Network Science Assignment

Total marks: 30 marks (30%)

Submission and due date: email the source code and report (as one zip file) to Chengbin 11756007@mail.sustc.edu.cn by 16:00pm 23/July/2018.

Aim: to analyse a citation network using PageRank algorithm and other centrality measures. We need to find out which papers are the most important papers in terms of PageRank scores and centrality measures.

Requirements:

1. Import the dataset from Stanford Network Analysis Portal (http://snap.stanford.edu./data/cit-HepTh.html), which is the High-energy physics theory citation network from January 1993 to April 2003 (124 months).
2. Implement the in/out Degree and Eigenvector centralities from scratch (You are not allowed to use other people's code or library)
3. Implement the PageRank algorithm from scratch based on the Power method (You are not allowed to use other people's code or library)
4. Visualise the results from PageRank. You can use any software or write code yourself
5. Write a report to report your results. In the report, you should briefly introduce the PageRank algorithm by using a flowchart and pseudo-code. Discuss what the parameters are and how you select them. List the top 10 papers based on PageRank scores, degree and Eigenvector centralities, respectively. You can also find out the paper by looking at the paper meta-information file (i.e., cit-HepTh-abstracts.tar.gz). Discuss the results, e.g., the relationships between the three sets of top papers.

Marking scheme (Total 30 marks):

1. Correct and efficient Implementation of the centrality measures and the PageRank algorithm (15 marks)
2. Informative visualisation of the results (5 marks)
3. Report (total 10 marks): Introduce to the PageRank algorithm using a flowchart and pseudo-code (2 marks). List all the results (3 marks) and informative discussion about the results (5 marks).

Bonus (3 marks): Analysing the dynamic citation network:

1. Use the time stamps of each node to construct 10 networks (the first network consists of 16 months' citation, then 12 month each)
2. Calculate the PageRank scores of the 10 networks, report and discuss the results
3. Dynamic visualisation of the 10 networks along with the PageRank