Web Mining

Pajek

Oliver Lehmberg

FSS 2016
Outline

1. Intro

2. Overview
   1. Basic Concepts in Pajek
   2. Range of Functions

3. Tasks
   1. Visual Exploration
   2. Basic Graph Measures
   3. Community Detection
   4. Facebook Graph Analysis

4. Other useful tools

5. Datasets
Facebook Friend Network - A real world graph

- Facebook is a social network platform

- Connections between different objects are created
  - Friends: People → People
  - Followers: People → Pages
  - Fans: People → Bands
  - ...

- Facebook provides users with an API to get at least the directly connected data
  - Friends + Information + Connection between friends
  - Pages which are public or they like
  - Events + Participants
  - ...

Task 5: Explore your Facebook Graph

- **Basic Network Measures**
  - Apply the measures learned in the lecture to characterize your network

- **Display your graph and reorder it with Energy layout**
  - Are there any sub-networks you can detect by looking at the graph?

- **Are there differences between your male and female friends?**
  - Load the _sex partition and draw the graph. (0 = female, 1 = male)
  - Create new networks for both types of nodes
  - Operations > Extract from Network > Partition
  - Draw the new network – are there new components?

- **Are your friends also active outside your ego-network?**
  - Load the fc vector and draw the graph
  - Can you think about a possibility to find out your friends are beside your own eco-network linked to other networks (fc > degree)
Task 5: Explore your Facebook Graph

Although the graph we can gather from Facebook is a little bit limited in terms of size, we can use it for community detection:

- Find out what different sub-networks exist in your friend network. Do you find anything surprising?
  - Net > Partitions > Core > All
  - Operations > Transform > Remove Lines > Between Clusters

- Try to identify the social-gateways (bridges) in your network between your components.
  - Net > Vector > Centrality > Betweenness
Task 6: Identifying potential opinion leader

- Imaging you are a marketer who wants to detect persons in your network which are worth giving an incentive to distribute information about your new product. As you cannot get message flows of the different persons from Facebook to analyze them you are interested in persons, who have a lot of friends, but whose friends are also well connected to increase the chance that your message is distributed extensively and does not stuck somewhere.

- Think about the different measures you have heard about in the lecture. What measure would help you finding those persons in your network and why?

- Calculate the measure(s) for your Facebook network using Pajek and interpret the results. If you are unsure which measure is the best, apply both (or more) possibilities and observe if they correlate positively?
Other useful tools

- **Network Workbench (NWB)**
  - [http://nwb.cns.iu.edu/](http://nwb.cns.iu.edu/)
  - Large-Scale Network Analysis Tool
  - Last update 2011

- **UCInet**
  - [https://sites.google.com/site/ucinetsoftware/home](https://sites.google.com/site/ucinetsoftware/home)
  - Visualization and modelling

- **Gephi**
  - [https://gephi.org/](https://gephi.org/)
  - Primary nice visualization

- **GUESS**
  - [http://graphexploration.cond.org/](http://graphexploration.cond.org/)
  - Network analysis and visualization
  - Java-based, script driven (jython)
Datasets

- **Stanford Large Network Dataset Collection**

- **Archive.org**
  - Friendster social network (2011)

- **Infochimps**
  - CiteULike

- **WebDataCommons**
  - .de subgraph on PLD level: [http://data.dws.informatik.uni-mannheim.de/hyperlinkgraph/2012-08/pajek/graph_de.net.gz](http://data.dws.informatik.uni-mannheim.de/hyperlinkgraph/2012-08/pajek/graph_de.net.gz)