Network Science 2022/23
Coursework

Matthew Johnson

Submit by 16 February 2023 at 2pm

Attempt all questions. Partial credit for incomplete solutions may be given.

Submit a single zip file that contains four folders `q1`, `q2`, `q3`, `q4`. Each should contain a pdf document containing the write-up of your answer to the corresponding question, and also any code used in obtaining your answers.

Plagiarism

Please remember that you should not share your work or make it available where others can find it as this can facilitate plagiarism and you can be penalised. This requirement applies until the assessment process is completed which does not happen until the exam board meets in June 2023. If you base any code you write on sources other than the materials provided within the module, you must clearly cite the source.

---

1. The networks in this question are directed.

   (a) From the course materials, obtain "citation_graph" and the PA graph "EX_GRAPH_PA1" used to model it. Plot the normalized distribution of the local clustering coefficients for each graph where each value is rounded to two decimal places. **[8 marks]**

   (b) For a PA graph on 10,000 nodes with $m = m_0 = 20$, label the nodes from 0 to 9,999 corresponding to the order in which they are added to the graph. For nodes 0, 30, 60 and 90, plot, on the same axes, their in-degree against the number of nodes added to the graph. (That is, how do their in-degrees increase as further nodes are added?). Comment on what you find with reference to a real network that is believed to have the preferential attachment property. **[10 marks]**

   (c) Consider the following network model that constructs a *C graph* with parameters n, m and p_1, p_2, p_3 and p_4 where $p_1 + p_2 + p_3 + p_4 = 1$.
   - Construct a complete graph on m nodes
   - Repeat the following $n - m$ times:
     - with probability p_1, add a new node v joined by m edges to existing nodes chosen with probability proportional to their degree;

- with probability p_2, add m random edges between existing nodes where the endpoints of the new edges are chosen with probability proportional to their degree;
- with probability p_3, delete an existing node chosen uniformly at random;
- with probability p_4, delete m existing edges chosen uniformly at random.

Investigate C graphs by generating instances and looking at their properties. You can choose $n = 2000$, $m = 20$ and vary the probabilities. Comment on your findings. Do C graphs provide an intuitive model for any real networks? **[16 marks]**

2. Consider the centrality measure C that is defined as follows. For each node v in a network, $C(v)$ is equal to the degree of v multiplied by one minus the clustering coefficient of v.

   (a) Is C a reasonable proxy for betweenness centrality? Provide an argument and test it on a small number of networks.

   (b) Why might C be preferred to betweenness centrality.

   [13 marks]

3. Consider the following undirected network model. To create a network N:

   (a) Begin with an initial network N_0.

   (b) Add additional nodes. Each new node w selects an existing node u uniformly at random and then uniformly at random chooses m neighbours of u. An edge is added from w to each of the m neighbours

   For each of the following cases, analyse the networks created by this model.

   (a) $N_0 = K_2$ and $m = 1$.

   (b) $N_0 = K_3$ and $m = 2$

   [18 marks]

4. Let N be an undirected network and C be a centrality measure. The *relatedness index* of N with respect to C, denoted $r(N, C)$ is a measure of the extent to which the centrality values of adjacent nodes in N are correlated. Formally it is the value of the Pearson's Product-Moment Correlation Coefficient over $C(u)$ and $C(v)$ for all edges uv in N. Thus it has a value between -1 and 1.

   (a) Calculate $r(N, C)$ where N is the karate network defined in `zachary.txt` and C is degree centrality and closeness centrality **[5 marks]**

   (b) Investigate the value of the relatedness index for degree centrality and closeness centrality for two of the network models studied in the course and a range of real world networks (chosen from those available on Ultra or otherwise). Comment on your findings. **[30 marks]**