Dutch School Example

Jeremy Foote

3/24/2020

## Example

I’m doing an in-depth example, primarily to give you examples of the sorts of things that you could do with the data, and so you’ll have code to work with.

This is above and beyond what I would expect for this assignment. That being said, I encourage you to be ambitious in how you approach this assignment. This is at the heart of seeing how networks can help us to understand things about the world.

## My question

Do stronger friendships influence drinking behavior more?

While these are friendships ties, they are not all reciprocal Sometimes A nominates B as a friend, but B doesn’t nominate A. We might expect that when an edge is reciprocal (both people nominate each other), then the friendship is stronger, and therefore their behavior is more likely to be the same.

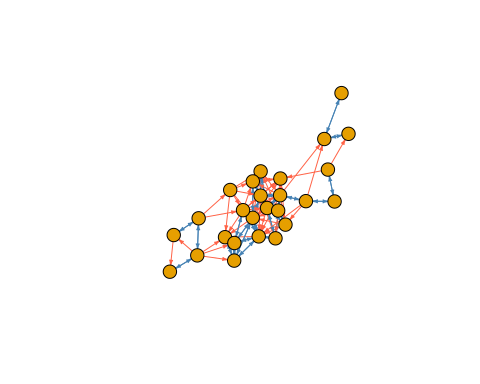
So, I want to find out if that’s true. My goal is to highlight drinking behavior and reciprocal ties.

First, I color the reciprocal ties. The which\_mutual() function tells me which edges are reciprocal, and I make them a different color.

library(igraph)  
library(tidygraph)  
library(ggraph)  
load(url('https://github.com/jdfoote/Communication-and-Social-Networks/raw/master/activities/school\_graph.Rdata'))  
  
which\_mutual(friend\_net)

## [1] TRUE TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE  
## [13] TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE  
## [25] TRUE FALSE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE  
## [37] TRUE FALSE FALSE TRUE FALSE TRUE TRUE TRUE TRUE FALSE FALSE FALSE  
## [49] TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE TRUE TRUE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE TRUE TRUE  
## [73] TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE  
## [85] FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE TRUE TRUE TRUE  
## [97] TRUE TRUE TRUE TRUE FALSE TRUE FALSE FALSE FALSE TRUE TRUE FALSE  
## [109] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE

E(friend\_net)[which\_mutual(friend\_net)==F]$color = 'tomato'  
E(friend\_net)[which\_mutual(friend\_net)==T]$color = 'steelblue'  
E(friend\_net)$arrow.size = .25 # Setting the arrow size much smaller  
V(friend\_net)$label = NA # Removing the labels since they are meaningless  
  
plot(friend\_net)

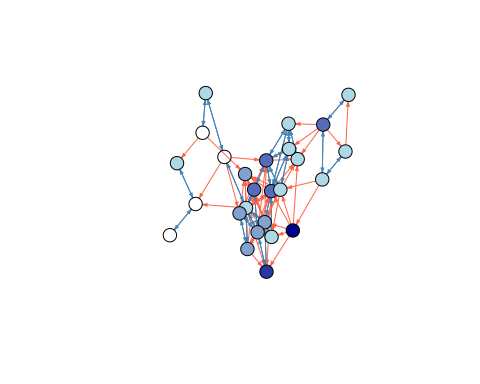


Now, I want to add colors for drinking behavior to the vertices.

# Make a list of colors  
# This gives us a function that will produce a set of colors from light blue to dark blue  
colors\_func = colorRampPalette(c("lightblue","darkblue"))  
# But we need to figure out how many colors we need.  
# We'll make one color for each unique item in alcohol\_use.  
# If we look at it, we can see how many there are.   
vertex.attributes(friend\_net)$alcohol\_use

## [1] 4 2 1 1 1 1 3 1 3 2 2 5 3 3 1 NA 1 1 NA 2 NA 1 NA 1 1  
## [26] 2

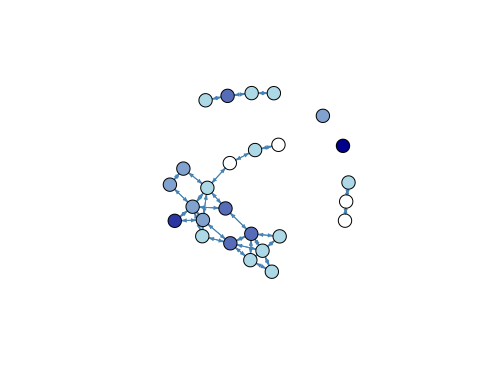
# It looks like alcohol\_use goes from 1 to 5, which is perfect.   
# We can just use the values as an index to the colors list.  
# First, we make the colors list - this will make 5 different colors,  
# going from light to dark blue  
color\_list = colors\_func(5)  
# Next, we set the colors of the nodes based on alcohol\_use  
V(friend\_net)$color = color\_list[V(friend\_net)$alcohol\_use]  
plot(friend\_net)



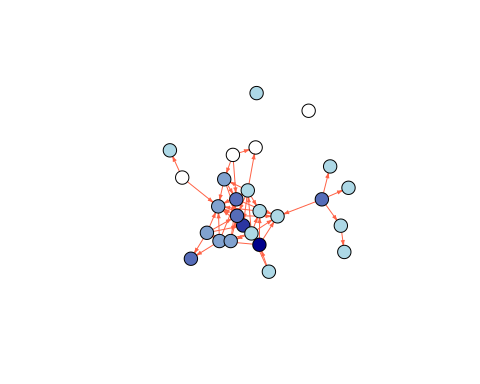
This is a pretty good plot, and helps us to get information about the question we were asking. One surprising finding is that there are a lot of non-mutual edges. Lots of people considered the other person a friend, but not vice versa. There isn’t clear evidence from this plot that mutual friends are more likely to have similar behavior, but we can try to draw that out a bit.

First, we can create plots that are a subset. The following plots show only reciprocal or non-reciprocal edges.

recip = delete\_edges(friend\_net, E(friend\_net)[which\_mutual(friend\_net)==F])  
non\_recip = delete\_edges(friend\_net, E(friend\_net)[which\_mutual(friend\_net)==T])  
plot(recip)



plot(non\_recip)

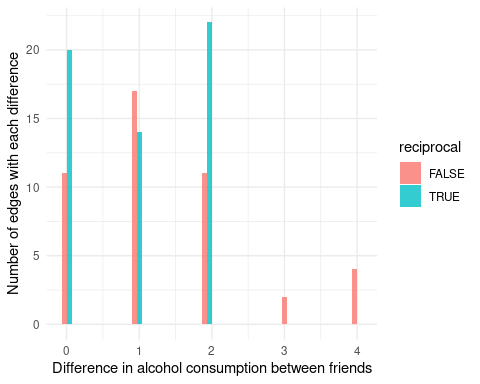


By splitting up the graphs, we can start to see some patterns. The mutual friend graph does appear to be a bit more clustred, eith the V23, V18, V4, V26 group all being non-drinkers and all being friends, for example.

Finally, if you really wanted to get fancy, you could try to plot information about the relationship between tie type and similarity in behavior.

I’m doing this based on the answer to a question that I asked on Stack Overflow [here](https://stackoverflow.com/questions/60817618/how-do-you-color-the-edges-of-a-graph-based-on-whether-they-are-homophilous-in-i/60818946#60818946), which uses the tidygraph library.

# Uncomment the following line and run it one time if you need to install tidygraph  
# install.packages('tidygraph')  
library(tidygraph)  
library(ggplot2)  
  
# The following code creates 2 new edge attributes. The first is 'difference', which is the  
# absolute value difference in alcohol\_use from the two nodes of each edge  
# The second is 'reciprocal', which just says if an edge is reciprocal. We already  
# figured this out when we colored the edges, so I just use the color attribute  
friend\_net = friend\_net %>%  
 as\_tbl\_graph() %>%  
 activate(edges) %>%  
 mutate(difference = abs(.N()$alcohol\_use[from] - .N()$alcohol\_use[to]),  
 reciprocal = color=='steelblue'  
 )  
  
# Next, we get the edges into a dataframe, so that we can analyze it  
  
df = as.data.frame(edge.attributes(friend\_net))  
  
# Next, I'm going to plot the distribution of difference for reciprocal and non-reciprocal edges  
  
df %>%  
 ggplot(aes(x=difference, fill=reciprocal)) +   
 geom\_histogram(alpha=0.8, position="dodge") + theme\_minimal() +   
 xlab('Difference in alcohol consumption between friends') +   
 ylab('Number of edges with each difference')



This last analysis is far beyond anything that we have done in class, but the example code might give you some hints about how you could do something similar.