Social Network Analysis Home Assignment 3

Chagev Vladislav

due date - 05.06.2016 23:59

Contents

Network communities

Please send your reports to network.hse.2016@gmail.com with the subject of of the following structure: *[MAGOLEGO SNA 2016] {LastName} {First Name} HW{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.

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

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

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

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.

```
g <- read.graph("music2K.gml", format = "gml")
artists <- scan("artists.txt", what="", sep="\n")
V(g)$label = artists
```

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?

```
before_ver <- vcount(g)
before_edg <- ecount(g)

g <- delete.vertices(g, degree(g) == 0)
simplify(g)
```

```
## IGRAPH U-W- 1996 6668 --
## + attr: id (v/n), label (v/c), weight (e/n)
## + edges:
##  [1]  1-- 639  1--1741  2-- 365  2-- 871  2--1403  2--1516  2--1837
##  [8]  2--1881  3-- 324  3-- 827  3--1425  3--1593  4-- 427  4-- 697
## [15]  4-- 755  4-- 837  4--1629  5-- 102  5-- 272  5-- 358  5-- 723
## [22]  5-- 759  5--1118  5--1217  5--1386  5--1513  5--1558  5--1892
## [29]  6-- 760  6-- 870  6--1246  6--1704  7-- 614  7-- 820  7--1010
## [36]  7--1097  7--1110  7--1408  7--1544  7--1607  7--1877  7--1928
## [43]  8--  56  8-- 190  8-- 487  8-- 512  8-- 841  9-- 271  9-- 526
## [50]  9-- 935  9-- 997 10-- 216 10-- 291 10-- 384 10--1003 10--1292
## + ... omitted several edges
```

```
after_ver <- vcount(g)
after_edg <- ecount(g)

before_ver - after_ver
```

```
## [1] 4
```

```
before_edg - after_edg
```

```
## [1] 0
```

2. Clustering

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

```
c <- multilevel.community(g)
```

Compute density of obtained clusters:

```
dens <- 1:length(c)
inner_con <-  1:length(c)
for (i in 1:length(c)) {
  temp_g <- induced.subgraph(g, which(c$membership == i))
  inner_con[i] <- ecount(temp_g)
  dens[i] <- graph.density(temp_g)
}
dens
```

```
##  [1] 0.53846154 0.13368984 0.03487859 0.10033670 0.06915114 0.10109290
##  [7] 0.05372405 0.17028986 0.03851516 0.17815126 0.04314134 0.05025052
## [13] 0.05708548 0.08980213 0.25362319 0.05326733 0.34640523 0.09813665
## [19] 0.05833905 0.03556658 0.06637807 0.11311475 0.64285714 0.09675325
## [25] 0.13636364 0.11035730 0.13846154 0.10862471
```

Compute the ratio of inner clusters connections to outer ones:

```
sum(inner_con) / ecount(g)
```

```
## [1] 0.854979
```

3. Visualization & interpretation

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

```
top_dens_clusters <- which(dens %in% sort(dens, decreasing = TRUE)[1:5])
top_vert <- which(c$membership %in% top_dens_clusters)
top_graph <- induced.subgraph(g, top_vert)
```

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

I used gephi tool to visualize graph. The image is attached. Artists in the segments have similar Genres. For example freestyle, jazz and others