Exercises for Statistical analysis of network data-Sheet 12

1. Assume you observe a network on six nodes $V = \{1, \ldots, 6\}$. Assume you have observed the edges (12), (16), (23), (24), (36), (45) and (56). Assume you have also observed the edge variables $A_{13} = A_{14} = A_{15} = A_{34} = A_{35} = 0$, as well as $A_{25} = A_{26} = 0$. There are therefore $\binom{6}{2} = 15$ edge variables and $15 - 14 = 1$ missing variables.

   (a) Compute the shortest path distance between all nodes, giving the answer in a table.

   (b) Compute the neighbourhood score, giving the answer in a table.

   (c) Compute the Jaccard coefficient, giving the answer in a table.

   (d) Compute the Liben-Nowell score, giving the answer in a table.

2. For the network in the previous problem explore for the four methods how different values of s^* impact the prediction of edges. What considerations would go into the choice of s^* in a social network setting? For your choice of s^* predict the missing edge variables, and critique your prediction.

3. Assume you have taken as $\mathbf{z}$ the observed degrees from i and j from the observed edges. Describe how to use a logistic regression classifier to predict the missing edges in problem 1 of this sheet.

4. Assume you have a network on four nodes $\{1, 2, 3, 4\}$ assume you have observed edges (12), (13). and non-edges (23) as well as (14). How would one predict edges involving node 4 ?