Data Science Challenge

Centrality metrics

Delermando Branquinho Filho

10 de novembro de 2016

Fist Challenge

In this challenge, suppose we are looking to do social network analysis for prospective customers. We want to extract from their social network a metric called "closeness centrality".

Centrality metrics try to approximate a measure of influence of an individual within a social network. The distance between any two vertices is their shortest path. The *farness* of a given vertex v is the sum of all distances from each vertex to v. Finally, the *closeness* of a vertex v is the inverse of the *farness*.

The first part of the challenge is to rank the vertices in a given graph by their *closeness*. The graph is provided in the attached file; each line of the file consists of two vertex names separated by a single space, representing an edge between those two nodes.

Load Libraries

```
library(GGally)
library(network)
library(sna)
library(ggplot2)
```

Loadding Dataset

```
myData <- as.matrix(read.table("D:/Data Science/Semantix/edges.dat", quote="\"",
                    comment.char="", stringsAsFactors=FALSE))
```

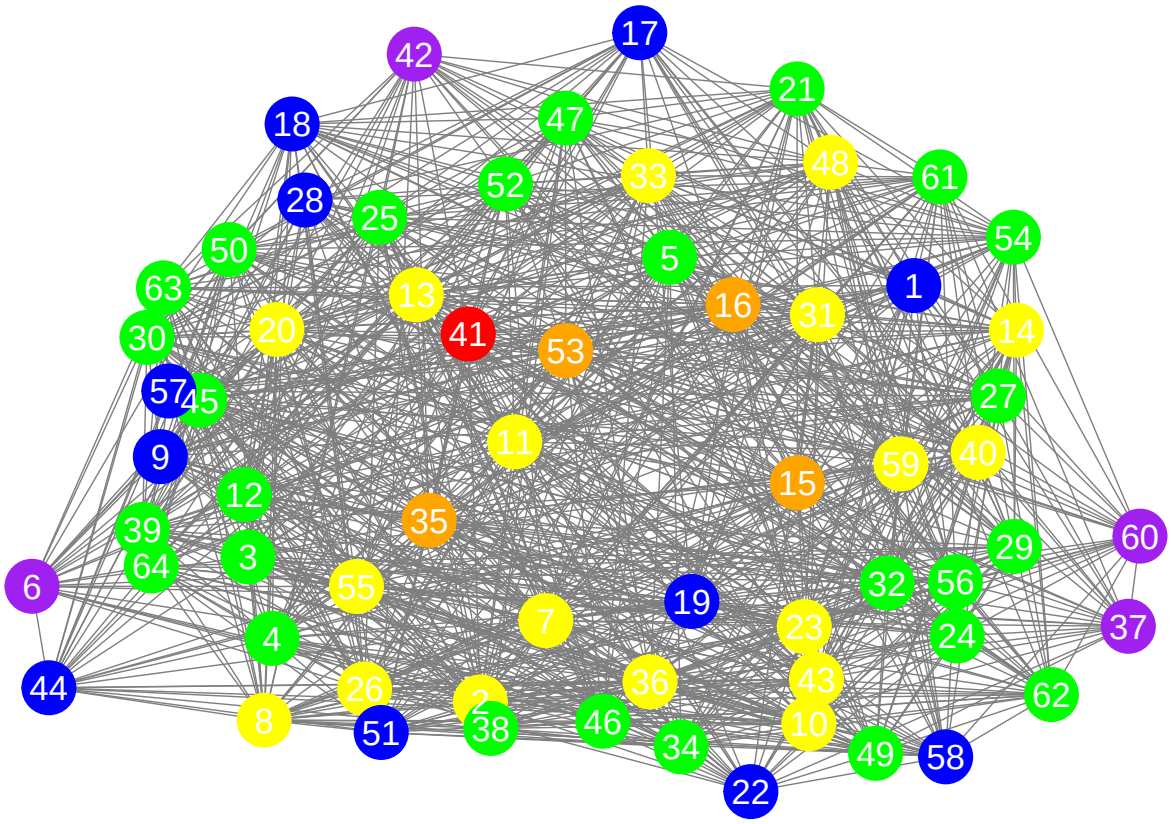
Creating Network Data

The nodes are numbered by one to number of nodes.

```
net = rgraph(myData, mode = "graph")
net = network(net, directed = FALSE)
network.vertex.names(net) = 1:dim(myData)[1]
```

Plotting

```
c.d   <- degree(net)
col<- as.integer(5*(c.d-min(c.d))/diff(range(c.d))+1)
palette <- c("purple","blue","green","yellow","orange","red")
g<-ggnet2(net, label = TRUE, label.color = "white",color = palette[col])
plot(g,vertex.color=palette[col],main="Degree Centrality",
     layout=layout.fruchterman.reingold)
```



Second Challenge

The second part of the challenge is to create a RESTful web server with endpoints to register edges and display the centrality of the graph.

Degree

The node degree is the number of relations (edges) of the nodes.

```
degree(net)
```

```
##  [1] 54 74 60 60 66 48 72 70 54 68 74 66 72 68 76 76 52 56 54 68 62 54 74
## [24] 60 62 72 60 52 64 64 72 64 72 62 76 74 44 62 62 70 82 50 70 52 64 62
## [47] 60 68 64 62 58 66 76 60 68 66 56 54 70 44 60 60 60 62
```

Betweenness centrality is even more important statistical property of a network. This is applied in a lot of real-world problems, such as finding influential people in a social network, finding crucial hubs in a computer network, finding border crossing points which have a largest traffic or trade flow.

```
degree(net,cmode="indegree")
```

```
##  [1] 27 37 30 30 33 24 36 35 27 34 37 33 36 34 38 38 26 28 27 34 31 27 37
## [24] 30 31 36 30 26 32 32 36 32 36 31 38 37 22 31 31 35 41 25 35 26 32 31
## [47] 30 34 32 31 29 33 38 30 34 33 28 27 35 22 30 30 30 31
```

Closeness (farness) centrality indicates how long it will take for information from a given node to reach other nodes in the network. The smaller the value, the more central role the node plays in the network.

```
degree(net,cmode="outdegree")
```

```
##  [1] 27 37 30 30 33 24 36 35 27 34 37 33 36 34 38 38 26 28 27 34 31 27 37
## [24] 30 31 36 30 26 32 32 36 32 36 31 38 37 22 31 31 35 41 25 35 26 32 31
## [47] 30 34 32 31 29 33 38 30 34 33 28 27 35 22 30 30 30 31
```

The centrol point of the map is:

```
centralization(net,degree,cmode="indegree")
```

```
## [1] 0.149156
```

REFERENCES

Alex, B. 1950. Communication patterns in task-oriented groups. J. Acoust. Soc. Am. 22 (6): 725-730.
Sabidussi, G. 1966. The centrality index of a graph. Psychometrika. 31: 581-603.