

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. Most of the exercises require you to check presentation slides for the definitions.

Task 1

In this task, we will use the techniques of Social Network analysis to study how virus spread. Let us assume terrified biologists came to the Institute of Computer Science seeking for a help. Their email network was infected by the virus that was created by the student that received 'B' for his Master's Thesis and got offended. You are trying poor biologists to estimate the worst case scenario of virus spread. Biologists observed that if virus infects a node, it always infects all its immediate neighbors if they are not already infected (it has 100% of infection rate). Also we know that virus spreads only along the edge direction (e.g. if virus infects node A, which only has an incoming edge from node B, node B will not be infected).

Biologists provided you with their directed anonymous email network that you can access on the course web-page.

Load the data. To get the first insights about biologists' network, calculate the list of the following 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
- degree distribution of the given network

- optionally calculate whatever measure you deem appropriate for better understanding

What intuition you can gather from these numbers?

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