Final Project Guide

Prof. Daning Hu
Department of Informatics
University of Zurich
Outline

- Final Project Report consists of:
  - Project Overview (5%)
  - Web Mining Using Open Hub API (Optional)
  - Data Collection and Processing (10%)
  - Network Modeling and Visualization (20%)
  - Node Level Analysis (20%)
  - Group and Link Level analysis (20%)
  - Network Level analysis (20%)
  - Discussions and Conclusions (5%)

- General Requirements
Final Project Structure (Option 1)

The Final Project provides two options. It consists of:

- Project Overview
- Web Mining Using Open Hub API (Optional)
- Data Collection and Processing
- Network Modeling and Visualization
- Node Level Analysis
- Group and Link Level Analysis
- Network Level Analysis
- Discussions and Conclusions

The deliverable is an project report which should follow the style of the ICIS2017 submission template:

Project Overview

In this section you have to write an overview about the study you want to report finally. It mainly includes the following parts.

- The data used in your study.
- The motivation of your study.
- The expected insights from your study.

You can develop some research questions in this section.

- The research questions are open questions about the study you want to explore.
- At least two research questions. E.g.,:
- What is the topology of the network?
Web Mining Using Open Hub API (Optional)

- Web mining using the Open Hub API:
  - [https://github.com/blackducksoftware/ohloh_api](https://github.com/blackducksoftware/ohloh_api)
  - Collect information about the 10 projects
  - Use the project_id to crawl contributors (account_id)
  - Use account_id to crawl account information

- Please describe the detail web mining process in your report. E.g.:
  - The programming language for web mining.
  - The duration of your web mining.
  - The information of the data you collected (attributes)
Data Collection and Processing

- You need to:
  - Choose the relation database you want (MySQL, PostgreSQL, etc.)
  - Create a database for the data storage.
  - Import the given files into the database.
  - Make sure all data items are in the appropriate data format.
  - Give a detail description about the process of the data storage.

- If you do not choose to use the relational database, please describe how do you store your data.
  - The files of storing your data (format, name).
  - The relationships among the files of your data.
  - How to import the data into your working environment.
  - Please provide the description of the mentioned items above in your final report.
Network Modeling and Visualization

- You have to construct networks based on the data you collected for your following analysis.
  - The network data formats used for your analysis.
  - Why do you choose that format for network analysis?
  - The direction of the network constructed.

- If there is a two-mode network, please describe how to change it to a one-mode network (optional)

- Please provide a detail description about the process of the network construction in your final report!
  - You can use a graph to present the process. It would be more clear.
Node Level Analysis

Node Centrality Analysis.

You need to report the top 20 nodes in terms of their:

- Degrees.
- Betweenness.
- Closeness.
- Eigenvector.

Please also explain what these measures mean in your data context.

- Comparing the centrality measures calculated above and provide some reasonable explanations.
- You can follow the “Comparison of Centrality Measures” table in the slide Lecture2.
Group and Link Level Analysis

- Identify and visualize the largest component in your network.
  - Please make the visualization as more clear as you can in the final report.

- You need to report:
  - The number of components.
  - Size of the largest component.

- You can use R, Netdraw, Gephi, etc. to locate cut points in the largest component in the network.
  - Please visualize the cut points in the largest component.
Network Level Analysis

- Analyze the whole network using the igraph library and calculate all the following network measures. Compare them with the features of the three network topologies.

- You need to report:
  - Size of the network.
  - Centralization score (degree).
  - Average degree.
  - Average path length.
  - Clustering coefficient.
  - Degree distribution (Plot the degree distribution and Explain).

- You need to compare the measures against the three network models and categorize your network.
Discussions and Conclusions

- Summarize your network analysis and draw some conclusions, for example:
  - Whether the research questions proposed before have been addressed?
  - Explain how your analysis help you to answer your research questions.
  - Provides your hands-on experiences in analyzing real-world social networks in a systematic manner.

- Provide some future research questions.
  - At least three research questions proposed here.
Final Project Structure (Option 2)

- The second option is to collect, model and analyze networks using the dataset you find by yourself.

- The other requirements are the same in the option 1.
General Requirements

- This report must be written in English.

- The format of the final submission file is as follows:
  - Yourname_StudenID_BNAReport

- The deadline of submission is 24:00PM 18th October 2017.