

research_report_01

```
library(tidyverse)
library(igraph)
library(netseg)

all_edges <- read_table(here::here("data", "High-School_data_2013.csv"))

sub_edges <- read_table(here::here("data", "Contact-diaries-network_data_2013.csv"))

attributes <- read_tsv(here::here("data", "HighSchool2013_metadata.txt"))

filtered_edges <- filter(all_edges, i %in% sub_edges$i | j %in% sub_edges$j)

edges <- filtered_edges |>
  distinct(i, j, .keep_all = T) |>
  relocate(t, .after = Cj)

net <- graph_from_data_frame(d = edges,
                             directed = F,
                             vertices = attributes)

isolated <- which(degree(net) == 0)
net <- delete_vertices(net, isolated)

attributes <- filter(attributes, i %in% filtered_edges$i | i %in% filtered_edges$j)
```

This is the high school interaction data I am using for my second year paper. Specifically, the network I am analyzing here is of the monitored face-to-face interactions. However, because the project is largely focused on comparing the physical interaction data to the diary-log data, I have filtered down the network to only show the nodes who provided the diary data and the individuals they reference in their diary-log. In this network, a tie is registered if the individuals were in face-to-face proximity for at least 20 seconds. The data logged every interaction as a

distinct tie, so I also narrowed down the data to show a maximum of one tie per dyad (I am working on turning the quantity of interactions into a tie weight). As most, but not all (86%) of the students in the school wore the physical sensors, and only 120 (approx. 1/3) provided diary logs, this network is not a perfectly accurate representation of the social space, but it does accurately represent the social space of the individuals of interest for this project.

```
cols <- ifelse(attributes$Gi == "F", "lightpink", "lightblue")
V(net)$color <- cols

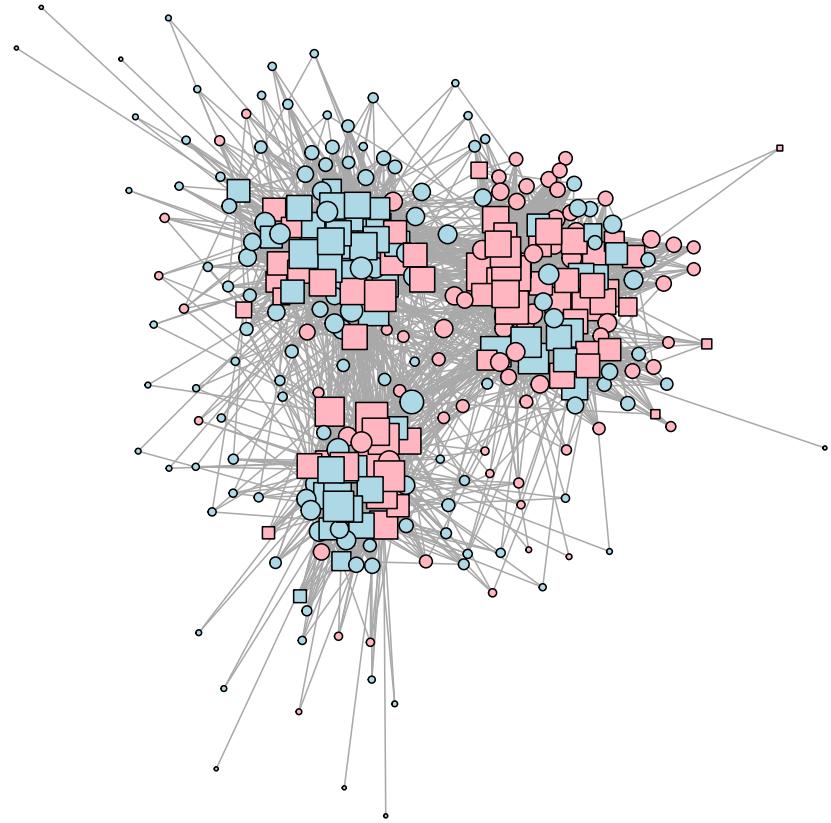
V(net)$shape <- ifelse(attributes$i %in% sub_edges$i, "square", "circle")

deg <- degree(net, mode = "all")

layout <- layout_with_fr(net)

par(mar = c(0, 0, 0, 0) + 0.1)

plot(net,
      vertex.label = NA,
      vertex.size = sqrt(deg),
      layout = layout)
```



```
edge_density(net)
```

```
[1] 0.06924191
```

```
transitivity(net, type = "global")
```

```
[1] 0.3671614
```

```
mixingm(net, "Gi")
```

		ego		
alter		F	M	Unknown
F		1048	1412	49
M		0	921	60
Unknown		0	0	0

```
mixingm(net, V(net)$shape)
```

		ego	
alter		circle	square
circle		0	2365
square		0	1125

The network is visualized with nodes sized proportional to their degree centrality, colored based on gender (pink = female, blue = male), and shaped based on if they provided diary-log data (square = yes, circle = no).

There are 3,490 ties present among the 318 nodes, for a density of 0.07, suggesting that while the network looks dense due to the sheer number of interactions, it is relatively quite sparse. The network seems to largely exist in 3 main clusters—one of which seems to have more women than men, another more men than women, and the third a fairly even split between the two. When actually calculating, there are 1,048 F-F interactions, 921 M-M interactions, and 1,412 mixed-gender interactions, suggesting that interactions do not lean strongly homogenous or heterogenous overall (though I suspect the results may be different if we separated them based on the observed sub-groups/communities). Most of the degree centrality seems to be associated with individuals who provided diary-log data—which makes sense here, as the only ties are those involving those individuals. 32% of the ties in the network are between two individuals who both provided diary data, while 68% of dyads who interacted at least once involve someone who did not report their contact diary. The manufactured absence of non-diary to non-diary ties makes triad analysis nonsensical, but the focus of this project is on 1-degree-out ego-nets within the broader structure.