# Asset Pricing using a Network Approach

Isaza, Esteban[*] jisazac@eafit.edu.co
Restrepo-Tobón, Diego[†] drestr16@eafit.edu.co
Pulido, Sergio[‡] spulidot@eafit.edu.co

March 17, 2019

**Abstract**

This paper approaches the U.S. stock market as a network and explains stocks' returns by taking into account community formation among securities and its centrality inside the network. This approach differs from the ones previously reported in that it analyzes complex systems of connected assets and considers characteristics generally ignored in financial markets.

***Keywords:*** Network theory, Financial Markets

## 1 Introduction

In recent times, researchers have characterized financial markets as networks in which securities correspond to nodes and links that relate to several different features. Some works define these links as a correlation of returns (Peralta and Zareei, 2016; Billio, Getmansky, Lo, and Pelizzon, 2012; Bonanno et al., 2004) , cross-holdings assets (Elliott, Golub, and Jackson, 2014) , overlapping portfolios (Caccioli, Shrestha, Moore, and Farmer, 2014) and volatility (Diebold and Yilmaz, 2015). According to several authors, limited knowledge of the networks formed by financial institutions was one reason why regulators and market participants were not able to identify the circumstances that caused shocks during the 2007–2008 financial crisis (Caccioli et al. (2014); Elliott et al. (2014)). In this regard, systemic phenomena in a financial system can be more easily identified by characterizing its networks and by having a deeper understanding of the connections among the agents within the web.

We depart from other studies by understanding links between assets as correlations that provide structural validity and non-spurious financial interdependence to the network.Therefore, the purpose of this research is to extract centrality measures that determine whether the level of interconnectedness in a network can explain an

[*]Msc Finance Candidate, Universidad EAFIT
[†]PhD Director, Universidad EAFIT
[‡]Analytics Director, Protección SA

asset's returns. This is relevant owing to at least two reasons. First, it allows us to understand the role of assets as an integrating agent. Second, it identifies channels through which firms' interconnections can spread co-movements in the market. Since interconnections in a market can provide opportunities for diversification within a given network, this information is of considerable relevance to investors and policy-makers.

This research investigates the extent to which a financial market, specifically a stock market as a network, can be an effective tool to describe the market as a system defined by interactions between its members. Relying on U.S. data, the obtained factor differ from previous ones because they enable the analysis of a complex system of connected assets after considering their individual features and financial interdependencies.

In the light of previous research and findings about networks in financial markets, this paper attempts to test a hypothesis **(H1)** suggesting that a network's structural metrics capture panel variation in stock returns.

This paper makes a twofold contribution. First, it adds to the empirical evidence of the usefulness of network theory in financial market analysis, thereby contributing to the emerging literature on financial networks. Second, it highlights network-based market analysis, uncovering valuable information about investments by capturing previously unobserved market traits and topologies, such as clusters and peer groups.

The remainder of the paper is organized as follows: Section 2 presents a definition of networks; Section 3 describes the data used in the empirical applications; Section 4 characterizes networks; Section 5 provides an econometric evaluation of the networks obtained in Section 4; and Section 6 offers a conclusion.

# 2 Connecting Networks with Financial Markets

Network theory, a methodology relying on graphs, statistics, and algebra, is used frequently in finance to illustrate different phenomena. We use network theory to represent a set of relationships as a graph that contains nodes (stocks) connected by edges (the financial interdependencies among them). In this research, a network's incidence function is defined as the cross-correlation of assets inside a stock index. Every node in an index has a set of edges connecting it to other stocks inside it.

Let $G = (N, \omega, \gamma)$ be a network composed by a set of nodes $N(G)$; a set of edges $\omega(G)$ connecting pairs of nodes and an incidence function $\gamma$ which associates every edge to a pair of nodes. If a link between nodes $i$ and $j$ exists, we indicate it as $(ij) \in \omega$. The information provided by the network could be rearranged in a $n$ x $n$ adjacency matrix $\Omega = [\Omega_{ij}]$, whose element $\Omega_{ij} \neq 0$ if $(ij) \in \omega$ . Since the links proposed represent a relation with variable strength, our network is said to be weighted, meaning $\Omega_{ij} \in \mathbb{R}$.

Figure 1 shows a network $G = \{N, \omega\}$, comprised by the nodes $N = \{1, 2, ..., 7\}$ and the edges $\omega = \{e_1, e_2, ..., e_7\}$

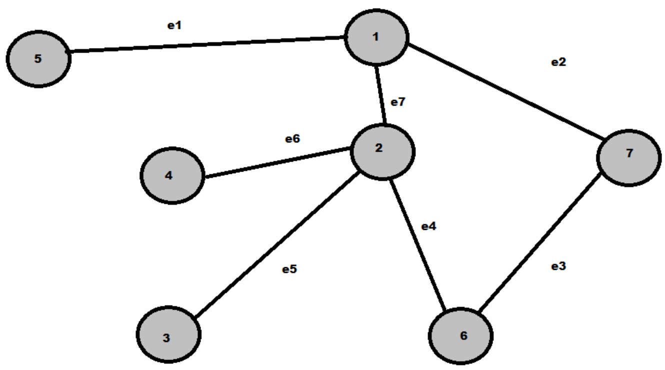


Figure 1: A 7-nodes network

The network is the entire collection of securities held by an agent. The network must allow the comparison of active investors, who select portfolios using securities contained in their information sets, with passive investors, who make no judgments about the stock market value and passively hold indices as their market portfolio (Merton, 1987). To ensure that the network allows such comparisons, it will be designed using publicly available information and an investable universe to permit replications.

# 3 Data

We use the SP 500 Index, a market capitalization-weighted equity index, to represent the U.S. Our data is derived from publicly available stock prices (from Bloomberg) at the end of every trading day, from 2010 to 2017 for 444 stocks from 500 inside the index after controlling for survivorship bias. The SP 500 Index measures the performance of large cap companies, excluding ADRs, ADS, preferred stocks, redeemable shares, warrants, rights, and trust receipts.

# 4 Network characterization

Analyzing an integrated market by means of a network allows us to take into account stocks' individual and systemic dimensions and provides a tool to analyze a market's features and topologies. As Pozzi, Di Matteo, and Aste (2013) suggest, the topology of a network could encode the dependency structure of financial equities.

We use daily returns for every stock in the SP 500 and obtain the correlation coefficient between all possible pairs of stocks inside the investable universe. The resulting matrix is obtained weekly, yielding 416 correlation matrices, one for every week between January 2010 and December 2017. Through simulations, we obtain a threshold correlation that ensures a valid, structural network. Finally, we characterize the 416 networks by using each correlation matrix as an adjacency matrix.

## 4.1 Network structure's validity

Any adjaceny matrix $\Omega$ cannot be immediately considered as a realistic representation of a network. It has been shown that, under some weak constraints, the connections distribution also known as degree distribution of the resulting networks should follow a Power-Law distribution, as seen in real world networks like Internet (Faloutsos, Faloutsos, & Faloutsos, 1999) and co-authorship networks (Newman, 2001). Power-law degree distribution is considered as a significant structural characteristic observed in many real-world complex networks, therefore as a necessary condition, the stock market network we construct should follow a Power-Law degree distribution:

$$p(d) = Pr(X = d) = ad^{-k} \tag{1}$$

Where d represent the quantity whose distribution we are interested in; i.e., a stock's degree (number of relevant financial connections to other entities) (Chattopadhyay & Murthy, 2017). Where $a$ is a normalization constant and $k$ is the Power-Law exponent. The empirical results for our networks should show that $p(d)$ in these networks decays as a Power-Law for large $d$. In other words, a financial network should have just a few nodes with high degree and most other nodes with small degree. In order to ensure this condition, we simulated a set of networks for the S&P 500 looping through several values of the 8-year correlation threshold to label a connection as significant, this is achieved by discarding any connection $\Omega_{ij}$ outside the range specified by the threshold.

Although least squares is a common method for analyzing power-law data, it can produce substantially inaccurate estimates of parameters for this distribution, thus we will follow the statistical framework proposed by Clauset, Shalizi, and Newman (2009) and the package implemented by Alstott, Bullmore, and Plenz (2014).Their approach combines maximum-likelihood fitting methods with a goodness-of-fit test based on the Kolmogorov-Smirnov statistic and likelihood ratios.

First we determine for each $\rho$ what portion of the data to fit. Since a heavy-tailed distribution's interesting feature is the tail and its properties we disregard those small values of the data (stocks with low degree) which do not follow a power law distribution. Clauset et al. (2009) find this optimal value of $d_{min}$ by creating a power law fit starting from each unique value in the data, then selecting the one that results in the minimal Kolmogorov-Smirnov distance, $D$, between the data and the fitted curve.

However, finding a $d_{min}$ does not necessarily implies that a power law is a good description of the degree distribution. Therefore, we evaluate the goodness of fit of a exponential distribution against the power law distributed data hypothesis. For this purpose, we use the loglikelihood ratio $R$ between the candidate distributions. This number will be positive if the data is more likely in the power law distribution, and negative if the data is more likely in the exponential distribution (Alstott et al., 2014).

In order to test the robustness of our results we consider three indexes for the structure's validity analysis, the S&P 500 as mentioned before, the TWSE, and the FTSE 100 Indexes. The TSWE Index is the capitalization-weighted index of all listed common shares traded on the Taiwan Stock Exchange. Finally, the FTSE 100 Index comprises the 100 most highly capitalized blue chip companies listed on the London Stock Exchange weighted by their current market capitalization.

Table 1 summarizes the results for every $\rho$ and every index, it shows that there is no a unique $\rho$ to characterize a network following a power law distribution. In the case of S&P 500 index, the most suitable $|\rho|$ is over 0.5, due to the exhibited combination of a small $D$ statistic and a both statistically significant and positive likelihood ratio. The FTSE index provides evidence of a power law distribution in $|\rho| > 0.6$ , however the data is not statistically significant most likely in this distribution than in an exponential distribution. Finally, the TSWE shows that a $|\rho|$ set over 0.3 retrieves a combination of low D and a significant likelihood ratio.

Table 1: Power-Law distribution parameters under $|\rho|$ thresholds

| S&P 500 | $\|\rho\| > 0.3$ | $\|\rho\| > 0.4$ | $\|\rho\| > 0.5$ | $\|\rho\| > 0.6$ | $\|\rho\| > 0.7$ |
|---|---|---|---|---|---|
| $d_{min}$ | 5 | 3 | 2 | 5 | 3 |
| $k$ | 7.35 | 2.66 | 1.92 | 2.1 | 2.21 |
| $D$ | 0.00 | 0.03 | 0.04 | 0.06 | 0.09 |
| $R$ exponential | -1.47 | -0.48 | 1.75* | 1.36 | 3.35*** |
| Connections | 44898 | 17406 | 5256 | 1830 | 833 |
| FTSE 100 | $\|\rho\| > 0.3$ | $\|\rho\| > 0.4$ | $\|\rho\| > 0.5$ | $\|\rho\| > 0.6$ | $\|\rho\| > 0.7$ |
| $d_{min}$ | 4 | 5 | 6 | 2 | 7 |
| $k$ | 6.28 | 5.6 | 4.03 | 2.22 | 7.13 |
| $D$ | 0.02 | 0.07 | 0.11 | 0.05 | 0.24 |
| $R$ exponential | -1.06 | -0.1 | -0.01 | 0.96 | -5.96 |
| Connections | 2120 | 891 | 341 | 125 | 46 |
| TSWE | $\|\rho\| > 0.3$ | $\|\rho\| > 0.4$ | $\|\rho\| > 0.5$ | $\|\rho\| > 0.6$ | $\|\rho\| > 0.7$ |
| $d_{min}$ | 1 | 1 | 7 | 1 | 1 |
| $k$ | 1.53 | 1.50 | 2.38 | 1.63 | 1.56 |
| $D$ | 0.06 | 0.07 | 0.12 | 0.08 | 0.15 |
| $R$ exponential | 1.83* | 1.08 | 1.75* | 1.78* | 0.01 |
| Connections | 4418 | 731 | 181 | 50 | 10 |

Table 1 allow us to say that our observations for 2 of 3 indices are consistent with the hypothesis that the stocks' degree distribution ($d$) is drawn from a distribution of the form 1 as figure 2 shows. Therefore, it is possible to characterize the adjacency matrix of cross-correlations as a network.

## 4.2 Longitudinal network

Longitudinal networks are simply networks that evolve chronologically. In our context, the 416 weekly networks yield a network that represents the market. In other words, attributed values regarding the number of stocks, correlations, and prices can be understood as a dynamic network. Figure 3 shows two correlation matrices for a sample of 200 stocks in the SP 500, from January 2010 to January 2017. Stock prices resemble a dynamic system because they tend to move together. As stated previously, we seek to assess these movements with a network-based factor.

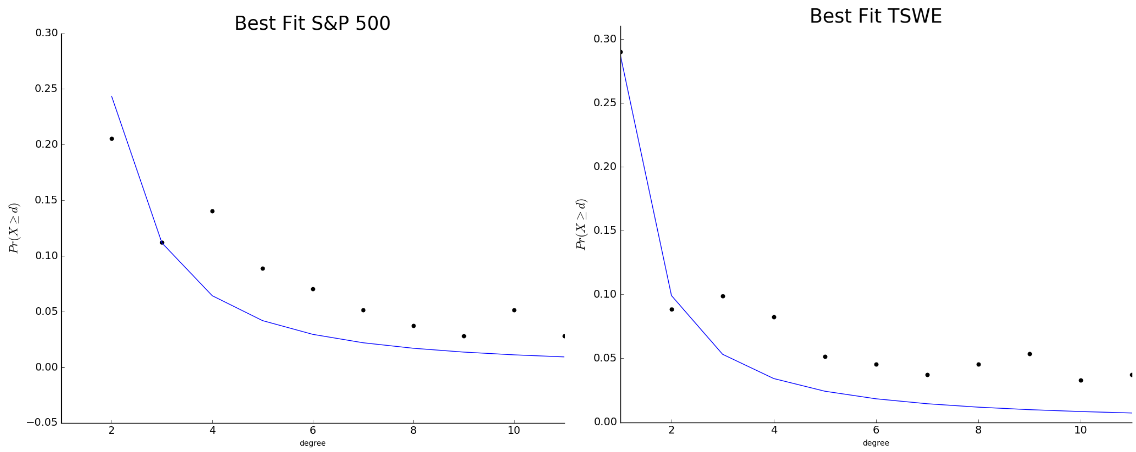


Figure 2: Degree distribution


Figure 3: Correlation Heatmap 2016-2017

An early precedent for the idea represented in figure 3 was provided by King (1966), ), who applied factor analysis to the observed covariance matrix of a series of monthly price changes. His paper, which explained the degree of cross-sectional interdependence exhibited by a series, found that about half of the total variation in a stock's price was accounted for by a market index and an average of 10% was accounted for by industry factors. Thus, it is relevant to quantify the strength of connection and to compare it. However, since the connection can be negative, we rewrite it with a transformation.

We consider relevant to quantify the strength of connection and compare it easily. However as our connection $\rho_{ij}$ can be negative, we will re-express them following this transformation:

$$z_{ij} = 4 - \sqrt{8 \cdot (1 - \rho_{ij})^2} \tag{2}$$

Where $i$ and $j$ are stocks and $z_{ij} \in [0, 4]$, $z_{ij}$ is a proposed translation of the a distance metric used by Bonanno et al. (2004) between a pair of stocks inside a correlation network of equities, our objective is instead to asses the strength of connection.

Once this transformation is carried out on every element of a given matrix we obtain a weighted adjacency matrix $\Omega'$ from which is possible to retrieve a weighted undirected network, since $z_{ij} = z_{ji}$, as fig 4.2 shows. We also use a weighted network besides an unweighted version since we consider fundamental to understand financial interdependencies as weighted relations. Besides, several techniques applied to the study of unweighted networks can be carried over with little modification to weighted networks (Newman, 2004).

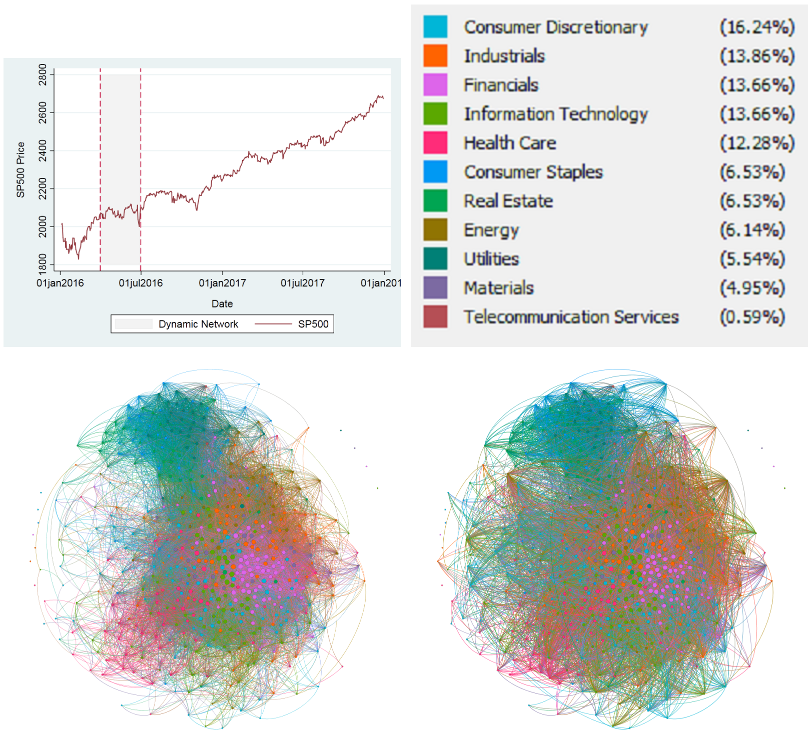


Figure 4: Network's dynamics

A desired result for our work will be the separation of a large set of individual stocks into a smaller set of clusters delivered by the network's topology that tend to move as homogeneous groups. In this paper we interpret clusters as Newman (2010), who defines a cluster as a naturally occurring group in a network regardless of their number or size, which is used primarily as a tool for discovering and understanding the large-scale structure of networks. The clusters that the empirical network could reveal will not be necessarily composed by stocks in the same industry, instead they will belong to the same group provided by the network characteristics.

## 4.3 Local Importance Metrics in a Market

A network's characteristics allow us to describe and quantify a set of properties regarding the importance of entities in the network whose properties attach a different status to each stock. The first network metric obtained is the importance of a stock inside a market, implemented as a value and derived from a historical correlation between assets. In other words, a stock correlating with others inside the network will have high importance. This measurement uses a degree centrality approach independent of network size that easily compares the relative centrality of points from different networks (Freeman, 1978).Thus, we advocate an initial bridge between centrality and the weighting of a stock inside a traditional portfolio. Following this reasoning, we expect that a stock with high centrality in a network will behave like a stock with significant weight inside a market portfolio. This hypothesis follows from the key results in Peralta and Zareei (2016) who designed portfolios from a network perspective in the Markowitz framework.

Assuming that we have n stocks in the network, a given stock, $s_k$, can at most be related with $n-1$ other stocks in the network. Extending this notion to our specific case for S&P 500, the maximum number of relations that each stock can hold is 444, and their centrality will be denoted by:

$$C(s_k) = \frac{\sum_{i=1}^{n} a(s_i, s_k)}{n-1} \tag{3}$$

Where $a(s_k, s_i) = 1$ if stock $k$ holds a filtered correlation with stock $i$ ensuring the power-law distribution. Although a binary representation of edges, the connection between two stocks exist or not, it is quite relevant to additionally include a weighted centrality measure. Following that our network's edges have associated weights $z_{ij}$ keeping record of their strength relative one to another. We denote the weighted centrality for a given stock as:

$$C(s_k) = \frac{\sum_{i=1}^{n} z_{ik}}{n-1} \tag{4}$$

Table 2 indicates the average centrality and highlights that on average, financials exhibit a higher centrality (0.636) than stocks from any other sector, followed by industrials (0.628) and materials (0.621). On the other hand, stocks in the utilities (0.592), telecommunications (0.594), and consumer staples sectors (0.596) show lower centrality levels. Figure 5 shows the time series of the average centrality per sector in the time sample.

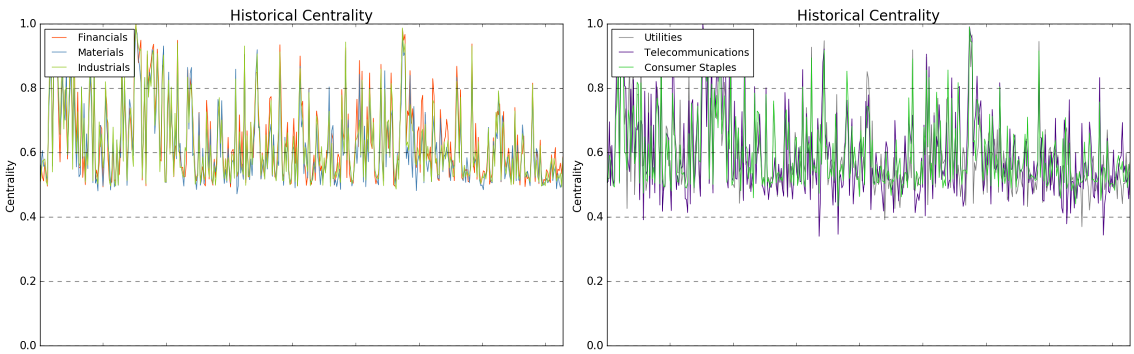


Figure 5: Centrality in Time

In the case of weighted centrality, Financial stocks exhibit higher weighted centrality (0.316) than stocks in any other sector, followed by industrials (0.309) and materials (0.302). Stocks in the utilities (0.265), telecommunications (0.269), and consumer staples (0.275) sectors are similar to those with lower centrality levels. The average values of weighted centrality shown in Table 2, for each sector, are roughly close to half the centrality values. As explained earlier, since these do not control the actual value of correlation, the centrality measure could inflate the importance of a particular stock inside the network.

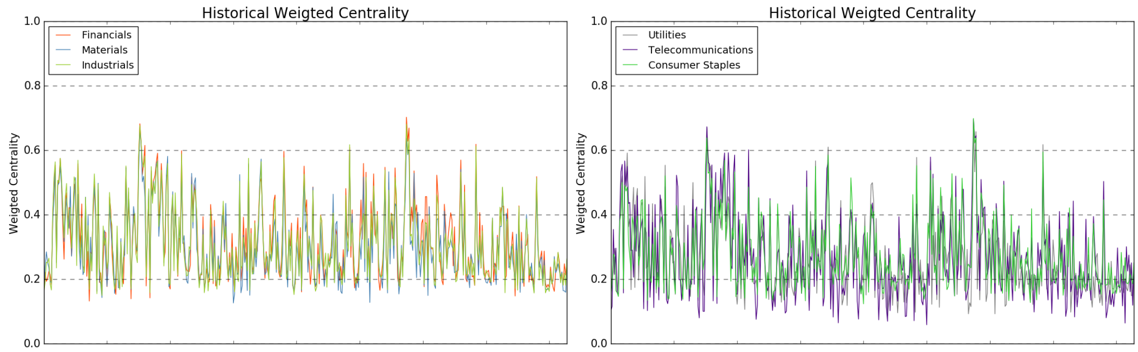


Figure 6: Weighted Centrality in Time

## 4.4 Global importance metrics in a market

However, as Opsahl, Agneessens, and Skvoretz (2010) stated, node strength could be a blunt measure as it only takes into consideration an equity's total level of involvement in the network, and not the number of other equities that it connects with. For this reason, we do not limit the importance analysis to just local metrics.

Hence, these two formerly mentioned measures, centrality and weighted centrality (i.e. the number and weight of links, respectively), lack the abilities to address the global features of a network; since the S&P index member's centrality is not accounted as a centrality inducer. To achieve this, we will rely on the eigenvector centrality, a global and non-path-based measure of centrality, where the centrality of an equity is directly related to the sum of the centralities of its adjacent equities (León & Pérez, 2014). Hence, a stock's centrality will be the weighted sum of centrality at all possible order adjacencies. Hence, the centrality of one equity will arise from (i) being related to many others; (ii) being related to central stocks; (iii) or both (Wasserman & Faust, 1994). Soramäki and Cook (2013) provide a simple interpretation for this measure: eigenvector centrality may be understood as the proportion of time spent in each stock in an infinite random walk through the network.

We decompose the adjacency matrix to obtain the eigenvector centrality for each stock. With $\Omega$ as the adjacency matrix (either weighted or binary), $\Lambda$ the diagonal matrix containing the $\Omega$ matrix eigenvalues , and $\Gamma$ an orthogonal matrix required that $\Gamma\Gamma' = \Gamma\Gamma = I_n$ such that:

$$\Omega = \Gamma\Lambda\Gamma' \tag{5}$$

After sorting the diagonal matrix of eigenvalues $\Lambda$ so that $\lambda_1 \geq \lambda_2 ... \lambda_n$ the first column in $\Lambda$ corresponds to the principal eigenvector of $\Gamma$ The principal eigenvector $(\Gamma_1)$ is considered as the leading vector of the system, the one that is able to explain

Table 2: Overall statistics by sector

| Sector | | S&P 500 | | |
|---|---|---|---|---|
| | % Sample | Weigthed Centrality | Eigencentrality | Centrality |
| Industrials | 13.7% | 0.309 | 0.0445 | 0.628 |
| Financials | 13.4% | 0.316 | 0.0443 | 0.636 |
| Materials | 4,8% | 0.302 | 0.0442 | 0.621 |
| Information | 13.6% | 0.298 | 0.0446 | 0.617 |
| C. Discretionary | 16.1% | 0.285 | 0.0444 | 0.604 |
| Telecommunication | 0.6% | 0.269 | 0.0441 | 0.594 |
| Real Estate | 6.9% | 0.290 | 0.0445 | 0.614 |
| C. Staples | 6.2% | 0.275 | 0.0441 | 0.596 |
| Health Care | 12.8% | 0.286 | 0.0443 | 0.606 |
| Energy | 6,1% | 0.284 | 0.0445 | 0.603 |
| Utilities | 5.8% | 0.265 | 0.0445 | 0.592 |

the most of the underlying system, in which the positive n-scaled scores corresponding to each element may be considered as their weights within a network (León & Pérez, 2014).

Given that the largest eigenvalue and its corresponding eigenvector provide the biggest explanatory power to reproduce the initial matrix, $\Gamma_1$ it is considered a global measure of centrality within a network, due to its ability to capture the main features of networks Bonacich (1972). The eigenvector centrality framework allows to capture the impact of an specific connection pattern on a global scale. Therefore, all stocks that are connected to the most important nodes, either directly or indirectly, inherit some degree of eigencentrality, after introducing the intensity of the connections. The average for global importance metrics are also presented in table 2 for each sector, figure 7 presents the time series for this metric separated by sector as well, the figure indicates a higher volatility in eigencentrality in those sectors with a low local centrality measure.

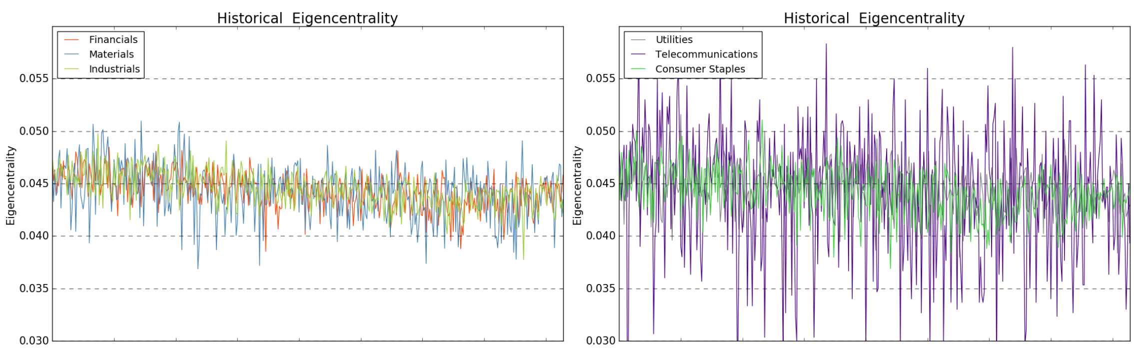


Figure 7: Eigencentrality in Time

# 5 Econometric Evaluation

Although a network provides a useful representation, the statistical properties of links connecting different entities of the system are really the solid ground used to prove our propositions about networks properties in stock markets, going farther than just relying in descriptive statements.

A number of approaches are available to calculate the return of a given security. We will locate in the factor models framework, which seeks to reduce the variance of the return by explaining its variation using additional factors provided by the network's topology. We are indeed considering the possibility that there are influences beyond the market that cause stock prices to move.

First, following closely the design of Fama and French (1993) we will evaluate the relevance of a factor regarding the differences exhibited by stocks in their centrality measures. We define *HLC* (High-Low Centrality) as the return of a equally weighted portfolio comprised of stocks with high centrality (top 80%) minus the return of the portfolio composed by stocks with low centrality (lowest 20%). The addition of this factor is induced by the results exhibited in table 3, which do an assessment of the differences in historical returns of stocks belonging to one of these two portfolios for every measure of importance proposed in this paper. Table 3 shows that stocks classified as assets with high weighted centrality and high eigencentrality tend to exhibit in average statistically significant higher returns (with 1% and 5% significance levels, respectively) that those stocks with low weighted centrality and low eigencentrality measures.

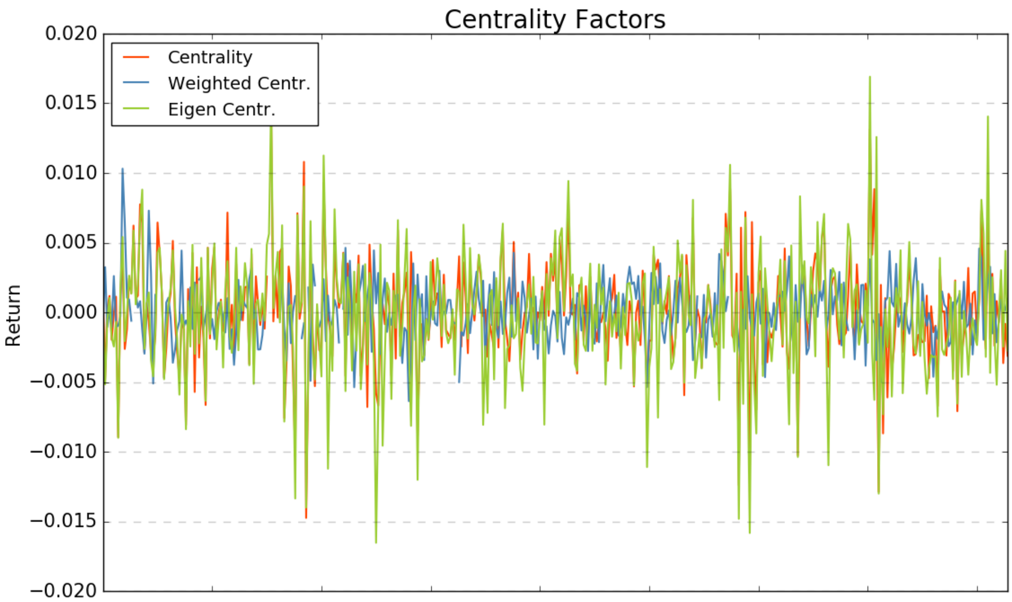


Figure 8: HLC Factor in Time

Additional factors commonly considered in applications are portfolios of traded securities and industry indexes to capture common movement between securities. This paper departs from this stream by considering factors as portfolios, selected trough centrality ranking or resulting from the clustering topology of the network.

Table 3: HLC Portfolios in S&P 500

| Centrality | | High | Low | p-value |
|---|---|---|---|---|
| N | | 36571 | 76248 | |
| Stock's return % | | 0.3 $(-1.4, 2.1)$ | 0.3 $(-1.7, 2.3)$ | 0.2 |
| Sector | C. Discretionary | 14.4% | 16.7% | <0.01 |
| | C. Staples | 5.0% | 6.8% | |
| | Energy | 3.9% | 6.7% | |
| | Financials | 17.3% | 12.5% | |
| | Health | 13.0% | 12.2% | |
| | Industrials | 17.1% | 13.4% | |
| | Information Techn. | 16.9% | 13.1% | |
| | Materials | 5.7% | 4.9% | |
| | Real Estate | 4.7% | 7.0% | |
| | Telecommunication | 0.5% | 0.7% | |
| | Utilities | 1.5% | 0.7% | |
| **Weighted Centrality** | | **High** | **Low** | **p-value** |
| N | | 37876 | 76878 | |
| Stock's return % | | 0.4 $(-1.5, 2.1)$ | 0.3 $(-1.7, 2.3)$ | <0.01 |
| Sector | C. Discretionary | 16.1% | 16.1% | <0.01 |
| | C. Staples | 6.1% | 6.3% | |
| | Energy | 6.2% | 6.1% | |
| | Financials | 13.4% | 13.5% | |
| | Health | 12.3% | 12.9% | |
| | Industrials | 13.8% | 13.4% | |
| | Information Techn. | 14.0% | 13.4% | |
| | Materials | 4.7% | 4.9% | |
| | Real Estate | 7.1% | 6.9% | |
| | Telecommunication | 0.6% | 0.7% | |
| | Utilities | 5.7% | 5.8% | |
| **Eigencentrality** | | **High** | **Low** | **p-value** |
| N | | 28810 | 73209 | |
| Stock's return % | | 0.4 $(-1.4, 2.2)$ | 0.3 $(-1.5, 2.2)$ | <0.05 |
| Sector | C. Discretionary | 14.4% | 16.7% | <0.01 |
| | C. Staples | 5.0% | 6.8% | |
| | Energy | 3.9% | 6.7% | |
| | Financials | 17.3% | 12.5% | |
| | Health | 13.0% | 12.2% | |
| | Industrials | 17.1% | 13.4% | |
| | Information Techn. | 16.9% | 13.1% | |
| | Materials | 5.7% | 4.9% | |
| | Real Estate | 4.7% | 7.0% | |
| | Telecommunication | 0.5% | 0.7% | |
| | Utilities | 1.5% | 0.7% | |

We will consider this factor as an additional source of covariance between securities in a equation for risk and return. We hypothesize that the return of any stock is a linear function of the return of the market and the factor capturing the performance of these centrality constructed portfolios, as the figure 8 shows, the HLC factor behaves in time as a risk factor. Hence, we propose the the following expression:

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i} HLC_t + \beta_{2i} R_{mt} + \varepsilon_{it} \tag{6}$$

where $R_{it}$ is the return on the asset $i$ (i=1,.., N) at time $t$, $R_{ft}$ is the risk-free return and $\varepsilon_{it}$ is the idiosyncratic return with mean zero and covariance matrix $\Sigma_\varepsilon$. If the model 6 is well-specified, the equilibrium pricing equation according to Shanken (1992) would be

$$E(r_{i,t}) = \gamma' \beta_i \tag{7}$$

where $\gamma$ is a (2x1) risk premia vector, $r_{i,t}$ is the return of asset i in excess of the riskless rate of return. The pricing equation 7 has the restriction that the intercepts should be zero. Thus, if $\alpha_i = 0$ for all assets, this would be evidence that supports the validity of our factor model. We use the GRS test, a joint F-tests developed by Gibbons, Ross, and Shanken (1989), to test whether the intercept estimate of the time series regression 6 are jointly different from zero. Thus GRS F-statistic test the joint hypothesis $H_0 : \alpha = (\alpha_1, \alpha_1, ..., \alpha_N)'$, GRS suggest the following test statistic for the null hypothesis:

$$W = \left[\frac{T(T-N-K)}{N(T-K-1)}\right] \frac{\hat{\alpha}' \hat{\Sigma_\varepsilon^{-1}} \hat{\alpha}}{1 + \hat{\vartheta}^2} \sim F(N, T-N-K) \tag{8}$$

where $\hat{\Sigma_\varepsilon^{-1}}$ is the residual covariance matrix assuming that the error terms are jointly normally distributed with mean zero, $\hat{\vartheta}^2 = \bar{F}' \hat{\Sigma}_F^{-1} \bar{F}$, with $\bar{F}$ the ($K$x1) being the vector of average returns of the factor portfolios. To test the validity of 6 we proceed as follows: i) From the SP500 we construct equally weighted portfolios with its members chosen randomly , ii) We estimate 6, iii) We apply the the GRS test and finally iv) we repeat this process 10 thousand times for several portfolios sizes. The results are presented in the table 4 for every underlying measure used for the three instances of the HLC factor (centrality, weighted centrality and eigencentrality).

Table 4: Simulated portfolios and Centrality Factor HLC

| **Underlying Measure** | | $n=5$ | $n=10$ | $n=15$ | $n=20$ | $n=25$ |
|---|---|---|---|---|---|---|
| Centrality | *** | 91.90% | 87.7% | 91.3% | 93.8% | 93.1% |
| | ** | 96.80% | 95.0% | 97.1% | 98.2% | 98.1% |
| | * | 99.94% | 99.7% | 99.7% | 99.8% | 99.8% |
| Weighted Centrality | *** | 90.7% | 89.6% | 87.1% | 93.3% | 93.8% |
| | ** | 96.8% | 96.8% | 95.4% | 97.4% | 98.1% |
| | * | 99.8% | 99.3% | 99.2% | 99.9% | 99.9% |
| Eigencentrality | *** | 88.3% | 89.7% | 91.3% | 94.9% | 94.1% |
| | ** | 94.4% | 96.6% | 95.7% | 98.6% | 97.8% |
| | * | 99.6% | 99.6% | 99.7% | 99.9% | 100% |

*** p<0.01, ** p<0.05, * p<0.1

Table 4 should be read as follows: The simulated portfolios with 5 randomly chosen stocks showed that with a p-value of 0.01 the 91.9% of them accepted the HLC factor as a significant covariate. If the rest of traditional significance levels these values rise as the exigence loosen, up to 96.8% and 99.94%, for 5% and 10% respectively.

Summarizing, the results show that after simulating 10000 portfolios of 5, 10, 15,20 and 25 stocks chosen randomly, the centrality is significant as a factor explaining stocks returns in approximately 90% of constructed portfolios. Thus, we find evidence that hypothesis 1 holds true in this sample under the GRS F-test.

# 6 Robustness

We test the robustness of our primary results applying the GRS F-test to the following specification:

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i} HLC_t + \beta_{2i} R_{mt} + \beta_{3i} HML + \beta_{3i} SML + \varepsilon_{it} \tag{9}$$

This specification introduces as additional factors those developed in Fama and French (1993) seeking to test if the HLC factor is still significant after having to compete for relevance inside the model with two commonly used and recognized factors.

Table 5: Simulated portfolios including Fama and French Factors

| **Underlying Measure** | | $n=5$ | $n=10$ | $n=15$ | $n=20$ | $n=25$ |
|---|---|---|---|---|---|---|
| Centrality | *** | 91.84% | 86.92% | 89.7% | 91.8% | 92.4% |
| | ** | 96.12% | 94.12% | 95.7% | 96.9% | 97.6% |
| | * | 99.76% | 99.54% | 94.6% | 98.1% | 98.3% |
| Weighted Centrality | *** | 88.3% | 87.6% | 87.4% | 92.5% | 91.5% |
| | ** | 96.8% | 96.8% | 95.4% | 97.4% | 96.4% |
| | * | 99.4% | 99.1% | 98.8% | 99.5% | 99.1% |
| Eigencentrality | *** | 86.7% | 89.1% | 91.0% | 93.2% | 92.2% |
| | ** | 94.1% | 96.2% | 94.9% | 98.2% | 97.5% |
| | * | 98.9% | 99.1% | 99.1% | 99.5% | 99.9% |

*** p<0.01, ** p<0.05, * p<0.1

Table 5 shows a reduction in portfolios with significant HLC factor across every instance of the simulation. Nevertheless, these decreases are marginal and in our opinion do not erode the results supporting evidence for our hypothesis.

# 7 Conclusions

After conceiving the components of the S&P 500 as a network, this paper developed the *HLC* factor which took into account the phenomenon of community formation among securities under its design. The developed factor differs from previously existing ones since it provides fresh insight for representing and analyzing complex systems of connected assets, taking into account usually unconsidered factors in a financial market.

This was achieved first, showing the feasibility of characterizing an adjacency matrix of cross-correlations as a network, after finding supporting evidence that the correlation structure of the S&P is consistent with the hypothesis that stocks' degree distribution is drawn from a power-law distribution. Secondly, we defined the $HLC$ factor (High-Low Centrality) as the return of a equally weighted portfolio comprised of stocks with high centrality (top 80%) minus the return of the portfolio composed by stocks with low centrality (lowest 20%).The $HLC$ factor comes from the separation of a large set of individual stocks into a smaller set of clusters delivered by the network's topology summarized through the centrality measures that tend to move as an homogeneous groups.

The aforementioned process showed that stocks classified as assets with high weighted centrality and high eigencentrality tend to exhibit in average higher returns (with 1% and 5% significance levels, respectively) that those stocks with low weighted centrality and low eigencentrality measures.

Finally, we implemented the GRS F-test on several simulated portfolios, selecting assets randomly and looping over portfolio's size, to determine the statistically significance of the HLC factor. The results show at a significance level of 1% that approximately 90% of $S\&P500$ simulated portfolios exhibits the centrality factor as a valid covariate of the proposed econometric specification. Hence, we find evidence that the centrality measure of a stock is a significant factor explaining its return. Furthermore, a robustness check was implemented introducing additional factors in the simulations to check if the explanatory power of the HLC held relevant. The results show that the HLC factor explanatory relevance was slightly diminished about a 1% in average for every portfolio simulated under the three centrality measures implemented.

The implications of our results are many. First, our results add to a body of literature on the network theory applications to financial networks. Besides, our findings also support findings in literature that stocks tend to move as homogeneous groups. Second, our results indicate that evidence from a single market is likely generalizable across countries.

## References


Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, *9*(1), e85777.

Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, *104*(3), 535–559.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, *2*(1), 113–120.

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B*, *38*(2), 363–371.

Caccioli, F., Shrestha, M., Moore, C., & Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, *46*, 233–245.

Chattopadhyay, S., & Murthy, C. (2017). Generation of power-law networks by employing various attachment schemes: Structural properties emulating real world networks. *Information Sciences*, *397*, 219–242.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661–703.

Diebold, F. X., & Yilmaz, K. (2015). Trans-atlantic equity volatility connectedness: Us and european financial institutions, 2004–2014. *Journal of Financial Econometrics*, *14*(1), 81–127.

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, *104*(10), 3115–53.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the internet topology. In *Acm sigcomm computer communication review* (Vol. 29, pp. 251–262).

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215–239.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

King, B. F. (1966). Market and industry factors in stock price behavior. *the Journal of Business*, *39*(1), 139–190.

León, C., & Pérez, J. (2014). Assessing financial market infrastructures' systemic importance with authority and hub centrality. *Journal of Financial Market Infrastructures*, *2*(3), 67–87.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The journal of finance*, *42*(3), 483–510.

Newman. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, *98*(2), 404–409.

Newman. (2004). Analysis of weighted networks. *Physical review E*, *70*(5), 056131.

Newman. (2010). *Networks: an introduction.* Oxford university press.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, *32*(3), 245–251.

Peralta, G., & Zareei, A. (2016). A network approach to portfolio selection. *Journal of Empirical Finance*, *38*, 157–180.

Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, *3*, 1665.

Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies*, *5*(1), 1–33.

Soramäki, K., & Cook, S. (2013). Sinkrank: An algorithm for identifying systemically important banks in payment systems. *Economics: The Open-Access, Open-Assessment E-Journal*, *7*(2013-28), 1–27.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.