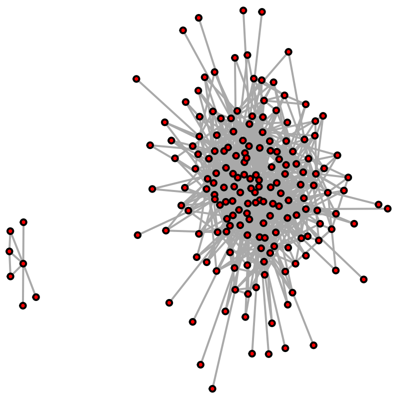
Homework 3 Assignment

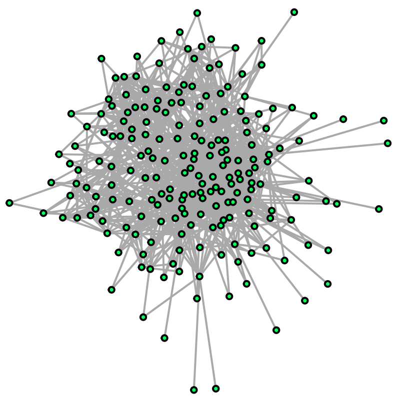
Each question is 10 points. Bonus is up to 5 points.

Starter code

```r
## microfinance network
## data from BANERJEE, CHANDRASEKHAR, DUFLO, JACKSON 2012

## data on 8622 households
hh <- read.csv("microfi_households.csv", row.names="hh")
hh$village <- factor(hh$village)

## We'll kick off with a bunch of network stuff.
## This will be covered in more detail in lecture 6.
## get igraph off of CRAN if you don't have it
## install.packages("igraph")
## this is a tool for network analysis
## (see http://igraph.sourceforge.net/)
library(igraph)
edges <- read.table("microfi_edges.txt", colClasses="character")
## edges holds connections between the household ids
hhnet <- graph.edgelist(as.matrix(edges))
hhnet <- as.undirected(hhnet) # two-way connections.

## igraph is all about plotting.
V(hhnet) ## our 8000+ household vertices
## Each vertex (node) has some attributes, and we can add more.
V(hhnet)$village <- as.character(hh[V(hhnet),'village'])
## we'll color them by village membership
vilcol <- rainbow(nlevels(hh$village))
names(vilcol) <- levels(hh$village)
V(hhnet)$color = vilcol[V(hhnet)$village]
## drop HH labels from plot
V(hhnet)$label=NA

# graph plots try to force distances proportional to connectivity
# imagine nodes connected by elastic bands that you are pulling apart
# The graphs can take a very long time, but I've found
# edge.curved=FALSE speeds things up a lot.  Not sure why.

## we'll use induced.subgraph and plot a couple villages
village1 <- induced.subgraph(hhnet, v=which(V(hhnet)$village=="1"))
village33 <- induced.subgraph(hhnet, v=which(V(hhnet)$village=="33"))

# vertex.size=3 is small.  default is 15
plot(village1, vertex.size=3, edge.curved=FALSE)
```



```r
plot(village33, vertex.size=3, edge.curved=FALSE)

######  now, on to your homework stuff

library(gamlr)
```



```r
## match id's; I call these 'zebras' because they are like crosswalks
zebra <- match(rownames(hh), V(hhnet)$name)

## calculate the `degree' of each hh:
##  number of commerce/friend/family connections
degree <- degree(hhnet)[zebra]
names(degree) <- rownames(hh)
degree[is.na(degree)] <- 0 # unconnected houses, not in our graph

## if you run a full glm, it takes forever and is an overfit mess
# > summary(full <- glm(loan ~ degree + .^2, data=hh, family="binomial"))
# Warning messages:
# 1: glm.fit: algorithm did not converge
# 2: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Question 1

I'd transform degree to create our treatment variable d. What would you do and why?

Question 2

Build a model to predict d from x, our controls. Comment on how tight the fit is, and what that implies for estimation of a treatment effect.

Question 3

Use predictions from Q2 in an estimator for effect of d on loan.

Question 4

Compare the results from Q3 to those from a straight (naive) lasso for loan on d and x. Explain why they are similar or different.

Question 5

Bootstrap your estimator from Q3 and describe the uncertainty.

Bonus

Can you think of how you'd design an experiment to estimate the treatment effect of network degree?