Physica A 523 (2019) 747–756

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Structural properties of statistically validated empirical information networks

Rui-Qi Han [a,b,c], Ming-Xia Li [b,d], Wei Chen [e], Wei-Xing Zhou [a,b,f,*], H. Eugene Stanley [c]

[a] Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China
[b] Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China
[c] Department of Physics and Center for Polymer Studies, Boston University, Boston, MA 02215, USA
[d] Research Institute of Sports Economics, East China University of Science and Technology, Shanghai 200237, China
[e] Shenzhen Stock Exchange, Shenzhen 518010, China
[f] Department of Finance, East China University of Science and Technology, Shanghai 200237, China

HIGHLIGHTS

- We construct the empirical information network (EIN) using SZSE stock data.
- The statistical validated EINs are obtained.
- The raw EIN exhibits a disassortative mixing pattern.
- The statistically validated EIN is assortative.
- The properties of the giant components of EINs are investigated.

ARTICLE INFO

Article history:
Received 8 December 2018
Received in revised form 14 January 2019
Available online 9 March 2019

Keywords:
Econophysics
Empirical information network
Statistically validated networks
Mixing patterns
Trading behavior

ABSTRACT

We construct the empirical information network (EIN) of traders using the order flow data of the constituent stocks of SZSE 100 Index in 2013. A statistical validation method is applied to the edges of the network to filter out noises and uncover the intrinsic interaction behaviors of traders. We investigate the correlation between topological structures and statistical properties for their largest connected components. We find that the statistical validated network shows an assortative mixing pattern while the original network exhibits a disassortative mixing pattern. We consider two definitions of edge weight for comparison but there is no significant difference in a same network. We also analyze the mutual relationships among node degree, edge weight and node strength.

1. Introduction

Financial markets are heterogeneous complex systems where the trading behaviors of traders are influenced by the structure of the underlying network [1,2]. Network theory provides us a promising tool to detect how diverse information signals diffuse over time among traders [3,4]. Information networks play a crucial role in the functioning of financial markets [3,5–7]. Ozsoylev et al. proposed an empirical information network (EIN) to identify the trading behavior in

* Correspondence to: 130 Meilong Road, P.O. Box 114, School of Business, East China University of Science and Technology, Shanghai 200237, China.
E-mail address: wxzhou@ecust.edu.cn (W.-X. Zhou).

https://doi.org/10.1016/j.physa.2019.03.010
0378-4371/

the entire stock market. They validated that the EIN captures information diffusion, which is suitable for mapping out the structure of true information networks. They found extensive evidence of frequent communication information exchange among stock market traders [4]. Feng and Seasholes found that Chinese trades are highly correlated when divided geographically, consistent with local communication among traders [8]. Chung et al. considered information flow as directional and compute the centralization of each firm's trader network as the proxy for information diffusion [9].

The concept of EIN is relevant to but different from trading networks, which are constructed based on equity transactions [10,11]. Some structural properties of trading networks have been investigated [10,12], which are correlated with some financial variables [13]. Trading behaviors are hidden in trading networks [14,15]. Hence, trading network analysis has the potential to distinguish manipulated stocks [16], identify abnormal network motifs suggesting possible stock manipulations [17], and predict stock price [18].

The availability of high frequency data and computational capacity of processing them provide us the opportunity to the large complex network analysis. However, uncovering informative structures of the underlying system has been constrained by the problem of filtering out spurious information in large-scale weighted network with strong heterogeneity [19]. Tumminello et al. introduced a filtering method for the spurious links by a statistical validation test in bipartite complex networks [20]. This method has been used in many different complex systems. Based on the co-occurrence of the trading activity of the traders of Nokia stock, the statistically validated method has been applied to identifying the clusters of traders and inferring their trading strategies [1]. Li et al. utilized the approach to the communication network constructed from mobile call records and performed a comparative analysis for the statistical properties of the original and the Bonferroni networks [21,22].

In this paper, we follow Ref. [4] to construct the empirical information network (EIN) based on traders' trading records on the constituent stocks of the SZSE 100 Index in 2013. Our investigations focus on the correlation between topological structures and statistical properties. We apply the Bonferroni filtering [20] to the empirical information network and find significant differences in the statistical property between the original and Bonferroni networks. For example, the statistically validated network becomes assortative, while the original network exhibits a disassortative mixing pattern. The rest of this paper is organized as follows. Section 2 describes the data set and the Bonferroni network method used in this work. Section 3 investigates the basic network characteristics and Section 4 studies the mutual correlations among them. Section 5 discusses and summarizes the results.

2. Statistically validated empirical information network

We use the order flow data of the constituent stocks of the SZSE 100 Index over a period of 12 months (238 trading days), from 4 January 2013 to 31 December 2013. The SZSE 100 Index focuses on the large-cap sector of the market and consists of the top 100 A-share listed companies listing and trading on the Shenzhen Stock Exchange (SZSE) of China ranked by total market capitalization, free-float market capitalization and turnovers. There are 381,345 active traders trading the 100 stocks during the sample period. To protect the privacy of traders, each trader is identified by a surrogate number following the order of trading time. For each trading day, we construct the daily trader information network where nodes are traders. We connect two traders with an undirected link if and only if traders i and j trade the same stock in the same direction at least three times within a time window of $\Delta t = 1$ min [4]. We can quantify the weight of edge (i, j) as the number of times of trading the same stock in the same direction within $\Delta t = 1$ min for the two traders.

We construct the empirical information network with nodes covering all the traders during the whole year. An edge is drawn between the traders i and j as long as they were connected for at least once in any trading day of 2013. There are two definitions for the edge weight. The number-of-days based edge weight w_{ij}^D is total number of daily networks in which the traders i and j are connected. The number-of-times based edge weight w_{ij}^T is the cumulative sum of the edge weights between i and j in all the daily networks. The empirical information network has 381,345 nodes and 8,134,541 edges, containing 2357 components. There are 376,242 nodes and 8,130,953 edges in the giant component (GCEIN) of the EIN.

For each edge in the network, we perform a statistical test to check whether the edge is statistically validated against a null hypothesis assuming the heterogeneity of random matching among traders. Edges that fail to reject the null hypothesis are removed together with the nodes that become isolated. The statistical test is implemented as follows. We define N as the sum total of all the edge weights in the network. Let us denote N_i (N_j) as the sum of weights of the edges incident to the node i (j). N_{ij} is the weight of the edge (i, j). Assuming that X is the number of days (times) when we observe the co-occurrence of trading behavior of two traders, the probability of observing X days (times) for the traders i and j is described by the hypergeometric distribution [20,23]

$$H(X|N, N_i, N_j) = \frac{C_{N_i}^{X} C_{N-N_i}^{N_j-X}}{C_N^{N_j}}, \tag{1}$$

where $C_{N_i}^{X}$ is a binomial coefficient. We can associate a p-value to the observed N_{ij} as follows:

$$p(N_{ij}) = 1 - \sum_{X=0}^{N_{ij}-1} H(X|N, N_i, N_j). \tag{2}$$

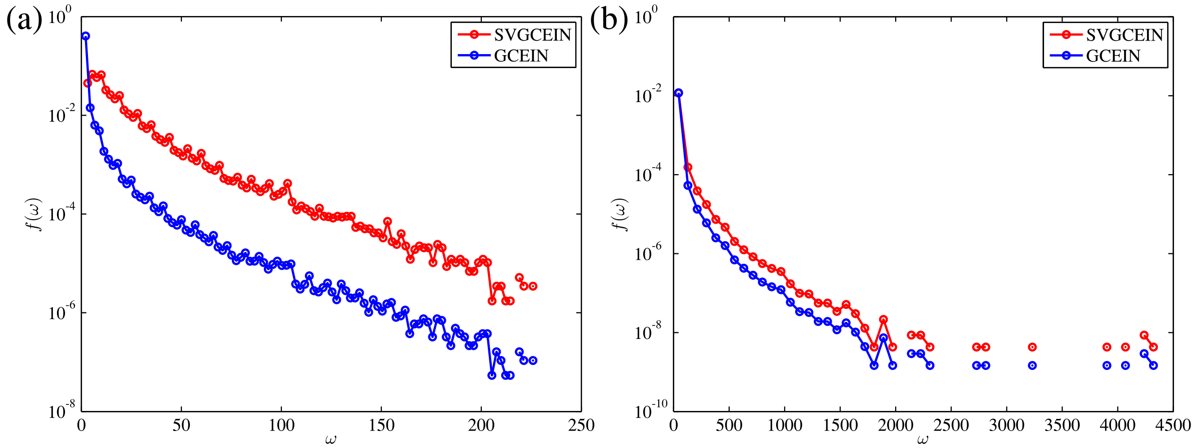

Fig. 1. Edge weight distributions for the GCEIN (blue) and the SVGCEIN (red) based on two different edge weights. (a) Distributions of number-of-days based edge weight w^D. (b) Distributions of number-of-times based edge weight w^T.

The test assigns a p-value to each pair of traders. We compare the p-values with a statistical threshold p. When a large number of statistical tests are performed simultaneously, the validity of the statistical test can be decreased by massive false positives unless a multiple-hypothesis test correction is used. In this case, we use the Bonferroni correction which is the strictest amongst all possible corrections controlling the familywise error rate [22]. The Bonferroni correction for the multiple testing hypothesis is $p_b = 0.01/N_E$, where N_E is the number of performed tests. If the estimated $p(N_{ij})$ is less than the statistical threshold p_b, we keep the edge between i and j in the statistically validated network. When the test does not reject the null hypothesis, the edge between two nodes is removed.

Comparing the different influences to the Bonferroni network, we perform the statistical validation on the edges of GCEIN based on the two definitions of edge weights. In Fig. 1(a) and (b), we show the edge weight distributions of w_{ij}^D and w_{ij}^T respectively. Fig. 2(a) presents number of links as a function of the p-value for the EIN. The red symbols describe the histogram for all links. Symbols of different color refer to the number of links of pairs of traders with weight w_{ij}^D equal to 1 (green), 5 (purple), 10 (blue), 20 (yellow) and 40 (black). The vertical line indicates the Bonferroni threshold. Links located to the left of the threshold are reserved in the Bonferroni network. The network is obtained by considering the entire period. We find that the links with weight $w_{ij}^D = 1$ are characterized by a p-value which is larger than the Bonferroni threshold (indicated as a vertical line). The links that are filtered out from the GCEIN comprise essentially all the links with weight 1. In Fig. 2(b), only part of the links with minimum weight $w_{ij}^T = 3$ are filtered out from the GCEIN. The links with minimum weight cannot reflect the interaction between the traders. We choose the edge weight w_{ij}^D for the statistical tests. The statistically validated network (SVGCEIN) of GCEIN has 25,244 nodes and 254,281 edges. The giant component GCSVGCEIN of SVGCEIN has 20,542 nodes and 227,855 edges.

3. Node and link properties

3.1. Degree distribution

For the undirected networks, The degree is the number of links connected to the node. In terms of the adjacency matrix $\mathbf{A} = (a_{ij})$, the degree of a node i is defined as

$$k_i = \sum_j a_{ij}. \tag{3}$$

Fig. 3 shows the degree distributions of the two giant components GCEIN and GCSVGCEIN. We can find that the two degree distributions have no power-law property. Due to the fat tails, neither normal distribution nor exponential distribution is suitable for fitting them. For the $k < 100$, the two distribution curves are very similar and share quite a few common features. The degree distribution of the GCSVGCEIN declines fast at the tail and the largest node degree of the GCSVGCEIN is much less than the GCEIN. Compared with the GCSVGCEIN, the GCEIN has a larger proportion of the high-degree nodes.

3.2. Mixing patterns

In most cases social networks are considered to be assortative, which means that people with many friends are connected to others who also have many social contacts. This gives rise to degree-degree correlations in the network,

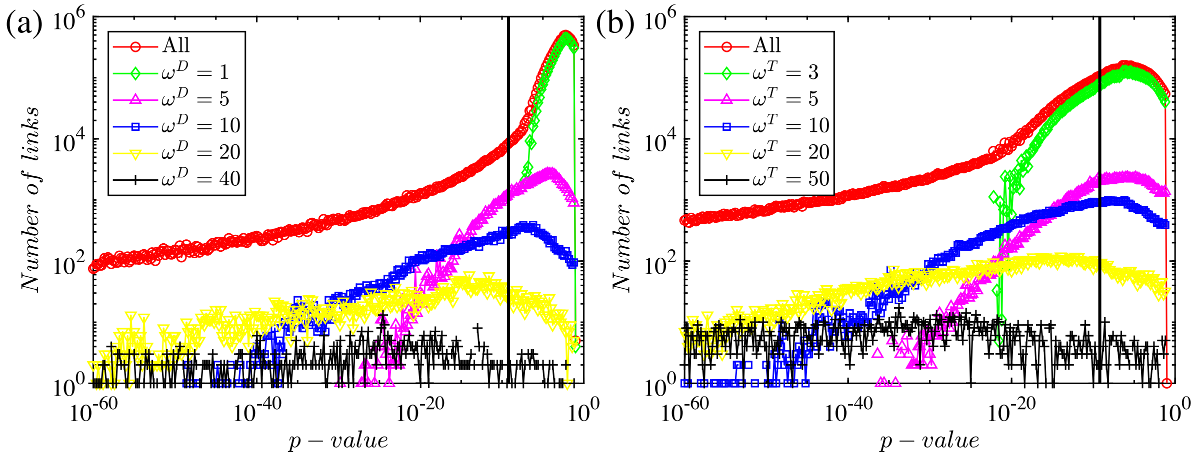

Fig. 2. Number of links as a function of the p-value for the GCEIN based on two different edge weights. The red symbols describe the histogram for all links. The vertical line indicates the Bonferroni threshold. (a) number-of-days based edge weight w^D. Symbols of different color refer to the number of links of pairs of traders with weight w_{ij}^D equal to 1 (green), 5 (purple), 10 (blue), 20 (yellow) and 40 (black). (b) number-of-times based edge weight w^T. Symbols of different color refer to the number of links of pairs of traders with weight w_{ij}^T equal to 3 (green), 5 (purple), 10 (blue), 20 (yellow) and 50 (black).

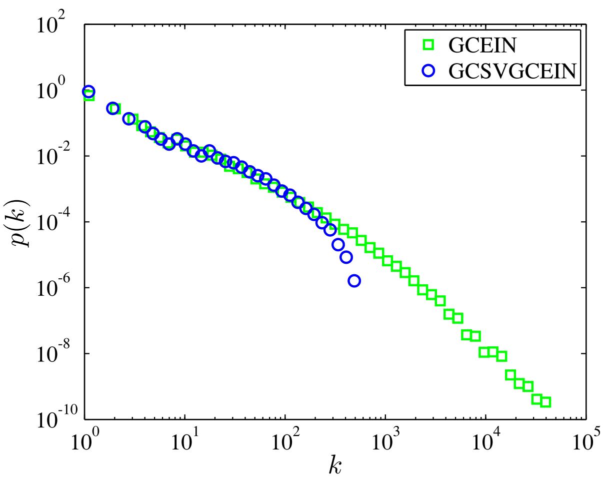

Fig. 3. Degree distributions of the largest connected components of the EIN and the SVGCEIN.

suggesting that the degree of a node positively correlates to the average degree of its neighborhood. The average nearest neighbors degree of a node i is defined as

$$k_{nn,i} = (1/k_i) \sum_{j \in \mathcal{N}_i} k_j, \tag{4}$$

where $\mathcal{N}_i$ denotes the neighborhood of node i. By averaging this value over all nodes in the network for a given degree k, one can calculate the average nearest neighbors degree denoted by $\langle k_{nn}|k \rangle$ [24,25]. A network is said to be assortatively mixed if $\langle k_{nn}|k \rangle$ increases with k and disassortatively mixed if it decreases as a function of k [26].

Fig. 4(a) presents the dependence of $\langle k_{nn}|k \rangle$ as a function of k for the giant components of GCEIN and GCSVGCEIN. We find that GCEIN exhibits a disassortative mixing pattern for the decrease of $\langle k_{nn}|k \rangle$ with the increasing degree k. For the GCSVGCEIN, the $\langle k_{nn}|k \rangle$ function shows a slow rise tendency and the curve becomes horizontal gradually. The statistically validated network GCSVGCEIN becomes assortative indicating that it has characteristics similar with social networks. The disassortative behavior of the GCEIN implies that the majority of the links are randomly formed so that highly active traders trade usually simultaneously with less active retailer traders. In contrast, the GCSVGCEIN contains statistically validated links which very possibly represent the mutual contacts between informed traders. The traders in the statistically validated EIN form a true social network that could be institutions or individuals with frequent information exchanges and active traders are more likely to exchange information with other active traders [17,27].

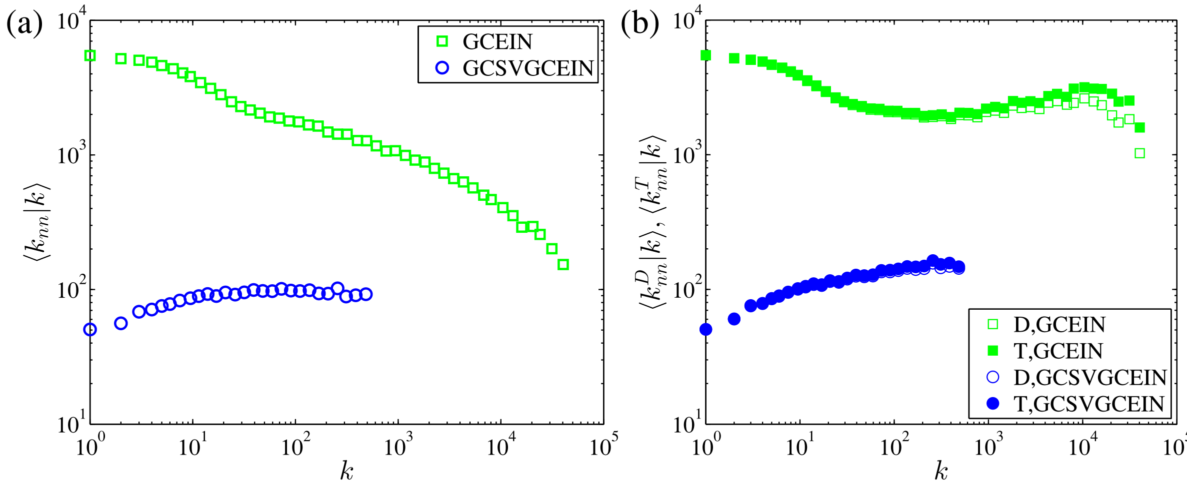

Fig. 4. (a) Average nearest neighbor degree $\langle k_{nn}|k\rangle$ as a function of degree k for the GCEIN and the GCSVGCEIN. (b) Weighted average nearest neighbor degree $\langle k^D_{nn}|k\rangle$ and $\langle k^T_{nn}|k\rangle$ as a function of degree k for the GCEIN and the GCSVGCEIN.

We also calculate two weighted average nearest neighbors degrees defined as [28]

$$k^D_{nn,i} = \sum_{j\in\mathcal{N}_i} \frac{k_j w^D_{ij}}{s^D_i}, \tag{5}$$

and

$$k^T_{nn,i} = \sum_{j\in\mathcal{N}_i} \frac{k_j w^T_{ij}}{s^T_i}, \tag{6}$$

where w^D_{ij} is total number of daily networks where the traders i and j are connected, and w^T_{ij} is the cumulative sum of the edge weights in the daily networks where the two traders are connected. We average these two weighted degrees over all nodes with the same degree k to get $\langle k^D_{nn}|k\rangle$ and $\langle k^T_{nn}|k\rangle$, which measure the effective affinity to connect with neighbors of a given degree while taking the magnitude of the interactions into account. Fig. 4(b) shows the relationship between weighted average nearest neighbors degrees and the degree k. We note that there is no significant difference between the two curves based on two definitions of edge weight. The weighted curve of the GCSVGCEIN shows the similar trend as that in Fig. 4(a). For the GCEIN, weighted average nearest neighbors degrees $\langle k^D_{nn,i}|k\rangle$ and $\langle k^T_{nn,i}|k\rangle$ decrease slowly as k grows. The curve tends to be steady in the middle and then slowly increases, showing a decline trend near the tail.

3.3. Link weight

We have defined two kinds of weight for each edge to quantify the strengths of the link (i, j) between two traders. The edge weight w^D_{ij} (w^T_{ij}) denotes the number of days (times) when we observe the co-occurrence of trading behavior for the traders i and j. Fig. 5(a) presents distributions of edge weight w^D for the GCEIN and the GCSVGCEIN. It can be seen that the distribution for the GCEIN exhibiting a decreasing trend cannot be exactly described by power law. The distribution for the GCSVGCEIN shows a short increasing trend for small values of $w^D < 4$. After reaching the maximum, the distribution curve declines as w^D increases. The proportion of the minimum weighted links is much lower in comparison with the GCEIN. It is revealed that the statistical validation test has filtered out the majority of links with very small weight.

The distributions of number-of-times based edge weight w^T are shown in Fig. 5(b). We notice that the overall shapes of the distributions are qualitatively similar to that of number-of-days based weight w^D for each network. The distribution for the GCEIN presents an general decreasing trend and fluctuates slightly for the smaller values, whereas the distribution for the GCSVGCEIN is approximately increasing with slight fluctuations and then shows a pronounced decrease after reaching the peak.

3.4. Node strength

For each node i in the weighted networks, we define two strengths of the node corresponding to the number-of-days and number-of-times based edge weights respectively:

$$s^D_i = \sum_{j\in\mathcal{N}_i} w^D_{ij}, \tag{7}$$

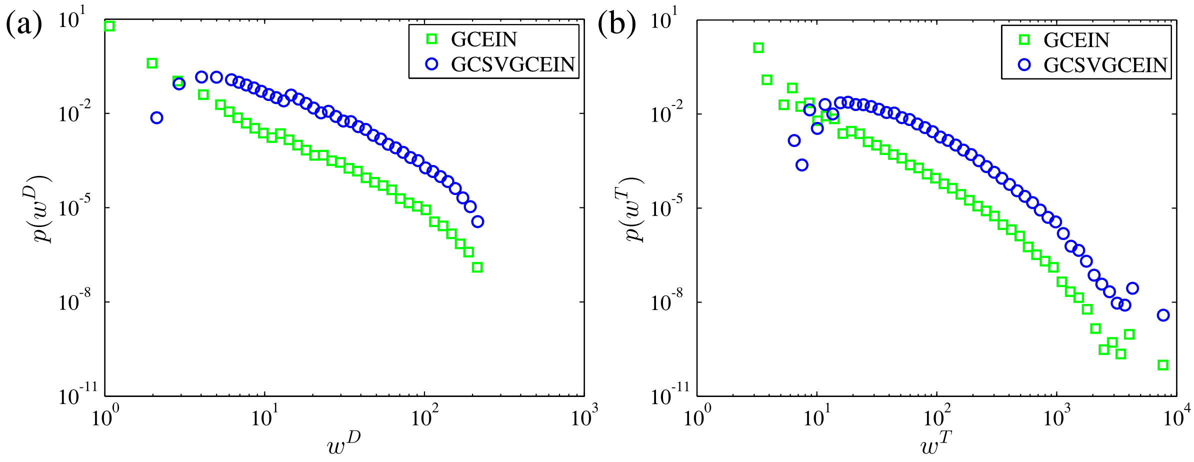

Fig. 5. Edge weight distributions for the GCEIN and the GCSVGCEIN based on two different edge weights. (a) Distributions of number-of-days based edge weight w^D. (b) Distributions of number-of-times based edge weight w^T.

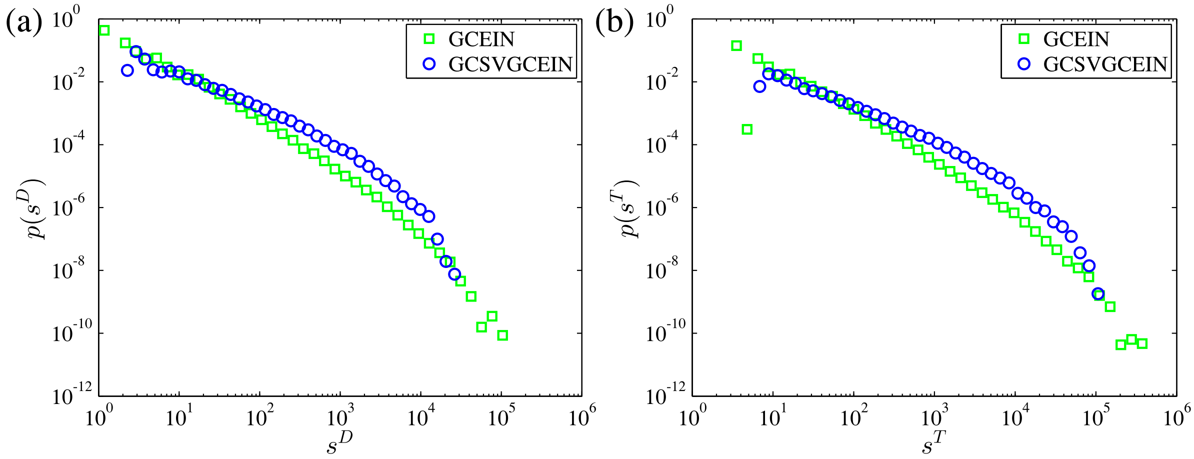

Fig. 6. Node strength distributions for the GCEIN and the GCSVGCEIN based on two different edge weights. (a) Distributions of number-of-days based node strength s^D. (b) Distributions of number-of-times based node strength s^T.

and

$$s_i^T = \sum_{j \in \mathcal{N}_i} w_{ij}^T. \tag{8}$$

We investigate the node strength distributions for the two networks.

The distributions of number-of-days based node strength s^D are shown in Fig. 6(a). Both distributions of node strength exhibit a decline trend with fat tails. We observe a broader tail for the distribution of the GCSVGCEIN with a rapidly decaying trend close to the end when compared to the GCEIN. The nodes with minimum strength in the original network were removed by the statistical validation method. Thus the proportion of the nodes with larger strength for the statistically validated network increased. Given that more than 90% of links have been filtered out from the original network, the largest value of node strengths for the GCSVGCEIN is far less than that of the GCEIN.

Fig. 6(b) shows the distributions of number-of-times based node strength s^T for the two networks. Qualitatively similar to that of number-of-days based node strength s^D in Fig. 6(a), both distributions of number-of-times based node strength decrease with respect to s^T. The distribution of the GCEIN exhibits an obvious breakpoint at $s_{ij}^T \approx 5$. It indicates a sharp decline for the proportion of the nodes with strength $s_{ij}^T \approx 5$.

3.5. Summary statistics

The giant component GCEIN contains 376,242 traders classified into three categories: institutional traders, individual traders and others according to the trader type. There are 7843 institutional traders and 360,922 individual traders

Table 1
Summary of the average values and the standard deviations of the statistics for different categories of traders.

		GCEIN			GCSVGCEIN		
		Institution	Male	Female	Institution	Male	Female
num		7843	216840	144082	996	11876	7217
k_i	mean	273	40	37	17	24	21
	std	1179	202	181	27	42	38
k_{nn}	mean	3854	3948	3930	47	81	78
	std	4753	6344	6341	47	53	53
k_{nn}^D	mean	4184	4047	4025	55	95	91
	std	4879	6377	6376	56	60	61
k_{nn}^T	mean	4308	4074	4052	54	97	93
	std	4957	6397	6398	56	61	62
s^D	mean	534	78	66	396	378	324
	std	2360	656	526	874	1178	1011
s^T	mean	2040	269	228	1806	1463	1262
	std	8760	2452	1962	3883	4882	4223
s_{nn}^D	mean	7645	8111	8055	1433	1701	1625
	std	9891	13259	13243	1883	1663	1653
s_{nn}^T	mean	27970	29733	29529	6246	6943	6597
	std	36357	48744	48698	8110	7017	6914
EC	mean $\times 10^3$	0.252	0.135	0.110	0.086	1.124	0.990
	std $\times 10^3$	1.246	1.781	1.428	1.118	7.343	6.826

including 216,840 male traders and 144,082 female traders. The GCSVGCEIN contains 996 institutional traders and 19,093 individual traders including 11,876 male traders and 7217 female traders. We summarize the average value and standard deviation of the statistics for each type of traders in Table 1. The average node degree and node strength of institutional traders in the GCEIN are significantly larger than that of individual traders, implying that the nodes corresponding to institutional traders tend to be connected with more nodes, but the similar results cannot be observed in the GCSVGCEIN.

4. Mutual relationship

4.1. Node strength and its average nearest node strength

In addition to the degree-degree correlations for nodes, we also consider to study the correlations between node strengths. The average nearest neighbor strengths of a node i are defined as follows,

$$s_{nn,i}^D = (1/k_i) \sum_{j \in \mathcal{N}_i} s_j^D, \tag{9}$$

and

$$s_{nn,i}^T = (1/k_i) \sum_{j \in \mathcal{N}_i} s_j^T, \tag{10}$$

By averaging the values of s^D and s^T over all nodes with a given strength s in the network, we can calculate the average strength of nearest neighbors $\langle s_{nn}^D | s^D \rangle$ and $\langle s_{nn}^T | s^T \rangle$.

Fig. 7(a) shows the dependence of $\langle s_{nn}^D | s^D \rangle$ as a function of s^D for the GCEIN and the GCSVGCEIN. We observe a decreasing trend for the $\langle s_{nn}^D | s^D \rangle$ curve of the GCEIN. It suggests that the nodes with large values of s^D prefer to be connected to those of small strength. The curve of the GCSVGCEIN increases sharply for small $s^D < 10$ and then shows a very slow rise for large values of s^D.

We also plot $\langle s_{nn}^T | s^T \rangle$ as a function of s^T for the two trader networks in Fig. 7(b). The $\langle s_{nn}^T | s^T \rangle$ curve shares a similar shape with the $\langle s_{nn}^D | s^D \rangle$ curve in Fig. 7(a) for each network. We can see different behaviors between the $\langle s_{nn}^T | s^T \rangle$ curves of the original network and its statistically validated network. Whereas the $\langle s_{nn}^T | s^T \rangle$ curve of the GCEIN declines with s^T, the $\langle s_{nn}^T | s^T \rangle$ function for the GCSVGCEIN goes up rapidly for $s^T < 20$ and exhibits a slight increasing trend for large values. For the statistically validated network, The average numbers of times of the trading behavior co-occurrence for the high-strength traders are larger than that of low-strength traders.

The quantitative differences between the corresponding curves in Figs. 4 and 7 are naturally due to the differences in the variables. We stress that qualitative similarity between the corresponding curves and argue that the patterns in Fig. 7 have the same origins as these in Fig. 4.

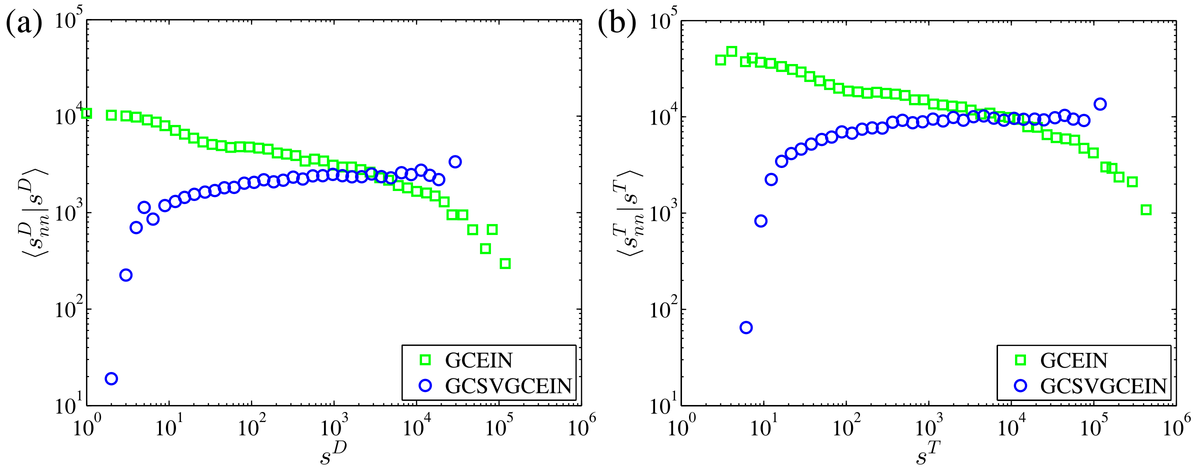

Fig. 7. (a) Average number-of-days based node strength $\langle s_{nn}^D|s^D\rangle$ as a function of s^D for the GCEIN and the GCSVGCEIN. (b) Average number-of-times based node strength $\langle s_{nn}^T|s^T\rangle$ as a function of s^T for the GCEIN and the GCSVGCEIN.

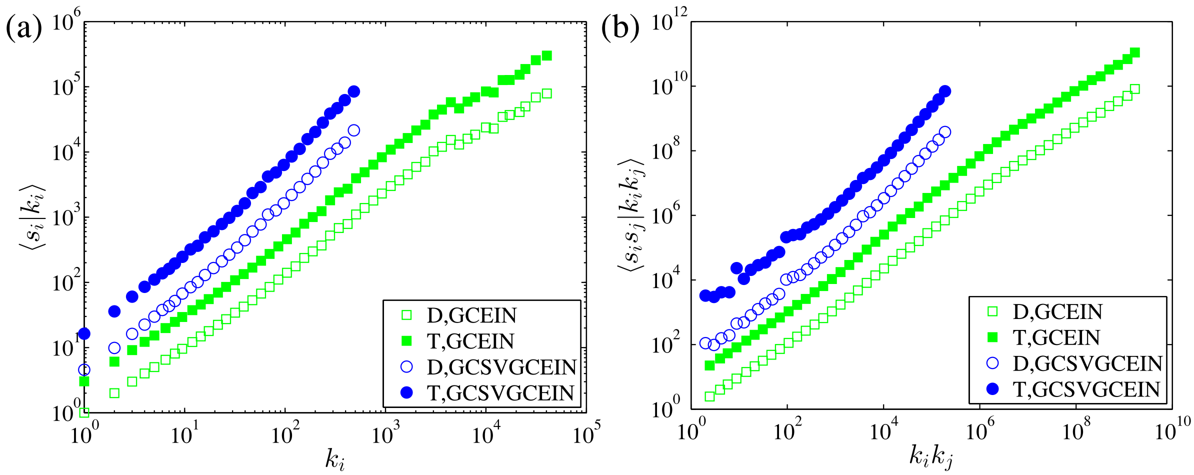

Fig. 8. (a) Dependence of the average node strength $\langle s^D|k\rangle$ and $\langle s^T|k\rangle$ on the node degree for the GCEIN and the GCSVGCEIN. (b) Dependence of $\langle s_i^D s_j^D|k_ik_j\rangle$ and $\langle s_i^T s_j^T|k_ik_j\rangle$ on the degree product k_ik_j for the GCEIN and the GCSVGCEIN.

4.2. Node degree versus node strength

We investigate the correlation between node strength and node degree. Fig. 8(a) illustrates the dependence of the node strengths on the node degrees for the GCEIN and the GCSVGCEIN. Average node strength conditional on degree measured in terms of the number of days $\langle s^D|k\rangle$ and times $\langle s^T|k\rangle$ exhibit almost the same behaviors for each network. The curves exhibit a clear power-law dependence $\langle s|k\rangle \sim k^{\alpha}$. The high-degree traders have larger numbers of days and times for the trading behavior co-occurrence on average. For any given degree k, the average node strength of the network GCSVGCEIN is larger than that of the GCEIN. This observation is due to the fact that the statistical validation approach has removed the nodes with very small strength.

We plot average strength product $\langle s_is_j|k_ik_j\rangle$ as a function of k_ik_j in Fig. 8(b). The curves based on the two different edge weights are very similar for each network. In this case we also observe a power-law dependence $\langle s_is_j|k_ik_j\rangle \sim (k_ik_j)^{\beta}$. The fitting power-law exponents are estimated as $\beta^D \approx 1.11$ and $\beta^T \approx 1.13$ for the GCEIN, $\beta^D \approx 1.34$ and $\beta^T \approx 1.29$ for the GCSVGCEIN. If there is no correlation between node degree and the weights of the edges adjacent to the node, we expect that $\langle s_i\rangle = k_i\langle w\rangle$ and obtain $\langle s_is_j|k_ik_j\rangle = \langle w\rangle^2\langle k_ik_j\rangle$, where $\langle w\rangle$ is the average edge weight in the network. However, the discrepancy of $\beta \neq 1$ indicates the existence of correlations.

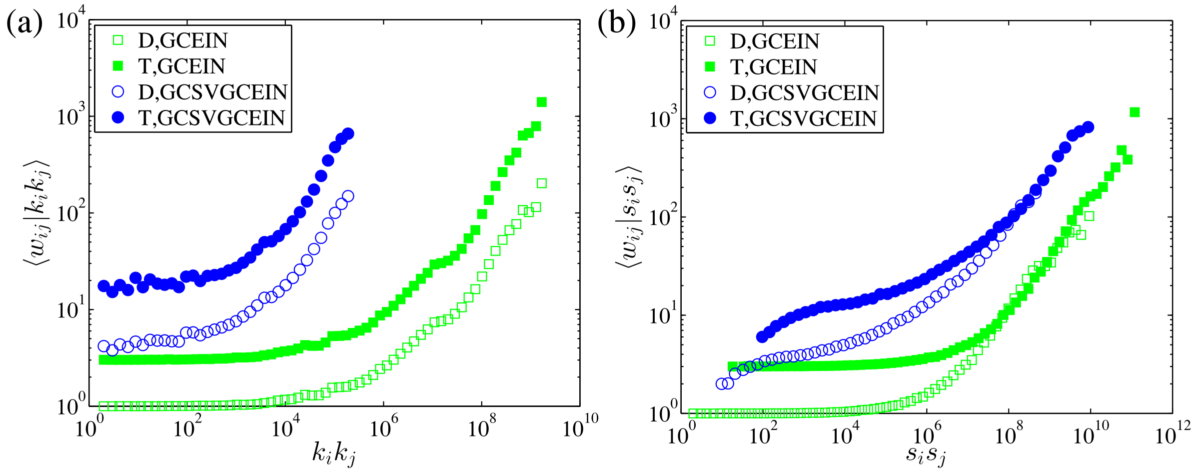

Fig. 9. (a) Average edge weight $\langle w_{ij}^D|k_ik_j\rangle$ and $\langle w_{ij}^T|k_ik_j\rangle$ as a function of degree product k_ik_j for the GCEIN and the GCSVGCEIN. (b) Average edge weight $\langle w_{ij}^D|s_i^Ds_j^D\rangle$ and $\langle w_{ij}^T|s_i^Ts_j^T\rangle$ as a function of strength product s_is_j for the GCEIN and the GCSVGCEIN.

4.3. Link weight versus node degree and strength

We present the correlation between edge weight and degree product k_ik_j for adjacent nodes in Fig. 9(a). The $\langle w_{ij}^D|k_ik_j\rangle$ curve and the $\langle w_{ij}^T|k_ik_j\rangle$ curve in the same network are similar. The dependence of $\langle w_{ij}|k_ik_j\rangle$ on k_ik_j reveals a positive correlation between them for the GCSVGCEIN. We note that the curves for the GCEIN can be divided into two parts. For the small $k_ik_j < kk_x$, the edge weights are independent of the degree product k_ik_j. The values of $\langle w_{ij}|k_ik_j\rangle$ increase as a function of k_ik_j for the large values, indicating the larger average numbers of days and times for the trading behavior co-occurrence between two traders with higher degree product.

We also study the dependence of $\langle w_{ij}|s_is_j\rangle$ on s_is_j as shown in Fig. 9(b). The plots for $\langle w_{ij}|s_is_j\rangle$ measured in terms of w_{ij}^D and w_{ij}^T exhibit similar behaviors. We observe a general trend of increase for the $\langle w_{ij}|s_is_j\rangle$ curves of the GCSVGCEIN. The adjacent trader pairs with higher node strength product have larger numbers of days and times for the trading behavior co-occurrence on average. For the network GCEIN, the behavior of the $\langle w_{ij}|s_is_j\rangle$ curve is different. Similar to Fig. 9(a), independence can be observed for small values of s_is_j, whereas the increase of $\langle w_{ij}|s_is_j\rangle$ is observed for large values.

5. Conclusion

In this paper we have studied the empirical information network based on the co-occurrence of the trading behaviors of the traders over the sample period. We have constructed statistically validated networks and investigated the correlation between topological structures and statistical properties of their largest connected components. We found that the size of statistically validated network is significantly smaller than the original network. We have analyzed the distributions of the degree, the edge weight, the node strength and their mutual correlations for the giant components GCEIN and GCSVGCEIN. Both distributions of degree are fat tailed without power-law property. The GCEIN exhibits a disassortative mixing pattern, while we observe the opposite situation for the statistical validated network that the traders with many information partners are more likely to interact with others who also have large numbers of counterparts. This finding shows the importance of statistical validation method that makes it easier to uncover the true interaction behaviors of traders. The statistically validated network has a larger proportion of the high-degree nodes due to the fact that the nodes with minimum strength were removed from the original network. We also considered the dependence of average node strength on node degree that revealed a positive correlation between them. The similar results can be observed for the correlation between edge weight and node degree product as well as the correlation between edge weight and node strength product.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (71571121, 71532009, U1811462) and the Fundamental Research Funds for the Central Universities, China (222201818006).

References

[1] M. Tumminello, F. Lillo, J. Piilo, R.N. Mantegna, Identification of clusters of investors from their real trading activity in a financial market, New J. Phys. 14 (2012) 013041, http://dx.doi.org/10.1088/1367-2630/14/1/013041.
[2] F. Lillo, S. Micciche, M. Tumminello, J. Piilo, R.N. Mantegna, How news affects the trading behaviour of different categories of investors in a financial market, Quant. Finance 15 (2) (2015) 213–229, http://dx.doi.org/10.1080/14697688.2014.931593.
[3] H.N. Ozsoylev, J. Walden, Asset pricing in large information networks, J. Econ. Theory 146 (6) (2011) 2252–2280, http://dx.doi.org/10.1016/j.jet.2011.10.003.
[4] H.N. Ozsoylev, J. Walden, M.D. Yavuz, R. Bildik, Investor networks in the stock market, Rev. Financ. Stud. 27 (5) (2014) 1323–1366, http://dx.doi.org/10.1093/rfs/hht065.
[5] P. Colla, A. Mele, Information linkages and correlated trading, Rev. Financ. Stud. 23 (1) (2010) 203–246, http://dx.doi.org/10.1093/rfs/hhp021.
[6] B. Han, L.-Y. Yang, Social networks, information acquisition, and asset prices, Manag. Sci. 59 (6) (2013) 1444–1457, http://dx.doi.org/10.1287/mnsc.1120.1678.
[7] F. Emmert-Streib, A. Musa, K. Baltakys, J. Kanniainen, S. Tripathi, O. Yli-Harja, H. Jodlbauer, M. Dehmer, Computational analysis of structural properties of economic and financial networks, J. Netw. Theory Financ. 4 (3) (2019) 1–32, http://dx.doi.org/10.21314/JNTF.2018.043.
[8] L. Feng, M.S. Seasholes, Correlated trading and location, J. Financ. 59 (5) (2004) 2117–2144, http://dx.doi.org/10.1111/j.1540-6261.2004.00694.x.
[9] S.-L. Chung, W.-C. Liu, W.-R. Liu, K. Tseng, Investor network: Implications for information diffusion and asset prices, Pacific-Basin Financ. J. 48 (2018) 186–209, http://dx.doi.org/10.1016/j.pacfin.2018.02.004.
[10] Z.-Q. Jiang, W.-X. Zhou, Complex stock trading network among investors, Physica A 389 (2010) 4929–4941, http://dx.doi.org/10.1016/j.physa.2010.07.024.
[11] L. Adamic, C. Brunetti, J. Harris, A.A. Kirilenko, Trading networks, Econom. J. 20 (3) (2017) S126–S149, http://dx.doi.org/10.1111/ectj.12090.
[12] J.-J. Wang, S.-G. Zhou, J.-H. Guan, Characteristics of real futures trading networks, Physica A 390 (2011) 398–409, http://dx.doi.org/10.1016/j.physa.2010.09.027.
[13] M.-X. Li, Z.-Q. Jiang, W.-J. Xie, X. Xiong, W. Zhang, W.-X. Zhou, Unveiling correlations between financial variables and topological metrics of trading networks: Evidence from a stock and its warrant, Physica A 419 (2015) 575–584, http://dx.doi.org/10.1016/j.physa.2014.10.039.
[14] E. Cohen-Cole, A. Kirilenko, E. Patacchini, Trading networks and liquidity provision, J. Financ. Econ. 114 (2) (2014) 235–251, http://dx.doi.org/10.1016/j.jfineco.2014.04.007.
[15] W.-J. Xie, M.-X. Li, H.-C. Xu, W. Chen, W.-X. Zhou, H.E. Stanley, Quantifying immediate price impact of trades based on the *k*-shell decomposition of stock trading networks, EPL 116 (2) (2016) 28006, http://dx.doi.org/10.1209/0295-5075/116/28006.
[16] X.-Q. Sun, X.-Q. Cheng, H.-W. Shen, Z.-Y. Wang, Distinguishing manipulated stocks via trading network analysis, Physica A 390 (2011) 3427–3434, http://dx.doi.org/10.1016/j.physa.2011.04.006.
[17] Z.-Q. Jiang, W.-J. Xie, X. Xiong, W. Zhang, Y.-J. Zhang, W.-X. Zhou, Trading networks abnormal motifs and stock manipulation, Quant. Financ. Lett. 1 (1) (2013) 1–8, http://dx.doi.org/10.1080/21649502.2013.802877.
[18] X.-Q. Sun, H.-W. Shen, X.-Q. Cheng, Trading network predicts stock price, Sci. Rep. 4 (2014) 3711, http://dx.doi.org/10.1038/srep03711.
[19] F. Radicchi, J.J. Ramasco, S. Fortunato, Information filtering in complex weighted networksn, Phys. Rev. E 83 (4) (2011) 046101, http://dx.doi.org/10.1103/PhysRevE.83.046101.
[20] M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, R.N. Mantegna, Statistically validated networks in bipartite complex systems, PLoS One 6 (2011) e17994, http://dx.doi.org/10.1371/journal.pone.0017994.
[21] M.-X. Li, Z.-Q. Jiang, W.-J. Xie, S. Miccichè, M. Tumminello, W.-X. Zhou, R.N. Mantegna, A comparative analysis of the statistical properties of large mobile phone calling networks, Sci. Rep. 4 (2014) 5132, http://dx.doi.org/10.1038/srep05132.
[22] M.-X. Li, V. Palchykov, Z.-Q. Jiang, K. Kaski, J. Kertész, S. Miccichè, M. Tumminello, W.-X. Zhou, R.N. Mantegna, Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data, New J. Phys. 16 (8) (2014) 083038, http://dx.doi.org/10.1088/1367-2630/16/8/083038.
[23] M. Tumminello, S. Miccichè, F. Lillo, J. Varho, J. Piilo, R.N. Mantegna, Community characterization of heterogeneous complex systems, J. Stat. Mech. Theory Exp. (2011) P01019, http://dx.doi.org/10.1088/1742-5468/2011/01/P01019.
[24] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the internet, Phys. Rev. Lett. 87 (25) (2001) 258701, http://dx.doi.org/10.1103/PhysRevLett.87.258701.
[25] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M.A. de Menezes, K. Kaski, A.-L. Barabási, J. Kertész, Analysis of a large-scale weighted network of one-to-one human communication, New J. Phys. 9 (2007) 179, http://dx.doi.org/10.1088/1367-2630/9/6/179.
[26] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701, http://dx.doi.org/10.1103/PhysRevLett.89.208701.
[27] W.-X. Zhou, G.-H. Mu, J. Kertész, Random matrix approach to the dynamics of stock inventory variations, New J. Phys. 14 (9) (2012) 093025, http://dx.doi.org/10.1088/1367-2630/14/9/093025.
[28] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (11) (2004) 3747–3752, http://dx.doi.org/10.1073/pnas.0400087101.