COMP0123 Complex Networks and Web

Instructions for Coursework 1 (CW1)

1. Marking.

CW1 accounts for 30% of the final total mark.
Your coursework will be graded as a mark between 0 and 100.

2. Coursework Instruction

- This is an **individual** coursework. Students should work independently.
- Data source:
  - Academic Social Network Data at https://www.aminer.cn/aminernetwork
- **Task 0 -** Preparation of network data (0 mark)
  - Among the millions of authors, firstly you extract all authors whose research interests are related to 'social network', including those with interests in 'Social network', 'social networks', 'human social network' etc.
  - Secondly, you extract all links (co-authorships) among the above authors. You should discard any authors with degree 0 (that have no link with any of the other authors) and any repeated links, so that you create a simple graph.
  - Thirdly, you extract the giant component in the above simple graph. For your information, the giant component should contain 2,068 unique nodes and 5,163 unique links. Note that link a-b is same as link b-a, which is a single unique link, not two links.
  - In this coursework, you only work on this giant component. For convenience, in the following, we call this giant component "**the author network**".
  - Generate a random network and a BA network with the same numbers of nodes and links. Please make sure the random network is a connected graph. The BA network should grow from a small random graph, say, with 20 nodes and 30 links, and the small random graph should be a connected graph. You should figure out the parameter values of n and m. Note you can use a different m value in the last x steps.
  - Note: You do NOT need to include this Task 0 in your coursework. This task is not marked.
  - Most of the following tasks will be conducted on the 3 networks, i.e. the author network, the random network and the BA network.
  - In each task, you should calculate properties and show them in a table. That is, each task has a table. When plotting a property, draw curves of the 3 networks in the same plot for comparison.
- **Task 1 - (15 marks)**
  - Calculate the average node degree and the maximum node degree of the 3 networks.
  - Plot their degree distribution P(k) on linear-linear scale and log-log scale, respectively.
  - Estimate the power-law exponent of the degree distribution P(k) of the author network only.
    - You can fit a curve by using the function polyfit from the numpy library.

        - Ideally, you can do the fitting on CCDF (the complementary cumulative distribution function) on log-log scale.
    - Briefly discuss your results, e.g., difference of the networks.
- **Task 2 - (15 marks)**
    - Calculate and plot the nearest neighbour's average degree knn as a function of node degree k, on log-log scale.
    - Calculate the assortative coefficient of the networks.
    - Briefly discuss your results.
- **Task 3 - (15 marks)**
    - Calculate the diameter and the average shortest path length of the network.
    - Calculate and plot the average node betweenness of k-degree nodes as a function of node degree k, where node betweenness is normalised, on log-log scale.
    - Briefly discuss your results.
- **Task 4 - (15 marks)**
    - Calculate and plot the rich-club coefficient as a function of node rank on log-log scale.
    - Calculate and plot the rich-club coefficient as a function of node degree on log-log scale..
    - Briefly discuss your result.
- **Task 5 - (15 marks)**
    - Obtain the community structure (with the largest modularity value) of the 3 networks.
    - Give the number of communities and the size (i.e. number of nodes) of the top 3 largest communities in each network.
    - Visualise the 3 networks. In each network, show every community with a different colour.
    - Briefly discuss your result.
- **Task 6 - (25 marks)**
    - Randomly rewire the 3 networks while preserving the degree distribution; and obtain the maximal random case of each network.
    - For the 3 randomised networks, plot their degree distribution.
    - For each of the randomised networks, calculate the average clustering coefficient, the assortative coefficient, the size of the giant component, and the average shortest path length in the giant component. Show these results and compare with those of the 3 original networks in a table.
    - Briefly discuss your result.

3. Methods

- In Task 0, you may need to write your own code to extract the simple graph from the source file(s).
- You can use the network analysis tools Gephi and/or NetworkX (Python) to obtain the giant component and calculate most of the properties.
- You may use any plotting software to create plots. Each plot should be clearly labelled and with a legend.
- Note: please round all property values to 2 digits after the decimal point.
- Coursework report format:
    - Cover page is optional.

- Copy and paste the text of tasks 1-6 in your report, and then insert your results below each task.
- Limit your discussion in each Task to 100 words.
- At the end of your coursework, please list any tools, packages, libraries, and the programming language that you have used for each task.

4. Online Submission

- You should submit your coursework as a single **PDF** file.
  - Do not include your code. Other file format or handwriting will not be marked.
- You can submit your report many times before the deadline. Only the last submission will be recorded in the system.
- Please be aware of UCL's policy on late submission penalties. Please note the mark shown on Moodle will be your original mark BEFORE any penalties applied. Your final mark (after any penalties applied) will be shown in Portico. Please do submit a version before the deadline in case of network jam near the deadline.