

## Lab 13: Diffusion

For this lab we will use R to test the diffusion of behaviors and evolution of ties in a longitudinal data set. We will examine ways to discern the statistical significance and model fit for a diffusion model.

Complete all tasks and answer all questions for full credit.

**NOTE:** This lab session uses the statistical program R. Before starting, you will need to open R and install the social network analysis packages in R which we will need for the exercises. The first step in R will be to change the directory (File->Change dir...). Set the directory to your folder in the shared drive.

### Dataset:

For this exercise, you will be working with the friendship data from a single school in the Add Health study (Class 126). The class is made up of 7<sup>th</sup> through 9<sup>th</sup> grade students in a middle school. The dataset for the exercise consists of the 50 students who participated.

**Task 1.** Open the R program. Install the RSIENA and statnet packages.

- a) Open a new session window in R.
- b) Copy and paste the following code into the session window.

```
setwd("PATHNAME FOR FOLDER")
install.packages("igraph")
library(igraph)
```

**Task 2.** Read in network data to R.

- a) Class126.txt is an adjacency matrix for the Add Health class 126 data at time 1. Class126\_t2.txt is the adjacency matrix for class 126 at time 2, one year after time 1. Attr126.txt is a matrix of attributes for class 126, including the behavior variables 'w1anyalc' and 'w2anyalc' measures of whether the adolescent has initiated alcohol use at time 1 or time 2.
- b) Read the data table into R with the following commands:

```
friend.data.w1 <- as.matrix(read.table("class126.txt"))
friend.data.w2 <- as.matrix(read.table("class126_t2.txt"))
attr126<-read.table("attr126.txt",header=TRUE,stringsAsFactors=FALSE)
```

```
# Set the wave 2 friendship data to missing if the wave 2 data are missing
for (i in 1:nrow(attr126)) {if (attr126[i,50] > 98) {friend.data.w2[i,]=9}}
friend.data.w1[friend.data.w1 %in% c(6,9)] <- NA
friend.data.w2[friend.data.w2 %in% c(6,9)] <- NA
```

```
# Create the wave 1 and wave 2 alcohol frequency matrix
alcinit<-cbind(attr126[,60],attr126[,51])
```

```

alcinit[alcinit %in% c(99)] <- NA

# Check dimensions to verify that adjacency and attribute matrices match
dim(friend.data.w1)
dim(alcinit)

# Plot the wave 1 connections
g1<-graph_from_adjacency_matrix(friend.data.w1)
plot(g1)

```

**Task 3.** Install netdiffuseR and create the netdiffuseR data object.

```

install.packages("netdiffuseR")
library(netdiffuseR)

# create the 'time of adoption' toa vector

# attr126[,60] indicates student alcohol drinking status at wave 1
attr126[,60]

# attr126[,51] is an indicator of drinking at wave 2
attr126[,51]

# The following code creates a vector that is valued 1 if student has started
# drinking prior to wave 1, valued 2 if student was nondrinker at wave 1 and
# drinker at wave 2, and valued NA if the student was a nondrinker
# throughout. Students lost to follow-up at wave 2 had alcohol use value 99
# at wave 2. The statements below recode toa to 1 if such a student recorded
# drinking at wave 1 and NA otherwise.

toa <- 2*attr126[,51] - attr126[,60]
toa[toa %in% c(197)] <- 1
toa[toa %in% c(198)] <- 0
toa[toa %in% c(0)] <- NA
toa

ah_pvt <- array(attr126[,36])
age <- attr126[,38]
male <- attr126[,40]
parentdrkfrq <- attr126[,53]
tobacco <- attr126[,56]

diffnet <- as_diffnet(friend.data.w1, toa)
diffnet

X <- cbind(ah_pvt = ah_pvt, age=age, male=male,
parentdrkfrq=parentdrkfrq, tobacco=tobacco)

dn <- new_diffnet(friend.data.w1, toa=toa, vertex.static.attrs = X)

```

```
dn[["cohesive_expo"]] <- cbind(NA, exposure(dn)[-nslices(dn)])
dn[["adopt"]] <- dn$cumadopt
```

```
dn
```

**Task 4.** Examine the diffusion plot.

```
summary(diffnet)
plot_diffnet(diffnet)
```

### Questions:

Period	Adopters	Cum Adopt. (%)	Hazard Rate	Density	Moran's I (sd)
1	12	12 (0.24)	-	0.05 0.01 (0.07)	
2	12	24 (0.48)	0.32	0.05 0.14 (0.07) **	

---

Left censoring : 0.24 (12)  
 Right censoring : 0.52 (26)  
 # of nodes : 50

**1. What is the prevalence of adolescent alcohol use initiation at wave 1?**

12 people

**2. How many adolescents in the school initiate alcohol use between wave 1 and wave 2?**

12

**Task 5.** Run the netdiffuseR null model.

```
# Generating the data and running the null model
dat <- as.data.frame(dn)
ans <- glm(adopt ~ cohesive_expo,
  data = dat,
  family = binomial(link="probit"),
  subset = is.na(toa) | (per <= toa))
```

```
summary(ans)
```

### Questions:

**3. What is the significance of alcohol use exposure in the null model? Interpret the findings.**

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7607    0.2846  -2.673  0.00752 **
cohesive_expo  1.3247    0.8458   1.566  0.11729
```

The exposure is not significant. maybe we need more sample and more time period to find significance.

**Task 6.** Add additional model effects. Rerun the netdiffuseR model.

```
# Generating the data and running the full model
dat <- as.data.frame(dn)
ans <- glm(adopt ~ cohesive_expo + ah_pvt + male + age + parentdrkfrq +
tobacco,
          data = dat,
          family = binomial(link="probit"),
          subset = is.na(toa) | (per <= toa))
summary(ans)
```

### Questions:

**4. Which effects are significant in the subsequent model? Interpret the findings.**

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.172613   5.466596  -0.580   0.562
cohesive_expo  1.416168   0.921847   1.536   0.124
ah_pvt        0.006487   0.016611   0.391   0.696
male         -0.158728   0.500265  -0.317   0.751
age          0.109623   0.373340   0.294   0.769
parentdrkfrq  0.134673   0.203519   0.662   0.508
tobacco       0.382613   0.511077   0.749   0.454
```

There is no significance in any of the effects. So we can say that the data does

**Task 7. Summary write-up.** Write up a short description of your netdiffuseR analysis of students in the Class126 dataset. Suggest directions for future research.

IN wave 2, some started to smoke who were not smoker in first wave. Interestingly, some of them were surrounded by some people who were already drinkers in the first wave. We have also seen that according to the data, exposure of alcohol use does not have any significance on the null model.

