Social network analysis hw1: export-network

Yijia Lin, Diego Paroli

2025-04-23

Libraries

```
rm(list = ls())
library(tidyverse)
library(httr2)
library(igraph)
library(tidygraph)
library(ggraph)
library(visNetwork)
```

Get the data

I have commented below to avoid calling it every time we render the document

```
# zip_data <- request("https://networks.skewed.de/net/product_space/files/SITC.csv.zip") |>
#   req_perform()
#
# writeBin(resp_body_raw(zip_data), "exports_SITC.csv.zip")
#
# unzip("exports_SITC.csv.zip", exdir = "network-data")
#
# file.remove("exports_SITC.csv.zip", )
```

```
nodes <- read_csv("network-data/nodes.csv")
links <- read_csv("network-data/edges.csv")
```

```
head(nodes)
```

```
## # A tibble: 6 x 9
##   `# index`   pid community  size pos                   leamer name  color `_pos`
##       <dbl> <dbl>     <dbl> <dbl> <chr>                  <dbl> <chr> <chr> <chr>
## 1         0  6932         0  48.8 array([4551.8996582~      8 WIRE~ "#9c~ array~
## 2         1  7362         0  65.2 array([ 216.8350982~      9 META~ "#40~ array~
## 3         2  7911         0  54.0 array([ 538.9149017~      9 RAIL~ "#40~ array~
## 4         3  8946         0  57.7 array([ 696.3942565~      7 NON-~ "#40~ array~
## 5         4  7264         0  73.3 array([  57.2840652~      9 PRIN~ "#40~ array~
## 6         5  2783         0  58.3 array([4662.2502441~      2 COMM~ "#ff~ array~
```

```r
head(links)
```

```
## # A tibble: 6 x 4
##   `# source` target width color
##        <dbl>  <dbl> <dbl> <chr>
## 1          1    328  5.58 "#727272\n"
## 2          4    475  6.36 "#7b7b7b\n"
## 3          6     69  5.71 "#737373\n"
## 4          8     18  5.12 "#6c6c6c\n"
## 5          8      9  3.72 "#545454\n"
## 6         10    480  8.92 "#949494\n"
```

Description of the dataset

Network of economic products, where a pair of products are connected if they are exported at similar rates by the same countries. The data are a projection from a bipartite network of nations and the products they export. Edges weights represent a similarity score (called "proximity"). Data based on UN Comtrade worldwide trade patterns. SITC network based on the Standard International Trade Classification.

Properties:

Weighted, Undirected

Graph:

```r
graph <- graph_from_data_frame(links, directed = FALSE, vertices = nodes)
graph
```

```
## IGRAPH 15427f0 UN-- 774 1779 --
## + attr: name (v/c), pid (v/n), community (v/n), size (v/n), pos (v/c),
## | leamer (v/n), color (v/c), _pos (v/c), width (e/n), color (e/c)
## + edges from 15427f0 (vertex names):
## [1] METAL FORMING MACHINE TOOLS                     --CONVERTERS,LADLES,INGOT MOULDS AND CASTING MACH
## [2] PRINTING PRESSES                                --OTHER MACH.-TOOLS FOR WORKING METAL OR MET.CARB
## [3] OTHER FOOD PROCESSING MACHINERY AND PARTS       --PARTS OF THE MACHINERY OF 744.2-
## [4] PRODUCER GAS AND WATER GAS GENERATORS AND PARTS--OTHER PUMPS FOR LIQUIDS & LIQUID ELEVATORS
## [5] PRODUCER GAS AND WATER GAS GENERATORS AND PARTS--CINEMATOGRAPHIC CAMERAS,PROJECTORS,SOUND-REC,PA
## + ... omitted several edges
```

Questions

1. What is the number of nodes and links?

```r
vcount(graph)
```

```
## [1] 774
```

```r
ecount(graph)
```

```
## [1] 1779
```

There are in total 774 nodes and 1779 links in this network.

2. What is the average degree in the network? And the standard deviation of the degree?

```r
mean(degree(graph))
```

```
## [1] 4.596899
```

```r
sd(degree(graph))
```

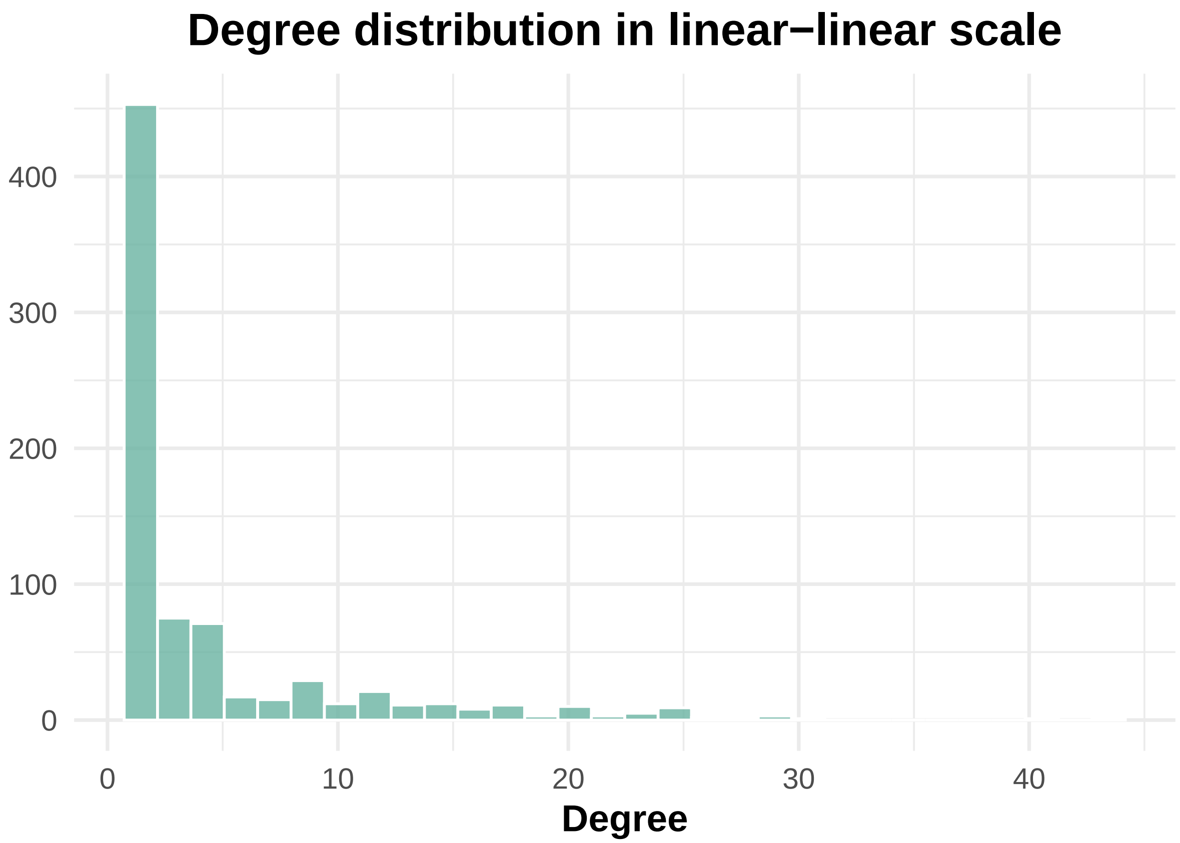
```
## [1] 5.994848
```

The average degree is 4.5969 in this network, with a standard deviation of 5.9948.

3. Plot the degree distribution in linear-linear scale and in log-log-scale. Does it have a typical connectivity? What is the degree of the most connected node?

```r
# In linear-linear scale
ggplot() +
  geom_histogram(aes(x = degree(graph)),
                 fill = "#69b3a2", color = "white", alpha = 0.8) +
  labs(x = "Degree", y = "", title = "Degree distribution in linear-linear scale")+
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    axis.title = element_text(face = "bold")
  )
```

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

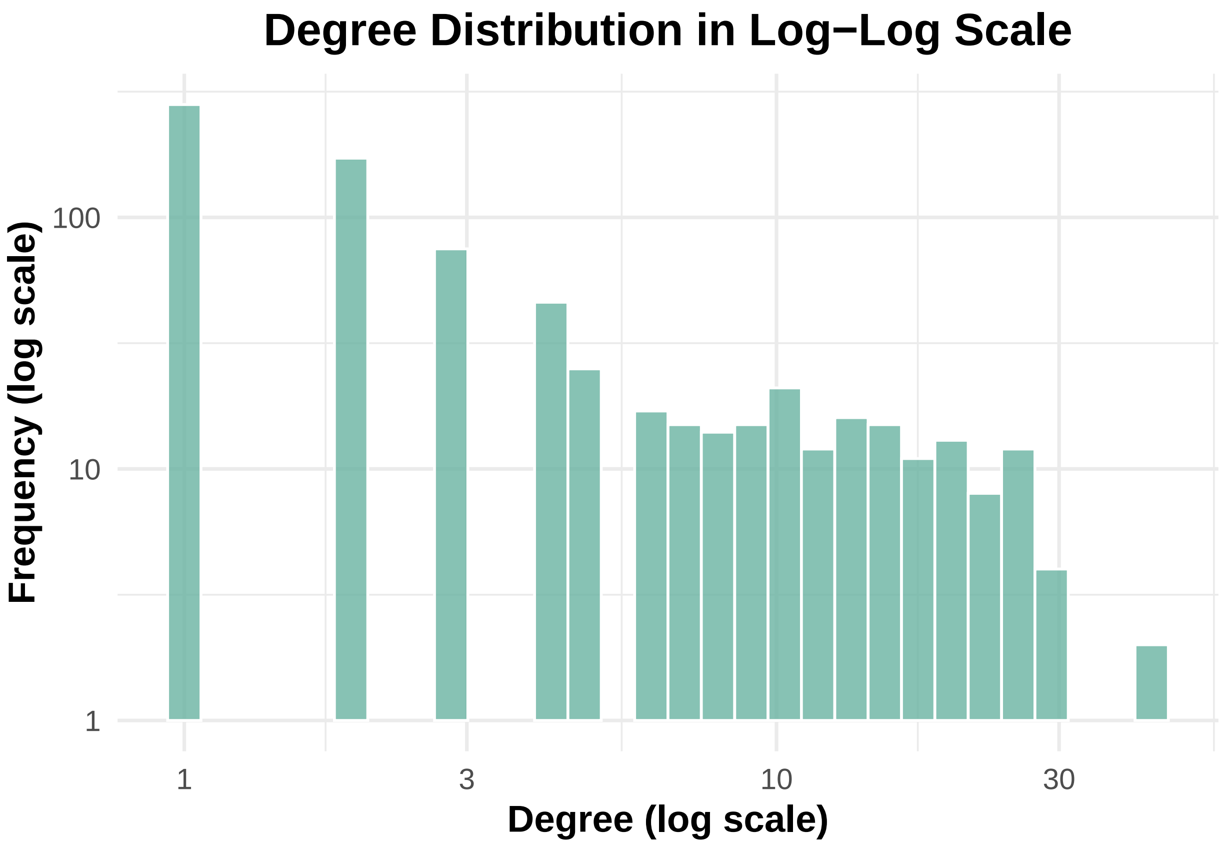


```r
# In log-log scale
deg <- degree(graph)
ggplot(data.frame(deg = deg), aes(x = deg)) +
  geom_histogram(fill = "#69b3a2", color = "white", alpha = 0.8) +
  scale_x_log10() +
  scale_y_log10() +
  labs(
    x = "Degree (log scale)",
    y = "Frequency (log scale)",
    title = "Degree Distribution in Log-Log Scale"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    axis.title = element_text(face = "bold")
  )
```

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

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
```

```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



```r
# Max_degree
max_degree(graph)
```

```
## [1] 43
```

We can observe that this network **does not exhibit typical connectivity**: Its degree distribution is highly skewed and lacks a clear peak. Most nodes have a very low degree, while a few have very high degree. In the log-log scale plot, we can observe a power-law-like distribution, which is not consistent with a Poisson-like distribution, where most nodes would have approximately the same number of links and no hubs.

The most connected node here has a degree of 43.

4. What is the clustering coefficient (transitivity) in the network?

```r
transitivity(graph)
```

```
## [1] 0.429691
```

The global transitivity of this network is 0.4297, which is closer to 0 than to 1, indicating a relatively low tendency of clustering.

5. What is the assortativity (degree) in the network?

```
assortativity_degree(graph)
```

```
## [1] 0.4571059
```

The assortativity coefficient of this network is 0.4571 (greater than 0), indicating a moderate to strong tendency for nodes to connect with others that have a similar degree. In other words, high-degree nodes tend to connect with other high-degree nodes, and low-degree nodes tend to connect with other low-degree nodes. This is a sign of assortative mixing.

6. Using the Louvain method, does the network have a community structure? If so, what is its modularity?

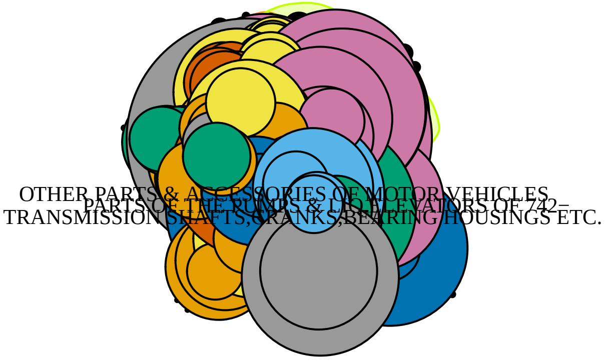
```
louvain_cluster <- cluster_louvain(graph, weights = E(graph)$width)

sizes(louvain_cluster)
```

```
## Community sizes
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20 
##  57 110  10  58 140  46  66  97  14   8  18   4  17  16   3  13  13   8  16  11 
##  21  22  23  24  25  26  27  28  29 
##   6   8   6   9   6   3   5   3   3
```

```
modularity(louvain_cluster)
```

```
## [1] 0.7407729
```

```
# Only display labels for nodes of a degree higher than or equal to 30
V(graph)$label <- ifelse(deg >= 30, V(graph)$name, NA)
# Generate random angles for lables to avoid overlapping
random_angles <- runif(length(V(graph)), 0, 2 * pi)
# Plotting
plot(louvain_cluster, graph,
     vertex.label = V(graph)$label,
     vertex.label.cex = 0.7,
     vertex.label.color = "black",
     vertex.label.dist = 2,
     vertex.label.degree = random_angles  )
```



Yes, the network has a clear community structure. Using the Louvain method, the network was partitioned into multiple communities, as the plot indicates. The modularity value is **0.7466**, which is considered high and indicates a strong modular (community) structure within the network.

7. Test that the clustering coefficient in the network cannot be statistically explain by a configuration model in which the nodes have the same degree distribution as the original.

```r
original_clustering <- transitivity(graph, type = "global")
```

```r
# Create 100 random configuration models and register the clustering coefficient
num_simulations <- 100
simulated_clustering <- numeric(num_simulations)
for (i in 1:num_simulations) {
  config_graph <- sample_degseq(degree(graph), method = "vl")
  simulated_clustering[i] <- transitivity(config_graph, type = "global")
}
```

Here we use the method “vl” (Viger–Latapy) instead of the traditional one, “configuration”, in order to avoid multi-edges and self-loops. As a consequence, we don’t need to add too many more checks in the loop like is.simple().

```r
# Evaluation and statistical test
# Calculate the p value clustering coefficient of the simulated network VS random configuration model
p_value <- mean(simulated_clustering >= original_clustering)

# Output
cat("Clustering coefficient of the original network: ", original_clustering, "\n")
```

```
## Clustering coefficient of the original network:  0.429691
```

```r
cat("Mean clustering coefficient of the configuration network: ", mean(simulated_clustering), "\n")
```

```
## Mean clustering coefficient of the configuration network:  0.03444034
```

```r
cat("p value:", p_value, "\n")
```

```
## p value: 0
```

We tested whether the clustering coefficient of the original network can be explained solely by its degree distribution, by comparing it to 1000 configuration model networks with the same degree sequence. The original network's clustering coefficient was **0.4297**, while the average clustering coefficient from the configuration models was **0.0342**. The p-value is **0**, indicating that none of the simulated networks reached the original clustering level.

Therefore, we reject the null hypothesis and conclude that the clustering structure in the original network **cannot be explained** by degree distribution alone — it has significant non-random structure.

8. Visualize the neighborhood of the node with the largest centrality (closeness)

```r
which.max(closeness(graph))
```

```
## SLAG WOOL.ROCK WOOL AND SIMILAR MINERAL WOOLS
##                                           453
```

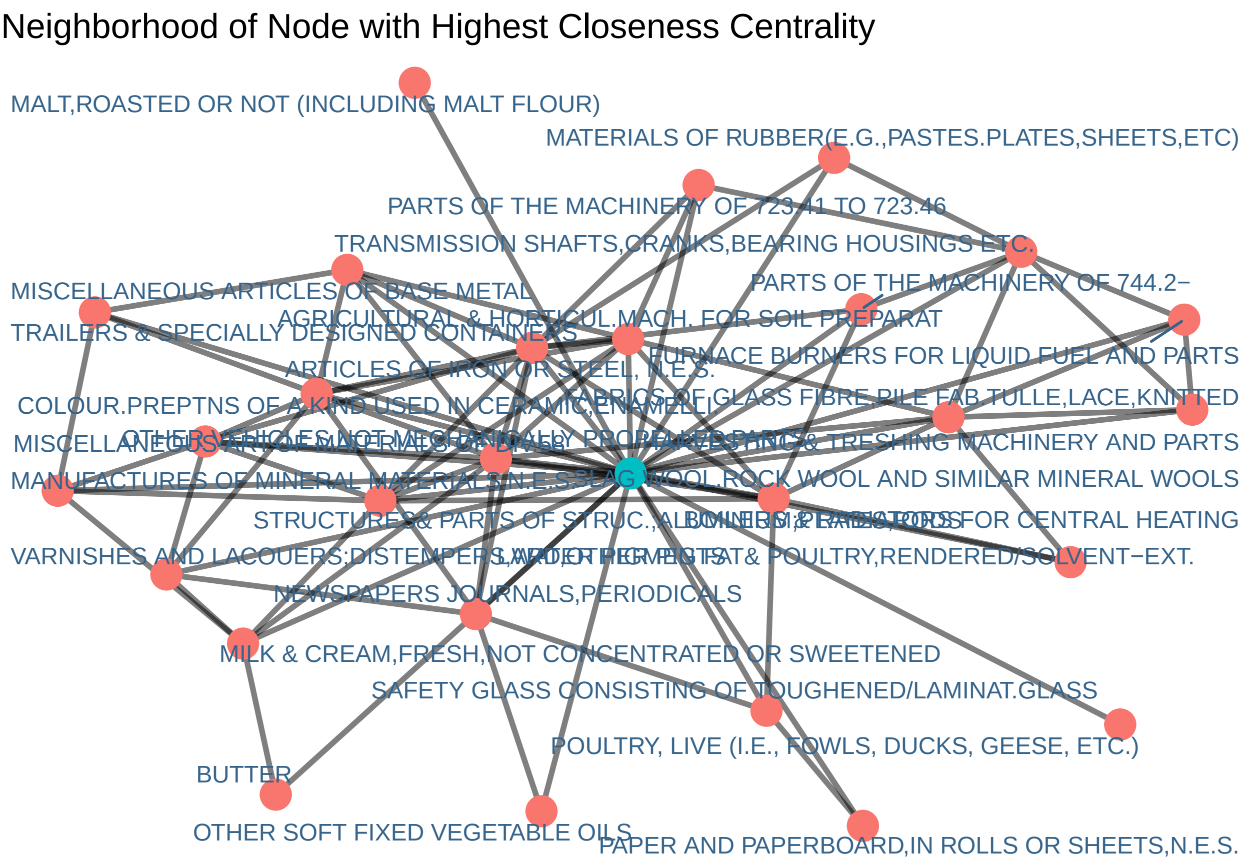
We discovered that the node with the largest centrality/closeness is "SLAG WOOL.ROCK WOOL AND SIMILAR MINERAL WOOLS".

```r
neighbors(graph,"SLAG WOOL.ROCK WOOL AND SIMILAR MINERAL WOOLS")
```

```
## + 27/774 vertices, named, from 15427f0:
##  [1] TRAILERS & SPECIALLY DESIGNED CONTAINERS
##  [2] PARTS OF THE MACHINERY OF 723.41 TO 723.46
##  [3] MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)
##  [4] OTHER VEHICLES,NOT MECHANICALLY PROPELLED,PARTS
##  [5] PARTS OF THE MACHINERY OF 744.2-
##  [6] MISCELLANEOUS ART.OF MATERIALS OF DIV.58
##  [7] POULTRY, LIVE (I.E., FOWLS, DUCKS, GEESE, ETC.)
##  [8] COLOUR.PREPTNS OF A KIND USED IN CERAMIC,ENAMELLI.
##  [9] VARNISHES AND LACOUERS;DISTEMPERS,WATER PIGMENTS
## [10] FABRICS OF GLASS FIBRE,PILE FAB.TULLE,LACE,KNITTED
## + ... omitted several vertices
```

```
neigh_graph <- make_neighborhood_graph(graph, order = 1, "SLAG WOOL.ROCK WOOL AND SIMILAR MINERAL WOOLS

# without the following plot won't work
V(neigh_graph)$color <- trimws(V(neigh_graph)$color)
E(neigh_graph)$color <- trimws(E(neigh_graph)$color)

plot(neigh_graph)
```

```
g <- as_tbl_graph(neigh_graph)

ggraph(g, layout = 'fr') +
  geom_edge_link(alpha = 0.5, width = 1) +
  geom_node_point(size = 5, aes(color = name == "SLAG WOOL.ROCK WOOL AND SIMILAR MINERAL WOOLS")) +
  geom_node_text(aes(label = name), repel = TRUE, size = 3, color = "steelblue4") +
  guides(color = "none") +
  theme_void()+
  labs(title = "Neighborhood of Node with Highest Closeness Centrality")
```

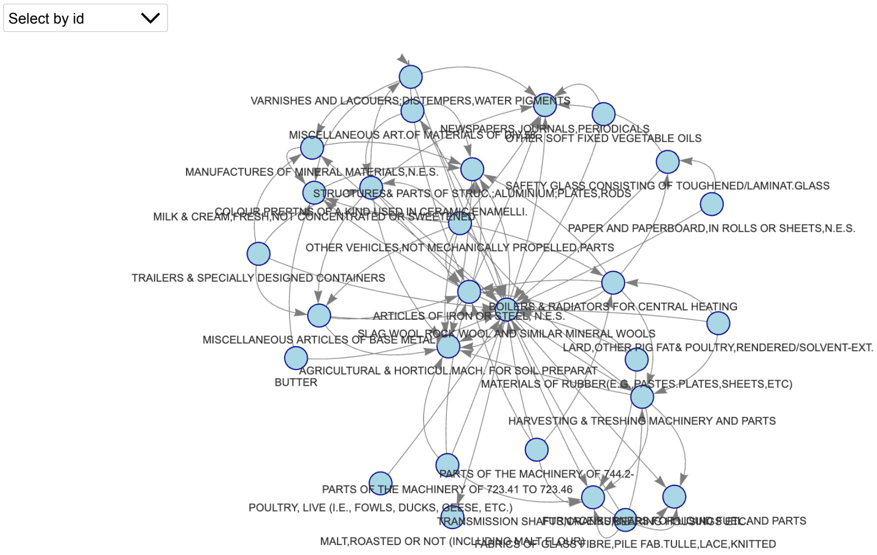

Neighborhood of Node with Highest Closeness Centrality

MALT,ROASTED OR NOT (INCLUDING MALT FLOUR)
MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)
PARTS OF THE MACHINERY OF 723.41 TO 723.46
TRANSMISSION SHAFTS,CRANKS,BEARING HOUSINGS ETC.
MISCELLANEOUS ARTICLES OF BASE METAL
PARTS OF THE MACHINERY OF 744.2−
AGRICULTURAL & HORTICUL.MACH. FOR SOIL PREPARAT
TRAILERS & SPECIALLY DESIGNED CONTAINERS
FURNACE BURNERS FOR LIQUID FUEL AND PARTS
ARTICLES OF IRON OR STEEL, N.E.S.
FABRICS OF GLASS FIBRE,PILE FAB.TULLE,LACE,KNITTED
COLOUR.PREPTNS OF A KIND USED IN CERAMIC,ENAMELLI.
OTHER VEHICLES,NOT MECHANICALLY PROPELLED;PARTS
MISCELLANEOUS ARTIC.OF MATERIALS OF DIV.58
HARVESTING & TRESHING MACHINERY AND PARTS
MANUFACTURES OF MINERAL MATERIALS,N.E.S.
SLAG WOOL,ROCK WOOL AND SIMILAR MINERAL WOOLS
STRUCTURES& PARTS OF STRUC.,ALUMINIUM;PLATES,RODS
BOILERS & RADIATORS FOR CENTRAL HEATING
VARNISHES AND LACOUERS;DISTEMPERS,WATER PIGMENTS
LARD;OTHER PIG FAT& POULTRY,RENDERED/SOLVENT−EXT.
NEWSPAPERS JOURNALS,PERIODICALS
MILK & CREAM,FRESH,NOT CONCENTRATED OR SWEETENED
SAFETY GLASS CONSISTING OF TOUGHENED/LAMINAT.GLASS
POULTRY, LIVE (I.E., FOWLS, DUCKS, GEESE, ETC.)
BUTTER
OTHER SOFT FIXED VEGETABLE OILS
PAPER AND PAPERBOARD,IN ROLLS OR SHEETS,N.E.S.

We also used the "visNetwork" package to make an interactive graph:

```r
library(visNetwork)
nodes <- data.frame(id = V(g)$name, label = V(g)$name)
edges <- data.frame(from = as.character(ends(g, E(g))[,1]), to = as.character(ends(g, E(g))[,2]))

visNetwork(nodes, edges) %>%
  visEdges(arrows = 'to', color = list(color = "gray", hover = "red")) %>%
  visNodes(size = 15, color = list(background = "lightblue", border = "darkblue")) %>%
  visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE) %>%
  visLayout(randomSeed = 123)
```

```
## file:////private/var/folders/5k/zh67w_yx2fj10t7z34l2ct7m0000gn/T/Rtmp7TOv65/file3bdf3f1d26e4/widget3
```


Select by id
VARNISHES AND LACOUERS;DISTEMPERS,WATER PIGMENTS
NEWSPAPERS,JOURNALS,PERIODICALS
MISCELLANEOUS ART.OF MATERIALS OF DIV.58
OTHER SOFT FIXED VEGETABLE OILS
MANUFACTURES OF MINERAL MATERIALS,N.E.S.
SAFETY GLASS CONSISTING OF TOUGHENED/LAMINAT.GLASS
STRUCTURES& PARTS OF STRUC.;ALUMINIUM;PLATES,RODS
COLOUR PREPTNS OF A KIND USED IN CERAMIC,ENAMELLI.
MILK & CREAM,FRESH,NOT CONCENTRATED OR SWEETENED
PAPER AND PAPERBOARD,IN ROLLS OR SHEETS,N.E.S.
OTHER VEHICLES,NOT MECHANICALLY PROPELLED,PARTS
TRAILERS & SPECIALLY DESIGNED CONTAINERS
BOILERS & RADIATORS FOR CENTRAL HEATING
ARTICLES OF IRON OR STEEL, N.E.S.
SLAG WOOL,ROCK WOOL AND SIMILAR MINERAL WOOLS
MISCELLANEOUS ARTICLES OF BASE METAL
LARD,OTHER PIG FAT& POULTRY,RENDERED/SOLVENT-EXT.
AGRICULTURAL & HORTICUL.MACH. FOR SOIL PREPARAT
BUTTER
MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)
HARVESTING & TRESHING MACHINERY AND PARTS
PARTS OF THE MACHINERY OF 744.2-
PARTS OF THE MACHINERY OF 723.41 TO 723.46
POULTRY, LIVE (I.E., FOWLS, DUCKS, GEESE, ETC.)
TRANSMISSION SHAFTS,CRANKS,BEARING HOUSINGS ETC.
FURNACE BURNERS FOR LIQUID FUEL AND PARTS
MALT,ROASTED OR NOT (INCLUDING MALT FLOUR)
FABRICS OF GLASS FIBRE,PILE FAB.TULLE,LACE,KNITTED