Which stocks are profitable? A network method to investigate the effects of network structure on stock returns

Kun Chen *, Peng Luo, Bianxia Sun, Huaiqing Wang

Department of Financial Mathematics and Engineering, South University of Science and Technology of China, Shenzhen, China

**Highlights**

- A network method is developed to investigate the structural evolution of Chinese stock market.
- A modularity method is proposed to detect the inter-effect and intra-effect of industries.
- Regression models are developed to determine the effect of network metrics and stock returns.

**Abstract**

According to asset pricing theory, a stock's expected returns are determined by its exposure to systematic risk. In this paper, we propose a new method for analyzing the interaction effects among industries and stocks on stock returns. We construct a complex network based on correlations of abnormal stock returns and use centrality and modularity, two popular measures in social science, to determine the effect of interconnections on industry and stock returns. Supported by previous studies, our findings indicate that a relationship exists between inter-industry closeness and industry returns and between stock centrality and stock returns. The theoretical and practical contributions of these findings are discussed.

© 2015 Elsevier B.V. All rights reserved.

**1. Introduction**

According to asset pricing theory (APT), a stock’s expected returns are determined by its exposure to environmental and systematic risk [1–4]. However, traditional pricing models are mostly defined based on the features of individual stocks, such as firm size, book-to-market equity ratio, and price-to-cash flow ratio. Thus, the inter-stock and inter-industry effects on stock returns remain unknown.

Recent financial research has used network models to represent the underlying relations between industries. Most of these networks are quite simple, with industries as nodes and business trades between industries as edges. Systematic risks contained in the industry network are then investigated to explain industry returns. For example, Acemoglu et al. [5] argue that sectoral risks can be transmitted to other sectors through a network of input and output linkages in a system [5]. Starting from the idea of network structure, Aobdia et al. [6] construct an industry network based on trade flows across different industries and find that firms in central industries are more exposed to systemic risks than other firms [6]. Ahern and Jarrad (2014) further shows that systematic risks constitute the aggregation of idiosyncratic shocks and that sectors that are more central in a network of intersectoral trade usually have higher returns because they experience greater exposure to systematic risk [7].
Complex networks based on individual stock indicators, such as price, volume, and volatility, have also attracted attention from researchers in physics and statistics. In such a network, individual stocks are nodes, and the correlations between stock returns are usually used as edges. Such studies focus on the features and structures of the complex network itself, and few financial problems are discussed. In our study, we apply the network construction method to a complex network in order to investigate the effects of network structure on stock returns. Specifically, we investigate the following two detailed research questions:

(1) Do industry relations in a network affect industry returns?
(2) Do stock relations in a network affect stock returns?

Regarding the first question, previous studies [5–7] have shown that risk can be transmitted between industries, leading to a return framework based on industry relations. Asness, Porter, and Stevens [8] include both within-industry and across-industry variables to model stock returns [8]. We similarly construct two industry-related components based on the network structure of the stock market, namely, intra-industry closeness and inter-industry closeness, to explain returns.

Regarding the second question, few studies have examined the relationship between individual stock returns and stock relations. However, Ahern and Jarrad (2014) has shown that industries in more central network positions tend to have higher returns [7]. To determine whether this finding also applies to individual stocks, we measure stocks' network centrality in this study.

The rest of the paper is organized as follows: Section 2 reviews relevant prior work on complex networks. Section 3 describes the analytical method. Detailed results are then presented in Section 4. Finally, we discuss the findings and conclude the paper in Section 5.

2. Complex networks

2.1. Complex network theory and its application

Complex network theory originates from discrete mathematics and graph theory, and it has developed over many decades as a theoretical framework for understanding the structural characteristics of networks [9]. Two well-known classes of complex networks are scale-free networks [10] and small-world networks [11,12]. Both of these types of networks are characterized by specific structural features—power-law degree distributions for the former and short path lengths and high clustering for the latter. As the study of complex networks has continued to increase in importance and popularity, many other aspects of network structure have attracted research attention, and research on complex networks has been used in fields such as computer science, biology, communications and engineering, and the social sciences [13].

2.2. Complex network analysis on stock markets

The dynamic nature of a financial market can be mapped as a complex network. For a stock market, networks are constructed based on the correlations of stock price returns. Previous research has examined the common topological and statistical features of such networks (Table 1). For example, Lee et al. [14] show that a network based on the Korean stock exchange is a scale-free network [14]. Boginski et al. [15] and Huang et al. [16] further show that stock networks follow a power-law model [15,16]. Given these findings, cliques and independent sets in stock networks have been identified [15], providing a technical method for determining stock clustering. In addition to these studies, much recent work uses network structure to elucidate the underlying properties of stock markets. For example, Tabak et al. [17] find that the relative importance of different industries within a stock network varies [17]. Tse et al. [18] study different stock network structures by using different construction methods, ultimately developing an index method [18]. Lee et al. [19] use the minimum spanning tree method on Korean stock market data and find that higher market volatility is associated with a denser minimum spanning tree of stocks [19]. Peron et al. [20] view stock markets as evolving systems [20]. They propose an entropy-related measure to analyze the topological and dynamic evolution of financial networks and to identify shared behaviors between stocks during financial crises.

In addition to price return networks, networks based on relations between investors and transactions in a stock market are also constructed. For example, Li and Wang [21] use networks to represent various patterns of HSI fluctuation, and based on important topological nodes, they extract hidden patterns of such fluctuation [21]. Moreover, Bakker et al. [22] consider the psychological factors that affect market valuation to construct “investor trust networks” [22]. Their simulation results show that real-life trust networks can significantly delay the stabilization of a market. Jiang et al. [23] construct a trading network between investors of a highly liquid stock and find that the trading network has a power-law degree distribution and disassortative architectures [23]. Furthermore, they extend their work on trading networks to investigate financial research problems. One representative work is to identify the abnormal trading motifs and stock manipulations based on the network topology [24]. It contributes to the field of market surveillance. Another work is to study the correlations between market indicators and network topological metrics to unveil how the trading behaviors affect the market performance [25]. It contributes to the behavioral finance research.
The basic idea is to use positive networks to illustrate the co-gain relations between stocks and to use negative networks if defined as follows:

\[ V = \text{stock network}, \quad \theta \geq \text{threshold value} \]

where \( \theta \) is the threshold value. Let the set of stocksberepresented by the vertices ofthenetwork. For the connectionsof the network, we havetwo different ways to construct networks. In the first method, we construct a network by using the abnormal returns of stocks in each year. The edge connecting two vertices is determined by the correlation between two abnormal return series corresponding to two stocks. Specifically, the correlations between two individual stocks are considered in terms of the matrix \( C \), which is obtained from the following formula:

\[
c_{ij} = \frac{\langle ar_i, ar_j \rangle - \langle ar_i \rangle \langle ar_j \rangle}{\sqrt{(\langle ar_i^2 \rangle - \langle ar_i \rangle^2) (\langle ar_j^2 \rangle - \langle ar_j \rangle^2)}}
\]

where \( ar \) denotes the abnormal return and the bracket indicates a temporal average over the period. Moreover, a certain threshold value \( \theta, 0 \leq \theta \leq 1 \) is specified, and an undirected edge is added between i and j if the absolute value of \( c_{ij} \) is greater than or equal to \( \theta \). In line with Zhang et al. [30], we keep the anti-relationship whose absolute value exceeds the threshold and use the absolute value of the correlation as the weight assigned to the edge in network. Let \( G = (V, E, W) \) represent the stock network, where \( V \) denotes the set of vertices, \( E \) represents the edges, and \( W \) is the edge's weight. \( W \) is defined as follows:

\[
W = \begin{cases} 
    w_{ij} = |c_{ij}|, & i \neq j \text{ and } |c_{ij}| \geq \theta \\
    w_{ij} = 0, & \text{else}
\end{cases}
\]

If \( w_{ij} \neq 0 \), there is an edge between node i and node j.

In the second method, we construct positive and negative networks separately by selecting part of the abnormal return data. The basic idea is to use positive networks to illustrate the co-gain relations between stocks and to use negative networks.
Table 2
Abnormal returns of two stocks.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>000006.SZ</td>
<td>0.0066</td>
<td>0.0370</td>
<td>−0.0134</td>
<td>0.0053</td>
<td>−0.0158</td>
<td>−0.0054</td>
<td>−0.0075</td>
<td>−0.0133</td>
<td>0.0154</td>
<td>0.0319</td>
</tr>
<tr>
<td>000007.SZ</td>
<td>−0.0155</td>
<td>0.0803</td>
<td>−0.0036</td>
<td>0.0603</td>
<td>0.01415</td>
<td>−0.0262</td>
<td>−0.0122</td>
<td>0.0602</td>
<td>−0.0179</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

Table 3
Returns for positive networks.

<table>
<thead>
<tr>
<th>Date</th>
<th>Jan. 8</th>
<th>Jan. 10</th>
<th>Jan. 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>000006.SZ</td>
<td>0.0370</td>
<td>0.0053</td>
<td>0.0319</td>
</tr>
<tr>
<td>000007.SZ</td>
<td>0.0803</td>
<td>0.0603</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

Table 4
Returns for negative networks.

<table>
<thead>
<tr>
<th>Date</th>
<th>Jan. 7</th>
<th>Jan. 9</th>
<th>Jan. 14</th>
<th>Jan. 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>000006.SZ</td>
<td>0.0066</td>
<td>−0.0134</td>
<td>−0.0054</td>
<td>−0.0075</td>
</tr>
<tr>
<td>000007.SZ</td>
<td>−0.0155</td>
<td>−0.0036</td>
<td>−0.0262</td>
<td>−0.0122</td>
</tr>
</tbody>
</table>

to illustrate the co-loser relations between stocks. In Table 2, we use daily abnormal returns of two stocks (000006.SZ and 000007.SZ) from January 7, 2013, to January 18, 2013, to demonstrate our method.

Days with positive abnormal returns for the two stocks are included in a positive network (as shown in Table 3). Thus, the positive network includes 3 days (Jan. 8, Jan. 10, and Jan. 18) in this example. In contrast, days with negative abnormal returns for both stocks are included in a negative network (as shown in Table 4). The negative network hence includes 4 days (Jan. 7, Jan. 9, Jan. 14, and Jan. 15). Three other days (Jan. 11, Jan. 16, and Jan. 17) are disregarded because no simultaneous relationship between the two abnormal returns exists for these days.

Next, we calculate the correlations between the two stocks based on formula (5). Here, two correlations can be obtained: \( (c'_ij) \) is the correlation for the days on which the abnormal returns of the two stocks are both positive, and \( (c''_ij) \) is the correlation for the days on which abnormal returns of the two stocks are both negative. In addition, we use a threshold \( \theta' \). When the total number of days \( (n', n'') \) on which two stocks' abnormal returns are both positive (negative) exceeds the threshold, the correlation between the returns is considered to be effective. In this way, we obtain two networks: the positive network \( G' = (V, E', W') \) and the negative network \( G'' = (V, E'', W'') \), where

\[
W' = \begin{cases} 
  w'_{ij} = c'_ij, & i \neq j \text{ and } n' \geq \theta' \\
  w'_{ij} = 0, & \text{else}
\end{cases}
\]

\[
W'' = \begin{cases} 
  w''_{ij} = c''_ij, & i \neq j \text{ and } n'' \geq \theta' \\
  w''_{ij} = 0, & \text{else}
\end{cases}
\]

If \( w'_{ij} \neq 0 \) or \( w''_{ij} \neq 0 \), there is an edge between node \( i \) and node \( j \) in the positive network or in the negative network, respectively.

3.2. Centrality

Centrality refers to the location of points in a network [31]. The point at the center of a star or at the hub of a wheel is the most centrally located position. Accordingly, we can use a centrality measure to determine the core nodes in a stock network and to identify the most connected stocks in a stock market.

As one of the most widespread centrality measures, degree centrality is defined as the number of direct ties that involve a given node [31]. In analyses of weighted networks, the degree of centrality has generally been extended to the sum of weights. This measure has been formalized as follows [32]:

\[
C^W_D(i) = \sum_{j=1}^{N} w_{ij}
\]

where \( W \) is the weighed adjacency matrix, in which the value of \( w_{ij} \) is the weight of the tie between node \( i \) and node \( j \).

3.3. Modularity

Modularity is defined as the fraction of edges that fall within communities minus the expected value of the same quantity if the edges are assigned at random, conditional on the given community memberships and the degree of the vertices [33]. In previous research, modularity has mainly been used to evaluate community detection [34,35]. In our paper, we introduce modularity to measure the strength of the connection between nodes in a group. According to the industry to which the stocks belong, we divide stocks into different groups and use modularity to calculate the intra- and inter-group closeness.
When computing the closeness of two groups, we treat the two groups as a whole to yield the modularity value of the whole group.

Let $c_i$ be the community to which node $i$ is assigned, and let $k_i$ denote the degree of node $i$. Then, the modularity $Q$ is given by [36]

$$Q = \frac{1}{2m} \sum_{ij} \left[ w_{ij} - \frac{k_i k_j}{2m} \right] \delta (c_i, c_j)$$

(8)

where the $\delta$-function $\delta (u, v)$ is 1 if $u = v$ and 0 otherwise and $m = 0.5 \sum_{ij} w_{ij}$ is the sum of weights in the whole network. However, this formula for modularity cannot be used directly in our paper; thus, we modify it to measure intra- and inter-industry closeness:

$$Q'_x = \frac{1}{2m} \sum_{ij} \left[ w_{ij} - \frac{k_i k_j}{2m} \right] \delta' (c_i, c_j)$$

(9)

$$Q''_{xy} = \frac{1}{2m} \sum_{ij} \left[ w_{ij} - \frac{k_i k_j}{2m} \right] \delta'' (c_i, c_j)$$

(10)

where $Q'_x$ is the closeness within industry $x$ and $Q''_{xy}$ is the closeness between industry $x$ and industry $y$. The $\delta'$ and $\delta''$ functions must obey the following:

$$\delta' (c_i, c_j) = \begin{cases} 1, & \text{if } c_i = c_j = x \\ 0, & \text{else} \end{cases}$$

$$\delta'' (c_i, c_j) = \begin{cases} 1, & \text{if } c_i = x, c_j = y \text{ or } c_i = y, c_j = x \\ 0, & \text{else} \end{cases}$$

4. Results

4.1. Data sets

We focus on all A-Shares stocks listed on both Shenzhen and Shanghai Stock exchanges to uncover the potential connections between industry/stock positions and their returns. Chinese stock market was established in 1992. There are only 28 stocks on market till the end of 1993. But from 1994 to 2013, the stock number increases up to 2445. Therefore, we use the daily stock closing price from January 1, 1994, to December 31, 2013, to construct the networks. In total, 20 stock networks are constructed for the Chinese stock market over the 20 years of the sturdy period: one network for each year. These time periods almost cover the entire development of the capital market in China. The value of 0.4 is selected as the optimal threshold value $\theta$: if the value is too small, there will be too many edges in the networks and the random effects will increase, but if the threshold value is too large, the edges in the networks will be very sparse. Table 5 presents the number of nodes and edges in the stock market network for each year. The connections between stocks are extremely dense in 2007 and 2008, as indicated by the large numbers of nodes and edges in the networks for these two years. Apparently, the Chinese stock market experienced a turbulent change from a bull market to a bear market during this period, similar to the majority of stock markets worldwide.

4.2. Network effects on industries

4.2.1. An overview of industries and networks

In these 20 networks, we choose the top-100 stocks based on the measure of centrality. The top-100 stocks are further divided into 28 industrial sectors in line with the Shenwan Industry Classification Standard, which is widely used in industrial analysis of Chinese financial markets. We average and scale the centrality values by year and present them with China’s yearly GDP growth rate in Fig. 1. Generally, the trend of the average centrality value of core stocks tracks that of the GDP growth rate, i.e., national economic development. Such a correlation indicates that if the economy is growing, the connections between stocks may be strong, and when the economy is shrinking, stocks may only be loosely connected with each other.
Among other information, Fig. 1 reveals some interesting phenomena. First, the real estate industry, shown in green, is quite large for most of the years in the 20-year sample. An industry’s centrality in a stock market network can reveal, to some extent, the position or the power of that industry in the whole market. Therefore, real estate stocks are highly important for the Chinese stock market, especially during the period in which China’s economy is growing.\(^1\) Since 2011, the centrality of real estate stocks has been decreasing, which coincides with the slow development of this industry in the real economy.

Second, the computer industry emerges suddenly in 2013, as shown in blue in Fig. 1. In 2013, the centrality of the computer industry constitutes almost half of the centrality of all industries, implying a strong connection between the computer industry and other industries in the market. By contrast, the computer industry’s centrality was not notable before 2013. This observation matches what occurred in the Chinese economy. Along with the weak performance of traditional industries, such as real estate, in the last two years for the sample, the computer industry has developed rapidly. Reflecting such development, stocks in the computer industry have become more active and attractive to investors, leading to much higher centrality in the stock market network.

We further compare the average centrality value of each industry, as shown in Fig. 2. Unsurprisingly, we find that the real estate industry holds a very central position in the Chinese stock market for most of the years in the sample period and that its influence has decreased only in the last two years. The computer industry is another core industry of the Chinese stock market, with a significantly increasing influence in recent years. In addition, Fig. 2 shows the important position of the healthcare industry in China, whose centrality in the network is only lower than that of the real estate industry. Moreover, the healthcare industry begins to hold market power only in 2008. This industry can be expected to become increasingly more important given the aging Chinese population.

4.2.2. Modularity effect on industries

Using the modularity equations defined in Section 3.3, we calculate the intra-industry closeness and inter-industry closeness, which measure the connections of stocks within an industry and the connections of stocks between industries, respectively. Examples are presented in Figs. 3 and 4. Fig. 3(a) displays a network of the computer industry in 2013. The network includes 129 nodes and 1202 edges, and the intra-industry closeness measure \(Q'_x = 0.01118\). Fig. 3(b) displays a network of the national defense industry in 2013. The network includes 103 nodes and 27 edges, and the intra-industry closeness measure \(Q'_x = 0.00233\). Clearly, more connections between stocks exist in the computer industry than in national defense industry. Fig. 4(a) displays a network between the computer industry and the electronics industry in 2013. The network includes 129 nodes from the computer industry and 137 nodes from the electronics industry; further, 872 edges are established between two industries, and the inter-industry closeness measure \(Q'_{xy} = 0.00620\). Fig. 4(b) displays a network between the computer industry and the real estate industry in 2013. The network includes 129 nodes from the computer

---

1 The centralities of real estate in 2006, 2007, and 2008 are small. This is because Chinese stock market is a bull market in these periods. In most cases, if the stock market is booming, the real-estate market is shrinking, because money is flowing to capital market and invested on various stocks.
industry and 144 nodes from the real estate industry. However, only 2 edges exist between two industries; the inter-industry closeness measure $Q_{xy}^{''} = -0.00146$.

Next, we investigate the relations between industry returns and industry inter-connections. Because we use the Shenwan industry classification standard and industry index, which was first released in 2000, the experimental data set covers the period from 2000 to 2013. Using the method of calculating intra- and inter-industry closeness described above, we obtain the intra- and inter-industry closeness for 28 industries in 13 networks. Because we have to determine which model (fixed-effect model or mixed-effect model) is suitable for the analysis, we first calculate the fixed-effect model. The F-test for individual effects is not significant ($p = 0.64$), so we adopt the mixed-effect model, which can be written as follows:

$$IAR_{i,t} = \alpha + \beta \text{Intra\_industry}_{i,t} + \gamma \text{Inter\_industries}_{i,t} + \delta IAR_{i,t-1} + \epsilon$$

where $IAR_{i,t}$ is the abnormal return of industry $i$, $(i = 1, 2, \ldots, 28)$ in year $t$, $(t = 2001, \ldots, 2013)$; $\text{Intra\_industry}_{i,t}$ denotes the intra-industry closeness for industry $i$ in the stock network in year $t$, calculated by Eq. (9); $\text{Inter\_industries}_{i,t}$ denotes the inter-industry closeness between industries $i$ and $j$ at time $t$, calculated by Eq. (8); and $\epsilon$ is the error term. 

Fig. 2. A comparison of average centrality between industries.

Fig. 3. Examples of intra-industry networks.
denotes the average value of closeness between industry $i$ and all the other industries in year $t$, calculated by Eq. (10); $IAR_{i,t-1}$ is the abnormal return of industry $i$ in year $t - 1$; and $\varepsilon$ denotes all other relevant factors.

Table 6 displays the regression results. The regression model is valid, as the F-test result is 9.47, which is highly significant. Interestingly, $\text{Inter\_industries}$ has a significant positive effect on an industry's abnormal return, whereas the effects of $\text{Intra\_industry}$ and $IAR_{i,t-1}$ are not significant. Moreover, the coefficient for the variable $\text{Inter\_industries}$ is 0.94, indicating that a one-unit increase in an industry’s closeness across industries results in a nearly 1-percentage-point increase in the industry’s abnormal returns, when other factors are held constant. This result indicates that an industry’s abnormal returns increase as the connections between the industry and other industries increase. This finding is consistent with previous studies that have constructed networks based on industry trades. Further, Acemoglu et al. [5] show that system risk can be transmitted between industries [5]. Therefore, a more connected industry aggregates more risk, which is generally compensated by higher returns. Our work differs from that of Asness, Porter, and Stevens [8], who adopt within-industry and across-industry variables to measure differences in firms’ characteristics within and between industries [8]. In this study, our modularity method measures the closeness of industries by stock performance.

4.3. Network effects on individual stocks

4.3.1. An overview of stocks and networks

To investigate stock distributions in networks, we first divide stocks into three groups according to their market capitalization, where the top 1/3 stocks in market capitalization are defined as large market capitalization stocks, the middle 1/3 stocks are defined as medium market capitalization stocks, and the bottom 1/3 stocks are defined as small market capitalization stocks. We then choose the top 10% of stocks in each group according to their centrality in a network. As Fig. 5 shows, among stocks in a core position in a network, the average centrality of small market capitalization stocks is larger than that of medium and large market capitalization stocks. Thus, small market capitalization stocks are more connected to other stocks in the Chinese stock market. This trend has been especially notable since 2002. The same results are also obtained when we use the top 20% and top 30% of stocks according to their centrality. However, the differences in centrality between small market capitalization stocks and large market capitalization stocks decrease as the percentage threshold increases.
4.3.2. Centrality effects on stocks

To investigate the relations between stock returns and centrality, we construct a positive network and a negative network according to the method described in Section 3.1. The threshold $\theta'$ is set at 60 for two reasons: 60 data points are sufficient to generate meaningful correlations, and 60 days cover 1/4 of the observations for a year, as there are approximately 240 trading days in a year. Fig. 6 presents two example networks of the top 10% of stocks based on centrality. The black node in Fig. 6(a) holds the central position in the network; it has a centrality value of 269.71 and 2167 edges linking it with other nodes. The blue node has a centrality value of 217.04, and there are 1875 edges linked with it. In the negative networks shown in Fig. 6(b), the red node holds the central position; it has a centrality value of 874.95 and 2054 edges linking it with other nodes. The blue node has a centrality value of 703.66, and there are 1699 edges linked with it. These results show that a higher centrality value is associated with a higher number of connections for a stock.

We also investigate the relations between a stock’s centrality and its abnormal returns in positive and negative networks. Abnormal returns in positive networks are the sum of the positive abnormal returns for a stock in each year. Conversely, the abnormal returns in negative networks are the sum of the negative abnormal returns for a stock in each year. Stock centrality is also calculated in positive and negative networks. First, we perform a random-effect test on the data. As shown in Table 7, the random effect is not significant, and the Hausman specification test rejects the null hypothesis that the difference in coefficients between the random-effect and fixed-effect estimators is not systematic.
networks. From the results in Table 9, it is obvious that the differences are significant when network rolling ahead one month. Also, the modularity effect is significant, as the overall Q value between industries. The results are shown in Table 9. We find the differences of one industry’s modularity value, including intra-effect and inter-effect, are not significant when network rolling ahead one month. So the modularity effect is quite stable in one month period.

Secondly, we do experiments on individual stocks’ centrality value to measure the stability of stocks’ positions in networks. From the results in Table 9, it is obvious that the differences are significant when network rolling ahead one month. Also, the modularity effect is significant, as the overall Q value between industries. The results are shown in Table 9. We find the differences of one industry’s modularity value, including intra-effect and inter-effect, are not significant when network rolling ahead one month. So the modularity effect is quite stable in one month period.

4.4. Stability of networks

To further investigate the stability of network structure, we compare aforementioned network metrics in different time periods with a sliding time window (as shown Fig. 7). Networks are build based on 12 months’ data as before. The time window is initially chosen as 1 month. We do experiments using data in 2012 and 2013.

In the experiments, we do pair-wise t test on the same variables in different networks. Firstly, we examine the modularity Q value between industries. The results are shown in Table 9. We find the differences of one industry’s modularity value, including intra-effect and inter-effect, are not significant when network rolling ahead one month. So the modularity effect is quite stable in one month period.

Secondly, we do experiments on individual stocks’ centrality value to measure the stability of stocks’ positions in networks. From the results in Table 9, it is obvious that the differences are significant when network rolling ahead one month. Also, the modularity effect is significant, as the overall Q value between industries. The results are shown in Table 9. We find the differences of one industry’s modularity value, including intra-effect and inter-effect, are not significant when network rolling ahead one month. So the modularity effect is quite stable in one month period.

Table 7
Fixed-effect versus random-effect estimator: Diagnostic results.

<table>
<thead>
<tr>
<th></th>
<th>Positive networks</th>
<th>Negative networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch–Pagan LM test</td>
<td>$P = 1.000$</td>
<td>$P = 1.000$</td>
</tr>
<tr>
<td>Hausman specification</td>
<td>$P = 0.000^{**}$</td>
<td>$P = 0.000^{**}$</td>
</tr>
</tbody>
</table>

Denotes significance at the 1% level.

Table 8
Fixed-effect static panel data regression results.

<table>
<thead>
<tr>
<th></th>
<th>Positive networks</th>
<th>Negative networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.2105*** (66.68)</td>
<td>−0.0595*** (−21.70)</td>
</tr>
<tr>
<td>Centrality</td>
<td>1.9570*** (166.97)</td>
<td>−2.0909*** (−210.61)</td>
</tr>
<tr>
<td>Abnormal return (−1)</td>
<td>0.4293*** (132.01)</td>
<td>0.3143*** (102.23)</td>
</tr>
<tr>
<td>Overall $R^2$</td>
<td>0.7682</td>
<td>0.8390</td>
</tr>
<tr>
<td>Between $R^2$</td>
<td>0.9669</td>
<td>0.9694</td>
</tr>
<tr>
<td>Number of observations</td>
<td>50740</td>
<td>50740</td>
</tr>
<tr>
<td>Number of groups</td>
<td>2537</td>
<td>2537</td>
</tr>
<tr>
<td>F-test (all $u_i,t = 0$)</td>
<td>1.87***</td>
<td>1.32***</td>
</tr>
</tbody>
</table>

Denotes significance at the 1% level.

These findings suggest that the random-effect model is not suitable for our regression; thus, we choose the fixed-effect model. Then, we have the following regression formula:

\[
AR_{i,t}^P = \alpha_i^P + \beta_i^P C_{i,t}^P + \gamma_i^P AR_{i,t-1}^P + \epsilon_{i,t}^P
\]

\[
AR_{i,t}^N = \alpha_i^N + \beta_i^N C_{i,t}^N + \gamma_i^N AR_{i,t-1}^N + \epsilon_{i,t}^N.
\]

These two equations provide a linear specification of the positive abnormal return ($AR^P$) and negative abnormal return ($AR^N$) of stock $i$ in year $t$. The $AR^P$ or $AR^N$ is a function of the stock's fixed effect ($\alpha_i^P$, $\alpha_i^N$), the stock's centrality in positive and negative networks ($C_i^P$, $C_i^N$), the positive and negative abnormal returns in year $t - 1$ ($AR_{i,t-1}^P$, $AR_{i,t-1}^N$), and all other factors ($\epsilon_i^P$, $\epsilon_i^N$) that affect the positive or negative abnormal returns.

The regression results are presented in Table 8. Both regression models for the positive and negative networks are highly significant, as the overall $R^2$ and the between $R^2$ are 0.768 and 0.967, respectively, for positive networks and 0.839 and 0.969, respectively, for negative networks. Further, the centrality of a stock has a significant effect on the stock’s abnormal returns in both positive and negative networks; thus, the inter-connections between stocks or the positions of stocks in networks determine stock returns. In addition, the identified centrality effects are different in positive and negative networks. In positive networks, stock centrality has a positive effect on abnormal returns, whereas in negative networks, stock centrality has a negative effect on abnormal returns. Notice that the coefficients for the variable centrality in the two regressions have almost the same absolute value (1.957 and 2.091, respectively), implying a similar magnitude of impact of centrality on stock returns, even though the direction of influence is different.

These results differ from previous findings on industry returns. For instance, Ahern and Jarrad (2014) uses a direct graph to simulate the economic flows between industries and finds that centrality has a positive effect on industry returns [7]. In our research, we use the correlations between stock abnormal returns to construct networks. Because the networks are non-directed, we separate then into sub-networks, positive networks and negative networks, to perform regressions on stock returns. Therefore, our network construction method is completely different from that of Ahern and Jarrad [7]. Regarding our findings, we investigate the co-gain and co-lose effect in stock networks, and we find that the stocks with a more central position in a positive network gain more abnormal returns when the market is booming, whereas stocks with more links with other stocks in a negative network lose more when the market is down.
month. But when we reduce the time window to one week (as shown in Table 10), many differences become not significant. It means the stocks’ positions are more stable in a short time period. This finding is reasonable, because interactions between industries are mostly caused by the nature and states of industries, while individual stocks’ correlations are quite random based on stock price, so industrial effect should be more stable than individual stock effect.

5. Discussion and conclusion

Based on APT and research on industry networks in the financial domain, we expect industry and stock returns to be related to connected structures among industries and stocks in network models. In this study, we construct a complex network based on the correlations between stocks’ abnormal returns to investigate such a relation. For this purpose, we use stock data on the Chinese stock market for the period from 1994 to 2013 to construct networks. Using the 20 years of stock data, we find that the average centrality value of the top-100 stocks tracks China’s GDP growth rate; thus, the degree of connection between stocks in the stock market can reflect the development of the real economy in China to some extent. Another finding is that stocks with smaller market capitalization tend to be located in more central positions in networks.

Further regression analyses show that an industry’s returns are significantly positively affected by the degree of inter-industry connection for that industry in a network. In addition, stock returns are significantly influenced by a stock’s position, as measured by centrality, in a network. Moreover, the direction of the effect of a stock’s centrality on the stock’s returns differs depending on whether the network is positive or negative, but the magnitude of the effect is the same regardless of whether the network is positive or negative. Our findings on the industry level are supported by Ahern and Jarrad [7]. Ahern and Jarrad [7] employed the economic trade-flow data on 479 industries in America to construct the network, and finds that more central industries have more stock returns. We use complex networks and modularity methods on individual stocks, and also conclude that more interactions one industry has with other industries, it will have more stock returns. On the stock level, our findings are novel because few previous works have been done on relations between individual stock returns and stock network topologic metrics.

Our findings have both theoretical and practical implications. Theoretically, introducing a network model to APT could provide a special perspective for investigating the structural features of industries and stocks based on an overview of the
market. By providing an innovative method for measuring the inter-relations between industries and individual stocks, this work may inspire a new APT model that incorporates the interaction factors related to stock markets. Practically, the findings are important for portfolio management. In contrast to other portfolio construction strategies (e.g., picking stocks based on their performance or market information), we provide an alternative method to select stocks based on their positions in networks. This method is important because it provides a relative measure that encompasses all market stocks.

Our study nevertheless has limitations. The study aims to explain not stock returns but the impact of stock and industry interaction effects on stock returns. In addition, many other factors in standard APT models (e.g., firm size, book-to-market equity ratio, and price-to-cash flow ratio) are not included in our regression models. Notably, the findings are drawn exclusively from Chinese stock market data, and further studies using data from other stock markets are thus needed to support our findings.

References