|  |
| --- |
|  |
| Descriptive details of Assignment:   * This is an individual work. * Justify every answer with all relevant details. Include and mention explicitly all methods, software you have used and references when appropriate, including from every dataset you use. Failure to do so will lead to the loss of marks. * Any source code you have used should be presented ***only on appendices*** (not in the middle of your main solution). You should clearly indicate where it was used. * Source code will only be checked if the requested solution of the exercise has been presented correctly. **Source code without the solution of the exercise presented in a readable way will not be awarded any marks.** A correct source code will not improve the marks if the main solution is not correct. * Your source code should be properly commented, with all variables and functions clearly explained. * Whenever a question asks you to relate quantities to the nature of the network, you should explain what the values mean by relating them to the meaning of vertices and edges for that particular network. Just describing the values without connection with the physical role of the network is not enough. * Wherever third-party sources are used, you are obliged to include the appropriate reference. **Failure to do so might lead to a plagiarism process.** * Analysing a network different from the one you were asked to will imply a penalty of 50% on the marks of the corresponding exercise after marking. * Use the appropriate technical language taught during the module. Do not invent new terms, unless you carefully and rigorously define them |
| beforehand. Do not use non-technical language to describe mathematical entities as it is not precise enough and prone to ambiguity.  • Please submit **only one PDF file** with the detailed solutions. **If more than one file is uploaded, they will not be considered as part of your submission.** |
| Recommended reading/ online sources:  • Reading list, recorded lectures, lecture slides and any other material on the module’s Blackboard page. |
| Key Dates:  Submission: 30/11/2021 (4 weeks after release)  Feedback: 28/12/2021 (up to 4 weeks after submission) |
| Submission Details:  • Online on Blackboard |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Marking Rubric:   * The whole coursework is worth 100 marks: 80 - content, 20 - presentation • This coursework is worth 50% of the module’s total marks.   **Marking Criteria for Presentation**   |  |  | | --- | --- | | 20 | Perfect presentation. Readable, well organised, professional looking. | | 19 - 14 | Good presentation. Some issues with organisation and/or format. | | 13 - 8 | Acceptable presentation. Several issues with organisations and/or format. Readability is somewhat compromised. | | <7 | Below-standards presentation. Too many issues of organisation and/or format. Difficult to read and follow. |     **General Advice**  Presentation is a question of common sense. You should be critical about your work by putting yourself in the place of the marker and question what you would think of your presentation: how many marks would you assign to it if you were a professional educator marking it?    **Points to Notice**  Some particular mistakes are often repeated. The following non-exhaustive list provides advice on some of the most common points of concern:   * Solutions should be presented in the order they were asked. A different order not only compromises readability, but also shows lack of care. * Graphs should be of professional standards: all axes should be appropriately labelled, captions should be informative, every annotation should be readable (pay attention to the size of the fonts!) and any provided graph should be referred to in the main text. |

**Network Science**

# Coursework 1 – 2021/22

|  |
| --- |
| **Notice:** unless explicitly stated on its text, exercises can be solved using any piece of software you choose. In every instance, you have to explicitly mention what software you are using and provide any code. |

**1. (32 Marks)** For this exercise, choose an **undirected** network from any online repository with more than 1000 nodes. Some things to pay attention:

* Provide the appropriate reference for the dataset.
* There are some repositories linked on the module’s Blackboard page, but you are free to choose networks from other repositories.
* When choosing your network, it is a good idea to first read the exercises and seek for a network that will be appropriate for the required analysis.

1. (2 marks) Describe the nature of the network by explicitly indicating what kind of objects are its nodes and their number, what kind of relation the links represent and their number, and what actual function does the network represent.
2. (6 marks) Using the maximum likelihood method, adjust a power law distribution to the degree distribution of your network and test the quality of the result by calculating the root mean square error between the estimated and the actual degree distribution. Write down explicitly the expression for the adjusted power law.
3. (5 marks) Calculate the root mean square error obtained by estimating the degree distribution of your network using a Poisson approximation of a 𝐺(𝑁, 𝑝) model with the appropriated parameter obtained from your data.
4. (6 marks) Plot the original degree distribution, the adjusted power law and the Poisson estimate in the same graph for the sake of comparison. Interpret your results by comparing the shape of the distributions and their root mean square errors. Indicate which of the latter two approximates better the actual distribution of the network and justify whether it makes sense for the kind of network you have.
5. (7 marks) Create a plot where the horizontal axis represents the average degree and the vertical axis the number of nodes in the giant component of the network. Each point of your plot should be obtained by the following method:
   * The first point corresponds to the average degree of your network and the fraction of nodes in its giant component.
   * Each subsequent point is obtained by removing the node with largest degree from your network and recalculating the above quantities (average degree and fraction of nodes in giant component).
   * You should obtain a plot with the same number of points as the number of nodes in your network.

Does the same phase transition explained in your lectures happen here?

1. (6 marks) Plot a graph of the average distance versus the size of the network using the following procedure:
   * The first point should be the average distance of the **giant component** of your network versus its number of nodes.
   * Remove randomly half of the nodes of your network, calculate the same quantities as above and plot a new point. • Repeat the above bullet point until you have just one node.

Analyse the result relating it to the nature of your network. Can you observe the small world property? Justify.

**2. (28 marks)** Choose a directed network with at least 500 nodes from any repository and describe its nature as well as its number of nodes and links (you will use it in the following items).

1. (5 marks) Explain what is assortativity by degree and calculate it for your chosen network **ignoring the link directions**. Analyse your result taking into consideration the nature of the network.
2. (6 marks) Generate 10 networks with the same number of nodes using the Barabasi-Albert model with 𝑚 = 1 and calculate the assortativity by degree for each one of them and also their average. Is the result similar or different from that you obtained for your network? What conclusion can you reach from this comparison.

|  |
| --- |
| For the next 3 items, you should use the network you chose in the beginning of this exercise. |

1. (8 marks) Plot the in and out degree distributions. Explain the meaning of each distribution and analyse your results with respect to the nature of the network.
2. (6 marks) Plot the distribution of hubs and authorities using the HITS algorithm. Explain how the algorithm works and the meaning of both distributions. Compare them with those obtained on the previous item.
3. (3 marks) Plot the distribution of PageRank centrality (directed). Explain what this centrality measures.

**3. (20 marks)**

a) (8 marks) By generating an appropriate sequence of random networks using the Barabasi-Albert model for a fixed value of 𝑚 (you should choose it), plot a graph of the ratio 〈𝐶〉/〈𝑘〉 (average clustering divided by average degree) as a function of the number of nodes and comment on how it compares to real networks.

c) (6 marks) Generate a 𝐺(𝑁, 𝐿) network with 200 nodes and 1000 links. Using degree configuration of this network, generate 200 networks using the configuration model and plot the distribution of the average distance of their **giant component**. What can you conclude from your results? Explain why we use the giant component and not the whole network.

b) (6 marks) Find the formula for a Random Exponential Graph where the constraint is the average value of the number of nodes (follow the calculations on your lecture slides). Choose a fixed value of the average number of nodes and plot the probability of a graph with 𝑁 nodes. Exclude graphs with zero nodes from the model.