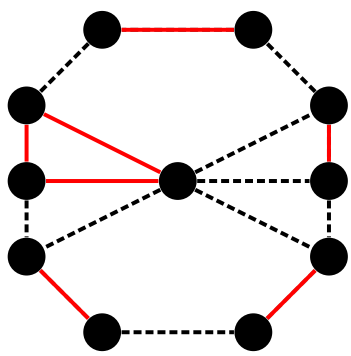
2. In a complete graph K_n of order n we delete each edge with probability q. Compute the clustering coefficient of this network as a function of n and q.

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 $1 - q$.

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.

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



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.

An edge (u, v) is a local bridge if and only if u and v do not have any common neighbors. Call X_w the event that w is a common neighbor of u and v, i.e., that $(u, w) \in E$ and $(v, w) \in E$, where E is the edge set of $G(n, p)$. Note that for any pair of distinct vertices w_1 and w_2, X_{w_1} and X_{w_2} are independent, because the two events depend on disjoint pairs of edges, and all edges are independent in the $G(n, p)$ model. Then

$$\begin{aligned} P(u \text{ and } v \text{ have no common neighbor}) &= P(\cap_{w \in V \setminus (u \cup v)} \overline{X_w}) = & (1) \\ &= \Pi_{w \in V \setminus (u \cup v)} P(\overline{X_w}) = & (2) \\ &= (1 - p^2)^{n-2}. & (3) \end{aligned}$$

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)?

By symmetry, the clustering coefficient of the graph is identical to that of every node. Without loss of generality, we look at some node x, which has neighbors $x - k, \ldots, x - 1, x + 1, \ldots, x + k($ mod $n)$ (we assume $n \gg k$).

Consider a node $x + i$ to the "right" of x, ie, $x < x + i \leq x + k$, and let n_i denote the number of neighbors of $x + i$ within the window $[x - k, x + k] \setminus x$. 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 to the "left" of x. 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 $(3/2)k(k-1)$.

$$c_G = c_0 = \frac{3k(k-1)}{2k(2k-1)}. \quad (4)$$

Note that this quantity is independent of n.

The equivalent density p, which is the expected clustering coefficient of the $G(n, p)$, is $p = 2k/(n-1)$. This is asymptotically vanishingly small compared to the clustering coefficient of the small world model.