

HomeWork 5

{Your Name}

deadline - 08.05.2016

Contents

Graph models. Centrality metrics	1
Task 1. Flickr network	1
Task 2. Yahoo Music Network	2

Graph models. Centrality metrics

Use this file as a template for your report.

Support your computations with figures and comments. Send ONLY hw5-6_Surname.Rmd versions of your report.

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

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

Task 2. Yahoo Music Network

For this assignment, you have to load [the part of Yahoo Music Network](#). Archive contains network in GML format and corresponding list of artists.

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

Edges in this network appear if enough number of users have given ratings to both music bands. Note, that edges are weighted with similarity of the ratings.

1. Preprocessing

This network needs to be preprocessed. Do the steps, described on seminar, to make it look properly. How many vertices and edges were deleted?

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

2. Clustering

Define clusters for this networks using one of the algorithms described on lectures and seminars:

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

Compute density of obtained clusters:

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

Compute the ratio of inner clusters connections to outer ones:

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

3. Visualization & interpretation

Visualize five of the most dense clusters. Use names of artists as node labels on the graph.

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

(Extra task) Try to interpret (using Wikipedia or other resources) obtained clusters.

Insert comments here