Social Network Analysis Home Assignment 2

{Your Name}

due date - 22.05.2017 23:59

Contents

Graph models. Centrality metrics

Please send your reports to hse.ntwks@gmail.com with the subject of of the following structure: *[MAGOLEGO SNA 2017] {LastName} {First Name} HA{Number}*

Late submission policy: -1 point per day

Use this file as a template for your report.
Support your computations with figures and comments. Send ONLY .Rmd versions of your report with datasets for your social network. Anonymize your data if necessary.

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

Task 1. Your social network

For the first task, you have to load your vk.com network. Please follow the instructions posted on the course wiki or user VK Application. For FB users try to use NetVizz. If you did it correctly, you should have a GraphML file with your own network. Read it to R:

```
## Put your code here
```

1. Degree distribution

First, plot degree distribution of your network in log-log scales:

```
## Put your code here
```

Is there any correspondence between actual degree distribution of your network and the Power Law distribution? If not, explain why.

Put your comments here

Now, let's see how it would look if it was random. Produce Erdos-Renyi graph matching your real network (same number of nodes and same average degree). Compare it with your degree distribution.

```
## Put your code here
```

Put your comments here

2. Compute centrality metrics

Compute for your network:

- degree centrality
- closeness centrality
- betweenness centrality
- eigenvector centrality
- Bonacich power centrality
- Alpha centrality

```
## Put your code here
```

Output six plots corresponding to six centrality metrics you've computed:

- Use first names of your friends as node labels on the graph (you may hide this information if you wish – change it by integer ID)
- Keep the same layout of the network
- Make node sizes and colours proportional to the respective centrality metrics

Hint: If you get stuck, lab 4 script might be helpful.

```
## Put your code here
```

Now, output top ten nodes in each ranking. Again, print only first names in your table to keep privacy:

```
## Put your code here
```

Comment on your results here - for example, why some nodes have high betweenness centrality and the others have high closeness centrality. Is this what you would expect to see?

Task 2. Flickr network

In the second task, you will work with a large directed graph.

Please download flickr.mat

Data contains sparse matrix A and list of user names. This is a denser part of the Flickr photo sharing site friendship graph from 2006. Edge direction corresponds to friendship requests (following). Some of the links are reciprocal, others not

It's a Matlab file. How to deal with it in R? There is a package `R.matlab`. Please install it and call `library(R.matlab)`

Now use `readMat` function to read the file and extract adjacency matrix and a list of user names:

```
flickr = readMat("YOUR_PATH/flickr.mat")
fmatrix=as.matrix(flickr[1]$A)
fnames=flickr[2]$names
```

If you have trouble loading large mat file on your laptop — try to use HSE computer classes with installed R+RStudio.

Look at user names. You might want to remove spaces from the names. Use a function `gsub` to remove them:

```
fnames=gsub(" ", "", fnames)
```

Now create a graph, output the number of vertices and edges:

```
## Put your code here
```

Compute in- and out- degree centralities, PageRank, Hubs and Authorities for this network:

```
## Put your code here
```

Print top ten names in each ranking:

```
## Put your code here
```

Produce the following plots:

- In-degree centralities versus out-degree centralities
- In-degree centralities versus authorities
- Out-degree centralities versus hubs
- Hubs versus authorities
- PageRank versus hubs
- PageRank versus authorities

```
## Put your code here
```

Comment on the relationships between different centrality metrics