Statistical comparison of male prairie voles based on centrality measures from their prefrontal cortex brain region in a network

Maths Application in Data Science Report One

George E. Quaye   gequaye@miners.utep.edu

University of Texas, El Paso (UTEP)

Department of Mathematical Sciences

Contents

1

Introduction

This study primarily provides statistical comparative analysis to answer an objective question between two groups of prairie voles by means of a network (graph analysis). The satellite region of interest from these prairie voles was the pre-frontal cortex region of their brain.

Graph theory concepts

A graph theory approach was used to generate network models and analyze the functional network connectivity between all brain regions for the less social (KI) and more social groups (IL) that are under study. Graph theory is a branch of mathematics and computer science that explores the patterns of connectivity between multiple vertices within a system. Graphs or networks are collections of elements (nodes, vertices) and their pairwise links (edges, connections) which, in their simplest form, can be summarized in the form of a connection (or adjacency) matrix (Sporns 2018). The graph approach allowed for an investigation of the most important nodes within the social decision-making network with regards to whether or not a group is less social or more social. A graph representation can be either directed or undirected. Directed graphs are a class of graphs that don't presume symmetry or reciprocity in the edges established between vertices whereas undirected graphs are more specific. For them, there's an extra assumption regarding the reciprocity in the relationship between pairs of vertices connected by an edge.

Numerous measures quantify the potential of individual nodes and edges to influence the global state of the network. Many of them allow the identification of network hubs, (Sporns, Honey, and Kötter 2007), (Heuvel and Sporns 2013), generally defined as highly central parts of the network. The number of connections maintained by a node (its degree centrality) often provides a strong indicator of influence or centrality. Other measures of centrality take advantage of the layout of the shortest paths within the network and record the number of such paths that pass through a given node or edge, this measure is called the betweenness centrality. Yet another way to approach centrality measure is closeness (or closeness centrality) of a node in a graph (network), which is estimated as the sum of the length of the shortest paths between the node and all other nodes in

the graph. Also, comes another influential centrality measure called eigen centrality, which assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

However, most measures of node centrality are mutually related (and hence statistically correlated); for example, in most (but not all) networks nodes with many connections (high degree) also serve as intermediaries on many short paths (high betweenness). Since different measures of centrality index different aspects of network organization, it can be beneficial to rank nodes by aggregating multiple centrality measures.(Sporns, Honey, and Kötter 2007)

Objective

In this analysis, the four centrality measures such as degree, closeness, betweenness, and Eigen centrality were estimated from the undirected graph of the more social (IL) or less social (KI) group. Given these measures, the pre-frontal cortex subregion of the brain from these two groups of voles was spotlighted for statistical comparison.

The main objective of this analysis then was to justify statistically that, given these centrality measures from the pre-frontal cortex region of the brain in the IL and KI male voles, we can strongly agree that there is a significant difference in their social behavior as indicated.

Exploratory Analysis

In this section, a Graph theory analysis was performed using an R programming language for each group understudy from the pre-frontal cortex region amongst all other regions. Graphical plots such as boxplot, barplot, and density plots were generated to compare the centrality measures from the IL and KI groups specific to the pre-frontal cortex region. A boxplot is a standardized way of displaying the distribution of data based on a five-number summary ("minimum", first quartile (Q1), median, third quartile (Q3), and "maximum") for these centrality measures. Also,

a probability density plot and bar plots were used to further assess the distribution of the sample centrality measure for further analysis.

Data description and source

A resting-state fMRI data from a more social group (IL) and less social group (KI) were used. KI group is a resultant offspring of a prairie vole from the Kansas region and Illinois region. The prairie vole (Microtus ochrogaster) is a socially monogamous rodent that forms a long-term pair-bond between breeding pairs, which provides a unique model to characterize functionality in this network associated with social commitment (Aragona and Wang 2004). The prairie vole has been used for over three decades to study the neurobiology of pair bonding, which has led to the well-defined behavioral characterization of the commitment signals of the pair bond, including partner affiliation and stranger aggression (Anacker, Smith, and Ryabinin 2014). This creates an ethologically valid model to measure social behaviors that constitute the behavioral expression of a pair bond. Furthermore, a number of brain regions in the SDMN have been individually identified to be involved in the regulation of partner affiliation, stranger aggression, or both in prairie voles (Walum and Young 2018).

Visualization of graph from entire brain
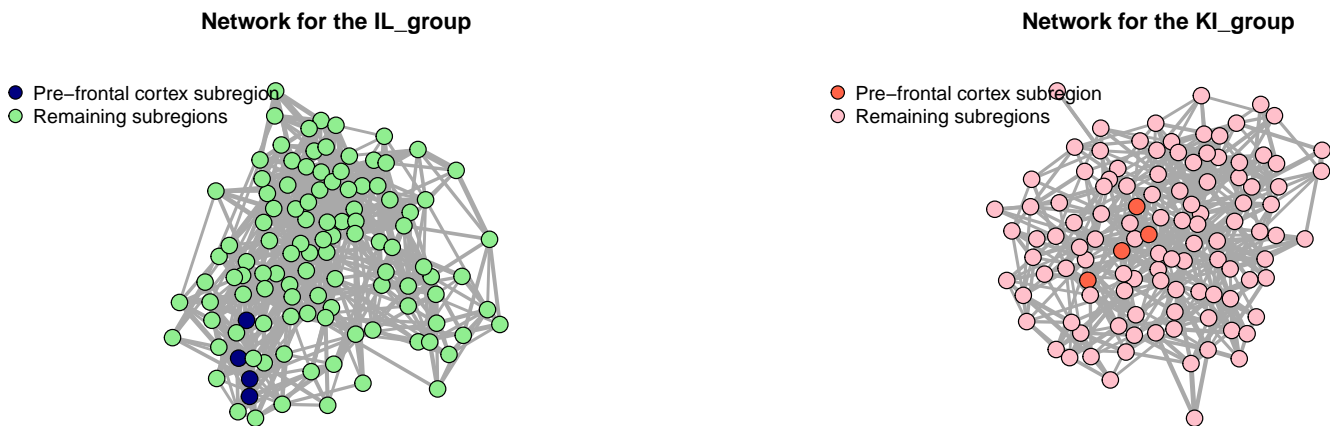
Above are undirected Graph theory plots from the resting-state fMRI data for both more social (IL) and less social (KI) groups. Given the layout from the two, it was generally seen that the

pre-frontal cortex region in the groups tends to have some substantial connectivity in the brain network. From these plots, the centrality measures as explained earlier were obtained for further analysis.

Distribution of degree centrality between groups

Distribution of degree centrality measure between IL and KI groups
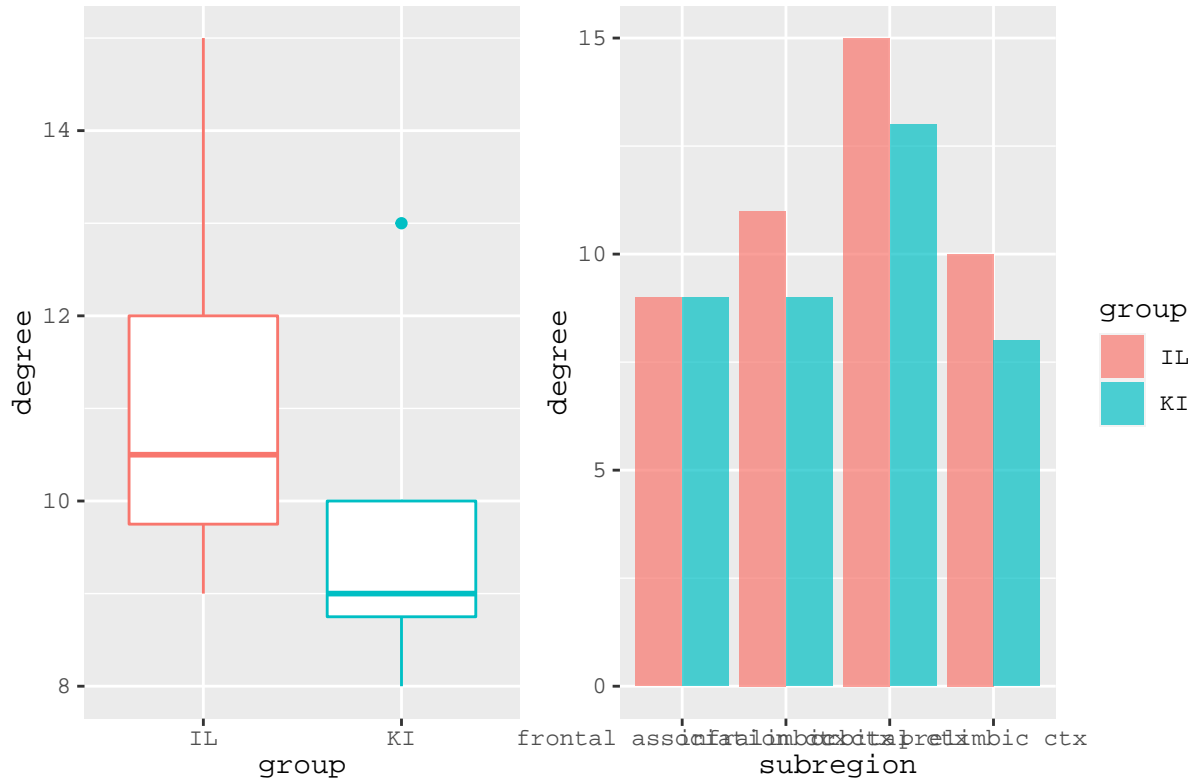
From the plot of degree centrality between the two groups, it was observed that there is greater variability in IL group area_average whereas there is a higher outlying observation in the KI group. On average from the boxplot, the degree measure from the IL group is greater than the KI group as it is also depicted in the graph bar plot by the vertical plots. The two samples have skewed distribution, that is they are not normally distributed as clearly shown in the box plot. Also, since the notches in the boxplot do not overlap, one can graphically conclude that with 95% confidence, the true averages do differ. However, it only needs a valid statistical analysis to confirm.

Distribution of betweenness centrality between groups

Betweenness centrality measure between IL and KI groups
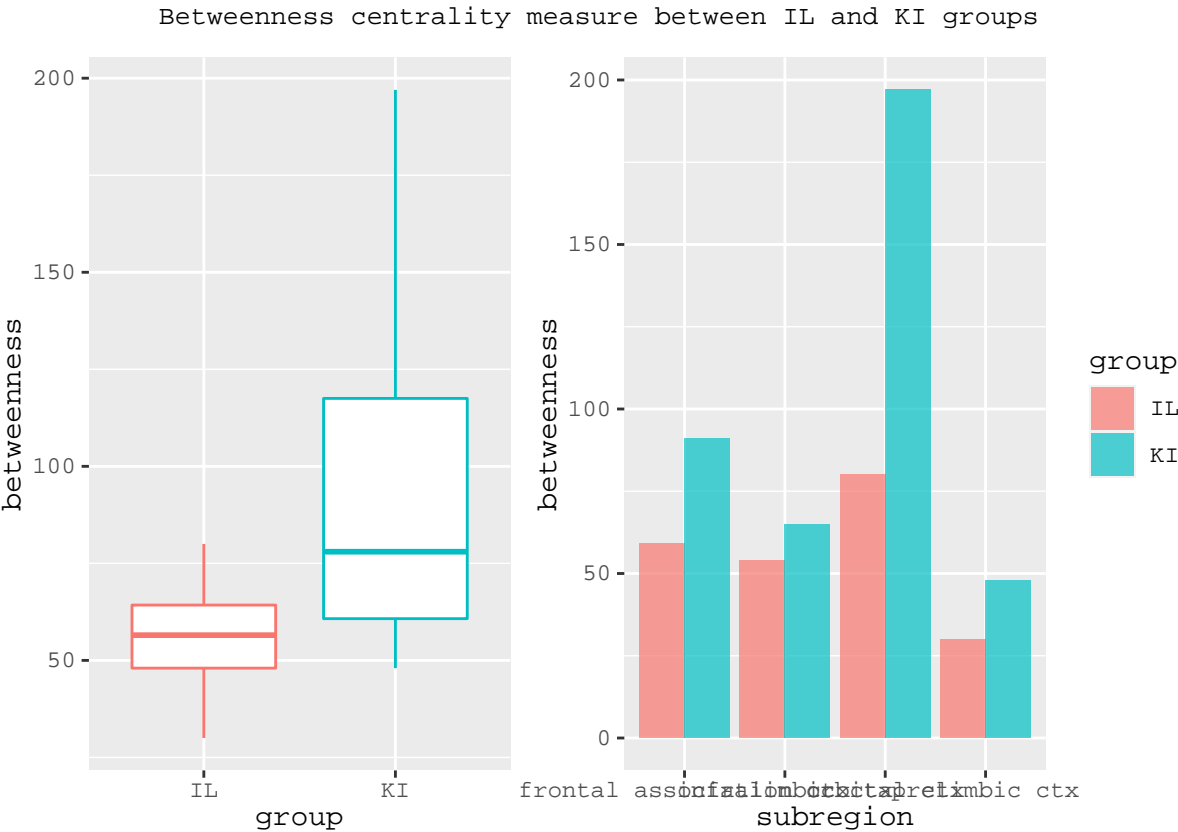
Given the plot of the betweenness centrality measure between the two groups, we observe that there is greater variability in the KI group area_average compared to the IL group. On average from the boxplot, the degree measure from the KI group is greater than the IL group as it is also depicted in the graph of bar plot. The KI group has some skewed distribution, that is, it is not normally distributed, however from the boxplot once again one can assume the IL group to be normally distributed. Also, since the notches in the boxplot do not overlap, one can graphically conclude that with 95% confidence, the true averages do differ. However, once again it only needs a strong statistical analysis to confirm.

Distribution of closeness centrality between groups

Closeness centrality measure between IL and KI groups

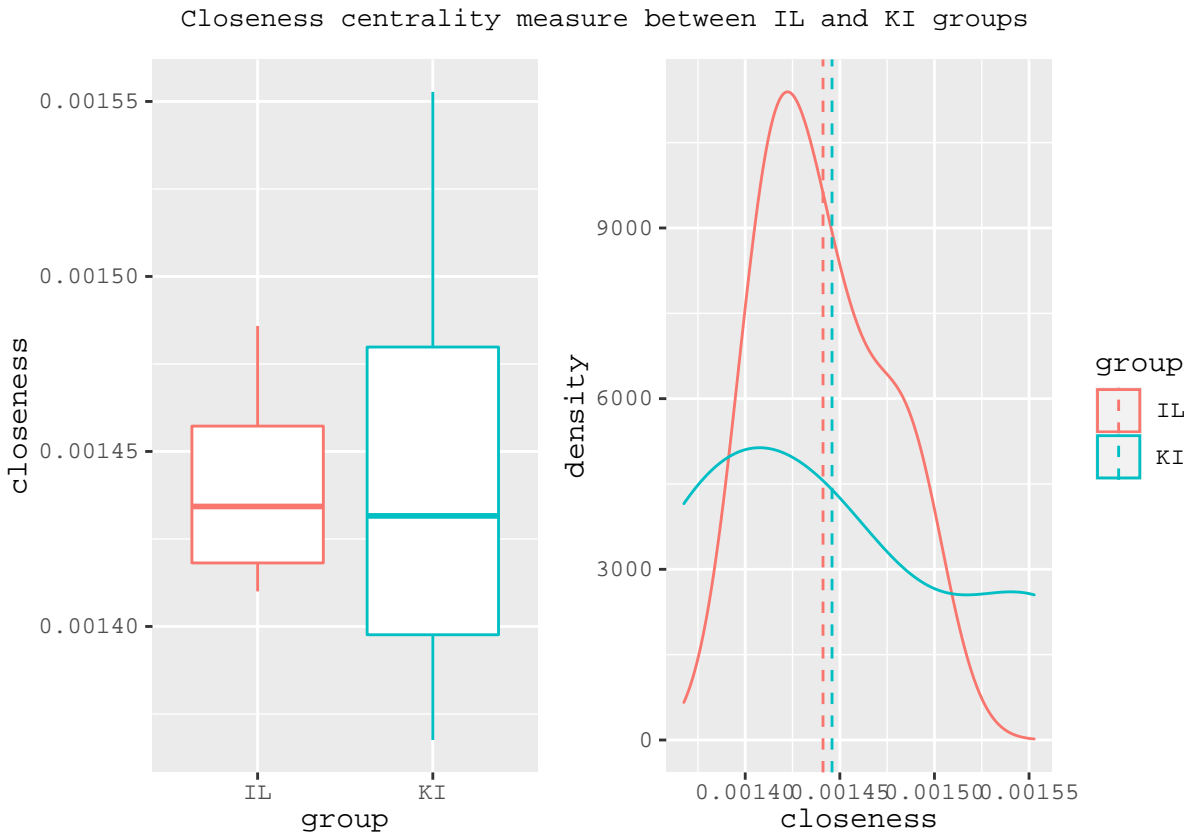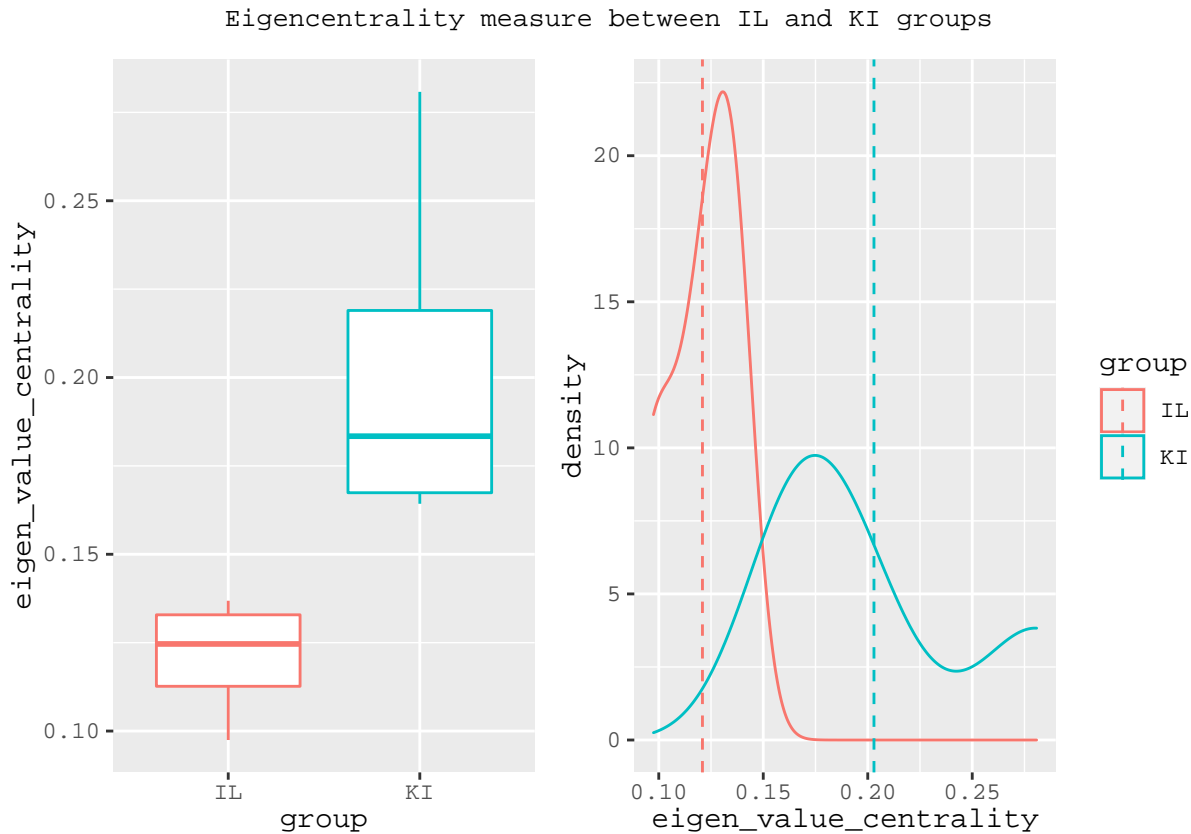Given the plot of closeness centrality between the two groups, we observe that there is greater variability in KI group area_average to the KI group. The KI group has the highest and least closeness observation between the two groups given the connectivity of the prefrontal cortex to others. On average from the boxplot, the closeness measure from the IL group is somehow equal to the KI group as it is also depicted in the graph density plot by the vertical dash lines. The two samples' closeness measures have some kind of normal distribution in the boxplot but not clearly shown in the density plot. Since the notches in the boxplot, somehow tend to overlap as well as the dash lines, one can graphically conclude with some certainty that, at 95% confidence, the true averages do not differ. An accurate statistical analysis is required to accept this graphical conclusion.

Distribution of eigen value centrality between groups

Eigencentrality measure between IL and KI groups

Given the plot of eigen centrality between the two groups, we observe that there is greater variability in KI group area_average to the KI group. The minimum value of the KI group however still higher than the maximum value or observation of the IL group from the pre-frontal cortex region. On average from the boxplot, the eigen centrality measure from the KI group is much greater than the IL group as it is also depicted in the graph density plot by the vertical dash lines. The two samples have some skewed distribution, that is they are not normally distributed as clearly shown in the density plot. Again, the notches in the boxplot do not overlap as well as the dash lines, one can graphically conclude that with 95% confidence, the true averages do differ. Statistical analysis will further be used to confirm this.

Methodology and Approach

Given the resting-state fMRI data obtained from the IL(male voles) and KI (male offspring voles), it is clearly identified that the two samples are dependent. Hence a paired sample test was appropriate for the analysis.

Statistical assumptions

After the exploratory analysis and graphical conclusions, one could say given the mean centrality measure such as the degree, the betweenness, and the eigen centrality the two groups differ. However, such a conclusion can only be concrete after a suitable and accurate test is carried out on each measure between the two groups.

Assumptions

- Paired data sample (hence a paired test was appropriate)

- Small sample size ($n<$ 30) : Since the sample size was not large (less than 30), normality assumption was carried on the differences between the two groups to decide the appropriate test.

- Normality check : The Shapiro-wilk normality test was used to test for the normality of the difference data. This test was viable given that the sample size $n<$ 30 (that is a small sample size).

Test for normality on the differences

Centrality	Shapiro P-value	Normality
Degree	0.0012	No
Betweenness	0.0722	Yes
Closeness	0.2357	Yes

Centrality	Shapiro P-value	Normality
Eigenvalue	0.7477	Yes

From the output, the p-values for betweenness, closeness and eigen value centralities are greater than the significance level 0.05 implying that the distribution of the differences in the IL and KI groups are not significantly different from normal distribution. In other words, we can assume normality, hence paired t-test was appropriate for their comparison.

However, the p-value for the degree centrality was less than the significance level 0.05 implying that the distribution of the differences in the IL and KI groups are significantly different from normal distribution. In other words, we cannot assume normality, hence a non-parametric test is appropriate.

Statistical analysis and Results

For the purpose of comparison and uniformity, all statistical tests involving the four centrality measures were subject to the follow;

Problem

Decide at .05 significance level, if there is a significant difference in the social behavior of less social (IL) and more social (KI) groups of male praire voles based on their centrality measures estimated.

- $\alpha = 0.05$ is the critical region (value).

- $n_{IL}= 4$ is the sample size for the IL group for the Pre-frontal cortex region.

- $n_{KI} = 4$ is the sample size for the IL group for the Pre-frontal cortex region.

- $n_d = 4$, is the sample size for the differences in group (i.e $n < 30$).

Hypothesis

- H_0 = There is no significant difference in social behaviour of the IL and KI groups of male prairie voles based on their centrality measures estimated.

- H_1 = There is a significant difference in social behaviour of the IL and KI groups of male prairie voles based on their centrality measures estimated

Decision rule

- Reject the null hypothesis (H_0) if the obtained $P_{value} < \alpha = 0.05$.

Wilcoxon signed-rank test (Paired) on Degree

- This test does not rely on assumption that, the data is drawn from a normal distribution, thus the use of variance becomes meaningless.

- Dependent samples: the two samples are dependent observations of the cases.

- Independence: The Wilcoxon signed-rank test assumes independence, meaning that the **paired** observations are randomly and independently drawn.

Degree centrality between groups

Wilcoxon signed–rank test on the two groups using their degree
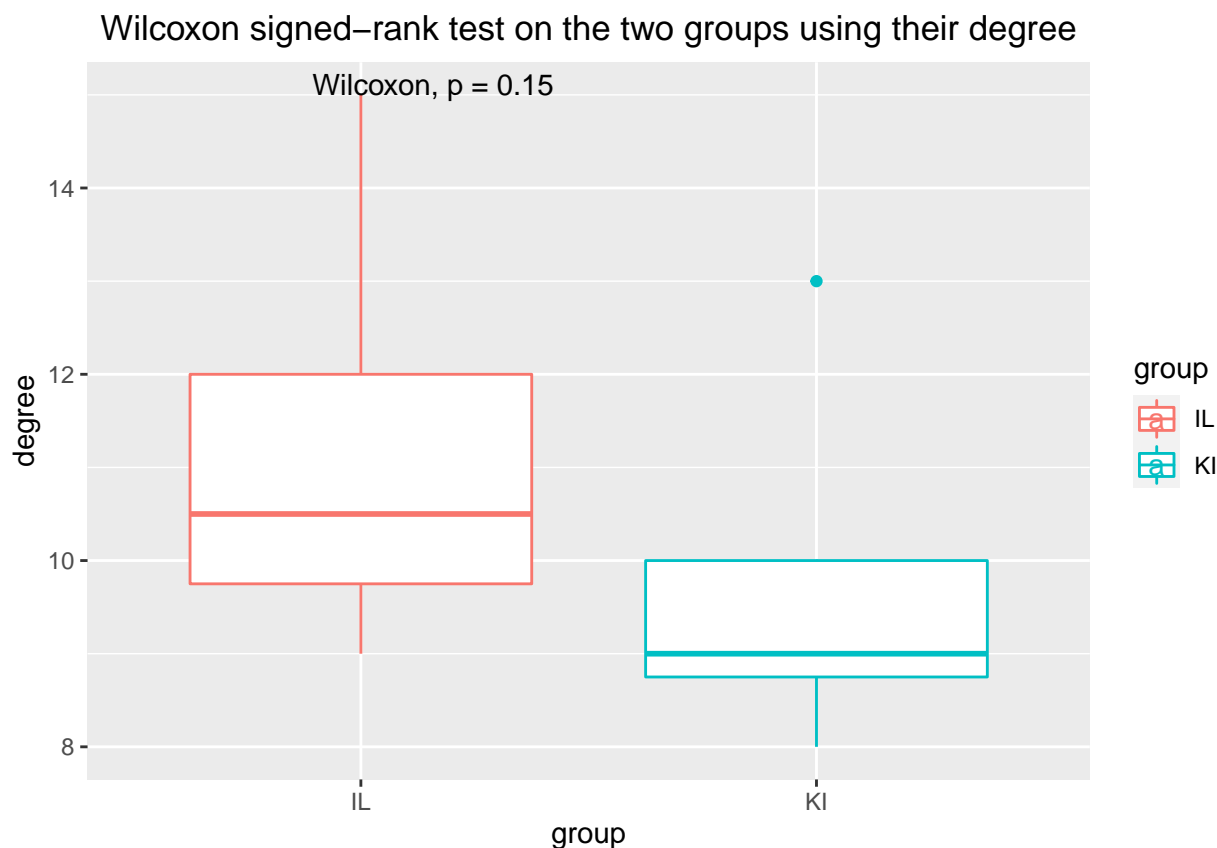
In the statistical analysis by the Wilcoxon signed-rank test, the p-values turn out to be 0.15, which is greater than the 0.05 significance level, hence we fail to reject the null hypothesis. There is insufficient evidence at 0.05 significance level to conclude that, based on the degree centrality there is a significant difference in the social behavior of the IL and KI groups of male prairie voles given their pre-frontal cortex region.
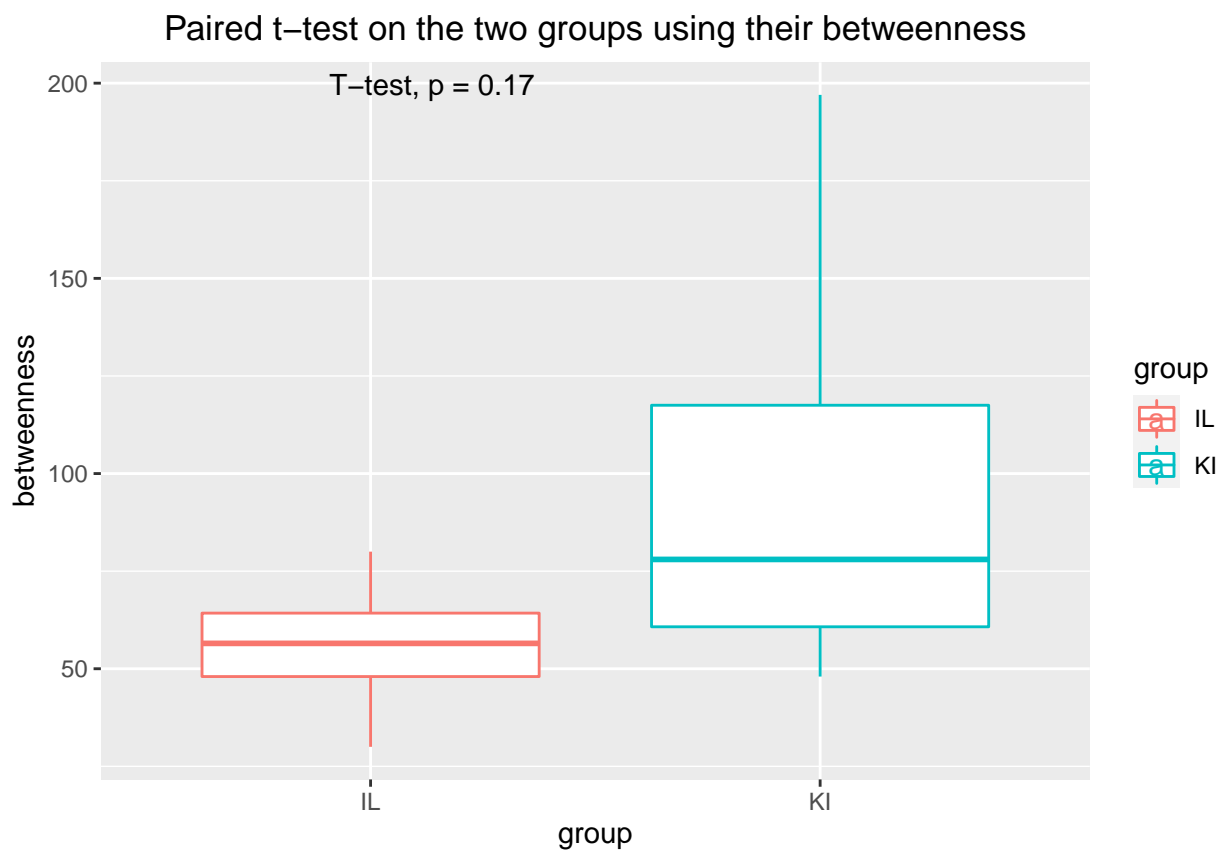
**Paired t-test on Betweenness, Closeness and Eigen Value**

- The sample size is less than 30 ($n < 30$).

- The two groups are dependent (Paired data)

- The distribution of the population of which the sample are drawn from are normally distributed (from the shapiro test).

Betweenness centrality between groups

Paired t–test on the two groups using their betweenness

In the statistical analysis by the paired sample t-test, the p-value from the test of betweenness turn out to be 0.17 which is greater than the 0.05 significance level, hence we fail to reject the null hypothesis. There is insufficient evidence at 0.05 significance level to conclude that, based on the betweenness centrality there is a significant difference in the social behavior of the IL and KI groups of male prairie voles given their pre-frontal cortex region.

Closeness centrality between groups

Paired t–test on the two groups using their closeness
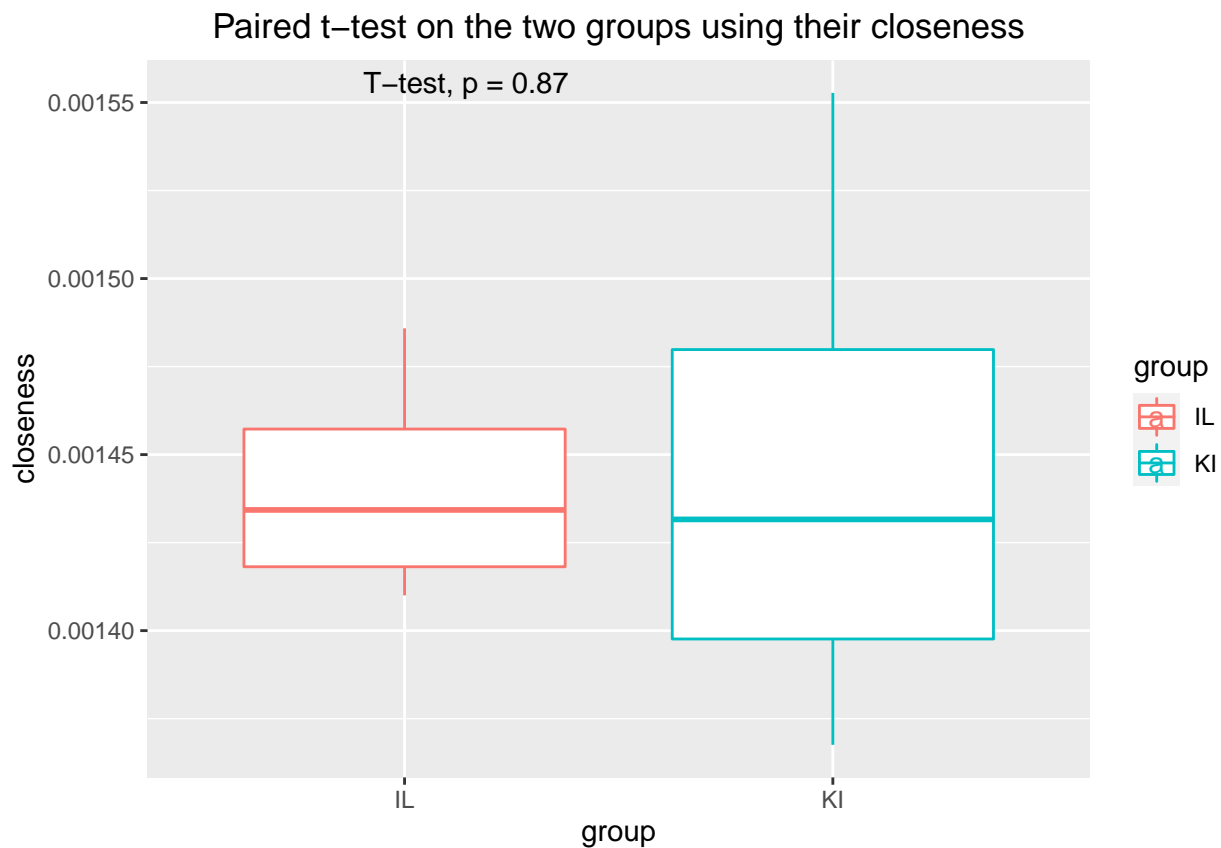
In the paired t-test for the closeness centrality between the groups, the p-value from the test turn out to be 0.87 which is greater than the 0.05 significance level, hence we fail to reject the null hypothesis. There is insufficient evidence at 0.05 significance level to conclude that, based on the closeness centrality there is a significant difference in the social behavior of the IL and KI groups of male prairie voles given their pre-frontal cortex region.

Eigen centrality between groups

Paired t–test on the two groups using their eigen_centrality

T–test, p = 0.047
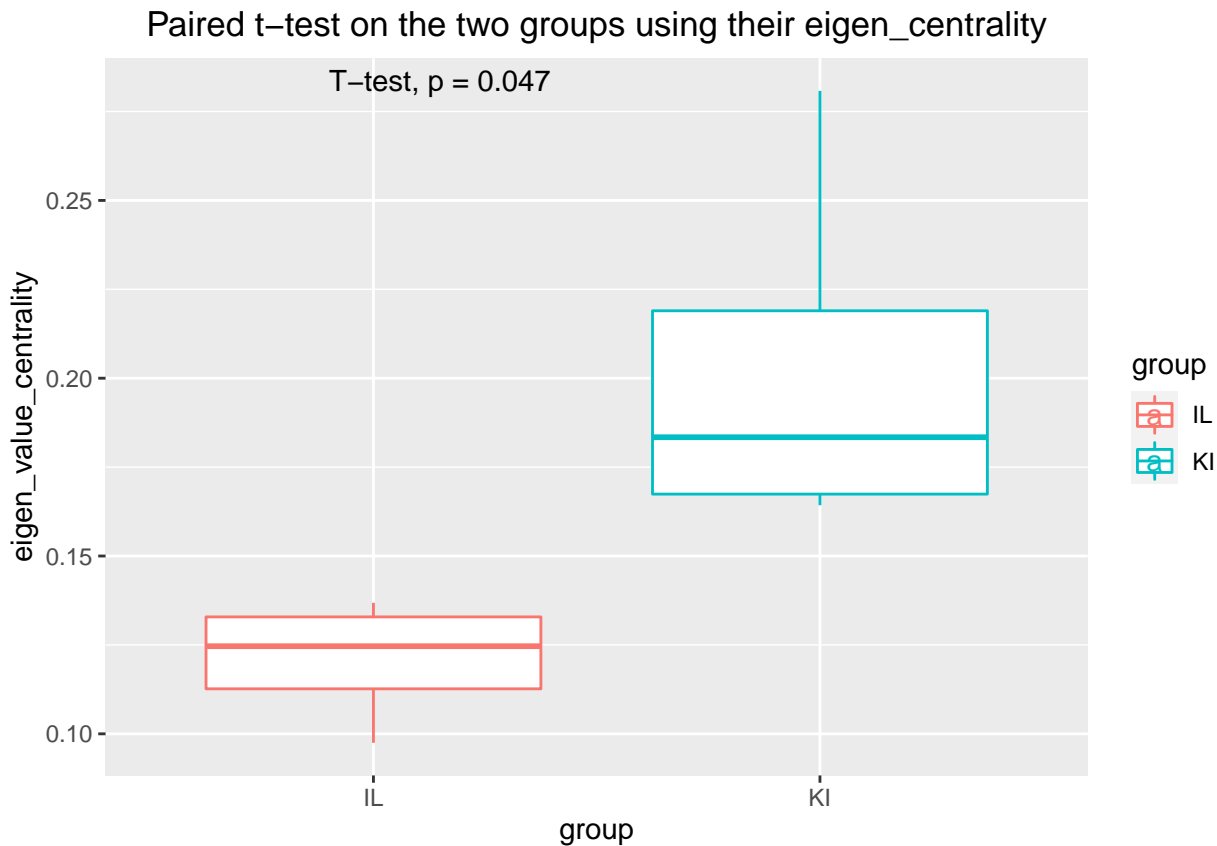
In the paired t-test for the Eigen centrality between the groups, the p-value from the test turns out to be 0.047 which is less than the 0.05 significance level, hence we reject the null hypothesis. There is sufficient evidence at a 0.05 significance level to conclude that, based on the Eigen centrality there is a statistically significant difference in the social behavior of the IL and KI groups of male prairie voles given their pre-frontal cortex region.

Conclusion

Praire vole is a rodent noted to exhibit some kind of social networking behavior based on geographical location. A test has earlier shown that if an offspring is obtained from the two groups of voles (that is a KS and IL) to obtain KI, it turns to exhibit less social behavior. However, there is a need to determine if there is a significant difference in the social behavior of more social (IL) and less social (KI) groups of male prairie voles based on their centrality measures from

a social graph network by means of statistical analysis. Specifically, centrality estimates from the prefrontal cortex region of the brain were highlighted for comparison. Sound statistical tests were carried out on these measures, and it turns out that, given the highlighted region, the only measure which indeed distinguished the IL and KI groups of voles was the eigenvalue centrality. Therefore the two groups are indeed significantly different in their social behavior given the Eigen centrality. However, not enough information available to draw the same conclusion in the case of the degree, betweenness and closeness measure. Hence, if the significance level is varied or the entire brain is considered for the comparison, there may be a higher chance of the other measures sufficiently distinguishing the two groups. Therefore there is statistical significance difference in the two groups of voles given the Eigen centrality from the prefrontal cortex region.

Recommendation

Given that, the statistical analysis was based on a particular subregion of the brain (that is the prefrontal cortex), decisions made here might be fairly biased. Therefore, I recommend that in order to have a sound conclusion, measures obtained from all subregions between the two groups should be statistically analyzed together between groups. This is because the pre-frontal alone may or may not be a factor for differentiating, however, may be influenced by other regions.

Reference

Anacker, Allison MJ, Monique L Smith, and Andrey E Ryabinin. 2014. "Establishment of Stable Dominance Interactions in Prairie Vole Peers: Relationships with Alcohol Drinking and Activation of the Paraventricular Nucleus of the Hypothalamus." *Social Neuroscience* 9 (5): 484–94.

Aragona, Brandon J, and Zuoxin Wang. 2004. "The Prairie Vole (Microtus Ochrogaster): An Animal Model for Behavioral Neuroendocrine Research on Pair Bonding." *Ilar Journal* 45 (1): 35–45.

Heuvel, Martijn P van den, and Olaf Sporns. 2013. "Network Hubs in the Human Brain." *Trends in Cognitive Sciences* 17 (12): 683–96.

Sporns, Olaf. 2018. "Graph Theory Methods: Applications in Brain Networks." *Dialogues in Clinical Neuroscience* 20 (2): 111.

Sporns, Olaf, Christopher J Honey, and Rolf Kötter. 2007. "Identification and Classification of Hubs in Brain Networks." *PloS One* 2 (10): e1049.

Walum, Hasse, and Larry J Young. 2018. "The Neural Mechanisms and Circuitry of the Pair Bond." *Nature Reviews Neuroscience* 19 (11): 643–54.

Appendix

Codes

- Loading Package

```
#### Loading required packages
library(patchwork)

library(asnipe)

library(igraph)

library(ggraph)

library(plyr)

library(ggpubr)

library(dplyr)
```

- Loading in data

```
IL_group <- read.csv(file="IL_group.csv", header=TRUE, check.names =FALSE,
        na.strings = c("NA", "", " "))


KI_group<- read.csv(file="KI_group.csv", header=TRUE, check.names =FALSE,
        na.strings = c("NA", "", " "))
```

- Defining Prefrontal region

```
target = c("'frontal association ctx'", "'infralimbic ctx'",
          "'orbital ctx'","'prelimbic ctx'")


rownames(IL_group)<-IL_group$Region

il<- as.matrix(IL_group[,-1])
```

```
rownames(KI_group)<-KI_group$Region

ki<- as.matrix(KI_group[,-1])
```

- Network Analysis with entire data set from both group of voles.

```
g=graph_from_adjacency_matrix(il, mode="undirected", weighted=T)

V(g)$color <- ifelse(colnames(il) %in%target, "navyblue", "lightgreen")


g_ki = graph_from_adjacency_matrix(ki, mode="undirected", weighted=T)

V(g_ki)$color <- ifelse(colnames(ki) %in%target, "tomato", "pink")


par(mfrow=c(1,2))

plot(g, edge.width=E(g)$weight, main = "Network for the IL_group",

     vertex.label= NA, vertex.size = 10 )

legend("topleft", legend = c("Pre-frontal cortex subregion",

                             "Remaining subregions")

       ,pt.cex=1.5, pt.bg = c("navyblue", "lightgreen"), pch = 21, cex = 0.5,

       bty = "n", ncol = 1)

plot(g_ki, edge.width=E(g_ki)$weight, main = "Network for the KI_group",

     vertex.label= NA, vertex.size = 10)

legend("topleft", legend = c("Pre-frontal cortex subregion",

                             "Remaining subregions")

       ,pt.cex=1.5, pt.bg = c("tomato", "pink"), pch = 21, cex = 0.5,

       bty = "n")
```

- Obtaining centrality measures

```r
#### Ilinois
be_IL=betweenness(g)

de_IL=degree(g)

cl_IL=closeness(g)

ei_IL=eigen_centrality(g)

eigen<-ei_IL$vector

names(eigen)<-NULL


#### for the kansas
be_ki=betweenness(g_ki)

de_ki=degree(g_ki)

cl_ki=closeness(g_ki)

ei_ki=eigen_centrality(g_ki)

eigen_ki<-ei_ki$vector

names(eigen_ki)<-NULL
```

- Subsetting (prefrontal cortex) region from graph centrality measures

```r
IL_group_data<- as.data.frame(cbind("subregion" = rownames(il),

                                    "degree" = de_IL, "betweenness"= be_IL,

                                    "closeness"= cl_IL,

                                    "eigen_value_centrality" = eigen))


KI_group_data<- as.data.frame(cbind("subregion" = rownames(ki),

                                    "degree" = de_ki, "betweenness"= be_ki,

                                    "closeness"= cl_ki,

                                    "eigen_value_centrality" = eigen_ki))
```

```r
IL_result<- IL_group_data %>%

  filter(subregion %in% target) %>%

  mutate(group = c("IL","IL","IL","IL")) %>%

  relocate(group, .before = subregion)


KI_result<- KI_group_data %>%

  filter(subregion %in% target) %>%

  mutate(group = c("KI","KI","KI","KI")) %>%

  relocate(group, .before = subregion)
```

- Combining the two table results for further analysis

```r
data<- bind_rows(IL_result, KI_result)

data$subregion <- gsub("'", "",data$subregion)

rownames(data)<- gsub("'","",rownames(data))

data[,c(3:6)]<- apply(data[,c(3:6)], 2, as.numeric)

head(data)
```

- Degree measure centrality distribution

```r
###Degree measure centrality distribution


# Plotting the first graph

p1<-ggplot(data, aes(x = group, y = degree, color=group))+

  geom_boxplot()+

  theme(legend.position = "none")


#Plotting the second graph(density plot)

p2<-ggplot(data, aes(fill=group, y=degree, x=subregion)) +
```

```
    geom_bar(position="dodge", stat="identity",alpha=0.7)


p1 + p2 + plot_annotation(title ='Distribution of degree centrality measure

                          between IL and KI groups',

                theme = theme(plot.title = element_text(size = 10))) &

  theme(text = element_text('mono')) & theme(plot.title =

                                      element_text(hjust = 0.5))
```

- Betweenness centrality measure distribution

```
###Betweenness centrality measure


# Plotting the first graph

p3<-ggplot(data, aes(x = group, y = betweenness, color= group, palette = "jco"))+

  geom_boxplot()+

  theme(legend.position = "none")


#Plotting the second graph(density plot)

p4<-ggplot(data, aes(fill=group, y=betweenness, x=subregion)) +

    geom_bar(position="dodge", stat="identity" ,alpha=0.7)


p3 + p4 + plot_annotation(title = 'Betweenness centrality measure between IL

                          and KI groups',

                theme = theme(plot.title = element_text(size = 10))) &

  theme(text = element_text('mono')) & theme(plot.title =

                                      element_text(hjust = 0.5))
```

- Closeness centrality measure between the groups distribution

```
# Plotting the first graph

p5<-ggplot(data, aes(x = group, y = closeness, color= group,

                    palette = "jco"))+

  geom_boxplot()+

  theme(legend.position = "none")


#Obtaining summary statistucs for the two groups

mu <- ddply(data, "group", summarise, grp.mean=mean(closeness))


#Plotting the second graph(density plot)

p6<-ggplot(data, aes(x=closeness, color=group)) +

  geom_density(alpha=0.4)


# Add mean lines

p66<-p6+geom_vline(data=mu, aes(xintercept=grp.mean, color=group),

            linetype="dashed")


p5 + p66 + plot_annotation(title = 'Closeness centrality measure

                        between IL and KI groups',

              theme = theme(plot.title = element_text(size = 14))) &

  theme(text = element_text('mono'))
```

- eigen centrality measure between the groups distribution

```
# Plotting the first graph

p7<-ggplot(data, aes(x = group, y = eigen_value_centrality, color= group,

                    palette = "jco"))+

  geom_boxplot()+

  theme(legend.position = "none")
```

```r
#Obtaining summary statistucs for the two groups
mu <- ddply(data, "group", summarise, grp.mean=mean(eigen_value_centrality))


#Plotting the second graph(density plot)
p8<-ggplot(data, aes(x=eigen_value_centrality, color=group)) +

  geom_density(alpha=0.4)


# Add mean lines
p88<-p8+geom_vline(data=mu, aes(xintercept=grp.mean, color=group),

              linetype="dashed")


p7 + p88 + plot_annotation(title = 'Eigen value centrality measure

                          between IL and KI groups',

              theme = theme(plot.title = element_text(size = 10))) &

  theme(text = element_text('mono'))
```

- Test for normality on the differences

```r
##Test for normality on the differences


my_data<- data[,c(1,3:6)]
diff_deg<-with(my_data, degree[group == "IL"] - degree[group == "KI"])
diff_bet<-with(my_data, betweenness[group == "IL"] - betweenness[group == "KI"])
diff_clo<-with(my_data, closeness[group == "IL"] - closeness[group == "KI"])
diff_eig<-with(my_data, eigen_value_centrality[group == "IL"] -

              eigen_value_centrality[group == "KI"])


diff_table<- as.data.frame(cbind("Degree"= diff_deg, "Betweenness"= diff_bet,
```

```
                                      "Closeness"= diff_clo,

                                           "Eigenvalue"= diff_eig))

test<- Map(function(x)cbind(shapiro.test(x)$p.value), diff_table)



com_test<- as.data.frame(cbind("Centrality"= c("Degree", "Betweenness",

                         "Closeness","Eigenvalue"), "Shapiro P-value" = lapply(c(te

                         "Normality"= c("No", "Yes", "Yes", "Yes")))

knitr::kable(com_test, align = "lcc")
```

- Statistical comparison for degree

```
set.seed(125)

p<-ggplot(data, aes(x = group,y = degree, color= group))+

  geom_boxplot()+

  ggtitle("Wilcoxon signed-rank test on the two groups using their degree")+

  theme(legend.position = "right", plot.title = element_text(hjust = 0.5))

p + stat_compare_means(paired = T)
```

- Statistical comparison for betweenness

```
set.seed(125)

p<-ggplot(data, aes(x = group, y = betweenness, color= group))+

  geom_boxplot()+

  ggtitle("Paired t-test on the two groups using their betweenness")+

  theme(legend.position = "right",plot.title = element_text(hjust = 0.5))

p + stat_compare_means(method = "t.test", paired = T)
```

- Statistical comparison for closeness

```r
set.seed(125)

p<-ggplot(data, aes(x = group, y = closeness, color= group))+

  geom_boxplot()+

  ggtitle("Paired t-test on the two groups using their closeness")+

  theme(legend.position = "right",plot.title = element_text(hjust = 0.5))

p + stat_compare_means(method = "t.test",paired = T)
```

- Statistical comparison for eigen centrality

```r
set.seed(125)

p<-ggplot(data, aes(x = group, y = eigen_value_centrality, color= group))+

  geom_boxplot()+

  ggtitle("Paired t-test on the two groups using their eigen_centrality")+

  theme(legend.position = "right",plot.title = element_text(hjust = 0.5))

p + stat_compare_means(method = "t.test", paired = T)
```