**Lab 8: Modeling Staff Communication Networks in R: ERGM**

Introduction. This lab focuses on developing true statistical models of clinic staff communication networks using R and the ergm package that is contained within statnet. To learn more about how to use ergm, please read through papers by Goodreau, O'Malley, and Morris.

Complete all tasks and answer all questions for full credit.

**Task 1.** Set-up. The data used here are configured as an R dataset. To be able to complete the following analyses, you need to install the statnet package and load the data.

Click on ‘Install package’ and choose a CRAN mirror. Any mirror will do, but I usually choose one from the US. Find the ‘statnet’ package and install it. This will make statnet available to you in the commands below.

Copy and paste the following commands into the R editor. Highlight and right click ‘Run line or selection’.

setwd("Your Own Directory Here")

library(statnet)

F2FData <- read.table("PrimaryCareF2FNet.txt", header=TRUE, row.names=1, check.names=FALSE)

F2FMatrix <- as.matrix(F2FData)

StaffAttr<-read.table("PrimaryCareAttributes.txt",header=TRUE,

stringsAsFactors=FALSE)

**Task 2.** Create the clinic network in R. There are 44 nodes in the staff dataset. A pair of clinic employees has a tie if they reported daily or more frequent face-to-face communication about patient care as part of the clinic network.

PrimaryCareF2FNet contains a number of node characteristics, including:

1. Job\_Category - (1-MD, 2-RN, 3-MA, 4-Office, 5-Lab)

2. MD/RN/MA/Office/Lab- binary indicators for job category

3. Years\_Clinic – number of years working at this clinic

4. FTE- Percent of full time employment

5. Female – employee gender

6. Job\_Satisfaction – job satisfaction scale score

Create the network matrix and attach the attribute data to the adjacency matrix:

F2FNet=network(F2FMatrix,matrix.type="adjacency",directed=TRUE)

F2FNet%v%'vertex.names'<- StaffAttr$SubjID

F2FNet%v%'Job\_Category'<- StaffAttr$Job\_Category

F2FNet%v%'Years\_Clinic'<- StaffAttr$Years\_Clinic

F2FNet%v%'FTE'<- StaffAttr$FTE

F2FNet%v%'Female'<- StaffAttr$Female

F2FNet%v%'Job\_Satisfaction' <- StaffAttr$Job\_Satisfaction

**Task 3:** The statnet package contains all the usual procedures for calculating the most commonly used node-level and graph-level network statistics. For example, the following commands calculate the density, clustering coefficient, and betweenness centralization of the ClinicNet communication network.

gden(F2FNet)

gtrans(F2FNet)

centralization(F2FNet, betweenness)

**Questions:**

**1: What are the density, transitivity, and betweenness centralization of the F2FNet communication network?**

**density - 0.2444**

**transitivity - 0.5197**

**centralization - 0.2547**

**Task 4.** The following syntax illustrates some of the options for plotting a network. Note that we are using color to indicate the job category, and the size of the node is scaled to its betweenness.

F2FBetw <- (betweenness(F2FNet)/max(betweenness(F2FNet))\*2)+0.5

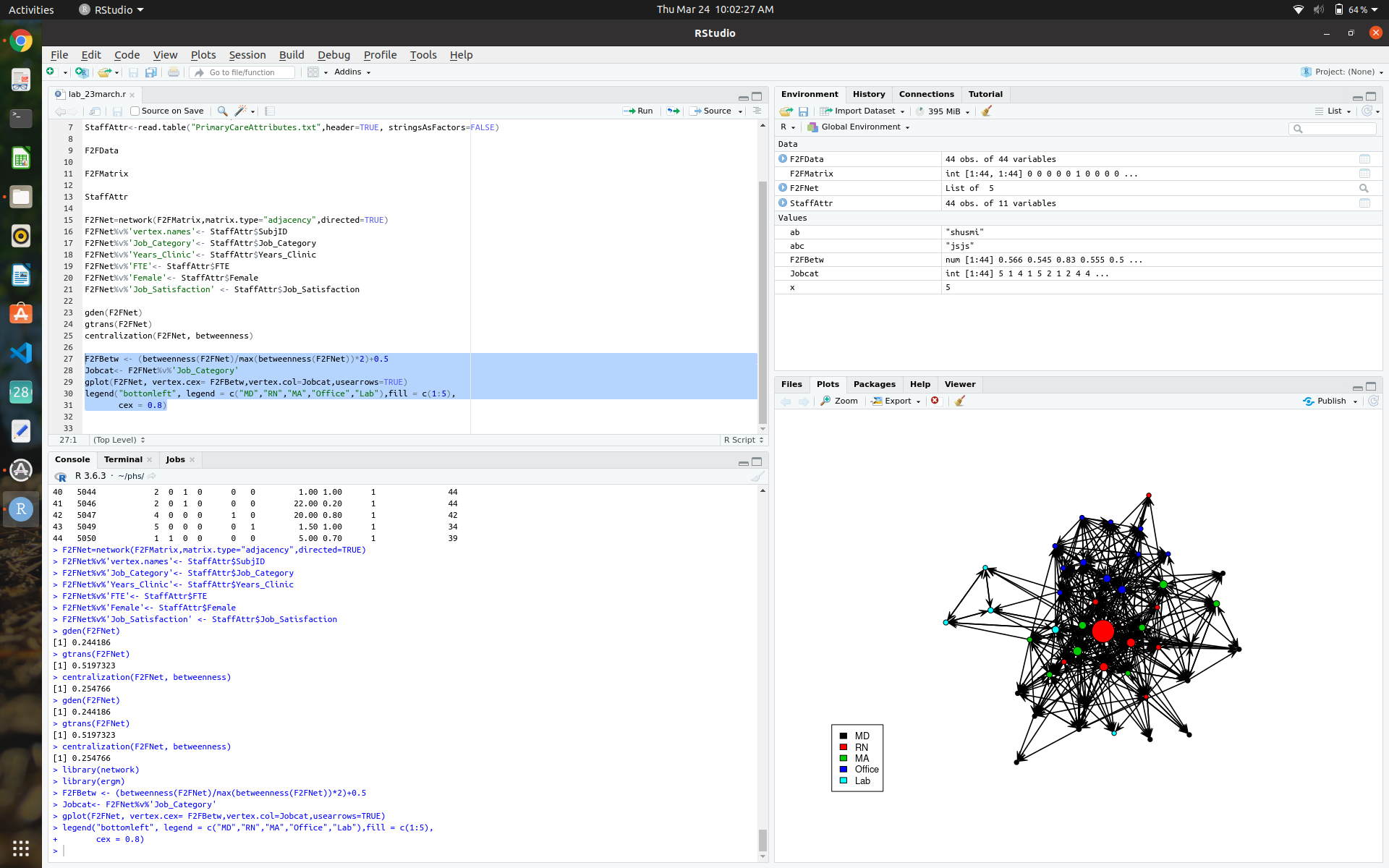
Jobcat<- F2FNet%v%'Job\_Category'

gplot(F2FNet, vertex.cex= F2FBetw,vertex.col=Jobcat,usearrows=TRUE)

legend("bottomleft", legend = c("MD","RN","MA","Office","Lab"),fill = c(1:5),

cex = 0.8)

Provide a screenshot of the communication network.



**Task 5.** We will use node characteristics, as well as other network relationships measured on the clinic network, to build a statistical model predicting collaboration ties between clinic employees. In addition to visualizing the network, it is important to explore the characteristics of the network as a whole and the attributes of the actors in the network to help guide network modeling. For example, the following commands show the pattern of communication ties within and between the five different types of job categories within the organization.

table (Jobcat)

mixingmatrix (F2FNet, "Job\_Category")

**Questions:**

**2: What do the graph and the mixing matrix show about intra- vs inter-job category communication?**

2,3,4 - nurses, medical assistants, office people have higher connectivity with other job categories.

4 (office staff) has lots of connections within themselves.

**Task 6.** ERGM modeling. In the rest of this lab, we will learn how to build, compare, and test the adequacy of statistical network models, using exponential random graph procedures contained in the ergm package. The first step in building a useful model is to construct a baseline null model. This is a random model with no predictors that will be used as a comparison to the more interesting models that follow. The important part of the model specification is in the first line. The rest of the call to ergm contains more technical specifications that can be ignored for now.

F2Fmodel.0 <- ergm(F2FNet ~edges, verbose=TRUE)

summary(F2Fmodel.0)

**Task 7.** Now examine the effect of adding job category to the network connection model:

F2Fmodel.1 <- ergm(F2FNet ~edges + nodefactor("Job\_Category"), verbose=TRUE)

summary(F2Fmodel.1)

**Questions:**

**3: Which categories of employees are more likely to send communication ties? Why do you think this is?**

**2,3,4 -** nurses, medical assistants, office people **very low p-value compared to the doctors.**

**We have seen earlier that these categories have many face to face communication ties.**

**Task 8.** Check the significance of the model addition:

anova(F2Fmodel.0, F2Fmodel.1)

Model 1: F2FNet ~ edges

Model 2: F2FNet ~ edges + nodefactor("Job\_Category")

Df Deviance Resid. Df Resid. Dev Pr(>|Chisq|)

NULL 1892 2622.9

1 1 519.51 1891 2103.4 < 2.2e-16 \*\*\*

2 4 172.23 1887 1931.1 < 2.2e-16 \*\*\*

we have added 4 degrees of freedom which has added significant predictive power.

**Task 9.** Now add the within job category matching variable ‘nodematch’ to the model:

F2Fmodel.2 <- ergm(F2FNet ~edges + nodefactor("Job\_Category") + nodematch("Job\_Category",diff=TRUE), verbose=TRUE)

summary(F2Fmodel.2)

anova(F2Fmodel.1, F2Fmodel.2)

**Questions:**

**4: Which employees are more likely to communicate to their colleagues within the same job category?**

**4,5 - office people and lab people**

**Task 10.** Next add job satisfaction and difference in number of years in the clinic:

F2Fmodel.3 <- ergm(F2FNet ~edges + nodefactor("Job\_Category") + nodematch("Job\_Category",diff=TRUE) + nodecov("Job\_Satisfaction") +

absdiff("Years\_Clinic"), verbose=TRUE)

summary(F2Fmodel.3)

anova(F2Fmodel.2, F2Fmodel.3)

Model 1: F2FNet ~ edges + nodefactor("Job\_Category") + nodematch("Job\_Category",

diff = TRUE)

Model 2: F2FNet ~ edges + nodefactor("Job\_Category") + nodematch("Job\_Category",

diff = TRUE) + nodecov("Job\_Satisfaction") + absdiff("Years\_Clinic")

Df Deviance Resid. Df Resid. Dev Pr(>|Chisq|)

NULL 1892 2622.9

1 10 812.69 1882 1810.2 < 2.2e-16 \*\*\*

2 2 37.50 1880 1772.7 7.211e-09 \*\*\*

**Questions:**

**5: Are network connections related to job satisfaction or number of years in the clinic? Explain why you might or might not expect this to be the case.**

**network connections related to the number of years in the clinic is connected to the number of ties. Similar longevity means more possibility of connection.**

**Task 11.** Goodness-of-fit of ERGM models. An important step for any statistical model analysis is determining how well the final (or candidate) model fits the data. We prefer simple models that do a good job of describing or predicting real-world behavior. This is more complicated and challenging for network analysis, where the objects being modeled are complex, and the statistical algorithms used for building models are iterative. Fortunately, ergm has a nice set of built-in tools for assessing model goodness- of-fit. Examine the results of the following:

gof(F2Fmodel.3, GOF = ~ distance + espartners + idegree + odegree + triadcensus, verbose=TRUE, burnin=1000, interval=1000)

**Questions:**

**6: How well is the model doing in terms of goodness of fit? Which aspects of the model are poorly fit?**

**Not that good. The Triadcensus, edgewise shared partner, geodesic distance fit poorly.**

**Task 12.** The model has some problems with goodness of fit. We can try to address these by adding structural parameters to the model. Add mutuality and a general transitivity parameter of ties:

F2Fmodel.4 <- ergm(F2FNet ~edges + mutual + gwesp(0.25,fixed=TRUE) + nodefactor("Job\_Category") + nodematch("Job\_Category",diff=TRUE)+ nodecov("Job\_Satisfaction") + absdiff("Years\_Clinic"), verbose=TRUE)

summary(F2Fmodel.4)

gof(F2Fmodel.4, GOF = ~ distance + espartners + idegree + odegree + triadcensus, verbose=TRUE, burnin=1000, interval=1000)

**Questions:**

**7: Did adding the structural parameters improve the model fit? What other model factors might be important to add to the model?**

**yes, not that much significant p values as the previous one. There might be a star structure or other structures that we may consider.**

**Task 13.** Summary. Summarize your findings about network connections in a clinic setting. Discuss additional hypotheses you would interested in testing with regard to clinic communication structure.

nurses and medical assistants are connected and clustered together, not too close and scattered a little bit. Office people are clustered together and the same goes for lab people. Physicians are scattered and peripheral. Also, nurses and office people have larger connections with other categories whereas lab and office people have larger connections within their own categories. I would be interested in seeing if gender of the persons have any connection with the number of ties. It might happen that connection is higher among female or maybe male people.