The research on dependancy between education level and employee's loyalty

Ilya Pavlov

I choose dependent variable Loyalty which represents mean of answers on questions 1, 2, 3 and mean of answer on reversed questions 18, 19, 20

```
ocb_att<-read.csv("ocb_attr.csv",header=TRUE,sep=";")
```

```
q18_rev<-ifelse(ocb_att$Q18==0,1,0)
q19_rev<-ifelse(ocb_att$Q19==0,1,0)
q20_rev<-ifelse(ocb_att$Q20==0,1,0)
```

```
loyalty<-rowMeans(cbind(ocb_att$Q1,ocb_att$Q2,ocb_att$Q2,q18_rev, q19_rev, q20_rev),na.rm=TRUE)
```

I select SupportNet as I hypothesize that if employee support their colleagues it could impact positively on his or her loyalty to the firm.

```
SupportNet<-read.csv("SupportNet.csv",header=TRUE,sep=";")
```

```
support_mat<-as.matrix(SupportNet)
```

```
library(network)
```

```
## network: Classes for Relational Data
## Version 1.13.0 created on 2015-08-31.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##                     Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Martina Morris, University of Washington
##                     Skye Bender-deMoll, University of Washington
##  For citation information, type citation("network").
##  Type help("network-package") to get started.
```

```
library(sna)
```

```
## Loading required package: statnet.common
```

```
##
## Attaching package: 'statnet.common'
```

```
## The following object is masked from 'package:base':
##
##     order
```

```
## sna: Tools for Social Network Analysis
## Version 2.4 created on 2016-07-23.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.
```

```
support_network<-as.network(support_mat, directed=TRUE)
geo.dist<-geodist(support_network)
indegree <- degree(support_network, gmode = 'digraph',
diag = FALSE, cmode = 'indegree',
```

```r
rescale = FALSE, ignore.eval = FALSE)
outdegree <- degree(support_network, gmode = 'digraph',
diag = FALSE, cmode = 'outdegree',
rescale = FALSE, ignore.eval = FALSE)
degree.f <- degree(support_network, gmode = 'digraph',
diag = FALSE, cmode = 'freeman',
rescale = FALSE, ignore.eval = FALSE)
between <- betweenness(support_network, gmode = 'digraph',
diag = FALSE, cmode = 'directed')
close <- closeness(support_network, gmode = 'digraph',
diag = FALSE, cmode = 'directed', rescale = FALSE)
eigen <- evcent(support_network, gmode = 'digraph',
diag = FALSE, rescale = FALSE)
```

```r
detach(package:sna)
detach(package:network)
library(igraph)
```

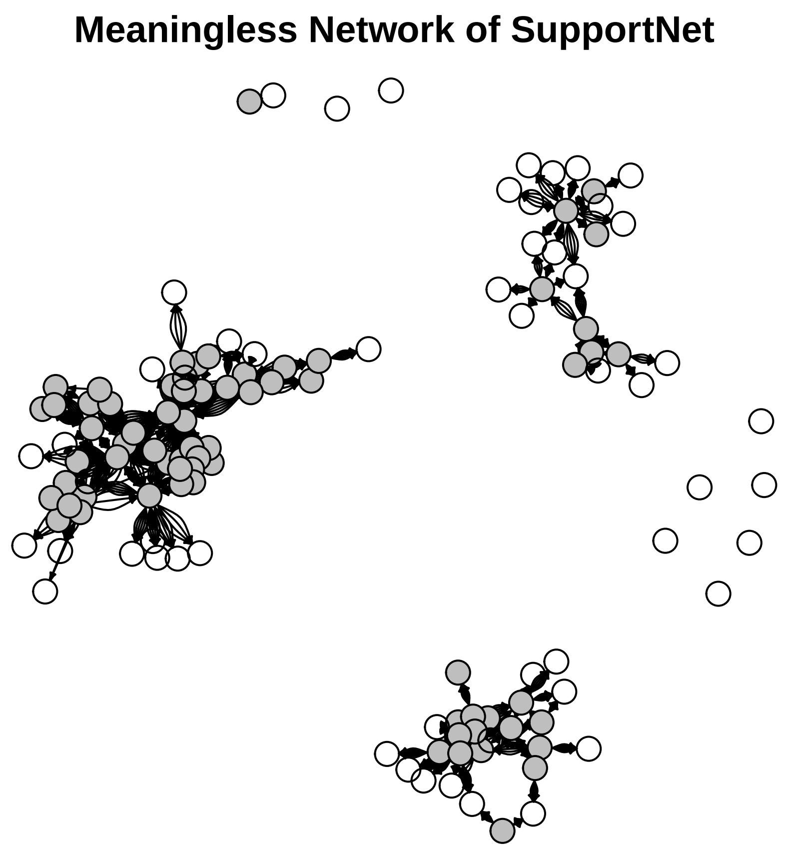
```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```r
all_graph<-graph_from_adjacency_matrix(support_mat)
```

```r
names<-ocb_att$Name
sex<-ocb_att$Sex
age<-ocb_att$Age
tenure<-ocb_att$WorkTitleYear+ocb_att$WorkTitleMonth/12
tenure_org<-ocb_att$WorkOrgYear+ocb_att$WorkOrgMonth/12
tenure_sup<-ocb_att$RepSupYear+ocb_att$RepSupMonth/12
ed1<-ifelse(ocb_att$Education==3,1,0)
ed2<-ifelse(ocb_att$Education==4,1,0)
gender_vector<-vector()
for(i in 1:122){
for(j in 1:68){
if(V(all_graph)$name[i]==names[j]){
gender_vector[i]<-sex[j]
break;}
else{gender_vector[i]<-NA}
}
}
gender_vector
```

```
##   [1]  3  3  3  3  3  3  3  3  3  3  3  3  3  2  3  3  3  3  3  3  3  3  3
##  [24]  3  3  3  3  3  3  2  2  3  2  2  3  3  3  3  3  3  3  3  3  2  3  2
##  [47]  2  2  2  2  3  3  2  3  3  3  3  2  3  3  2  2  3  2  3 NA NA NA NA
##  [70] NA NA NA NA NA NA NA NA  3 NA NA NA NA NA NA NA NA NA  2 NA NA NA NA
##  [93] NA NA NA NA NA NA NA NA NA NA NA NA  2 NA NA NA NA NA NA NA NA NA NA
## [116] NA NA NA NA NA NA NA
```

```
all_graph<-set_vertex_attr(all_graph, 'gender', value=c(gender_vector))
colors<-ifelse(gender_vector==1,"palevioletred",
ifelse(gender_vector==0,"blueviolet","gray"))
par(mar=c(0,0,1,0))
plot(all_graph, vertex.size=6.5, vertex.color=colors,
edge.arrow.size=.2, edge.color="black",
vertex.label=NA, main="Meaningless Network of SupportNet")
```




```
v_geodist<-vector()
v_indegree<-vector()
v_outdegree<-vector()
v_degree.f<-vector()
v_between<-vector()
v_close<-vector()
v_eigen<-vector()
for(i in 1:122){
for(j in 1:68){
if(V(all_graph)$name[i]==names[j]){
v_geodist[j]<-geo.dist[i]
v_indegree[j]<-indegree[i]
v_outdegree[j]<-outdegree[i]
v_degree.f[j]<-degree.f[i]
v_between[j]<-between[i]
v_close[j]<-close[i]
v_eigen[j]<-eigen[i]
break;}
```

```
else{}
}
}
```

```
cor_mat<-cbind(loyalty,sex,age,tenure,tenure_sup,tenure_org,v_indegree,v_outdegree)
```

```
## Warning in cbind(loyalty, sex, age, tenure, tenure_sup, tenure_org,
## v_indegree, : number of rows of result is not a multiple of vector length
## (arg 7)
```

```
cor(cor_mat)
```

```
##             loyalty         sex       age tenure tenure_sup tenure_org
## loyalty           1          NA        NA     NA         NA         NA
## sex              NA  1.00000000 0.1894667     NA         NA         NA
## age              NA  0.18946670 1.0000000     NA         NA         NA
## tenure           NA          NA        NA      1         NA         NA
## tenure_sup       NA          NA        NA     NA          1         NA
## tenure_org       NA          NA        NA     NA         NA          1
## v_indegree       NA  0.22914404 0.5651666     NA         NA         NA
## v_outdegree      NA -0.03634641 0.1685519     NA         NA         NA
##             v_indegree v_outdegree
## loyalty             NA          NA
## sex          0.2291440 -0.03634641
## age          0.5651666  0.16855193
## tenure              NA          NA
## tenure_sup          NA          NA
## tenure_org          NA          NA
## v_indegree   1.0000000  0.28014905
## v_outdegree  0.2801491  1.00000000
```

We can see that variable 'tenure' highly correlates with variables 'tenure_sup' and 'tenure_org'. Let's delete it from our modelling.

```
cor_mat<-cbind(loyalty,sex,age,tenure_sup,tenure_org,v_indegree,v_outdegree)
```

```
## Warning in cbind(loyalty, sex, age, tenure_sup, tenure_org, v_indegree, :
## number of rows of result is not a multiple of vector length (arg 6)
```

```
cor(cor_mat)
```

```
##             loyalty         sex       age tenure_sup tenure_org v_indegree
## loyalty           1          NA        NA         NA         NA         NA
## sex              NA  1.00000000 0.1894667         NA         NA  0.2291440
## age              NA  0.18946670 1.0000000         NA         NA  0.5651666
## tenure_sup       NA          NA        NA          1         NA         NA
## tenure_org       NA          NA        NA         NA          1         NA
## v_indegree       NA  0.22914404 0.5651666         NA         NA  1.0000000
## v_outdegree      NA -0.03634641 0.1685519         NA         NA  0.2801491
##             v_outdegree
## loyalty              NA
## sex         -0.03634641
## age          0.16855193
## tenure_sup           NA
## tenure_org           NA
## v_indegree   0.28014905
```

```
## v_outdegree  1.00000000
```

Also let's hide tenure_org strongly correlated with one of our most hopeful predictors tenure_sup.

```
cor_mat<-cbind(loyalty,sex,age,tenure_sup,v_indegree,v_outdegree)
```

```
## Warning in cbind(loyalty, sex, age, tenure_sup, v_indegree, v_outdegree):
## number of rows of result is not a multiple of vector length (arg 5)
```

```
cor(cor_mat)
```

```
##             loyalty         sex       age tenure_sup v_indegree
## loyalty           1          NA        NA         NA         NA
## sex              NA  1.00000000 0.1894667         NA  0.2291440
## age              NA  0.18946670 1.0000000         NA  0.5651666
## tenure_sup       NA          NA        NA          1         NA
## v_indegree       NA  0.22914404 0.5651666         NA  1.0000000
## v_outdegree      NA -0.03634641 0.1685519         NA  0.2801491
##             v_outdegree
## loyalty              NA
## sex         -0.03634641
## age          0.16855193
## tenure_sup           NA
## v_indegree   0.28014905
## v_outdegree  1.00000000
```

Now we do not have serious problems with strong correlation between predictors.

```
##lmout<-lm(loyalty~sex+age+tenure_sup+v_indegree+v_outdegree)
##summary(lmout)
```

We see that all our predictors have statistically insignificant impact on dependenet variable. Let's try to include variables 'ed1' and 'ed2' in modelling.

```
##lmout<-lm(loyalty~sex+age+tenure_sup+ed1+ed2+v_indegree+v_outdegree)
##summary(lmout)
```

Let's leave only statistically significant predictors in the model.

```
lmout<-lm(loyalty~ed1+ed2)
summary(lmout)
```

```
## 
## Call:
## lm(formula = loyalty ~ ed1 + ed2)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.47949 -0.14615  0.02051  0.52051  0.52051 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   1.0000     0.5543   1.804 0.075863 .  
## ed1           1.9795     0.5586   3.544 0.000736 ***
## ed2           2.5000     0.6789   3.682 0.000472 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```
## Residual standard error: 0.5543 on 65 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1821, Adjusted R-squared:  0.157
## F-statistic: 7.237 on 2 and 65 DF,  p-value: 0.001454
```

We can observe that having education marked 4 ('ed2') positively impact on employee's loyalty. And having education marked 3 ('ed1') also positively influence but no so high. Surely, we understand that these variables are strongly correlate with each other but only they could be statistically significant predictors of our dependent variable. R-squared of our model is rather small.