Adrian Varallyay
April 2021

Lab 2

Table of Contents

Find a complete social network, preferably one with at least some attributes about the nodes with it. (If you simply have a social network, but no real attributes, you will need to pick an additional network to compare that first one to.)

Describe the social network(s) to me, in terms of how it was collected, what it represents and so forth. Also give me basic topography of the network: the nature of the ties; direction of ties; overall density; and if attributes are with the network, the distribution of the categories and variables of those attributes.

The data I used is from David Krackhardt's high-tech managers' networks. Specifically, I used the advice network. This data represents the advice network from 21 management personnel in a machine manufacturing firm. The data was originally collected to analyze a management intervention program. Original data contains advice (advice network) and friendship (friendship network) ties in addition to the formal structure (reports to network).

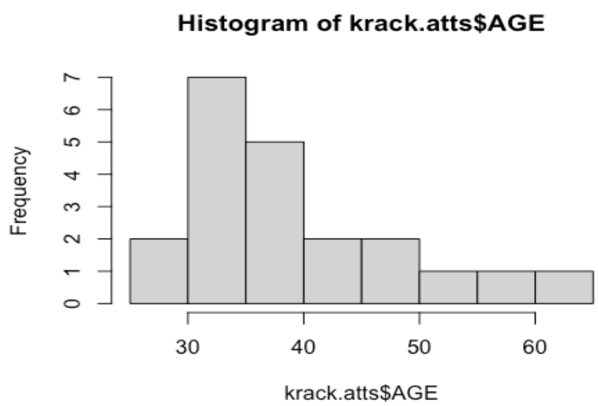
In the advice network, each individual was asked, "who does X go to for advice and help with work?". The nature of the ties is binary, the ties are directed, and the overall density is 0.9 (90%),

Potential Connections: PC = n*(n-1)/2 Network density: Actual connections/potential connections :: 190/210 = 0.9047619 (from: https://www.the-vital-edge.com/what-is-network-density/)

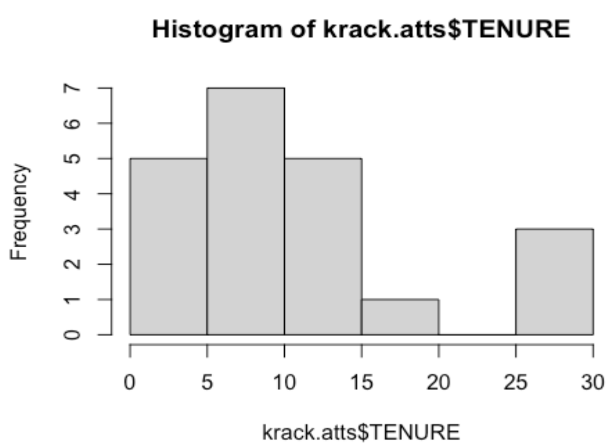
```
# overall network density
# actual connections (out of 210 possible connections  (21 * 20)/2 = 210)
actual_con <- ecount(krack_adv_graph)
net_den <- actual_con/210
net_den

## [1] 0.9047619
```

Adrian Varallyay
April 2021

```
psych::describe(krack.atts)

##        vars  n  mean   sd median trimmed  mad min max range  skew kurtosis
se
## ID        1 21 11.00 6.20     11   11.00 7.41   1  21    20  0.00    -1.37
1.35
## AGE       2 21 39.71 9.56     37   38.59 7.41  27  62    35  0.95    -0.16
2.09
## TENURE    3 21 11.71 8.11      9   10.88 5.93   0  30    30  1.00     0.01
1.77
## LEVEL     4 21  2.71 0.56      3    2.82 0.00   1   3     2 -1.65     1.72
0.12
## DEPT      5 21  2.19 1.17      2    2.18 1.48   0   4     4  0.19    -1.02
0.25

hist(krack.atts$AGE)
```




```
hist(krack.atts$TENURE)
```



Adrian Varallyay
April 2021

Calculate degree centrality (in- and out-degree, too, if you have such data); closeness centrality; betweenness centrality; and eigenvector centrality. Correlate those measures of centrality. Highlight which nodes are most central and least central, along different dimensions.

```
# prep centrality measures for correlation
krack_attributes <- tibble::tibble(krack_attributes)
krack_att_simple <- krack_attributes[,c(6:12)]

# correlate the measures
cor(krack_att_simple)
```

```
##             degree     in.deg    out.deg       btwn      close     vector
## degree   1.0000000  0.4114096  0.7334068  0.9292057  0.7783058  0.9931338
## in.deg   0.4114096  1.0000000 -0.3178644  0.2828835  0.2073325  0.4008942
## out.deg  0.7334068 -0.3178644  1.0000000  0.7556221  0.6549969  0.7341069
## btwn     0.9292057  0.2828835  0.7556221  1.0000000  0.8150296  0.8889104
## close    0.7783058  0.2073325  0.6549969  0.8150296  1.0000000  0.7582500
## vector   0.9931338  0.4008942  0.7341069  0.8889104  0.7582500  1.0000000
## bon     -0.5147375 -0.1086366 -0.4544307 -0.4067454 -0.3950058 -0.5290795
##                bon
##         -0.5147375
##         -0.1086366
##         -0.4544307
##         -0.4067454
##         -0.3950058
##         -0.5290795
##          1.0000000
```

```
psych::describe(krack_att_simple)
```

```
##         vars  n  mean   sd median trimmed  mad   min   max range skew kurt
osis
## degree     1 21 18.10 5.67  19.00   17.82 5.93  9.00 32.00 23.00 0.39    -
0.28
## in.deg     2 21  9.05 4.07   9.00    8.76 5.93  4.00 18.00 14.00 0.43    -
0.81
## out.deg    3 21  9.05 5.45   8.00    8.82 5.93  1.00 20.00 19.00 0.23    -
1.22
## btwn       4 21  3.10 2.57   3.53    2.83 3.32  0.20  9.15  8.95 0.53    -
0.68
## close      5 21  0.04 0.01   0.04    0.04 0.01  0.03  0.05  0.02 0.21    -
0.83
## vector     6 21  0.60 0.16   0.63    0.59 0.16  0.32  1.00  0.68 0.34    -
0.07
## bon        7 21 -0.92 0.41  -0.88   -0.93 0.35 -1.71 -0.05  1.66 0.17    -
0.36
##           se
## degree  1.24
## in.deg  0.89
```

```
## out.deg 1.19
## btwn    0.56
## close   0.00
## vector  0.04
## bon     0.09
```

As seen in the codes in the appendix, **for each measure of centrality, the most central row/nodes are as follows:**

degree = manager 18 (row 18)

in.degree = manager 2 (row 2)

out.degree = manager 15 (row 15)

btwn = manager 18 (row 18)

close = manager 15 (row 15)

eigen (vector) = manager 18 (row 18)
(this portion of the code may not have worked for me, I just get null returned if i look for the eigen measure. just in case, I kept the column named vector assuming this could be it)

bon = manager 11 (row 11)

For each measure of centrality, the least central nodes are as follows:

degree = manager 12 (row 12)

in.degree = managers 9, 13, 15, and 19 (rows 9, 13, 15, and 19)

out.degree = manager 6 (row 6)

btwn = manager 16 (row 16)

close = manager 12 (row 12)

eigen (vector) = manager 12 (row 12) (this portion of the code may not have worked for me, I just get null returned if i look for the eigen measure. just in case, I kept the column named vector assuming this could be it)

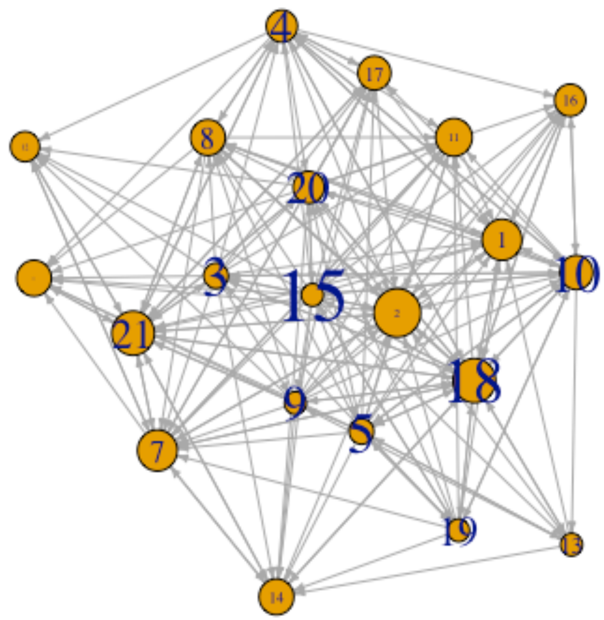
bon = manager 21 (row 21)

Adrian Varallyay
April 2021

Manager Advice Network Graph:

Size node: in-degree

Size of node label: out-degree



3a. If you have a network with attribute data, then state some hypothesis about how an attribute may be related to some (or all of the) measures of centrality. Explains why you think these two variables should be related.

I could not ascertain what the variables “level” and “dept” meant in the attributes, therefore I will rely on the remaining two attributes, “age” and “tenure”, seeing as these are the clearest to me in meaning (age and length of time at company).

I wanted to see if there was a relation between *TENURE* and degree centrality. In specific, I wanted to see how time at a company would predict *in.deg/out.deg*. I figured that as one’s time at a company increases, it would seem logical that being asked for advice also increases, as the seniors would have more experience. This would render one more central in the network (in.degree). Whereas the less time one is with a company, they would have less experience and have more out going advice centrality (out.degree).

```
summary(lm0)

##
## Call:
## lm(formula = in.deg ~ TENURE + AGE, data = krack_attributes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.4754 -1.9996 -0.6912  1.1643  6.2700
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.35882    3.14700   3.609   0.0020 **
## TENURE       0.36988    0.10246   3.610   0.0020 **
## AGE         -0.16730    0.08692  -1.925   0.0702 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.263 on 18 degrees of freedom
## Multiple R-squared:  0.4209, Adjusted R-squared:  0.3566
## F-statistic: 6.542 on 2 and 18 DF,  p-value: 0.00732
```

```
summary(lm1)
```

```
##
## Call:
## lm(formula = out.deg ~ TENURE + AGE, data = krack_attributes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6671 -3.7470  0.2465  3.6763  9.8857
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.31434    5.07396   2.230   0.0387 *
## TENURE      -0.28511    0.16519  -1.726   0.1015
## AGE          0.02702    0.14014   0.193   0.8493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.261 on 18 degrees of freedom
## Multiple R-squared:  0.1626, Adjusted R-squared:  0.0696
## F-statistic: 1.748 on 2 and 18 DF,  p-value: 0.2024
```

In either case, when you are done above, then consider alternate specifications of your variables and codings and decisions and models. What would you want to consider changing and why. If you can, report on what are the consequences of those changes?

There is one main thing I would like to tweak in my models. Upon viewing the results of the first two regressions, I intuited the need for a *TENURE x AGE* interaction. It occurred to me that maybe the effect of tenure is affected by age. What if one manager is particularly young but has been with the company for a while, would that change their advice *in.deg* centrality? Below I update the model to include the interaction term.

```
summary(lm2)
```

```
##
## Call:
## lm(formula = in.deg ~ TENURE + AGE + TENURE * AGE, data = krack_attributes
)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1796 -2.0582  0.0599  0.9743  5.8149
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.28726    5.94042   0.722   0.4803
## TENURE       1.10958    0.54136   2.050   0.0562 .
## AGE         -0.02119    0.13501  -0.157   0.8771
## TENURE:AGE  -0.01478    0.01063  -1.390   0.1824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.182 on 17 degrees of freedom
## Multiple R-squared:  0.4801, Adjusted R-squared:  0.3883
## F-statistic: 5.232 on 3 and 17 DF,  p-value: 0.009659
```

```r
summary(lm3)
```

```
##
## Call:
## lm(formula = out.deg ~ TENURE + AGE + TENURE * AGE, data = krack_attribute
s)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.673 -3.877  0.458  3.994  9.893
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.906173  10.099796   0.981    0.340
## TENURE      -0.137816   0.920403  -0.150    0.883
## AGE          0.056116   0.229538   0.244    0.810
## TENURE:AGE  -0.002943   0.018072  -0.163    0.873
##
## Residual standard error: 5.409 on 17 degrees of freedom
## Multiple R-squared:  0.1639, Adjusted R-squared:  0.01641
## F-statistic: 1.111 on 3 and 17 DF,  p-value: 0.372
```

Lastly, give your best conclusion as to what you learned from your analysis. Did it make sense, given your initial expectations? Why? Why not?

For *lm0*, *in.deg ~ TENURE + AGE*, the coefficient on Tenure is 0.37 and is statistically significant ($p<0.01$), meaning that on average, as one's tenure at the company increases by one, the value of in.deg increases by 0.37. This follows the logic I presented earlier.

Adrian Varallyay
April 2021

The coefficient on age is -0.17 and is almost statistically significant ($p<0.1$). Though not quite significant, it would insinuate that as one's age increases by one year, the value of in.deg decreases by 0.17. This also makes sense and kind of follows along the lines of my thinking (though I did not explicitly state/think it)

Interestingly, for *lm1*, *out.deg ~ TENURE + AGE*, neither coefficient for age and tenure were statistically significant for out.degree. However the direction of the effect for tenure was as predicted, the longer at the company the lower out.degree centrality (coefficient on TENURE = -0.29).

As both TENURE and AGE essentially represent one year for every increment, I already begin to sense that I must be wary of an interaction.

For *lm2*, *in.deg ~ TENURE + AGE + TENURE x AGE*, statistical significance all but disappears. The coefficient on tenure is now 1.11 and is almost statistically significant ($p<0.1$). Now the interaction term is beginning to explain more in the model, it attenuated the outcome for TENURE.

Like *lm1*, *lm3*, *out.deg ~ TENURE + AGE + TENURE x AGE*, does not produce statistical significance for tenure or age. The interaction term is also insignificant.

As such, I believe the results mainly make sense given my expectations. I did not initially think that my interaction term would affect the results of the *in.deg* models as much as it did. But even as I was in process of doing the lab, the intuitive awareness bloomed and I realized that though I was partially on to something, I had to tweak the model to the reality I was picking up on. Its interesting to think that one's "popularity"/"prestige" (*in.deg* centrality) for advice could be predicted/affected by one's tenure. Whereas tenure has no predictive power on the "social activity"/"seeking" for advice. But yes, in the course of doing these regressions, the models make sense to me. Despite the significance of the variables in the models, they all behaved in an anticipated way, so to speak; the direction of effects was (relatively) as I thought they would be.

Adrian Varallyay
April 2021

Appendix (code)

SET UP

```
# attach attributes to vertices
vertex_attr(krack_adv_graph, index=krack.atts$ID) <- krack.atts

# calculate degree for advice after Merging attributes with a new df
krack_attributes <- merge(krack.atts,
                          data.frame(ID=V(krack_adv_graph)$ID,
                          degree= degree(krack_adv_graph)),
                          by='ID')

# calculate centrality for adv
krack_attributes <- merge(krack_attributes,
                          data.frame(ID=V(krack_adv_graph)$ID,
                          in.deg= degree(krack_adv_graph, mode = c("in"), loo
ps = TRUE, normalized = FALSE),
                          out.deg= degree(krack_adv_graph, mode = c("out"), l
oops = TRUE, normalized = FALSE),
                          btwn= betweenness(krack_adv_graph, directed = F), #
not sure if this should have been "T", Laz and Krack were both directed but i
n example "F" is still used
                          close = closeness(krack_adv_graph, mode = c("all"))
,
                          eigen <- evcent(krack_adv_graph), # this portion of
the code didn't work for me, I just get null returned if i look for the eigen
measure. just in case, I kept the column named vector assuming this could be
it
                          bon <- bonpow(krack_adv_graph)),
                          by='ID')

# clean up  dataframe
krack_attributes <- krack_attributes[,c(1:11, 33)]

# change the bon centrality name
names(krack_attributes)[names(krack_attributes)== "bon....bonpow.krack_adv_gr
aph."] <- "bon"

names(krack_attributes)

##  [1] "ID"      "AGE"     "TENURE"  "LEVEL"   "DEPT"    "degree"  "in.deg"
##  [8] "out.deg" "btwn"    "close"   "vector"  "bon"
```

CENTRALITY

```
# max/min degree
slice_max(krack_att_simple, order_by = degree)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl> <dbl>
## 1     32     15      17  9.15 0.0435   1.00 -1.06
```

```
slice_min(krack_att_simple, order_by = degree)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1      9      7       2 0.267 0.0303  0.316 -0.688
```

```
# max/min in.deg
slice_max(krack_att_simple, order_by = in.deg)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     21     18       3  4.62 0.0476  0.665 -0.734
```

```
slice_min(krack_att_simple, order_by = in.deg)
```

```
## # A tibble: 4 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     17      4      13 4.50  0.0417  0.537 -0.732
## 2     10      4       6 0.239 0.0312  0.362 -0.126
## 3     24      4      20 7.28  0.05    0.753 -1.28
## 4     15      4      11 0.555 0.0345  0.536 -0.943
```

```
# max/min out.deg
slice_max(krack_att_simple, order_by = out.deg)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl> <dbl>  <dbl> <dbl>
## 1     24      4      20  7.28  0.05  0.753 -1.28
```

```
slice_min(krack_att_simple, order_by = out.deg)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     11     10       1 0.327 0.0333  0.382 -0.879
```

```
# max/min btwn
slice_max(krack_att_simple, order_by = btwn)
```

```
## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl> <dbl>
## 1     32     15      17  9.15 0.0435   1.00 -1.06
```

```
slice_min(krack_att_simple, order_by = btwn)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     12      8       4 0.198 0.0323  0.442 -0.678

# max/min close
slice_max(krack_att_simple, order_by = close)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl> <dbl>  <dbl> <dbl>
## 1     24      4      20  7.28  0.05  0.753 -1.28

slice_min(krack_att_simple, order_by = close)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1      9      7       2 0.267 0.0303  0.316 -0.688

# max/min vector
slice_max(krack_att_simple, order_by = vector)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl> <dbl>
## 1     32     15      17  9.15 0.0435   1.00 -1.06

slice_min(krack_att_simple, order_by = vector)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector    bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1      9      7       2 0.267 0.0303  0.316 -0.688

# max/min bon
slice_max(krack_att_simple, order_by = bon)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn  close vector     bon
##    <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>   <dbl>
## 1     14     11       3 0.807 0.0370  0.481 -0.0510

slice_min(krack_att_simple, order_by = bon)

## # A tibble: 1 x 7
##   degree in.deg out.deg  btwn close vector   bon
##    <dbl>  <dbl>   <dbl> <dbl> <dbl>  <dbl> <dbl>
## 1     26     15      11  5.67  0.04  0.807 -1.71
```

Adrian Varallyay
April 2021

Network Graph

```
## start the graph ##
set.seed(12)
l <- layout.kamada.kawai(krack_adv_graph)

# Plot undecorated first.
par(mfrow=c(1,1))
oldMargins<-par("mar")
par(mar=c(1,1,1,1))
### par(mar=oldMargins) ### to return to default ...


# Size node by in-degree.
V(krack_adv_graph)$size <- 4*sqrt(degree(krack_adv_graph, mode="in"))
V(krack_adv_graph)$size2 <- V(krack_adv_graph)$size * .5

# Size of node label by out-degree.
V(krack_adv_graph)$label.cex <- 2.5 * degree(krack_adv_graph, mode="out") / m
ax(degree(krack_adv_graph, mode="out"))

# Shrink arrows
plot(krack_adv_graph, layout=l, edge.arrow.size=.3)
```

MODELS

```
head(krack_attributes)

## # A tibble: 6 x 12
##      ID   AGE TENURE LEVEL  DEPT degree in.deg out.deg  btwn  close vector
##   <int> <int>  <int> <int> <int>  <dbl>  <dbl>   <dbl> <dbl>  <dbl>  <dbl>
## 1     1    33      9     3     4     19     13       6 2.88  0.0417  0.623
## 2     2    42     20     2     4     21     18       3 4.62  0.0476  0.665
## 3     3    40     13     3     2     20      5      15 3.53  0.0435  0.664
## 4     4    33      8     3     4     20      8      12 3.62  0.0385  0.663
## 5     5    32      3     3     2     20      5      15 4.94  0.0417  0.632
## 6     6    59     28     3     1     11     10       1 0.327 0.0333  0.382
## # … with 1 more variable: bon <dbl>

# regress attributes on centrality measures
lm0 <- lm(in.deg ~ TENURE + AGE, krack_attributes)
lm1 <- lm(out.deg ~ TENURE + AGE, krack_attributes)

lm2 <- lm(in.deg ~ TENURE + AGE + TENURE*AGE, krack_attributes)
lm3 <- lm(out.deg ~ TENURE + AGE + TENURE*AGE, krack_attributes)
```