Homework 2

Deadline: 2018.05.20 (Sunday) 23:59

Problem: Generating Realistic Social Networks

Realistic social network information networks satisfy three fundamental properties, i.e., short average path length, high clustering coefficient, and power-law degree distribution. To generate synthetic networks satisfying these properties, three well-known network generation models, *Erdos-Renyi* (ER) model, *Barabasi-Albert* (BA) model, and *Watts-Strogatz* (WS) model, had been proposed. However, each of these models generate synthetic networks possessing only a *subset* of the three properties. Your task in this homework is to develop one or more different algorithms that are able to generate synthetic but realistic networks by gradually adding nodes and edges to a certain time point so that *all of the three properties can be satisfied*.

To evaluate your network generation algorithms, you are asked to not only submit your source codes with clear comments (in either Python or R), but also describe the details of your methods in a report containing the following justification. The maximum length of the report is 10 pages using this template: https://www.acm.org/publications/proceedings-template . Note that you are encouraged to use LaTeX to compile your report and submit your report in PDF.

- **Description.** What are your methods for adding nodes and edges into the synthetic graphs? What are the main ideas, intuitions, and physical meanings of your methods? You are asked to write down the detailed procedures (e.g. algorithms) of your methods. Please also give a title of your report, and name your proposed methods in the report.
- **Simulation Comparison.** A direct approach to evaluate your methods is generate some synthetic graphs, and produce simulation comparison by examining the three properties in ways such as Page 33, 35, 52-56, 69, and 71-72 of the slides in Lecture 5. You can create tables and generate some figure plots to justify the graphs generated by your methods.
- **Visualization.** Seeing is believing. You are asked to plot some figures of the graphs generated by your methods (by varying some parameters), together with some textual description, to explain the similarity or the differences for the graph shapes between your generated graphs and the realistic graphs introduced in the course lecture. To have clear observation by eyes, you may want to use a limited number of nodes and edges in the visualizing graphs.
- **References.** If you methods are implementing some of existing network generation algorithms (note: you cannot directly call any graph generation functions in any packages written by others), you still need to include the abovementioned three parts in your report, and provide the references to the corresponding papers. If you totally have no ideas about how to generate

networks satisfying the three properties, we have provided you three papers for your references in the following. Note that it is worthwhile reading the first paper [1] since it well reviewed a number of network generation models; therefore, you can find more references to existing well-known methods.

- **Theoretical Justification.** [Optional+Bonus] By following the lecture slides, you can also provide some theoretical justification by showing the mathematical forms of average path length, clustering coefficient, and degree distribution derived from the synthetic graphs generated by your methods.

Note: If none of the methods that you had tried and developed can generate graphs satisfying the three properties, it's fine. But you also need to write down all the methods you have tried and show the abovementioned items, and analyze why it cannot generate the realistic networks in your report. With your report, we are able to understand which methods cannot work even though they possess some physical meanings.

References

[1] Deepayan Chakrabarti and Christos Faloutsos. Graph Mining: Laws, Generators, and Algorithms. *ACM Computing Survey*, 38(1), 2006. (517 cites)

[2] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg and Edoardo M. Airoldi. A Survey of Statistical Network Models. *Foundations and Trends in Machine Learning*, 2(2), 129-233, 2010. (442 cites)

[3] Leman Akoglu and Christos Faloutsos. RTG: A Recursive Realistic Graph Generator using Random Typing. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (ECML PKDD), 13-28, 2009. (Best Paper Award)

How to Submit Your Homework?

You will need to submit multiple files. One is your Python/R code, and the other is the report in PDF format. Please name the source code file as "hw2.py" or "hw2.R". If you have developed multiple methods, you can name them as "hw2_XXX.py" and "hw2_YYY.py", where XXX and YYY are the names of your methods. In addition, please also submit your report as "hw2.pdf". Finally, zip your files and submit the file with file name "姓名_hw2.zip" using Moodle.