Social Network Analysis (SNA) Part2

Nik Bear Brown

In this lesson we'll learn the how to implement and analyze Social network analysis (SNA) in R.

# Additional packages needed

To run the code you may need additional packages.

* If necessary install the followings packages.

install.packages("igraph");  
install.packages("tools");  
install.packages("visNetwork");

library(igraph)

##   
## Attaching package: 'igraph'

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

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

library(tools)  
library(visNetwork)

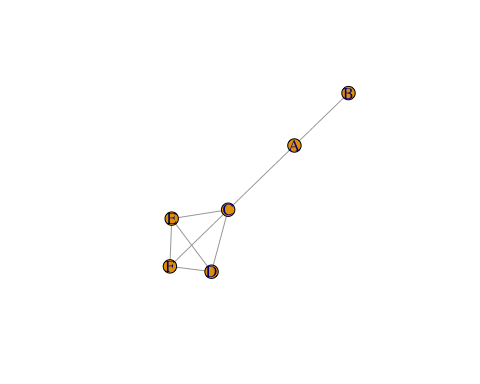
##   
## Attaching package: 'visNetwork'

## The following object is masked from 'package:igraph':  
##   
## %>%

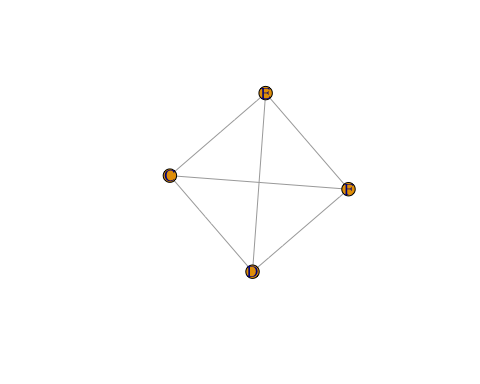
# Data

We will be creating graphs with the library(igraph).

# for reproducibility of graphs plots (plot.igraph uses random numbers)  
set.seed(3333)  
# create an example graph  
D <- read.table(header=T,text=  
 'from to  
A B  
A C  
C D  
C F  
C E  
D E  
D F  
E F')  
g1 <- graph.data.frame(D,directed=F)  
# plot the original graph  
plot(g1)



# find all the largest cliques (returns a list of vector of vertiex ids)  
a <- largest.cliques(g1)  
# let's just take the first of the largest cliques  
# (in this case there's just one clique)  
clique1 <- a[[1]]  
# subset the original graph by passing the clique vertices  
g2 <- graph.full(length(clique1))  
V(g2)$name <- V(g1)$name[clique1]  
# plot the clique  
plot(g2)



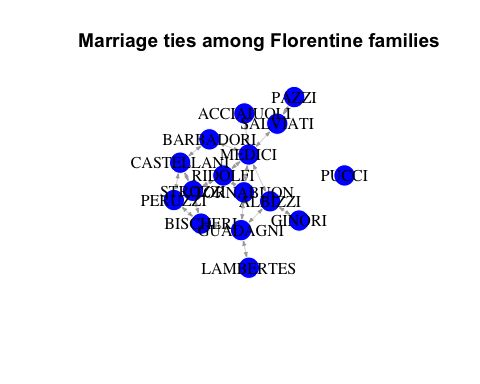
In addition we will be using social relations among Renaissance Florentine families (person aggregates) collected by John Padgett from historical documents. The two relations are business ties (PADGB - specifically, recorded financial ties such as loans, credits and joint partnerships) and marriage alliances (PADGM). See <http://moreno.ss.uci.edu/data.html#padgett>

# Studying marriage ties among Renaissance Florentine families  
data\_url <- 'http://nikbearbrown.com/YouTube/MachineLearning/M12/padgett-cleaned.txt'  
padgett.cleaned <- read.table(url(data\_url), quote="\"")

# Cliques and Motifs among Renaissance Florentine families

Find and plot all the largest cliques seperately. Find any significant motifs.

florentine\_marriage\_ties <- padgett.cleaned[1:16,]  
rownames(florentine\_marriage\_ties) <- c("ACCIAIUOLI", "ALBIZZI", "BARBADORI",   
 "BISCHERI", "CASTELLANI", "GINORI",   
 "GUADAGNI", "LAMBERTES","MEDICI",   
 "PAZZI", "PERUZZI", "PUCCI",   
 "RIDOLFI", "SALVIATI", "STROZZI",   
 "TORNABUON")  
colnames(florentine\_marriage\_ties) <- rownames(florentine\_marriage\_ties)  
florentine\_graph <-   
 graph\_from\_adjacency\_matrix(as.matrix(florentine\_marriage\_ties))   
plot(florentine\_graph, layout=layout.fruchterman.reingold,  
 main = "Marriage ties among Florentine families",  
 vertex.label=V(florentine\_graph)$name, vertex.size=25,   
 vertex.color="blue", vertex.frame.color="white",  
 vertex.label.color="black", vertex.label.cex=1,  
 edge.arrow.size=0.25, edge.width=0.25)



count\_motifs(florentine\_graph, size = 3)

## [1] 41

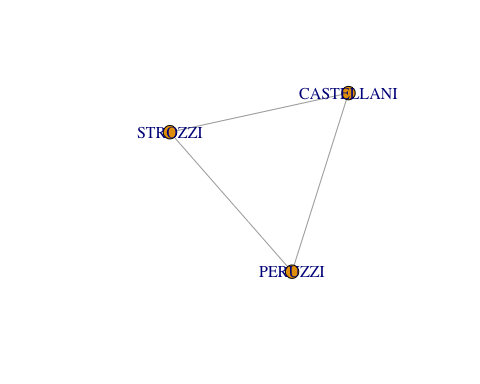
count\_motifs(florentine\_graph, size = 4)

## [1] 91

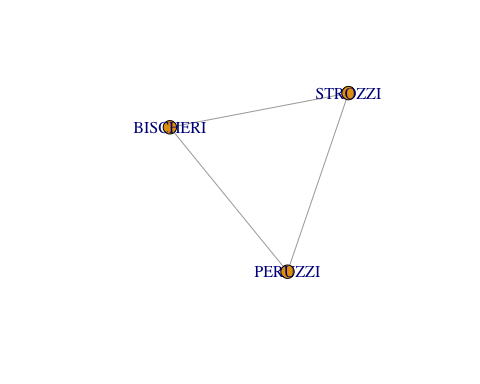
# find and plot all the largest cliques seperately  
florentine\_graph\_undirected <- as.undirected(florentine\_graph)  
# following code taken from  
# http://stackoverflow.com/questions/26222659/identifying-cliques-in-r   
all <- largest\_cliques(florentine\_graph\_undirected)  
all

## [[1]]  
## + 3/16 vertices, named:  
## [1] STROZZI PERUZZI CASTELLANI  
##   
## [[2]]  
## + 3/16 vertices, named:  
## [1] STROZZI PERUZZI BISCHERI  
##   
## [[3]]  
## + 3/16 vertices, named:  
## [1] MEDICI RIDOLFI TORNABUON

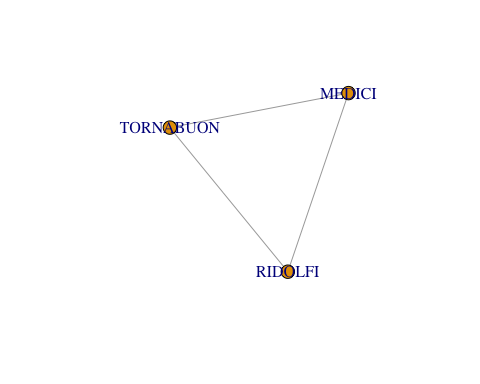
clique\_graph1 <- make\_full\_graph(length(all[[1]]))  
V(clique\_graph1)$name <- V(florentine\_graph\_undirected)$name[all[[1]]]  
plot(clique\_graph1)



clique\_graph2 <- make\_full\_graph(length(all[[2]]))  
V(clique\_graph2)$name <- V(florentine\_graph\_undirected)$name[all[[2]]]  
plot(clique\_graph2)



clique\_graph3 <- make\_full\_graph(length(all[[3]]))  
V(clique\_graph3)$name <- V(florentine\_graph\_undirected)$name[all[[3]]]  
plot(clique\_graph3)



# plot all largest cliques  
# following code taken from R and Data Mining: Examples and Case Studies   
# by Yanchang Zhao, page 118  
florentine\_largest\_cliques <- largest\_cliques(florentine\_graph\_undirected)  
length(florentine\_largest\_cliques)

## [1] 3

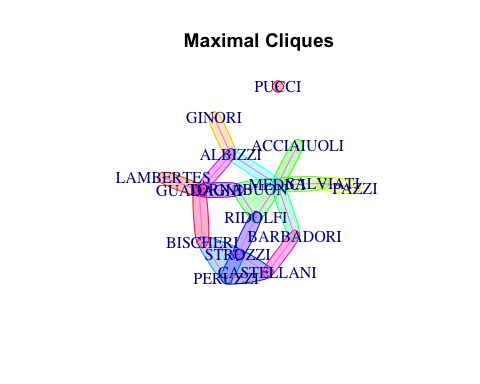
colorbar <- rainbow(length(florentine\_largest\_cliques) + 1)  
for (i in 1:length(florentine\_largest\_cliques)) {  
 V(florentine\_graph\_undirected)[florentine\_largest\_cliques[[i]]]$color <- colorbar[i+1]  
}  
plot(florentine\_graph\_undirected, mark.groups=florentine\_largest\_cliques,  
 vertex.size=0.3, vertex.label.cex=1, edge.color=rgb(1,0.5,1,1),  
 main = "Largest Cliques")



# plot all maximal cliques  
# following code taken from R and Data Mining: Examples and Case Studies   
# by Yanchang Zhao, page 117  
florentine\_max\_cliques <- max\_cliques(florentine\_graph\_undirected)  
length(florentine\_max\_cliques)

## [1] 16

colorbar <- rainbow(length(florentine\_max\_cliques) + 1)  
for (i in 1:length(florentine\_max\_cliques)) {  
 V(florentine\_graph\_undirected)[florentine\_max\_cliques[[i]]]$color <- colorbar[i+1]  
}  
plot(florentine\_graph\_undirected, mark.groups=florentine\_max\_cliques,  
 vertex.size=0.3, vertex.label.cex=1, edge.color=rgb(1,0.5,1,1),  
 main = "Maximal Cliques")



# Calculate PageRank of the vertices  
ranks\_of\_families <- page\_rank(florentine\_graph)$vector  
ranks\_of\_families

## ACCIAIUOLI ALBIZZI BARBADORI BISCHERI CASTELLANI GINORI   
## 0.03035390 0.07833886 0.04980296 0.06818000 0.06864374 0.03209700   
## GUADAGNI LAMBERTES MEDICI PAZZI PERUZZI PUCCI   
## 0.09742360 0.03060350 0.14437347 0.03569690 0.06720328 0.00990099   
## RIDOLFI SALVIATI STROZZI TORNABUON   
## 0.06888541 0.06069627 0.08722618 0.07057395

# most important or influential family  
which.max(ranks\_of\_families) # 9 is the index of the vertex

## MEDICI   
## 9

# Resources

* [Network visualization in R with the igraph package](https://rulesofreason.wordpress.com/2012/11/05/network-visualization-in-r-with-the-igraph-package/)
* [Making prettier network graphs with sna and igraph via @rbloggers](<http://www.r-bloggers.com/making-prettier-network-graphs-with-sna-and-igraph/>)
* [igraph R manual pages](http://igraph.org/r/doc/)

# References

The data, R code and lessons are based upon:

Graph theory/data structures:  
\* <http://math.tut.fi/~ruohonen/GT_English.pdf>  
\* <http://www.cl.cam.ac.uk/teaching/1011/PrincComm/slides-lpr/graph_theory_1-11.pdf>  
\* <http://www.researchgate.net/publication/228300013_Graph_Theory_A_Primer_for_Using_R_Visualization_Techniques_in_the_Applications_of_the_Adjacency_Matrix>  
\* <http://www.boost.org/doc/libs/1_59_0/libs/graph/doc/graph_theory_review.html>  
SNA:  
\* <http://files.meetup.com/1406240/sna_in_R.pdf>  
\* <http://www2.unb.ca/~ddu/6634/Lecture_notes/Lec1_intro_handout.pdf>  
\* <http://www.faculty.ucr.edu/~hanneman/nettext/>  
\* <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.1960&rep=rep1&type=pdf>  
\* <http://www.rdatamining.com/examples/social-network-analysis>

igraph:  
\* <http://statmath.wu.ac.at/research/friday/resources_WS0708_SS08/igraph.pdf>  
\* <http://blog.revolutionanalytics.com/2014/11/a-look-at-the-igraph-package.html>  
\* <http://www.r-bloggers.com/igraph-and-sna-an-amateurs-dabbling/>  
\* <http://www.r-bloggers.com/going-viral-with-rs-igraph-package/>  
\* <https://cran.r-project.org/web/packages/igraph/igraph.pdf>

Other famous packages for SNA:  
\* <http://www.r-bloggers.com/must-have-r-packages-for-social-scientists/>  
\* <https://cran.r-project.org/web/views/SocialSciences.html>  
\* <https://cran.r-project.org/web/packages/sna/sna.pdf>  
\* <https://cran.r-project.org/web/packages/RSiena/RSiena.pdf>  
\* <https://cran.r-project.org/web/packages/network/network.pdf>  
\* <https://www.bioconductor.org/packages/release/bioc/manuals/graph/man/graph.pdf>  
\* <http://www.statnet.org/>

In-depth SNA tutorials:  
\* <http://sna.stanford.edu/rlabs.php>  
\* <http://www.stats.ox.ac.uk/~snijders/sna_course.htm>  
\* <http://www.shizukalab.com/toolkits>

Sample projects:

* <http://www.orgnet.com/cases.html>

Motifs:  
\* <http://igraph.org/r/doc/motifs.html>  
\* <https://en.wikipedia.org/wiki/Network_motif>  
\* <http://www.cs.columbia.edu/4761/notes07/chapter8.2-topology.pdf>  
\* <https://sites.google.com/site/networkanalysisacourse/schedule/networkmotifs>  
Cliques:  
\* <http://igraph.org/r/doc/cliques.html>  
\* <http://faculty.ucr.edu/~hanneman/nettext/C11_Cliques.html>  
\* <https://courses.cs.washington.edu/courses/cse527/01au/oct25/oct25.html>  
\* ttp://www.mathcove.net/petersen/lessons/get-lesson?les=29  
\* <http://news.stanford.edu/news/2014/november/cliques-high-school-110514.html>  
PageRank:  
\* <http://igraph.org/r/doc/page_rank.html>  
\* <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>  
\* <http://www.cs.princeton.edu/~chazelle/courses/BIB/pagerank.htm>  
\* <http://www.stat.cmu.edu/~ryantibs/datamining/lectures/03-pr-marked.pdf>  
\* <http://smallstats.blogspot.com/2014/04/from-random-walks-to-personalized.html>  
\* <http://blog.revolutionanalytics.com/2014/12/a-reproducible-r-example-finding-the-most-popular-packages-using-the-pagerank-algorithm.html>  
\* Mining Massive Datasets on Coursera - Week 1, Videos 5 through 11 explain PageRank elegantly. The course maybe unavailable (or archived) by the time this module is out

Dataset:  
\* <http://moreno.ss.uci.edu/data.html#padgett>  
\* <http://home.uchicago.edu/~jpadgett/papers/unpublished/maelite.pdf>

Other SNA:  
\* <http://www.r-bloggers.com/experiments-with-igraph/>  
\* <http://cran.us.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf>