Exercises for Statistical analysis of network data-Sheet 13

1. We are studying customers $\{1, \ldots, n\}$ who buy products $\{1, \ldots, m\}$. An edge variable takes the value one if customers buy a particular product, this forming the data matrix X which is $n \times m$.

   (a) Assume that the expectation of X_{ij} is a product of a term capturing the customer activity α_i and a term capturing product popularity β_j. Write down $\mathrm{E}X_{ij} = P_{ij}$.

   (b) Calculate the in and out degrees of the data matrix X. Determine its distribution, assuming every entry of the data matrix is independent.

   (c) Describe how to estimate α and β using the degrees.

   (d) Calculate the mean and variance of the estimated parameters.

   (e) Describe how to test that $\alpha_i = \alpha$ a constant.

2. Referring back to question 1, the proposed model captures an effect due to each individual, combined with an effect due to each product. What could be clusters in such data ?

   Assume that there are rows r_i belonging to cluster i and columns c_i that belong to the same cluster. Describe how we would estimate the cluster expected value θ_{r_i,c_j}, i.e. the probability of i and j being connected while in their respective cluster, for $i \in [n]$ and $j \in [m]$. Is it more important to estimate cluster membership or cluster probabilities correctly?

3. Define the adjacency tensor to be (A_{rmj}). A_{rmj} is unity if interactions between 3 nodes r, m and j are present, and is zero otherwise. We assume that the interaction has probability $\rho > 0$ of occurring and that the hyperedge variables are independent. We define

$$k_{2,0}(r) = \sum_{m<j} A_{rmj}$$
$$k_{2,1}(r, m) = \sum_{j} A_{rmj}.$$

   Calculate the expectation and variance of $k_{2,0}(r)$ and $k_{2,1}(r, m)$.

4. We study interactions between Jack, Joe, Jill, James and Jonathan. We know that Jack, Joe and Jill study together, and also that Jill, James and Jonathan study together. Enumerate the adjacency tensor for the study interactions of the five.

5. We shall study the adjacency matrix when A_{rmj} is independent. We assume

$$\mathrm{E}\{A_{rmj}\} = \pi_r \pi_m \pi_j.$$

   Calculate the expectation and variance of $k_{2,0}(r)$ and $k_{2,1}(r, m)$.