A Semantic Network Analysis of the Stock Market

Seth Kimmel **Benjamin Klimko** **Matthew Mutammara**

Abstract

We propose a semantic network analysis of the stock market. By constructing a relational graph of stocks mentioned together within human-written articles, we seek to expose underlying relationships among the stocks which otherwise may not be apparent. We will perform multiple network analyses on this graph and investigate what, if any, demonstrative power it may have in understanding the market and its constituents.

All relevant code for this project can be found at:
https://github.com/sethkimmel3/SemanticNetworkStockMarketAnalysis.

1 Introduction

There is a lot of work done trying to make sense of the countless datapoints produced by the stock market every day. One focus of this work is to better understand the relationships between stocks: how one stock's performance influences other stocks. There are several ways of defining these relationships, including by industry, by stock value thresholds, or even geographically.
Our goal is to avoid these rigidly defined relationships and instead capture the publicly perceived relationships between stocks. Specifically, we chose stocks as nodes and define edge weights as the frequency with which stocks are mentioned together. We capture this information by scraping financial news articles and checking how often stocks are mentioned in the same article.
We then perform multiple analyses on the resulting graph and look for correlations between graph properties and financial data in hopes of uncovering new understandings of the community of stocks.

2 Problem Statement

We seek to solve the problem of market understanding and market prediction through building a semantic co-occurrence network from financial news articles on stocks. Once the network is complete, we will perform both network specific analyses such as community detection as well as correlation of network information to market time-series data to understand relationships between stocks and its relationship to individual stock performance.

3 Literature review

We are not the first to consider looking at the stock market as a graph. We have gathered a number of prior research publications which analyze the stock market in a similar fashion. These publications include:

1. On Structural Properties of the Market Graph (1)

In this paper, the authors are creating a network analogous to those of a call graph of telephone numbers, or the Internet graph. Their network representation of the stock market takes stocks as vertices and forms an edge between them if the correlation in price fluctuation

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

exceeds a specific threshold. In a sense, this is the opposite task we are doing, as we are looking to study price correlations as an effect of other data, instead of the cause. Hence, this paper serves as an interesting reference and potential comparison article when we generate our results.

2. A Network Analysis of the Chinese Stock Market (4)

The primary purposes of this paper was to formulate the Chinese Stock Market as a graph. They then study its properties including connected components, cliques, independent sets, as well as the topological stability of the network. This last point allows them to detect critical nodes in the network which may lead to market instability or catastrophe. These insights could help portfolio managers understand their holdings and protect against potential risk.

3. Mining Market Data: A Network Approach (2)

This paper also created a network by calculating cross correlations of opening stock prices for pairs of stocks and then drew conclusions from how the network structure changed day to day about the dynamics of the stock market. This is relatively straightforward, and gives us similar insights to that of the first paper we looked at. However the inverse relationship to our problem only goes so far in providing valuable ideas.

4. Identifying influential stock indices from global stock markets: a Social Network Analysis Approach (5)

This paper analyzed the changes to the correlation network and minimum spanning tree of global indices both before and after the collapse of Lehman Brothers, allowing them to determine which stock indices are most influential, regional influence on co-movement, and how the collapse of Lehman impacted markets both in the US and abroad. This is valuable in providing the idea that we can extract economic importance from our network. This might allow us to see which nodes in the network - if they failed - could lead to market collapse.

5. Network Analysis of the Stock Market (6)

The authors of this paper also sought to generate a graph based on correlations of stock returns, allowing them to perform measurements on the network and provide potentially valuable insights to investors. While not entirely novel compared to the other papers, this acts as a valuable comparative measure, and gives us more insight and ideas about the inverse relationship that we are studying in our work.

6. Impact of dynamic corporate news networks on asset return and volatility (3)

This paper builds a network of the STOXX 50 index, pairing each stock within it, and generating edge weights as the sum of the number of news articles on a topic in a set the authors devised. They then attempt to find correlation between network properties including centrality with the return and volatility of the index. Finally they attempt to use these insights in a predictive manner for future returns. While this is a different manner of constructing the graph, it is potentially the most similar study we have come across to ours. This is a valuable reference to understand how they tackled the problem, and may be helpful in bench-marking our results when finished. While not exact to what we are doing, it is helpful nonetheless.

Additionally, it is quite likely that there is significant work that has been done in this area within industry, but would be proprietary and not accessible for reference.

4 Network Construction

Our network is constructed entirely from data that we collected. As explained in our proposal, our network consists of stocks as nodes and edge weights as the frequency with which the companies are mentioned together. We capture this information by scraping financial news articles and checking how often stocks are mentioned in the same article.

4.1 Design Choices

Before collecting data, we needed to decide what companies we would include and what news sources would be queried. We decided to start by only including companies in the S&P 100 because they would be most likely to appear in news articles. For news sources, we somewhat arbitrarily chose Bloomberg, CNBC News, MSNBC News, NBC News, The New York Times, Reuters, TechCrunch, The Washington Post, The Washington Times, Wired, and The Verge.

4.2 Data Collection

Our data collection process consists of two steps:

1. URL Collection
We use the NewsAPI Python library to compile news articles published in the past month from the news sources above. The free version of the API limits us to this 1-month timespan, but we believe that recent articles should suffice for the purposes of this project. The API returns a list of URL's to news articles for each company.

2. Article Search
We use the Newspaper API to read each news article into a Python string. We then use basic substring search for the name of each company in each news article. We do our best to replace full company names with their colloquial names when searching for mentions (e.g. replace "The Walt Disney Company" with "Disney"). We then, for each pair of companies, count the number of articles that mention both companies.

4.3 Network Construction

Finally, we simply create a `Pandas` dataframe with our data. The dataframe includes pairs of companies and the number of articles that mention them both. These mention counts will be the edge weights in our network, forming the adjacency matrix. Using the networkx `from_pandas_dataframe` function, we get our final network.

5 Visualization and Interpretation

Our graph is fully connected except for one disconnected node: Southern Company. Below, we have plotted the graph using Spring Layout and colored companies by sector. Zooming in slightly, we begin to see some meaningful patterns:

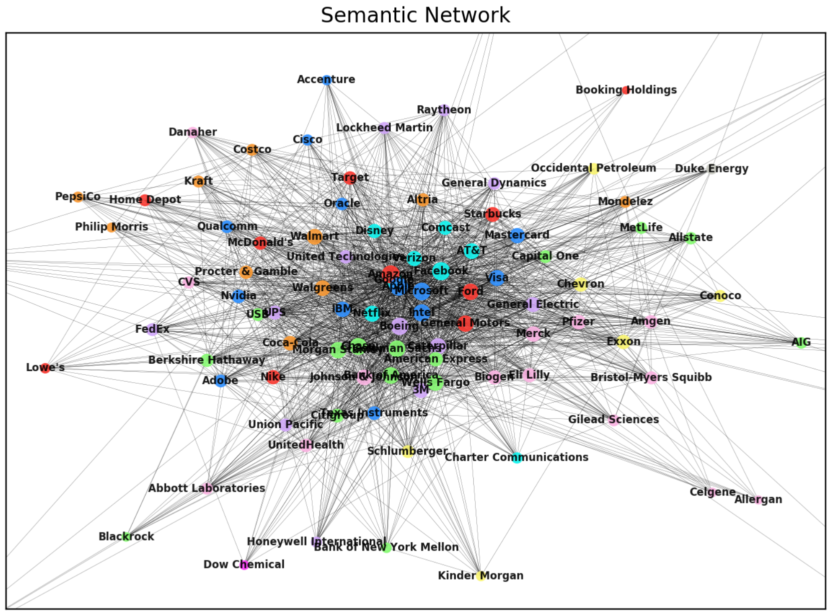


Figure 1: Zoom-In of network in spring layout

Initially, we see sectors begin to cluster together. We also see very high connectivity across the graph.

Looking closer toward the center, we again uncover even more of what we expect to see. Many of the leading companies in the US have connections to each other. Many companies in similar industries appear to be closely related within the graph. For example, we see several financial institutions in green group together at the bottom of the graph. Similarly, we see several healthcare related companies in pink on the right side of the graph. And, of course, the large tech companies like Microsoft and Apple are in the middle with very high degree.

While it is encouraging to see companies from similar sectors cluster together, one exciting feature of our network actually lies in the companies from *different* sectors that appear close together. Take, for example, Target and Walmart near the top left side of the graph. The two companies are classified into different sectors by the S&P's official metrics. However, because the companies are commonly associated with each other, they appear fairly close to each other in our graph. Another example is Capital One on the right side of the image. While it is in the same "green" sector as banks like Wells Fargo and Goldman Sachs, it seems to be more closely associated with Visa and Mastercard (likely because of Capital One's emphasis on their credit card offerings). This distinction would be lost by simply following the company's sector classification. And again, we see insurance companies Metlife, Allstate, and AIG on the right side of the graph. They separate from other financial institutions, distinguishing themselves with more detail than their sector designation would give.

6 Experiments

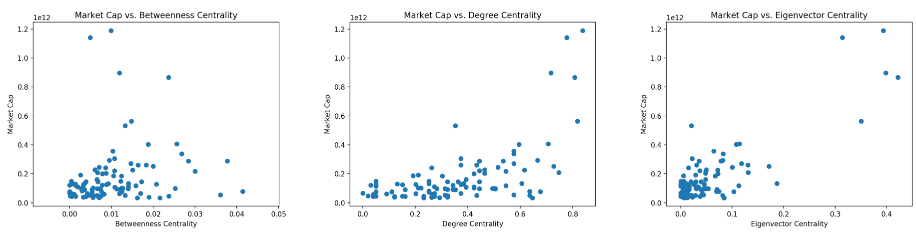
With these insights in mind, we performed a wide variety of experiments. We will first discuss experiments on the graph itself and then move on to examining the graph's correlation to time series data of stock prices.

6.1 Analyses on the Network Alone

6.1.1 Finding economically important stocks

An important understand which we hoped to glean from our network was a sense of which stocks were the most critical. That is, could our network tell us which stocks were most economically important? Doing such analyses are important for macro- and micro-economic studies, policy making, and much more.

To do this, we needed to correlate individual node measures to some standard, current measure of economic importance. We decided to use betweenness, degree, and eigenvector centralities to explain the market capitalization of the stocks in the network. We chose these to correlate these two measures as centrality acts as a fairly strong indicator of a node's importance to a network, and market capitalization is a simple way to understand how important a stock might be to a larger economy. Below are the results of this analysis:



(a) Betweenness (b) Degree (c) Eigenvector

Figure 2: Centrality measures

We find some interesting results from this analysis. Clearly, we see an upward sloping trend in degree centrality vs. market capitalization, and two upward trending clusters form in eigenvector centrality vs. market capitalization. Hence, we can conclude that if we assume market capitalization to be an indicator of economic importance, node centrality measures (degree and eigenvector, specifically) also correlate to economic importance in our network. Perhaps our measures are an even more telling indicator, but further work on this would need to be done to understand whether or not this is the case.

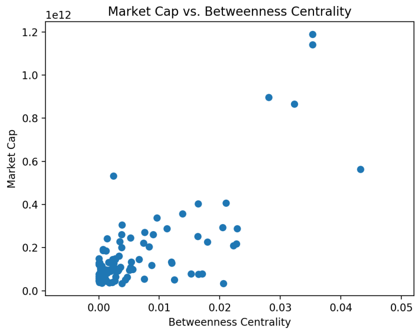
However, the question remains as to why betweenness centrality does not seem to produce any sort of correlation. Looking at how the measure is actually defined, it is the number of shortest paths which pass through the node (weighted or unweighted). Because our network is weighted and defined such that a high edge weight is actually an indication of relative "closeness," this measure should fail when taking weights into account. We can use the unweighted version of the measure to produce the following result instead, which gives a result similar to that of eigenvector centrality. However, eigenvector centrality is also a weighted measure. Below we see the results of both measures calculated without weights:

This correlation now appears similar to that of degree centrality - an unweighted measure.

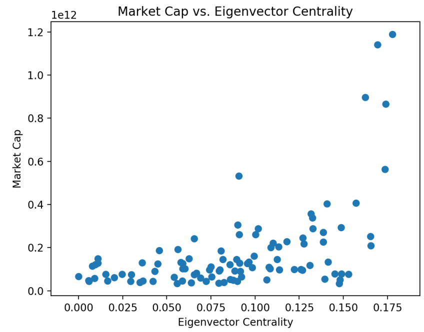
Hence we see that all centrality measures, when being careful relative to the involvement of edge weights, do show correlations to market capitalization, and hence indicate economic importance.

6.1.2 Creating indexes from communities/cliques

Another important notion when it comes to the stock market is forming subsets of the market to create indexes or portfolios. We wanted to see whether we could extract valuable subsets from our network. Specifically we wanted to see if we could extract subsets which have stocks that either move together or separate in price. However, we wanted to simply form the subsets first without any comparison to price data whatsoever. We created these indices in two ways.



(a) Betweenness centrality



(b) Eigenvector centrality

Figure 3: Unweighted centrality experiments

"Bottom-Up" Index Formation

In this method, we choose a number of desired indices (n=10, normally), and divide the range of weights in the network (in our case 0 to 117) by n. We then have n thresholds for allowing stocks into an index. For a stock to be included in an index, it must have an adjacent vertex where the weight of the edge between them is greater than the threshold value for that index. Below are some results for n=10 indices:

1. ['Amgen', 'Danaher', 'Biogen', 'Altria', 'Occidental Petroleum', 'Celgene', 'Schlumberger', 'Caterpillar', 'Johnson & Johnson', 'Cisco', 'Coca-Cola', 'Bank of America', 'Boeing', 'Berkshire Hathaway', 'Charter Communications', 'Paypal', 'Starbucks', 'Procter & Gamble', 'Comcast', 'Oracle', 'Union Pacific', 'Intel', 'Adobe', 'Allergan', 'Walmart', 'Exelon', "Lowe's", 'AT&T', 'General Dynamics', 'Colgate-Palmolive', 'PepsiCo', 'Allstate', 'Raytheon', 'Abbott Laboratories', 'Google', 'NextEra', 'Dow Chemical', 'Goldman Sachs', 'Booking Holdings', 'United Technologies', 'Capital One', 'Bristol-Myers Squibb', 'Exxon', 'General Electric', 'Philip Morris', 'Conoco', 'Kinder Morgan', 'Nvidia', 'Wells Fargo', 'Visa', 'Simon Property Group', 'Costco', 'Amazon', 'Kraft', 'Honeywell International', 'USB', 'Medtronic', 'Walgreens', 'Pfizer', 'Ford', 'Target', "McDonald's", 'General Motors', 'American Express', 'Mondelez', 'UPS', 'Bank of New York Mellon', 'Gilead Sciences', 'CVS', 'Citigroup', 'Netflix', 'MetLife', 'Disney', 'Mastercard', 'AIG', 'Nike', 'IBM', 'Texas Instruments', 'Microsoft', 'Merck', 'Chevron', 'DuPont de Nemours', 'Home Depot', 'Blackrock', 'Lockheed Martin', 'Morgan Stanley', 'Chase', 'FedEx', 'Verizon', 'UnitedHealth', 'Duke Energy', 'Abbvie', 'Apple', 'Accenture', '3M', 'Southern Company', 'Emerson Electric', 'Qualcomm', 'Facebook', 'Eli Lilly']
2. ['Altria', 'Apple', 'Caterpillar', 'Boeing', 'Intel', 'Amazon', 'Texas Instruments', 'Microsoft', 'Johnson & Johnson', 'Bank of America', 'Goldman Sachs', 'Wells Fargo', 'Chase', 'UnitedHealth', 'Coca-Cola', 'American Express', 'Citigroup', 'Netflix', 'Morgan Stanley', 'Ford', 'Facebook', 'Comcast', 'AT&T', 'Disney', 'Oracle', 'Google', 'IBM', 'USB', 'Walmart', 'Target', 'Verizon', 'Nvidia', 'Exxon', 'Chevron', 'Visa', 'Mastercard', 'Walgreens', 'CVS', 'Pfizer', 'Merck', 'General Motors', '3M', 'Qualcomm']
...
5. ['Google', 'Amazon', 'Microsoft', 'Apple', 'Facebook', 'Netflix']
6. ['Google', 'Amazon', 'Microsoft', 'Apple', 'Facebook']
7. ['Google', 'Amazon', 'Microsoft', 'Apple', 'Facebook']
8. ['Google', 'Amazon', 'Apple', 'Facebook']
9. ['Google', 'Amazon', 'Facebook', 'Apple']
10. ['Google', 'Amazon']

Obviously, the groupings get more restrictive and thus smaller as thresholding increases. Interestingly, at index 5, our method has generated a known stock grouping: FAANG + M (Facebook, Apple, Amazon, Netflix, Google, and Microsoft). This grouping of stocks is very popular for use in finance, indicating our method does capture public sentiment. This could be due to source bias or simply the fact that these stocks are talked about more, but is still interesting nonetheless.

"Top-Down" Index Formation

In this method, we start with the same process as before by choosing n and dividing the weight range to form thresholds. However, instead of building up and allowing stocks into thresholds, we start by initializing all indices to include the full S&P 100, and then systematically remove them. If a node has an adjacent node with an edge weight between them greater than the threshold value for that index, it is removed. Below are some of the results:

1. []
2. ['Amgen', 'Danaher', 'Biogen', 'Occidental Petroleum', 'Celgene', 'Schlumberger', 'Cisco', 'Berkshire Hathaway', 'Charter Communications', 'Paypal', 'Starbucks', 'Procter & Gamble', 'Union Pacific', 'Adobe', 'Allergan', 'Exelon', "Lowe's", 'General Dynamics', 'Colgate-Palmolive', 'PepsiCo', 'Allstate', 'Raytheon', 'Abbott Laboratories', 'NextEra', 'Dow Chemical', 'Booking Holdings', 'United Technologies', 'Capital One', 'Bristol-Myers Squibb', 'General Electric', 'Philip Morris', 'Conoco', 'Kinder Morgan', 'Simon Property Group', 'Costco', 'Kraft', 'Honeywell International', 'Medtronic', "McDonald's", 'Mondelez', 'UPS', 'Bank of New York Mellon', 'Gilead Sciences', 'MetLife', 'AIG', 'Nike', 'DuPont de Nemours', 'Home Depot', 'Blackrock', 'Lockheed Martin', 'FedEx', 'Duke Energy', 'Abbvie', 'Accenture', 'Southern Company', 'Emerson Electric', 'Eli Lilly']
3. ['Amgen', 'Danaher', 'Biogen', 'Altria', 'Occidental Petroleum', 'Celgene', 'Schlumberger', 'Johnson & Johnson', 'Cisco', 'Coca-Cola', 'Berkshire Hathaway', 'Charter Communications', 'Paypal', 'Starbucks', 'Procter & Gamble', 'Comcast', 'Oracle', 'Union Pacific', 'Adobe', 'Allergan', 'Exelon', "Lowe's", 'General Dynamics', 'Colgate-Palmolive', 'PepsiCo', 'Allstate', 'Raytheon', 'Abbott Laboratories', 'NextEra', 'Dow Chemical', 'Booking Holdings', 'United Technologies', 'Capital One', 'Bristol-Myers Squibb', 'Exxon', 'General Electric', 'Philip Morris', 'Conoco', 'Kinder Morgan', 'Nvidia', 'Simon Property Group', 'Costco', 'Kraft', 'Honeywell International', 'USB', 'Medtronic', 'Walgreens', 'Pfizer', 'Ford', 'Target', "McDonald's", 'General Motors', 'American Express', 'Mondelez', 'UPS', 'Bank of New York Mellon', 'Gilead Sciences', 'CVS', 'Citigroup', 'MetLife', 'AIG', 'Nike', 'Texas Instruments', 'Merck', 'Chevron', 'DuPont de Nemours', 'Home Depot', 'Blackrock', 'Lockheed Martin', 'Morgan Stanley', 'FedEx', 'UnitedHealth', 'Duke Energy', 'Abbvie', 'Accenture', '3M', 'Southern Company', 'Emerson Electric', 'Qualcomm', 'Eli Lilly']

...

10. ['Amgen', 'Danaher', 'Biogen', 'Altria', 'Occidental Petroleum', 'Celgene', 'Schlumberger', 'Caterpillar', 'Johnson & Johnson', 'Cisco', 'Coca-Cola', 'Bank of America', 'Boeing', 'Berkshire Hathaway', 'Charter Communications', 'Paypal', 'Starbucks', 'Procter & Gamble', 'Comcast', 'Oracle', 'Union Pacific', 'Intel', 'Adobe', 'Allergan', 'Walmart', 'Exelon', "Lowe's", 'AT&T', 'General Dynamics', 'Colgate-Palmolive', 'PepsiCo', 'Allstate', 'Raytheon', 'Abbott Laboratories', 'NextEra', 'Dow Chemical', 'Goldman Sachs', 'Booking Holdings', 'United Technologies', 'Capital One', 'Bristol-Myers Squibb', 'Exxon', 'General Electric', 'Philip Morris', 'Conoco', 'Kinder Morgan', 'Nvidia', 'Wells Fargo', 'Visa', 'Simon Property Group', 'Costco', 'Kraft', 'Honeywell International', 'USB', 'Medtronic', 'Walgreens', 'Pfizer', 'Ford', 'Target', "McDonald's", 'General Motors', 'American Express', 'Mondelez', 'UPS', 'Bank of New York Mellon', 'Gilead Sciences', 'CVS', 'Citigroup', 'Netflix', 'MetLife', 'Disney', 'Mastercard', 'AIG', 'Nike', 'IBM', 'Texas Instruments', 'Microsoft', 'Merck', 'Chevron', 'DuPont de Nemours', 'Home Depot', 'Blackrock', 'Lockheed Martin', 'Morgan Stanley', 'Chase', 'FedEx', 'Verizon', 'UnitedHealth', 'Duke Energy', 'Abbvie', 'Apple', 'Accenture', '3M', 'Southern Company', 'Emerson Electric', 'Qualcomm', 'Facebook', 'Eli Lilly']

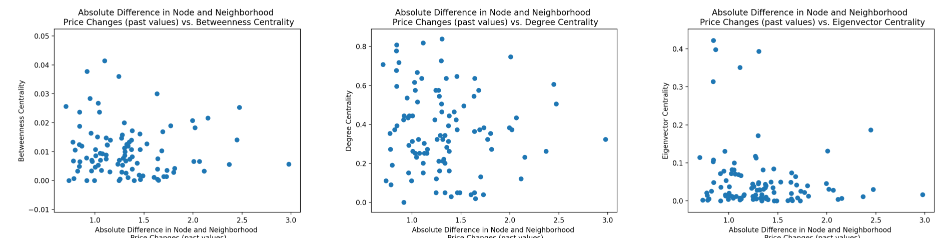
Here we can see groupings getting less restrictive as the index thresholds increase. The idea behind this method is find uncorrelated groupings - those which have little similarity between them. This will be explained when tested against price data.

6.2 Correlations to Stock Prices

6.2.1 How Centrality Affects Price Movements of Neighborhoods

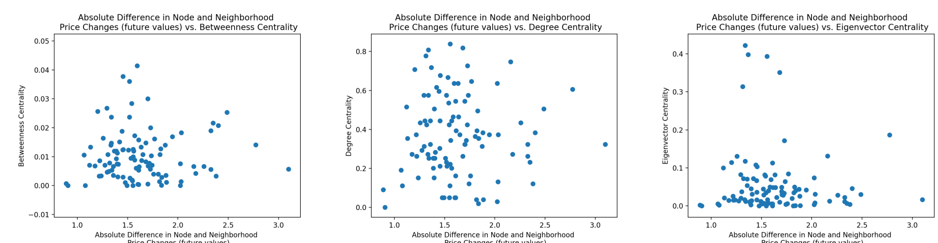
The first analysis that was performed which included stock price data was to understand how the centrality of a node has an influence on the average price movements of its neighborhood. Specifically, we wanted to know whether an increase in centrality for a node would indicate more uniform movements in its neighborhood. Hence, we correlated the centrality of nodes to the average price change difference between it and its neighbors over some period of time.

We used daily closing stock prices and performed analyses which correlated past stock prices to past stock prices of its neighbors (for example, prices in the range $t - n$ and t for both node and neighbors) as well as future indicators ($t - n$ to t values explaining the t to $t + k$ prices for its neighbors, as an example). We used these same conventions for all analyses with price data.



(a) Betweenness (b) Degree (c) Eigenvector

Figure 4: Centrality measures for past values, $n = 3$



(a) Betweenness (b) Degree (c) Eigenvector

Figure 5: Centrality measures for future values, $n = 3,\ k = 1$

As you can see from the figures, there is fairly weak correlation between the centrality of a node and the average price movements of its neighborhood. This holds for both past and future values. This is somewhat surprising, but perhaps there is simply too much noise in the data, or not enough data to reveal any correlation.

6.2.2 Pairwise Distance (Degree Separation) vs. Pairwise Average Price Difference

The next analysis we performed was to understand whether a relationship exists between unweighted pairwise distance of nodes and the average price change difference of the nodes. Again, we used the same conventions in terms of price data, and performed both past and future value correlations.

For past values (n=1), we have the following:

For Distance = 1, Average Price Change Difference = 0.0123%.
For Distance = 2, Average Price Change Difference = 0.0127%.
For Distance = 3, Average Price Change Difference = 0.0131%.

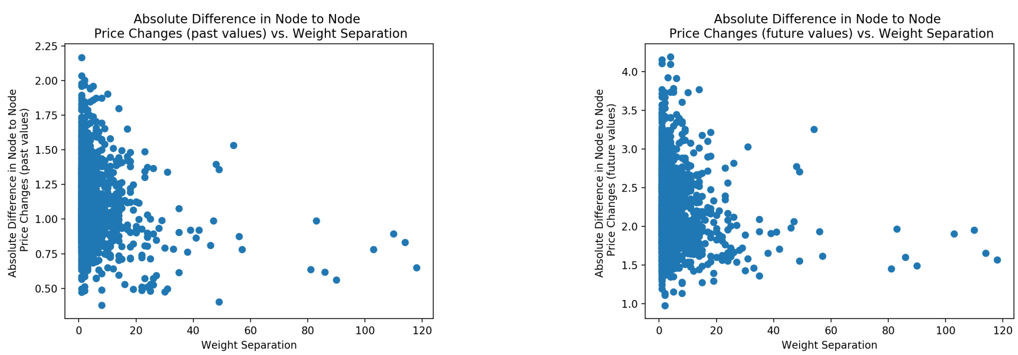
For future values (n=3,k=1), we have the following:

For Distance = 1, Average Price Change Difference = 0.0252%.
For Distance = 2, Average Price Change Difference = 0.0256%.
For Distance = 3, Average Price Change Difference = 0.0256%.

We can see that pairwise degree separation actually does play an interesting role in explaining price differences. It appears that the further the unweighted distance is for two nodes in a network, the more separated their prices are. This is what one might expect and is interesting to see in practice. While it may not be extremely significant looking, an idea such as this could lay the foundation for a predictive price strategy.

6.2.3 Pairwise Edge Weights (Weight Separation) vs. Pairwise Average Price Change Difference

In this analysis, we sought to understand how weights between adjacent nodes explained the average price change difference. This is a fundamentally different question than the previous analysis, and only takes adjacent nodes (not all pairs) into account. Below are the results:



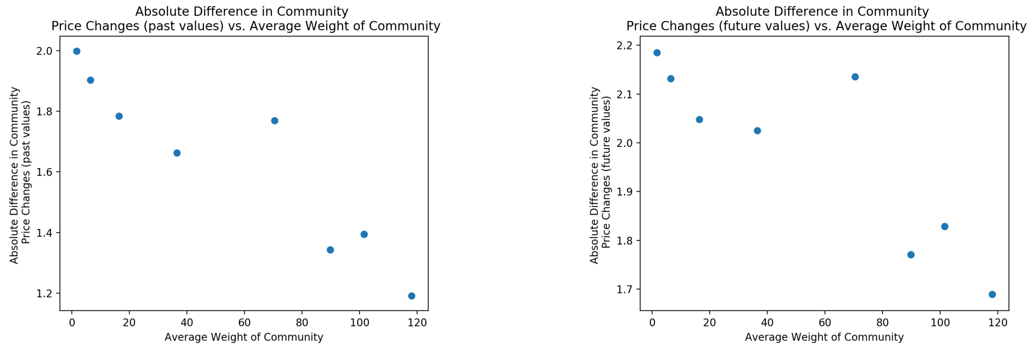
(a) Past values, $n = 3$ (b) Future values, $n = 3,\ k = 1$

Figure 6: Pairwise edge weights explaining average price changes

As you can see, we have interesting clusters which form in both the past and future analysis. However, trend identification is difficult here, indicating few substantial results.

6.2.4 Average Weight of Index vs. Index Price Changes

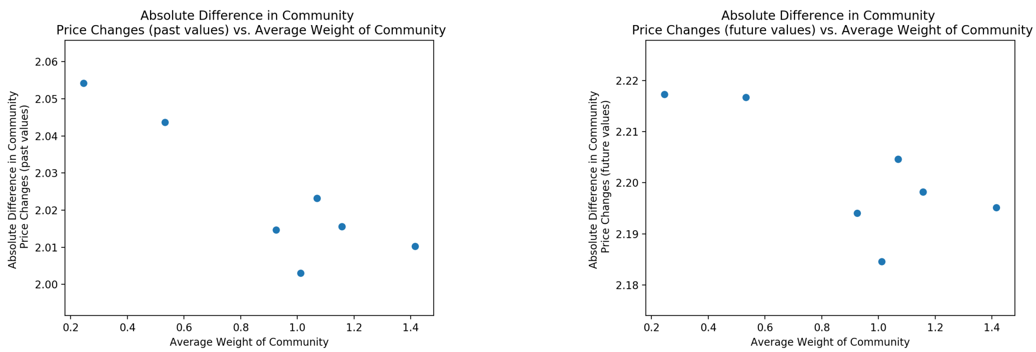
For our final analysis, we return to the index formation analysis we did earlier and correlate their average weight of the index with price change differences of contained stocks. We do this for both the top-down as well as bottom-up index formation methods. For each we do a past and future analysis.



(a) Past values, $n = 3$

(b) Future values, $n = 3,\ k = 1$

Figure 7: Average index weight (bottom-up formation) explaining average price changes



(a) Past values, $n = 3$

(b) Future values, $n = 3,\ k = 1$

Figure 8: Average index weight (top-down formation) explaining average price changes

Interestingly, we obtain the result in both the bottom-up and top-down cases that average weight of a community (index) has a strong influence on the average price change differences of the community. Hence, we can conclude that as groupings of stocks are formed with increasing average weight in the group, we can expect that the group will move more as a unit in price.

This is extremely helpful for selection of portfolios where the manager may want very strong correlations among the stocks, or very uncorrelated stocks for diversification. This appears to be one of the strongest results of our analyses.

7 Limitations

A few limitations from this work that could improve results are:

- Removing bias from news source selection and content.
- Using NLP methods for language extraction rather than just substring search.
- Extending our news article search beyond just the prior 30 days.
- Extending our financial time-series data and correlations past 90 days.
- Using more than just S&P 100 stocks.

Making such modifications could allow us to see a clearer picture from our analyses.

8 Conclusion

Using our semantic stock market network, we have shown the power of network science applied to the problem of market understanding, and found several notable results. Clearly, by using networks to capture public sentiment about the stock market, we can extract trends that allow us to glean insights from the market. This would not have been possible years ago before the proliferation of network science methods, tools, and relevant sources of data.

Among the most interesting results from the analyses we performed were:

- Index generation via "top-down" and "bottom-up" methods to find correlated and uncorrelated groupings of stocks.
- The notion of centrality measures of nodes as indicators of economic importance.
- Pairwise degree separation explaining price change differences.

While most of these analyses are presented in a purely academic sense, they can form the basis for important work to be done in economics and finance. Additionally, this is only a subset of possibly analyses which could have been performed and encourage those reading to consider other useful properties of this network.

9 Future Work

In potential future work, we would of course like to address some of the limitations discussed above. We would like to capture such fine trends from this network so as to build predictive pricing strategies. This may require far more data and extensive computing capabilities, as well as implementing graph deep learning algorithms, such as DeepWalk or similar. We may also want to incorporate social media data to gather sentiment in an even finer detail for such purposes.

Finally, understanding network dynamics could be interesting in this case. How the markets evolve over time with respect to public sentiment could lend itself to powerful analyses and understanding.

References

[1] V. Boginski, S. Butenko, and P. Pardalos. On structural properties of the market graph. In A. Nagurney, editor, *Innovations in Financial and Economic Networks*. Edward Elgar, 2003.

[2] V. Boginski, S. Butenko, and P. Pardalos. Mining market data: A network approach. *Computers & Operations Research*, 33, 2006.

[3] G. G. Creamer, Y. Ren, and J. V. Nickerson. Impact of dynamic corporate news networks on assets return and volatility. *Howe School Research*, (2013-4.), 2013.

[4] Q. Huang, X. Zhuang, and S. Yao. A network analysis of the chinese stock market. *Physica A*, 388, 2009.

[5] R. B. Roy and U. K. Sarkar. Identifying influential stock indices from global stock markets: a social network analysis approach. *Procedia Computer Science*, 5, 2011.

[6] W. Sun, C. Tian, and G. Yang. Network analysis of the stock market. Stanford CS224W project report, 2015.