Network Analysis Assignment 5

Camilla Casamento Tumeo, Guirong Fu, Melanie Gaillochet, Ritu Sriram

2 Dec 2018

Prepare the study object

```
rm(list = ls())
setwd("E:/courses/Network analysis/lab/lab5")
library(igraph)
library(ggplot2)
library(ggraph)
library(ggthemes) # To plot figures
library(reshape2)
library(Rmisc) # To plot figures
library(grid) # To plot figures
set.seed(1996)

# Import data
countries <- read.csv("./data/countries.csv", header = T, sep=",", stringsAsFactors = F)
exports <- data.matrix(read.csv("./data/exports.csv", row.names = 1,
                                header = T, sep=",", stringsAsFactors = F))

# Rename and dealing with NA.
colnames(exports) <- countries$iso
rownames(exports) <- countries$iso
exports[is.na(exports)] <- 0
```

```
# Build the graph and set vertex attributes.
trade_graph <- graph.adjacency(exports,mode = 'directed',weighted = T)
trade_graph <-set_vertex_attr(trade_graph,  "iso", index=countries$X, value=countries$iso)
trade_graph <-set_vertex_attr(trade_graph,  "gdp", index=countries$X, value=countries$gdp)
trade_graph <- set_vertex_attr(trade_graph, "name", index=countries$X,
                               value=countries$country)
summary(trade_graph)
```

```
## IGRAPH d6417d0 DNW- 174 18416 --
## + attr: name (v/c), iso (v/c), gdp (v/n), weight (e/n)
```

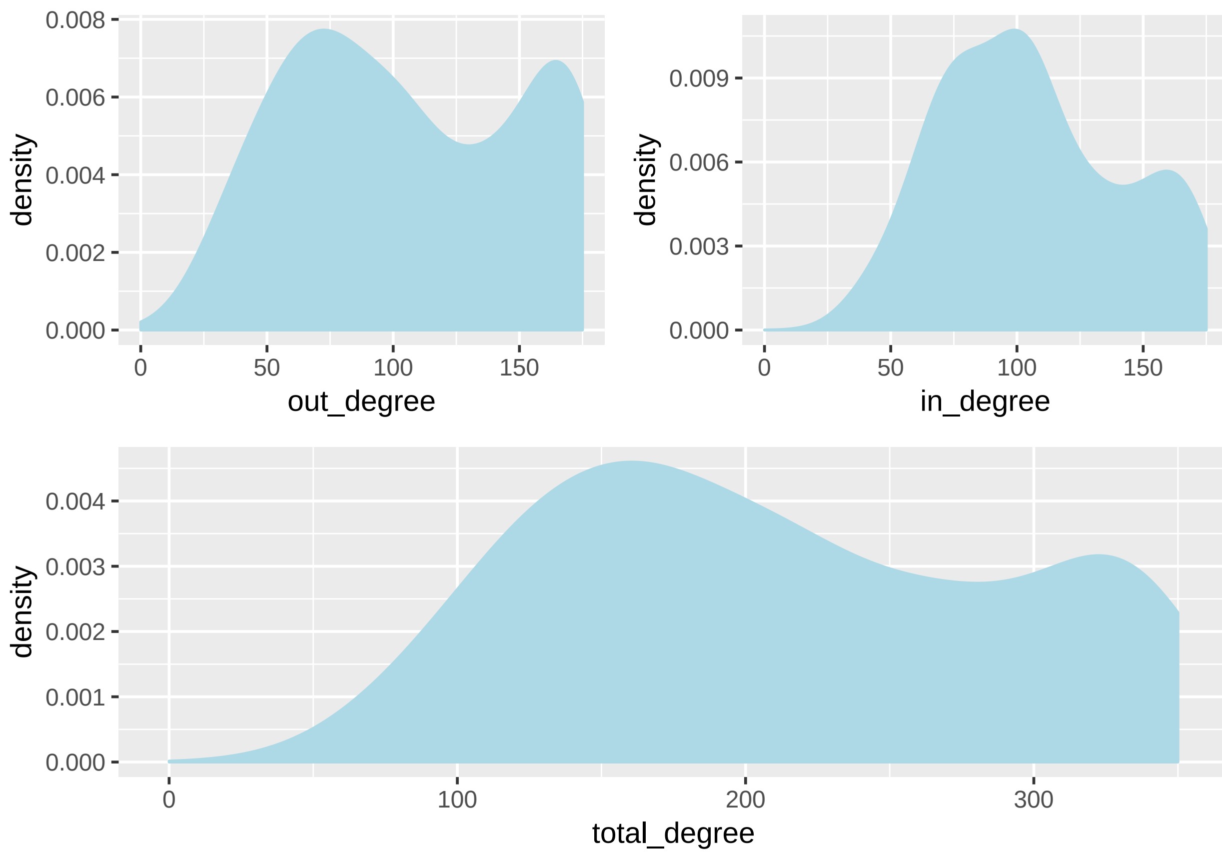
This is a directed graph with 174 nodes and 18416 edges. The density of the network is 0.6117866.

Individual features.

Centralities

```
# degree centrialities
out_degree <- degree(trade_graph, mode = 'out')
in_degree <- degree(trade_graph, mode = 'in')
total_degree <- degree(trade_graph, mode = 'total')

trade_graph <-set_vertex_attr(trade_graph,  "OutDeg", index=countries$X, value=out_degree)
```

```r
trade_graph <-set_vertex_attr(trade_graph,  "InDeg", index=countries$X, value=in_degree)

out_degree_g<- ggplot(data.frame(out_degree), aes(x=out_degree)) +
  xlim(0,175)+
  geom_density(fill = "lightblue", colour = "lightblue")
in_degree_g<- ggplot(data.frame(in_degree), aes(x=in_degree)) +
  xlim(0,175)+
  geom_density(fill = "lightblue", colour = "lightblue")
tot_degree_g<- ggplot(data.frame(total_degree), aes(x=total_degree)) +
  xlim(0,350)+
  geom_density(fill = "lightblue", colour = "lightblue")

grid.newpage()
pushViewport(viewport(layout = grid.layout(2, 2)))
vplayout = function(x, y) viewport(layout.pos.row = x, layout.pos.col = y)
print(out_degree_g, vp = vplayout(1, 1))
print(in_degree_g, vp = vplayout(1, 2))
print(tot_degree_g, vp = vplayout(2, 1:2))
```

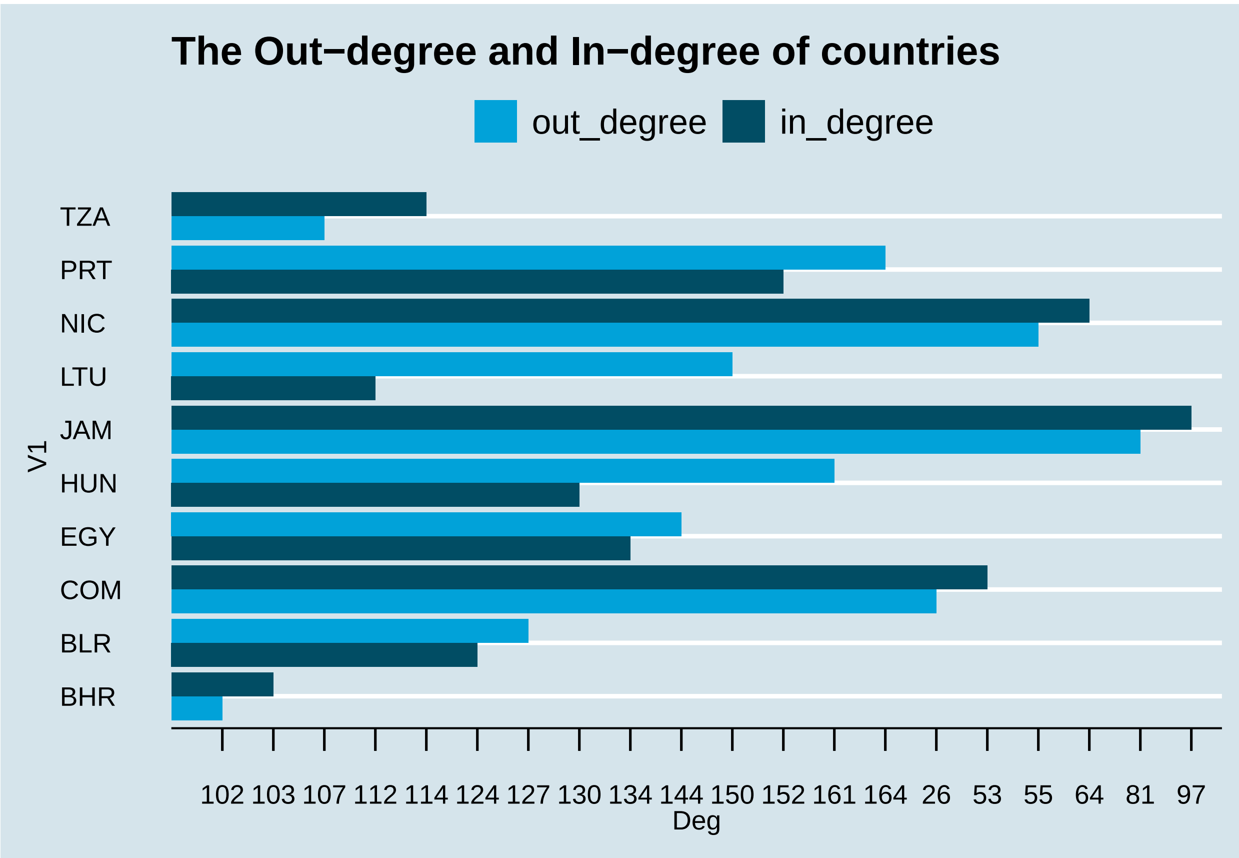


From the density of degree distribution, we see that the out-degree distributions are much uniform than in-degree and total-degree. In in-degree distribution, more countries' import from around the average number of countries. There isn't an obvious tendency in total-degree distribution. The interactions of a country with other countries change a lot among different countries.

```r
deg_data=data.frame(t(rbind(countries$iso,out_degree, in_degree)))
deg_data<-melt(deg_data, id.vars = "V1", variable.name = "Type", value.name = "Deg")
```

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

```
samples<-sample(1:174,10)
samples<-c(samples,174+samples)

ggplot(deg_data[samples,], aes(V1,Deg,fill=Type))+
  geom_bar(stat="identity",position="dodge")+
  ggtitle("The Out-degree and In-degree of countries")+
  theme_economist(base_size=10)+
  scale_fill_economist()+
  theme(axis.ticks.length=unit(0.3,'cm'))+
  guides(fill=guide_legend(title=NULL))+
  coord_flip()
```

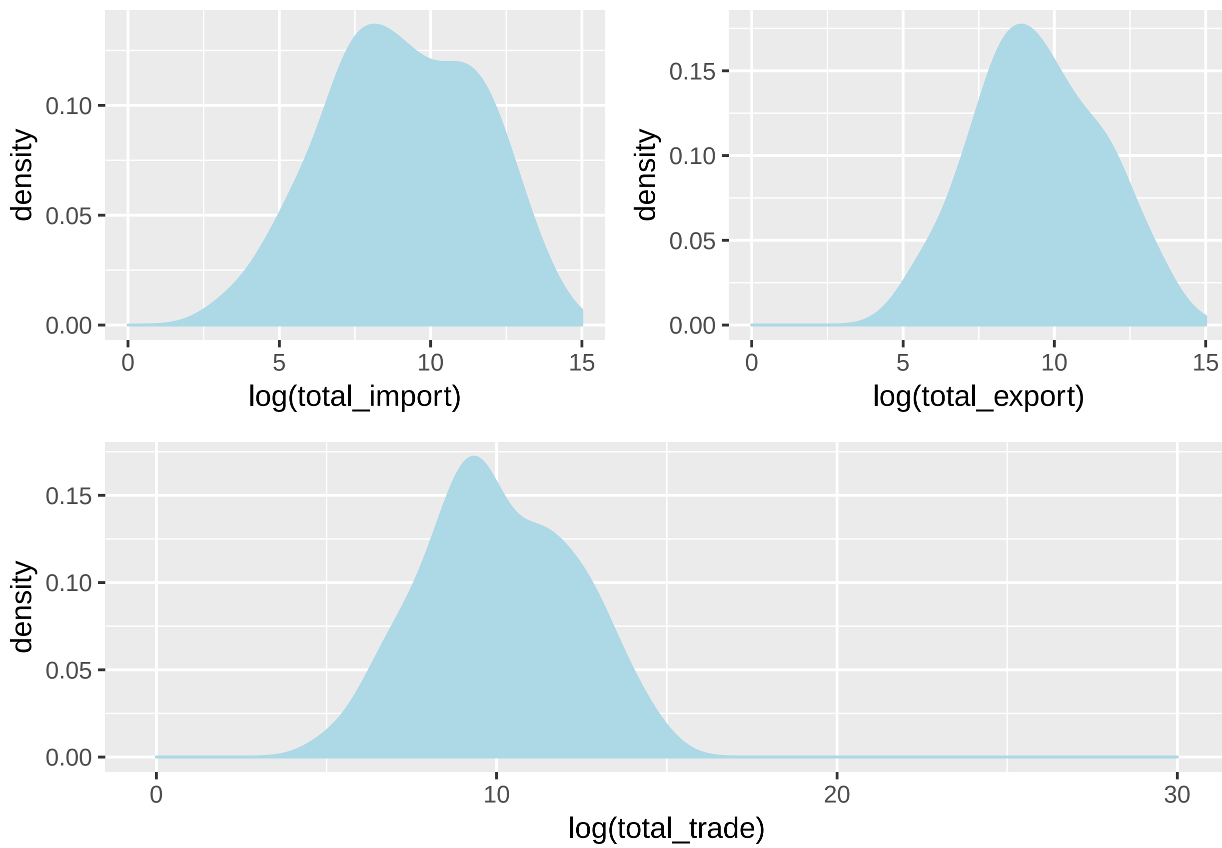


We randomly select 10 countries to see their out-degrees and in-degrees. From this figure, we see these two degrees are almost at the same level for a certain country, (e.g. the numbers are similar) with a little difference. In another word, if a country has a large out degree, its in degree is also very likely to be large.

Then, we try to explore on the nodes' strength. That is, the weighted node degrees. Here we define three kinds of node's strength by import, export and total trade, $s_{im}(i) = \Sigma_{j \in N} w_{ij}$, $s_{ex}(i) = \Sigma_{j \in N} w_{ji}$ and $s_{tot}(i) = \Sigma_{j \in N}(w_{ij} + w_{ji})$. w_{ij} is the item's value in exports.csv file.

```
# weighted node strength
total_export <- c(0, times=174)
total_import <- c(0, times=174)

for(i in c(1:174)){
```

```r
  total_import[i] <- sum(exports[i,])
  total_export[i] <- sum(exports[,i])
}
total_trade<-total_export+total_import

trade_graph <-set_vertex_attr(trade_graph,"Export",index=countries$X,value=total_export)
trade_graph <-set_vertex_attr(trade_graph,"Import",index=countries$X,value=total_import)
trade_graph <-set_vertex_attr(trade_graph,"TotTrade",index=countries$X,value=total_trade)

# Add new attribute, splitting countries upon their amount of import and export
countries$type[which(total_export >= total_import)] <- 1
countries$type[which(total_export < total_import)] <- 2
trade_graph <-set_vertex_attr(trade_graph,"TradeType",
                              index=countries$X,value=countries$type)

import_g <- ggplot(data.frame(log(total_import)),aes(x=log(total_import))) +
  xlim(0,15)+
  geom_density(fill = "lightblue", colour = "lightblue")
export_g <- ggplot(data.frame(log(total_export)),aes(x=log(total_export))) +
  xlim(0,15)+
  geom_density(fill = "lightblue", colour = "lightblue")
total_trade_g <- ggplot(data.frame(log(total_trade)),aes(x=log(total_trade))) +
  xlim(0,30)+
  geom_density(fill = "lightblue", colour = "lightblue")

grid.newpage()
pushViewport(viewport(layout = grid.layout(2, 2)))
print(import_g, vp = vplayout(1, 1))
print(export_g, vp = vplayout(1, 2))
print(total_trade_g, vp = vplayout(2, 1:2))
```



Different trade types We now try to compute the densities of full graph, that of the graph induced by Type1 (export more than or equal to import) vertices and that of the graph induced by Type2 (export less than import) vertices.

```
# density of yeast_graph
igraph::graph.density(trade_graph)
```

```
## [1] 0.6117866
```

```
# density of subgraph induced by type_1
t1_vert <- V(trade_graph)[TradeType==1]
sub_t1 <- induced_subgraph(trade_graph, t1_vert)
graph.density(sub_t1)
```

```
## [1] 0.5439533
```

```
# density of subgraph induced by TradeType2
t2_vert <- V(trade_graph)[TradeType==2]
sub_t2 <- induced_subgraph(trade_graph, t2_vert)
graph.density(sub_t2)
```

```
## [1] 0.7269345
```

The given graph is fairly dense. But we can see that the subgraph of countries having higher exports than imports is less dense as compared to the subgraph of countries having higher imports than exports. (We have already established that there is a positive correlation between the gdp of a country and for it to have higher exports than imports which indicates self-sufficiency to some extent.) This says that countries with higher gdp interact less among themselves and probably more with the lower gdp countries as one would expect.

Relations of Trade with GDP

Hypothesis: If a country's export is greater than its import, it will have a higher gdp.

```
diff <- total_export - total_import
cor.test(diff,countries$gdp)
```

```
##
##  Pearson's product-moment correlation
##
## data:  diff and countries$gdp
## t = 6.9795, df = 172, p-value = 6.159e-11
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3451480 0.5781567
## sample estimates:
##      cor
## 0.469795
```

There is a positive correlation between difference in export and import and gdp of the countries. This shows that if this difference is positive, then the country is economically more forward.

Hypothesis: If a country's out-degree is greater than its in-degree, it will have a higher gdp.

```
diff <- out_degree-in_degree
cor.test(diff,countries$gdp)
```

```
##
##  Pearson's product-moment correlation
##
## data:  diff and countries$gdp
## t = 0.86048, df = 172, p-value = 0.3907
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.08411868  0.21217373
## sample estimates:
##        cor
## 0.06547047
```

This hypothesis is not true. No strong correlation between gdp of a country and number of export relations being greater than number of import relations.

Since the range of import(export) amount is very large, we see its log-density. All of the three (total import, total export and total trade) graphs look like a normal distribution. That means more countries have small export(import, total trade) amout, while few countries have pretty large amount of export(import. total trade).

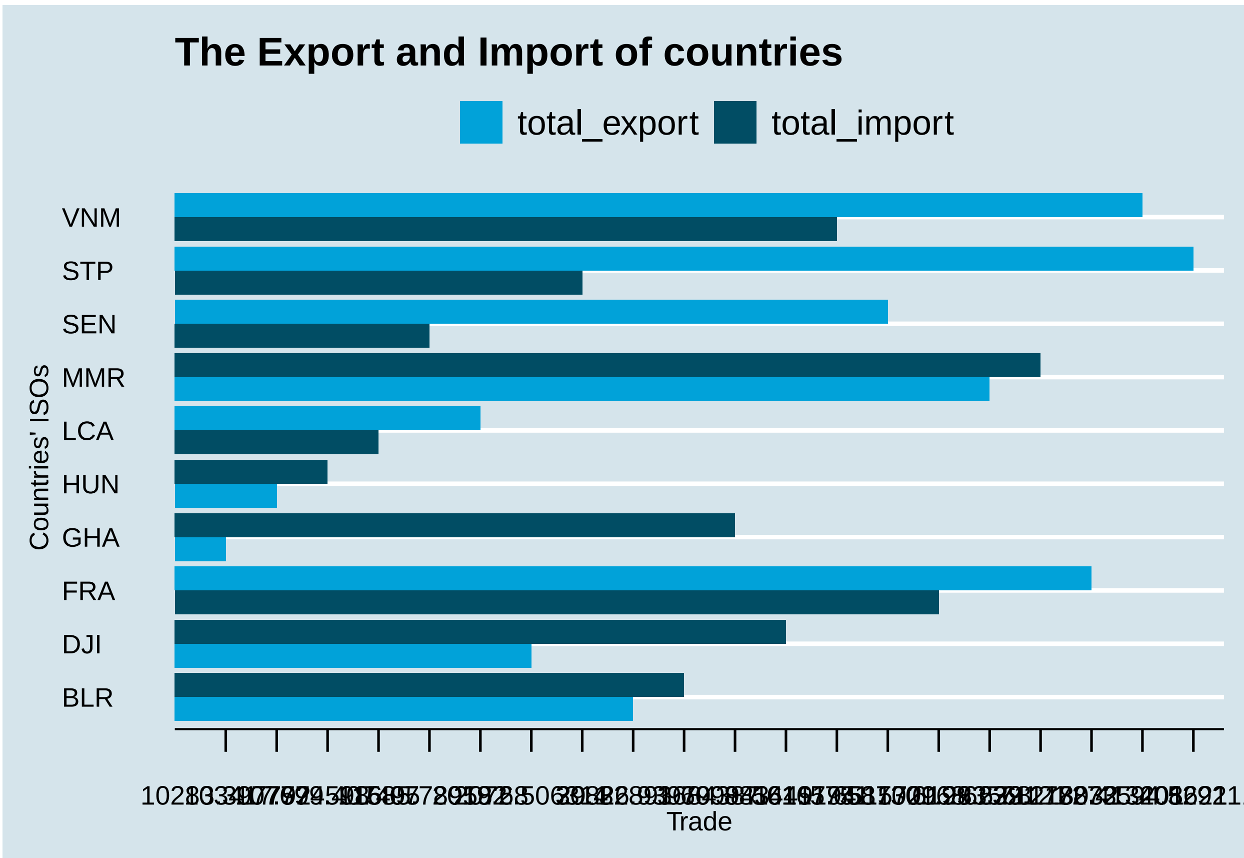
Like what we do in the degree analysis, we also randomly choose 10 samples to analyse their trade amount distribution.

```
trade_data=data.frame(t(rbind(countries$iso,total_export, total_import)))
trade_data<-melt(trade_data, id.vars = "V1", variable.name = "Type", value.name = "Trade")
```

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

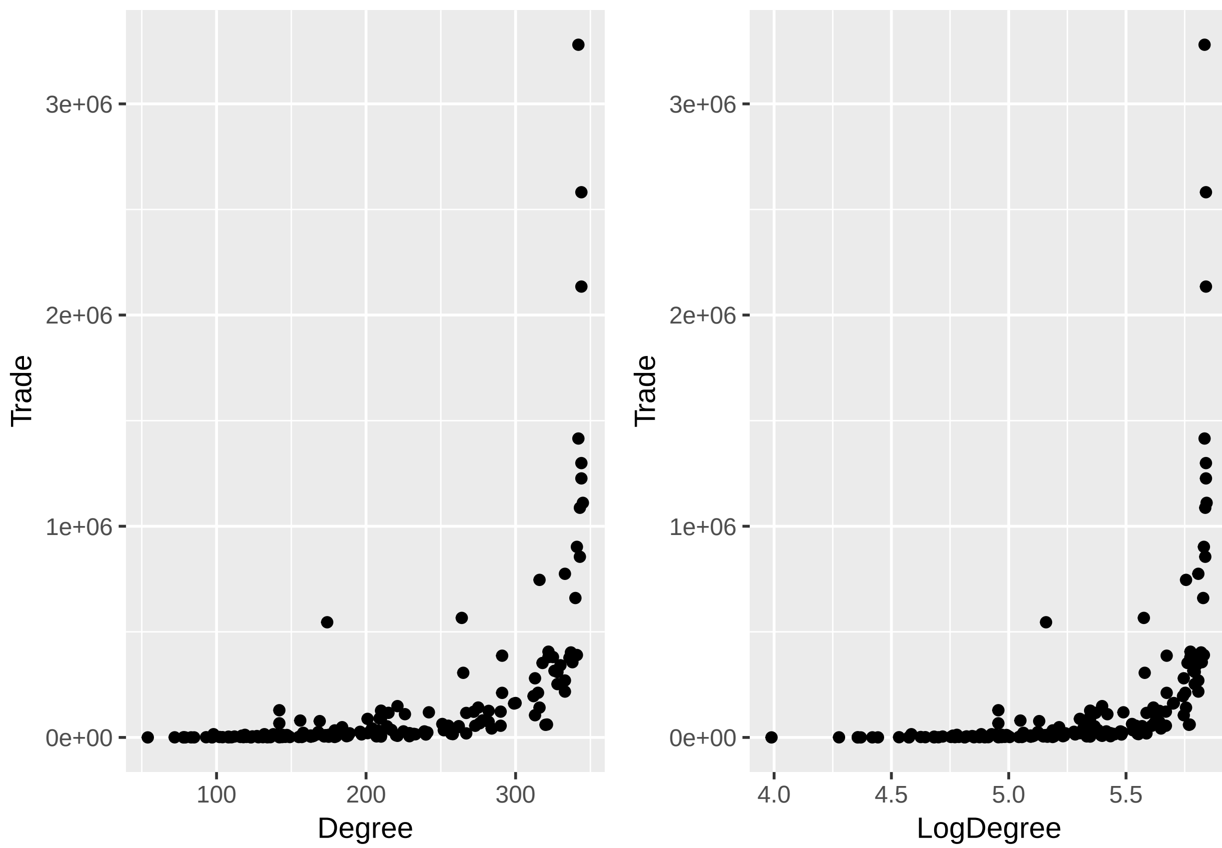
```
samples<-sample(1:174,10)
samples<-c(samples,174+samples)
```

```
ggplot(trade_data[samples,], aes(V1,Trade,fill=Type))+
  geom_bar(stat="identity",position="dodge")+
  ggtitle("The Export and Import of countries")+
  theme_economist(base_size=10)+
  scale_fill_economist()+
  xlab("Countries' ISOs")+
  theme(axis.ticks.length=unit(0.3,'cm'))+
  guides(fill=guide_legend(title=NULL))+
  coord_flip()
```



Different from the degree graph, it is worth noticing that some countries have big difference between import and export amount (e.g. in our 10 samples, STP's export amount is more than 2 times of its import amount; GHA's import amount is more than 10 times of its export amount). We believe it is not a random phenomenan and try to explore the relation of degree and weighted degree.

```
relation_data<-data.frame(cbind(total_trade, total_degree))
colnames(relation_data)<-c("Trade","Degree")
relation_data_log<-data.frame(cbind(total_trade, log(total_degree)))
colnames(relation_data_log)<-c("Trade","LogDegree")
td_graph<-ggplot(relation_data, aes(x=Degree, y=Trade))+geom_point()
td_log_graph<-ggplot(relation_data_log, aes(x=LogDegree, y=Trade))+geom_point()

grid.newpage()
pushViewport(viewport(layout = grid.layout(1, 2)))
print(td_graph, vp = vplayout(1, 1))
print(td_log_graph, vp = vplayout(1, 2))
```



The left figure describes the relation of degree and trade amout. The right figure describes the relation of the log of degree and trade amount. Their shapes are similar, which means that the huge amount of trades tightly happen on countries with large degrees.

In the following part, we calculate three centrialities, betweeness centriality, degree centrialities and closeness centrialities for each country and make some comparisons to figure out sepcial countries' roles.

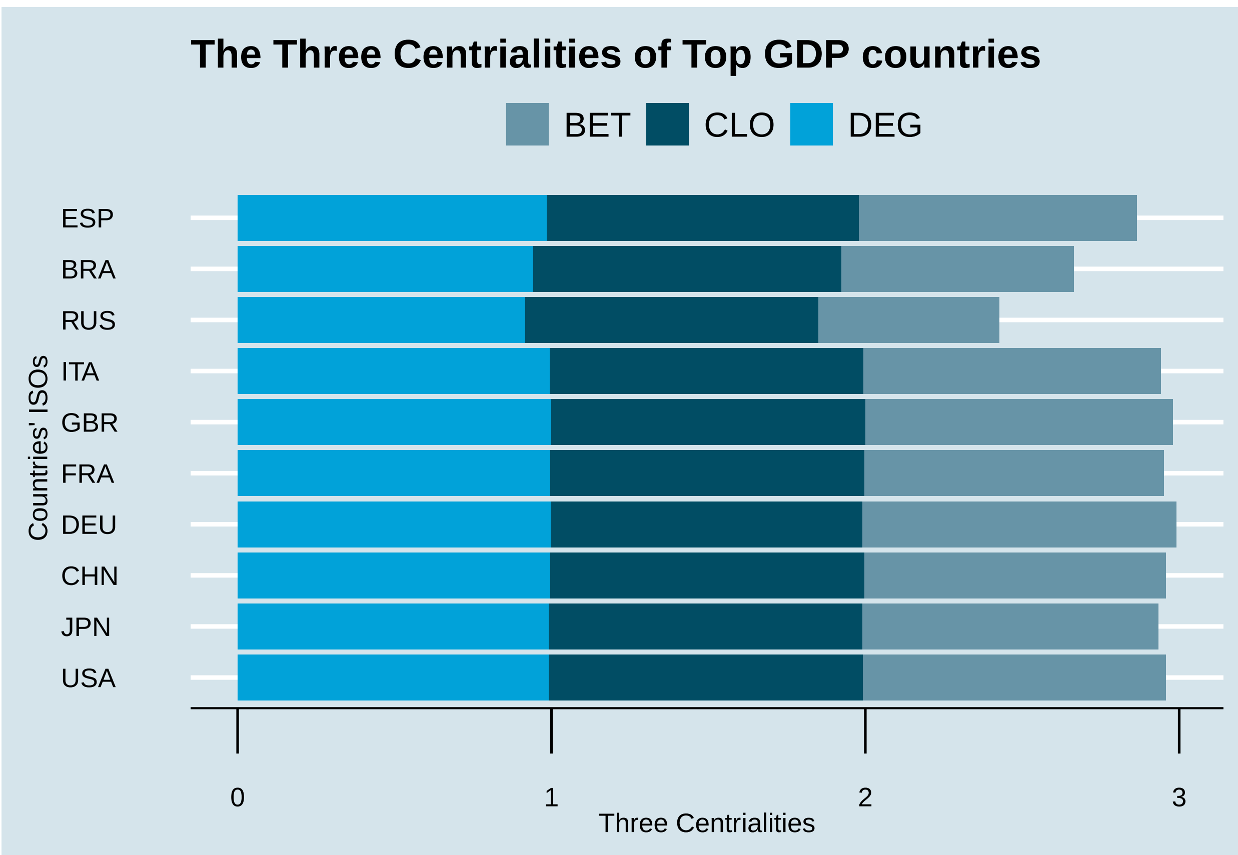
- Betweeness Centriality: $c_B(i) = \Sigma_{s,t \in N} \frac{\sigma(s,t|i)}{\sigma(s,t)}$, describes a node's role as an information medium. It calculated the times of a node in the shorteset paths between two other nodes. In some way, a node with high betweeness centriality controls the information broadcast among clusters. If a node connecting two seperate clusters, it can have a high betweeness centrialities even with low degree centrialities.
- Degree Centriality: $c_D(i) = \frac{\Sigma_j (x_{ij} + x_{ji})}{2(N-1)}$. It describes the number of neighbours of a node, directly reflecting a node's active level.
- Closeness Centriality: $c_C(i) = \frac{R(i)}{N-1} \cdot [\frac{\Sigma_{t \in i \to *t} dist(i,t)}{R(i)}]^{-1}$. It evaluates the average shortest path of a node to the others.

```
# Degree Centrialities
bet_cent<-centr_betw(trade_graph)
clo_cent<-centr_clo(trade_graph)
deg_cent<-centr_degree(trade_graph)

trade_graph <-set_vertex_attr(trade_graph,  "BetCent", index=countries$X, value=bet_cent)
trade_graph <-set_vertex_attr(trade_graph,  "CloCent", index=countries$X, value=clo_cent)
trade_graph <-set_vertex_attr(trade_graph,  "DegCent", index=countries$X, value=deg_cent)

# Select the top10 countries based on gdp and make the figures to see their standarized centrialities.
top10_gdp<-order(countries$gdp, decreasing = TRUE)[1:10]
top10_gdp_bet<-bet_cent[[1]][top10_gdp]
```

```r
top10_gdp_clo<-clo_cent[[1]][top10_gdp]
top10_gdp_deg<-deg_cent[[1]][top10_gdp]

cent_data=data.frame(t(rbind(countries$iso[top10_gdp],top10_gdp_bet/max(bet_cent[[1]]),
                             top10_gdp_clo/max(clo_cent[[1]]),top10_gdp_deg/max(deg_cent[[1]]))))
colnames(cent_data)<-c("ISO","BET","CLO","DEG")
cent_data<-melt(cent_data, id.vars = "ISO", variable.name = "Type", value.name = "Centriality")
```

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

```r
ggplot(cent_data, aes(factor(ISO,levels = countries$iso[top10_gdp]),as.numeric(Centriality),fill=Type))+
  geom_bar(stat="identity",position="stack")+
  ggtitle("The Three Centrialities of Top GDP countries")+
  theme_economist(base_size=10)+
  scale_fill_economist()+
  xlab("Countries' ISOs")+
  ylab("Three Centrialities")+
  guides(fill=guide_legend(title=NULL))+
  theme(axis.ticks.length=unit(0.6,'cm'))+
  coord_flip()
```



We select 10 countries with top10 GDP and show their standarized (the absolute value divied by the maximum) three centrialities index. From the graph, we find those countries with high GDP also have high centrialities in the international trade network.

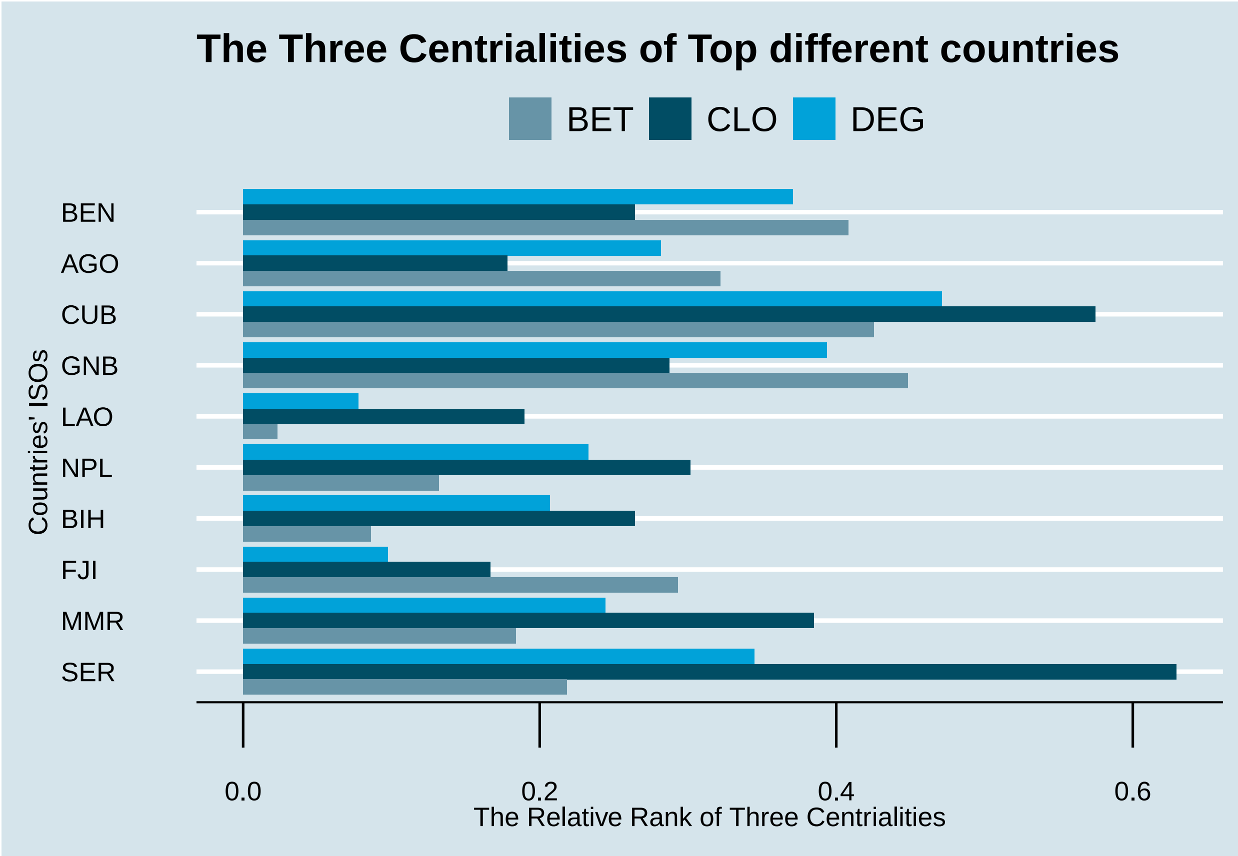
We notice that RUS (Russian Federation), whose GDP (1784.514) ranks 7 worldwide has the lowest summed centrialities compared with the other 9 countries, even obviously lower than ESP (which GDp is 1642.765) and BRA (which GDP is 1694.856). That means RUS plays a relatively inactive role within the similar economics amount countries.

Next, we try to find countries with the most different roles in comparison of three centrialities. We define the role difference by the average absolute value of rank difference in each two of centrialities.

```r
# get the rank of countries in each centrialities, decreasing order
bet_rank<-rank(-bet_cent[[1]])
clo_rank<-rank(-clo_cent[[1]])
deg_rank<-rank(-deg_cent[[1]])
diff_rank<-(abs(bet_rank-clo_rank)+abs(clo_rank-deg_rank)+abs(deg_rank-bet_rank))/3

top10_diff<-order(diff_rank,decreasing = TRUE)[1:10]

# Draw the figure of the most centrialities different countries
top10_diff_bet<-bet_rank[top10_diff]
top10_diff_clo<-clo_rank[top10_diff]
top10_diff_deg<-deg_rank[top10_diff]

# Define the difference by the average absolute value of rank diffs.
diff_data=data.frame(t(rbind(countries$iso[top10_diff],1-top10_diff_bet/174,
                            1-top10_diff_clo/174,1-top10_diff_deg/174)))
colnames(diff_data)<-c("ISO","BET","CLO","DEG")
diff_data<-melt(diff_data, id.vars = "ISO",
                variable.name = "Type", value.name = "Centriality")
```

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

```r
ggplot(diff_data, aes(factor(ISO,levels = countries$iso[top10_diff]),
                      as.numeric(Centriality),fill=Type))+
  geom_bar(stat="identity",position="dodge")+
  ggtitle("The Three Centrialities of Top different countries")+
  theme_economist(base_size=10)+
  scale_fill_economist()+
  xlab("Countries' ISOs")+
  ylab("The Relative Rank of Three Centrialities")+
  guides(fill=guide_legend(title=NULL))+
  theme(axis.ticks.length=unit(0.6,'cm'))+
  coord_flip()
```



We plot the top10 centrality different countries of their relative ranks, which is defined by $relaRank(i) = 1 - \frac{abRank(i)}{174}$. From this graph, we find several insteresting things: - SER (Serbia, Republic of) has a pretty high closeness centriality compared with its betweeness centrialities. It may be a hint that SER has direct links to some centers in this international trade network while itself is not that centered. LAO (Lao People's Democratic Republic), BIH (Bosnia and Herzegovina) are similar to SER in this aspect. - BEN (Benin), AGO (Angola) and GNB (Guinea) have a relatively high betweeness centriality compared to degree/closeness centrialities. We guess they play a hub role to link a cluster to other clusters.

We look some of the important attributes of SER and GNB.

ISO	GDP	OutDeg	InDeg	Export	Import	BetCent	CloCent	DegCent
SER	49.165	127	41	16814.17	8437.52	6.218542	0.7899543	168
GNB	0.869	72	103	2173.452	1345.915	20.00696	0.6313869	175
AGO	84.178	60	96	18052.83	61983.11	9.835055	0.6048951	156

An interesting finding is that GNB is a small country accoring to its GDP but has a much higher betweeness centrality. We refer it has links to some periphery countries or relatively closed clusters and look its ego-graph. In the past part, we do some study on periphery countries. However, there are no more evidence to support this guess.

who is the graph center

First, we have a look at eccentricity.

```
ec_out <- eccentricity(trade_graph, vids = V(trade_graph),
                       mode = c("out"))
```

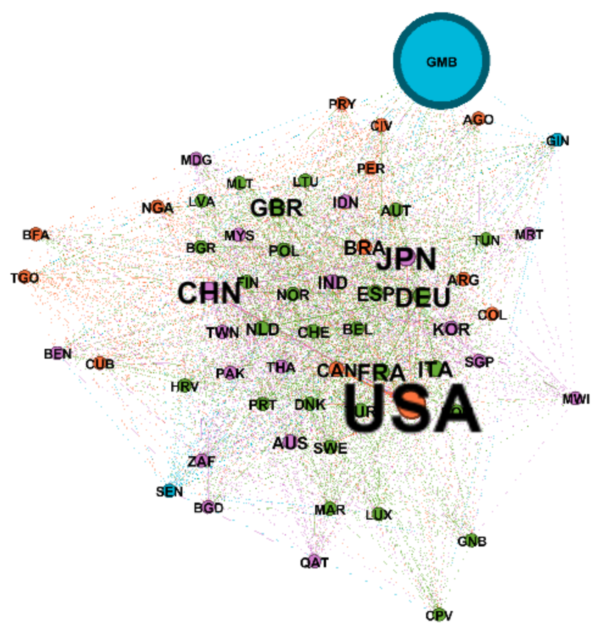
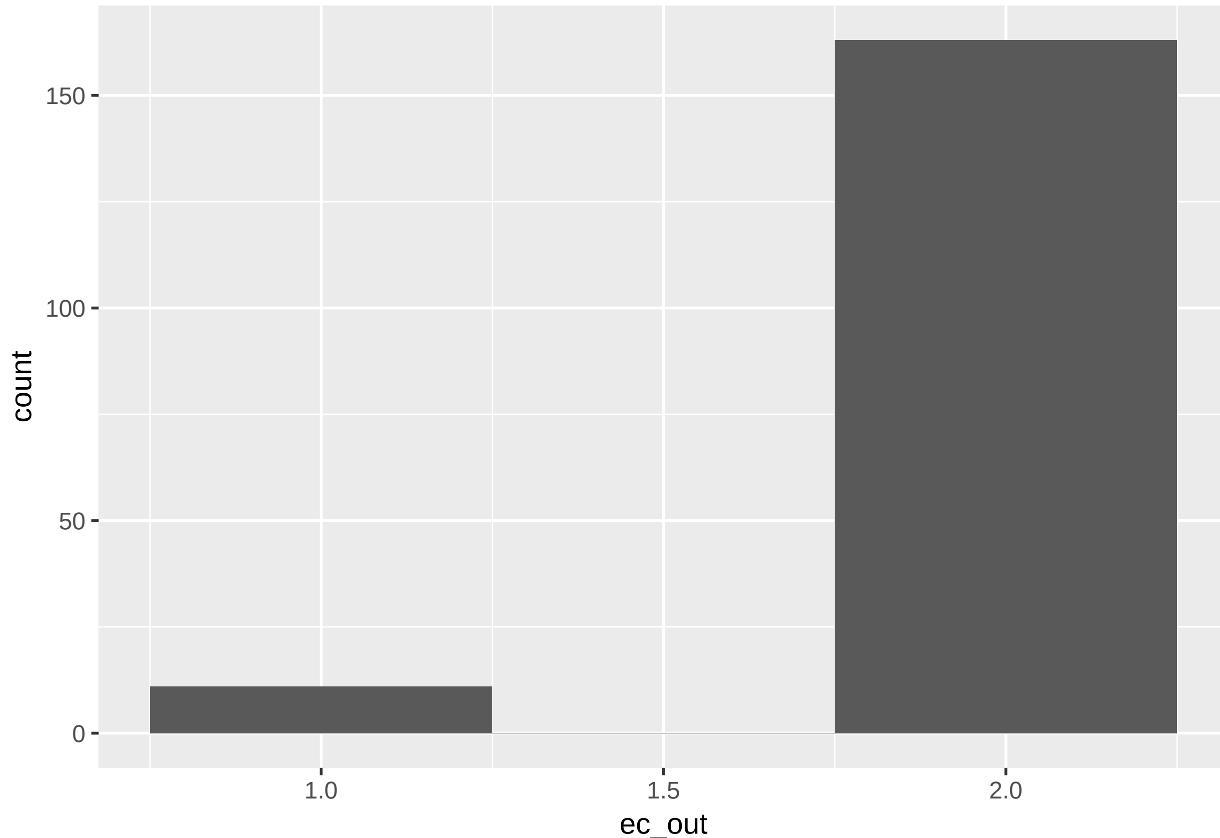


Figure 1: GNB location

```r
ggplot(data.frame(ec_out),aes(x=ec_out))+
  stat_bin(bins=3)+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



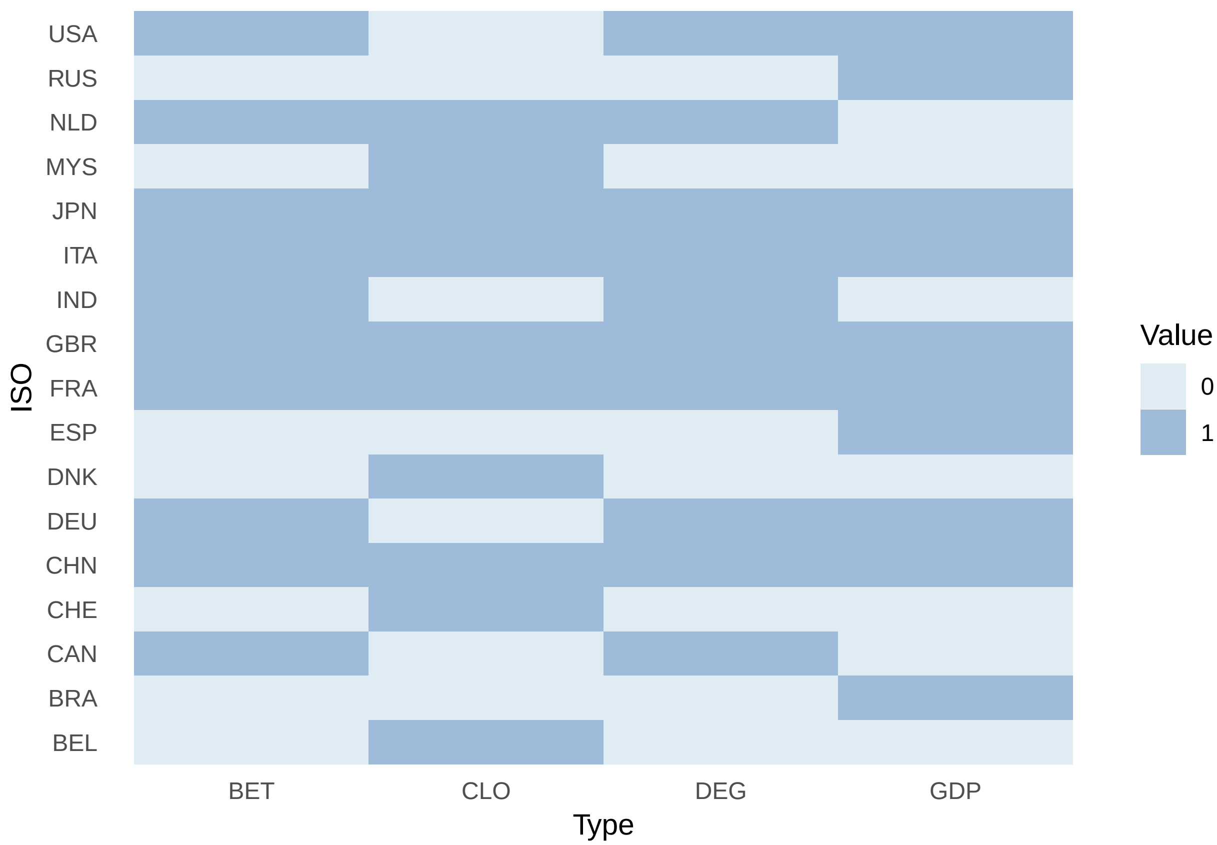

```r
which(ec_out == min(ec_out))
```

```
##             Belgium China, P.R.: Mainland               Denmark
##                  15                    32                    43
##              France                 Italy                 Japan
##                  55                    77                    79
##            Malaysia           Netherlands           Switzerland
##                  97                   110                   150
##      United Kingdom         United States
##                 165                   166
```

We can see that the majority maximum distance between any pair of nodes is just 2. Every node is atmost 2 hops away from each other. By comparision, we can see 11 countries which have a directed edge link with every other country in the graph. In graph theory, vertices with minimum eccentricity are referred to as the centre of the graph. So we find BEL, CHN, DNK, FRA, ITA, JPN, MYS, NLD, CHE, GBR, USA to be the 11 central nodes in this trade graph.

```r
# compare four indices
# select top 10 betweeness/closeness/degree centrality countries
top10_bet<-order(bet_cent[[1]], decreasing = TRUE)[1:10]
top10_bet_country<-countries$iso[top10_bet]
top10_clo<-order(clo_cent[[1]], decreasing = TRUE)[1:10]
```

```r
top10_clo_country<-countries$iso[top10_clo]
top10_deg<-order(bet_cent[[1]], decreasing = TRUE)[1:10]
top10_deg_country<-countries$iso[top10_deg]
top10_gdp_country<-countries$iso[top10_gdp]

source("utils.R")
topn<-data.frame(topn_countries_attries(10,top10_deg_country,top10_clo_country,
                              top10_bet_country, top10_gdp_country))

colnames(topn) <- c("ISO","Type","Value")
ggplot(data = topn) +
  geom_tile(aes(x = Type, y = ISO, fill = Value)) +
  theme_classic() +
  theme(axis.ticks = element_blank(),
        axis.line = element_blank()) +
  scale_fill_brewer(palette = 3)
```



We compare the Top10 countries in betweeness centrialities, closeness centrialities, degree centrialities and GDP. The results is shown in the above figure.

- JPN (Japan), ITA (Italy), GBR (United Kingdom), FRA (France) and CHN (China, P.R.: Mainland) are in all of the four Top10 ranks. They play significant roles in international trades.
- USA (United States) and DEU (Germany) only miss the roles in closeness centriality rank while MYS (Malaysia), DNK (Denmark), CHE (Switzerland) and BEL (Belgium) only appear in this rank. Since closeness centriality describes the average shortest paths, the high value means countries not only trade with important countries but also be closed to edged countries. In other words, USA and DEU may

pay little attention to few countries.
- BRA (Brazil), ESP (Spain) and RUS only appear in Top10 GDP rank. They have strong economics basis but play less active roles in international trades.

In the following analysis, we define "core countries" as countries who appear at least twice in these Top10 ranks. They are USA, NLD (Netherlands), JPN, ITA, IND (India), GBR, FRA, DEU, CHN and CAN (Canada).

positonal dominance

Exports dominance

We say that i dominates j if the sum of all outgoing edges from i is greater that the sum of all outgoing edges from j, $\Sigma_{k\in N} x_{ik} \geq \Sigma_{k\in N} x_{jk}$. We think the countries who export more are among the most important players.

```
# 5 countries that export more
sort(total_export, decreasing = TRUE)[1:5]
```

```
##
## 2010543.8 1138467.8  853683.2  711738.4  674853.2
```

```
# 5 countries that export less
sort(total_export)[1:5]
```

```
##
##  94.52921 143.82782 208.16677 208.88014 253.00481
```

Import dominance

Defined as before, but considering import (ingoing) ties rather than export ones.

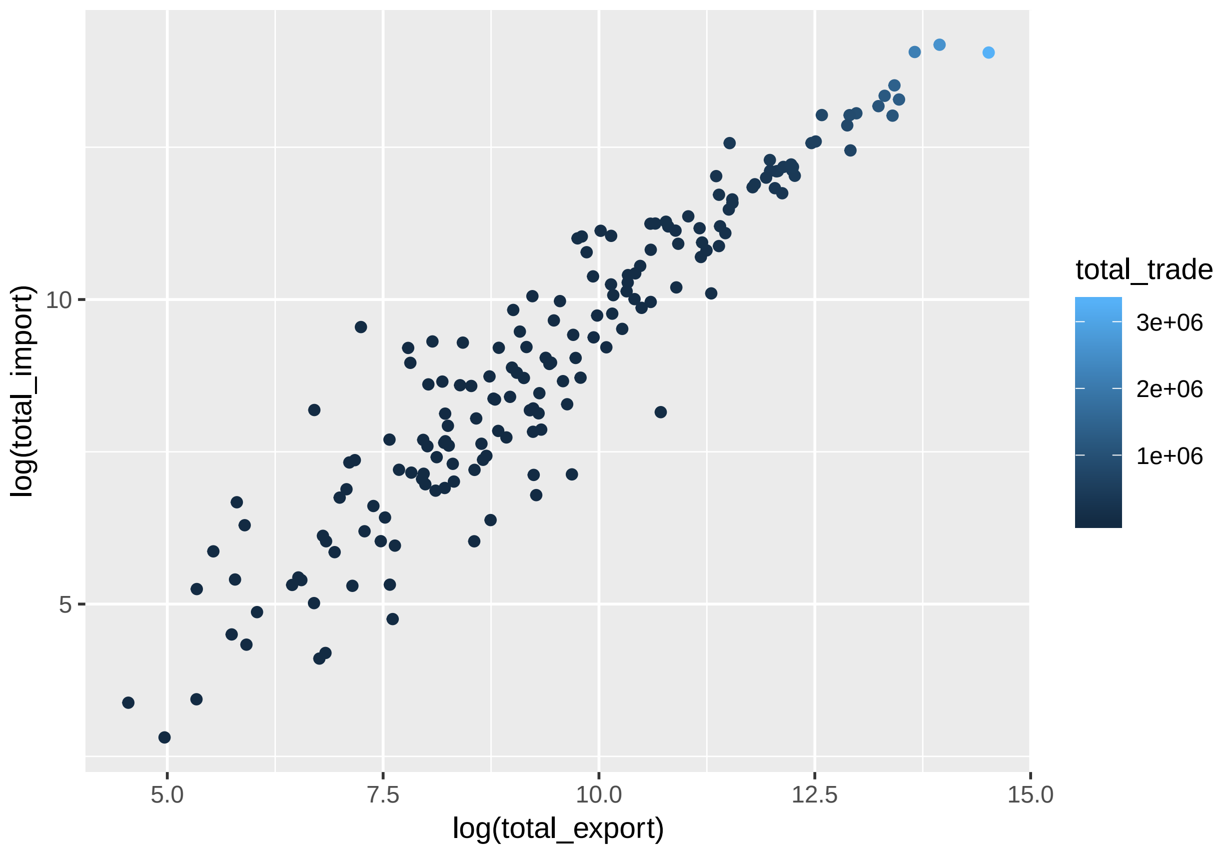
```
# 5 countries that import more
sort(total_import, decreasing=TRUE)[1:5]
```

```
##
## 1443143.9 1281131.2 1269380.6  740526.9  624106.0
```

```
# 5 countries that import less
sort(total_import)[1:5]
```

```
##
## 16.63222 29.40000 31.11000 60.68429 66.53003
```

```
trade_plot<-data.frame(cbind(total_export, total_import,total_trade))
ggplot(trade_plot,aes(x=log(total_export),y=log(total_import),
                      color=total_trade))+
  geom_point()
```



GDP dominance

Here we will simply say that i dominates j if i has a higher GDP than j has.

```
low_threshold <- 200
high_threshold <- 1000

poor_GDP <- V(trade_graph)[V(trade_graph)$gdp <low_threshold]
sort(poor_GDP)[1:5]
```

```
## + 5/174 vertices, named, from d6417d0:
## [1] Afghanistan, Islamic Republic of Albania
## [3] Algeria                             Angola
## [5] Armenia, Republic of
```

```
high_GDP <-  V(trade_graph)[V(trade_graph)$gdp >high_threshold]
sort(high_GDP, decreasing=TRUE)[1:5]
```

```
## + 5/174 vertices, named, from d6417d0:
## [1] United States      United Kingdom     Spain
## [4] Russian Federation Mexico
```

We divide countries in 3 groups based on their GDP level

We divide the countries in three groups based on their GDP level. We will assign a "GDP level" attribute to the countries: this attribute may assume values "low", "medium" or "high" based on the GDP of the considered countrie. The level will be low if $GDP < 200$, medium if $200 \leq GDP \leq 1000$ or high if $GDP > 1000$.

```r
threshold_low <- 200
threshold_high <- 1000
# Set attribute
vertex_attr(trade_graph, "GDP_level") <-
  ifelse(get.vertex.attribute(trade_graph, "gdp")<threshold_low, "low",
         ifelse(get.vertex.attribute(trade_graph, "gdp")
                >threshold_high, "high", "medium"))
```

Positional dominance with respect to different groups (who exports more towards each group)

Here we want to understand who exports more in each group. This way we understand who gets more richness from each group.

We will say that i dominates j with respect to group k if the sum of exports ties from i to all vertices in group k is greater or equal to the corresponding sum of export ties from j to vertices in group k, $\Sigma_{t \in K} x_{it} \geq \Sigma_{t \in K} x_{jt}$

```r
# Scores based on each group
getID_low <- as.vector(countries$X[countries$gdp<threshold_low])
group1 <- as.matrix(exports[1:174,getID_low])
group1 <- rowSums(group1)
trade_graph<-set_vertex_attr(trade_graph,
                             "exports_to_low", index=countries$X, value = group1)

getID_medium <- as.vector(countries$X[countries$gdp>= threshold_low & countries$gdp <= threshold_high])
group2 <- as.matrix(exports[1:174,getID_medium])
group2 <- rowSums(group2)
trade_graph<-set_vertex_attr(trade_graph,
                             "exports_to_medium", index=countries$X, value = group2)

getID_high <- as.vector(countries$X[countries$gdp>threshold_high])
group3 <- as.matrix(exports[1:174,getID_high])
group3 <- rowSums(group3)
trade_graph<-set_vertex_attr(trade_graph,
                             "exports_to_high", index = countries$X, value = group3)

# getting some results
# first 5 exporters to low group
sort(group1, decreasing=TRUE)[1:5]
```

```
##       CHN       USA       DEU       RUS       JPN
## 200159.08 155488.10 139428.78 119020.55  97680.43
```

```r
# first 5 exporters to medium group
sort(group2, decreasing=TRUE)[1:5]
```

```
##      DEU      CHN      USA      NLD      JPN
## 616087.5 289901.6 278990.6 204245.1 190159.4
```

```r
# first 5 exporters to high group
sort(group3, decreasing=TRUE)[1:5]
```

```
##      USA      CHN      DEU      JPN      CAN
## 834901.9 791070.5 687627.6 452687.0 415085.5
```

Positional dominance with respect to different groups (who imports more from each group)

Similar as before, but this time we want to understand who imports more from each group.

```
# Scores based on each group
getID_low <- as.vector(countries$X[countries$gdp<threshold_low])
group1 <- as.matrix(exports[getID_low, 1:174])
group1 <- colSums(group1)
trade_graph<-set_vertex_attr(trade_graph,
                             "imports_from_low", index=countries$X, value = group1)

getID_medium <- as.vector(countries$X[countries$gdp
                                      >=threshold_low & countries$gdp <= threshold_high])
group2 <- as.matrix(exports[getID_medium, 1:174])
group2 <- colSums(group2)
trade_graph<-set_vertex_attr(trade_graph,
                             "impors_from_medium", index=countries$X, value = group2)

getID_high <- as.vector(countries$X[countries$gdp>threshold_high])
group3 <- as.matrix(exports[getID_high, 1:174])
group3 <- colSums(group3)
trade_graph<-set_vertex_attr(trade_graph,
                             "imports_from_high", index = countries$X, value = group3)

# getting some results (get first 5 countries for each group)
# first 5 importers from low group
sort(group1, decreasing=TRUE)[1:5]
```

```
##       USA       CHN       DEU       JPN       ITA
## 218076.56 144701.30 117141.25 110964.75  99382.23
```

```
# first 5 importers from medium group
sort(group2, decreasing=TRUE)[1:5]
```

```
##      DEU      USA      FRA      GBR      CHN
## 571788.7 418348.4 260818.6 260613.6 232935.0
```

```
# first 5 importers from high group
sort(group3, decreasing=TRUE)[1:5]
```

```
##       USA       CHN       DEU       FRA       NLD
## 1374118.8  476046.9  449537.9  378844.4  367461.8
```

Who exports more within each group

Now we are interested in understanding who are the important players WITHIN each group. We say that for every two countries i, j in group k, i dominates j with respect to export if the sum of all outgoing ties from i to countries in group k is greater than the same value for j, $\Sigma_{t \in K} x_{it} \geq \Sigma_{t \in k} x_{jt}, i, j \in K$

```
# Scores based on each group
getID_low <- as.vector(countries$X[countries$gdp<threshold_low])
group1 <- as.matrix(exports[getID_low,getID_low])
group1 <- rowSums(group1)


getID_medium <- as.vector(countries$X[countries$gdp
```

```
                                      >=threshold_low & countries$gdp <= threshold_high])
group2 <- as.matrix(exports[getID_medium,getID_medium])
group2 <- rowSums(group2)


getID_high <- as.vector(countries$X[countries$gdp>threshold_high])
group3 <- as.matrix(exports[getID_high,getID_high])
group3 <- rowSums(group3)

sort(group1, decreasing=TRUE)[1:5]
```

```
##      SGP      UKR      HUN      KWT      SYR
## 42468.24 21172.42 17342.64 12061.67 10090.21
```

```
sort(group2, decreasing=TRUE)[1:5]
```

```
##       NLD       BEL       SWE       NOR       SAU
## 204245.09 130223.98  83150.54  51699.05  50822.84
```

```
sort(group3, decreasing=TRUE)[1:5]
```

```
##      USA      CHN      DEU      JPN      CAN
## 834901.9 791070.5 687627.6 452687.0 415085.5
```

Who imports more within each group

Defined as before, but with respect to import.

```
# Scores based on each group
getID_low <- as.vector(countries$X[countries$gdp< threshold_low])
group1 <- as.matrix(exports[getID_low, getID_low])
group1 <- colSums(group1)


getID_medium <- as.vector(countries$X[countries$gdp
                                      >=threshold_low & countries$gdp <= threshold_high])
group2 <- as.matrix(exports[getID_medium, getID_medium])
group2 <- colSums(group2)


getID_high <- as.vector(countries$X[countries$gdp> threshold_high])
group3 <- as.matrix(exports[getID_high, getID_high])
group3 <- colSums(group3)


# getting some results
sort(group1, decreasing=TRUE)[1:5]
```

```
##      SGP      UKR      HUN      VNM      PHL
## 22992.42 20432.52 10726.96 10525.91 10320.65
```

```
sort(group2, decreasing=TRUE)[1:5]
```

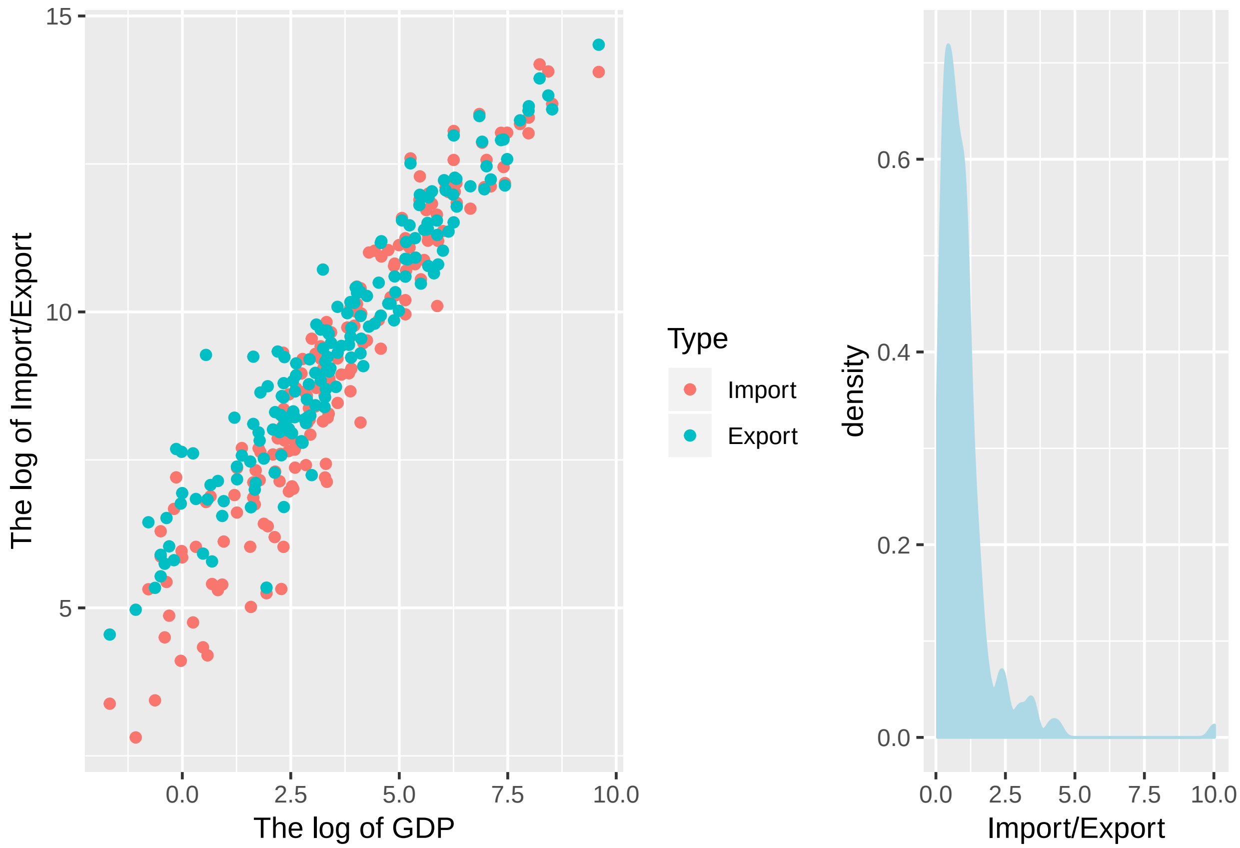
```
##       NLD       BEL       SWE       POL       TUR
## 172120.77 164435.83  73534.61  57292.45  46546.99
```

```r
sort(group3, decreasing=TRUE)[1:5]
```

```
##       USA       CHN       DEU       FRA       GBR 
## 1374118.8  476046.9  449537.9  378844.4  353748.1
```

Correlations

```r
in_ex_gdp<-data.frame(get.vertex.attribute(trade_graph, "gdp"),
                   get.vertex.attribute(trade_graph, "Import"),
                   get.vertex.attribute(trade_graph, "Export"))
colnames(in_ex_gdp)<-c("gdp","Import", "Export")
in_ex_gdp<-melt(in_ex_gdp, id.vars = "gdp",
                variable.name = "Type", value.name = "Amount")

in_ex_gdp_g<-ggplot(in_ex_gdp,aes(x=log(gdp),y=log(Amount), color=Type))+
  geom_point()+
  xlab("The log of GDP")+
  ylab("The log of Import/Export")

in_ex<-data.frame(get.vertex.attribute(trade_graph,"Import"),
                  get.vertex.attribute(trade_graph, "Export"))
colnames(in_ex)<-c("Import","Export")

in_ex_g<-ggplot(in_ex, aes(x=Import/Export)) +
  geom_density(fill = "lightblue", colour = "lightblue")

grid.newpage()
pushViewport(viewport(layout = grid.layout(1, 3)))
print(in_ex_gdp_g, vp = vplayout(1, 1:2))
print(in_ex_g, vp = vplayout(1, 3))
```



Overall conclusions about this section

There are countries that dominate the trade scene both for total import and total export: among these the most important are China, Germany, USA (we showed there is positive correlation between total import and total export, so the ones who export more are also the ones who import more).

Whe defined the simple concept of GDP dominance and divided the countries in 3 main cathegories (low, medium, high GDP).

It turned out that the leading countries for import and export were also the ones who exported, imported more with respect to each GDP group.

In the end we tried to understand who are the most important countries within the groups. China, Japan, USA and Germany remain the most important in the high GDP group. We found out that the most important countries in the medium level GDP group are Netherlands, Belgium, Sweden, Poland, Turkey, Saudi Arabia (they are important both for import and export).

Finally, the most important import/export countries in the low-level group are Singapore, Ukraine, Hungary, Vietnam Philippines: these are among the most important countries in the lowGDP group.

Whole network structure

First, we filter the trade networks, leaving some unimportant nodes out. The unimportant nodes are defined as the intersect set of the coutries with total trade less than 2000 and those with GDP less than 10. We plot the filtered graph to get a basic idea.

```
unimport_countries<-intersect(countries$X[countries$gdp<10],countries$X[total_trade<2000])
filtered_countries<-setdiff(countries$X, unimport_countries)
```

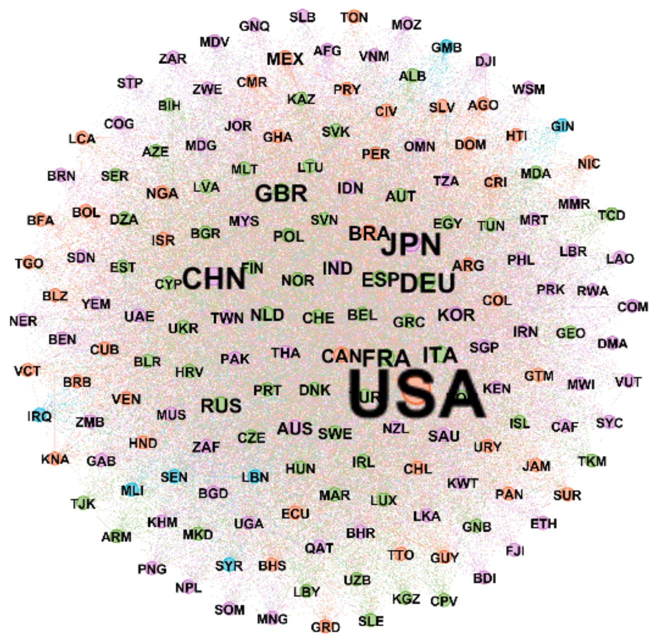


Figure 2: filtered network graph

```
core_countries_iso<-c("USA","NLD","JPN","ITA","IND","GBR","FRA","DEU","CHN","CAN")
core_countries<-NULL
for (e_count in core_countries_iso){
  core_countries<-c(core_countries, getID(e_count))
}

filtered_v <- V(trade_graph)[filtered_countries]
filtered_graph <- induced_subgraph(trade_graph, filtered_v)
walk_comm<-cluster_walktrap(filtered_graph)
members<-membership(walk_comm)
filtered_graph<-set.vertex.attribute(filtered_graph,"member", value = members)

write.graph(filtered_graph, "./filtered.graphml","graphml")
```

The node size is based on the TotalTrade, the colors of nodes represent their memberships got from a community dectection realized by Vincent D Blomdel ect. and the colors and sizes of edges are related to edges' weights.

Cliques Analysis

```
countries$coreness <- graph.coreness(trade_graph, mode='out')
table(countries$coreness)
```

```
##
## 20 26 30 31 32 34 35 36 37 38 41 42 46 47 48 51 52 53 54 55 57 58 59 60 61
##  1  1  1  1  1  2  1  1  1  3  1  1  4  1  1  1  3  2  1  1  4  2  1  1  2
## 62 63 64 65 66 67 68 69 70 71 72 73
##  3  3  3  5  2  6  6  2  3  7  3 92
```

```
# Experiment using largest.cliques
export_clique<-largest.cliques(trade_graph)
```

```
## Warning in largest.cliques(trade_graph): At cliques.c:1087 :directionality
## of edges is ignored for directed graphs
```

```
i1<-intersect(export_clique[[1]],export_clique[[2]])
i2<-intersect(i1,export_clique[[3]])
i3<-intersect(i2,export_clique[[4]])
i4<-intersect(i3,export_clique[[5]])
i5<-intersect(i4,export_clique[[6]])
i6<-intersect(i5,export_clique[[7]])
length(i6)
```

```
## [1] 70
```

Looking at the table, we can see that computing the coreness is not very informative, since 92 of the contries have a coreness of 73, meaning that 93 countries are in a maximal subgraph in which every vertex has at least degree 73.

Is it A Small World?

A small-world network is a type of mathematical graph in which - most nodes are not neighbors of one another, - but the neighbors of any given node are likely to be neighbors of each other - and most nodes can be reached from every other node by a small number of hops or steps, - while the clustering coefficient is not small.

Specifically, a small-world network is defined to be a network where the typical distance L between two randomly chosen nodes (the number of steps required) grows proportionally to the logarithm of the number of nodes N in the network.

From the former observation, nodes's degree distribution tends to be even and some nodes have nearly full degrees. The network density is also high 0.6117866. These are against "most nodes are not neighbors of one another". However, We see this networks shares some other features with a small-world network.

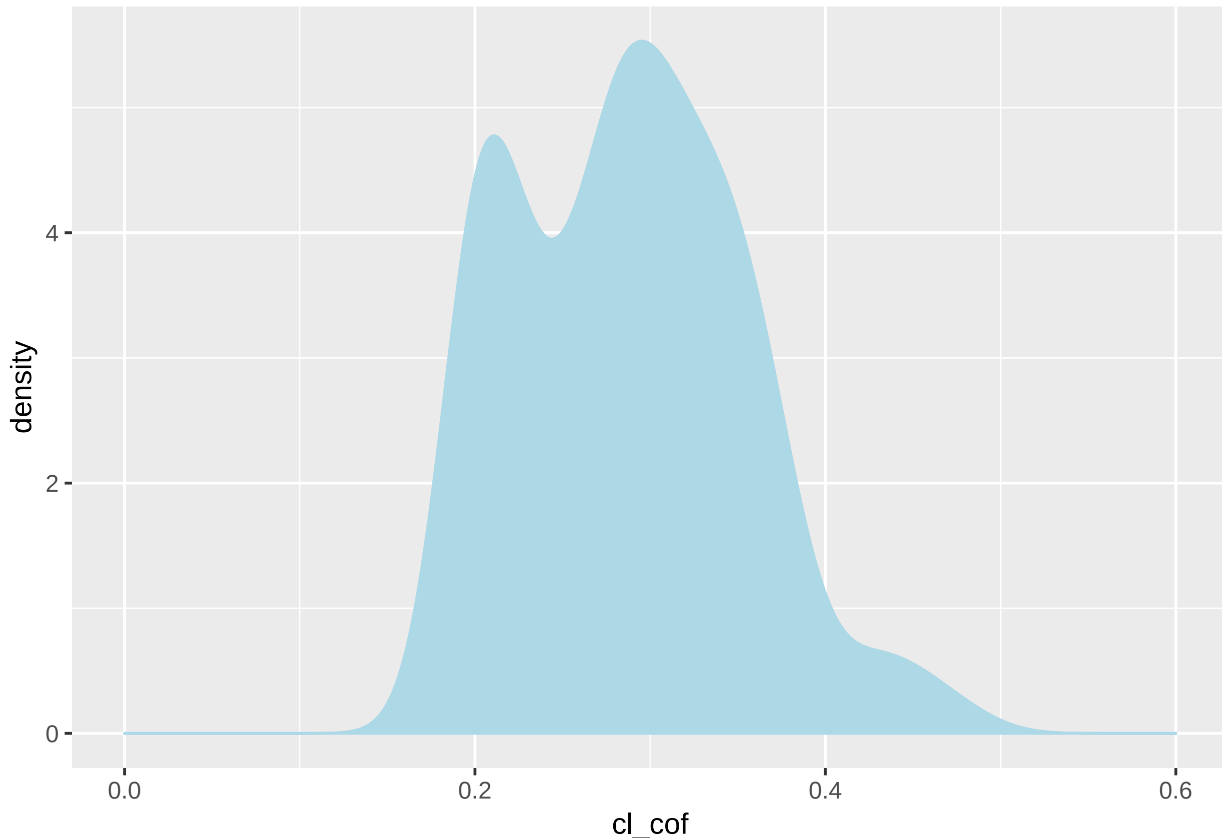
clustering coefficient

First, we look at the clustering coefficient, $weightedC_i = \frac{1}{s_i}\frac{1}{k_i-1}\Sigma_{j,h}(\frac{w_{ij}+w_{ih}}{2}a_{ij}a_{ih}a_{jh})$,$k_i$ is the vertex degree, w_{ij} are the weights, a_{ij} are elements of the adjacency matrix, and $s_i = \Sigma_{j \in N}(w_{ij} + w_{ji})$ is the strength of vertex i. It measures the probability that the adjacent vertices of a vertex are connected. The local clustering coefficient describes the interlap level of two neighbours' neighbourhood.

We see the clustering coefficient for the whole trade network is 0.790237 and for the filtered network is 0.8450512. They are pretty high, closed to 1. The following is the distribution of each node's clustering coffecient. It is a concentrated distribution with the minimum 0.1956929 and the maximum 0.4793239.

```
cl_cof<-transitivity(filtered_graph, type="local")
ggplot(data.frame(cl_cof), aes(x=cl_cof)) +
```

```r
xlim(0,0.6)+
geom_density(fill = "lightblue", colour = "lightblue")
```



Average Trade Path

The second feature of a samll world network is that it has short average path. We check this feature on our network.

```r
(apl<-average.path.length(filtered_graph))
```

```
## [1] 1.293951
```

```r
shortestp<-shortest.paths(filtered_graph, weights = NA)
samples<-sample(1:151,10)
shortestp[samples,samples]
```

```
##               Cuba Italy Pakistan United States Slovenia Cameroon Israel
## Cuba             0     1        1             1        1        2      2
## Italy            1     0        1             1        1        1      1
## Pakistan         1     1        0             1        1        1      2
## United States    1     1        1             0        1        1      1
## Slovenia         1     1        1             1        0        1      1
## Cameroon         2     1        1             1        1        0      1
## Israel           2     1        2             1        1        1      0
## Australia        1     1        1             1        1        1      1
## Bangladesh       2     1        1             1        1        1      1
## Switzerland      1     1        1             1        1        1      1
```

```
##               Australia Bangladesh Switzerland
## Cuba                  1          2           1
## Italy                 1          1           1
## Pakistan              1          1           1
## United States         1          1           1
## Slovenia              1          1           1
## Cameroon              1          1           1
## Israel                1          1           1
## Australia             0          1           1
## Bangladesh            1          0           1
## Switzerland           1          1           0
```

The average shortest path is pretty low, only 1.2939514, which corresponds with our hypothesis. Countries can easily touch the other countries. We randomly choose 10 countries and see their mutual shortest path distances. All of them are 1 or 2 (except for 0 to themselves).

In all, the above features show that the worldwide trade environment is much open nowadays and countries are more free to trade with each other.

"Core-Periphery" Network

Core-periphery model assumes that there are two classes of nodes. - The first consists of a cohesive core sub-graph in which the nodes are highly interconnected - and the second is made up of a peripheral set of nodes that is loosely connected to the core.

In an ideal core–periphery matrix, core nodes are adjacent to other core nodes and to some peripheral nodes while peripheral nodes are not connected with other peripheral nodes.

Core countries

We try on the *largest.cliques* function in igraph package. Since it defines the largest cliques on undirected networks, merge the directed edges and give us very large cliques (the size of intersect of all of them is 70), we decide not take its results as the core-networks. As we discussed in centralities, we use USA, NLD (Netherlands), JPN, ITA, IND (India), GBR, FRA, DEU, CHN and CAN (Canada) as the core coutries.

```r
# Analyse the core.
core_v <- V(trade_graph)[core_countries]
core_graph <- induced_subgraph(trade_graph, core_v)
transitivity(core_graph)
```

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

```r
graph.density(core_graph)
```

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

```r
write.graph(core_graph,"./core.graphml","graphml")
```

We find that the core-countries's network if fully-connected. They have bi-directional trades with each of the other country. In the graph,

The size of nodes represents gdp, the size of edges represents the total trade amount between the two countries. It is clear that USA has the largest trade amount with other countries, and links to CAN. There is an obvious triangles among USA, JPN and CHN. Meanwhile, DEU plays an important role in connecting Europe with the rest of the world. Also, from the import graph,

we can find DEU's import amount similar to CHN and USA. It can be another evidence of its hub role in Europe.

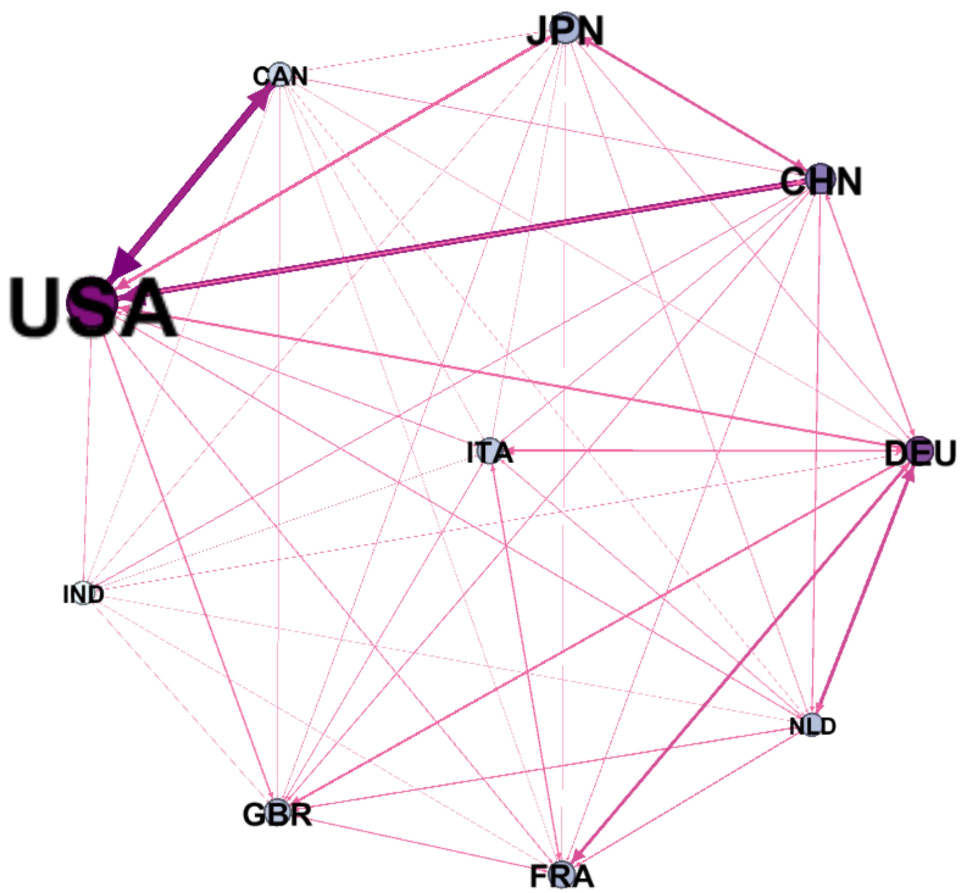


Figure 3: Core Countries

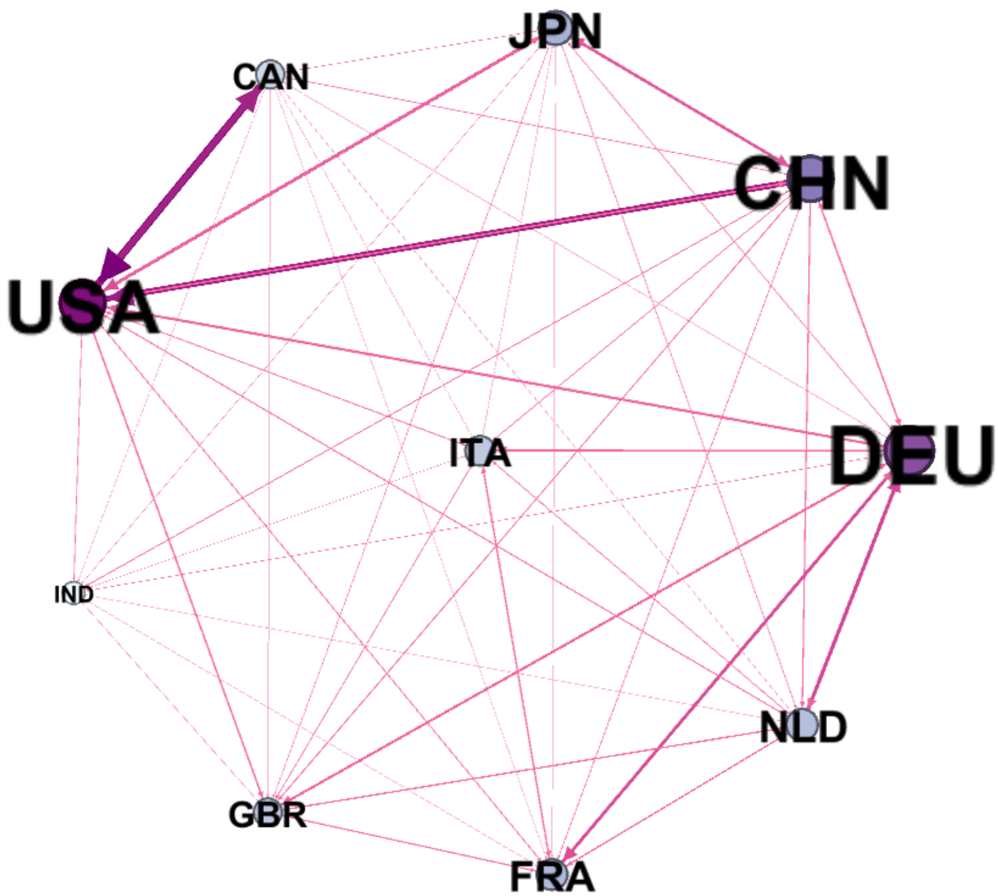


Figure 4: Import Network of Core Countries

Semi-periphery countries

It is composed of countreis who have both import and export relations with all of the core countries. The other nodes are periphery countries. Then, we study the ego-graph of each country.

```
peri_exp<-rowSums(exports[,core_countries]==0)
peri_imp<-colSums(exports[core_countries,]==0)
peri_countries<-setdiff(countries$X[(peri_exp+peri_imp)>0],core_countries)

semi_countries<-setdiff(countries$X,core_countries)
semi_countries<-setdiff(semi_countries, peri_countries)

# study ego-graphs of each core country
ego_results_ex<-NULL
for (iso in core_countries_iso){
  ego_results_ex<-rbind(ego_results_ex,ExploreEgo(iso, "out"))
}
ego_results_in<-NULL
for (iso in core_countries_iso){
  ego_results_in<-rbind(ego_results_in,ExploreEgo(iso, "in"))
}
ego_results<-NULL
for (iso in core_countries_iso){
  ego_results<-rbind(ego_results,ExploreEgo(iso, "all"))
}
ego_results
```

```
##       ISO   Size  Density             Bet Ego
##  [1,] "USA" "174" "0.611786592253006" "265.745751054268"
##  [2,] "NLD" "174" "0.611786592253006" "262.950392508483"
##  [3,] "JPN" "174" "0.611786592253006" "259.386259135849"
##  [4,] "ITA" "174" "0.611786592253006" "260.908014876819"
##  [5,] "IND" "174" "0.611786592253006" "253.162995549843"
##  [6,] "GBR" "174" "0.611786592253006" "269.545783141376"
##  [7,] "FRA" "174" "0.611786592253006" "262.285375542603"
##  [8,] "DEU" "173" "0.608079042882108" "281.491221193633"
##  [9,] "CHN" "174" "0.611786592253006" "264.339518879355"
## [10,] "CAN" "173" "0.613254469686786" "263.812315994954"
##       Bet Cent             Constraint
##  [1,] "0.00703087467830588" "0.132750179016961"
##  [2,] "0.00703087467830588" "0.208300652958078"
##  [3,] "0.00703087467830588" "0.193907166482238"
##  [4,] "0.00703087467830588" "0.17072484818593"
##  [5,] "0.00703087467830588" "0.166486508412764"
##  [6,] "0.00703087467830588" "0.181216752295018"
##  [7,] "0.00703087467830588" "0.195629342208345"
##  [8,] "0.00732100749624789" "0.134477582290934"
##  [9,] "0.00703087467830588" "0.162114607299819"
## [10,] "0.00695468839915401" "0.636583448883311"
```

```
ego_results_ex
```

```
##       ISO   Size  Density             Bet Ego
##  [1,] "USA" "174" "0.611786592253006" "265.745751054268"
##  [2,] "NLD" "174" "0.611786592253006" "262.950392508483"
##  [3,] "JPN" "174" "0.611786592253006" "259.386259135849"
```

```
##  [4,] "ITA" "174" "0.611786592253006" "260.908014876819"
##  [5,] "IND" "172" "0.616381068951448" "249.67922445412"
##  [6,] "GBR" "174" "0.611786592253006" "269.545783141376"
##  [7,] "FRA" "174" "0.611786592253006" "262.285375542603"
##  [8,] "DEU" "173" "0.608079042882108" "281.491221193633"
##  [9,] "CHN" "174" "0.611786592253006" "264.339518879355"
## [10,] "CAN" "173" "0.613254469686786" "263.812315994954"
##       Bet Cent              Constraint
##  [1,] "0.00703087467830588" "0.132750179016961"
##  [2,] "0.00703087467830588" "0.208300652958078"
##  [3,] "0.00703087467830588" "0.193907166482238"
##  [4,] "0.00703087467830588" "0.17072484818593"
##  [5,] "0.00687247926383591" "0.166562373144185"
##  [6,] "0.00703087467830588" "0.181216752295018"
##  [7,] "0.00703087467830588" "0.195629342208345"
##  [8,] "0.00732100749624789" "0.134477582290934"
##  [9,] "0.00703087467830588" "0.162114607299819"
## [10,] "0.00695468839915401" "0.636583448883311"
```

```
ego_results_in
```

```
##       ISO   Size  Density             Bet Ego
##  [1,] "USA" "170" "0.619979115906718" "29.0632710053369"
##  [2,] "NLD" "172" "0.620767033863729" "255.045220662585"
##  [3,] "JPN" "170" "0.624155934563174" "71.4939361079663"
##  [4,] "ITA" "171" "0.623701410388717" "236.22036781009"
##  [5,] "IND" "171" "0.623047815617475" "32.627756284818"
##  [6,] "GBR" "173" "0.616279069767442" "262.123171537464"
##  [7,] "FRA" "172" "0.621549027607779" "253.409031501266"
##  [8,] "DEU" "173" "0.608079042882108" "281.491221193633"
##  [9,] "CHN" "172" "0.62015503875969"  "105.555266839663"
## [10,] "CAN" "172" "0.617537059703522" "260.689002585127"
##       Bet Cent              Constraint
##  [1,] "0.00681583617699316" "0.44271670194827"
##  [2,] "0.00658096467167667" "0.208304163076754"
##  [3,] "0.00660056557430155" "0.13806058507609"
##  [4,] "0.00658205357074668" "0.193942108855732"
##  [5,] "0.00659932913204827" "0.191105401286562"
##  [6,] "0.00683417134034527" "0.132750662739776"
##  [7,] "0.00652897792950224" "0.195630550140667"
##  [8,] "0.00732100749624789" "0.134477582290934"
##  [9,] "0.00670154307194626" "0.343620353743911"
## [10,] "0.00675713360664856" "0.636585588158099"
```

Since the network are highly-connected, we see many similar values of each attribute among all of the core countries. We plot a table of each attributes with the network modes of "all", "in" and "out". "Size" is the number of nodes in the ego-graph, "Density" is the ego-graph's density $\frac{2E}{N(N-1)}$, "Bet Ego" is the core's betweeness centriality in its ego-graph, "Bet Cent" is the betweeness centrality of the ego-graph and "Constraint" is the Burt's constraintwiki link of the core country in its ego-graph. Burt's constraint measures the extent to which the manager's network of colleagues is like a straitjacket around the manager, limiting his or her vision of alternative ideas and sources of support. It is higher if ego has less, or mutually stronger related (i.e. more redundant) contacts.

There are still some details worthy noticing. - CAN has a much higher constraint in all of the three types compared to other core countries. - USA's constraint on import 0.44271670194827 is much higher than its on export 0.132750179016961 and total trade 0.132750179016961, and also higher than most of other countries

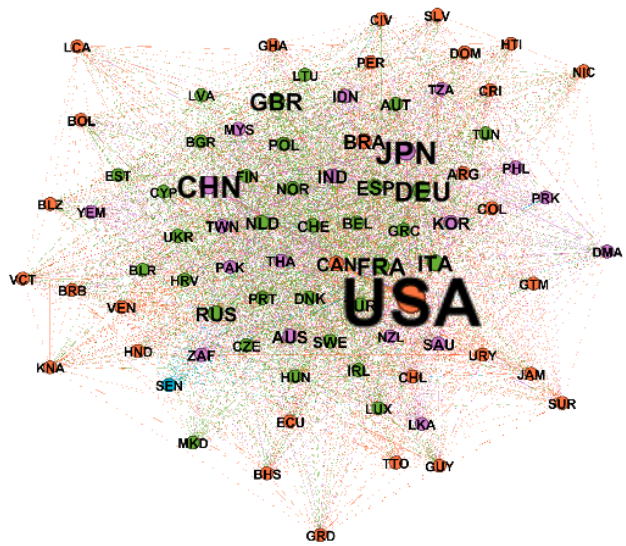


Figure 5: Djibouti graph

in import type. It can be referred that USA's economy doesn't rely much on import or it limits its import resources. - JPN has the few constraint on import 0.13806058507609. We argue that it is because Japan highly rely on import to develop. - DEU owns the fewest constraint on import 0.134477582290934 and nearly the fewest constraint on export 0.134477582290934. This evidence is consistent with our former guess of its hub role in Europe.

Periphery Coutries

In the last, we turn our attention to thest periphery countries. We try to find their location in the graph.

We select several countries' ego-graph and find in these countries' graph, colors (which are the result of walktrap cluster) have a fantastic distribution.

In Djibouti's ego-graph, there are some red countries. Djibouti is in the north-east of Africa. We guess the colors are related to continents. The case is the same to Guinea-Bissau, another African country.

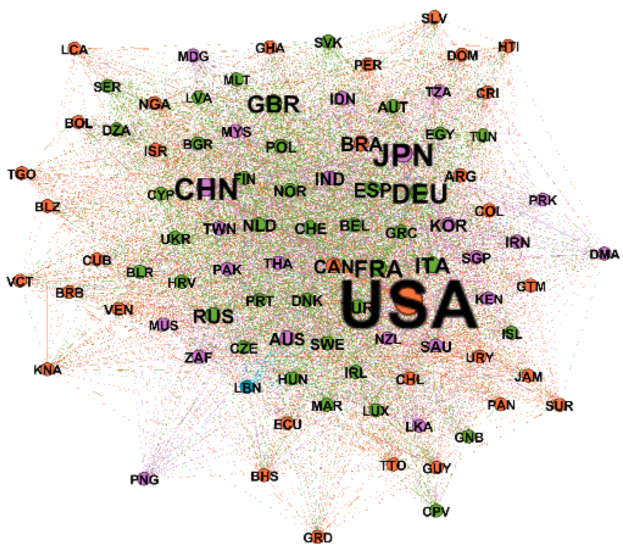
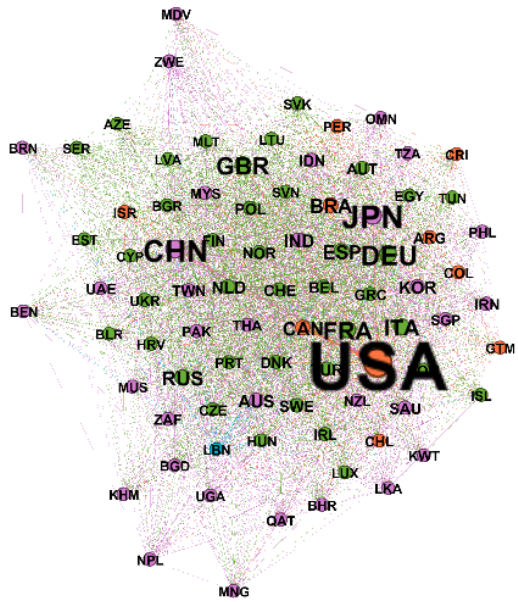


Figure 6: Guinea-Bissau



In Myanmar's ego-graph, most of the countries are green and purple. Myanmar is located in south-east of Asia, same to CHN, JPN, TWN, IND and THA, which are its neighbours and also in Asia.

To conclude, when analysing this international trade network, we want to share the following feelings: - this network is very dense. It means if we remove the edges' weights and only consider the numbers, it is hard to find meaningful structures. In other words, to a dense network, its unweighted version will lose a lot of valuable information, and should be the second to the weighted version to study. What's more, some common network analysis techniques lose power in the face of a dense network, e.g. degrees, closeness centrality. - We filter the countries by virtue of GDP and trade amount. It is also worthy to dropping edges under some threshold. This way may perform better to simplify the given network, without using additional information. - International trade environment is quite open nowadays. Almost all of the countries can trade with each other. Some countries, like USA, JPN, CHN, and DEU play top roles on linking different markets and contributing to trade amount. - Some periphery countries are significant to identify the local trade circles, for they are more likely to stuck in the local markets.