

Data Mining

Homework Assignment #11

Dmytro Fishman, Anna Leontjeva and Jaak Vilo

May 8, 2014

The goal for this homework is to get you started with basics of social network analysis. There are many different network analysis tools. Choose the one you prefer. Some of the well-known tools and packages are: NetworkX and igraph for Python, JUNG for Java, igraph for R, Gephi for the visualization with some built-in calculations and even NodeXL free template for Excel. Most of the exercises require you to check presentation slides for the definitions.

Task 1

To start with, open SNAP page by Jure Leskovec with the collection of some graphs. Take a look at the data and the description:

<http://snap.stanford.edu/data/wiki-Vote.html>.

Find the subsection files and download WikiVote.txt.gz. Load the data using one of the chosen tools. Note that the network is directed. Calculate the list of the statistics:

- number of nodes in the network
- number of edges in the network
- number of nodes with a self-loop
- number of mutual connections or *reciprocated* edges, i.e if there is a directed edge from node a to node b, there is also an edge from b to a.
- number of nodes with zero indegree
- number of nodes with zero outdegree
- number of nodes with the total degree (both in and out) greater than 30.

Task 2

Task 3

write your own function that generates erdos-renyi random graph with input parameter p , where p is the probability of edge creation. generate the graph with 50 nodes and at least five different p values. Plot the result.

Task 4

Task 5

Community detection

Task 6