Rubin *et al. Applied Network Science* (2019) 4:31
https://doi.org/10.1007/s41109-019-0132-5

Applied Network Science

RESEARCH Open Access

Uncovering dynamic stock return correlations with multilayer network analysis

Danielle N. Rubin[1,2], Danielle S. Bassett[1,3*] and Robert Ready[4]

*Correspondence:
dsb@seas.upenn.edu
[1]Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia PA, USA
[3]Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia PA, USA
Full list of author information is available at the end of the article

Abstract
We apply recent innovations in network science to analyze how correlations of stock returns evolve over time. To illustrate these techniques we study the returns of 30 industry stock portfolios from 1927 to 2014. We calculate Pearson correlation matrices for each year, and apply multilayer network tools to these correlation matrices to uncover mesoscale architecture in the form of communities. These communities are easily interpretable as groups of industries with highly correlated stock returns. We observe that the flexibility, or the likelihood of industries to switch communities, exhibits a statistically significant increase after 1970, and that the communities evolve in ways consistent with changes in the structure of the U.S. economy. We find that these patterns are not explained by changes in average pairwise correlations or industry market betas. These results therefore underscore the potential for using multilayer network tools to study time-varying correlations of financial assets.

Keywords: Stocks, Industries, Networks, Clustering, Modularity, Community structure, Flexibility

Introduction

Correlations of asset returns are central to our understanding of financial markets. However, the high dimensionality and time-varying nature of correlation matrices poses challenges for both estimation and interpretation. A large set of tools exists to address these issues (Kenett et al. 2015; Uechi et al. 2014; Musmeci et al. 2015; Song et al. 2012), but it remains difficult to answer simple intuitive questions about which groups of assets are highly correlated with each other, and how these groups change through time. For instance, Dynamic Conditional Correlation (DCC) models (Engle et al. 1992) allow for estimation of changing correlation matrices, and principal components analysis reduces the dimensionality of return correlation matrices (Connor and Korajczyk 1993). However, neither of these technique yields results that allow for an intuitive understanding of which assets share a high degree of correlation at a given point in time.

Network science provides methods for reducing the high-dimensionality of a correlation matrix to yield easily interpretable summaries. The approach can be used to assess structure within the correlation matrix that exists over a range of topological scales, from the local pattern of correlations relevant to a single asset to the global pattern of correlations between all assets. At the intermediate scale, tools from network science can be used to assess so-called mesoscale architecture in the correlation matrix (Tilak et al. 2011; Pollet and Wilson 2010; Yang et al. 2014; Coats and Fant 1993). In general, these mesoscale techniques allow for the representation of a high-dimensional structure which falls between the microscale, which in our context would be the full correlation matrix, and the macroscale, such as a single number representing the average correlation across the entire matrix[1]. The mesoscale network structure of a correlation matrix can be characterized by a variety of features, with the most common and well-studied being the presence of communities or modules, which can be thought of as clusters of highly correlated assets.

Network models in finance have a long history, going back several decades (Odom and Sharda 1990). Despite their consistent utility, network-based tools are most often applied to studying the business relations of firms (e.g., (Boss et al. 2004) or (Diebold and Yilmaz 2014)), or the impact of shocks on the financial system (JBookstaber and Kenett 2016; Korniyenko et al. 2018). More recently, similar tools have been applied to analyze stock comovements (Raddant and Kenett 2016; Tilak et al. 2011; Achard et al. 2008) and how their dynamics may relate to volatility in financial returns (Aste et al. 2010; Isogai 2016), or to cluster dynamics in the foreign exchange market (Fenn, et al. 2009). Here we expand upon the contributions of these prior studies by utilizing network tools to characterize the evolution of mesoscale architecture (Jalili and Perc 2017) in correlation matrices, which we represent as an ordered multilayer network.

Specifically, in this paper we use recent developments in network science that have taken tools to identify, extract, and characterize the mesoscale architecture of networks and extended those tools into the time domain (Mucha et al. 2010; Holme and Saramaki 2015; Zalesky et al. 2014). This approach is particularly useful for assessing dependencies in complex financial data sets (Musmeci et al. 2017; Fenn et al. 2010). To demonstrate how the statistical tools to uncover mesoscale network reconfigurations can be applied to financial markets, we use them to study the relationships between 30 different industries' stock returns over the last 88 years. The estimation yields a time series of highly-correlated industry groups, or *communities*. The identified communities change over time, demonstrating nontrivial dynamics in the mesoscale structure in the industry stock return correlation matrix. The communities we identify are also intuitive and adjust in response to changes in the underlying structure of the U.S. economy. We find that the community structure is constant for the first 38 years of our sample, with a large community of manufacturing industries and intermediate goods industries forming the dominant block. This dominant community is accompanied by many smaller ones, including several composed of just a single industry.

Around 1970, these communities begin to change. In the context of our method, the associated networks become more *flexible* as the communities shift and reorganize in ways that reflect changes in the structure of the U.S. Economy. As an example, many industries that produce intermediate goods leave the manufacturing community, and join first an information technology community in 2000, and then an energy producing community by 2010. Also of note, prior to 1970, Finance and Banking is assigned to a community with both Utility and Telecommunications firms, suggesting that financial firms behaved more as a consumer utility. However, for all years after 1990, Finance and Banking is assigned its own community reflecting the finance sector's increased independence and importance in the U.S. economy.

An important contribution relative to the existing applications of mesoscale techniques to financial data is that our method also allows for statistical inference. By generating a bootstrapped null model we are able to show that the flexibility of the network structure has increased over time by a statistically significant amount, suggesting a more dynamic structure of industry returns in the latter half of the sample. Critically, these intuitions about the dynamic nature of financial assets are not possible using more standard measures of changes in stock market correlations. Specifically, we do not see a trend in the average correlation across industries, nor do we find that our communities are well described by their market beta, suggesting that simply relying on changing exposures to a single principal component would not uncover these results. Following a description of the approach used to analyze the mesoscale network architecture present in our data source's time-dependent correlation matrices, we present the results of the analyses performed, followed by a discussion of the findings and the implications of these techniques for the field.

Mesocale network architecture in time-dependent correlation matrices

Initially, the study of network models of real-world systems predominantly focused on models that considered each relationship between two system parts and summarized that relationship in a single value. However, many systems are characterized by different sorts of relationships between system parts, or relationships that vary over time. The multiplicity of potential edges between two nodes called for new tools. To address this rather pervasive complexity, recent work in the field of applied mathematics has begun focusing on *multilayer networks, which provide* a framework that can be used to understand collections of networks in which the structure of individual networks are interdependent (Kivela et al. 2014). We use this approach to represent the time-dependent correlation matrices as an ordered multilayer network. Note that we use the word *ordered* to reflect the fact that each correlation matrix has a location in time, and those locations are ordered from earlier time to most recent time.

One useful way in which to study ordered multilayer networks reflecting the temporal progression of internode interaction patterns is to identify and characterize the network's mesoscale architecture. One common mesoscale architecture studied in the literature is that of modular architecture (Fortunato 2011): the presence of clusters of nodes that have denser connectivity to other nodes in the same cluster than to nodes in other clusters (Newman 2006). Such modular architecture – and its reconfiguration over time – can be identified using data-driven clustering techniques developed for network representations of relational data (Mucha et al. 2010). A particularly common approach that has proven useful in the context of diverse complex systems is the clustering technique known as multilayer modularity maximization. This approach extends the modularity maximization technique for static networks into the temporal domain, allowing the investigator to detect dense clusters of nodes whose constituencies can change over time.

Here we utilized this multilayer modularity maximization approach, and we operationalize the technique by using a Louvain-like locally greedy algorithm to maximize a multilayer modularity quality function for ordered multislice networks. This procedure assigns the 30 industries in the network to time-evolving communities by maximizing the modularity quality index:

$$Q = \frac{1}{2\mu} \sum_{ijlr} \left[\left(A_{ij_l} - \gamma_l P_{ij_l} \right) \delta_{lr} + \delta_{ij} \omega_{j_{lr}} \right] \delta \left(g_{i_l}, g_{j_r} \right), \tag{1}$$

where A_{ij_l} is the weighted adjacency matrix of the network, γ_l is the structural resolution parameter that tunes the number of modules to which the algorithm is sensitive, P_{ij_l} is the adjacency matrix of an appropriate null model, $w_{j_{lr}}$ is the temporal resolution parameter which tunes the time scale of the dynamics to which the algorithm is sensitive, g_{i_l} is the community i in time layer l, g_{j_r} is the community j in time layer r, and δ is the Kronecker delta (Blondel et al. 2008). Due to the known near-degeneracies of the modularity landscape (Good et al. 2010), we maximized this quality function 100 times (Mucha et al. 2010). Following prior work (Bassett and Porter, et al. 2013), we chose γ and ω to be constants, and set their values to unity. To confirm robustness of our results to reasonable variations in this choice, we confirmed our main findings at both $\gamma = \omega = 0.95$ and $\gamma = \omega = 1.05$.

After performing multilayer modularity maximization, it is often of interest to extract summary statistics from the resulting partitions of network nodes into time-evolving communities. Such statistics can address diverse features of the network reconfiguration, including the stationarity of the communities, the variability of the communities, and time-dependent functions of community size, among others. Based on prior studies of multilayer network models of time-varying correlation matrices, we chose to summarize the nature of module reconfiguration over time by calculating the network flexibility (Bassett et al. 2011). Network flexibility is equal to the number of times that a node switches module allegiance over the set of time windows, divided by the number of possible times that a node could change module allegiance. Note that the number of possible times is equal to the number of time windows minus 1. Flexibility is then normalized by dividing by the number of modules observed.

Null model construction and statistical testing

When implementing community detection techniques in the context of real-world data, it is important to keep in mind that these techniques always produce a solution, and it is the investigator's responsibility to evaluate the validity of that solution. To assess the statistical significance of the patterns we identify in the data, we compare the empirically derived statistics to those expected in an appropriate dynamic network null model. The null model we elect to consider is constructed by permuting the order of layers in the multilayer network uniformly at random (Bassett and Porter, et al. 2013). We maximize the multilayer modularity quality function of this null model network, and obtain associated values of the quality index Q.

As mentioned in the previous section, to confirm robustness of our results to reasonable variations in the choice of γ and ω, we compared the community assignments obtained at these values to those obtained at $\gamma = \omega = 0.95$ and $\gamma = \omega = 1.05$ (Bassett and Porter, et al. 2013). Specifically, to compare community assignments, we calculated the z-score of the Rand coefficient (Traud et al. 2011). To compare two partitions α and β, we calculate the z-score of the Rand index

$$z_{\alpha\beta} = \frac{1}{\sigma_{w_{\alpha\beta}}} \left(w_{\alpha\beta} - \frac{M_\alpha M_\beta}{M} \right), \tag{2}$$

where M is the network's total number of node pairs, M_α is the number of pairs that are in the same module in partition α, M_β is the number of pairs that are in the same module in partition β, $w_{\alpha\beta}$ is the number of pairs that are assigned to the same module both in partition α and in partition β, and $\sigma_{w_{\alpha\beta}}$ is the standard deviation of $w_{\alpha\beta}$ (as in (Traud et al. 2011)).

Using the formulation above, we can assess the similarity between any two partitions. Next, we define the *mean partition similarity* z to be the mean value of $z_{\alpha\beta}$ over all possible partition pairs for $\alpha \neq \beta$. We observed that the community assignments obtained at $\gamma = \omega = 1$ were statistically similar to those obtained at $\gamma = \omega = 0.95$ and $\gamma = \omega = 1.05$. Specifically, we observed that the z-score of the Rand coefficient was consistently greater than 1.96 ($p < 0.05$): to be exact, $z = 22.46$ between $\gamma = \omega = 0.95$, and $z = 235.75$ between $\gamma = \omega = 1.05$. These findings indicate that the results that we report are robust to reasonable variation in parameter choices.

Applications for industry portfolios

Our primary focus for this analysis is industry portfolios of stock returns. While our methods are very general, and could in theory be used for any set of financial assets, including individual firms, we view our focus on industry portfolios as a natural application of our methodology for several reasons. The first reason is that using diversified portfolios reduces the noise coming from firm-specific shocks, allowing us to uncover information about the structure of the macroeconomy. The second reason is that using portfolios of stocks in a given industry allows for straightforward economic interpretation of our communities. For instance, if one were to generate communities of individual stocks, these communities would require a second step of analysis to understand what common characteristics link stocks in each community.

However, there are some limitations inherent to this approach. The most notable is that the results will be sensitive to the initial classification of stocks to various industries. We therefore proceed using a standard set of industry portfolios (French 2016). Nevertheless, it is important to note that our findings may represent changes in the relation between industries, or changes in the composition of firms classified to a given industry.

Constructing correlations

Using Ken French's Data Library, we obtain daily stock return time series over 88 years (from July 1, 1927 and ending on December 31, 2014) for 30 different industry portfolios (French 2016). These industry portfolios are constructed as the value weighted average returns for a set of firms in a given industry. The industry classifications are based on firms' 4-digit SIC codes as assigned by COMPUSTAT. For instance, the returns to the industry "beer" in 1995 are calculated as the value-weighted portfolio return of all firms classified with 4-digit SIC codes 2080-2085 in 1995. These industry portfolios are standard in the finance literature, and so we view them as a natural testing ground for our methodology.

Using these returns, we then construct correlation matrices – or weighted undirected networks or adjacency matrices – in which network nodes represent industries, and network edges represent the Pearson correlation coefficient between financial returns in pairs of industries. We construct these matrices for each calendar year in our sample, giving us 88 individual correlation matrices[2].

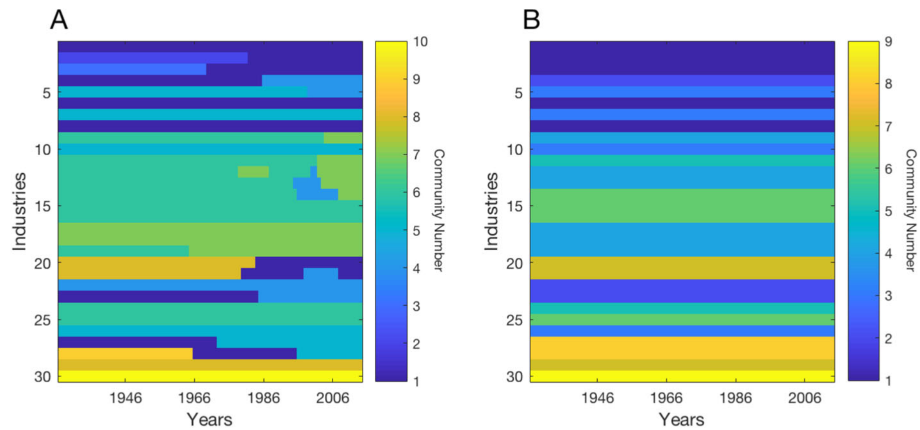

Fig. 1 *Community Assignments: Data and Null Model.* The left hand plot shows community assignments for the 30 industry portfolios over the 88 years from 1927 to 2014. This structure is constant prior to 1960, but shows substantial changes in the latter half of the sample. The right hand plot shows a representative run of the "null model". The null model is constructed by randomizing the order of the 88 sample years, and is constant over the entire sample

Mesoscale network architecture in time-dependent correlation matrices

To determine whether the time-dependent correlation matrices of industry stock returns displayed mesoscale architecture, we performed a network-based clustering technique known as community detection (described earlier) using our 88 yearly correlation matrices. The quality with which communities (or modules) could be observed was quantified by the modularity quality index Q, and the statistical significance of these modules was determined by comparing the empirically derived Q to that expected in an appropriate dynamic network null model. We observed that the Q values of the real data were significantly greater than those of the null model (permutation testing: $p = 4.2 \times 10^{-118}$). This result indicates that time-dependent correlation matrices of industry stock returns display strong modular architecture, with industries in one cluster having a high correlation with industries in that same cluster and a low correlation to industries in other clusters.

Figure 1 shows the identified communities of industries both for the data and for a representative null model network. The left hand panel shows the community structure in the data. This structure is constant prior to 1960, but shows substantial changes in the latter half of the sample. The right panel shows the community structure for a representative null model network. This structure is constant over the entire sample.

The evolution of industry correlations

After revealing significant modular structure in these correlation matrices, we next turned to asking how that modular structure changed across our sample. Figure 2 shows the network partitions at 6 different points in time. One of the most striking patterns in the data is that the composition of communities is constant over the first 38 years of the sample. From 1927 to 1964, the detection algorithm allocates the 30 industries to the same 10 modules in each year. Several industry portfolios are allocated to their own unique modules, including Personal and Business Services (Servs), Tobacco Products (Smoke), Beer and Liquor (Beer), and Restaurants and Hotels (Meals), as well as the Other Industries (Other) portfolio, which takes all industries not readily classified into the other 30

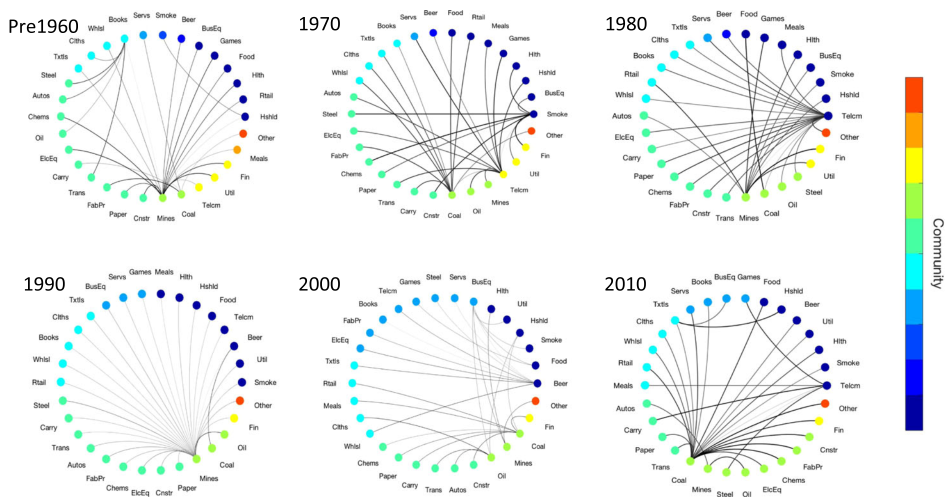

Fig. 2 *Snapshots of Community Assignments from the Data.* This figure plots snapshots of industry communities every 10 years for the 1960 to 2014 sample. Note that the community structure was constant before 1960. Lines indicate a pairwise correlation of industry returns greater than 0.7. Darker lines indicate a stronger connection. By 1970 the communities begin to change modestly, with increasing movement through the 80s, 90s, and 00s

portfolios. This industry is unique in that it stays in its own module for the entirety of the sample.

The largest module over this period contains 10 industries including Automobiles and Trucks (Autos), Aircraft, Ships, and Railroad Equipment (Carry), Construction and Construction Materials (Cnstr), Transportation (Trans), Fabricated Products and Machinery (FabPr), Electrical Equipment (ElcEq), Oil and Natural Gas (Oil), Business Supplies and Shipping Containers (Paper), Steel Works (Steel), and Chemicals (Chems). These mostly manufacturing industries form a dominant block in the correlation structure prior to 1965. The next two largest communities include consumer goods industries, with one containing Clothing (Clths), Textiles (Txtls), Wholesale (Whlsl) and Printing and Publishing (Books), and the other containing Household Goods (Hshld), Food Products (Food), Recreation (Games), Retail (Rtail), and Business Equipment (BusEq). The last two communities include one with Coal Production (Coal), and Mining (Mines), as well as one with Telecommunications (Tlcm), Utilities (Util), and the Financial and Banking Industry (Fin).

By 1970 these communities begin to change, with only modest changes at first. Oil and Natural gas join the community with Coal and Mining to create a clear energy sector. This community expands to include Steel Works in 1980. By 1990, following the rising share of the financial industry in the U.S. economy throughout the 1980s, the Financial and Banking Industry has been left in its own community, as Telecommunications and Utilities have migrated to the Household consumer module. A similar observation was reported by Di-Matteo et al. (2009) publication describing the central role that the Financial sector plays in the hierarchical organization of financial market sectors, and a year later by Kenett et al. (2010) in their publication revealing that Financial sector stocks were found to be the most influential in the correlation profile of the New York Stock Exchange system. It is interesting to note that it was later found that this centrality decreased in the years leading up to the financial crisis (Aste et al. 2010).

More changes can be seen by 2000. With the rise of information technology, the Services Industry is joined by Telecommunications and the intermediate goods producing industries of Electrical Equipment, Fabricated Products, Steel Works, and Business Equipment to form a dominant community. This marks the end of the large manufacturing block as the largest community, a role it held for the prior 70 years. The information technology community however proves short-lived, as by 2010 many of the intermediate goods industries have joined the energy producing sector corresponding with a period of volatile global oil prices and the rise of onshore U.S. production of shale oil and natural gas. While not shown, this structure has remained constant through the end of our data (2014).

Community flexibility

To statistically quantify the degree of network module changes in this system, and to prove that those changes are not expected in an appropriate dynamic network null model, we calculated a network statistic known as *flexibility*. Flexibility describes the frequency with which industries switch modules over time (as described earlier). Figure 3 shows our results. The top panel in the figure plots our estimate for the overall flexibility, which is the total number of community changes through time. The estimated value for the real data is plotted as well as the distribution of this value for the null model networks. We observe that network flexibility was significantly greater in the real data than in the null model networks (t-test with $\alpha < 0.05$: $p = 2.5 \times 10^{-4}$). These statistical tests indicate that the temporal variation in mesoscale network structure that we observe in industry stock returns is non-trivial.

We can also calculate flexibility for each time period (averaged over industries), and examine the presence of flexible *versus* rigid time periods (middle panel of Fig. 3). We observed that flexibility has substantially increased over time, and that this increase is statistically significant relative to a null model which randomizes industries within each time period. Finally, we can calculate flexibility for each industry to determine which industries are flexible, and which industries are rigid. We find that relative to the Null Model, the most flexible industries are Pharmaceuticals, Oil and Gas, Wholesale, and Business Equipment.

It is important to note, however, that changes within industry sectors may also impact the perceived flexibility of the industry over time. For example, information services was not a significant portion of the Personal and Business Services industry sector for the first several decades of the data set, but in recent years has become a major constituent of the economy. The same can be said for computers, which today constitute a significant portion of the Business Equipment industry sector, though that was not the case in 1927. Therefore, our estimates of flexibility may capture either changes in the relations between industries, or changes in the internal compositions of firms within an industry.

Discussion of other statistical methods

Our method of community detection is able to reveal time variation in the correlation of industry stock returns. Estimating correlations between financial assets is a central problem in financial markets, but the high dimensionality of correlation matrices makes the problem quite difficult. The most common method for estimating changes in correlations is the DCC method of (Engle et al. 1992). However, DCC quickly becomes

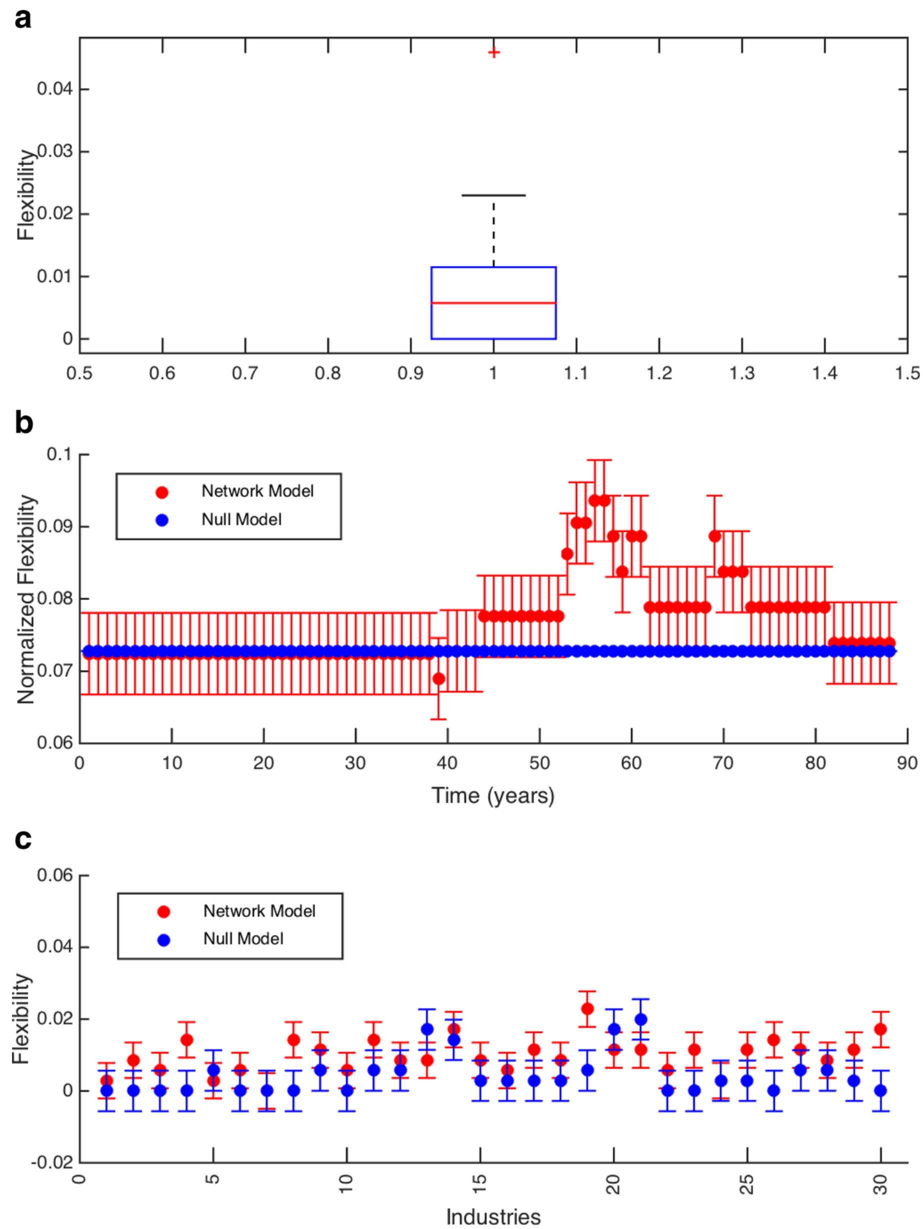

Fig. 3 *Flexibility of Communities in Multilayer Networks.* **a** The box plot shows our estimate of the distribution of flexibilities for the entire data set, which is the total number of community changes through time. **b** The flexibility of the annual null matrix is 0, so all of the industries in the data set are more flexible than their counterparts in the null model, though one standard deviation below the mean for some of the industries fell below 0. The differences observed between the network and null matrix flexibilities were highly significant ($p = 2.5 \times 10^{-4}$). These statistical tests indicate that the temporal variation in mesoscale network structure that we observe in industry stock returns is non-trivial. **c** The quarterly flexibilities are much more variable in both the network matrix and the null matrix. Some industries in the network matrix are more flexible than their counterparts in the null matrix e.g., Pharmaceuticals, Oil and Gas, Wholesale, and Business Equipment, and others are less flexible. The differences observed between the network and null matrix flexibilities were also highly statistically significant ($p = 8.4 \times 10^{-5}$)

computationally infeasible as the number of assets increases, and is more typically used in contexts where there are only two assets. For instance, for our data we found that the standard DCC estimation technique failed to converge when using either daily or monthly return data. Moreover, while this technique can yield time-varying correlation matrices among a few assets, the economic meaning of these matrices is not easily interpretable.

Index models (Sharpe 1963), often implemented via principal components analysis (Connor and Korajczyk 1993), allow for a reduction of the dimensionality of stock return correlations. Unfortunately this technique, as applied to stock returns, yields estimated principal components are not easily interpreted in terms of their economic content. The exception to this is the first principal component, which is typically very close to the return to the aggregate market portfolio. Market betas are then roughly equivalent to loadings on this first principal component, and are a key statistic for capturing risk in financial markets. One potential concern is that our method is simply sorting industries based on their market beta, and that calculating rolling market betas would be enough to uncover the patterns we find. Another concern is that the increased flexibility over the latter half of the sample is simply driven by an increase in the overall correlation of industry returns, which is a proxy for the variation explained by the first principal component.

To address these concerns, we construct two intuitive measures for each year. The first is simply the average pairwise correlation across all industries constructed using daily returns. The second is the average industry market beta of each community in each year. Figure 4 shows the results. The first panel shows that there is no discernible trend in overall industry correlation over our period. While industries do become less correlated on average in the 2000s, this correlation rises in the period since the global financial crisis. This finding suggests that our changes in community flexibility are not being driven simply by changes in overall industry correlation. The second panel of the figure shows the average market beta of each community within each year. While there are differences in market beta across communities, these differences are not persistent. For instance, considering the first half of the sample in which our community structure is completely stable, we see that the betas of individual communities is highly volatile. These findings suggest that time-varying loadings on the first principal component of returns will not be enough to explain our findings.

Conclusion

In this study, we exercise recently-developed tools from the field of network science to characterize financial returns in the stock market. Over the course of 88 years from July 1, 1926 to December 31, 2014, we demonstrate that 30 common industries change in their patterns of return correlation with one another. We study these changes using an ordered multilayer network representation of the data, and apply a dynamic community detection technique to reveal clusters of industries (or modules) that change in their composition over time. We demonstrate that these tools can be used to quantify a novel statistic to describe the dynamics of industry return correlations – flexibility – which has steadily increased from 1926 until 2014. Moreover, we show that the communities have evolved in ways that reflect the changing structure of the U.S. Economy. Finally, we show that these insights are not revealed using aggregate correlations or changing market betas. Taken together, these results underscore the potential for using network science tools to better understand mesoscale collective dynamics in financial markets.

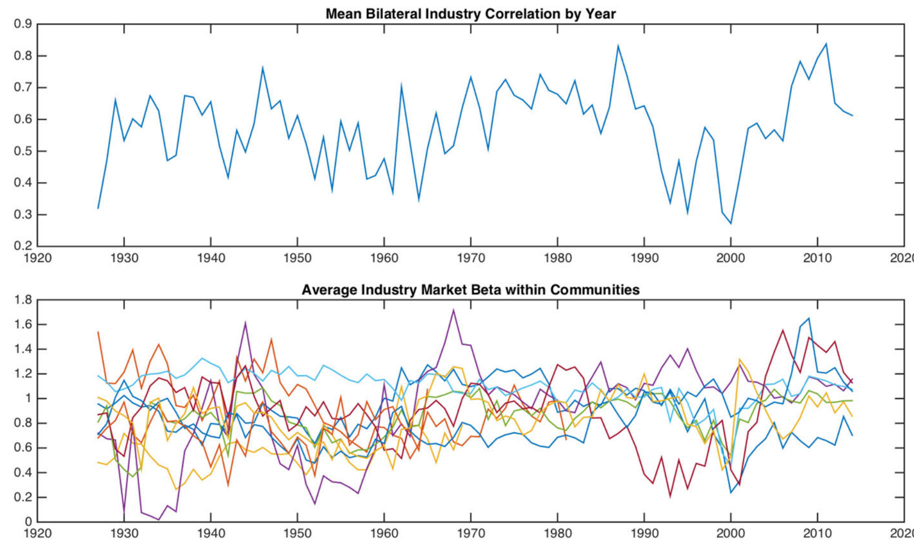

Fig. 4 *Aggregate Correlation and Industry Betas.* The top panel plots the average pairwise correlation of all 30 industries for each year in the sample. We observe that there is no discernible trend in overall industry correlation over the period that we study. The bottom panel plots the average industry market beta for each community in each year. While there are differences in market beta across communities, these differences are not persistent, suggesting that time-varying loadings on to the first principal component of returns are not sufficient to explain our findings

While the results of the analysis of industry returns in this study are interesting in their own right, this study also highlights the potential utility of these techniques for understanding time-varying correlation matrices of asset returns. It is plausible that this dynamic network-based approach could have important applications for portfolio analysis and our understanding of financial risks.

Endnotes

[1] For example, computing the column and row mean to identify the general level of comovement across the whole system yields the global network connectedness. This statistic has been used to study the global banking system (Diebold and Yilmaz 2014).

[2] These correlations are nearly all positive, but a small number (458 of the 79,200) are negative. Since most of these negative correlations are close to zero, we proceed by setting all negative correlations equal to zero.

Abbreviations

Autos: Automobiles/Trucks; Beer: Beer/Liquor; Books: Printing/Publishing; BusEq: Business equipment; Carry: Aircraft/Ships/Railroad equipment; Chems: Chemicals; Clths: Apparel; Cnstr: Construction/Construction materials; Coal: Coal; ElcEq: Electrical equipment; FabPr: Fabricated products/machinery; Fin: Banking/Insurance/Real estate/Trading; Food: Food products; Games: Recreation; Hlth: Healthcare/Medical equipment/ Pharmaceutical products; Hshld: Consumer goods; Meals: Restaurants/Hotels/Motels; Mines: Precious metals/Non-metallic/Industrial metal mining; Oil: Petroleum/Natural gas; Other: All other firms; Paper: Business supplies/Shipping containers; Rtail: Retail; Servs: Personal/Business services; Smoke: Tobacco products; Steel: Steel works; Telcm: Communication; Trans: Transportation; Txtls: Textiles; Util: Utilities; Whlsl: Wholesale For the full list of industry definitions, visit the following website and click on "Download industry definitions" next to "Portfolios": http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_30_ind_port.html

Acknowledgments

Richard Betzel (Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA) assisted in generating the visual displays.

Funding

DSB acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, and the ISI Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Availability of data and materials
The datasets analyzed during the current study are available in Ken French's data repository, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library.html.

Authors' contributions
DNR analyzed and interpreted the data with the guidance and assistance of DSB. RR played a significant role in determining the implications of the findings, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia PA, USA. [2]Department of Operations, Information and Decisions, Wharton School, University of Pennsylvania, PA, USA. [3]Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia PA, USA. [4]Department of Finance, Lundquist College of Business, Mail: Lillis Hall, University of Oregon, Eugene, Oregon, USA.

Received: 30 October 2018 Accepted: 8 April 2019
Published online: 13 June 2019

References
Achard S, Bassett D, Meyer-Lindenberg A, Bullmore E (2008) Fractal Connectivity of Long-Memory Networks. Phys Rev E Stat Nonlin Soft Matter Phys 77(3 Pt 2):036104
Aste T, Shaw X, Di Matteo T (2010) Correlation structure and dynamics in volatile markets. New J Phys 12(8):085009. https://doi.org/10.1088/1367-2630/12/8/085009
Bassett D, Wymbs N, Porter M, Mucha P, Carlson J, Grafton S (2011) Dynamic Reconfiguration of Human Brain Networks During Learning. PNAS 108:7641–7646
Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, Mucha PJ (2013) Robust detection of dynamic community structure in networks. Chaos 23(1):013142
Boss M, Elsinger H, Summer M, Thurner 4 S (2004) Network topology of the interbank market. Quant Finan 4(6):677–684
Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theor Exp 10:P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008
JBookstaber R, Kenett D (2016) Looking Deeper, Seeing More: A Multilayer Map of the Financial System. OFR Brief 6(16). https://www.financialresearch.gov/briefs/files/OFRbr_2016-06_Multilayer-Map.pdf
Coats K, Fant L (1993) Recognizing Financial Distress Patterns Using a Neural Network Tool. Financ Manag 22:142–155
Connor G, Korajczyk R (1993) A Test for the Number of Factors in an Approximate Factor Model. J Financ 48:1263–1291
Di-Matteo T, Pozzi F, Aste T (2009) The use of dynamical networks to detect the hierarchical organization of financial market sectors. Eur Phys J B 73:3–11
Diebold F, Yilmaz K (2014) On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. J Econ 182:119–134
Engle R, Ng V, Rothschild M (1992) A Multi-Dynamic-Factor Model for Stock Returns. J Econ 51:245
Fenn DJ, Porter MA, McDonald M, Williams S, Johnson NF, Jones NS (2009) Dynamic communities in multichannel data: an application to the foreign exchange market during the 2007-2008 credit crisis. Chaos 19(3):033119. https://doi.org/10.1063/1.3184538
Fenn D, Porter M, Mucha P, McDonald M, Williams S, Johnson N, Jones N (2010) Dynamical Clustering of Exchange Rates. Cornell University Library. https://arxiv.org/abs/0905.4912. Accessed 24 May 2019
Fortunato S (2011) Reuben Cohen and Shlomo Havlin: Complex Networks. J Stat Phys 142:640–641
French K (2016) Detail for 30 Industry Portfolios. Dartmouth Univ. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_30_ind_port.html
Good BH, de Montjoye YA, Clauset A (2010) Performance of modularity maximization in practical contexts. Phys Rev E Stat Nonlin Soft Matter Phys 81(4 Pt 2):046106
Holme P, Saramaki J (2015) Exploring Temporal Networks with Greedy Walks. Eur Phys J B 88
Isogai T (2016) Building Dynamic Correlation Network for Financial Asset Returns. Appl Netw Sci 1(1):7. https://doi.org/10.1007/s41109-016-0008-x
Jalili M, Perc M (2017) Information Cascades in Complex Networks. Journal of Complex Networks 5:665–693
Kenett D, Tumminello M, Madi A, Gur-Gershgoren G, Mantegna R, Ben-Jacob E (2010) Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. PLoS ONE 5:12
Kenett D, Huang X, Vodenska I, Havlin S, Eugene Stanley H (2015) Partial Correlation Analysis: Applications for Financial Markets. Quant Financ 15:569–578
Kivela M, Arenas A, Barthelemy M, Gleeson J, Moreno Y, Porter M (2014) Multilayer Networks. J Complex Networks 2:203–271
Korniyenko Y, Patnam M, del Rio-Chanon R, Porter M (2018) Evolution of the Global Financial Network and Contagion: A New Approach. Int Monet Fund. https://www.imf.org/en/Publications/WP/Issues/2018/05/15/Evolution-of-the-Global-Financial-Network-and-Contagion-A-New-Approach-45825
Mucha P, Richardson T, Macon K, Porter M, Onnela J (2010) Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328:876–878

Musmeci N, Aste T, Di Matteo T (2015) Relation between Financial Market Structure and the Real Economy: Comparison between Clustering Methods. PLoS ONE 10(3):e0116201. https://doi.org/10.1371/journal.pone.0116201
Musmeci N, Nicosia V, Aste T, Di Matteo T, Latora V (2017) The Multiplex Dependency Structure of Financial Markets. Complexity 2017
Newman MEJ (2006) Modularity and Community Structure in Networks. PNAS 103:8577–8582
Odom MD, Sharda R (1990) A Neural Network Model for Bankruptcy Prediction. IEEE, San Diego. https://ieeexplore.ieee.org/document/5726669
Pollet J, Wilson M (2010) Average Correlation and Stock Market Returns. J Financ Econ 96:364–380
Raddant M, Kenett D (2016) Interconnectedness in the Global Financial Market, Vol. 16. Office of Financial Research Working Paper. https://econpapers.repec.org/paper/zbwifwkwp/2076.htm
Sharpe WF (1963) A simplified model for portfolio analysis. Management Sci 2(9):277–293
Song W-M, Di Matteo T, Aste T (2012) Hierarchical Information Clustering by Means of Topologically Embedded Graphs. PLoS ONE 7(3):e31929. https://doi.org/10.1371/journal.pone.0031929
Tilak G, Szell T, Chicheportiche R, Chakraborti A (2011) Study of Statistical Correlations in Intraday and Daily Financial Return Time Series. Econophys-Kolkata VI. https://hal.archives-ouvertes.fr/hal-00827947/document
Traud A, Kelsic E, Mucha P, Porter M (2011) Comparing Community Structure to Characteristics in Online Collegiate Social Networks. SIAM Rev 53:526–543
Uechi L, Akutsu T, Eugene Stanley H, Marcus A, Kenett D (2014) Sector Dominance Ratio Analysis of Financial Markets. Physica A 421:488–509
Yang C, Chen Y, Hao W, Shen Y, Tang M, Niu L (2014) Effects of financial crisis on the industry sector of Chinese stock market - from a perspective of complex network. Modern Phys Lett B 28(13):1450102. https://doi.org/10.1142/S0217984914501024
Zalesky A, Fornito A, Cocchi L, Gollo L, Breakspear M (2014) Time-Resolved Resting-State Brain Networks. PNAS 111:10341–10346