Assignment 6: Due Thursday 2nd May at 5pm

Late assignments will not be accepted except by prior arrangement (for a good reason)

Please include your student number in your handed up work, as Canvas doesn't give this to me automatically.

1. The *configuration model* is a random graph model in which the *degree sequence* of a graph is replicated exactly. However, note that in its simplest form it generates a multi-graph.

   Assume that each node $i = 1, \ldots, n$ has degree k_i then the configuration model can be implemented by writing a sequence in which the i is repeated k_i times. For example, if $k_1 = 3$ and $k_2 = 1$ and $k_3 = 2$ we would write

   $$1, 1, 1, 2, 3, 3, ...$$

   Now pick two nodes uniformly at random from this list, without replication, and form a link between the two. Alternative, pick two nodes, link them, and then delete them from the list.

   You will find a new dataset `graph_A6_Q1_a1010101.graphml`, at `https://roughan.info/notes/Network_Modelling/10data.html`

   The file is in GraphML format. This is a standard graph format and is readable by many tools and libraries (you may use a package, e.g., igraph to read the graph). It can also be converted to other formats by various tools.

   The task this week is to analyse this, and develop a model of the data. Use the tools you have been taught, to

   (a) Generate a set of 100 configuration graphs that match the dataset.

   (b) Calculate the average degree, number of connected components, and clustering coefficient of your graphs and compare them to the original data.

   (c) Now adopt a sampling strategy where you sample only 1/4 of your graph (in the sense appropriate for your strategy) and recalculate these metrics.

   (d) Write a short report comparing the model to the data. Describe your work concisely. Focus on relevant details. Reporting irrelevant information will cost marks.

   [8 marks]

2. (a) Given an undirected weighted graph $G = (N, E)$ with positive edge weights w_e for $e \in E$, demonstrate that the distances d_{ij} created by

   - finding shortest *weighted* paths between nodes i and j;
   - adding up the weights along the paths,

   is a distance metric.

   (b) Show conversely that if the the distances are created by

   - finding shortest *hop* paths between nodes i and j;
   - adding up the weights along the paths,

   then this is not a metric.

   [2 marks]