

|  |
| --- |
| Progress Report |
|  |
| October 6  MXB362 – Advanced Visualization and Data Science  Student: Viet Hoang Do - n10329935 |

# Investigating the Correlation Network in Foreign Currency Exchange Market

## Summary

The foreign currency exchange (FOREX) market is the biggest financial market in the world. It exerts an enormous impact on all other markets (Ismaila and Dallah, 2010). It is because the currency presents the price of any asset and its value represents the country's economic status. A foreign exchange rate is a relative price of a region's currency in terms of the currency of another region. It can express the economic balance between two regions (Mizuno et al., 2006). The FOREX market is a typical complex system and can be presented by the correlation networks and minimum spanning tree (MST) maps. The measures of centrality (degree, closeness, eigenvector, and betweenness) are all used to describe the characteristics of the FOREX market (Kazemilari and Djauhari, 2013). They can illustrate the significance of each currency inside the FOREX market network by investigating the components and interrelationships.

The project will study a set of currencies rates during the COVID-19 period from 30 December 2019, until 29 June 2021. The expected audience of visualizations will be investors and companies who want to develop the new investment strategy during the pandemic period. There are many benefits for companies and investors when important currencies are identified. The influential currencies tend to exert an impact on others. The increase or decrease in the foreign exchange rate of influential currencies would lead to a change in the trend of related currencies. There are many investors using key currencies as an indicator to take advantage of market opportunities. Those currencies are monitored carefully. For example, the information about events and policies which could affect the trend of influential currencies will be the first priority for the investors to focus on.

The measures of centrality will explore the differences between the behavior of currencies and determine the most influential currencies. The analysis of the Foreign Currency Exchange Network aims to reveal the information about market's structure and the role of each particular currency. The project centers on quantifying the currencies' importance in the global financial market. To achieve the aim, some currency correlation networks and their minimum spanning tree which contain general currencies information will be constructed.

**Aim**: Better understand the relationship of currencies and find the most influential currencies.

## Project Description

### The Data

The daily closing prices of 58 currencies is retrieved from Exchange Rate Service (<http://fx.sauder.ubc.ca/>). A table is constructed to display the countries' name and their respective currency symbol. The project uses US Dollars as the base currency to state the daily exchange rate.

### Correlation Heatmap

The heatmap is a conventional method to illustrate the correlations between sectors [[Figure 1](#_Figure_1:_Illustration)]. Constructing the heatmap of the correlation matrix is necessary. It helps to validate the results from the correlation calculations and provides a high-level understanding of relationships between currencies.

### Currency network

Although heatmap is helpful, it can only present one dimension of data (the correlation between two currencies). As the outlined purpose of the project, even with a heatmap, crucial issues about the most influential currencies and their behavior remain unanswered. The currency network will be utilized to study further based on initial findings achieved from the correlation heatmap [[Figure 2](#_Figure_2:_Illustration)]. The visualization of the currency network also provides a more accessible method to convey meaningful messages.

### Minimum Spanning Tree with different bases

The minimum spanning tree is widely used to visualize the financial networks [[Figure 3](#_Figure_3:_Illustration)]. While the number of edges displayed in the currency network can be up to N(N – 1) (with N is the number of currencies), the minimum spanning tree only has (N-1) edges. It contains the edges that link all the nodes together, minimize the sum of edge weights, and do not include any cycles. The project will implement Kruskal's algorithm in building the minimum spanning tree, which would aid to understand, simplify and summarize the message obtained from the currency network.

### Expected Output

For all three visualizations, I expect to explore the capabilities of Python Libraries such as Numpy, Pandas, Networkx, Matplotlib, Seaborn to pre-process data and construct basic visualizations. I also expect to implement network drawing software Gephi in mapping the currency network in geographical layout. The custom color palette, the scale, the size of each element, and the layout of visualization will be studied carefully to display the information in the most efficient way.

The output obtained from analyzing the visualizations is expected to be particularly helpful in some real-world use cases. For example, the analysis of the correlation between currencies would help investors diversify their portfolios since it helps to identify uncorrelated currencies to invest in. Another example is that by pointing out the most influential currencies, it would be the suggestion for investors to focus more on these currencies.

## Proposed Task and Project Timeline

### The Gantt Chart

The list of tasks expected to be accomplished is displayed in the Gantt Chart [[Figure 5](#_Figure_5:_Gantt)].

### Expected outcomes by Week 11

By the end of Week 10, I expect to generate the primitive version of three visualizations. The table showing the statistic about currency network and minimum spanning tree will also be done. A brief analysis and example about practical use case of each visualization will be outlined.

### Progress to date

The project journal can be seen in the [Figure 4](#_Figure_4:_Project).

1. Background research, reading, and investigations carried out

* Many studies that proposed methods to analyze the correlation network of the world currency exchange rate. Kazemi Lari and Djauhari (2013) examined the network topology of the Forex market from 2009 to 2012 based on centrality measures. Their paper successfully pointed out nine influential currencies during this period and suggested investors and companies should monitor these currencies carefully. Kwapień et al. (2009) provide the analysis of a network structure of the foreign currency exchange market. This paper did identify the clusters of strongly connected currencies and study the network topology of minimum spanning trees constructed by different base currencies such as GPB or USD. The analysis indicates that EUR has played a crucial role in the global Forex market.
* There is an interesting reading that discusses visualizing asset price correlation written by Julian West (2019). In this article, the author showed how to visualize the correlation matrix of asset price as an interactive network graph that provides the audience with a high-level overview of the connection between asset classes.

1. Collected and generated data

* A CSV file named ‘country-currency-region.csv’ contains information about the symbol of currencies and their corresponded country name and region.
* Two CSV files ‘nodes\_network.csv’ and ‘edges\_network.csv’ contain node and edge data for network visualization using Gephi software. Those files are generated by MatLab script in ‘data\_process.mlx’ file.
* A CSV file named ‘fxdata.csv’ contains the daily closing prices of 58 currencies retrieved from Exchange Rate Service (<http://fx.sauder.ubc.ca/>)
* A table shows information about countries, respective symbols and centrality measures [[Figure 17](#_Figure_17:_Countries,)]
* A CSV file named ‘pythonData.csv’ contains the cleaned data from ‘fxdata.csv’. It is ready to be used and processed as a data frame by Pandas library.

1. Algorithm and visualization techniques

* Before constructing the heatmap, normalizing data and converting the absolute asset values into log-returns are of importance. This transformation is prevalent in financial time-series data since investors are more concerned with asset returns than their absolute values. Moreover, it is easier to compare the returns of two currencies when the data is normalized. First, we compute the daily returns of the logarithm rate: where (number of currencies), is the daily return at time t of currency and is the closing price of currency at time .
* After the closing prices are converted into log returns, the *corr*() function is used to compute the pairwise correlations between currencies. This function is available in both Python and MatLab, and it will return an NxN correlation matrix. Python provides a data visualization library called Seaborn, which visualizes the correlation matrix as a clustered heatmap via a useful function called *clustermap*(). The clustered heatmap is a conventional approach to illustrate correlations between variables in a dataset, especially if the data is multidimensional because it automatically organizes similar variables into clusters. This improves the structure and readability of the heatmap, making it simpler to notice connections between currencies that act similarly. The result of the visualization is shown in [Figure 10](#_Figure_10:_Clustered).
* The Marchenko–Pastur distribution is implemented to detect and remove insignificant eigenvalues and eigenvectors of the matrix, reduce noise and make the matrix display only meaningful messages. The next step is converting the cleaned correlation matrix into the lists of nodes and edges in order to visualize it as a network. In MatLab, this step is done by using the OutputNetwork.p function, which takes the correlation matrix and an array of currencies' labels as parameters and returns two CSV files: nodes.csv and edges.csv. Then, those files are loaded into Gephi to generate network visualization. To improve the visualization, insignificant edges are removed if their correlation value is below a 0.3 threshold value. Gephi uses the Louvain method to compute the modularity and detect community structure in networks. It successfully detects three major communities in the currency network as shown in [Figure 11](#_Figure_11:_Currency).
* To produce the minimum spanning tree (MST), a technique of MST introduced by Mizuno et al. (2006) is implemented. The correlation matrix is transformed to the distance matrix as follows: The three axioms of Euclidean distance are satisfied: (i) , (ii) , (iii) , where is the distance between the rate of currency . Using the distance matrix as a parameter, the OutputNetwork.p function outputs nodes and edges CSV files. Those files are imported to Gephi to generate MST visualization. In Gephi, a plugin called Minimum Spanning Tree implements Kruskal's algorithm to construct the FX network.
* Since centrality is one of the most researched topics in network analysis, a variety of measurements have been proposed such as closeness centrality, betweenness centrality, degree centrality, eigenvector centrality, flow betweenness, the rush index, etc. In this project, the network centrality was examined by evaluating four particular node-level centrality metrics that are essential in most network studies (Kazemi Lari and Djauhari, 2013). Degree, closeness, betweenness, and eigenvector centralities are calculated for each node in a network.
  + Degree Centrality shows the degree of importance of information for each currency (Niemincn, 1974). It is calculated as follows:
  + Betweenness Centrality illustrates the frequency for each currency in the shortest routes between indirectly linked nodes (Freeman, 1977). It is calculated as follows: Where is the total number of shortest paths from *s* to *t*, and is the number of those paths that pass through node *v*.
  + Closeness Centrality: measures the time to transfer information between a node *v* to each other reachable node. The higher the closeness centrality score is, the more important the currency is (Sabidussi, 1966). It is calculated as follows: Where *n* is the size of the connected component that node *v* can reach
  + Eigenvector Centrality identifies which currency is connected to the most connected nodes (Banocicha, 2007). The eigenvector centrality of node *i* is calculated as follows:

1. Visualization analysis
   * The clustered heatmap [[Figure 10](#_Figure_10:_Clustered)]
     + A diverging colorblind-friendly color palette is used to colorize the heatmap. The positive correlations are green, uncorrelated currencies are white and negative correlations are brown. Looking at the clustered heatmap visualization, there are interesting insights the audience could gain.
     + Overall, the strongly negative correlations are hardly seen between currencies. The currencies that are close to each other according to geographic regions tend to be grouped in the same cluster. For example, it is shown in the visualization that the list of currencies run from CHF (Switzerland) to PLN (Poland) are mostly European currencies. CAD (Canada), GPB (United Kingdom), SGD (Singapore), AUD (Australia), and NZD (New Zealand) are grouped into a cluster. It could be because they are Commonwealth countries that have political and cultural backgrounds in common. There is also a cluster of non-European regions, which consists of the list of currencies run from IDR (Indonesia) to ZAR (Africa). Most of those currencies come from the Asian region. It is also noticeable that HKD (Hong Kong) and JPY (Japan) are not correlated with Asian currencies; otherwise, they are correlated with European currencies. The event that the CNY (China) and JPY (Japan) joined XDR (the special draw rights) could be the reason for that. SAR (Saudi Arabia), ARS (Argentina), and PKR (Pakistan) are three currencies that are uncorrelated with all others.
   * The currency network [[Figure 11](#_Figure_11:_Currency)]
     + A qualitative color palette is used to colorize the nodes based on the communities detected in the currency network. The network visualization provides the audience with a better picture of the data. By removing insignificant edges with correlation values smaller than 0.3, the graph only displays the meaningful correlations between currencies. The edge thickness is scaled based on the magnitude of correlation. The Louvain method successfully detects which groups of currencies behave similarly. There are four clusters that can be seen from the graph: orange, green, blue, and grey. It is noticeable that all of the nodes in the orange cluster (except JPY) come from the European region. The European currencies are strongly correlated to each other and have a significant correlation to other currencies. Looking at the green cluster, there are four currencies namely COP, PEN, BRL, and CLP located in the left part of the cluster. All of them are from South/Latin America and they only connect to the currencies within the same cluster. Therefore, it is true to say that there is no correlation in the currency exchange rate between South/Latin America and Europe. It is also noteworthy that even TRY and RUB come from Europe, they have no correlation with European currencies. It might be because of their unique geographic position, both Turkey and Russia are lying partly in Asia and partly in Europe. Regarding the blue cluster, it contains 12 currencies (9 of them are in Europe and 3 of them are in Asia). It is shown that European currencies are grouped into two different clusters. Moreover, even though MYR, PHP, and THB are from South East Asia, they tend to be more correlated with European currencies.
   * The minimum spanning tree [[Figure 12](#_Figure_12:_Minimum), [Figure 13](#_Figure_13:_MST), [Figure 14](#_Figure_14:_MST), [Figure 15](#_Figure_15:_MST), [Figure 16](#_Figure_16:_MST)]
     + A qualitative color palette is used to colorize the nodes based on their region [[Figure 12](#_Figure_12:_Minimum)] (i.e., Europe, orange; Asia & Pacific, dark green; South/Latin America, blue; Africa, pink; North America, light green; Middle East, yellow). The minimum spanning tree is a more readable and well-structured version of the previous network graph. There is some knowledge obtained from the minimum spanning tree. The NOK is at the center of the FX network since it connects currencies from four different regions (i.e., Asia, South/Latin America, Middle East, and Europe). The majority of the currencies are grouped together based on geographical regions. The currencies of Commonwealth countries are connected together such as MYR, CAD, AUD, NZD, and SGD. The currencies of Europe are the most closely linked. EUR has a predominant position in the European monetary system. JPY is connected to European currencies, whereas CNY is more attached to the Asian cluster despite the fact that both currencies join the special drawing rights.
2. Interactive visualization
   * The code for the website can be access via my GitHub: <https://github.com/hoangqwe159/forex-analysis-covid19>
   * The live demo of the website can be access via the link: <https://forex-network.herokuapp.com>
   * The website allows the audience to interact with visualization such as hovering for additional information, highlight the connected nodes, drag, and zoom the network. The application is only displaying the draft version of the visualizations. An update version will be deployed as the output of the project

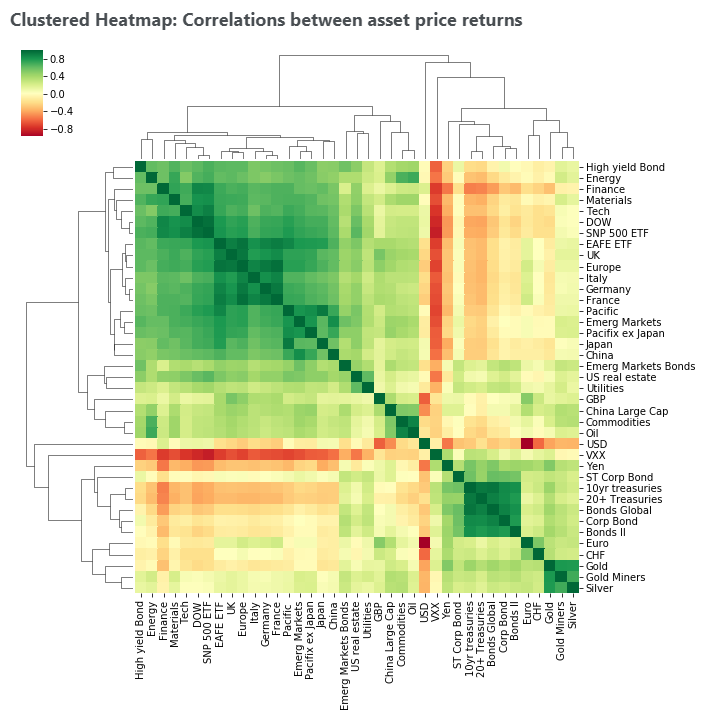
## Foreseen problems and risks

One of the most major problems of the project is choosing the optimal tools to prepare and process data. While Python provides a wide range of professional libraries and is supported by the larger community, MATLAB is the language I am more familiar with. There is a risk that some problems I might not solve by using Python, then the integration between two technologies is necessary for me to complete the project. Another problem is that mapping currencies node into their nation on the geographical map can only be done manually if I utilize Gephi Software.

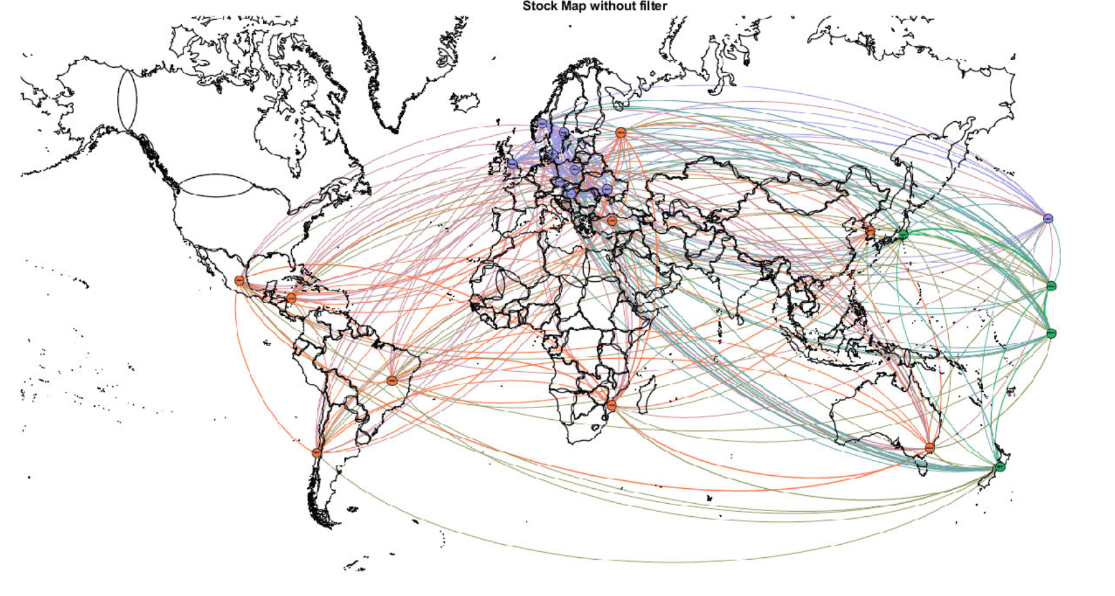
## Reference

1. Ismaila, A., Dallah, H. (2010). Time Series Forecasting of Naira against Major Currencies Exchange Rates. International Journal of Applied Mathematics and Statistics (IJAMAS), vol.19, pp.105-111.
2. Mizuno, T., Takayasu, H., Takayasu, M. (2006). Correlation networks among currencies. Physica A, 336-364.
3. Yang, X., Wen, S., Liu, Z., Li, C., & Huang, C. (2019). Dynamic properties of foreign exchange complex network. Mathematics, 7(9), 832.
4. Kazemilari, M., & Djauhari, M. A. (2013). Analysis of a correlation network in world currency exchange market. Int. J. of Applied Mathematics and Statistics, 44(14), 202-209.
5. Julian West. (2019). Visualizing asset price correlations. Retrieved from: <https://julian-west.github.io/blog/visualising-asset-price-correlations/>
6. Kwapień, Jarosław & Gworek, Sylwia & Drozdz, Stanislaw & Gorski, Andrzej. (2009). Analysis of a network structure of the foreign currency exchange market. Journal of Economic Interaction and Coordination. 4. 10.1007/s11403-009-0047-9.
7. Sabidussi, G. (1966). The centrality index of a graph. Psychometrika, 31(4), 581–603.
8. Freeman, L.C. (1977). A set of measures of centrality based on betweenness. Sociometry, 40(1), 35-41.
9. Niemincn, J. (1974). On centrality in a graph. Scandinavian Journal of Psychology 15:322-336.
10. Banocicha, P. (2007). Some unique properties of eigenvector centrality. Social Networks, 29: 555-564.

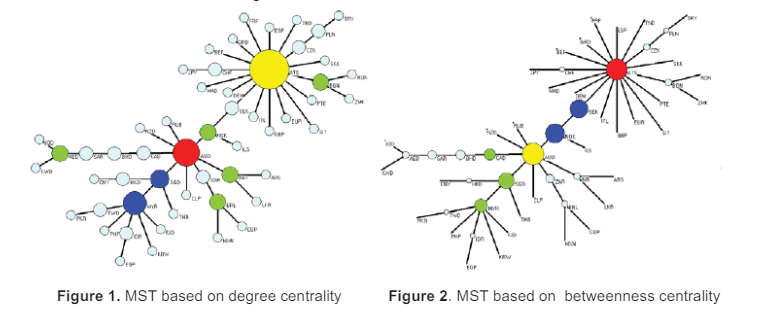
## Appendix



### Figure 1: Illustration of the Heatmap Correlation Matrix (Julian, 2019)



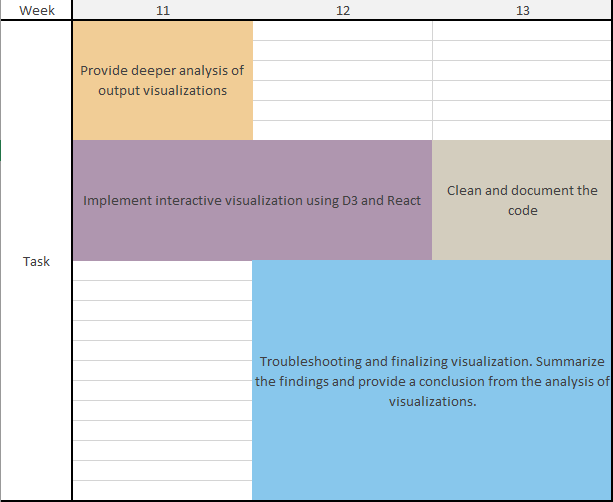
### Figure 2: Illustration of the Currency Network using Geographical Layout



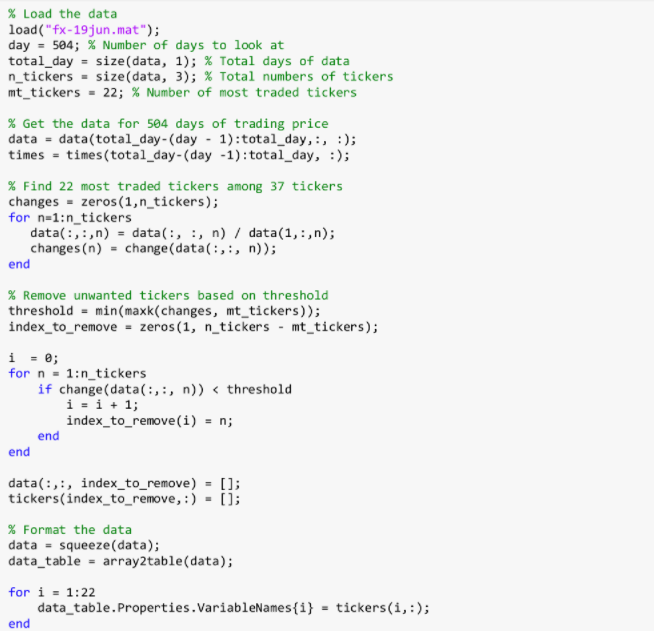
### Figure 3: Illustration of Minimum Spanning Tree with different bases (Kazemilari and Djauhari, 2013).

### 

### Figure 4: Project Journal to 02/10/2021

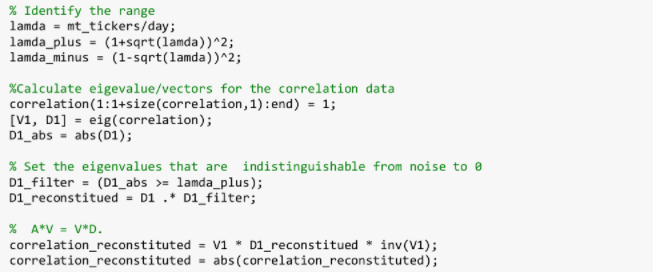


### Figure 5: Gantt Chart

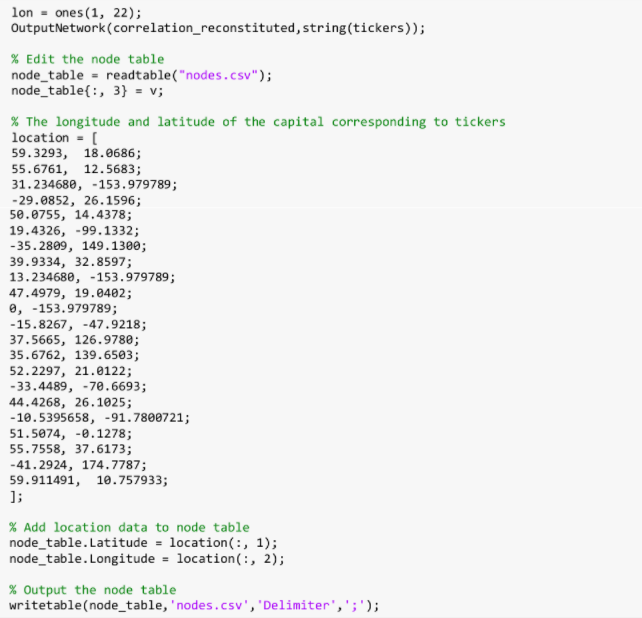


### Figure 6: Load and pre-process data

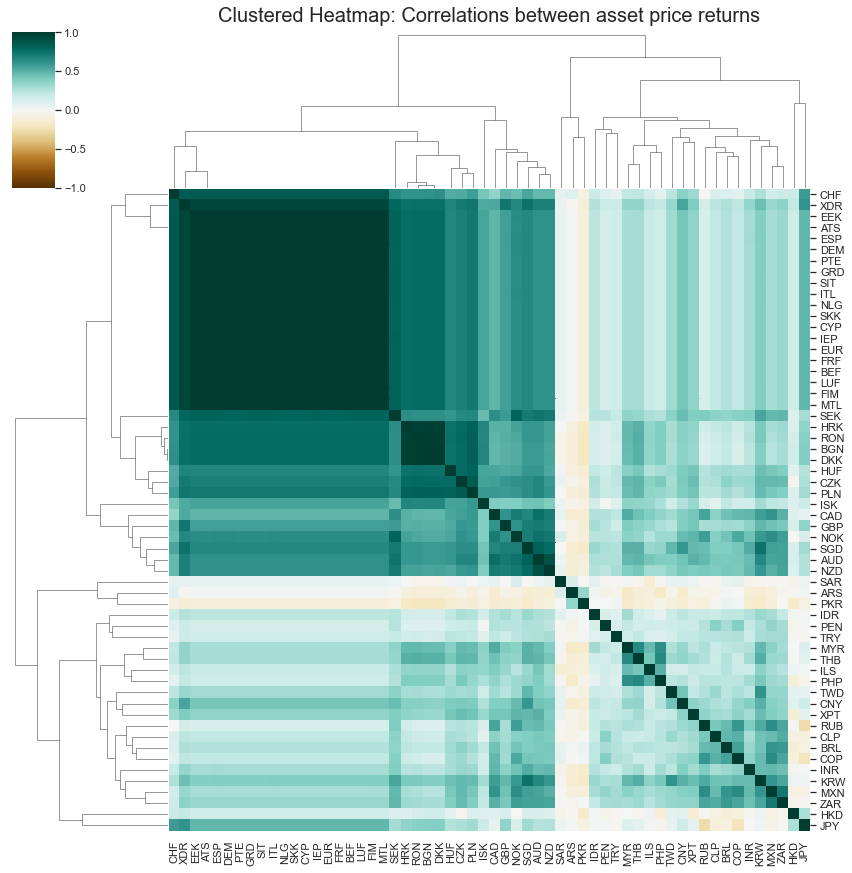
### Figure 7: Compute Daily return and Correlation matrix



### Figure 8: Reduce noise from Correlation Matrix



### Figure 9: Output nodes and edges file for Gephi



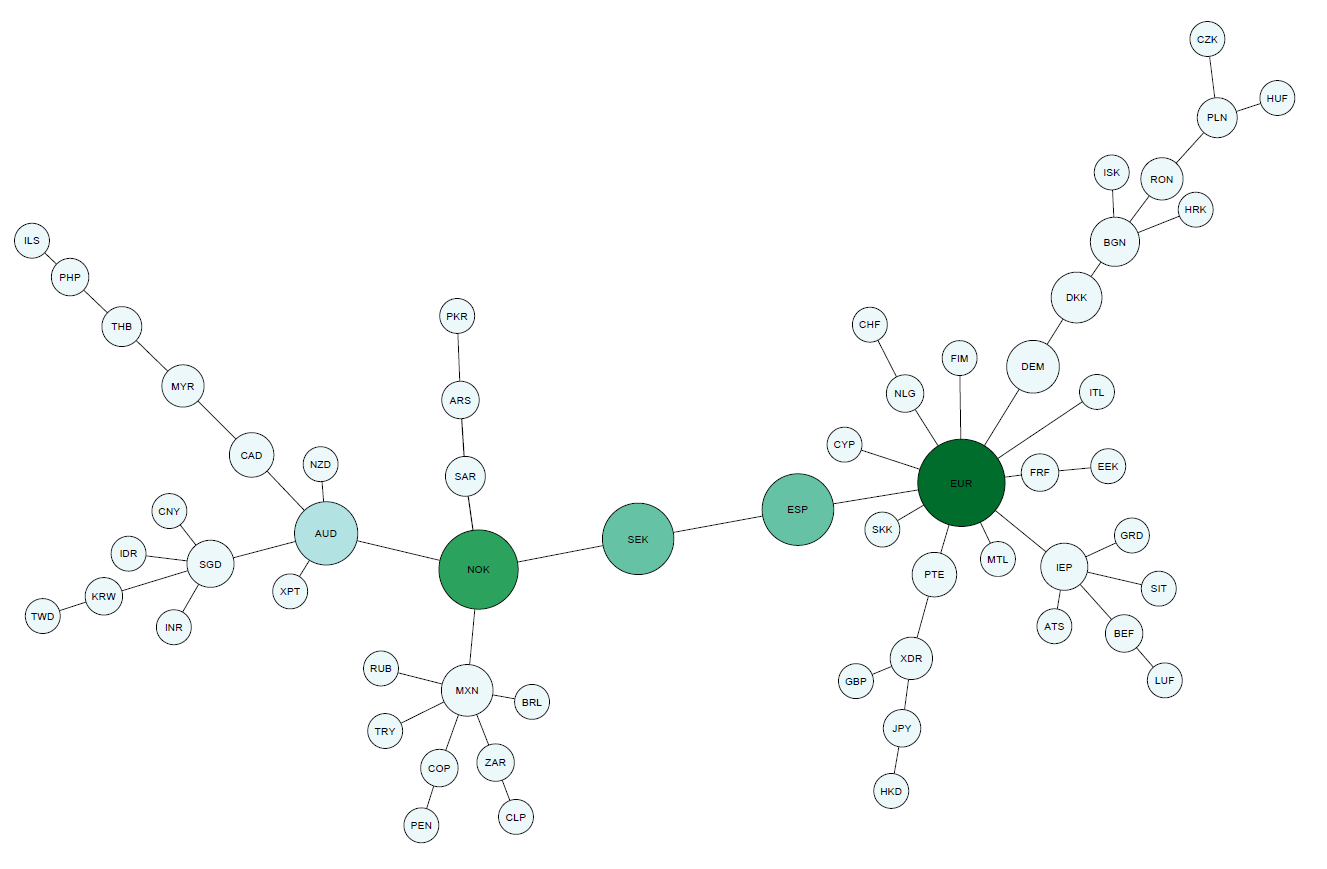
### Figure 10: Clustered Heatmap

### 

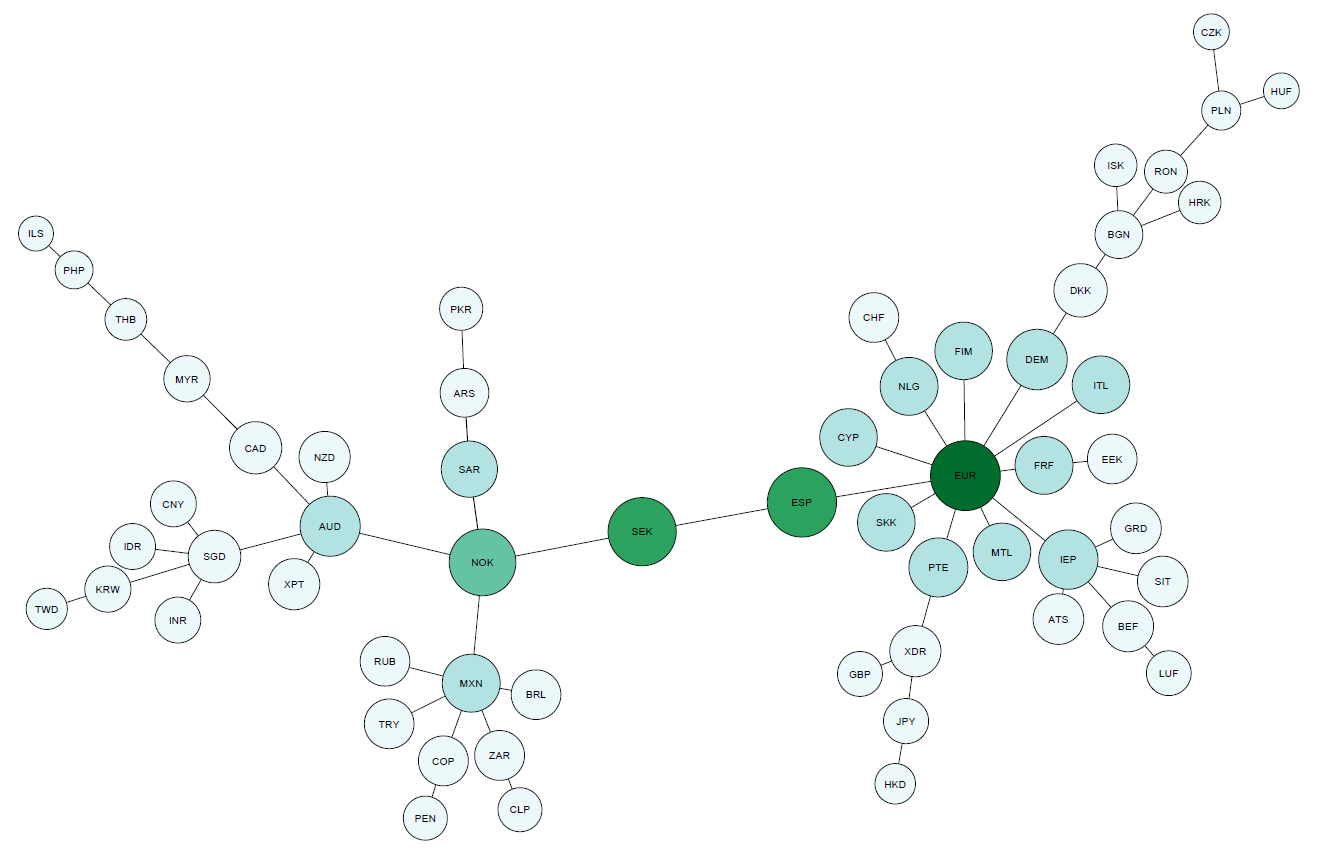
### Figure 11: Currency network. The currencies from the same cluster have same color.

### 

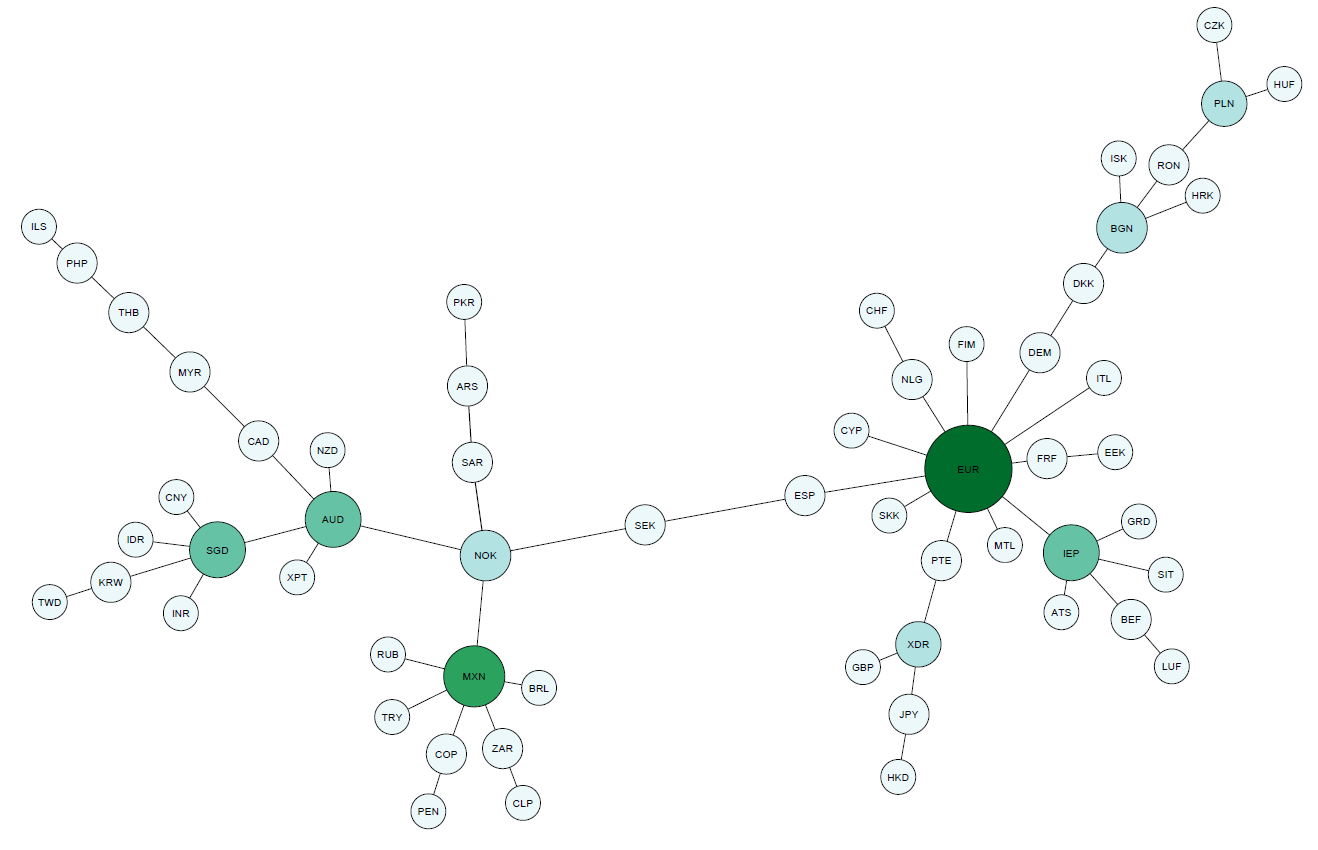
### Figure 12: Minimum Spanning Tree. The currencies from the same region have same color. (i.e., Europe, orange; Asia & Pacific, dark green; South/Latin America, blue; Africa, pink; North America, light green; Middle East, yellow)



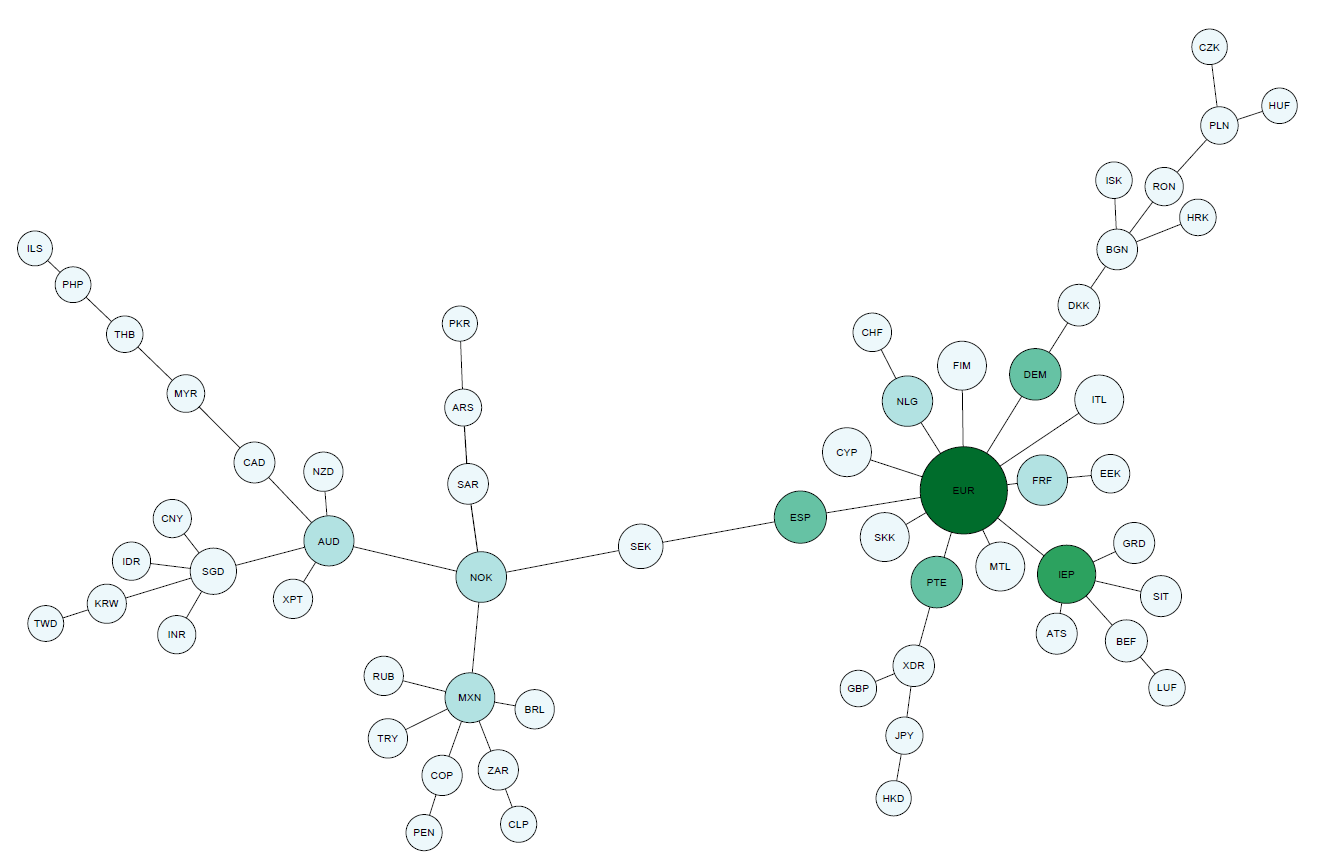
### Figure 13: MST based on betweenness centrality



### Figure 14: MST based on closeness centrality



### Figure 15: MST based on degree centrality



### Figure 16: MST based on eigenvector centrality



### Figure 17: Countries, respective symbols and centrality measures