CS4352 Assignment 1

Akhilesh Raj

September 2022

Solution 1:

(a)

Clustering coefficient C:

$$C = \frac{\text{number of links between neighbors}}{\text{maximum possible number of links between neighbors}} \tag{1}$$

For an Erdos-Reyni graph $G_{n,p}$ with n nodes, $N = \frac{n(n-1)}{2}$ is the maximum number of possible pair of nodes / edges that can exist. We choose m edges among the N available edges with a probability p. Therefore the number of links that exist are $m = pN = p\frac{n(n-1)}{2}$. Therefore the clustering coefficient for the given $G(n, p)$ is:

$$\frac{p\frac{n(n-1)}{2}}{\frac{n(n-1)}{2}} = p \tag{2}$$

(b)

Given that the average degree of a node in a graph is given by:

$$< k >= p(n-1) \approx pn. \tag{3}$$

From the previous solution 2 we know that the clustering coefficient is $C = p$ is has to vary inversely with respect to the number of nodes n in-order to make the average degree $< k >$ fixed.

$$\frac{< k >}{n} = p, \tag{4}$$

Taking limits on both the sides we get:

$$\lim_{n\to\infty} \frac{< k >}{n} = 0 = \lim_{p\to 0} p. \tag{5}$$

Since $C = p$ we can say that C will approach 0 as n approaches ∞.

Solution 2:

(a)

Number of triangles = ΔS_3, where Δ = probability for any particular set of three nodes to form a triangle and S_3 = number of possible sets of three nodes in a graph of n nodes. For an ER graph $G_{n,p}$ the value of S_3 can be determined as number of combinations in choosing 3 nodes from the set of n nodes which is given by ${}^nC_3$ while Δ is given by p^3 since it is the probability of three edges to exist among three nodes. Therefore the number of triangles in an E-R graph is given by

$$\text{number of triangles} = p^3\ {}^nC_3 = p^3\frac{n(n-1)(n-2)}{6} \tag{6}$$

(b)

When the value of $n \to \infty$ we can write

$$n \approx (n-1) \approx (n-2) \tag{7}$$

Therefore by substituting (7) in (6) we get

$$\begin{aligned} \text{number of triangles} = p^3\ {}^nC_3 &= p^3\frac{n(n-1)(n-2)}{6} \\ &= \frac{p^3n^3}{6} = \frac{(np)^3}{6} = \frac{<k>}{6} \end{aligned} \tag{8}$$

Solution 3:

Expected number of common neighbors for a configuration model is given by:

$$n_{ij}^{CN} = \sum_l \Big(\frac{k_i k_l}{2m}\Big)\Big(\frac{k_j(k_l-1)}{2m}\Big) \tag{9}$$

The above equation (9) can be rewritten as:

$$\begin{aligned} n_{ij}^{CN} =& \Big(\frac{k_i k_j}{2m}\Big)\sum_l \Big(\frac{k_l(k_l-1)}{2m}\Big) \\ =& p_{ij}\sum_l \Big(\frac{k_l(k_l-1)}{2m}\Big) \\ =& p_{ij}\sum_l \Big(\frac{k_l^2-k_l}{2m}\Big) \\ =& \frac{p_{ij}}{2m}\sum_l k_l^2 - \sum_l k_l \end{aligned} \tag{10}$$

Given $< k^2 >= \frac{1}{n}\sum_i k_i^2$ and $< k >= \frac{1}{n}\sum_i k_i$. Substituting these in the equation (10) we get:

$$\begin{aligned} =&\frac{p_{ij}}{2m}\sum_l k_l^2 - \sum_l k_l \\ =&\frac{p_{ij}}{2m}(n < k^2 > -n < k >) \\ =&\frac{p_{ij}n}{2m}(< k^2 > - < k >) \end{aligned} \tag{11}$$

Now we know that m is the expected number of edges in the random graph when p is the probability of edge sampling from the total number of nodes n. Therefore we know that:

$$\begin{gathered} m =p\frac{n(n-1)}{2} \\ \text{and} \\ < k > \approx pn. \end{gathered} \tag{12}$$

Combining the above two equations we can easily write:

$$2m = pn^2 =< k > n \tag{13}$$

Upon substituting (13) in (11) we get:

$$n_{ij}^{CN} = \frac{p_{ij}n}{2m}(< k^2 > - < k >) =p_{ij}\frac{(< k^2 > - < k >)}{< k >} \tag{14}$$

Solution 4:

(a)

The degree distribution of a network with degree sequence $D = k_1, k_2, ..., k_n$ such that $k_1 = k_2 = ... = k_n$ is as follows:

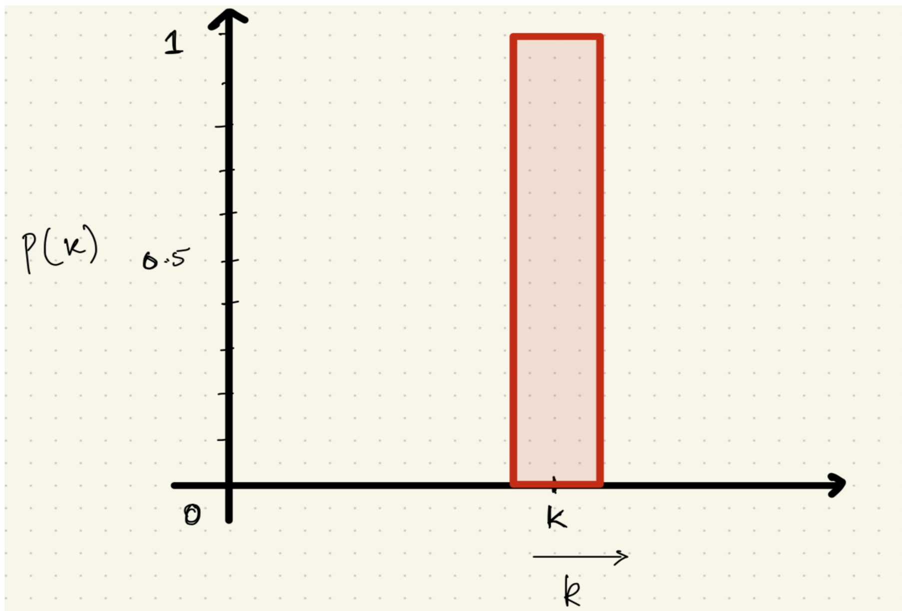

Figure 1: Degree distribution (P(k) vs k, where P(k) depicts the fraction of nodes and k depicts the degree.

(b)

When all node have degree = 1, then there is a possibility of one connection per node. As given in the note there cannot be self loops since it will result in some nodes having a degree = 2. Therefore the only possible way of arrangement is that, each node getting connected to one and only one another node who is not connected to anyone else. Therefore the network will consist of a number of disjoint sets which is equal to the half of total number of nodes within the network. Therefore a total of $n/2$ disjoint sets are possible. The analysis is simple. Consider a network consisting of n nodes and then divide it into two sets consisting of $n/2$ nodes each. Here $n/2$ will be an integer since n is an even number. Otherwise as per the note says the resulting one node will have either no degree (0) or a degree of 2. If you connect one node from a set to a node in the other set, what you get is the non trivial connected component. Like that we can do a one to one mapping of each nodes in the sets and we will end up in getting $n/2$ connected components.

Solution 7

(a)

Given the page rank matrix:

$$P^{"} = \alpha P^{'} + (1-\alpha)\frac{ee^T}{n} \tag{15}$$

where $e \in \mathbb{R}^{n\times 1}$ is a vector consisting on ones. Therefore ee^T is a square matrix with all elements as ones. When $\alpha = 0$ we have:

$$P^{"} = \frac{ee^T}{n} \tag{16}$$

Now the page rank matrix will end up with $P^{"} = \frac{1}{n}ee^T$ which mean each page will have a probability of $\frac{1}{n}$ to be chosen. Therefore the pages are likely to appear randomly when each time the random surfer starts surfing.

The Perron-Frobenius Theorem requires the given $P^{"}$ matrix to be:

- Stochastic (i.e., non-negative and rows sum to one)
- Irreducible (i.e., strongly connected)
- Aperiodic (if the greatest common divisor of the lengths of cycles is one and is guaranteed when self loops are present.)

In the given P" all the three conditions are satisfied since there is a positive value $(\frac{1}{n})$ at all the places of the $n \times n$ matrix which means that the rows/columns sum to one, strongly connected (a single connected component) and self loops are present. Therefore none of the conditions for the Perron-Frobenius Theorem gets violated when $\alpha = 0$.

(b)

When $\alpha = 1$, the page rank matrix will be:

$$P^{"} = P' \tag{17}$$

where,

$$P^{'} = D^{-1}A + \frac{se^T}{n} \tag{18}$$

where D is the matrix with $D_{ii} = diag(d_i^{out})$, s is absorbing nodes indicator vector and e is vector of 1s. The $P^{'}$ term having links from all the s will never be able to converge since there are no damping terms.

Checking the Perron-Frobenius Theorem we can say $P^{'}$ is:

- Stochastic (Since the rows sum to one having scaled by the degree vector.)

- Not irreducible (Since there is not guarantee of a non zero value at all the places in $P^{'}$. It may end up having some elements as 0 upon which a smaller connected component may exist.)
- Not aperiodic (Since there is no guarantee that non zero elements exists along the diagonal of $P^{'}$ indicating at least one self loop.)