

RESEARCH

Partial Correlation Financial Networks

Tristan Millington* and Mahesan Niranjan

*Correspondence:
t.millington@soton.ac.uk
School of Electronics and
Computer Science, University of
Southampton, University Road,
Southampton, UK

Abstract

Correlation based networks have been a popular way of inferring a financial network due to the simplicity of construction and the ease of interpretability. However two variables which share a common cause can be correlated, leading to the inference of spurious relationships. To solve this we can use partial correlation. In this paper we construct both correlation and partial correlation networks from S&P500 returns and compare and contrast the two. Firstly we show that the partial correlation networks have a smaller and much less variable intensity than the correlation networks, but in fact are less stable. We look at the centrality of the various sectors in the graph using degree centrality and eigenvector centrality, finding that sector centralities move together during the 2009 market crash and that the financial sector generally has a higher mean centrality over most of the dataset. Exploring the use of these centrality measures for portfolio construction, we show there is mild correlation between the in-sample centrality and the out of sample Sharpe ratio but there is negative correlation between the in-sample centrality and out of sample risk. Finally we use a community detection method to study how the networks reflect the underlying sector structure and study how stable these communities are over time.

Keywords: correlation network; financial networks; portfolio optimization; partial correlation; covariance estimation

1 Introduction

Financial markets are a critical part of modern economies and help us to understand the current consensus on how various constituent companies are performing. Asset prices are an equilibrium reached by the interactions between external information, historic performance and trader sentiments. Because of this, they are widely regarded as incredibly complex systems. Networks are a popular method of representing a complex system due to their ability to express interactions or relationships between components in a simple model that is widely applicable to many different fields. Networks inferred from these disparate fields share surprisingly common properties, including heavy tailed degree distributions, high clustering coefficients and community structure [1] [2]. Once constructed, the networks can be used for a wide variety of applications, for instance portfolio selection [3] [4], stability analysis [5] and to detect attempts to manipulate stock prices [6] [7].

A common and popular way of constructing a financial network is to infer it from returns data following the seminal work of Mantenga [8]. In this paper the author uses Pearson's correlation coefficient between the returns to weight the edge between two companies and constructs a minimum spanning tree from the resulting graph, finding companies in similar sectors are clustered in the tree. Onella et al [9] further investigate this, studying the effects of market crashes on the minimum

spanning tree. They find the length of the tree decreases in times of market disruption (i.e. things become more correlated). Boginski et al [10] create a network from the correlation matrix inferred from financial returns data and investigate what happens to the structure of the network when a threshold is set. Any correlation coefficients with an absolute value below the threshold are set to zero and any above are set to one. This model is applied to the US stock market, and they find that the degree distribution of the networks resembles a power law if the threshold is set to a sufficiently large value. Furthermore they also construct maximal independent sets from the network in an attempt to create diversified portfolios. Other authors have applied this model to the Russian [11], Chinese [5] and British [12] stock markets. These models have also been applied to other assets, including cryptocurrencies [13].

An important question is whether these correlation networks contain real information or whether they are merely picking up noise present at that moment of time. Plerou et al [14] use random matrix theory to analyze correlation matrices inferred from stock returns. By studying the eigenvalues and eigenvectors of the correlation matrix they obtain several interesting results. Firstly, only a few of the eigenvalues and eigenvectors contained in the correlation matrix differ significantly from those obtained from a random matrix. This implies many of the relationships in the correlation matrix are noise. However the largest eigenvalue tends to differ significantly, being an order of magnitude larger than the second largest. The eigenvector that corresponds to this eigenvalue tends to contain information that affects all stocks (for instance interest rate increases), with its components being significantly different from those obtained from a random matrix. They also study the next few eigenvectors corresponding to the next largest eigenvalues and find these tend to have significant values for related stocks - for instance those in the same sector or who have business in similar regions. Finally they show the eigenvector that corresponds to the largest eigenvalue tends to be preserved for correlation matrices inferred from data across different times, indicating it is stable.

While using the correlation coefficient does give a simple, interpretable model there are downsides. Two variables which share a common cause can be correlated, which could be considered a false relationship. We can use partial correlation in this context to attempt to remove these indirect correlations. Kennet et al [15] create a form of partial correlation network they term a 'dependency network'. Calculating partial correlations by removing one variable at a time, they subtract this off the correlation coefficient between two variables. This gives the contribution that a particular company has on the correlation between two others. These networks are directed and help us to understand which companies are influential in the market. They find that the financial sector is very influential in the US stock markets and its influence is maintained throughout the period of study.

Another example of the use of partial correlation is used by Wang et al [16] who compare minimum spanning trees constructed using the correlation and partial correlation coefficients inferred from various stock indices across the world. They calculate partial correlation by inverting the correlation matrix and find that the centrality structure in the minimum spanning tree constructed from the partial correlation matrix is more useful than that constructed from the correlation matrix,

with the USA, Germany and Japan clearly serving as hubs. Other authors have used sparse precision matrix estimators to infer these networks, for instance the graphical lasso [17] [4] and SPACE [18]. These sparse methods do however tend to produce quite unstable networks, possibly due to the instability of the lasso with highly correlated data [19] [20].

A covariance matrix can be used to estimate both correlation and partial correlation matrices. We can obtain the correlation matrix (C) from the covariance matrix (Σ) by normalizing the off-diagonal entries:

$$C_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} \quad (1)$$

Obtaining the partial correlation matrix can be done in a similar manner. We scale the off diagonal entries of the inverse of the covariance matrix (Θ), also called the precision matrix, to acquire the partial correlation matrix (P)

$$P_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} \quad (2)$$

Calculating the partial correlation matrix therefore requires that the covariance matrix is invertible. This is not always the case with financial data due to the presence of outliers and the issue of having more dimensions than samples. To solve these issues we use the Ledoit-Wolf Covariance shrinkage method [21]. This produces a well-regularized covariance matrix by combining the sample covariance matrix with the identity to reduce the off-diagonal values. This well formed covariance matrix can then be inverted to obtain the precision matrix. With these regularized covariance and precision matrices, we can use the equations above to obtain the correlation and partial correlation networks. Once obtained, we can then compare and contrast said networks. The shrinkage method is described further in section 2.

We hope that by using partial correlation we can discover latent relationships that are hidden in correlation networks by the overall movement of the market. Previous work has focused on removing the market mode from the data by various methods, including deleting the largest eigenvalue and eigenvector, or using a factor model [22] but to our knowledge a partial correlation approach like the one we take has not been used on financial data before. In our approach, the partial correlation between two variables is the correlation between the two once the linear effects of all other variables have been removed. To gain further insight, first consider that we can infer row i of a precision matrix from a dataset X via regressing variable i (\mathbf{x}_i) on the others (X_{-i})

$$\boldsymbol{\beta}^* = \arg \min_{\boldsymbol{\beta}} \|\mathbf{x}_i - X_{-i}\boldsymbol{\beta}\|_2^2 \quad (3)$$

The solution to this is

$$\beta_{ij}^* = -\frac{\Theta_{ij}}{\Theta_{ii}} = P_{ij} \sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}} \quad (4)$$

This means that the partial correlation between i and j (P_{ij}) is proportional to the weight that the least squares method would assign to j in a regression problem if we tried to predict i from the rest of the dataset (or since $P_{ij} = P_{ji}$ the weight that the least squares method would assign to i if we tried to predict j from the dataset)

Furthermore the precision matrix is intimately connected with the minimum variance portfolio. The problem, formulated as [23]

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \Sigma \mathbf{w} \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{w} = 1 \end{aligned} \quad (5)$$

has a solution of for one corner of the optimal mean-variance returns frontier

$$\mathbf{w}^* = \frac{1}{\mathbf{1}^T \Theta \mathbf{1}} \Theta \mathbf{1} \quad (6)$$

where $\mathbf{1}$ is a vector of all 1s and w_i is the amount to be invested in asset i . Due to the relationship between the precision matrix (Θ) and partial correlation (P) (see equation 2) we expect these networks will give insight into the benefits and drawbacks of these portfolios, however it is important to remember a partial correlation matrix discards the diagonal of the precision matrix

2 Ledoit-Wolf Covariance

Ledoit-Wolf covariance is based upon shrinkage, where we combine the sample covariance matrix S (which may have a high variance but low bias) with a known matrix with desirable properties (which has a low variance but high bias). The usual choice for this is the identity matrix I and we create a linear combination of the two

$$\Sigma_{\text{lw}} = (1 - \rho)S + \rho \text{tr}(S)I \quad (7)$$

To decide ρ we wish to minimize the Frobenius norm of the difference between Σ_{lw} and the true population covariance matrix Σ_*

$$\min_{\rho} E[||\Sigma_* - \Sigma_{\text{lw}}||_F^2] \quad (8)$$

The optimal solution of ρ is

$$\rho = \frac{E[||S - \Sigma_*||_F^2]}{E[||S - \text{tr}(S)I||_F^2]} = \frac{\beta^2}{\delta^2} \quad (9)$$

The interpretation here is that if S is very close to Σ_* (i.e. our estimate of the covariance is good) then we do not need to shrink much, or if our shrinkage choice does not seem accurate then we should not shrink much either. However the obvious flaw so far is that we need to know the true population covariance matrix to obtain the correct value for ρ - and if we did then we would not need to bother estimating it to begin with! We therefore require estimates of β^2 and δ^2 . We can estimate δ^2 as following:

$$\hat{\delta}^2 = ||S - \text{tr}(S)I||_F^2 \quad (10)$$

and β^2 as

$$\hat{\gamma}^2 = \frac{1}{n^2} \sum_{k=1}^n \|x_k x_k^T - S\|_F^2 \quad (11)$$

$$\hat{\beta}^2 = \min(\hat{\delta}^2, \hat{\gamma}^2) \quad (12)$$

$$\hat{\rho} = \frac{\hat{\beta}^2}{\hat{\delta}^2} \quad (13)$$

The constraint on $\hat{\beta}^2$ ensures that $\rho < 1$. While it is rarely necessary, it does help stop us accidentally making our estimate less well formed.

The Ledoit-Wolf method is guaranteed to give us a positive-definite invertible matrix, which is critical in this application as we require the inverse of the covariance matrix (the precision matrix) to acquire the partial correlation matrix.

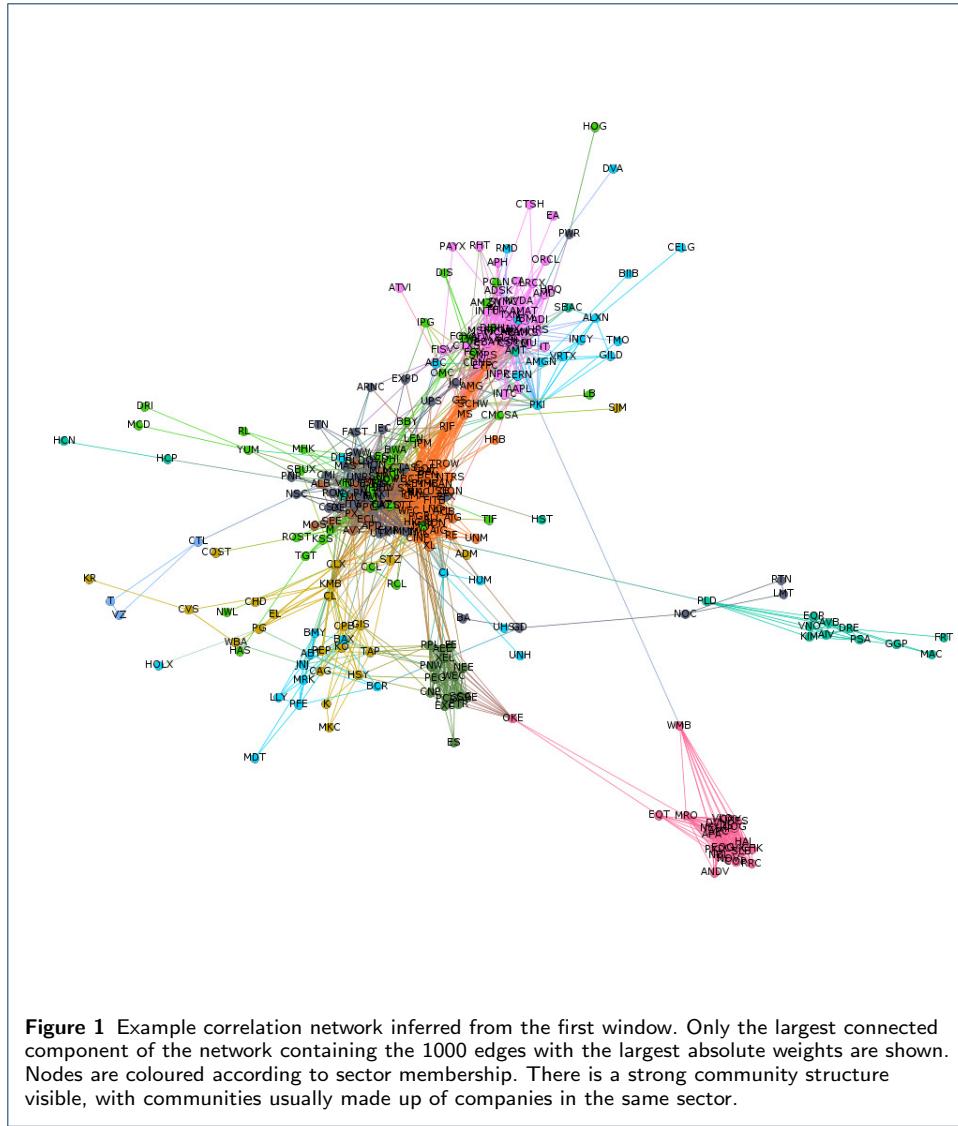
This covariance estimation method has been applied to the genomics field [24], for portfolio optimization [21] and in the neuroscience field [25] but to our knowledge has not actually been applied to create financial networks.

3 Data and Software

For our study we use daily log returns from the S&P500. If there is less than 10% of the data missing for a particular stock we fill it with the price from the previous day, or if the data is missing from the start, from the first day when the stock is traded. If there is more than 10% missing we discard the data for that stock. We use the close price on the day to calculate the return, from 2000-01-03 to 2017-12-05. Overall we have 4510 days of return data for 345 stocks. Since financial data is non stationary we use a window of 300 days and slide along this 30 days at a time to obtain a sample where we can assume the data is stationary, giving us 140 windows overall. The returns inside each window are normalized using the z-score to have a mean of 0 and a standard deviation of 1. While correlation is by definition normalized, this procedure is mostly for the benefit of the shrinkage procedure - normalizing reduces the amount of shrinkage required which allows us to capture more relationships.

Using this dataset we infer a network for each window by using the Ledoit-Wolf shrinkage methods to obtain a covariance matrix and inverting it to obtain a precision matrix. We then scale both of these matrices appropriately using equations 1 (Correlation) and 2 (Partial Correlation) to create the correlation and partial correlation matrices and use these as adjacency matrices to construct the networks. We then study the properties of these networks and how they change over time.

We make use of Python, NumPy and SciPy [26] for general scripting, pandas [27] for handing the data, sklearn [28] for the implementation of the Ledoit Wolf estimation methods, statsmodels [29] for some of the statistical analysis, matplotlib [30] for plotting, Networkx [31] for the network analysis and gephi [32] for some graph visualization.

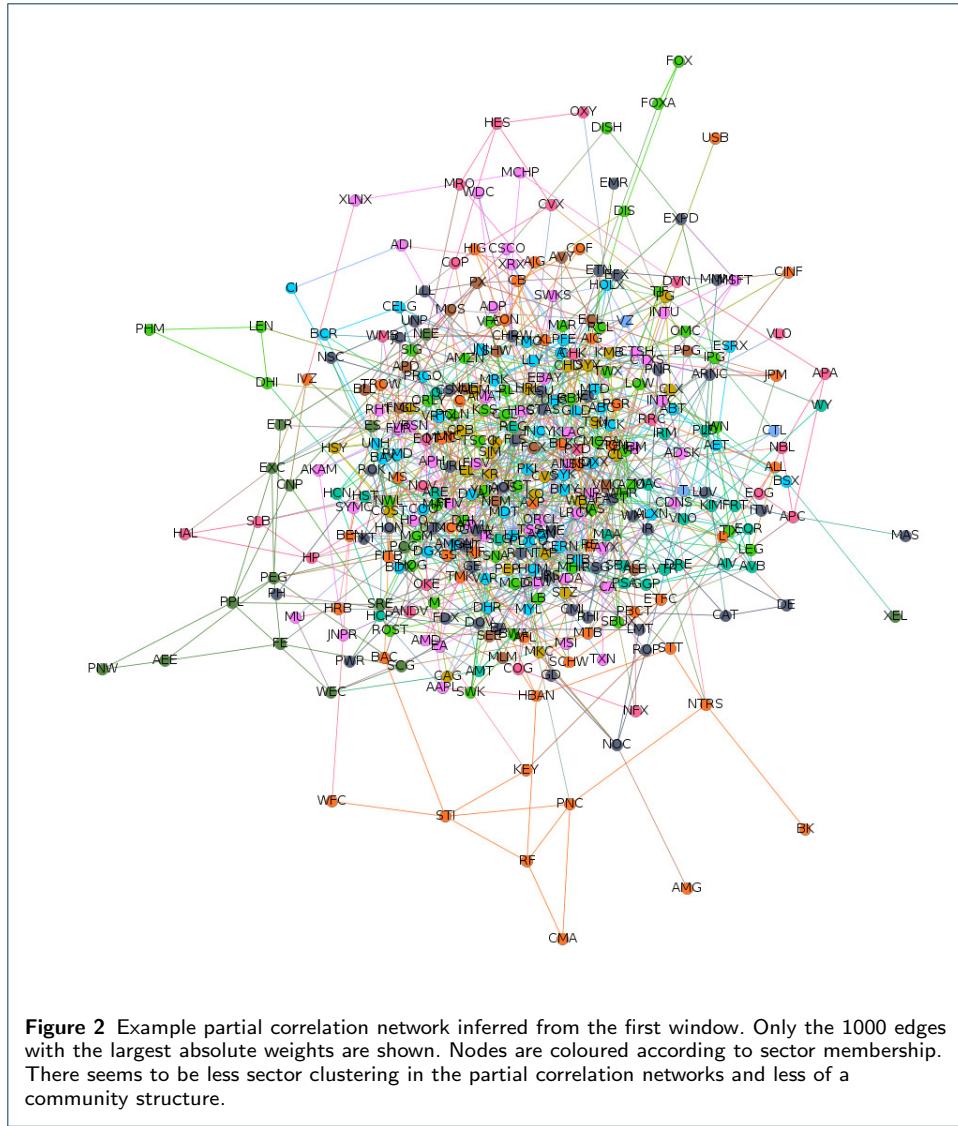


4 Results and Analysis

4.1 Network Analysis and Sector Centrality

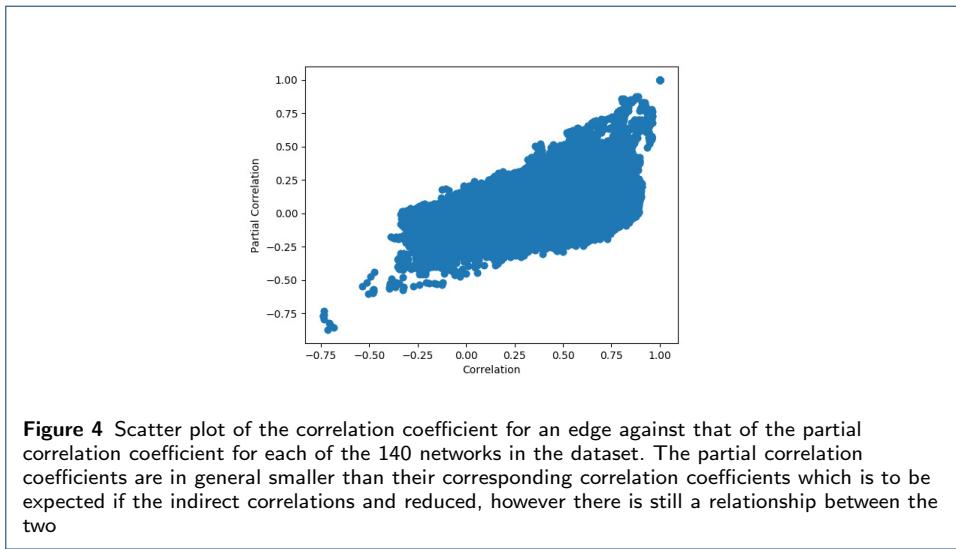
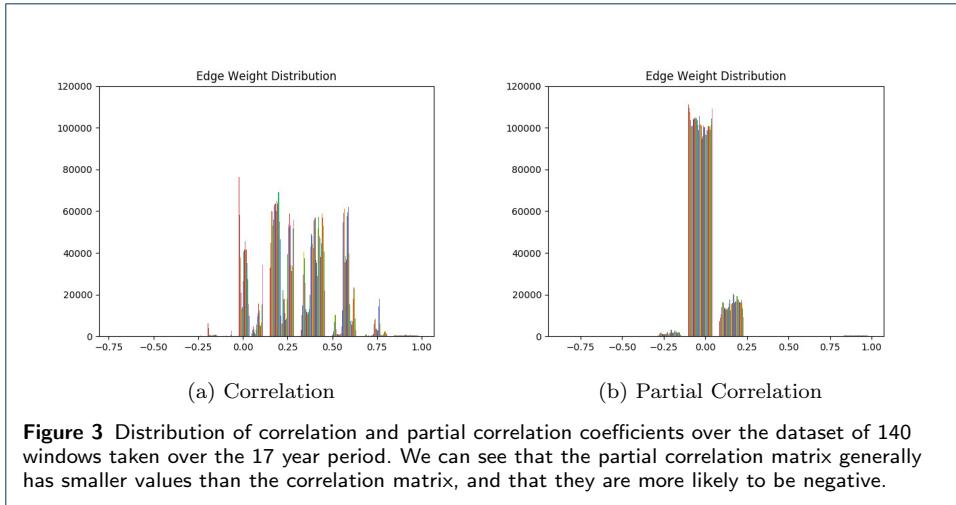
Firstly we display the networks constructed on the first window of this data. Since they are dense, we display the edges that correspond to the 1000 largest absolute values from the off-diagonal of the matrix. The correlation network has isolated nodes in this situation, so we only display the largest connected component, but the partial correlation network remains connected. These networks are displayed in Figures 1 (Correlation) and 2 (Partial Correlation). Both networks show a degree of sector clustering but it is far more prominent in the correlation networks compared to the partial correlation networks. The partial correlation networks also seem to have a more uniform degree distribution than the correlation ones, with less community structure.

To begin our analysis we look at the distribution of correlation coefficients to partial correlation coefficients in the network, and the difference in weight between the same edge in the two networks. A histogram of these is shown in Figure 3, and

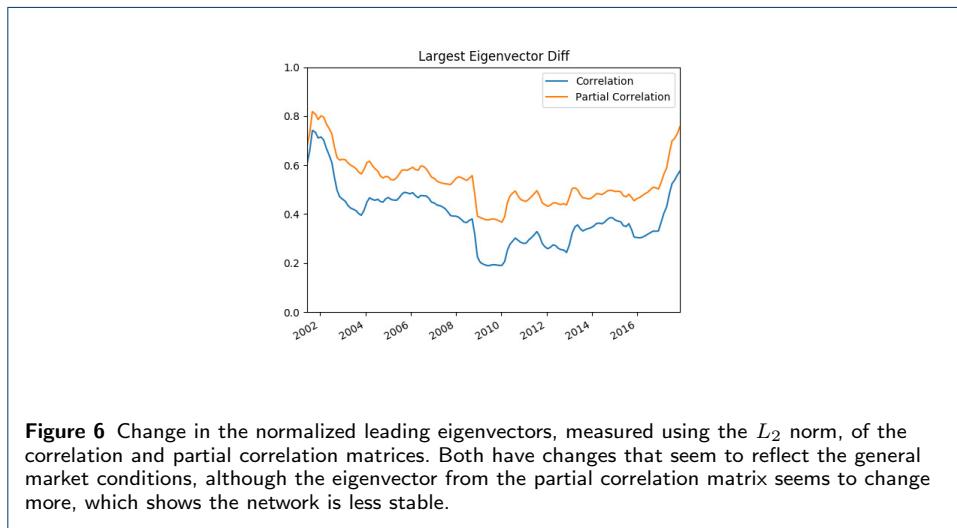
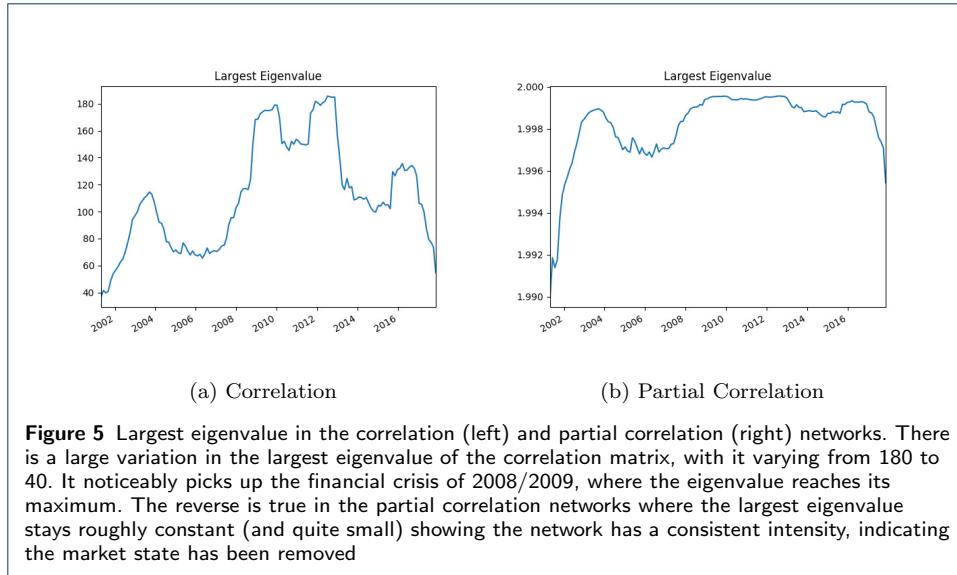


a scatter plot relating the two is shown in Figure 4. In general, partial correlation coefficients tend to be smaller than the corresponding correlation values and are more likely to be negative, but it is also clear that the two are related. This is likely to be due to the definition of partial correlation - if it is reducing the value of indirect correlations then we would expect some companies that are supposedly correlated to have these relationship strengths reduced.

Our next goal is to compare and contrast stability of the networks. In a correlation matrix the largest eigenvalue measures the intensity of the correlation present, and the corresponding eigenvector measures the ‘market mode’ and the effect the general market has on that particular company [14] [33]. Each entry of this eigenvector can also be used as a measure of centrality. Therefore we can study how this eigenvector changes over time to see if the networks regard the same nodes as important, a proxy for how stable the networks are. To measure this we normalize the eigenvectors so the components to add to 1 and then measure the difference between those from adjacent windows using the L_2 norm.



Firstly we look at how the largest eigenvalue varies over time. The results are shown in Figure 5. From this we can see that the largest eigenvalue of the partial correlation matrix is much smaller and varies relatively little, compared to the largest eigenvalue of the correlation matrix. This implies that the intensity of the partial correlation networks does not change much over the dataset, particularly compared to the correlation networks which have large changes. This perhaps indicates that the market mode has been removed. However if we look at the difference in the eigenvectors we get a slightly different story. From Figure 6 we can see there is a larger change in the corresponding leading eigenvector of the partial correlation matrix as opposed to the correlation matrix, signifying the partial correlation networks are less stable than the corresponding correlation networks and could indicate as to why minimum risk portfolios tend to require large changes in asset holdings [34]. Both seem to reflect macroeconomic changes, with the magnitude of the difference varying over time. Interestingly the differences between eigenvectors from adjacent windows drops during periods of disruption.



Next we study the centrality of the sectors in the networks, allowing us to quantify their influence in the economy. We use two measures, degree centrality and eigenvector centrality. Relating these specifically to financial networks, firstly we note that degree centrality is simply the sum of the edges of a node. The weight on the optimal portfolio is also proportional to weight of the edges on a node (see equation 6). Secondly eigenvector centrality is calculated using the eigenvector that corresponds to the largest eigenvalue, with its components normalized to sum to 1 in the same manner as above. This largest eigenvector reflects the market mode and the effect the general market has on a particular company.

The presence of negative edges in the networks makes calculating centrality more challenging. Negative edges can result in a node having negative centrality, which does not have an obvious interpretation. We can use the absolute values of the edges to solve this problem but this involves discarding the negative relationships, which are numerous in the partial correlation networks. In our experiments we found relatively little difference between permitting negative edges or using the absolute

values of edge weights and so permit the existence of negative edges. We normalize at the end so the sum of all node centralities is 1.

To measure the centrality of a sector we take the mean centrality of all the companies in said sector. We then normalize these mean sector centralities to add to 1 to make comparison easier. To start with we look at the mean degree centrality for the sectors for each network. A graph of this over time is shown in Figure 8. In this graph we can see that the partial correlation networks have a much lower difference and variance in the sector centrality than the correlation networks - each sector has roughly the same centrality and there is little variation over time. The telecommunications sector has relatively few companies, hence why its centrality has far more variance. In the correlation networks we firstly see a much larger variance in the mean centrality of a sector, with the financial sector having the highest mean centrality for the majority of the dataset. Interestingly all the centralities 'jump' together during the financial crisis, showing how suddenly the correlations between previously unrelated companies increases due to these macroeconomic effects.



Figure 7 Legend for the sector colours for Figures 8 and 9

Next we look at eigenvector centrality. A graph of this over time is shown in Figure 9. In this figure we see very different results to degree centrality for the partial correlation networks, which show a much larger variation in the centrality of the sectors. They also show a slightly larger variance in the mean centrality when compared with the correlation networks too. In particular the financial sector is far more dominant than we would expect. The macroeconomic conditions are also visible in these graphs, with the financial crisis again forcing the mean centralities towards a mean.

The movement of centrality measures towards a mean during times of disruption is particularly interesting. Preis et al [20] pointed out that the market tends to be more correlated during times of market disruption, which makes selecting truly diversified portfolios very challenging as suddenly supposedly unrelated assets become related during these periods. This may be relevant here too - here we have that during periods of disruption companies start having far more similar behaviour than they did during times of stability.

We have pointed out this connection between the minimum risk portfolios and the degree centrality in the partial correlation networks, but to what degree does it

hold? To further explore this we plot the L_2 difference between the optimal portfolio vector and the sum of the diagonal of the precision matrix (as this is effectively discarded by the partial correlation matrix) in Figure 10. From this we can see the difference can be quite large, although most of the difference seems to be explained by the size of the precision matrix diagonal, which is effectively discarded by the partial correlation networks.

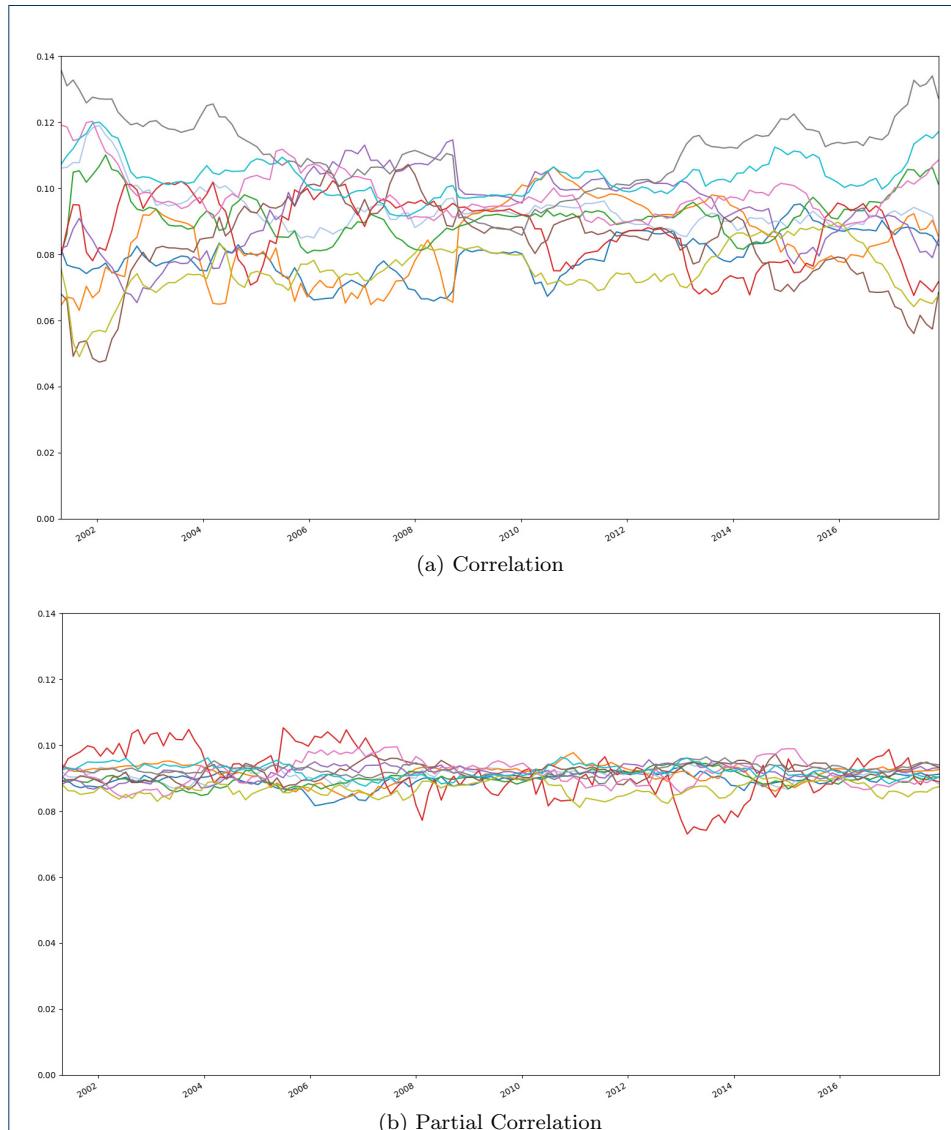


Figure 8 Mean degree centrality for each sector over time. It is noticeable that in the partial correlation networks the difference in centrality is much smaller for each sector than in the correlation networks. We can see the macroeconomic trends in the correlation networks with the centralities jumping together during the crash. The colour legend can be found in Figure 7

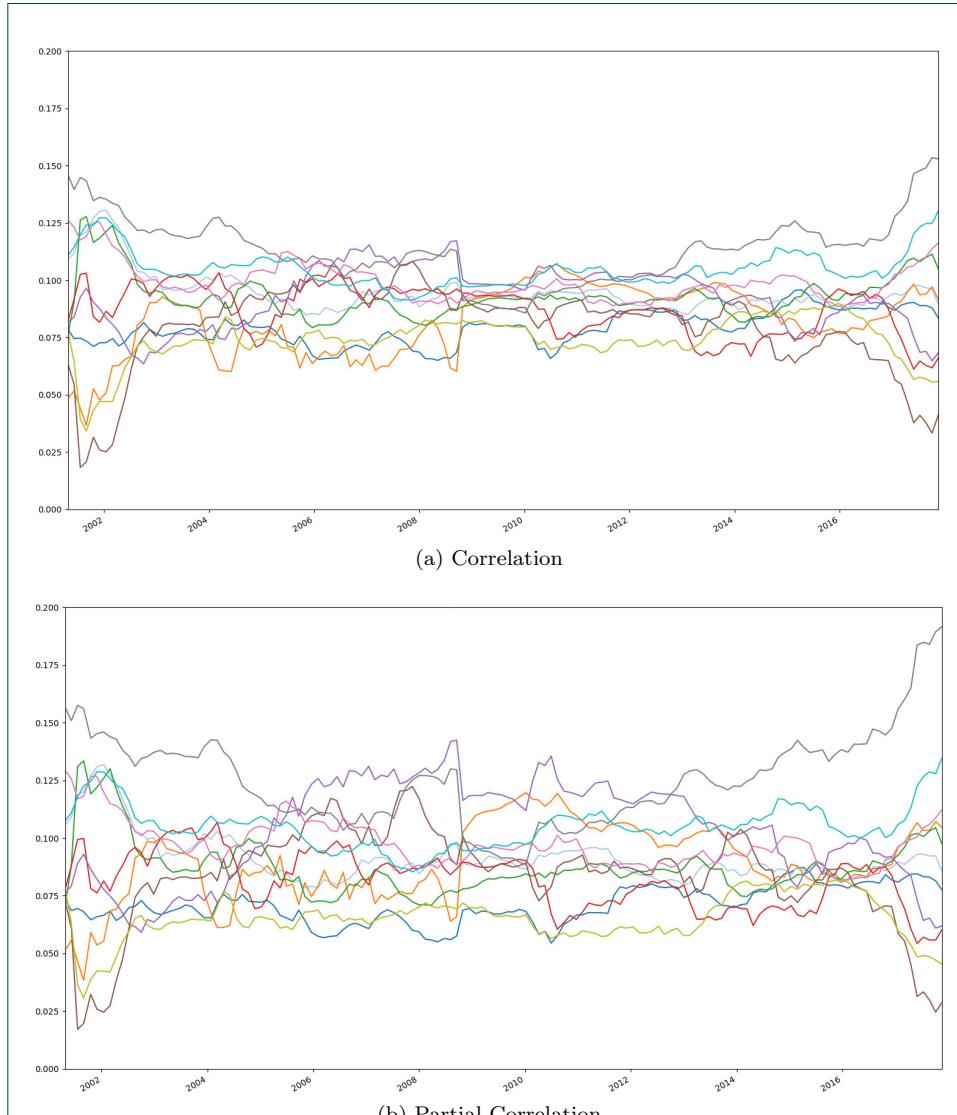


Figure 9 Mean eigenvector centrality for each sector over time. These centralities have a much larger variance than the degree centralities, particularly for the partial correlation networks. Here we can see the financial sector is the most central for the majority of the dataset, although the real estate does also become important in both networks. Macroeconomic effects are also much more visible, with the strong change from 2009 - 2011 as all the sector centralities move together. The colour legend can be found in Figure 7

4.2 Out of sample Portfolio Performance

Previous work with correlation networks has stated that companies who are on the fringes of the network have a better Sharpe ratio than those who are more central [4] [3]. We are curious as to how this applies in these networks. Therefore we study the centrality of a company against its out of sample Sharpe ratio (defined as the mean return over the standard deviation of the returns $\frac{\mu}{\sigma}$) and its risk (defined as the standard deviation of the returns) for the next window.

Using Spearman correlation we find that there is mild positive correlation between the centrality of a company and its out of sample Sharpe ratio in every network, and perhaps unsurprisingly mild negative correlation between out of sample risk and centrality. The exact results are shown in Table 1. All results are statistically significant at $p < 0.05$.

4.3 Community Detection

In Figures 1 and 2 it can be seen that there is some community structure in the networks, so we are interested in further studying this. To do so we use a community detection algorithm to divide each network into communities, and analyze how these change over time. Since we have a ground truth classification of the sector memberships of the various companies, we can also quantify how well these communities reflect the sector structure.

A popular method to detect communities is to attempt to maximize the modularity of the network. This is an NP-hard problem [35] and so various approximate methods have been proposed, including a spectral method [36] or the Louvain algorithm [37]. These methods have been applied previously to detect communities in financial networks constructed from stock data [38] [39].

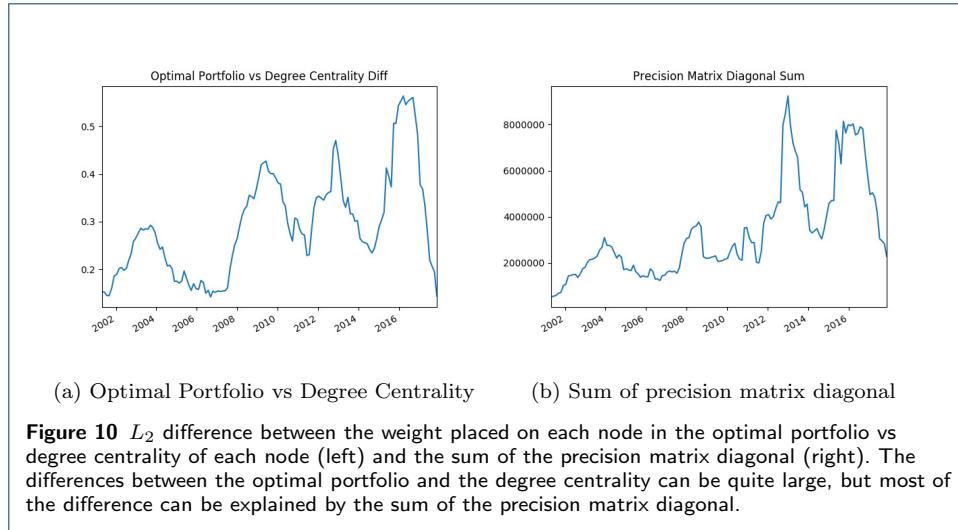
The classic formulation of modularity for a network with adjacency matrix A and a vector of community assignments c is [36]

$$Q = \frac{1}{m} \sum_i \sum_j (A_{ij} - \frac{k_i k_j}{m}) \delta(c_i, c_j) \quad (14)$$

where $m = \sum_i \sum_j A_{ij}$, $\delta(c_i, c_j)$ is the Dirac function, equaling 1 when $c_i = c_j$ (i.e. nodes i and j are in the same community and 0 otherwise) and k_i is the sum of weights of a node. However this is not appropriate when we are looking at graphs with negative edges. Here we use a definition designed for the presence of negative edges. Proposed by Gomez et al [40] we divide the network into positive edges

Sharpe Ratio	Degree	Eigenvector
Correlation	0.156	0.148
Partial Correlation	0.129	0.131
Risk		
Correlation	-0.189	-0.175
Partial Correlation	-0.145	-0.156

Table 1 Spearman correlation between the centrality measures and the out of sample risk and Sharpe ratio. In all situations there is a positive relationship between centrality and Sharpe ratio, and a negative relationship between centrality and out of sample risk. All relationships are statistically significant at $p < 0.05$ but the correlation is relatively mild.



(signified by a +) and negative edges (signified by a -)

$$A_{ij} = A_{ij}^+ - A_{ij}^- \quad (15)$$

where

$$A_{ij}^+ = \max(0, A_{ij}) \quad (16)$$

$$A_{ij}^- = \max(0, -A_{ij}) \quad (17)$$

and so the definitions of modularity are

$$Q^+ = \frac{1}{m^+} \sum_i \sum_j (A_{ij}^+ - \frac{k_i^+ k_j^+}{m^+}) \delta(c_i, c_j) \quad (18)$$

$$Q^- = \frac{1}{m^-} \sum_i \sum_j (A_{ij}^- - \frac{k_i^- k_j^-}{m^-}) \delta(c_i, c_j) \quad (19)$$

where

$$m^+ = \sum_i \sum_j A_{ij}^+ \quad (20)$$

$$m^- = \sum_i \sum_j A_{ij}^- \quad (21)$$

$$k_i^+ = \sum_j A_{ij}^+ \quad (22)$$

$$k_i^- = \sum_j A_{ij}^- \quad (23)$$

Total modularity is then a scaled version of these

$$Q = \frac{m^+}{m^+ + m^-} Q^+ - \frac{m^-}{m^+ + m^-} Q^- \quad (24)$$

We choose the Louvain method [37] to maximize modularity. This method works by maximizing modularity in a greedy bottom up manner. All nodes are initialized into their own random community. We then check the gain in modularity from moving node i from community a to community b . Doing so for all communities, we put node i in the community that maximizes the modularity gain the most, assuming the gain is positive. We then continue to do this over all nodes until we cannot put a node in another community and make a positive gain in modularity. Phase 1 of the algorithm is now complete. In phase 2 each community can then be treated as a node and the edges out to other communities collapsed into one edge per community. The algorithm can then be run again until we do not make a gain in modularity by collapsing the communities.

The gain in modularity from moving isolated node i into a community can be calculated by separately considering the positive and negative edges as follows

$$\delta Q^+ = \frac{\sum_{in}^+ + 2k_i^+}{m^+} - \left(\frac{\sum_{tot}^+ + k_i^+}{m^+} \right)^2 - \left(\frac{\sum_{in}^+}{m^+} - \left(\frac{\sum_{tot}^+}{m^+} \right)^2 - \left(\frac{k_i^+}{m^+} \right)^2 \right) \quad (25)$$

$$\delta Q^- = \frac{\sum_{in}^- + 2k_i^-}{m^-} - \left(\frac{\sum_{tot}^- + k_i^-}{m^-} \right)^2 - \left(\frac{\sum_{in}^-}{m^-} - \left(\frac{\sum_{tot}^-}{m^-} \right)^2 - \left(\frac{k_i^-}{m^-} \right)^2 \right) \quad (26)$$

where \sum_{in} is the sum of weights of all the edges inside the community node i is being moved into, \sum_{tot} is the sum of weights of the edges to the community. The gains are then scaled by the total weight of positive and negative edges in the graph and combined together

$$\delta Q = \frac{m^-}{m^+ + m^-} \delta Q^+ - \frac{m^+}{m^+ + m^-} \delta Q^- \quad (27)$$

A notable advantage of this algorithm is that we do not need to choose the number of communities. Since this algorithm is greedy, we randomise the order of the nodes each run through in phase 1. This of course means that we will achieve different results every time we run the algorithm. Therefore we run the algorithm 10 times on each network to get a mean and standard deviation for any measures taken.

To evaluate our clustering we use the Adjusted Rand Index. Given a set of elements S , and two partitions of this X and Y divided into subsets the Rand Index [41] is defined as

$$R = \frac{a + b}{a + b + c + d} \quad (28)$$

where a is the number of pairs of items in the same subset in X and Y , b is the number of pairs of items that are in different subsets in X and Y , c is the number of pairs of items in the same subset in X and a different subset in Y and d is the number of pairs of items in the same subset in Y and different subsets in X .

Basically it is very similar to a measure of accuracy if we have a ground truth labelling.

In the Adjusted Rand Index (ARI) a correction is made for chance, using the expected similarity for clustering under a random model:

$$ARI = \frac{R - E[R]}{\max(R) - E[R]} \quad (29)$$

Figure 11 shows the ARI over time, exhibiting how the networks reflect the known sector structure. The partial correlation networks have less success overall in discovering the sector structure, although both networks show large variations over the time period. We suggest this is due to the reduction in indirect correlations that the partial correlation coefficient provides - we would expect this to reduce intra-sector correlation strengths which would lead to a reduced success in recovering the sector structure. It is also noticeable that times of disruption seem to lower the ARI for the correlation networks. This could be due to the increased amount of correlation and volatility causing companies to behave more similarly, reducing the ability of the algorithm to separate them [20]. However for the partial correlation networks the ARI actually increases during these times.

We next look at the number of clusters produced. This is shown in Figure 12. The correlation networks have a smaller number of clusters than the partial correlation networks, averaging around 4 while the partial correlation networks have an average of around 20. Both choose roughly the same number of clusters throughout the entire dataset, although we can see there is a small dip for both networks in 2009. Again this could be due to the increasing correlations during market disruption, making the companies seem more similar. There are 11 actual sectors in the dataset, so neither method is particularly close to the true value.

Finally we study how stable the clusters are over time. Since we have 10 partitions per network, we use the adjusted rand index to compare the consistency of the clustering between those from the previous window and those from the next. The results of this are shown in Figure 13. The correlation networks have a much more stable structure than the partial correlation networks (i.e the ARI between each run is larger meaning more companies are in the same cluster), although the variance does seem to be higher. For the correlation networks there is a large 'break' in 2008 with the consistency dropping considerably, but then both networks have an increase in stable clustering over the crisis. This is consistent with our finding about network stability in Section 4.1, with the partial correlation networks being less stable than the correlation networks.

5 Conclusion and Future Work

In this paper we have constructed correlation and partial correlation networks from S&P500 returns data using the Ledoit-Wolf covariance estimator. This is designed to cope having more dimensions than samples and always gives us an invertible covariance matrix, which is required for estimation of the partial correlation matrix. We construct 140 networks using windows of data and contrast the correlation and partial correlation networks produced.

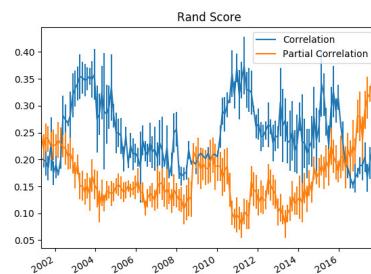


Figure 11 Adjusted Rand Index score over time using Louvain community detection algorithm. In general the partial correlation networks have less success in uncovering the sector structure than the correlation networks although there are time periods where this is not the case. This could be due to the reduction in supposedly indirect correlations by the partial correlation networks

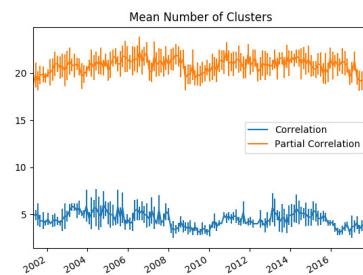


Figure 12 Number of clusters over time using Louvain community detection algorithm. The correlation networks generally seem to pick up a smaller number of clusters than the partial correlation networks, but both keep a similar number over the entire dataset. There is also a drop in the number of clusters detected for both networks during the financial crisis

Firstly we compare the edge weights in these networks. The partial correlation network has more negative edges than the correlation network and generally has smaller weights. There is however a clear relationship between the two - several edges with a high correlation also have a high partial correlation. Since partial correlation is designed to reduce the effect of indirect correlations this is something we would expect.

Secondly we use the largest eigenvalue and corresponding eigenvector to measure the intensity and stability of the networks. The largest eigenvalue of the correlation network varies significantly depending on the state of the market at that time while the largest eigenvalue of the partial correlation network remains roughly constant. This shows that the partial correlation networks do not have much change in intensity over the dataset. However, using the difference in the largest eigenvector to measure the stability of the network we find the partial correlation networks are significantly less stable than the correlation networks, perhaps showing why minimum risk portfolios tend to be less stable.

Exploring the mean centrality of various sectors using both degree and eigenvector centrality, we find that in the partial correlation networks the sectors all have relatively similar mean centrality with degree centrality. Furthermore macroeconomic

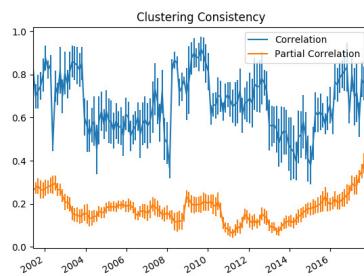


Figure 13 Clustering Consistency for the Louvain algorithm over time. The correlation network produces much more stable clusters than the partial correlation networks, although there is much more variance in this consistency than in the partial correlation networks

factors do not seem to effect the centrality, with all sectors have a fairly consistent mean degree centrality. This is not the case in the correlation networks, where there is clear variation in the centralities, notably during the financial crisis of 2008/2009. The eigenvector centralities show a very different story, with there being more variation in the partial correlation networks rather than less. Macroeconomic effects are also picked up in both networks here, again with the centralities of the sectors moving together during the financial crisis.

Utilizing these networks for portfolio selection, we find there is positive correlation between the centrality of a company and its out of sample Sharpe ratio but there is negative correlation between its centrality and risk. This result is statistically significant and is relevant for both the correlation and partial correlation networks.

Finally we run an altered Louvain community detection algorithm using a version of modularity that is designed for networks with negative edges to attempt to discover whether the sector assignments are replicated in the actual data. We find that the partial correlation networks are less successful than the correlation networks in uncovering these sector clusters. The correlation networks also produce more stable clustering with a lower number of clusters than the partial correlation networks. This indicates that in general the partial correlation networks have less stable structure than a correlation network constructed from the same data.

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TM designed the research and conducted the experiments. MN guided the research with valuable comments. All authors have read and approved the manuscript

Funding

TM Acknowledges PhD studentship funding from the School of Electronics and Computer Science, University of Southampton; MN's contribution to this work was partially funded by EPSRC grant Joining the Dots: From Data to Insight (EP/N014189/1)

Acknowledgements

Not applicable

Availability of data and materials

Data and code used to create this manuscript are available on Github at https://github.com/shazzm/partial_correlation_financial_networks

References

1. Barabási, A.-L.: Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**(1987), 20120375 (2013)
2. Barabási, A.-L.: *Linked: The new science of networks*. AAPT (2003)
3. Pozzi, F., Di Matteo, T., Aste, T.: Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports* **3**, 1665 (2013). doi:10.1038/srep01665
4. Peralta, G.: Network-based measures as leading indicators of market instability: the case of the spanish stock market. *Journal of Network Theory in Finance* (2015). doi:10.2139/ssrn.3400673
5. Huang, W.-Q., Zhuang, X.-T., Yao, S.: A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications* **388**(14), 2956–2964 (2009). doi:10.1016/j.physa.2009.03.028
6. Sun, X.-Q., Cheng, X.-Q., Shen, H.-W., Wang, Z.-Y.: Distinguishing manipulated stocks via trading network analysis. *Physica A: Statistical Mechanics and its Applications* **390**(20), 3427–3434 (2011). doi:10.1016/j.physa.2011.04.006
7. Shi, F.-B., Sun, X.-Q., Shen, H.-W., Cheng, X.-Q.: Detect colluded stock manipulation via clique in trading network. *Physica A: Statistical Mechanics and its Applications* **513**, 565–571 (2019). doi:10.1016/j.physa.2018.09.011
8. Mantegna, R.N.: Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems* **11**(1), 193–197 (1999)
9. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J.: Dynamic asset trees and black monday. *Physica A: Statistical Mechanics and its Applications* **324**(1), 247–252 (2003). doi:10.1016/S0378-4371(02)01882-4. Proceedings of the International Econophysics Conference
10. Boginski, V., Butenko, S., Pardalos, P.M.: Statistical analysis of financial networks. *Computational Statistics & Data Analysis* **48**(2), 431–443 (2005). doi:10.1016/j.csda.2004.02.004
11. Nagurney, A.: *Innovations in Financial and Economic Networks*. Edward Elgar Publishing, ??? (2003)
12. Chu, J., Nadarajah, S.: A statistical analysis of uk financial networks. *Physica A: Statistical Mechanics and its Applications* **471**(Supplement C), 445–459 (2017). doi:10.1016/j.physa.2016.12.073
13. Song, J.Y., Chang, W., Song, J.W.: Cluster analysis on the structure of the cryptocurrency market via bitcoin–ethereum filtering. *Physica A: Statistical Mechanics and its Applications* **527**, 121339 (2019). doi:10.1016/j.physa.2019.121339
14. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T., Stanley, H.E.: Random matrix approach to cross correlations in financial data. *Physical Review E* **65**(6), 066126 (2002)
15. Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershoren, G., Mantegna, R.N., Ben-Jacob, E.: Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLOS ONE* **5**(12), 1–14 (2010). doi:10.1371/journal.pone.0015032
16. Wang, G.-J., Xie, C., Stanley, H.E.: Correlation structure and evolution of world stock markets: Evidence from pearson and partial correlation-based networks. *Computational Economics* **51**(3), 607–635 (2018). doi:10.1007/s10614-016-9627-7
17. Millington, T., Niranjan, M.: *Robust Portfolio Risk Minimization Using the Graphical Lasso*, pp. 863–872. Springer, Cham (2017)
18. Millington, T., Niranjan, M.: Quantifying influence in financial markets via partial correlation network inference. In: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 306–311 (2019). doi:10.1109/ISPA.2019.8868437
19. Wainwright, M.J.: Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory* **55**(5), 2183–2202 (2009). doi:10.1109/TIT.2009.2016018
20. Preis, T., Kenett, D.Y., Stanley, H.E., Helbing, D., Ben-Jacob, E.: Quantifying the behavior of stock correlations under market stress. *Scientific Reports* **2**, 752 (2012). doi:10.1038/srep00752. Article
21. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**(2), 365–411 (2004)
22. Namaki, A., Shirazi, A.H., Raei, R., Jafari, G.R.: Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications* **390**(21), 3835–3841 (2011). doi:10.1016/j.physa.2011.06.033
23. Markowitz, H.M.: Portfolio selection. *The Journal of Finance* **7**(60), 77–91 (1952)
24. Schäfer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* **4**(1) (2005)
25. Brier, M.R., Mitra, A., McCarthy, J.E., Ances, B.M., Snyder, A.Z.: Partial covariance based functional connectivity computation using ledoit–wolf covariance regularization. *NeuroImage* **121**, 29–38 (2015). doi:10.1016/j.neuroimage.2015.07.039
26. Oliphant, T.E.: *A Guide to NumPy* vol. 1, (2006)
27. McKinney, W.: Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*, pp. 51–56 (2010)
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
29. Seabold, S., Perktold, J.: Statsmodels: Econometric and statistical modeling with python. In: *9th Python in Science Conference* (2010)
30. Hunter, J.D.: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**(3), 90–95 (2007). doi:10.1109/MCSE.2007.55
31. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, pp. 11–15 (2008). doi:10.25080/issn.2575-9752

32. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: *Third International AAAI Conference on Weblogs and Social Media* (2009)
33. Namaki, A., Shirazi, A., Raei, R., Jafari, G.: Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications* **390**(21-22), 3835–3841 (2011)
34. DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies* **22**(5), 1915–1953 (2007). doi:10.1093/rfs/hhm075
35. Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: Maximizing modularity is hard. *arXiv preprint physics/0608255* (2006)
36. Newman, M.E.J.: Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103**(23), 8577–8582 (2006). doi:10.1073/pnas.0601602103. <https://www.pnas.org/content/103/23/8577.full.pdf>
37. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**(10), 10008 (2008). doi:10.1088/1742-5468/2008/10/p10008
38. PICCARDI, C., CALATRONI, L., BERTONI, F.: Clustering financial time series by network community analysis. *International Journal of Modern Physics C* **22**(01), 35–50 (2011). doi:10.1142/S012918311101604X. <https://doi.org/10.1142/S012918311101604X>
39. Isogai, T.: Clustering of Japanese stock returns by recursive modularity optimization for efficient portfolio diversification*. *Journal of Complex Networks* **2**(4), 557–584 (2014). doi:10.1093/comnet/cnu023. <http://oup.prod.sis.lan/comnet/article-pdf/2/4/557/9130958/cnu023.pdf>
40. Gómez, S., Jensen, P., Arenas, A.: Analysis of community structure in networks of correlated data. *Physical Review E* **80**(1), 016114 (2009)
41. Rand, W.M.: Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**(336), 846–850 (1971). doi:10.1080/01621459.1971.10482356. <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1971.10482356>