A Study of Industry Networks and Information Flow

Jiyuan Ding, Quan Zhou, Yunhao Zhang

Practice of QCF, Fall 2016

Executive Summary

As the growth of modern technology and global economy, the world has gradually become a connected whole. So have the industries in the economy. The connections between different industries have become stronger in recent years, which is the reason of the fact that all industries are affected by other related industries. Financial economists have long recognized the importance of understanding how value relevant information disseminates to stock markets and how market participants incorporate this information into stock prices. So we have the assumption that the stocks prices of companies in different industries have some interesting relationships.

We examined whether the returns of different industries are partially driven by the difference in speed of information flow due to their inherent position in the industry network. We constructed such a network using the supply and demand data from Bureau of Economic Analysis and formed a long-short trading strategy based on the centrality of the industries. Finally, the returns of our portfolios are compared against several asset pricing models such as CAPM and Fama-French 5-factor model to examine whether there is an extra driving factor other than the risk factors in the models. We discovered that our central/ peripheral portfolios have little dependency on the risk factors specified in CAPM and FF5 models, indicating our returns may be partially driven by the information incorporated in industries' positions in the networks.

Introduction

In this paper, we try to find the relationship based on information flow, in the framework of networks. We question whether a node's position in the network affects the complexity of value-relevant information that investors process and thus influences the speed of information flow through the network. We divided the industries into central industries and peripheral industries. As the names suggested, the central industries are more central, which means that they have more related industries, while the peripheral industries have less related industries.

We build connections between related industries, and we could get a network of industries after combining all the connections.

Our assumption is that the peripheral industries would react quicker to the changes in their related industries as they have fewer related industries compared with central industries. The fundamental reason for this assumption is that to price the central industry, investors need to understand not only shocks to other related industries and their price movements but also how important the related industries are in the total revenue of the central industry. Processing more complicated information can slow down the information flow to the central industry.

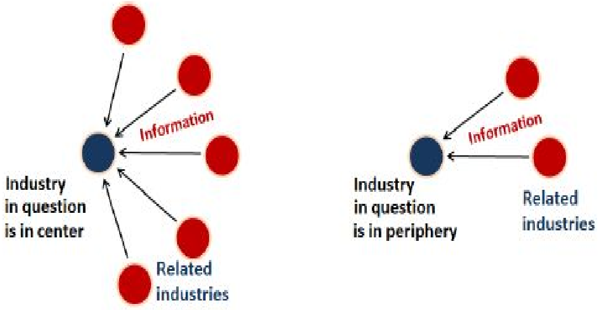
Information
Industry in question is in center
Related industries
Information
Industry in question is in periphery
Related industries

Figure 1. An illustration of the difference in central/peripheral network information flow

In order to prove our assumption, in each month, we form quintile portfolios of central industries sorted on the return of related industries of central industries in the previous month, go long the quintile portfolio with the highest past return of related industries, and go short the quintile portfolio with the lowest past return of related industries. This self-financing trading strategy is rebalanced every month and involves the buying and selling of central industries. To compare the two kinds of industries to prove our assumption, we form the same strategy on peripheral industries, and compare the return of two strategies.

Our project is consist of two parts. We firstly mimic the results of Joonki Noh's paper 'Industry Networks and the speed of Information Flow'. Then we used the network to develop strategies to see the performance of the strategy in recent stock market. Our main goal is not only to earn money with this strategy, but also try to prove the fact that the information flow to peripheral industries are quicker than to central industries.

Data Sources

For the Network Construction part, we used the industry Input-Output (IO) Tables produced by the Bureau of Economic Analysis (BEA). The BEA is a U.S. government agency that provides official macroeconomic and industry statistics including the gross domestic product of the United States.

Starting in 1947, the BEA reports domestic industry statistics every five years. The data format and details changed, but some basic elements remain the same over time. One main part of the BEA's industry report is the MAKE and USE tables. The MAKE table records the values of goods that each industry produces and the USE table gives us the commodities that are consumed by the industries. Each table describes a match between industries and commodities. As mentioned in Noh's paper, by combining these two tables, we can build an industry-to-industry supply-demand relationship. From there, we construct an industry network and compute the centrality of every industry. Details on network construction will be introduced in Methodology.

The industry IO data is available on BEA's official website. We used the IO tables in 1992 to replicate the result. The IO tables use self-defined six-digit industry codes together with a file that shows how to convert the industry codes to SIC (Standard Industry Classification) codes. The 1992 IO tables contain roughly 400 to 500 industries and commodities. The data from bea is saved in Coordinate Format (COO) in a text file. However, for some unknown reasons, the 2007 IO table is saved as dense matrix in an excel file. For this project, since our primary goal is to replicate the strategy in the paper, only the 1992 IO tables are used.

For the Strategy Backtesting part, we got the data from the Center for Research in Security Prices (CRSP). Monthly stock data from 1992 to 2007 was downloaded through WRDS in CSV format. To make the dataset as small as possible, only the following columns were selected when downloaded. CUSIP, a unique number for each company, doesn't change over time. SICCD, SIC code of the stock, indicating which industry it belongs to. SIC code is also the bridge between CRSP data and BEA IO tables. ASK, closing ask price on that day. BID, closing bid price of the stock on that day. SHROUT, number of outstanding shares. We use the average of ASK and BID to approximate closing price. Market value can be calculated by multiplying SHROUT and price.

Methodology

Constructing the industry network

To construct the industry network, we first extract data from the BEA IO tables. The IO tables all refer to the 1992 BEA detailed IO tables unless specified otherwise. The MAKE table has 4 columns: industry code, commodity code, table reference number and cell value. Except for table reference number, all other three columns are useful. It shows the dollar value of commodity that is produced by an industry. The USE table has 13 columns. Only 3 of them are useful, which are commodity code, industry code and cell value. Other columns such as Water Transportation Cost, Wholesale Margin and Retail Margin are not used in this project. The USE table shows the dollar value of commodity that is consumed by a certain industry. With MAKE table and USE table , we can skip commodity and generate an industry-to-industry supply-demand relationship.

We normalize the MAKE table along each column and multiply it by the USE table, producing the REVSHARE matrix. Then normalize REVSHARE long each row and we have the CUST matrix. The entry *(i, j)* in CUST indicates the fraction of industry *i*'s sales consumed by industry *j*. Therefor if *(i, j)* is large, it means industry *i* is very important as a customer to industry *j.* Normalize REVSHARE along each column and we have the SUPP matrix, which is similar to CUST. The SUPP matrix shows the fraction of industry *j*'s purchases produced by industry *i*. So large *(i,j)* in SUPP means industry *i* is very important as a supplier to industry *j*.

After computing CUST and SUPP, we combine these two matrix by averaging them into one square matrix, which is the COMB matrix. And we symmetrize COMB by taking the maximum value of *(i, j)* and *(j, i)* entries. The COMB matrix is an industry-to-industry matrix that indicated the strength of links among different industries. It is used to determine an industry's position in the network.

Determine the position of an industry in the network.

We first set all diagonals to zero. Then the eigenvalue centrality is defined as:

$$c_i = \frac{1}{\lambda} \sum_j A_{i,j} c_j,$$

where $A_{i,\,j}$ denotes the *(i, j)* entry of the adjacency matrix A and λ is a scaling constant. This equation is quite intuitive. To be more central in the network, an industry needs to be connected to more industries and/or connected strongly to other more central industries. Thus the eigenvector centrality captures not only the number of connections but also their strength in the industry network. In a matrix form, the equation is written as c = Ac, meaning that the principal eigenvector of A with the highest eigenvalue defines the eigenvector centralities of the industry network. The eigenvector centrality is the most appropriate centrality measure for the network of industry-to-industry trade.

Using this equation, an eigenvalue centrality can be computed for each industry. A higher value means the industry is more central and a lower value means the industry is less central. Based on these, we can find central and peripheral industries and build a long-short portfolio to test out hypothesis that central industries react slower and thus are more predictable. We expect that using the same strategy, central industries will generate higher returns than peripheral industries.

Build long-short portfolio

Since the COMB matrix still uses industry code but the stock data uses SIC code, we need to convert industry code to SIC code. BEA provides a text file that gives the corresponding SIC code to each industry code. However, this is not an one-to-one matching. Some industry codes have multiple SIC code. Also, since the file is generated in 1992, it is so poorly structured that we have to write a specific parser to extract information from this file.

The CRSP data contains SIC code, so equal weighted industry portfolios are constructed for every industry. The returns are one-month lagged to prevent peeking into the future. Then, for each central industry, find the returns of all its related industries. The trading strategy is to long the top 10 central industries whose related industries performed well in the past month and short 10 central industries whose related industries performed poorly in the past month. Based on the network effect, an industry will have good returns if its related industries have good returns. We build another long-short portfolio for peripheral industries using the same strategy. The self-financing portfolio is rebalanced every month.

Backtest/ Portfolio value calculation

For each month in the examined period, we searched CRSP data for the survival bias-free constituents of each industry. We only kept the companies that existed throughout the entire month. We then calculated the total return of the month on an equal-weight basis and divided each day's portfolio value by previous day's portfolio value. The resulting 1-based returns of each month are then concatenated into one entire time series and we took the cumulative product to reflect the total return of our industry throughout time.

Likewise for the backtester, using the signals generated according our trading rules at the end of each month, we calculated the equal-weighted total return of given industry portfolios for the following month. We then converted them to 1-based returns and calculated their cumulative product to evaluate the out-of-sample performance between 1992 and 2007.

Results and Discussion

As mentioned above, our first goal is to mimic Noh's paper to see whether the results are the same. After constructing the network and doing the backtest in the same time period as in the paper, we calculated the centrality vector and found the following statistics. The following table shows great resemblance between our results and those from Noh's paper.

Table 1. Comparison between Noh's centrality calculations and our results

	Our Result	**Paper**
Mean	0.037	0.038
Standard Deviation	0.028	0.038
Minimum	0.012	0.01
5th percentile	0.019	0.018
10th percentile	0.021	0.021
25th percentile	0.025	0.025
50th percentile	0.030	0.033
75th percentile	0.040	0.04

90th percentile	0.060	0.056
95th percentile	0.074	0.068
Maximum	0.341	0.347
Number of obs	462	465

For the network, our results of the most and least central industries are also almost the same as the results in the paper; we were able to match 17 out of the 21 industries on average. Detailed comparisons are presented in Appendix B Table 1 and Table 2.

Visualization

With the industry centrality vector and COMB matrix, we were able to determine the location of each industry within the network as well as the strength of all interconnections. We then proceeded to visualize this network using HTML, CSS and JavaScript (D3.js specifically) in a force-directed graph. The strength of the connections are simulated using a physics engine and a stronger connection means a stiffer bond between two industries on the graph. We used the value at location (i,j) of COMB matrix as the magnitude of the force between industry i and j, and filtered out bonds that have a connection strength below median strength. Due to the large number of industries, we only visualized the connection among 20 most central industries and 20 least central industries. The result can be seen in Appendix A Figure 1. Many of the observed patterns seemed fairly reasonable. For instance, the wholesale trade industry is connected to a large number of industries with strong bonds because it processes raw merchandises for many industries such as agriculture, manufacturing, mining and many more. In contrary, industries such as cigars and boots are outliers of the network because their products demand little from other industries and go directly to consumers. An interactive graph can be accessed via this link: http://ading999.github.io/networks

Backtest Performance

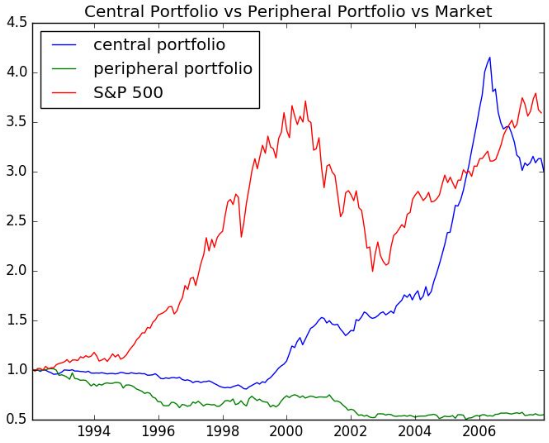

Figure 2. The value of central, peripheral and market portfolios

Between 1992 and 2007, the U.S. stock market as a whole on average increased 0.64% each month, or 7.64% annually, with a Sharpe ratio of 0.15. As expected, out central portfolio outperformed our peripheral portfolio by a significant margin. However, as can be seen in Figure 2 neither portfolios actually outperformed the benchmark. Interestingly enough, both portfolios seemed to have a downward trend between 1992 and 2000, opposite to what we observe of the market. However, after 2000, there seemed to be some kind of regime change, which caused the central portfolio to rise sharply and surpass the market at some point. Information on portfolio returns can be seen in Appendix B Table 3.

Test for significant risk factors

Some classical asset pricing theories consider the risk premium of stock returns to be a linear combination of various risk factors such as market, value (HML), growth (SMB) etc. These theories are well known because they effectively explain the underlying driving force for many assets and tarnish the active asset managers of their stock picking skills. In order to examine whether we truly extracted abnormal returns from our industry networks, we decided to run a statistical test on the residual error of our linear regression results. Preliminary results from the Kolmogorov–Smirnov test showed that our residual errors were in fact not normally distributed, which violates a weak assumption of our linear model as well as Student’s t-test. Thus, we chose to deploy Wilcoxon Signed-Rank test to test for median instead. The null hypothesis is that the residual errors of our returns have a median of zero. The central portfolio turned out with a p-value of 0.39, which fails to reject the null hypothesis for the Fama-French factors. The peripheral had a p-value of 0.0046, which is low enough to reject our null hypothesis. Detailed results can be seen in Appendix B Table 2.

Conclusion

Although the statistical results partially contradict our expectations and literature results, we observed that the returns have really low regression coefficients against the market factor as well as other Fama-French factors. This lead us to believe that the connectedness and difference in speed of information flow can indeed drive stock returns. The undesired results may be that linear models are not suitable for this kind of analysis given the violation of certain assumptions. In the future, we wish to examine this phenomenon further by using methods such as robust linear regression. We should also test our performance under different parameters, such as sample period, rebalance period, portfolio weights and number of industries to trade. In addition, industry classification and portfolio construction can definitely be improved. Last but not least, transaction costs and liquidity risk should also be considered to make our simulations more realistic.

References

Noh, Joonki. Industry Networks and the Speed of Information Flow (November 27, 2014). https://ssrn.com/abstract=2531450 or http://dx.doi.org/10.2139/ssrn.2531450

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. Econometrica, 80(5):1977{2016, Sept. 2012.

D. Aobdia, J. Caskey, and N. B. Ozel. Inter-industry network structure and the cross predictability of earnings and stock returns. Review of Accounting Studies, pages 1{34, Apr. 2014.

K. R. Ahern. Network centrality and the cross section of stock returns. SSRN Scholarly Paper, Dec. 2013.

L. Cohen and A. Frazzini. Economic links and predictable returns. The Journal of Finance, 63(4):1977{2011, Aug. 2008.

Appendix A. Graphs

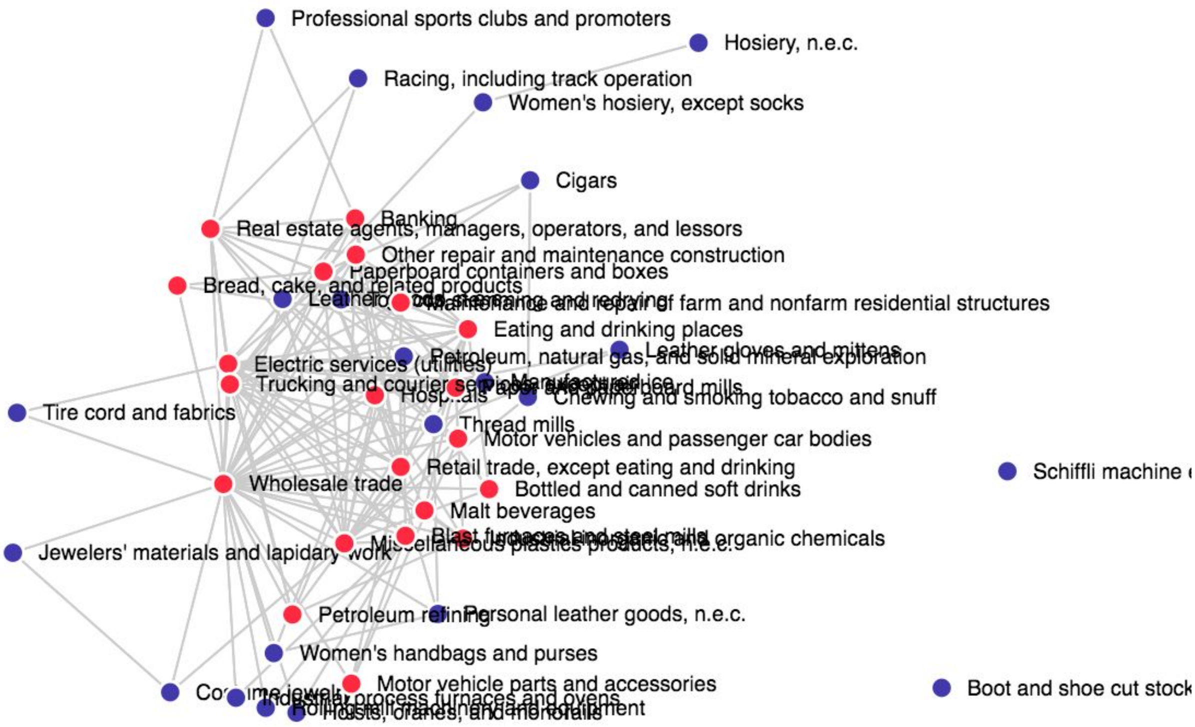
Professional sports clubs and promoters
Hosiery, n.e.c.
Racing, including track operation
Women's hosiery, except socks
Cigars
Banking
Real estate agents, managers, operators, and lessors
Other repair and maintenance construction
Paperboard containers and boxes
Bread, cake, and related products
Maintenance and repair of farm and nonfarm residential structures
Eating and drinking places
Leather gloves and mittens
Petroleum, natural gas, and solid mineral exploration
Electric services (utilities)
Trucking and courier services
Hospitals
Manufactured ice
Paper and paperboard mills
Chewing and smoking tobacco and snuff
Tire cord and fabrics
Thread mills
Motor vehicles and passenger car bodies
Retail trade, except eating and drinking
Schiffli machine
Wholesale trade
Bottled and canned soft drinks
Malt beverages
Jewelers' materials and lapidary work
Miscellaneous plastics products, n.e.c.
Blast furnaces and steel mills
Industrial inorganic and organic chemicals
Petroleum refining
Personal leather goods, n.e.c.
Women's handbags and purses
Motor vehicle parts and accessories
Industrial process furnaces and ovens
Hoists, cranes, and monorails
Boot and shoe cut stock

Figure 1. The connections between 20 most central industries and 20 least central industries, where blue indicates peripheral and red indicates central.

Appendix B. Tables

Table 1. Most Central Industries identify by Noh et. al vs our results (Match Rate: 17/21)

Result in the Paper	**Our Result**
Automotive repair shops and services	Banking
Banking	Blast furnaces and steel mills
Blast furnaces and steel mills	Bottled and canned soft drinks
Bread, cake, and related products	Bread, cake, and related products
Commercial construction industries	Eating and Drinking places
Construction industries	Electric services (utilities)
Eating and Drinking places	Hospitals
Electric services (utilities)	Industrial inorganic and organic chemicals
Hospitals	Maintenance and repair of farm and nonfarm residential structures

Industrial inorganic and organic chemicals	Malt beverages
Miscellaneous plastics products	Miscellaneous plastics products
Motor vehicle parts and accessories	Motor vehicle parts and accessories
Motor vehicles and passenger car bodies	Motor vehicles and passenger car bodies
Paper and paperboard mills	Other repair and maintenance construction
Paperboard containers and boxes	Paper and paperboard mills
Petroleum refining	Paperboard containers and boxes
Real estate agents, managers, and operators	Petroleum refining
Retail trade, except eating and drinking	Real estate agents, managers, and operators
Telephone, telegraph communications and communications	Retail trade, except eating and drinking
Trucking and courier services, except air	Trucking and courier services, except air
Wholesale trade	Wholesale trade

Table 2. Least central industries identify by Noh et. al vs our results (Match Rate: 17/21)

Result in the Paper	**Our Result**
Boot and shoe cut stock and findings	Boot and shoe cut stock and findings
Burial Caskets	Chewing and smoking tobacco and snuff
Chewing and smoking tobacco and snuff	Cigars
Cigars	Costume jewelry
Costume jewelry	Hoists, cranes, and monorails
Hosiery	Hosiery, n.e.c
Jewelers’ materials and lapidary work	Industrial Process Furnaces and ovens
Leather gloves and mittens	Jewelers' materials and lapidary work
Leather goods	Leather gloves and mittens
Manufactured ice	Leather goods
Nonferrous metal ores, except copper	Manufactured ice
Personal leather goods	Personal leather goods
Petroleum, natural gas, solid mineral exploration	Petroleum, natural gas, solid mineral exploration
Professional sports clubs and promoters	Professional sports clubs and promoters
Racing, including track operation	Racing, including track operation

Schiffli machine embroideries	Rolling mill machinery and equipment
Special product sawmills	Schiffli machine embroideries
Tobacco stemming and redrying	Thread mills
Women's handbags and purses	Tobacco stemming and redrying
Women's hosiery, except socks	Women's handbags and purses
X-ray apparatus and tubes	Women's hosiery, except socks

Table 3. Central vs Peripheral vs Market Portfolio performance

1992-2007	Central	Peripheral	Market
Monthly Return	0.29%	-0.59%	0.64%
Annualized Return	3.52%	-7.12%	7.64%
Monthly Sharpe	0.11	-0.21	0.15
Jensen’s Alpha	0.24%	-0.65%	0
Beta	0.09	0.09	1
FF5 Alpha	0.20%	-0.57%	-

Table 4. Regression Coefficients for CAPM and Fama-French Five Factor Model

1992-2007	Central	Peripheral	Market
Jensen’s Alpha	0.24%	-0.65%	0
Beta	0.09	0.09	1
Wilcoxon P-value	0.35	0.0009	-
FF5 Alpha	0.20%	-0.57%	-
FF5 Coeff	0.08636512, 0.12815593, 0.12633646, -0.00437628, -0.11337009	0.05371256, 0.01465303, -0.13792159, -0.01003158, 0.04160498	-
Wilcoxon P-value	0.39	0.0046	-