

Social Network Analysis Home Assignment 3

Lev Mazaev

due date - 03.06.2019 23:59

Contents

Yahoo Music Network	1
-------------------------------	---

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.

```
ymnRaw <- read_graph(file = 'music2K.gml', format = 'gml')
vLabels <- read.delim(file = 'artists.txt', header = FALSE, stringsAsFactors = FALSE)[, 1]
ymnRaw <- set_vertex_attr(graph = ymnRaw, name = 'label', value = vLabels)
```

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

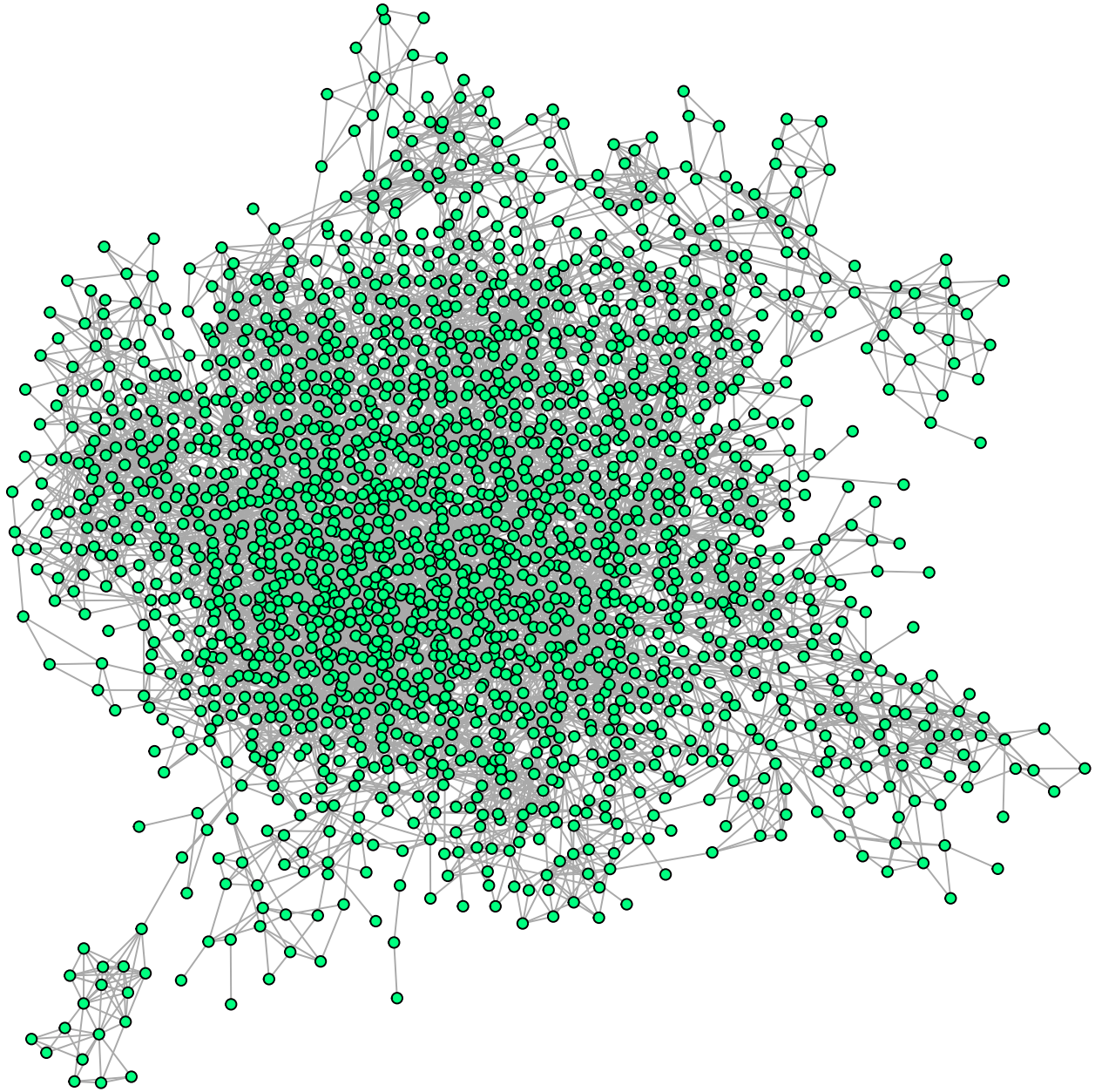
This network needs to be preprocessed. Do the steps, described on seminar, to make it look properly. How many vertices and edges were deleted?

```
ymn <- delete_vertices(ymnRaw, degree(ymnRaw) == 0)
ymn <- simplify(ymn)
vDiff <- vcount(ymnRaw) - vcount(ymn)
eDiff <- ecoun(ymnRaw) - ecoun(ymn)
```

4 vertices deleted, 0 edges deleted.

Let's plot the graph itself.

```
plot.igraph(ymn, layout = layout_with_lgl, vertex.label = NA,
            vertex.size = 2, vertex.color = 'springgreen')
```



2. Clustering

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

```
commYMN <- cluster_louvain(graph = ymn)
```

Compute density of obtained clusters:

```
clustDensity <- sapply(communities(commYMN), function(comm) {
  subYMN <- induced_subgraph(graph = ymn, vids = comm)
  return(edge_density(graph = subYMN))
})
```

Compute the ratio of inner clusters connections to outer ones:

```

clustRatio <- sapply(communities(commYMN), function(comm) {
  subYMN <- induced_subgraph(graph = ymn, vids = comm)
  innerConnections <- ecount(subYMN)
  # induced_subgraph does not keep outer edges
  outerConnections <- sum(degree(graph = ymn, v = comm)) - innerConnections * 2
  return(innerConnections / outerConnections)
})

```

Obtained stats for each cluster:

```

dfYMN <- data.frame(`cluster number` = names(clustDensity),
  density = clustDensity,
  ratio = clustRatio,
  check.names = FALSE,
  stringsAsFactors = FALSE)
knitr::kable(dfYMN[order(dfYMN$density, decreasing = TRUE), ], row.names = FALSE)

```

cluster number	density	ratio
23	0.6428571	4.500000
1	0.5384615	3.818182
17	0.3464052	3.117647
15	0.2536232	70.000000
10	0.1781513	2.864865
8	0.1702899	6.714286
27	0.1384615	2.700000
25	0.1363636	11.727273
2	0.1336898	3.571429
22	0.1131148	1.380000
26	0.1103573	5.083333
28	0.1086247	2.741176
6	0.1010929	2.402597
4	0.1003367	3.634146
18	0.0981366	19.750000
24	0.0967532	4.656250
14	0.0898021	9.440000
5	0.0691511	3.630435
21	0.0663781	4.075949
19	0.0583391	2.628866
13	0.0570855	1.519084
7	0.0537241	3.728814
16	0.0532673	3.955882
12	0.0502505	2.214286
11	0.0431413	1.621739
9	0.0385152	3.877193
20	0.0355666	3.071429
3	0.0348786	2.005076

3. Visualization & interpretation

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

```

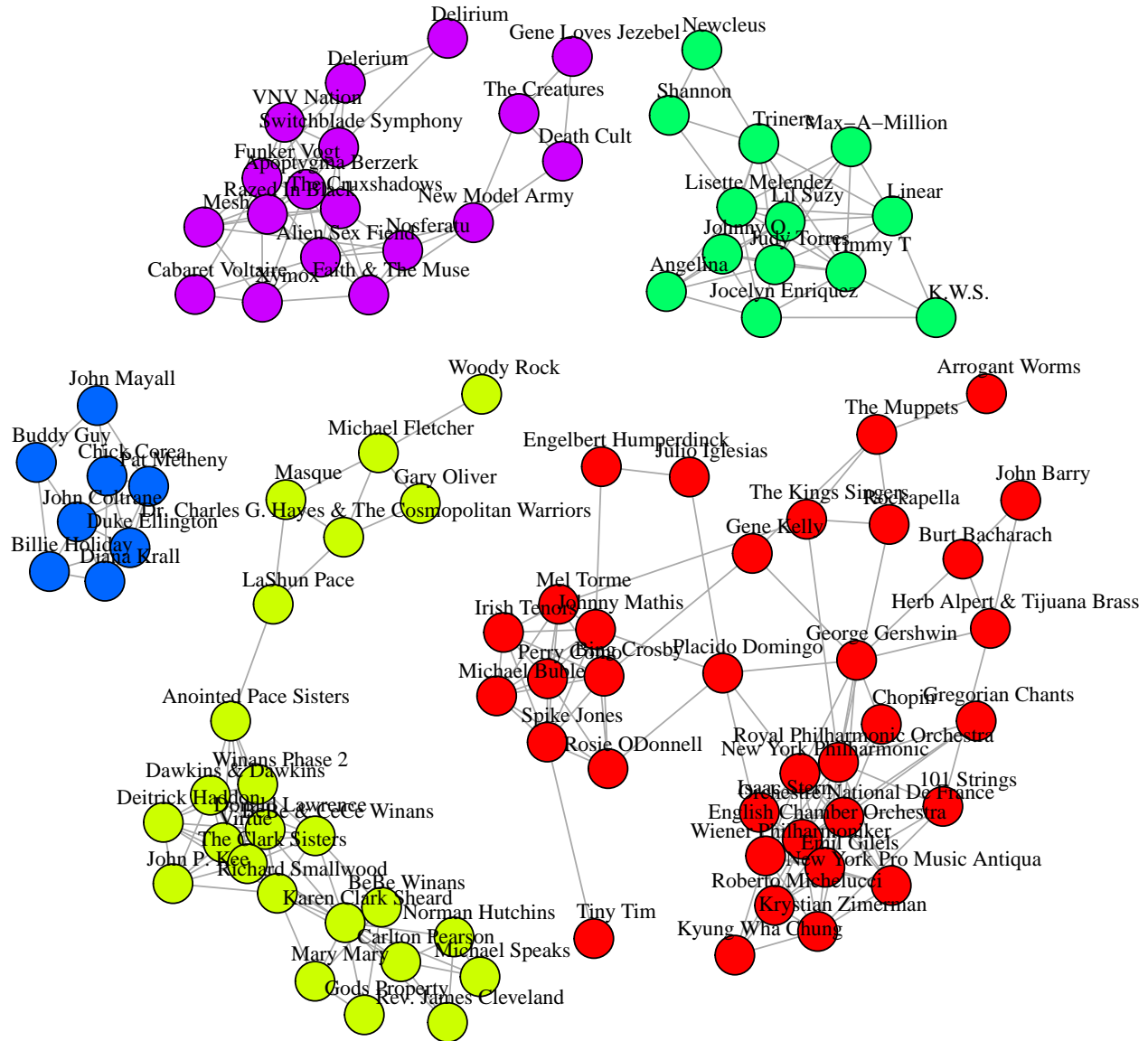
mostDenseClusters <- order(clustDensity, decreasing = TRUE)[1:5]
theirVertices <- unlist(communities(commYMN)[mostDenseClusters])

```

```

theseClustersGraph <- induced_subgraph(graph = ymn, vids = theirVertices)
V(theseClustersGraph)$color <- rainbow(5)[components(theseClustersGraph)$membership]
plot(theseClustersGraph,
     vertex.size = 8,
     vertex.label.color = 'black',
     vertex.label.dist = 1,
     vertex.label.cex = 0.8)

```



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

Genres of several chosen performers from each cluster are present below:

- Green cluster: pop, freestyle, dance
- Lightgreen cluster: gospel, soul, R&B
- Blue cluster: jazz
- Purple cluster: electronic rock, gothic rock, synthpop
- Red: classical, opera, comedian - mix of genres

As we know, edges in this network appear if enough number of users have given ratings to both music bands.

So, these clusters may correspond to particular sets of people's preferences in music. And these sets do not intersect much with each other (high inner to outer connections ratio) possibly revealing the fact that most people do not change their music preferences often.