Portfolio construction of stocks using social network analysis

Mubarrat Tajoar Chowdhury
Department of Industrial Engineering
UNIST
mubarrattajoar@unist.ac.kr

Md Khalequzzaman Chowdhury Sayem
Department of Industrial Engineering
UNIST
khalequzzamansayem@unist.ac.kr

Abstract
In this work, we form a series of optimized portfolios of stocks out of companies listed in the S&P500 index fund, using social network theories of centrality. Centrality measures for each stock in the network is used to decide the proportion of portfolio budget to be allocated to different stocks. And then, the performance of the optimized portfolios is tested against the overall performance of the S&P500 index funds at the specified duration. The objective is to analyze how centrality property of a stock impacts it's price changes within a particular duration.

I. Introduction

Traditional methods for quantitative evaluation and portfolio construction (Markowitz portfolio optimization for eg,) consider performance of a stock as somewhat independent by considering the price change of a stock as its return and variance in the price change as the stock's volatility.
Using social networks to model stocks allows for interrelation of different stocks within an asset universe to be considered. The general expectation is that performance of different stocks is correlated and hence, building a social network based on correlations between different asset allows one to consider this phenomenon. By then picking the 'most connected' stocks in the network the general trend of the market can then captured.
Whether such 'high-connectivity' stocks return high returns is explored in this work. Some of the methodologies in this work are inspired from the work[1]. However, stock selection methodologies and experiment setups are our own contribution. The conclusion reached and the trends detected from the series of experiments are also different.

II. Theoretical Background

Daily price growth of a particular stock is calculated by comparing the price of the particular stock A to it's previous day price. This can be formulated as:

$$return_{stock\_A,Day} = \log\left(\frac{Closing\_Price_{Stock\_A,Day}}{Closing\_Price_{Stock\_A,Day-1}}\right)$$

Network construction is based on correlation between daily return of stocks in the S&P500 index fund.

The general objecting for any portfolio construction is to maximize cumulative return over the investment period.
To decide on stocks to be selected during portfolio construction, three centrality values are considered for each stock; Betweenness Centrality, Degree Centrality and Closeness Centrality. A more intuitive understanding of centrality measures can be found in this paper [2].

Betweenness Centrality: Proportion of shortest paths between other node pairs other than node v that passes through node v.

$$C_b(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where, $\sigma_{st}(v)$ = number of shortest paths between nodes s and t that passes through node v
σ_{st} = number of shortest paths between nodes s and t

Degree Centrality: Ratio of number of existing connections of node v to max possible connections of node v. Calculated as:

$$C_d(v) = \frac{d(v)}{n-1}$$

where, n = number of nodes in the network.
d(v) = degree of node v.

Closeness Centrality: Ratio of number of nodes connected to node v to summation of shortest distance between node v and the other connected nodes

$$C_c(v) = \frac{n-1}{\sum_{j=1}^{n} shortest\_dist(v,j)}$$

Symbols in your equation should be defined before the equation appears or immediately following. Cite equations using "(1)," not Eq. (1)" or "equation (1)," except at the beginning of a sentence: "Equation (1) is …"

III. Method

Data source: Yahoo Finance website, for dates 1st January 2013 onwards.

Daily returns for all S&P500 stocks are calculated from 1 January, 2013 to 31 December, 2015, which shall be referred to as training period. A correlation matrix is then generated for daily returns within the training period. The correlation matrix is used to pick out stock pairs with correlation value of 0.7 or higher. The stock pairs are then used to form the social network of stocks. The network is illustration generated using Gephi is displayed in figure 1.

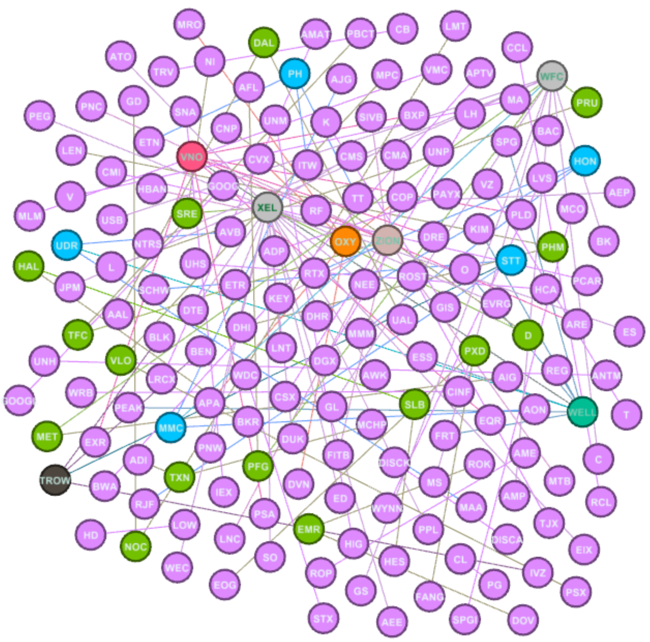

Fig. 1 *Network of highly correlated stocks*. Edges between node pairs exist only if the node pairs have correlation > 0.7.

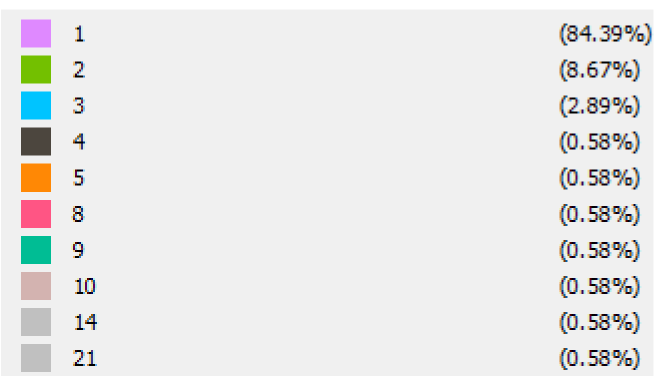

Fig. 2. *Colourmap of proportion of nodes(stocks) in Fig1 with the specified degree values*

The centrality values of edges are then calculated for the stocks in the 'high-correlation' network. To account for the overall centrality of each stock, weighted sum of the three centralities is calculated.

$$C(v) = \omega_b C_b(v) + \omega_d C_d(v) + \omega_c C_c(v)$$

To find the optimum weights for the three centrality measures the following objective function is defined, which is then optimized.

$$Maximize\ \text{return} * (\omega_b C_b + \omega_d C_d + \omega_c C_c)$$

Using the python CVXPY package the optimal weights are found.

The weighted sum of each stock using the optimal weight values is then calculated.

To select the stocks to invest in, either *n* stocks with highest values of weighted sum of centralities or *n* stocks with lowest values for weighted sum of centralities are chosen. The value of *n* is changed in different instances of experiments

The metric for proportion of budget to be allocated to different stocks is decided based on whether the 'bottom' few stocks are being chosen or top few stocks are being chosen from a sorted list of stock with descending measure of weighted sum of centralities.

If *n* top stocks are chosen from the sorted list, proportion of a particular stock that is included in the portfolio is proportional to the weight sum of centrality measure for that stock. This proportion for a stock A is calculated as:

$$Portfolio\_Weight_{stock\_A} = \frac{C(A)}{\sum_i^n C(i)}$$

If *n* bottom stocks are chosen from the sorted list, proportion of a particular stock that is included in the portfolio is inversely proportional to the weight sum of centrality measure for that stock. This proportion for a stock A is calculated as:

$$Portfolio\_Weight_{stock\_A} = \frac{C(A)^{-1}}{\sum_i^n C(i)}$$

For a particular experiment all the stocks in the network were chosen and proportion of each stocks in the constructed portfolio proportional to their weighted sum of centrality measure, similar to how proportion was calculated in the *n-top* case.

IV. Results and Discussion

Data from 1 January 2016 to 30 November 2021 is used to test how constructed portfolios perform in the test duration. The setup of the experiment therefore shows cumulative return upon using the specified strategies, had one invested starting from 1 January 2016 to 30 November 2021, using only the latest available data at the time up-to the period 31 December 2015 as training period to develop the discussed investing strategy.

Seven experiments are performed in total. The selection methods of stocks from sorted list of stocks with weighted sum of centralities descending down the list. Calculation of proportion of individual stocks to be included in the portfolio is done using the methods discussed towards the end of previous section:

- 30 Bottom stocks (Fig 3)
- 20 Bottom stocks (Fig 4)
- 10 Bottom stocks (Fig 5)
- All stocks (Fig 6)
- 30 Top stocks (Fig 7)
- 20 Top stocks (Fig 8)
- 10 Top stocks (Fig 9)

Below are series of charts showing how the worth of different portfolios grew to by the end of testing period. Benchmarks mentioned in the charts are overall growth in the same period for S&P500 index funds. The yellow traces in the charts are the benchmark "S&P500"index fund and the blue traces show growth of corresponding network portfolios.

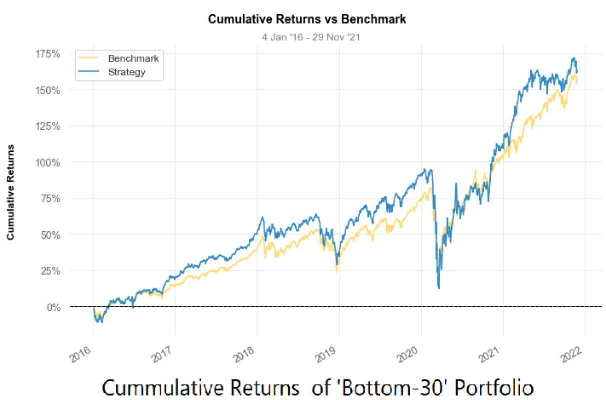

Fig 3

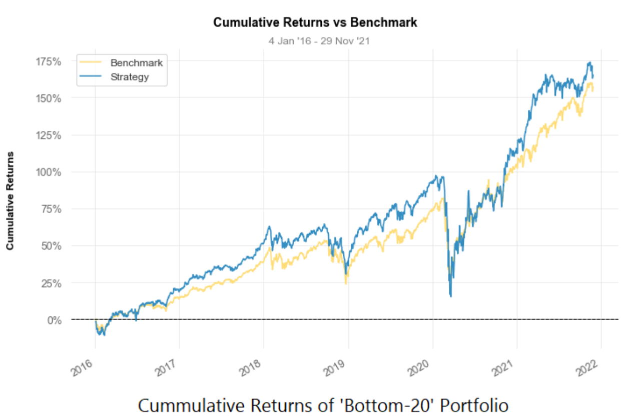

Fig 4

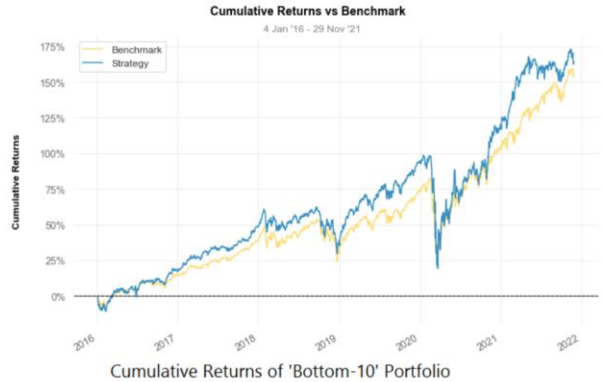

Fig 5

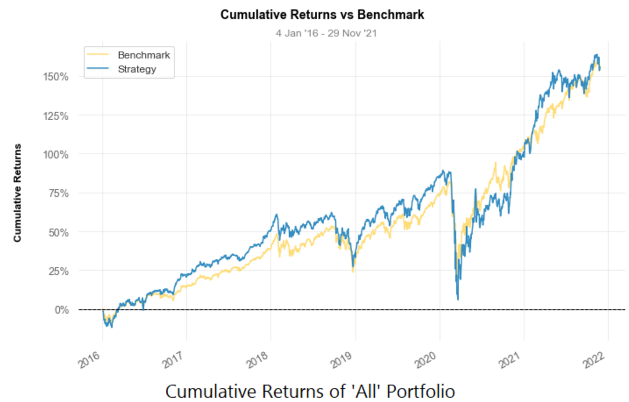

Fig 6

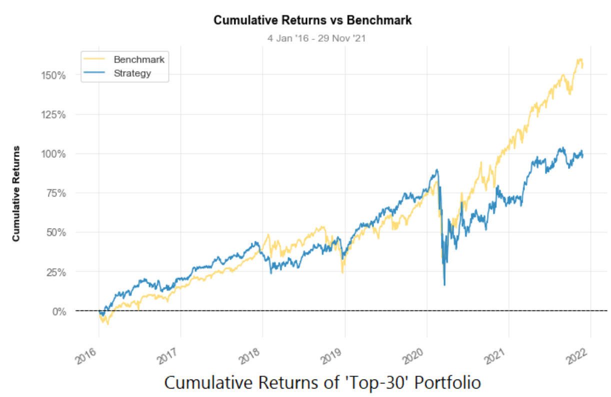

Fig 7

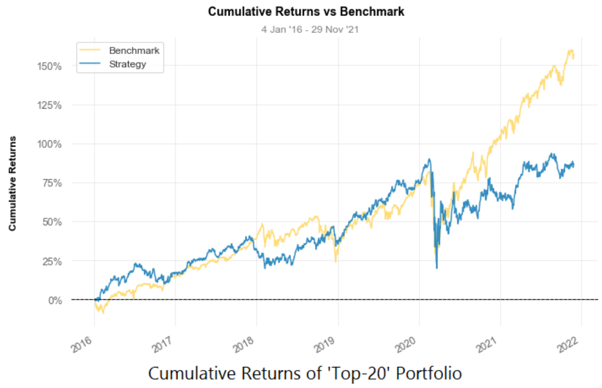

Fig 8

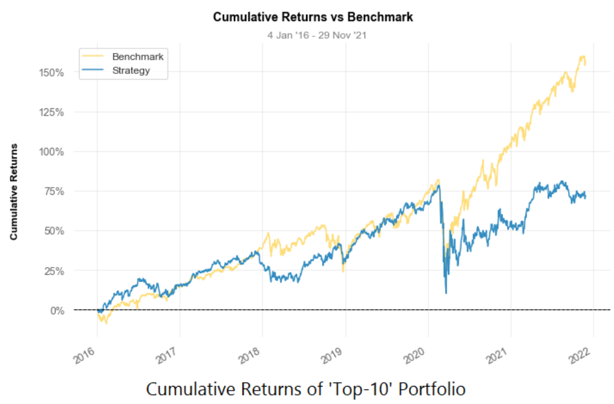

Fig 9

Below is Fig 10 showing Annual Return comparison for each of the seven portfolios.

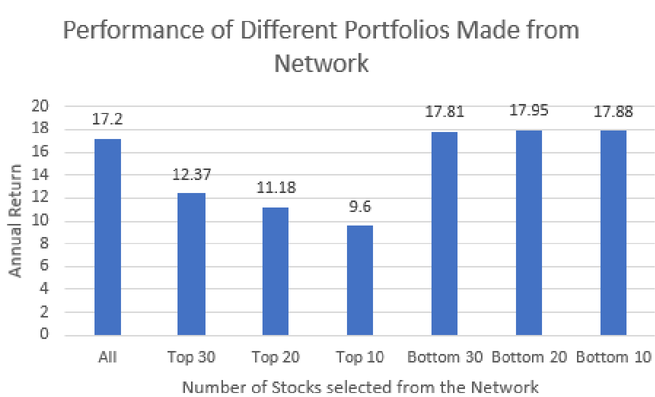

Fig 10

As illustrated in Fig 3,4,5 and the last three bars in Fig 10, returns are higher for stocks with lower weighted sum of centrality. Figure 7,8,9 and 2nd o 4th bars in figure 10 imply the opposite. Not only do the 'bottom' portfolios end up at higher cumulative return value than the benchmark, the growth of 'bottom *n*' stocks dominate for majority of the testing period whereas benchmark dominates for majority of the period in case of 'top *n*' stocks.

From a social network analysis perspective alone, it could be expected that the stocks with higher centrality measures would be the influential nodes and set the market trend considering their high connectedness.

However, as the results imply, the case might not be so. The expectation from highly connected stocks changes based on how high centrality measures are interpreted. Since, edges in the network portfolio indicate presence of a high correlation between pairs of stocks forming the edge, the stocks with high centrality measure nodes are those which have high correlations with many stocks in the network. Hence, higher connectivity in this scenario might not mean "trend-setter" but instead implies that price of highly connected nodes are dependent on price changes of many more stocks than prices of stocks with lower centrality measures,

V. Conclusion

Price of portfolios made up mostly of stocks with a higher measure of centrality are more volatile because of the stocks in such portfolios having high correlations with price changes with multitude of different stocks. Hence, returns from such portfolios tend to fluctuate more and lead to lower returns. Stocks that are part of high performing indexes and have a lower centrality measure yield returns that have lower volatility, ie, more robust. Hence, stocks such are better suited for portfolio construction as they are able to maintain their high performing trend

References

[1] Sun, Wenyue, Chuan Tian, and Guang Yang. "Network analysis of the stock market." (2015).

[2] Zhang, Junlong, and Yu Luo. "Degree centrality, betweenness centrality, and closeness centrality in social network." Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). Vol. 132. 2017.