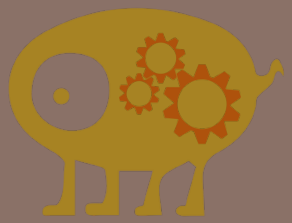

Data Mining
資料探勘

Project 3

Hung-Yu Kao, Fall 2018

Link Analysis Practice

- Please implement
  - HITS and PageRank (Lecture 7, P37, random jumping probability, i.e., damping factor=0.15) and calculate authority, hub and PageRank values for the following **8** graphs
    - 6 graphs in project3dataset
    - 1 graphs from project1 transaction data (connect items in each row, bi-directed or directed)
  - SimRank to calculate pair-wise similarity of nodes (choice any parameter C you like) , using
    - first **5** graphs of project3dataset.
- Find a way (e.g., add/delete some links) to increase hub, authority, and PageRank of Node 1 in first 3 graphs respectively.

Link Analysis Practice

- Please describe and analysis your results for each algorithm in each graph.
- Please also include your source code files in your uploaded file.
- Due: 12/25 9am

Requirement

- You should write a report for your system, including:
  - Implementation detail
  - Result analysis and discussion
  - Computation performance analysis
  - Discussion (what you learned from this project and your comments about this project)

Questions & Discussion (optional, but recommended)

- More limitations about link analysis algorithms
- Can link analysis algorithms really find the "important" pages from Web?
- What are practical issues when implement these algorithms in a real Web?
  - Performance discussion (time cost)
- What do the result say for your actor/movie graph?
- Any new idea about the link analysis algorithm?
- What is the effect of "C" parameter in SimRank?
- Design a new link-based similarity measurement