DATA SCIENCE MASTERS : MAJOR RESEARCH PROJECT (MRP)

Graph Perspective of Stock Market Behaviour

Somnath Kundu
Supervisor : Dr. Pawel Pralat

June 16, 2017

1 PROBLEM DESCRIPTION

It is often noticed in practice that the prices of different Stocks and other financial investment instruments move together. From economic and investment theoretical point of view this co-movements are very much logical. As many different companies engage in similar type of business. Also many a time one business depends on another business. So those companies are tied together by an invisible thread of relationship. Though it is difficult to find the actual relationship of the companies we can always measure the strength of their relationship by the similarity of movement of their attributes.

Correlation Coefficient is one of the measures of the strength of the relationship between two variables. We can assume if the correlation coefficient of one or more attributes of two Stocks or Financial instruments is more than a threshold value then those two stocks or the instruments are connected or we can say an Edge exists between them.

In this project our objective is to explore these relationship of the stocks from the Graph theory perspective and try to investigate different properties of those graphs.

1.1 PROBLEM DOMAIN & BASIC DATA REQUIREMENT

There are many different Financial Instruments which people invested on. Some of those instruments are formally traded in a more structured way. In Financial terms they are collectively called "Securities". The most frequently used Securities are as follows:-

1. Stocks (Equities) [Ex : IBM:NYSE, MSFT:NASDAQ, etc]
2. Published Index [DOW Jones Industrial Average, S&P 500]
3. ETF (Exchange Traded Fund)
4. Commodities (Cash & Future)
5. Bonds
6. Money-Market
7. Currencies

There are also many other similar entities which people invest on. ex.Real Estate, Private Funds, etc.

For this current work, We planned to confine our studies only within the domain of Stocks and Published Index. Though our studies can be very well used for any of the financial instruments or any time-series in general. Even within the domain of stocks there are about 100,000 (10^5) stocks worldwide And there are about frequently used 5000 Index (Indices) worldwide. With about 40 to 50 years of Data (about 250 Trading Days per year) these could generate a huge data volume. If we take one record per day per company, then approximately the total number of records would be around 40 * 250 * 10^5 or 10^9 (1 Billion Record)

Due to time and resource constraints we would start with about **30 different stocks of current Dow Jones Index for on an average 20 years of Data** [We took one extra so ours is 31 companies]. That is about 250 * 20 * 31 or about **155,000 records**. The list of the companies studied are listed in the appendix.

We are also planning to see the relationship between major indices around the world. The Indices depicts the overall market sentiment. So that would also give us an idea of the relationships at the macro level.

Different attributes of the Stocks can be studied. But we are confining our study **only based on the movement of prices of the stocks**. We are not going to use any Financial or Fundamental data. So our study will under the domain of "Technical Analysis" [In stock market any study which only uses the price data are collectively called "Technical Analysis"]

1.2 Data Collection & Data Exploration

For Stock Market Price, Data are very well defined. And the data comes more or less in the same format. So the Data exploration becomes very simple for this case.

There are two types of Price data that can be used for our study. One is Intra-day price and another is Everyday Price. Intra-day price is too volatile and it doesn't necessarily relates to the longterm behaviour of the stocks. So we are using the Daily (End of the Day) price for stocks. The End of Day Price consists of the following Data Elements:-

1. Date [Date of the Price]
2. Open Price [Starting Price of the Day]
3. Close Price [Ending Price of the Day]
4. High Price [Maximum Price of the Day]
5. Low Price [Ending Price of the Day]
6. Volume [Total No. of Shares that is traded during the day]

Out of that **Close Price** is the most important data elements of it.

Some Data Source Providers provide Weekly (End-Of-the Week) or Monthly (End-Of-the Month) Price Data.But we decided to use the End of the Day price data and smooth the data for different periodicity using different Smoothing Period and Smoothing Operators (Like different types of Moving Average)

We used **Yahoo Finance** to download the Historical Price Data from the Internet. This is free and readily available. Some other Free Data Providers for Stock Prices are Google Finance and MSN money. The Web link structure for downloading the Price data for a stock from Yahoo Finance Web Site is as follows:-

$$\begin{aligned}
&http://ichart.finance.yahoo.com/table.csv\\
&?s = \textbf{[Ticker]}\\
&\&d = \textbf{[Start Month-1]}\\
&\&e = \textbf{[Start Day]}\\
&\&f = \textbf{[Start Year]}\\
&\&g = d\\
&\&a = \textbf{[End Month-1]}\\
&\&b = \textbf{[End Day]}\\
&\&c = \textbf{[End Year]}\\
&\&ignore = .csv
\end{aligned} \tag{1.1}$$

The Price Data comes as a CSV (Comma Separated) file.

There are two important features that need to be mentioned here.Firstly the data are given in the descending order of date, i.e. the recent prices are at the top and the older prices are at the bottom. Secondly, Yahoo also gives a very important column for each row called **Adjusted Close**. This column gives the corrected value of Close Price after adjusting the price for stock split. For all our further analysis we are going to use the Adjusted Close price instead of the original Close Price.

1.3 Data Preparation

Price Data from Yahoo are very clean and it is already given in a nicely formatted table structure. So there are not much to be done for Data clean up. We needed to reverse the order of the obtained Data so that the Dates are ordered in ascending order (Chronologically). Also we added the Columns for the Adjusted Price for all the other Price components (High, Open, Low). The correction factor for the stock split is obtained from the ratio of Close Price to the Adjusted Close Price .

$$\begin{aligned}
&StockSplitAdjustmentFactor = [AdjustedClose]/[Close] \\
&AdjustedHigh = [High] * [StockSplitAdjustmentFactor] \\
&AdjustedLow = [Low] * [StockSplitAdjustmentFactor] \\
&AdjustedOpen = [Open] * [StockSplitAdjustmentFactor]
\end{aligned} \tag{1.2}$$

From now on we will mean Adjusted price, when we say price.

Now we added the Daily Range column, which is given by

[Daily Range] = [High]-[Low]

We also added the Daily Return Columns and Daily Max Return column. We used Percentage Return instead of Log Return, for consistency sake (Discussed later). So the formula for Daily Return and Daily Max Return are given by :-

$$\begin{aligned}
&DailyReturn = [Close]/[Open] - 1 \\
&DailyMaxReturn = [High]/[Low] - 1 = [DailyRange]/[Low]
\end{aligned} \tag{1.3}$$

It is obvious that Daily Return can be negative and would oscillate around 0, but the Daily Max Return can never be negative, as High Price is always greater than Low Price (will rarely be equal, unless the Stock is inactive).

1.4 Experiment Choices

We decided to experiment with different Price Periodicity. For that, we are planning to use Different Smoothing Indicator with different periodicity. The indicators we are planning to try are.

1. Simple Moving Average (SMA)
2. Weighted Moving Average (WMA)
3. Exponential Moving Average (EMA)
4. Moving Linear Regression Indicator (MLRI)

If the time-series has the values V_i at the **i** time **t** then the Generic Moving Average MA_t at time **t** for a period **n** is given by

$$MA_t = \frac{\sum_{i=t-n+1}^{t} V_i * f(i)}{\sum_{i=1}^{n} f(i)} \tag{1.4}$$

Where **f(i)** is the generic weight function,

Simple Moving Average the weight function is given by **f(i)=1**;

And for Weighted Moving Average (End Weighted) the weight function is given by **f(i)=i**;

So SMA is given by

$$SMA_t = \frac{\sum_{i=t-n+1}^{t} V_i * f(i)}{n} \tag{1.5}$$

And WMA is given by

$$WMA_t = \frac{\sum_{i=t-n+1}^{t} V_i * i}{n*(n+1)/2} = \frac{2*\sum_{i=t-n+1}^{t} V_i * i}{n*(n+1)} \tag{1.6}$$

For Exponential Moving Average EMA_t at time **t** it is given by

$$EMA_t = \alpha * V_t - (1-\alpha) * EMA_{t-1} \tag{1.7}$$

Where α is called the exponential coefficients

For Moving Linear Regression Indicator $MLRI_t$ at time **t** it can be proven that

$$MLRI_t = 3 * WMA_t - 2 * SMA_t \tag{1.8}$$

Also we should also try an indicator which take in to consideration the effect of Volume movement. In stock market the movement of price should agree with the volume. For example, a price can suddenly jump or dive because of some stray trading. in that case the volume for that corresponding price would be very minimal. But if the movement is supported by a strong rise in volume then that movement is more reliable. For this idea we decided to experiment with a Money-flow indicator called **Accumulation / Distribution Line**

The Steps for calculating the Accumulation / Distribution Line are given below

[Close Line Value] or CLV is given by

$$\begin{aligned} CLV &= [(Close - Low) - (High - Close)]/(High - Low) \\ &= 2*(Close - AVERAGE(High + Low))/[DailyRange] \end{aligned} \tag{1.9}$$

[Daily Money Flow] or DMF is given by

$$DMF = CLV * [Volume] \tag{1.10}$$

Accumulation Distribution Line (ADL) is the Cumulative Value of the Daily Money Flow. Daily Money Flow can be Positive or Negative, so ADL will also have ups and downs because of that.

$$ADL_t = \sum_{i=1}^{t} DMF_i = DMF_i + ADL_{t-1} \tag{1.11}$$

For calculating the movements of the Price we need to find the change of price from a base point. This base point can be the Starting Point of the day or a Price sometime in the history. The change is always measured as the ratio of the base price. Which is basically the Rate of Change (ROC) of the Price. This concept of movement is called **Returns**.

$$\begin{aligned} RateofChangeofPrice &= ([CurrentPrice - BasePrice])/[BasePrice] \\ &= [CurrentPrice]/[BasePrice] - 1 \end{aligned} \tag{1.12}$$

Sometime the Log Return is also used in place of Percentage Change. The formulas for those two are given below:-

$$\begin{aligned} LogReturn &= Log([CurrentPrice]/[BasePrice]) \\ &= Log([CurrentPrice]) - Log([BasePrice]) \end{aligned} \tag{1.13}$$

In our case the Base Price can be the price P time before the Current day. So the Base Price will also be moving as we move along the time-line. The difference of time between the Current time and the Base time would be called the **periodicity** of the Change. We are planning to experiment with different periodicity for measuring the movement of the Price. In this context it is to be noted that the Log Returns are very similar to the Rate of Change (ROC) values when the periodicity is very small (1 or 2 days). For higher periodicity the log returns give very different values of the movements. So for consistency sake we are going to use only Rate of Change. ROC at time t for a periodicity P is given by

$$ROC_t = (Price_i - Price_{t-P})/Price_{t-P} = (Price_i/Price_{t-P}) - 1 \tag{1.14}$$

As it is obvious that in calculating the Returns we are only using the two extreme end points. In this analysis we are completely ignoring price movements of the intermediate prices. This may be okay for small periodicity. But for bigger periodicity it won't necessarily depict the correct pattern of movements of the prices. To overcome this problem we can use **Linear Regression Slope (Coefficients) [LRS]** of the Prices within that time period P. It can be Proven that

$$LRS_t = \frac{6}{(P-1) * (WMA_t - SMA_t)} \tag{1.15}$$

For future work we could also adopt sophisticated indicators like **Moving Average Convergence Divergence (MACD)** or Lag Adjusted Moving Averages.

1.5 Graph Preparation

As discussed before the Stocks don't have any direct connections between them. But we can assume that they have virtual connections among them based on the similarity of their movements. The Movements of Stocks are measured by their returns (relative price changes) and not by their absolute price change. In the previous sections we have discussed different types of stock returns.

In this sections we are going to discuss how to measure the similarities between the Stock returns. We would use Pearson's Correlation Coefficient to measure the Strength of the Returns between two Stocks.

The Pearson Correlation coefficients $C_i j$ between two stocks i and j are given as

$$C_{ij} = \frac{< R_i R_j > - < R_i >< R_j >}{\sqrt{(< R_i^2 > - < R_i >^2)(< R_i^2 > - < R_i >^2)}} \tag{1.16}$$

The Correlation Coefficients between each stocks can be represented in a **Square Matrix**. The Size of the Row and the Columns of that Matrix are the number of the Stocks. In this case because it is undirected relationship the Matrix will be symmetrical around the Diagonal. The Diagonal would be always 1.

Heatmap chart with different intensity of color coding can be used to visualize the relationship strength for different Stocks.

We could approach this in two different ways. We can have a weighted graph or a Minimum Spanning Tree (MST) using the strength of their relationships. Or we can have a undirected, unweighted graph using the relationship as in indication of the edge between two stocks. if the stocks exceeds a **threshold value**.

First we intended to study the overall graph properties, the node properties and the Group properties of the undirected graph based on the threshold. The properties that we are planning to study are as follows :-

Overall Graph Properties

1. No. Of Total Nodes
2. No. Of Connected Nodes
3. No. of Edges
4. Power-law distribution
5. Density
6. Diameter (Based on the Biggest group of connected nodes)

7. Radius (Based on the Biggest group of connected nodes)
8. Centrality (Freeman's Centrality measure)

Node Properties

1. Degree
2. Eccentricity
3. Degree Centrality
4. Connection Centrality
5. Betweenness Centrality
6. Eigen-Vector Centrality

Group Properties

1. Cliques
2. Clusters

To investigate this we have created an Application to study those different properties with different options as discussed before. This Application is created using Microsoft C#.NET programming language. (Which is 99% similar to Java programming language). Everything is developed from the ground up, without using any Third party libraries.

The code for that can be seen or downloaded from here from the github repository.

For storing the Correlation Coefficients **Adjacency Matrix** Data Structure is used and for storing the Graph Structure **Adjacency List** structure is used. First all the Edges are detected. This gave us the density of the Graph. Density D of a Graph of N number of nodes and E number of Edges are given by

$$D = \frac{2 * E}{N * (N-1)} \tag{1.17}$$

Sparseness can be thought as just the inverse of the Density

Then the Degrees of each node is calculated. The number of links that is connected to a node is called the Degree of the Node.For a node i the degree k_i is given by

$$k_i = \sum_{i=1; i \neq j}^{n} E_{ij} \tag{1.18}$$

Where k_i denotes the Edges from node i to j

Using the Degrees for all the nodes we could calculate the distribution of the Degrees and the number of nodes associated to it. This would give us the indication if the Graph follows the Power Law distribution or not.If the Number of Nodes that has Degree 1 is A, then the assumed degree k_i using the Power law function for any Node i can be approximated as

$$k_i = A * i^{-x} = \frac{A}{i^x} \tag{1.19}$$

The value of x can be found which gives the total least squared deviation of the Actual Values from the Calculated Values. This can be found using Brute Force algorithm or finding the Numerical Method Solutions for the Equation that can be found equating the derivative of the total Least Squared deviation to 0.

$$\begin{aligned} &\frac{d}{dx}\sum_{i=1}^{n}(D_i - \frac{A}{i^x}) = 0 \\ &\sum_{i=1}^{n}\frac{D_i}{i^{x-1}} = A * \sum_{i=1}^{n}\frac{1}{i^{(2*x+1)}} \end{aligned} \tag{1.20}$$

Instead of solving the X we just made it an Experiment input.

Then the code for finding the shortest distance from one node to another node is calculated using the **Breadth First Search (BFS)** algorithm. Using that, all the possible routes from one node to all the nodes are calculated. This gave us the Eccentricity of each node and the diameter & radius of the overall graph.

For the Graph visualization we have used our own algorithm to draw the graph. For drawing the Graph we drew the nodes in concentric circles, according to their eccentricities. The Nodes with lowest eccentricities will be in the inner-most circle and the nodes with higher eccentricities will be in the outer circles. The nodes will be placed at an equal angle in one circle. The starting angle will be randomly chosen in order to minimize the overlap of lines under the nodes.We also gave the facility to move the nodes around the graph. So that the Graph can be better adjusted for visual display.

1.6 Experiments

We did some experiments with the process as discussed before. The results are as follows:-

Price	:	Close Price
Date Range	:	[19-Mar-2008] - [25-Jul-2016]
Return	:	Previous Day's Return
Similarity	:	Pearson Correlation Coefficient

	AAPL	AXP	BA	CAT	CSCO	CVX	DD	DIS	GE	GS	HD	IBM	INTC	JNJ	JPM	KO	MCD	MMM	MRK	MSFT	NKE	PFE	PG	TRV	UNH	UTX	VZ	V	WMT	XOM	YHOO
AAPL	1.00	0.48	0.45	0.49	0.51	0.46	0.49	0.47	0.48	0.46	0.47	0.49	0.52	0.39	0.47	0.35	0.42	0.48	0.36	0.48	0.43	0.40	0.37	0.46	0.40	0.52	0.40	0.44	0.32	0.42	0.39
AXP	0.48	1.00	0.57	0.61	0.54	0.54	0.62	0.63	0.63	0.66	0.59	0.50	0.57	0.48	0.71	0.43	0.46	0.61	0.45	0.53	0.53	0.52	0.48	0.61	0.45	0.63	0.50	0.57	0.36	0.52	0.40
BA	0.45	0.57	1.00	0.60	0.53	0.58	0.61	0.63	0.55	0.51	0.53	0.54	0.54	0.53	0.50	0.48	0.52	0.62	0.45	0.52	0.52	0.49	0.50	0.50	0.44	0.71	0.46	0.45	0.40	0.59	0.40
CAT	0.49	0.61	0.60	1.00	0.58	0.66	0.70	0.62	0.60	0.56	0.53	0.56	0.59	0.49	0.54	0.44	0.46	0.67	0.42	0.54	0.50	0.49	0.45	0.50	0.40	0.69	0.49	0.48	0.36	0.63	0.44
CSCO	0.51	0.54	0.53	0.58	1.00	0.57	0.60	0.60	0.52	0.54	0.54	0.58	0.62	0.51	0.53	0.45	0.48	0.60	0.45	0.60	0.50	0.47	0.49	0.51	0.40	0.62	0.50	0.47	0.41	0.56	0.44
CVX	0.46	0.54	0.58	0.66	0.57	1.00	0.66	0.64	0.56	0.55	0.52	0.57	0.57	0.61	0.51	0.52	0.52	0.64	0.53	0.56	0.47	0.57	0.56	0.58	0.50	0.65	0.55	0.48	0.43	0.88	0.42
DD	0.49	0.62	0.61	0.70	0.60	0.66	1.00	0.67	0.62	0.59	0.58	0.58	0.60	0.55	0.60	0.49	0.51	0.68	0.49	0.56	0.53	0.54	0.54	0.56	0.44	0.68	0.54	0.51	0.41	0.65	0.43
DIS	0.47	0.63	0.63	0.62	0.60	0.64	0.67	1.00	0.59	0.58	0.63	0.58	0.59	0.59	0.58	0.54	0.54	0.67	0.52	0.58	0.57	0.57	0.58	0.58	0.50	0.70	0.55	0.52	0.46	0.63	0.46
GE	0.48	0.63	0.55	0.60	0.52	0.56	0.62	0.59	1.00	0.54	0.56	0.54	0.56	0.50	0.66	0.42	0.48	0.65	0.44	0.46	0.50	0.52	0.49	0.55	0.42	0.65	0.52	0.47	0.38	0.54	0.41
GS	0.46	0.66	0.51	0.56	0.54	0.55	0.59	0.58	0.54	1.00	0.55	0.53	0.51	0.47	0.75	0.39	0.41	0.56	0.41	0.50	0.45	0.48	0.42	0.57	0.45	0.57	0.45	0.53	0.33	0.51	0.39
HD	0.47	0.59	0.53	0.53	0.54	0.52	0.58	0.63	0.56	0.55	1.00	0.53	0.55	0.52	0.58	0.48	0.55	0.60	0.45	0.52	0.57	0.50	0.52	0.54	0.47	0.62	0.54	0.50	0.50	0.50	0.41
IBM	0.49	0.50	0.54	0.56	0.58	0.57	0.58	0.58	0.54	0.53	0.53	1.00	0.59	0.52	0.54	0.46	0.48	0.58	0.44	0.56	0.50	0.47	0.49	0.52	0.41	0.61	0.51	0.46	0.41	0.56	0.41
INTC	0.52	0.57	0.54	0.59	0.62	0.57	0.60	0.59	0.56	0.51	0.55	0.59	1.00	0.51	0.53	0.46	0.48	0.59	0.47	0.64	0.49	0.49	0.48	0.53	0.42	0.62	0.52	0.46	0.38	0.56	0.43
JNJ	0.39	0.48	0.53	0.49	0.51	0.61	0.55	0.59	0.50	0.47	0.52	0.52	0.51	1.00	0.43	0.59	0.53	0.63	0.62	0.52	0.47	0.64	0.64	0.56	0.53	0.62	0.56	0.44	0.51	0.64	0.38
JPM	0.47	0.71	0.50	0.54	0.53	0.51	0.60	0.58	0.66	0.75	0.58	0.54	0.53	0.43	1.00	0.39	0.46	0.57	0.41	0.49	0.50	0.48	0.46	0.61	0.41	0.59	0.49	0.53	0.35	0.49	0.39
KO	0.35	0.43	0.48	0.44	0.45	0.52	0.49	0.54	0.42	0.39	0.48	0.46	0.46	0.59	0.39	1.00	0.52	0.52	0.47	0.48	0.44	0.50	0.58	0.49	0.46	0.55	0.51	0.33	0.44	0.53	0.33
MCD	0.42	0.46	0.52	0.46	0.48	0.52	0.51	0.54	0.48	0.41	0.55	0.48	0.48	0.53	0.46	0.52	1.00	0.55	0.46	0.49	0.53	0.49	0.55	0.51	0.43	0.56	0.49	0.43	0.48	0.53	0.35
MMM	0.48	0.61	0.62	0.67	0.60	0.64	0.68	0.67	0.65	0.56	0.60	0.58	0.59	0.63	0.57	0.52	0.55	1.00	0.50	0.57	0.53	0.57	0.59	0.59	0.47	0.74	0.56	0.48	0.46	0.65	0.44
MRK	0.36	0.45	0.45	0.42	0.45	0.53	0.49	0.52	0.44	0.41	0.45	0.44	0.47	0.62	0.41	0.47	0.46	0.50	1.00	0.44	0.40	0.63	0.53	0.48	0.46	0.50	0.49	0.38	0.41	0.54	0.32
MSFT	0.48	0.53	0.52	0.54	0.60	0.56	0.56	0.58	0.46	0.50	0.52	0.56	0.64	0.52	0.49	0.48	0.49	0.57	0.44	1.00	0.47	0.48	0.49	0.51	0.42	0.60	0.50	0.42	0.40	0.56	0.41
NKE	0.43	0.53	0.52	0.50	0.50	0.47	0.53	0.57	0.50	0.45	0.57	0.50	0.49	0.47	0.50	0.44	0.53	0.53	0.40	0.47	1.00	0.44	0.47	0.46	0.39	0.56	0.48	0.45	0.44	0.48	0.41
PFE	0.40	0.52	0.49	0.49	0.47	0.57	0.54	0.57	0.52	0.48	0.50	0.47	0.49	0.64	0.48	0.50	0.49	0.57	0.63	0.48	0.44	1.00	0.55	0.56	0.50	0.56	0.51	0.45	0.44	0.58	0.36
PG	0.37	0.48	0.50	0.45	0.49	0.56	0.54	0.58	0.49	0.42	0.52	0.49	0.48	0.64	0.46	0.58	0.55	0.59	0.53	0.49	0.47	0.55	1.00	0.53	0.42	0.59	0.56	0.40	0.52	0.60	0.32
TRV	0.46	0.61	0.50	0.50	0.51	0.58	0.56	0.58	0.55	0.57	0.54	0.52	0.53	0.56	0.61	0.49	0.51	0.59	0.48	0.51	0.46	0.56	0.53	1.00	0.51	0.60	0.50	0.48	0.45	0.57	0.40
UNH	0.40	0.45	0.44	0.40	0.40	0.50	0.44	0.50	0.42	0.45	0.47	0.41	0.42	0.53	0.41	0.46	0.43	0.47	0.46	0.42	0.39	0.50	0.42	0.51	1.00	0.51	0.37	0.35	0.37	0.48	0.34
UTX	0.52	0.63	0.71	0.69	0.62	0.65	0.68	0.70	0.65	0.57	0.62	0.61	0.62	0.62	0.59	0.55	0.56	0.74	0.50	0.60	0.56	0.56	0.59	0.60	0.51	1.00	0.56	0.52	0.49	0.65	0.44
VZ	0.40	0.50	0.46	0.49	0.50	0.55	0.54	0.55	0.52	0.45	0.54	0.51	0.52	0.56	0.49	0.51	0.49	0.56	0.49	0.50	0.48	0.51	0.56	0.50	0.37	0.56	1.00	0.41	0.47	0.57	0.32
V	0.44	0.57	0.45	0.48	0.47	0.48	0.51	0.52	0.47	0.53	0.50	0.46	0.46	0.44	0.53	0.33	0.43	0.48	0.38	0.42	0.45	0.45	0.40	0.48	0.35	0.52	0.41	1.00	0.38	0.45	0.39
WMT	0.32	0.36	0.40	0.36	0.41	0.43	0.41	0.46	0.38	0.33	0.50	0.41	0.38	0.51	0.35	0.44	0.48	0.46	0.41	0.40	0.44	0.44	0.52	0.45	0.37	0.49	0.47	0.38	1.00	0.45	0.29
XOM	0.42	0.52	0.59	0.63	0.56	0.88	0.65	0.63	0.54	0.51	0.50	0.56	0.56	0.64	0.49	0.53	0.53	0.65	0.54	0.56	0.48	0.58	0.60	0.57	0.48	0.65	0.57	0.45	0.45	1.00	0.40
YHOO	0.39	0.40	0.40	0.44	0.44	0.42	0.43	0.46	0.41	0.39	0.41	0.41	0.43	0.38	0.39	0.33	0.35	0.44	0.32	0.41	0.41	0.36	0.32	0.40	0.34	0.44	0.32	0.39	0.29	0.40	1.00

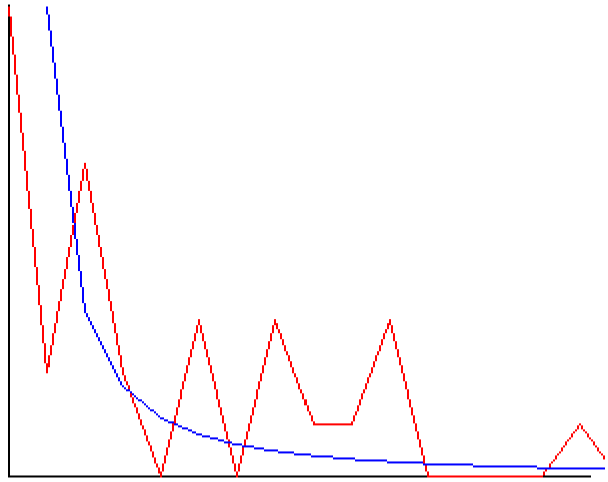
Heatmap for the Correlation Coefficients for different Stocks

Threshold : 0.6

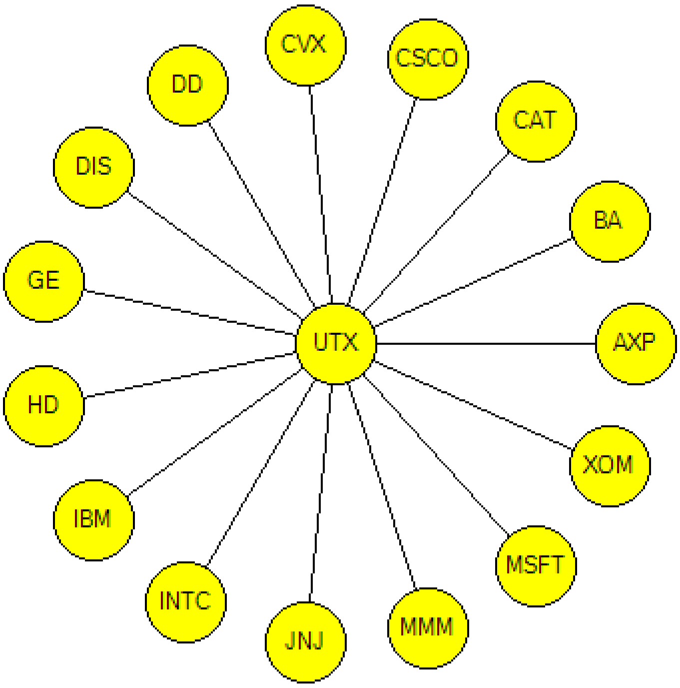
Node Properties

Ticker	Exchange	Degree	Eccentricity	Closeness
AAPL	NASDAQ	0	-	0
AXP	NYSE	9	4	8.25
BA	NYSE	5	4	7.42
CAT	NYSE	8	4	8.08
CSCO	NASDAQ	3	4	7
CVX	NYSE	7	4	8
DD	NYSE	10	4	8.42
DIS	NYSE	10	4	8.42
GE	NYSE	5	4	7.58
GS	NYSE	2	5	6.18
HD	NYSE	2	4	6.83
IBM	NYSE	1	4	6.67
INTC	NASDAQ	3	4	7
JNJ	NYSE	7	4	7.92
JPM	NYSE	5	5	6.93
KO	NYSE	0	-	0
MCD	NYSE	0	-	0
MMM	NYSE	10	3	8.67
MRK	NYSE	2	5	6.02
MSFT	NASDAQ	2	4	6.83
NKE	NYSE	0	-	0
PFE	NYSE	2	5	6.02
PG	NYSE	1	5	5.85
TRV	NYSE	2	5	6.18
UNH	NYSE	0	-	0
UTX	NYSE	15	3	9.5
VZ	NYSE	0	-	0
V	NYSE	0	-	0
WMT	NYSE	0	-	0
XOM	NYSE	7	4	8
YHOO	NYSE	0	-	0

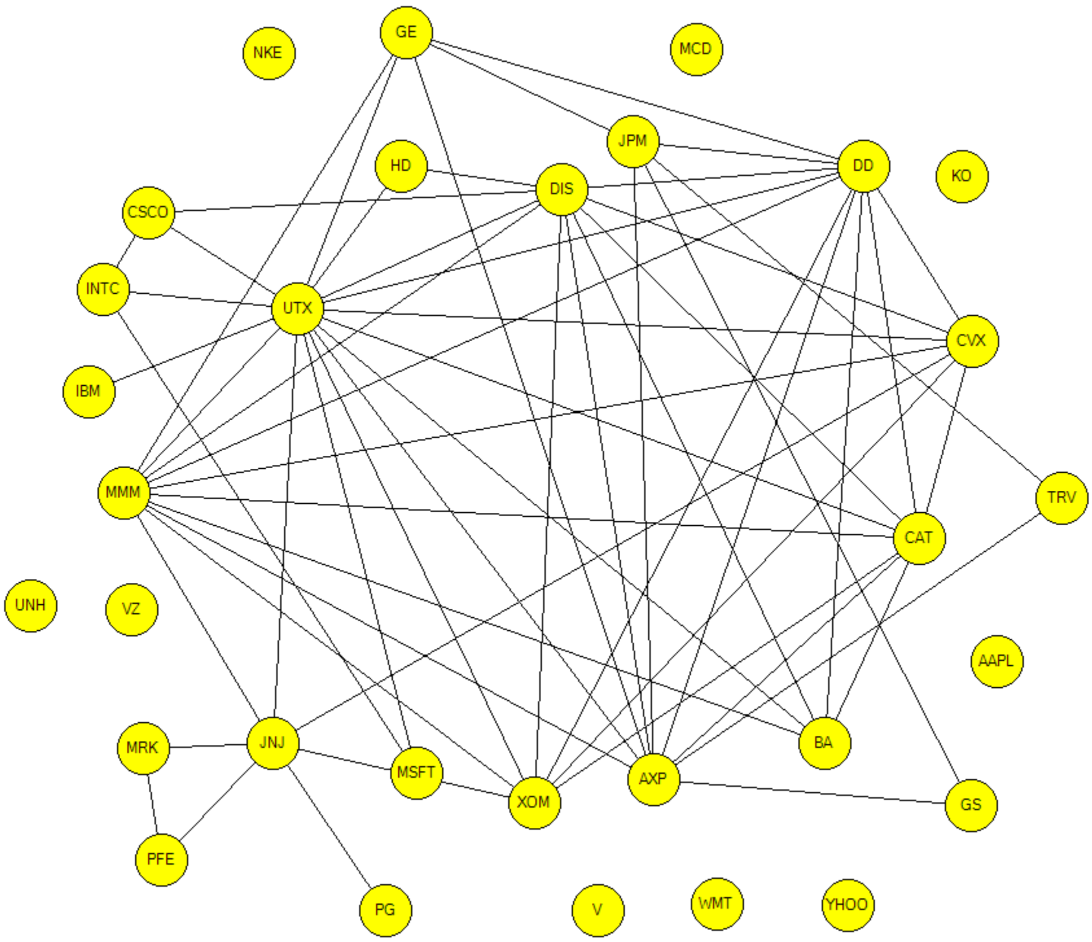
Power-law distribution of the Nodes. [Red Actual and Blue estimated; Exponent : 1.5]



The Degree distribution of one of the Stock; Example : UTX]



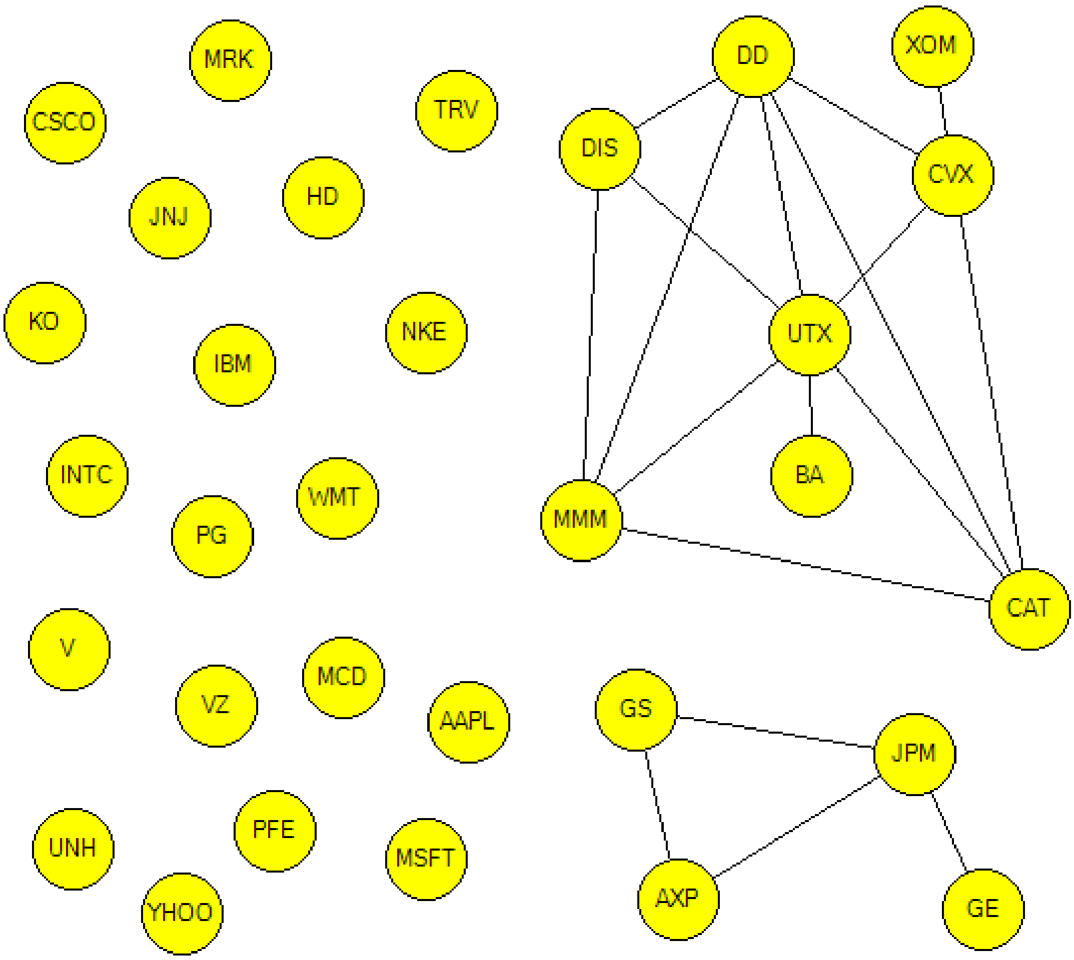
The Complete Graph



No. of Nodes	:	31
Connected Nodes	:	22
No. of Edges	:	59
Diameter *	:	5
Radius *	:	3

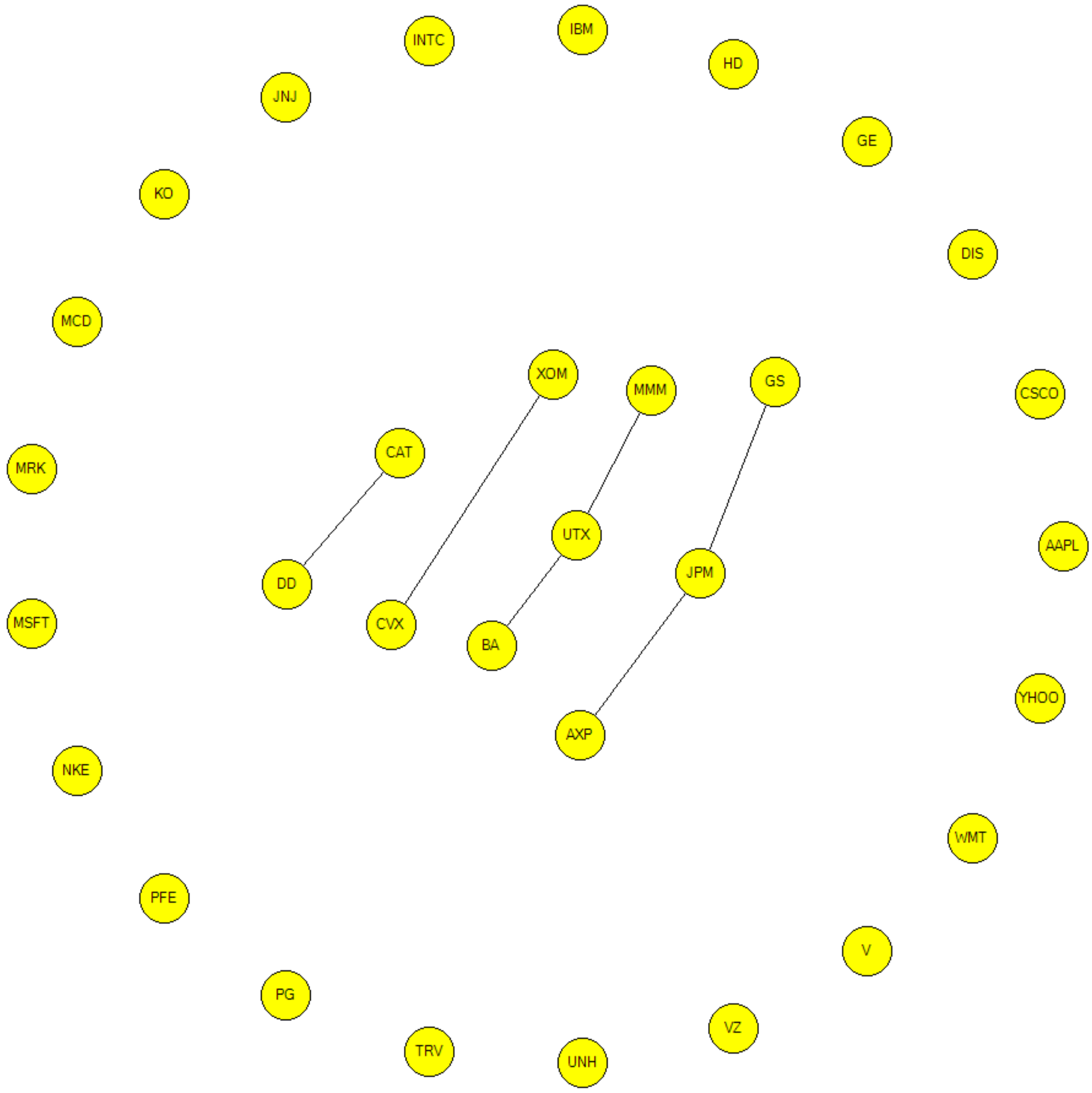
Threshold : 0.65

No. of Nodes	:	31
Connected Nodes	:	12
No. of Edges	:	18
Diameter *	:	4
Radius *	:	2



Threshold : 0.70

No. of Nodes	:	31
Connected Nodes	:	10
No. of Edges	:	6
Diameter *	:	3
Radius *	:	2



As we can see if we increase the threshold value the graph becomes more and more parse and less connected nodes will be there. And for higher threshold it is not following power law.