

Ant colony

Jan Lennartz & Andrei Chirita

11/22/2020

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

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

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

##
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

Contents

Introduction	3
The data	3
Preliminary data exploration	3
The original paper	5
Our questions	6
Descriptive Statistics of the network	6
For the general network	6
Average Path Length and Diameter	7
Measures of Centrality	7
Clustering Coefficient	12
By group	12
Histogram and Density	12
Connectivity and Components	16
Average Path Length and Diameter	16
Centrality Measures	17
Clustering Coefficient	23
K-means check	23
Preparing the data	23
Attempt 2	26

Introduction

Ant colonies have a complex and fascinating social structure that may bring answers to a multitude of scientific questions. Usually the nest are organized in a stratified manner with a queen at the center and numerous workers doing tasks needed for the upkeep of the colony. The study for which the data we worked on was collected sought to understand the social structure of *Camponotus fellah* ants, what are the groups inside the colonies and what factors define them.

The data

The original data of the study consist of more than 9 million observed interactions between ants collected for 41 days from ants belonging to 6 colonies. For our project we decided to work on

Preliminary data exploration

In the following section we will show how the data is structured for a single colony in a single day (namely day 17, colony 1). We will start by loading the data using the `get_graph` function that we built previously.

```
g <- get_graph(colony = 1, day = 17)
```

Next we will have a look at the vertices of the graph, as can be seen there are 99 of them in this graph.

```
vcount(g)
```

```
## [1] 99
```

Next we displayed the number of edges, as can be seen there are over 3300 edges. Thus in day 17, for the first colony there were over 3342 interactions between 99 ants.

```
ecount(g)
```

```
## [1] 3342
```

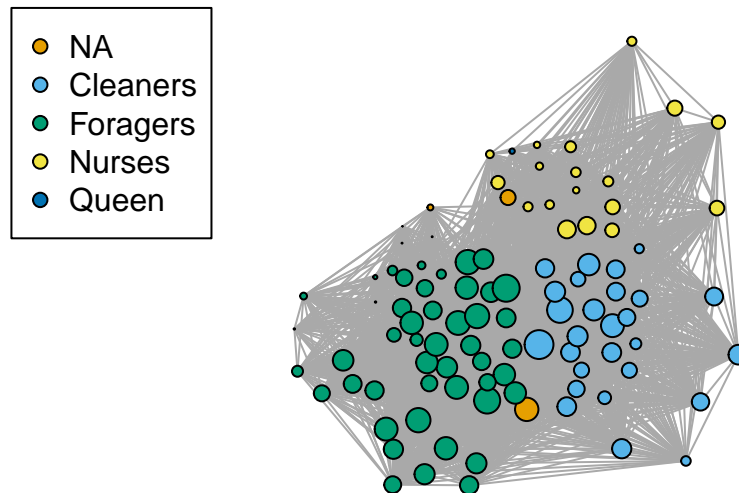
```
str(vertex.attributes(g))
```

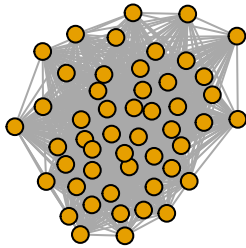
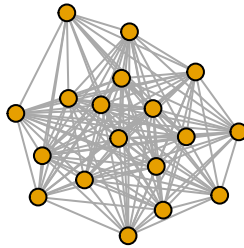
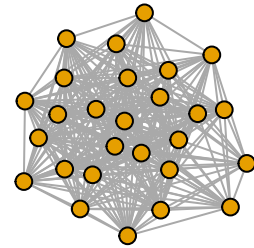
```
## List of 19
## $ nb_interaction_queen : num [1:99] NaN 8.636 0 1.091 0.0909 ...
## $ nb_interaction_foragers: num [1:99] NaN 68.4 681.3 86.5 88.2 ...
## $ nb_interaction_cleaners: num [1:99] NaN 169.4 65.4 126.8 143.7 ...
## $ nb_interaction_nurses : num [1:99] NaN 161.64 9.09 97.27 37.64 ...
## $ visits_to_rubbishpile : num [1:99] NaN 4.6364 0.2727 0.0909 50.2727 ...
## $ visits_to_nest_entrance: num [1:99] NaN 0.909 166 0 0.818 ...
## $ visits_to_brood : num [1:99] NaN 137.909 0.636 85.909 30.909 ...
## $ group_period4 : chr [1:99] "" " " " "F" "N" ...
## $ group_period3 : chr [1:99] "" " " " "F" "N" ...
## $ group_period2 : chr [1:99] "" "C" "F" "C" ...
## $ group_period1 : chr [1:99] "" "N" "F" "C" ...
## $ nb_foraging_events : num [1:99] NaN 0 15 0 0 0 0 0 43 ...
## $ age(days) : num [1:99] NaN 57 246 372 71 190 351 71 57 344 ...
```

```
## $ body_size      : num [1:99] NaN 141 215 149 124 ...
## $ tag_id         : num [1:99] NaN 620 356 540 115 117 113 191 190 217 ...
## $ colony         : num [1:99] NaN 4 4 4 4 4 4 4 4 4 ...
## $ id             : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
## $ group          : chr [1:99] "" "C" "F" "C" ...
## $ name           : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
```

Each vertex has also a set of attributes like:

- Several attributes that are useful for understanding the interactions of the studied
- Attributes that register the visits of the ant to important places of the colony (like the brood or the nest entrance)
- The groups fitted by the authors of the study
- The age of the ant (measured in days)
- The body size of the ant



Foragers**Nurses****Cleaners**

```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
```

As we can see the foragers and cleaners groups contain more ants than the nurses group and they also contains more interactions.

The original paper

The original paper was written by: Danielle P. Mersch, Alessandro Crespi and Laurent Keller and explores questions related how can we separate ant colonies into groups and what makes ants change the group they belong to. During their study they found 3 main groups based on the interactions between ants and concluded that age is the main factor that determines ants to change the group they are part of. All colonies studied had 4-years old queens and between 122 and 192 workers per colony. Each ant was marked and followed individually and an interaction between two ants were defined by the fact that “the front end of one ant was located within the trapezoidal shape representing the other ant”.

Our questions

The goal of this work is to conduct the given data set w.r.t. various aspects. Furthermore, a validation of the key results of the original paper is carried out. For this a sample colony is chosen on a given day. On this network the analysis is to be done. Certain properties will be evaluated over multiple days when required.

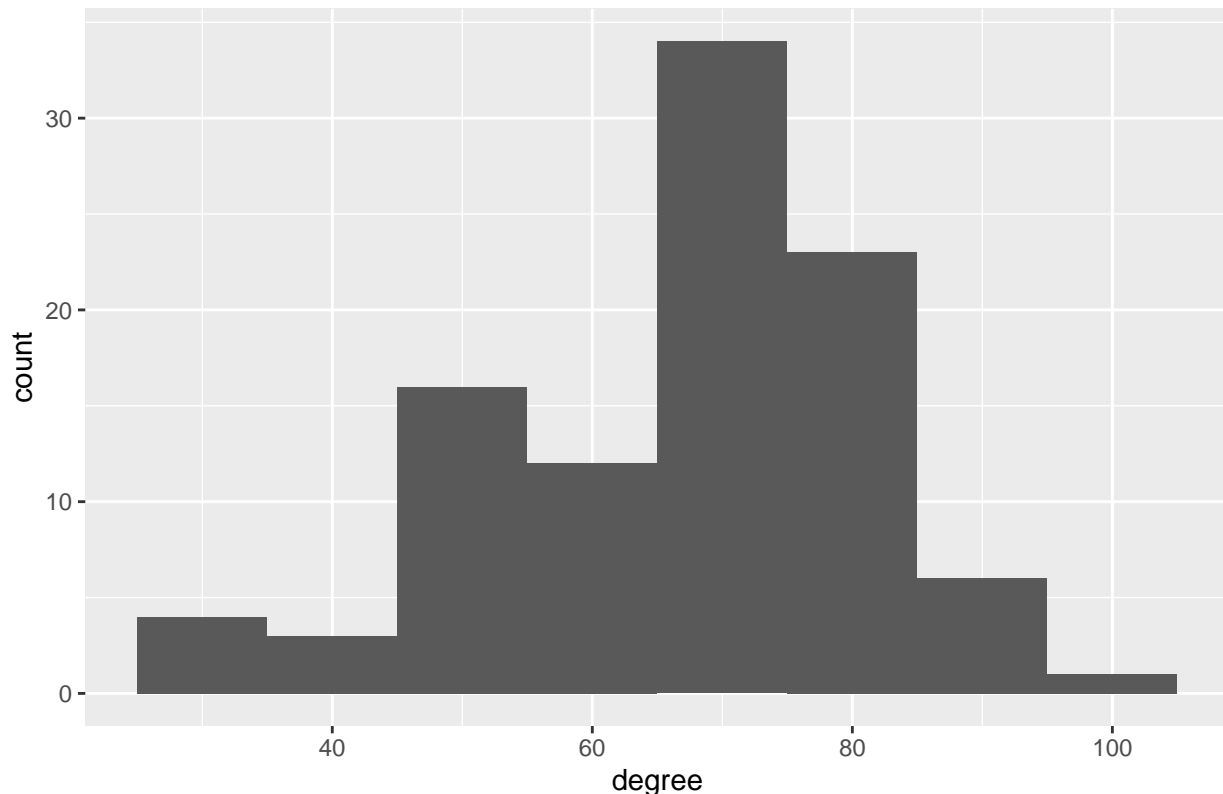
We will first explore the network in a descriptive manner. This includes characteristics like degree distribution, density, diameter and more. In the second step we have a closer look at the groups. First, we validate that the three groups are a valid proposal for the given network. This is done by running a clustering algorithm on the network to identify the groups which will be compared to the labeled groupings. Second, we investigate how frequently ants communicate within groups and compare this to the level of communication between groups. We answer the question how fast information can be spread in the network and compare this to the result of the paper. Additionally, we calculate the centrality of specific ant or groups (e.g. the queen) w.r.t. different measures. Furthermore, we review several properties of the ants and their correlation with the groups (e.g. age, size).

Descriptive Statistics of the network

As we noted above the data we are going to use for our analysis refers to the movements of the ants of the first colony in day 17 of the study. Thus the graph has 99 vertices and 3342 edges. We must also note that the graph is connected.

For the general network

Histogram of the degrees



```
## [1] "Distribution of the degrees: "

## [1] "[0-30] : 0"

## [1] "(30-60] : 27"

## [1] "(60-90] : 70"

## [1] ">90 : 2"

## [1] "min : 33"

## [1] "max : 96"
```

Most nodes have a rather higher degree (between 60 and 90) and there are some nodes with a lower value of the degree, thus we can infer that most ants had a higher number of contacts while some ants had less frequent meetings. There is a small number of nodes that has a very high degree (90 or over 90), they represent ants that had a lot of encounters. We must also note that no ant had less than 33 encounters.

Average Path Length and Diameter

The average path length can show how connected the analyzed network is. In the case of our graph the value is approximately 1.31. Another measure of the connectivity of the graph is its diameter (the maximum of the series of shortest paths), the value of the indicator is: 5

```
average.path.length(g)
```

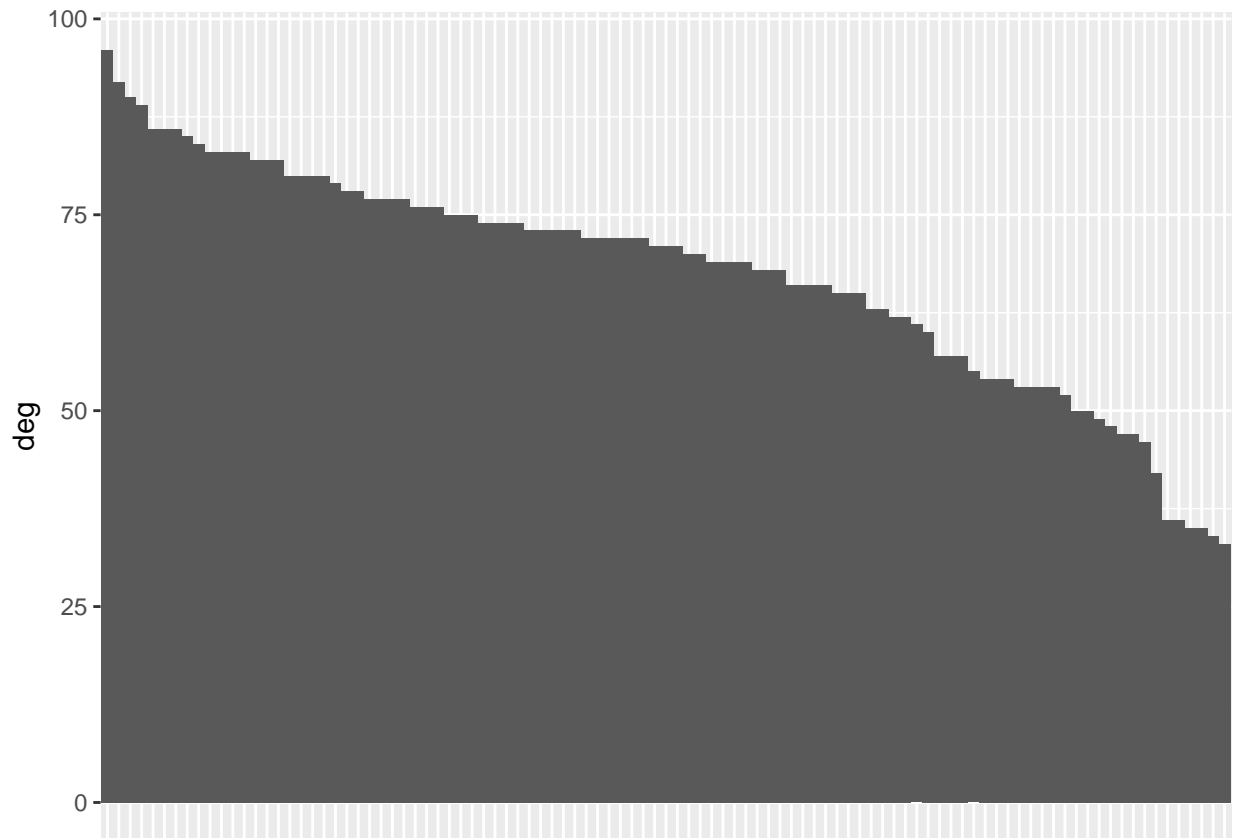
```
## [1] 1.31107
```

```
diameter(g)
```

```
## [1] 5
```

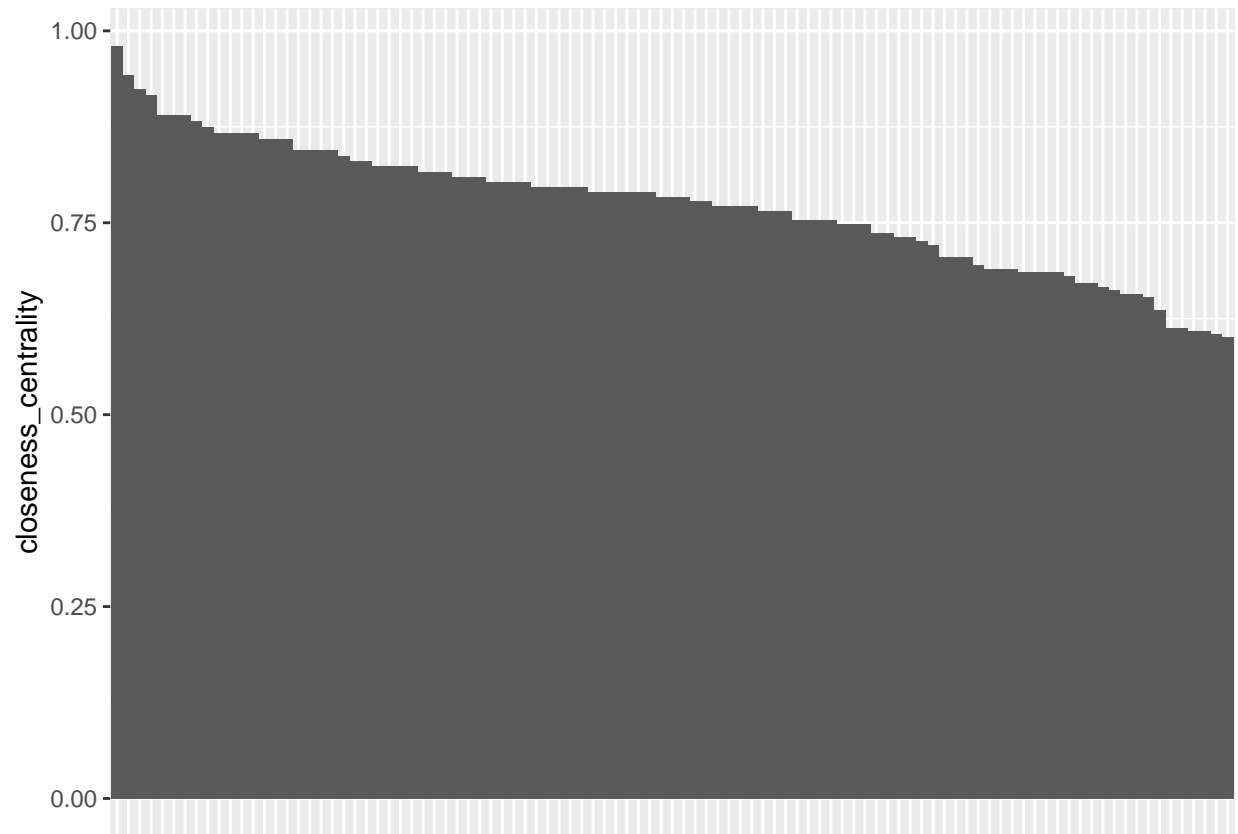
Measures of Centrality

```
deg <- centralization.degree(g)
deg <- data.frame(index=factor(1:length(deg$res),levels=1:length(deg$res)),deg=deg$res[order(deg$res, d
ggplot(deg)+aes(y=deg,x=index)+geom_col(width=1)+
  theme(axis.title.x = element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()))
```

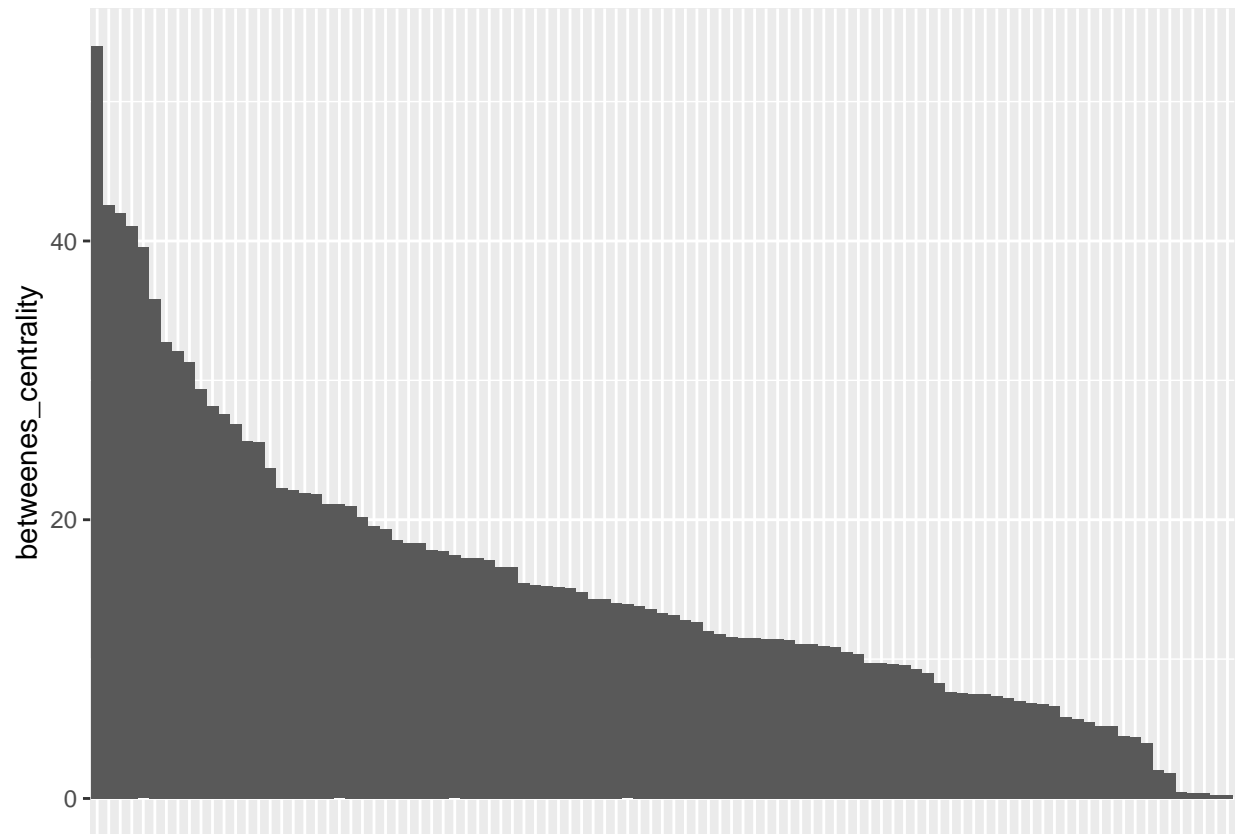


?geom_col

```
deg <- centralization.closeness(g)
deg <- data.frame(index=factor(1:length(deg$res),levels=1:length(deg$res)),closeness_centrality=deg$res)
ggplot(deg)+aes(y=closeness_centrality,x=index)+geom_col(width=1)+
  theme(axis.title.x = element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```

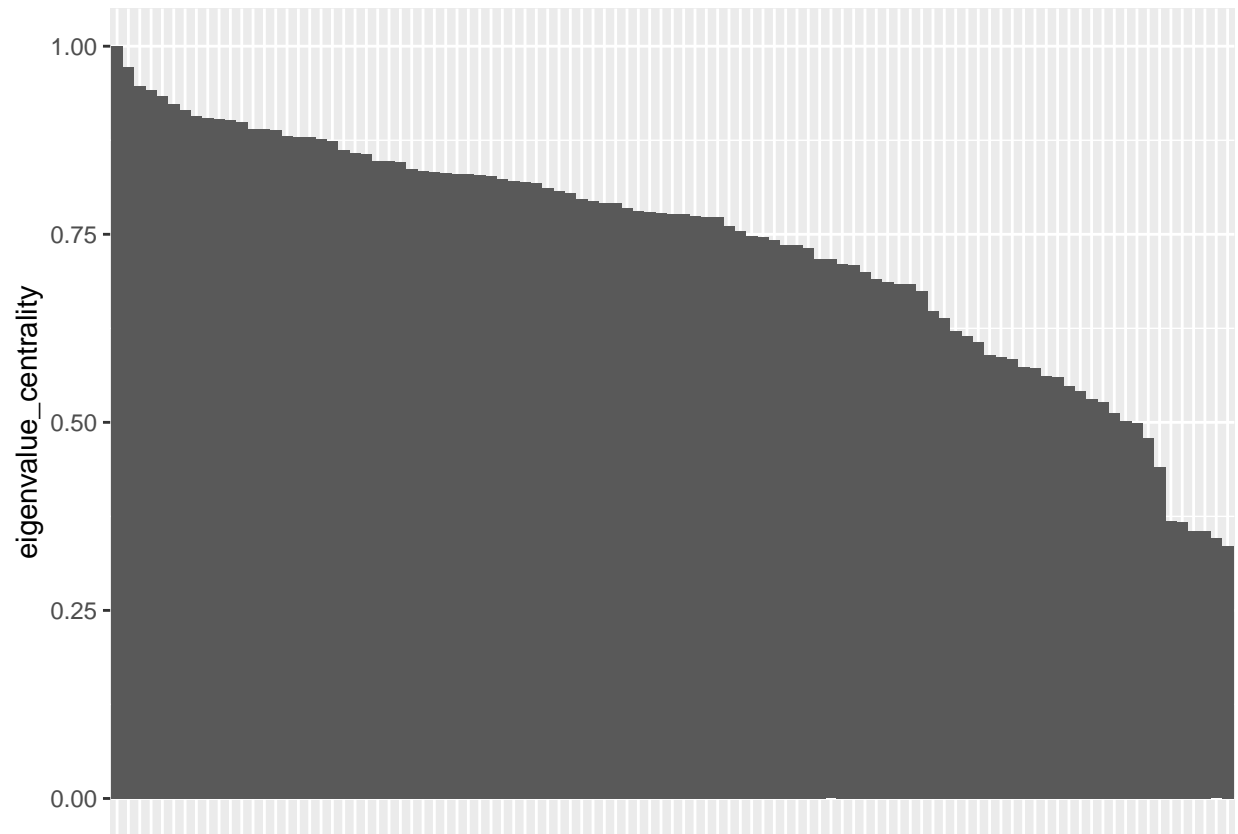



```
deg <- centralization.betweenness(g)
deg <- data.frame(index=factor(1:length(deg$res),levels=1:length(deg$res)),betweenes Centrality=deg$res)
ggplot(deg)+aes(y=betweenes Centrality,x=index)+geom_col(width=1)+
  theme(axis.title.x = element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```



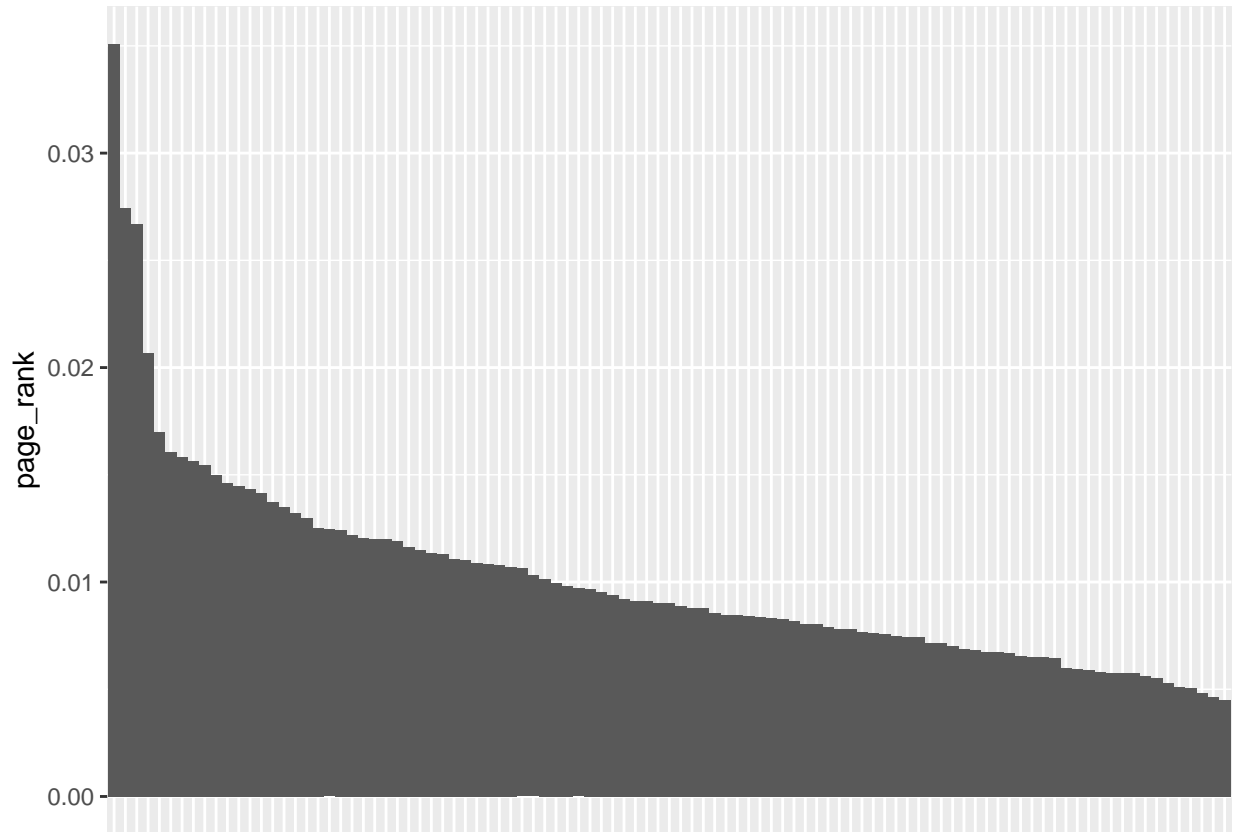
In regard to betweenness centrality we can observe that one ant ranks far over the others, thus we can conclude that there is one ant that has a very central role in the network a conclusion we couldn't draw based on the other measures of centrality.

```
deg <- centralization.evcent(g)
deg <- data.frame(index=factor(1:length(deg$vector),levels=1:length(deg$vector)),eigenvalue_centrality=
ggplot(deg)+aes(y=eigenvalue_centrality,x=index)+geom_col(width=1)+
  theme(axis.title.x = element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()))
```



The eigenvalue centrality presents a decreasing structure although it is a smoother decrease compared to the one of the betweenness centrality, in this case we can find no node that is much more “central” than the others.

```
deg <- page.rank(g)
deg <- data.frame(index=factor(1:length(deg$vector),levels=1:length(deg$vector)),page_rank=deg$vector[o
ggplot(deg)+aes(y=page_rank,x=index)+geom_col(width=1)+
  theme(axis.title.x = element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```



The page rank of the nodes of the graph presents a structure in which there are several dominating nodes that have a much higher rank than the others.

Clustering Coefficient

```
transitivity(g)
```

```
## [1] 0.7758672
```

The clustering coefficient of the whole graph is 0.77, we will compare it to those of the groups in the next sections.

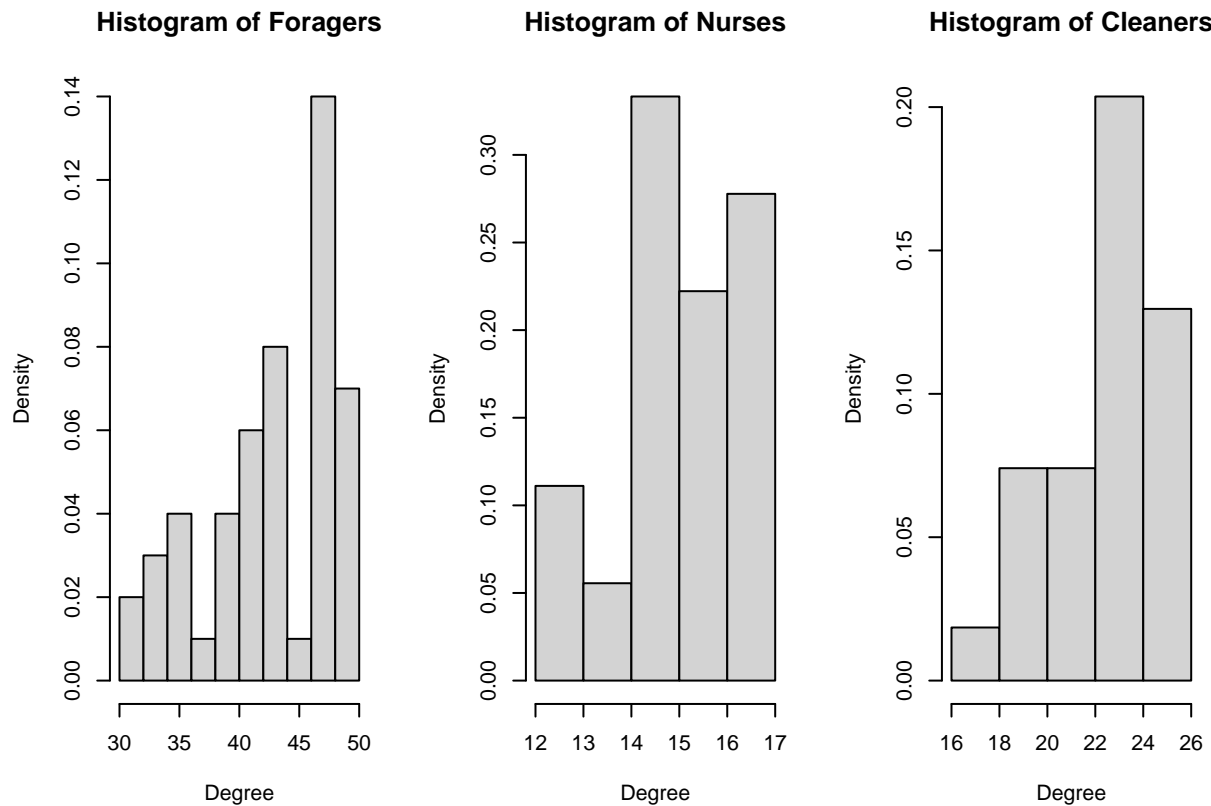
By group

Before looking at the following summary statistics per group it is important to know about the size of each group. There are 50 ants in the group foragers, 27 cleaners and 18 are nurses. This means the following results need to be considered under the fact of this unequal distribution of group members.

Histogram and Density

The foragers have a rather wider spread of degrees compared to the other two groups. Most of them have a degree greater than 40. The other two groups are distributed very homogeneously. It is notable though

that there is a hierarchy of degrees starting from the low degree group of nurses to the midrange group of cleaners over to the higher degree group of foragers.



```
par(mfrow=c(1,3))
max(degree(groups[[1]]))
```

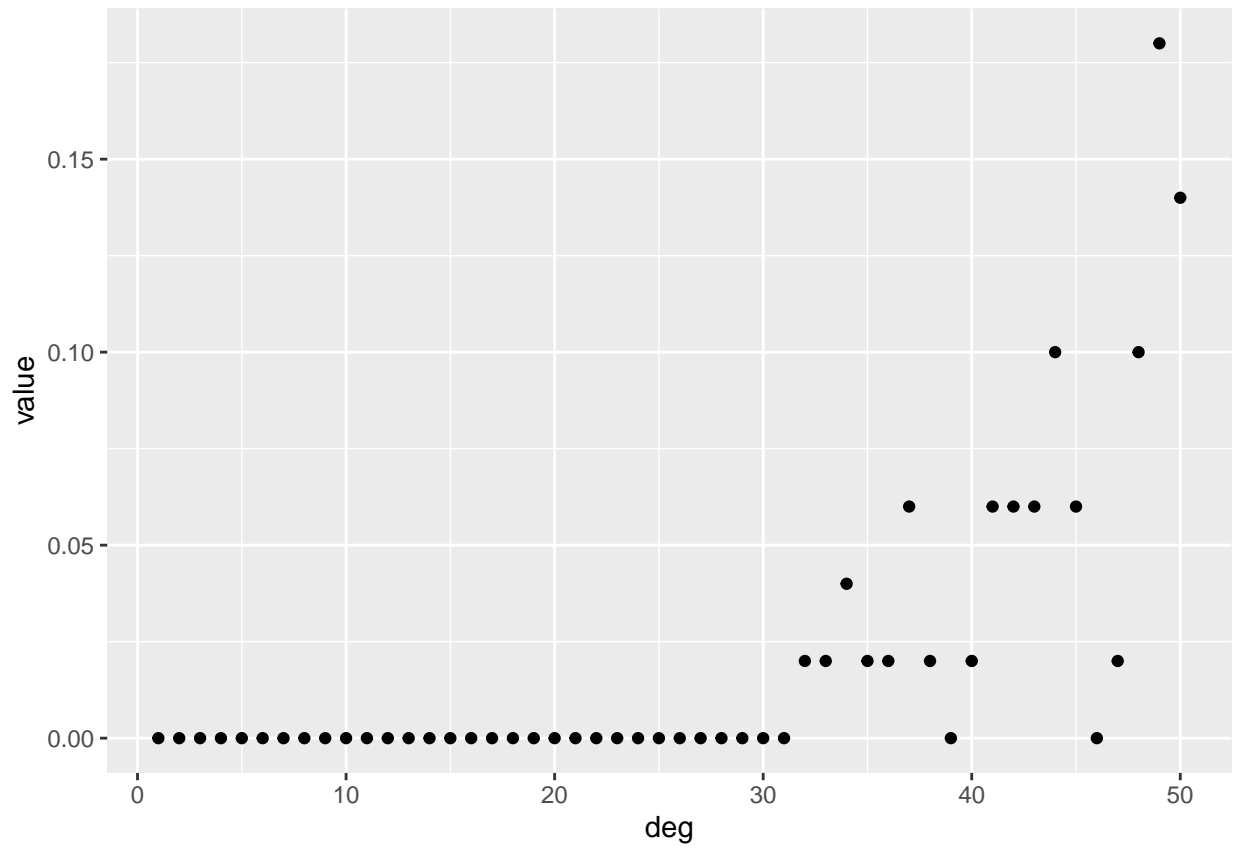
```
## [1] 49
```

```
length(degree_distribution(groups[[1]]))
```

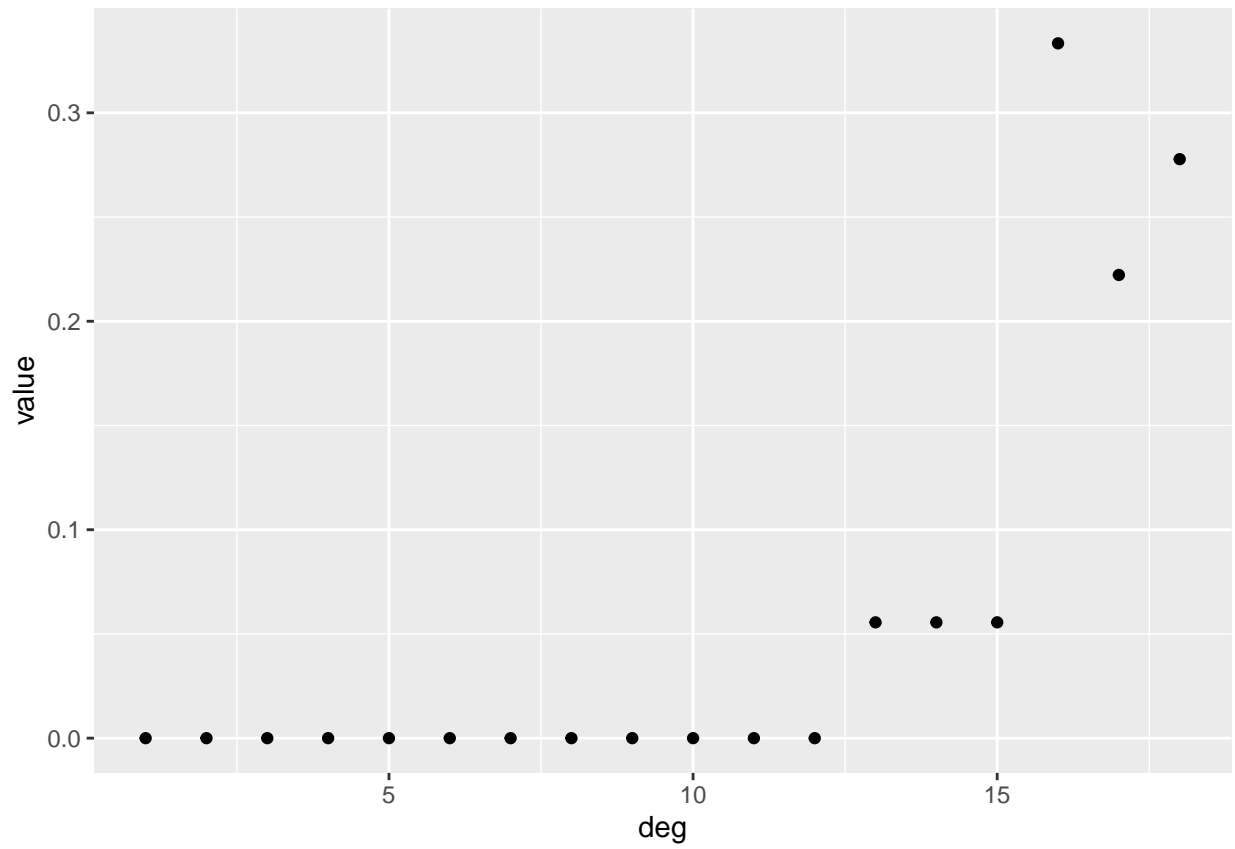
```
## [1] 50
```

```
tmp <- lapply(groups, function(u){
  deg_dist <- data.frame(value = degree_distribution(u), deg=1:(max(degree(u))+1))
  ggplot(deg_dist) + aes(x=deg, y=value) + geom_point()
})
tmp
```

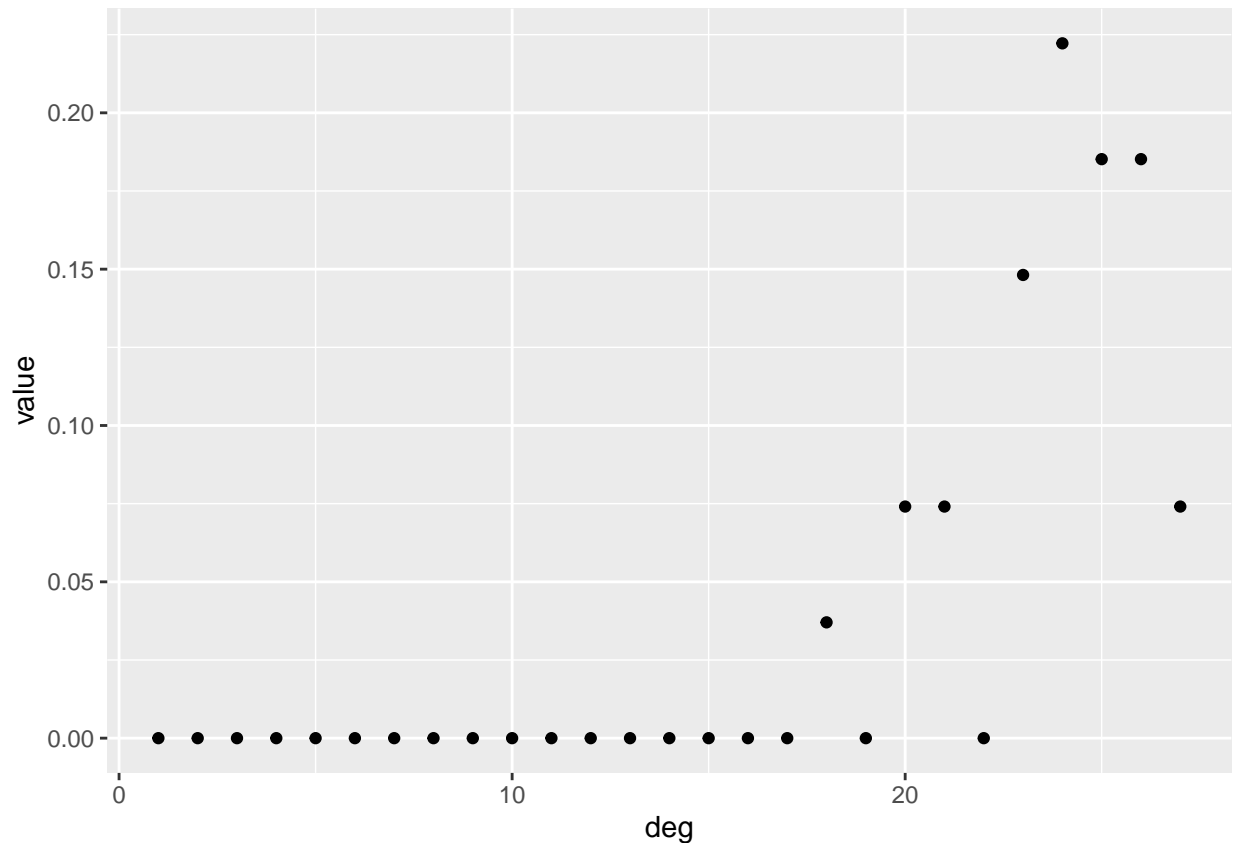
```
## $Foragers
```



\$Nurses



```
##  
## $Cleaners
```



Connectivity and Components

All separate groups are connected and thus, each of them form a big component.

```
lapply(groups, is_connected)
```

```
## $Foragers
## [1] TRUE
##
## $Nurses
## [1] TRUE
##
## $Cleaners
## [1] TRUE
```

Average Path Length and Diameter

The average path length can be an indicator of how connected a network is. Here the foragers and the cleaners have a similar average path length of around 1.12. In comparison the nurses have an average path length of 1.09, a bit smaller. All three groups are very well connected because most ants can reach any other ant of the same group in less than 2 steps. Moreover the groups have smaller average path lengths than the overall graph which suggests that the proposed grouping does indeed lead to more coherent groups. The diameter is the longest shortest path in a network and differs quite a bit in this network. The foragers have a diameter of 11, compared to the nurses with 7 and the cleaners with 4. This might be an indicator of a

lower connection level in the group of the foragers. There exists at least one ant who has to go over 11 other ants in order to connect with another specially chosen ant. This is quite extreme considering the average path length is 1.12 in this group.

```
lapply(groups, average.path.length)
```

```
## $Foragers
## [1] 1.120816
##
## $Nurses
## [1] 1.091503
##
## $Cleaners
## [1] 1.119658
```

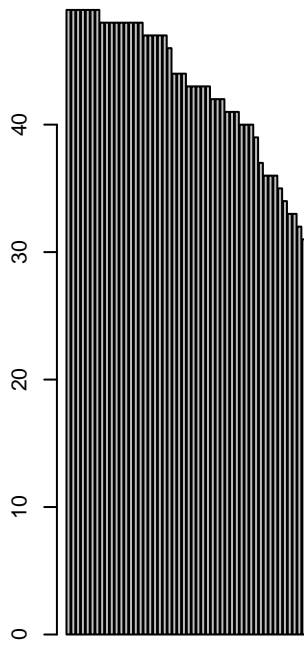
```
lapply(groups, diameter)
```

```
## $Foragers
## [1] 11
##
## $Nurses
## [1] 7
##
## $Cleaners
## [1] 4
```

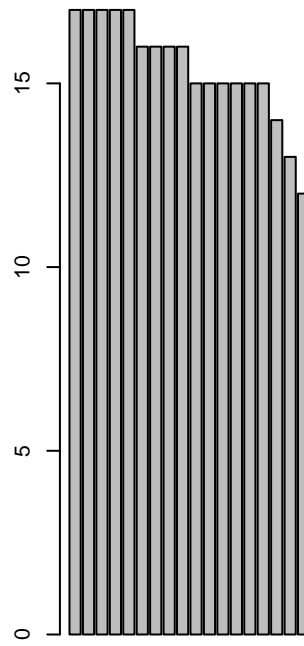
Centrality Measures

We have considered many different centrality measures. Most of them have shown a rather homogeneous distribution, meaning there was no distinction possible between central nodes and less central nodes. For each group the following centrality measures were decreasing smoothly without any particular exceptional occurrences: degree centrality, closeness centrality and eigenvalue centrality. While this can indicate that in general there are no real distinct central nodes in the network we can still try to find some special ants w.r.t. other centrality measures.

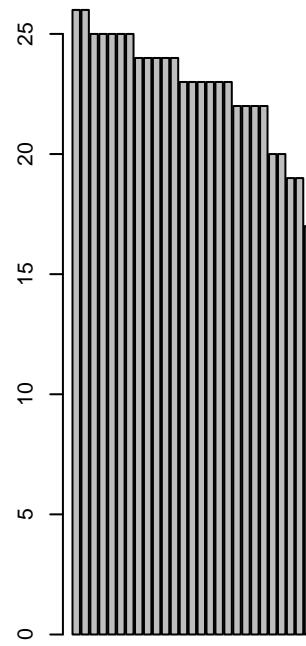
Degree centrality Foragers



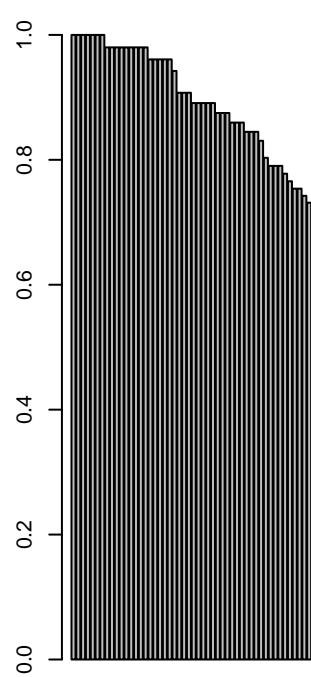
Degree centrality Nurses



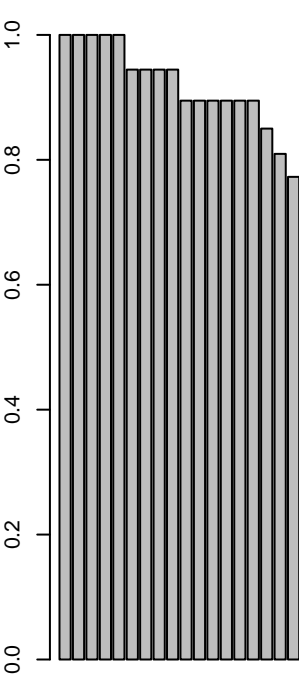
Degree centrality Cleaners



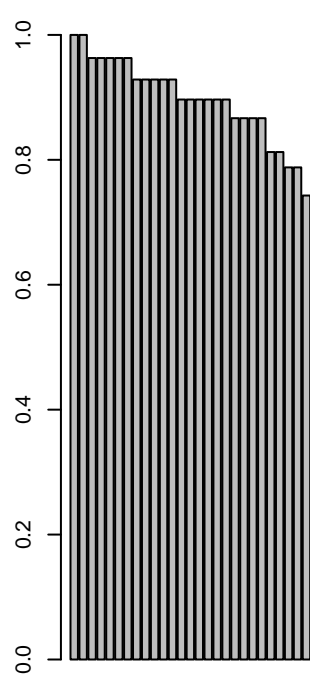
Closeness centrality Foragers

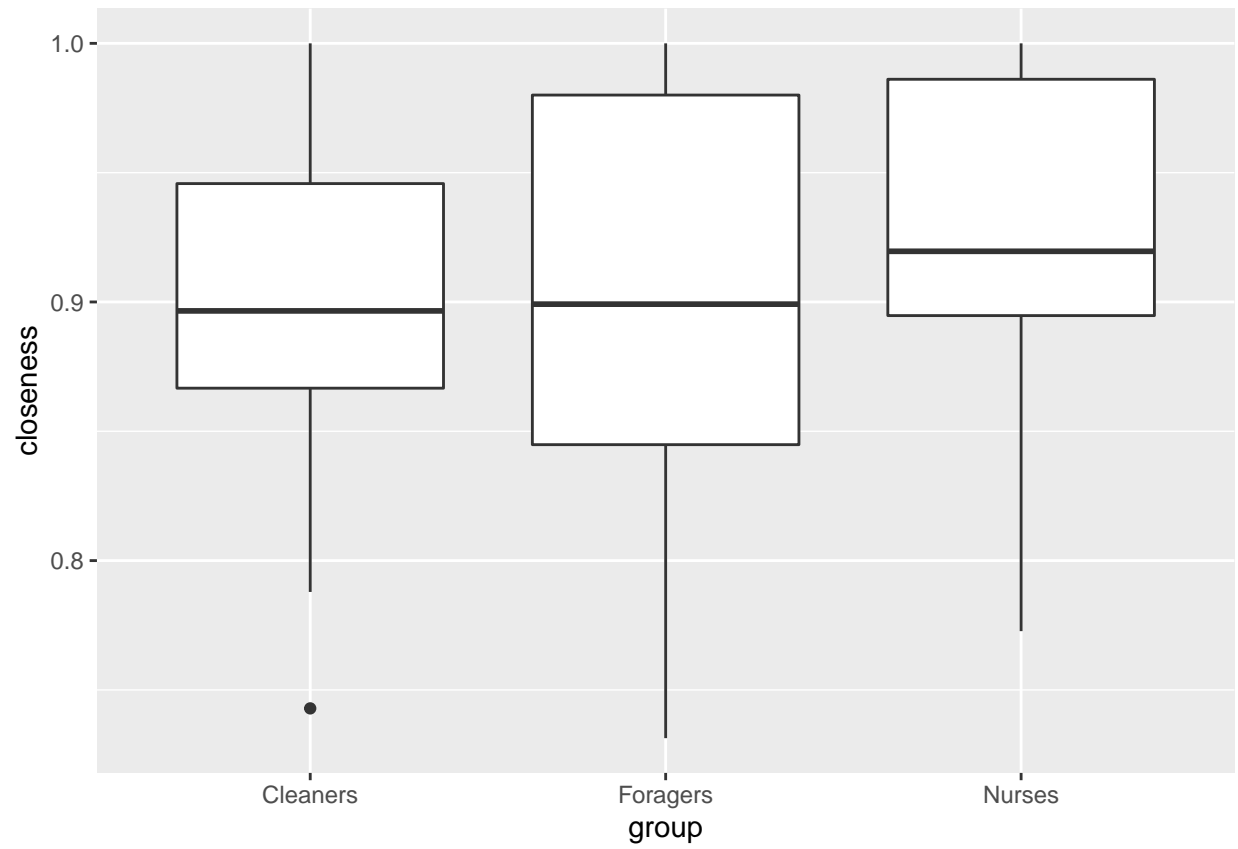


Closeness centrality Nurses

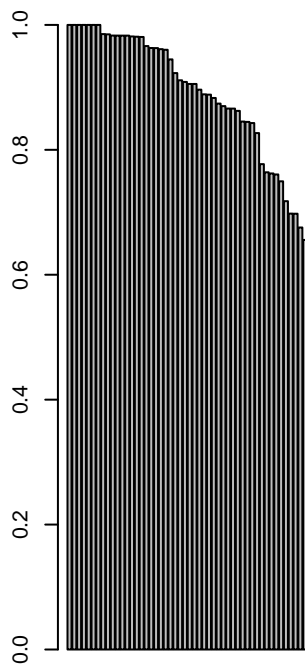


Closeness centrality Cleaners

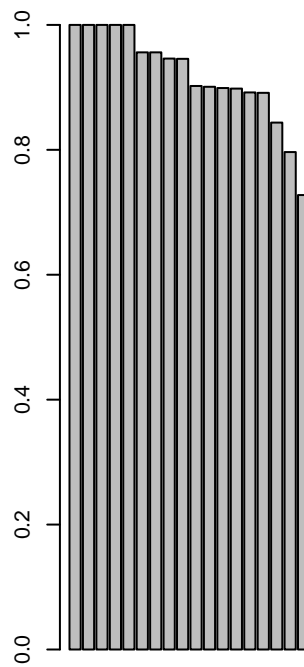




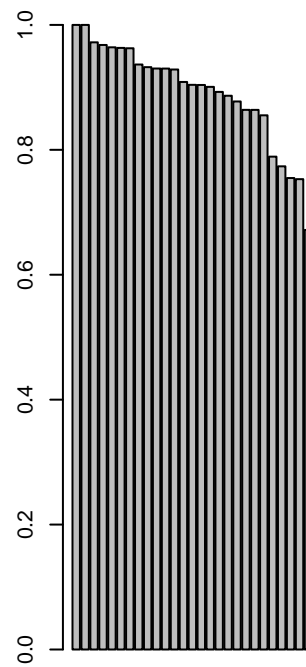
Eigenvalue centrality Forager:



Eigenvalue centrality Nurses

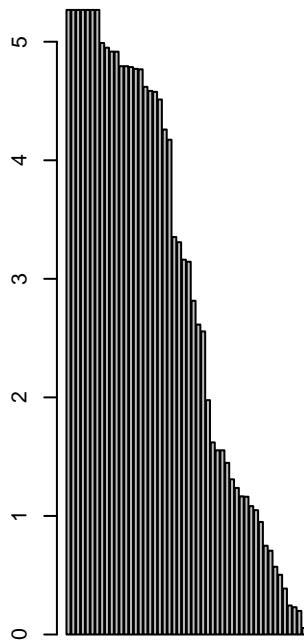
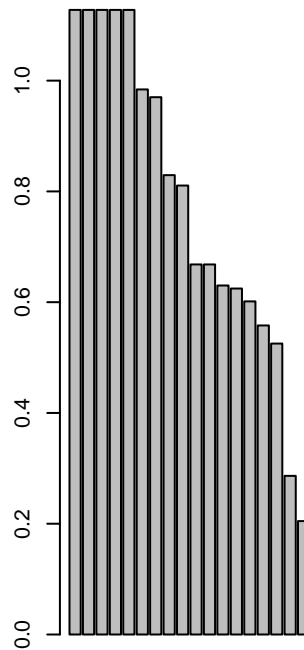
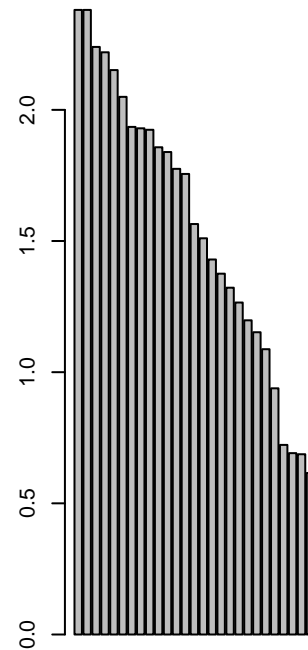


Eigenvalue centrality Cleaner:



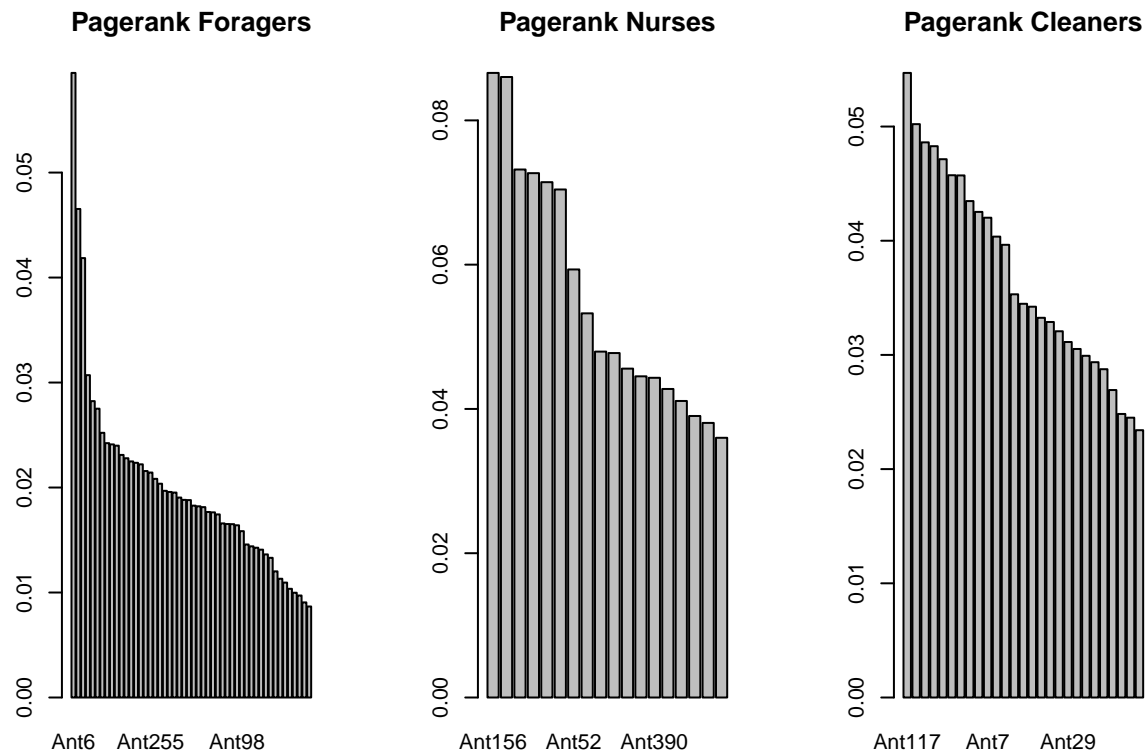
Betweenness centrality: Here we see actually a decreasing structure in all three groups. Some ants lie on the paths connecting other ants. And some ants are rather unimportant in the network.

```
par(mfrow=c(1,3))
betw <- lapply(groups, centralization.betweenness)
betw <- lapply(betw, function(u) u$res[order(u$res, decreasing = TRUE)])
tmp <- lapply(names(betw), function(u) barplot(betw[[u]], main = paste("Betweenness centrality", u)))
```

Betweenness centrality Forage**Betweenness centrality Nurse****Betweenness centrality Cleanse**

Pagerank: Here we can see that the foragers have a few ants who are very important w.r.t to this centrality measure. This could indicate that these ants are consulted by many other ants who themselves are consulted by many ants.

```
par(mfrow=c(1,3))
pagr <- lapply(groups, page.rank)
pagr <- lapply(pagr, function(u) u$vector[order(u$vector, decreasing = TRUE)])
tmp <- lapply(names(pagr), function(u) barplot(pagr[[u]], main = paste("Pagerank", u)))
```



Clustering Coefficient

Very similar results again for the clustering coefficient. All groups are very highly connected.

```
lapply(groups, transitivity)
```

```
## $Foragers
## [1] 0.9064523
##
## $Nurses
## [1] 0.9125926
##
## $Cleaners
## [1] 0.889019
```

K-means check

Preparing the data

```
atributes<-vertex.attributes(g)
str(atributes)
```

```
## List of 19
## $ nb_interaction_queen : num [1:99] NaN 8.636 0 1.091 0.0909 ...
## $ nb_interaction_foragers: num [1:99] NaN 68.4 681.3 86.5 88.2 ...
## $ nb_interaction_cleaners: num [1:99] NaN 169.4 65.4 126.8 143.7 ...
## $ nb_interaction_nurses : num [1:99] NaN 161.64 9.09 97.27 37.64 ...
## $ visits_to_rubbishpile : num [1:99] NaN 4.6364 0.2727 0.0909 50.2727 ...
## $ visits_to_nest_entrance: num [1:99] NaN 0.909 166 0 0.818 ...
## $ visits_to_brood : num [1:99] NaN 137.909 0.636 85.909 30.909 ...
## $ group_period4 : chr [1:99] "" " " "F" "N" ...
## $ group_period3 : chr [1:99] "" " " "F" "N" ...
## $ group_period2 : chr [1:99] "" "C" "F" "C" ...
## $ group_period1 : chr [1:99] "" "N" "F" "C" ...
## $ nb_foraging_events : num [1:99] NaN 0 15 0 0 0 0 0 43 ...
## $ age(days) : num [1:99] NaN 57 246 372 71 190 351 71 57 344 ...
## $ body_size : num [1:99] NaN 141 215 149 124 ...
## $ tag_id : num [1:99] NaN 620 356 540 115 117 113 191 190 217 ...
## $ colony : num [1:99] NaN 4 4 4 4 4 4 4 4 4 ...
## $ id : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
## $ group : chr [1:99] "" "C" "F" "C" ...
## $ name : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
```

```
to_cluster<-data.frame(
  id=attributes$name,
  group=attributes$group,
  i_queen=attributes$nb_interaction_queen,
  i_forag=attributes$nb_interaction_foragers,
  i_clean=attributes$nb_interaction_cleaners,
  i_nurse=attributes$nb_interaction_nurses,
  v_rubbish=attributes$visits_to_rubbishpile,
  v_entr=attributes$visits_to_nest_entrance,
  v_brood=attributes$visits_to_brood,
  foraging=attributes$nb_foraging_events,
  age=attributes$`age(days)`,
  size=attributes$body_size
)
to_cluster<-na.omit(to_cluster)
str(to_cluster)
```

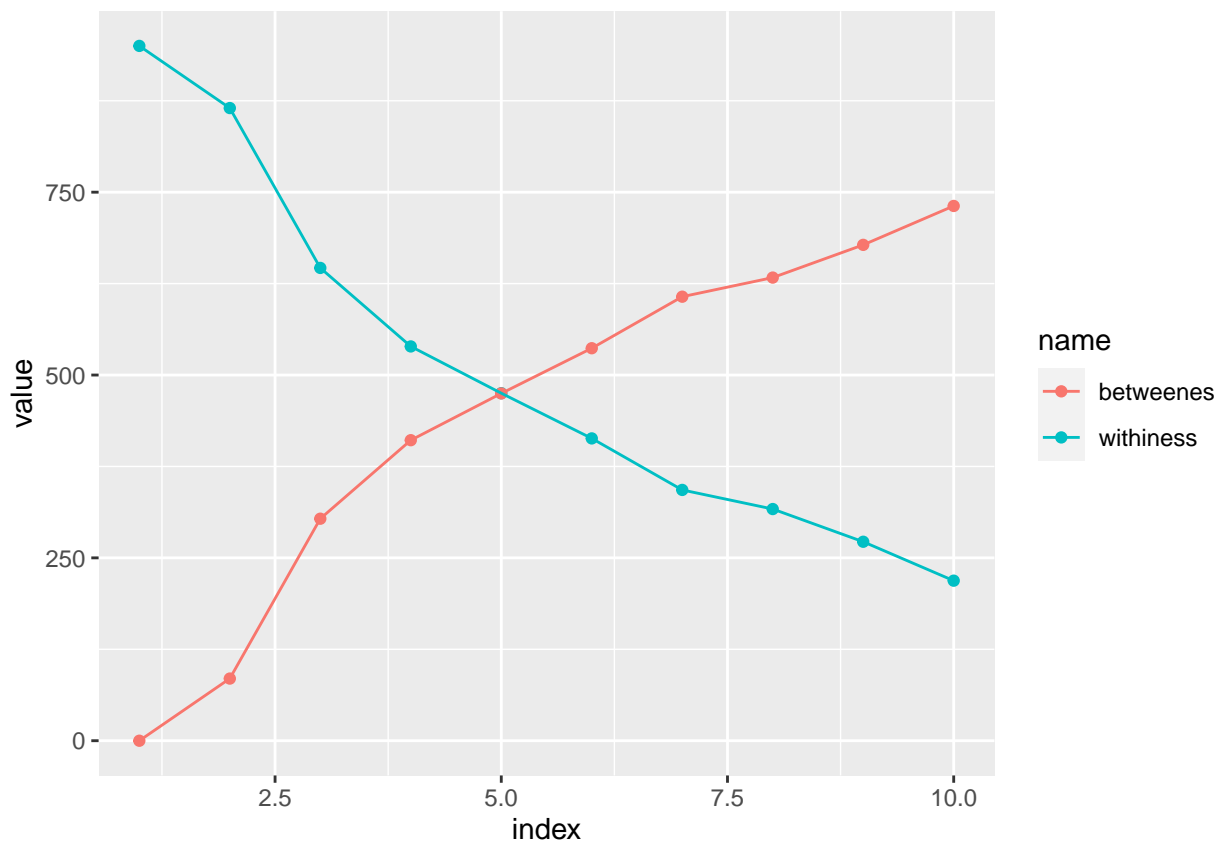
```
## 'data.frame': 96 obs. of 12 variables:
## $ id : chr "Ant620" "Ant356" "Ant540" "Ant115" ...
## $ group : chr "C" "F" "C" "F" ...
## $ i_queen : num 8.636 0 1.091 0.0909 6.727 ...
## $ i_forag : num 68.4 681.3 86.5 88.2 179 ...
## $ i_clean : num 169.4 65.4 126.8 143.7 198.2 ...
## $ i_nurse : num 161.64 9.09 97.27 37.64 128.09 ...
## $ v_rubbish: num 4.6364 0.2727 0.0909 50.2727 3.4545 ...
## $ v_entr : num 0.909 166 0 0.818 0.273 ...
## $ v_brood : num 137.909 0.636 85.909 30.909 88.454 ...
## $ foraging : num 0 15 0 0 0 0 0 0 43 32 ...
## $ age : num 57 246 372 71 190 351 71 57 344 330 ...
## $ size : num 141 215 149 124 160 ...
## - attr(*, "na.action")= 'omit' Named int [1:3] 1 62 63
## ..- attr(*, "names")= chr [1:3] "1" "62" "63"
```



```
nrow(to_cluster)
```

```
## [1] 96
```

```
to_cluster2<-scale(to_cluster[,3:ncol(to_cluster)])
models<-lapply(1:10,function(i){kmeans(to_cluster2,i)})
sum_within<-do.call(rbind,lapply(1:length(models),function(i){models[[i]]$tot.withinss}))
sum_within<-data.frame(index=1:nrow(sum_within),name=rep("withinss",nrow(sum_within)),value=sum_within)
sum_between<-do.call(rbind,lapply(1:length(models),function(i){models[[i]]$betweenss}))
sum_between<-data.frame(index=1:nrow(sum_between),name=rep("betweenss",nrow(sum_between)),value=sum_between)
#sum_total<-data.frame(index=1:nrow(sum_between),
#                        name=rep("total error",nrow(sum_between)),
#                        value=sum_between$value+sum_within$value)
to_plot<-rbind(sum_within,sum_between)
ggplot(to_plot)+aes(x=index,y=value,colour=name)+geom_point()+geom_line()
```



```
model3<-models[[3]]
to_cluster3<-data.frame(to_cluster,found_labels=model3$cluster)
table(to_cluster3$group,to_cluster3$found_labels)
```

```
##
##      1  2  3
## C 26  0  1
## F 12  6 32
```

```
##    N 18  0  0
##    Q  0  1  0
```

From the results above we may infer that the foragers group may have been misclassified and that there are a lot of foragers that present similar behaviour to that of the cleaners.

```
model4<-models[[4]]
to_cluster4<-data.frame(to_cluster,found_labels=model4$cluster)
table(to_cluster4$group,to_cluster4$found_labels)
```

```
##
##      1  2  3  4
##    C  0  9 18  0
##    F 29  0 21  0
##    N  0 14  2  2
##    Q  0  1  0  0
```

Atempt 2

```
atributes<-vertex.attributes(g)
str(atributes)
```

```
## List of 19
## $ nb_interaction_queen : num [1:99] NaN 8.636 0 1.091 0.0909 ...
## $ nb_interaction_foragers: num [1:99] NaN 68.4 681.3 86.5 88.2 ...
## $ nb_interaction_cleaners: num [1:99] NaN 169.4 65.4 126.8 143.7 ...
## $ nb_interaction_nurses : num [1:99] NaN 161.64 9.09 97.27 37.64 ...
## $ visits_to_rubbishpile : num [1:99] NaN 4.6364 0.2727 0.0909 50.2727 ...
## $ visits_to_nest_entrance: num [1:99] NaN 0.909 166 0 0.818 ...
## $ visits_to_brood : num [1:99] NaN 137.909 0.636 85.909 30.909 ...
## $ group_period4 : chr [1:99] "" " " " "F" "N" ...
## $ group_period3 : chr [1:99] "" " " " "F" "N" ...
## $ group_period2 : chr [1:99] "" "C" "F" "C" ...
## $ group_period1 : chr [1:99] "" "N" "F" "C" ...
## $ nb_foraging_events : num [1:99] NaN 0 15 0 0 0 0 0 43 ...
## $ age(days) : num [1:99] NaN 57 246 372 71 190 351 71 57 344 ...
## $ body_size : num [1:99] NaN 141 215 149 124 ...
## $ tag_id : num [1:99] NaN 620 356 540 115 117 113 191 190 217 ...
## $ colony : num [1:99] NaN 4 4 4 4 4 4 4 4 ...
## $ id : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
## $ group : chr [1:99] "" "C" "F" "C" ...
## $ name : chr [1:99] "Ant621" "Ant620" "Ant356" "Ant540" ...
```

```
to_cluster<-data.frame(
  id=atributes$name,
  group=atributes$group,
  i_queen=atributes$nb_interaction_queen,
  i_forag=atributes$nb_interaction_foragers,
  i_clean=atributes$nb_interaction_cleaners,
  i_nurse=atributes$nb_interaction_nurses,
  v_rubbish=atributes$visits_to_rubbishpile,
```

```

v_entr=attributes$visits_to_nest_entrance,
v_brood=attributes$visits_to_brood,
foraging=attributes$nb_foraging_events
)
to_cluster<-na.omit(to_cluster)
str(to_cluster)

```

```

## 'data.frame': 96 obs. of 10 variables:
## $ id : chr "Ant620" "Ant356" "Ant540" "Ant115" ...
## $ group : chr "C" "F" "C" "F" ...
## $ i_queen : num 8.636 0 1.091 0.0909 6.727 ...
## $ i_forag : num 68.4 681.3 86.5 88.2 179 ...
## $ i_clean : num 169.4 65.4 126.8 143.7 198.2 ...
## $ i_nurse : num 161.64 9.09 97.27 37.64 128.09 ...
## $ v_rubbish: num 4.6364 0.2727 0.0909 50.2727 3.4545 ...
## $ v_entr : num 0.909 166 0 0.818 0.273 ...
## $ v_brood : num 137.909 0.636 85.909 30.909 88.454 ...
## $ foraging : num 0 15 0 0 0 0 0 0 43 32 ...
## - attr(*, "na.action")= 'omit' Named int [1:3] 1 62 63
## ..- attr(*, "names")= chr [1:3] "1" "62" "63"

```

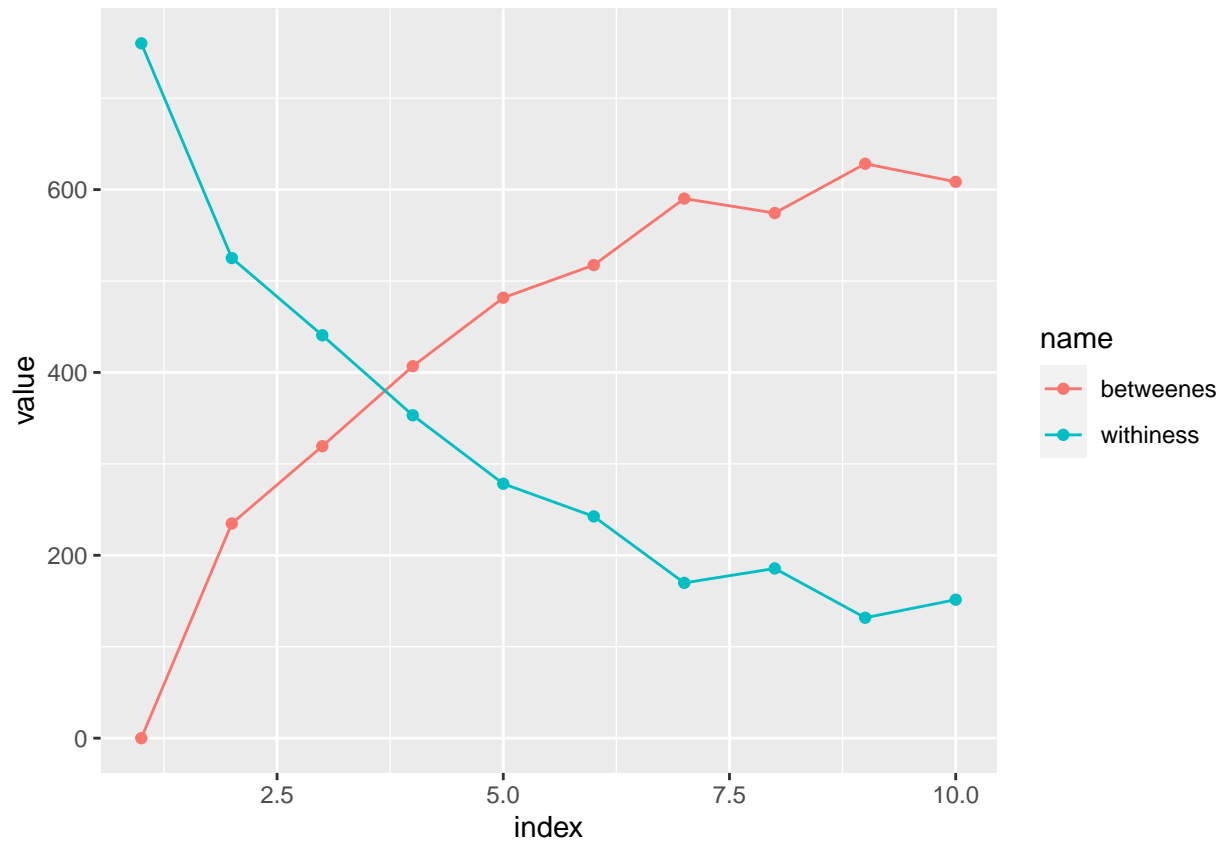
```
nrow(to_cluster)
```

```
## [1] 96
```

```

to_cluster2<-scale(to_cluster[,3:ncol(to_cluster)])
models<-lapply(1:10,function(i){kmeans(to_cluster2,i)})
sum_within<-do.call(rbind,lapply(1:length(models),function(i){models[[i]]$tot.withinss}))
sum_within<-data.frame(index=1:nrow(sum_within),name=rep("withinness",nrow(sum_within)),value=sum_within)
sum_between<-do.call(rbind,lapply(1:length(models),function(i){models[[i]]$betweenss}))
sum_between<-data.frame(index=1:nrow(sum_between),name=rep("betweenness",nrow(sum_between)),value=sum_between)
#sum_total<-data.frame(index=1:nrow(sum_between),
#
#               name=rep("total error",nrow(sum_between)),
#
#               value=sum_between$value+sum_within$value)
to_plot<-rbind(sum_within,sum_between)
ggplot(to_plot)+aes(x=index,y=value,colour=name)+geom_point()+geom_line()

```



```
model3<-models[[3]]
to_cluster3<-data.frame(to_cluster,found_labels=model3$cluster)
table(to_cluster3$group,to_cluster3$found_labels)
```

```
##
##      1  2  3
## C   0  0 27
## F   0 38 12
## N   2  0 16
## Q   0  0  1
```

Stil the same problem