Lab1

Amir Imani - ai2335

22 February 2016

1. Develop a hypothesis about how some ego-network measure (e.g., degree/size, density, diversity, average-level of alters, homophily, structural holes, or brokerage) may be related to some other variable of interest.

I have selected KRACKHARDT–HIGH-TECH MANAGERS data set. I am interested to see how and to what extend different attributes of managers (Age, Level, Department, Tenure) affect their friendship network. My hypothesis is that homophily in friendship ego networks have a positive corelation with your age group, your deperatment, tenureship and your level in corporate hierarchy. I define age group as if you are in the range of one standard deviation.

2. Explain why you think these two variables should be related.

I suggest that in a corporation, you tend to befriend colleagues who are in your own department, or in your own level, or in the same age group as you have more work/non-work things to share.

3. Tell me about your variables. What is your dependent variable? What are your independent variables? How are they coded? How are they recoded? How are the calculated, if appropriate?

My dependant variable is homophily in ego networks. My independent variables are age, level, tenure, and department. I will introduce more independant variable in the next sections.

My independent variables were collected in the attributes table. This consist of the managers age (in years), length of service or tenure (in years), level in the corporate hierarchy (coded 1,2 and 3; 1=CEO, 2 = Vice President, 3 = manager) and department (coded 1,2,3,4 with the CEO in department 0 ie not in a department).

Age	Tenure	Level	Department
33	9	3	4
42	20	2	4
...	...	...	...

Since I needed to see if two connected nodes with aare in the same level, same department, and same age and tenureship, I recoded them like below to have 1 or 0 result for each node.

```
friendship_edges_attributes$same_dept = ifelse(
  friendship_edges_attributes$Dept1==friendship_edges_attributes$Dept2, 1,
  0) # otherwise it's 0

#  Add columns  which inidcate if they have the same level
friendship_edges_attributes$same_level = ifelse(
  friendship_edges_attributes$Level1==friendship_edges_attributes$Level2, 1,
```

```
  0) # otherwise it's 0

#  Add columns  which inidcate if they are in the same age group (based on age SD)
friendship_edges_attributes$same_age = ifelse(
  (friendship_edges_attributes$Age1 < friendship_edges_attributes$Age2 + 5) &
    (friendship_edges_attributes$Age1 > friendship_edges_attributes$Age2 - 5),
   1,
  0) # otherwise it's 0

#  Add columns  which inidcate if they are in the same tenure group
friendship_edges_attributes$same_tenure = ifelse(
  (friendship_edges_attributes$Tenure1 < friendship_edges_attributes$Tenure2 + 1) &
    (friendship_edges_attributes$Tenure1 > friendship_edges_attributes$Tenure2 - 1),
  1,
  0) # otherwise it's 0
```

Basic ego netowork measures including in and out degree, densitiy, and reciprocity of friendship edges are:

Average and Standard Deviation of in-degrees

```
## [1] 4.857143
```

```
## [1] 2.220039
```

Average and Standard Deviation of out-degrees

```
## [1] 4.857143
```

```
## [1] 4.475329
```

Reciprocity

```
## [1] 0.4509804
```

Density

```
## [1] 0.2428571
```

4. Present your initial results from your first few models. What do they indicated about your hypothesis?

My initial finding on the relation of homophliy to attributes for each node is shown in the following table.

```
knitr::kable(ego_homophily_stats)
```

ID1	same_dept	same_level	same_age	same_tenure
1	0.6000000	0.8000000	0.6000000	0.2000000
2	0.3333333	0.6666667	0.0000000	0.0000000

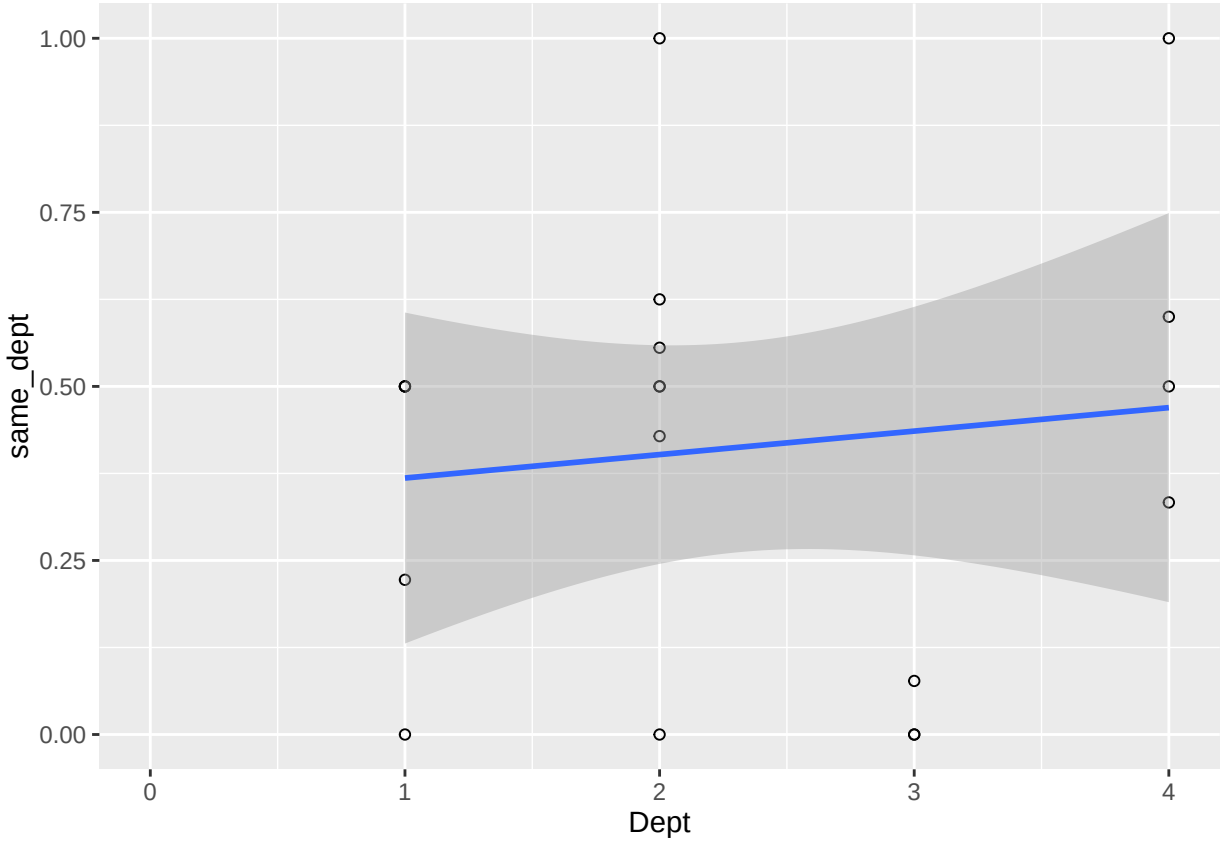
ID1	same_dept	same_level	same_age	same_tenure
3	1.0000000	0.5000000	0.5000000	0.0000000
4	0.5000000	0.8333333	0.6666667	0.0000000
5	0.4285714	0.5714286	0.4285714	0.0000000
6	0.5000000	0.5000000	0.3333333	0.0000000
8	0.0000000	1.0000000	1.0000000	0.0000000
10	0.0000000	1.0000000	0.5714286	0.1428571
11	0.0769231	0.8461538	0.1538462	0.0000000
12	0.5000000	0.7500000	1.0000000	0.2500000
13	0.5000000	1.0000000	0.5000000	0.0000000
14	0.5000000	0.0000000	0.5000000	0.0000000
15	0.6250000	0.8750000	0.2500000	0.0000000
16	1.0000000	0.5000000	0.0000000	0.0000000
17	0.2222222	0.7777778	0.3888889	0.0555556
18	0.0000000	1.0000000	0.0000000	0.0000000
19	0.5555556	0.7777778	0.3333333	0.0000000
20	0.0000000	0.5000000	0.0000000	0.0000000
21	0.5000000	0.5000000	0.5000000	0.0000000

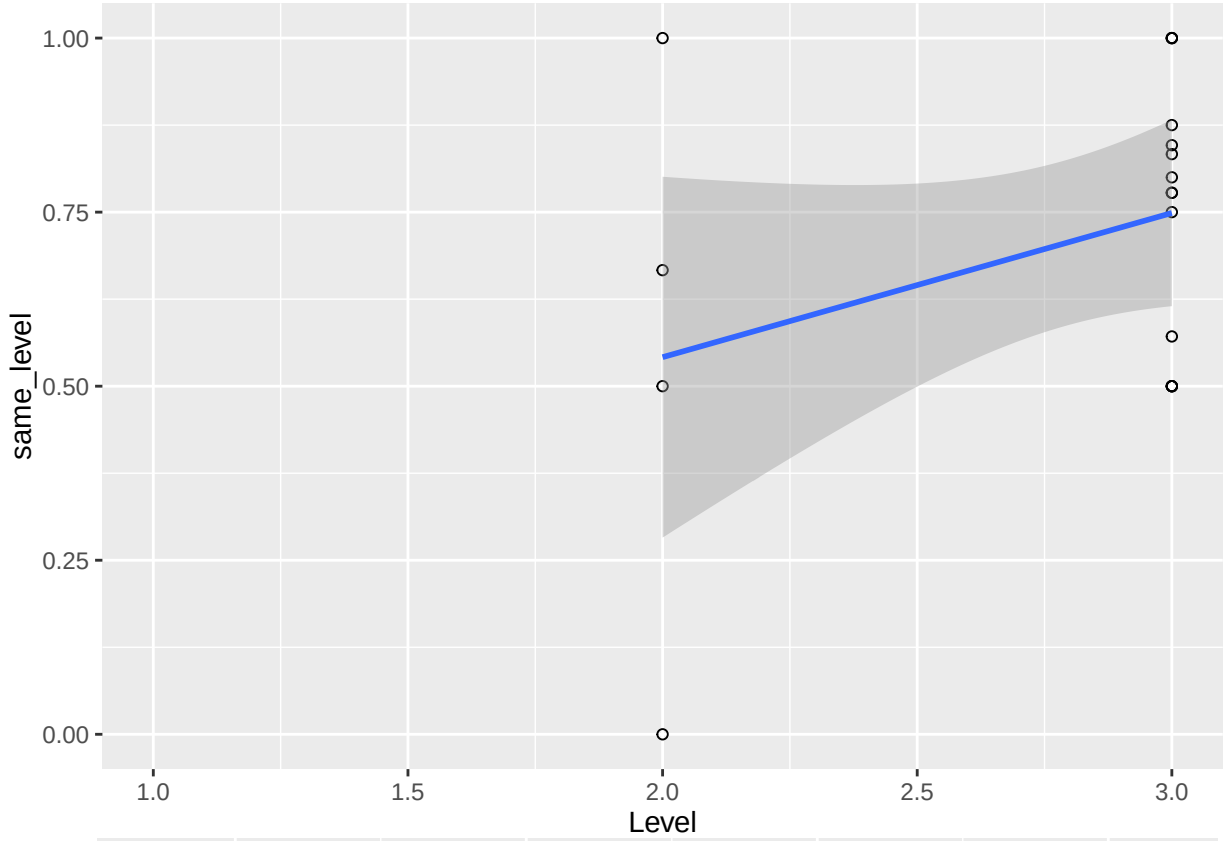
4.1 Basic statistics

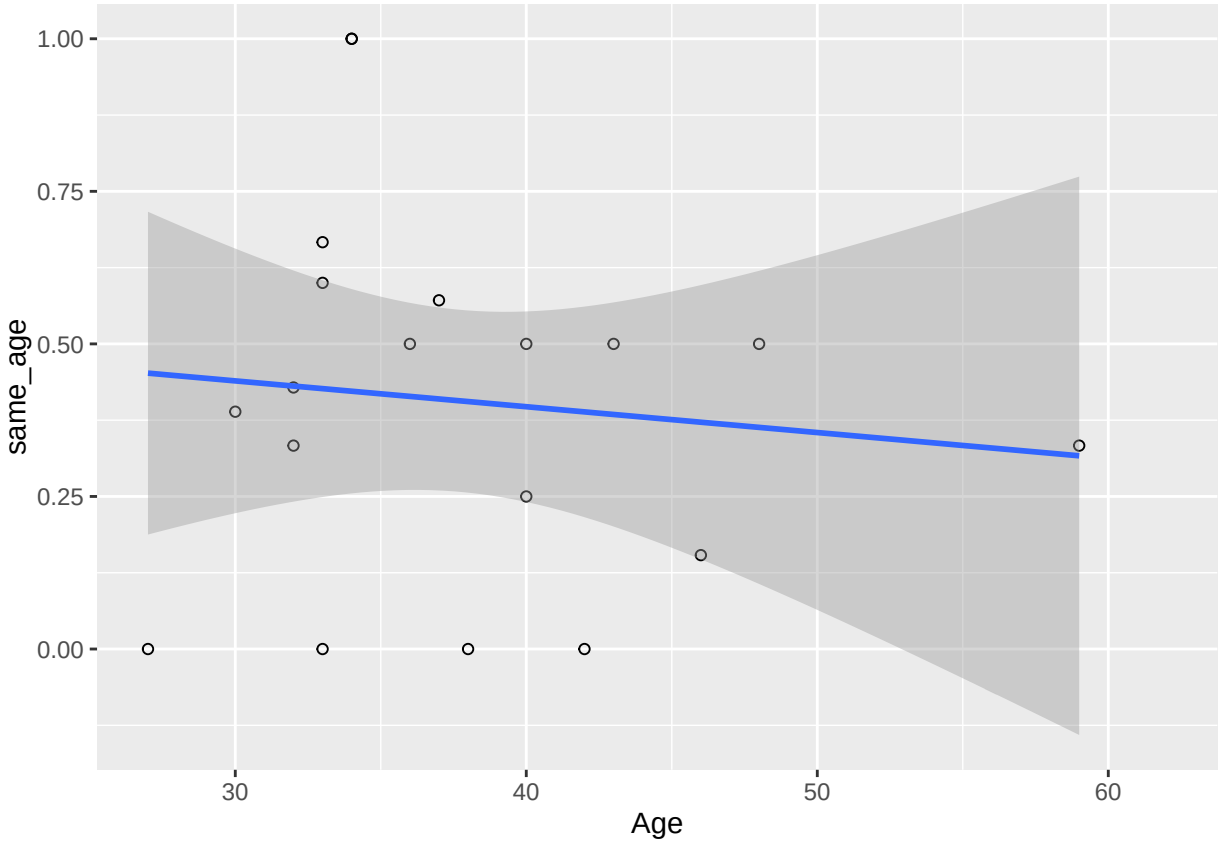
```
## [1] "Average homophily based on department: 0.412716085084506"

## [1] "Standard deviation for homophily based on department: 0.304867426629826"

## [1] "Average homophily based on level: 0.705165156480946"

## [1] "Standard deviation for homophily based on level: 0.25402288997073"

## [1] "Average homophily based on age: 0.406635177687809"

## [1] "Standard deviation for homophily based on age: 0.301162817566538"

## [1] "Average homophily based on tenure: 0.0341269841269841"

## [1] "Standard deviation for homophily based on tenure: 0.075962788094613"
```

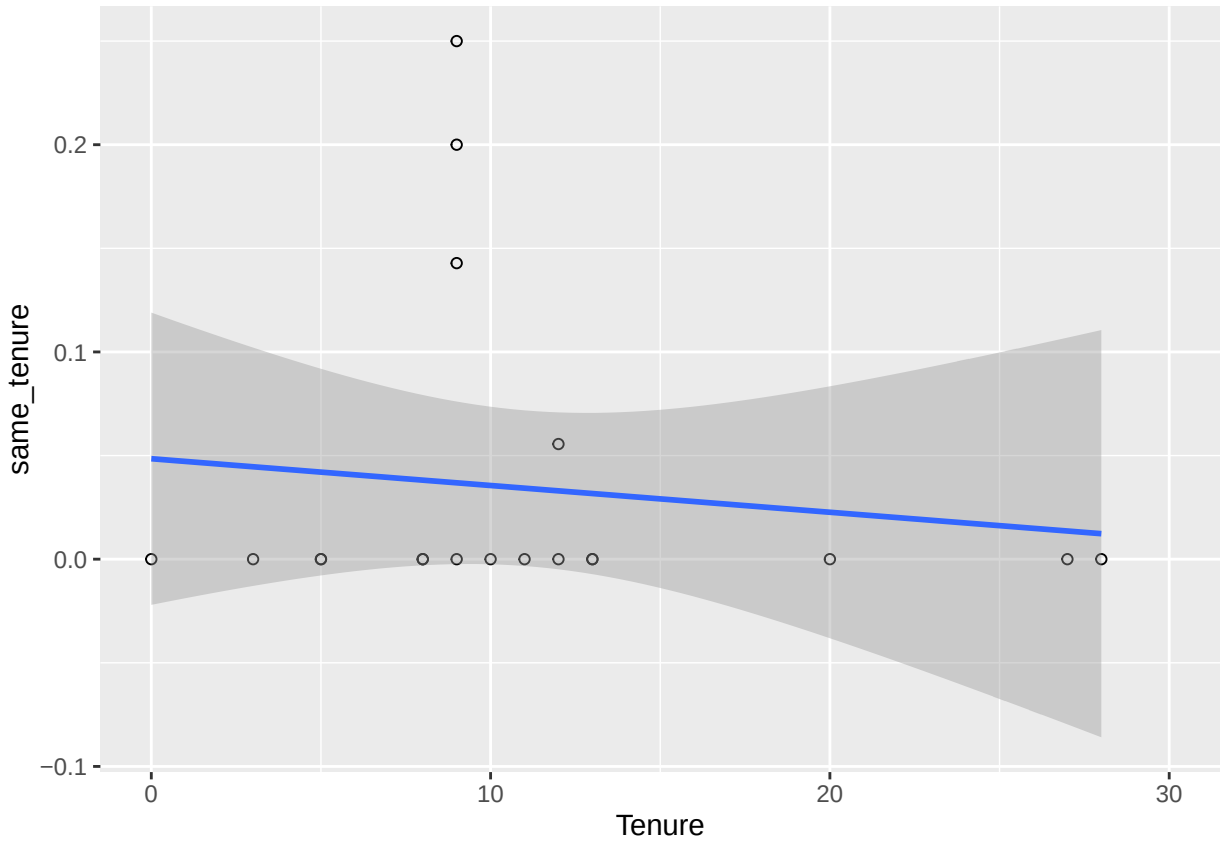
As can be seen from the means and standard deviations of homophily of different attributes, being in the same level means more homophily with less deviation comparing to other attributes. Also it seems that having the same tenureship is not a major factor in forming homophil network of friendship.

4.2 Plots I have plotted the realtion between each attribute and homophily based on those attributes to see if any visible relations exists.




same_level
1.00
0.75
0.50
0.25
0.00
1.0
1.5
2.0
2.5
3.0
Level


same_age
1.00
0.75
0.50
0.25
0.00
30
40
50
60
Age



As seen from the graphs, people from tend to form firendship network from their own department, and their own level in the hierarchy. I think it make sense as picking friends from your own department is convinient, you spend time more often with them, and usually there is alot more to discuss about. Forming your friendship network from people of the same rank aslo makes sense - particularly for higher positions like managers - as often there is a seperation between lower ranking staff with higher ones based on the hierarchical power structure. As it is seen from the graph, homophily among managers is 0.75.

However, interesting observation from these graphs is negetive corelation between network homophily and age/tenure. I think one reason could be that since the average age is around 40 with SD 9.5, there are not too many +50 staffs. They are also the same one with higher seniority in their tenure.

4.3 Model I have fitted linear models on each of those attributes to have a better understanding of correlations.

Homophiy based on the same **Department**:

```
summary(lm(same_dept ~ Dept, attributes))
```

```
##
## Call:
## lm(formula = same_dept ~ Dept, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43578 -0.25250  0.09793  0.13163  0.59793
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.33466    0.16917   1.978   0.0643 .
## Dept         0.03371    0.06622   0.509   0.6173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3113 on 17 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.01501,    Adjusted R-squared:  -0.04293
## F-statistic: 0.2591 on 1 and 17 DF,  p-value: 0.6173
```

Homophiy based on the same **Level**:

```
summary(lm(same_level ~ Level, attributes))
```

```
##
## Call:
## lm(formula = same_level ~ Level, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54167 -0.21305  0.02901  0.12562  0.45833
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1275     0.3897   0.327    0.748
## Level         0.2071     0.1382   1.498    0.152
##
## Residual standard error: 0.2457 on 17 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1166, Adjusted R-squared:  0.06464
## F-statistic: 2.244 on 1 and 17 DF,  p-value: 0.1525
```

Homophiy based on the same **Age**:

```
summary(lm(same_age ~ Age, attributes))
```

```
##
## Call:
## lm(formula = same_age ~ Age, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45207 -0.18244  0.01668  0.14923  0.57755
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.566322   0.370614   1.528    0.145
## Age         -0.004232   0.009641  -0.439    0.666
##
```

```
## Residual standard error: 0.3082 on 17 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.01121,    Adjusted R-squared:  -0.04696
## F-statistic: 0.1927 on 1 and 17 DF,  p-value: 0.6662
```

Homophiy based on the same **Tenure**:

```
summary(lm(same_tenure ~ Tenure, attributes))
```

```
##
## Call:
## lm(formula = same_tenure ~ Tenure, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04847 -0.03814 -0.03297 -0.01295  0.21315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.048471   0.033419   1.450    0.165
## Tenure      -0.001292   0.002547  -0.507    0.619
##
## Residual standard error: 0.07758 on 17 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.0149, Adjusted R-squared:  -0.04304
## F-statistic: 0.2572 on 1 and 17 DF,  p-value: 0.6186
```

The same conclusions can be made from these models that with unit increase in age and tenureship, homophily drops by 0.004 and 0.001 ofaverage. Also it can be seen that none of the variables are statistcially significant.

5. Consider alternate specifications of your variables (i.e., recodings of various kinds). Consider interactions among your variables.

I have fit the same models including **in-degree**, **out-degree**, and **transitivity** for each node.

I constructed those graph measures as below.

```
attributes <- merge(attributes, data.frame(ID=V(friendship_graph)$ID,
                                           degree.in = degree(friendship_graph, mode = "in")),
                    by='ID', all =TRUE)

attributes <- merge(attributes, data.frame(ID=V(friendship_graph)$ID,
                    degree.out = degree(friendship_graph, mode = "out")),
                    by='ID', all =TRUE)

#transitivity
attributes <- merge(attributes, data.frame(ID=V(friendship_graph)$ID,
                    transitivity=transitivity(friendship_graph, type="local") ) ,
                    by='ID', all = TRUE)
```

```r
summary(lm(same_dept ~ Dept + degree.in + degree.out + transitivity, attributes))
```

```
##
## Call:
## lm(formula = same_dept ~ Dept + degree.in + degree.out + transitivity,
##     data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49670 -0.24000  0.08422  0.18128  0.47823
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.042527   0.533964   0.080    0.938
## Dept          0.021120   0.073232   0.288    0.777
## degree.in     0.023739   0.037568   0.632    0.538
## degree.out   -0.002518   0.023202  -0.109    0.915
## transitivity  0.679017   1.032192   0.658    0.521
##
## Residual standard error: 0.3316 on 14 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.07989,    Adjusted R-squared:  -0.183
## F-statistic: 0.3039 on 4 and 14 DF,  p-value: 0.8705
```

```r
summary(lm(same_level ~ Level + degree.in + degree.out + transitivity, attributes))
```

```
##
## Call:
## lm(formula = same_level ~ Level + degree.in + degree.out + transitivity,
##     data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31664 -0.14375  0.01426  0.10860  0.40512
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.77626    0.48324   1.606   0.1305
## Level         0.26468    0.14644   1.807   0.0922 .
## degree.in    -0.02371    0.02550  -0.930   0.3682
## degree.out   -0.02185    0.01738  -1.257   0.2292
## transitivity -1.78930    0.71989  -2.486   0.0262 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.224 on 14 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.3951, Adjusted R-squared:  0.2223
## F-statistic: 2.286 on 4 and 14 DF,  p-value: 0.1116
```

```r
summary(lm(same_age ~ Age + degree.in + degree.out + transitivity, attributes))
```

```
##
## Call:
## lm(formula = same_age ~ Age + degree.in + degree.out + transitivity,
##     data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50747 -0.14960  0.02612  0.15918  0.61234
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.8741265  0.7330082   1.193    0.253
## Age          -0.0043262  0.0110466  -0.392    0.701
## degree.in    -0.0005437  0.0389960  -0.014    0.989
## degree.out   -0.0148168  0.0231359  -0.640    0.532
## transitivity -0.6896690  1.0414127  -0.662    0.519
##
## Residual standard error: 0.3329 on 14 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.0495, Adjusted R-squared:  -0.2221
## F-statistic: 0.1823 on 4 and 14 DF,  p-value: 0.9438
```

```
summary(lm(same_tenure ~ Tenure + degree.in + degree.out + transitivity, attributes))
```

```
##
## Call:
## lm(formula = same_tenure ~ Tenure + degree.in + degree.out +
##     transitivity, data = attributes)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.073153 -0.051215 -0.004376  0.015693  0.148963
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.137964   0.118005   1.169    0.262
## Tenure       -0.001938   0.002610  -0.743    0.470
## degree.in     0.006558   0.008542   0.768    0.455
## degree.out   -0.002720   0.005391  -0.504    0.622
## transitivity -0.310052   0.237326  -1.306    0.212
##
## Residual standard error: 0.07614 on 14 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.2186, Adjusted R-squared:  -0.004656
## F-statistic: 0.9791 on 4 and 14 DF,  p-value: 0.4501
```

6. And give your best conclusion as to whether your initial hypothesis held up - and if not, why not.

It seems that there is a small corealtion (positive for department and level, negative for age and tenure) and homophily. However, transitivity seems to be a larger contributor to homohpily. However, all those factors are statistically insignificant.