MSDM 5056 Network Modeling
Labs and Project (2023)

There are two labs and one project in this course. Each of the labs is worth 20%, and the project is worth 25% of the final grade. For the labs, you only need to paste your results onto a document and submit it. For the project, you need to write a short report and **discuss** the properties of the networks that you construct. The deadline of the two labs is November 22, 2023, and the deadline of the project is December 6, 2023.

Lab 1: Random Network (*Refer to Lecture 6*)

(a) Generate a random network with $N = 100$ nodes and an average degree $\langle k \rangle = 20$.

(i) Plot its degree distribution $P(k)$.

(ii) Compute its average shortest distance L.

(b) Now repeat (a) with $N = 500, 1000, 5000, 10000$.

(c) Plot L against $\ln N$. What is the relation between L and N?

Design your program wisely so that it will not get stuck at a large N. If you cannot improve its efficiency, you may consider estimating L with some statistical techniques, but you must report your estimation error as well.

Lab 2: Small-World Network (*Refer to Lecture 7*)

Generate some small-world networks each with 1000 nodes but a different rewiring probability p. Make a plot of the average shortest distance L and the average clustering coefficient C against p like Fig. 1 below.

You need to generate a few samples for each p and take the average in order to get a smooth curve.

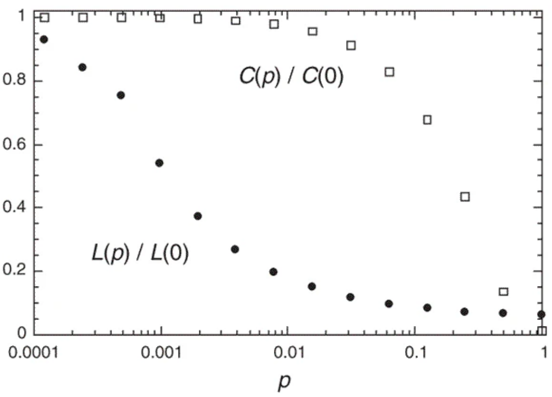

Fig. 1. Figure 2 from Watts and Strogatz [1].

Project

Choose two datasets from the website Network Science [2] or other reliable sources. One should be small and contain only 100 to 200 nodes. The other one should be large and contain at least 1000 nodes. You must not use the datasets that the tutorials have used.

(a) Plot their degree distributions, then compute their average shortest distances and average clustering coefficients.

(b) Community Detection (*Refer to Lecture 5*)

(i) For the small network, use the Girvan-Newman algorithm [3] to find its communities and visualize the results.

(ii) For the large network, you can use any algorithm you like, then you need to visualize the results with another network in which

- each node represents a community;
- the size of a node indicates the size of a community;
- the width of a node's border indicates the number of edges inside a community; and
- the width of an edge indicates the number of edges between two communities.

(c) Network Resilience (*Refer to Lecture 9*)

Perform (i) a random attack and (ii) a targeted attack on the large network by removing nodes from it. Plot the size of the giant component against the fraction of removed nodes like p.17–18 in Lecture 9.

For a random attack, you need to repeat your experiment at least a few times and take the average in order to get a smooth curve. For a targeted attack, you need to specify the strategy that you use.

Write a short report to show all the answers you have obtained. Briefly discuss what kind of networks you have (by, e.g., analyzing their statistical and topological properties) and what other information you may get from your results.

You must write your report in an academic style. You may follow the format of Lecture Notes in Computer Science (LNCS) if you like. You can download its template from LNCS's website [4]. You are also welcome to it for Labs 1 and 2 if you want to submit a report for the labs.

References

1. Watts, D. J., Strogatz, S. H.: Collective dynamics of ‘small-world’ networks. Nature 393(6684), 440–442 (1998).
2. Network Science, http://network-science.org/
3. Girvan, M., Newman, M. E. J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA 99(12), 7821–7826 (2002).
4. Information for Authors of Springer Computer Science Proceedings, https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines