

# Problem Set 2 (CSS)

Due by 05/12/25

## Data:

The data we use for this exercise set stems from the Teenage Friends and Lifestyle Study (Bush, West & Michell 1997, Michell and West 1996, Pearson and Michell 2000, Pearson and West 2003). [For more information and online download see here.](#)

It aims to identify processes by which attitudes towards smoking and smoking behavior itself change over early to mid adolescence, focusing particularly on the stage of experimentation, and the transition from experimental to regular smoking. It followed a single cohort of students from their 2nd to 4th year, with a focus on smoking attitudes and behavior.

Friendship data were collected at *three time points*, each represented by a 160×160 matrix. Rows indicate the ego (respondent), columns the alter (peer), and values denote the relationship: 1 = best friends, 2 = friends, 0 = no tie, and 10 = structural zeros (e.g., if a student had not yet joined the school). We retain only reported friendships (values 1 and 2) and exclude structural zeros.

```
library(igraph)

#exchange with your path
load("Data/Glasgow_data/glasgow_data/Glasgow_data.RData")

# load the data
load(file = "Data/Glasgow_data/Glasgow-friendship.RData")

# recode the structural zero's to NA
friendship.1[friendship.1 == 10] <- NA
friendship.2[friendship.2 == 10] <- NA
friendship.3[friendship.3 == 10] <- NA

# Recode regular friends (2) to 0.5 to reflect weaker tie strength
friendship.1[friendship.1 == 2] <- 0.5
```

```

friendship.2[friendship.2 == 2] <- 0.5
friendship.3[friendship.3 == 2] <- 0.5

friendmat_1 <- friendship.1
friendmat_2 <- friendship.2
friendmat_3 <- friendship.3

friends <- list("wave_1" = friendmat_1,
               "wave_2" = friendmat_2,
               "wave_3" = friendmat_3)

#complete cases only
dat_clean <- dat[complete.cases(dat[, c("age",
                                         "male",
                                         "money_1",

                                         "weed_1",
                                         "alc_1",
                                         "tob_1",
                                         "money_2",

                                         "weed_2",
                                         "alc_2",
                                         "tob_2",
                                         "money_3",

                                         "weed_3",
                                         "alc_3",
                                         "tob_3"
                                         )
                                         ],
               ),
               ]

friends <- list("wave_1" = friendmat_1[dat_clean$idstud,
                                         dat_clean$idstud],
               "wave_2" = friendmat_2[dat_clean$idstud,
                                         dat_clean$idstud],

```

```

        "wave_3" = friendmat_3[dat_clean$idstud,
                                dat_clean$idstud])

save(dat_clean,
     friends,
     file = "data/Glasgow_data/glasgow_data.RData")

graph1 <- graph_from_adjacency_matrix(adjmatrix = friends$wave_1,
                                     mode = "directed",
                                     weighted = TRUE)

graph2 <- graph_from_adjacency_matrix(adjmatrix = friends$wave_2,
                                     mode = "directed",
                                     weighted = TRUE)

graph3 <- graph_from_adjacency_matrix(adjmatrix = friends$wave_3,
                                     mode = "directed",
                                     weighted = TRUE)

# Add attributes to the graphs at each time point

V(graph1)$age <- dat_clean$age
V(graph1)$gender <- dat_clean$male
V(graph1)$money <- dat_clean$money_1
V(graph1)$smoking <- dat_clean$tob_1
V(graph1)$weed <- dat_clean$weed_1
V(graph1)$alc <- dat_clean$alc_1

V(graph2)$age <- dat_clean$age
V(graph2)$gender <- dat_clean$male
V(graph2)$money <- dat_clean$money_2
V(graph2)$smoking <- dat_clean$tob_2
V(graph2)$weed <- dat_clean$weed_2
V(graph2)$alc <- dat_clean$alc_2

V(graph3)$age <- dat_clean$age
V(graph3)$gender <- dat_clean$male
V(graph3)$money <- dat_clean$money_3
V(graph3)$smoking <- dat_clean$tob_3
V(graph3)$weed <- dat_clean$weed_3

```

```
V(graph3)$alc <- dat_clean$alc_3

graph1 <- igraph::simplify(graph1)
graph2 <- igraph::simplify(graph2)
graph3 <- igraph::simplify(graph3)
```

Variable	Description
<i>alc</i>	<b>Alcohol consumption:</b> 1 = none 2 = seldomly 3 = once a month 4 = once a week 5 = more than once a week
<i>smoking</i>	<b>Smoking (tobacco)</b> 1 = never 2 = occasionally 3 = regularly
<i>weed</i>	<b>Smoking (cannabis)</b> 1 = never 2 = tried it once 3 = occasionally 4 = regularly
<i>money</i>	Numeric value of how much pocket money the students had
<i>gender</i>	<b>Binary sex</b> 1 = boys 2 = girls
<i>age</i>	Age on Jan 1, 1995 (in years with one decimal), recoded from birthdates

Download the data from the website and run the code above in order to load the **igraph** objects to your R-environment.

---

## 1 Analyze the network in terms of the distribution of social capital:

**1.1** Identify cutpoints, bridges, and local bridges *(1 Point)*

**1.2** Are there forbidden triads in the sense of Granovetter? *(1 Point)*

**1.3** Are all local bridges weak ties? *(1 Point)*

**1.4** Calculate Burt's constraint measure for each node. Interpret the distribution. (2 Points)

**1.5** Why does Burt propose the constraint measure instead of traditional centrality measures? (2 Points)

---

## **2 Analyze graph3 with respect to segregation and cohesion**

**2.1** Calculate segregation based on gender and pocket money. Interpret the results. (2 Points)

**2.2** Create subnetworks for boys and girls. Which one is more cohesive? (1 Point)

**2.3** Which group is more cohesive: alcohol users or non-users? What can we make of this? (1 Point)

**2.4** Create a random network of the same size. Compare density, dyad census, and clustering. Reflect: Is this comparison helpful? (2 Points)

```
set.seed(1)

rand <- erdos.renyi.game(vcount(graph3),
                        ecount(graph3),
                        type = "gnm",
                        directed = TRUE
                        )
```

---

**3** Colin Cluster is a teacher at the respective school in Glasgow. Over time he observes the following:

1. The “cool kids” are more likely to smoke than others.
2. The older the students get, the more likely they are to smoke.
3. Over time, smokers seem to form tighter friend groups, than non smokers.

However, he's not quite sure how to verify these observations using appropriate measures. Can you help him?

**3.1** Choose a measure to identify *cool kids*. Justify your choice. (2 Points)

**3.2** Are the *cool kids* more likely to smoke at time  $t1$ ? (1 Point)

**3.3** Plot the network at time  $t_1$ , showing information on the relationship strength, the smoking behavior of the persons, gender and their ‘coolness factor’. (3 Points)

**3.4** Plot networks for  $t_1$ ,  $t_2$  and  $t_3$ . From the visual analysis, does smoking lead to stronger clustering over time? (3 Points)



**HE COULDN'T, BUT YOU CAN**