Journal of Industrial and Systems Engineering
Vol. 13, No. 1, pp. 136-146
Winter (January) 2020

The structure of stock markets as signed networks

Maryam Ehsani[1*]

[1]*Department of Electrical Engineering, Arak University of Technology, Arak, Iran*

m.ehsani@arakut.ac.ir

Abstract

Dynamism and evolution in financial markets and specifically stock markets represents a complex network with many relations between different finance agents and corporations. So there are many researches analyzing different aspects of stock markets in the field of complex networks. However, studying financial markets as signed networks maybe considered as a new perspective in this area. This paper proposes a new methodology for analyzing structure of stock markets as signed networks in the perspective of balance theory. For this purpose, some stock markets based on some data of Tehran stock market and Nasdaq are modeled as signed networks and some aspects of their structural properties were studied from the point of view of balance theory. The results show the whole pattern of the structure of stock networks approximately fit to a completely balanced structure. It is observed that the distance from structural balance rises abruptly in some unstable duration and so may be proposed as an index for forecasting overall function of stock markets or crisis conditions. The results also imply the existence and role of positive connection between two balanced partitions. The proposed methodology can lead directly to many applications in analyzing, evaluating and forecasting stock markets such as balanced clustering and determining the important companies and relations affecting the overall system function. The applications could be useful for system control and decisions either in micro level such as portfolio investment or macro level and regulating the market.

Keywords: Complex networks, signed networks, balance theory, financial markets

1-Introduction

Function and dynamism of a complex system are mainly affected by the interactions and its structure. Also, evolution of finance markets is controlled by many complicated parameters and interactions between many different agents. Specifically, fluctuations in stock prices are caused by interactions between companies that could be represented as a complex network in which each company indicated by its stock price.

Therefore, network analysis approach has been recently so powerful in stock market modeling, analysis and prediction. For this purpose, different methods and concepts were extended and applied in financial market analysis based on network approach such as Minimum Spanning Tree, Shortest Path and concepts like centrality measures including degree distribution and betweenness measure,

*Corresponding author

also methods for community detection, motif analysis and so on. Among so many studies in this field a few can be mentioned here, for instance: Bekiros et al. (2017), Garas et al. (2008), Onella et al. (2004), Korbel et al. (2017), Tse et al. (2010), Battistone et al. (2010), Arcangelis and Rotundo (2016), Allen and Babus (2009), Esfanipour and Zamanzadeh (2013).

As a pioneer study in this area, Mantegna (1999) proposed the idea of asset graph in which the vertices are companies' stocks and the edges indicate correlation between the stocks. In spite of different models and methods in this context, construction the networks corresponding to stock markets essentially follows this basic idea with some modifications.

Nevertheless, analysis of finance markets as signed networks can be considered as a new research area. In fact, the asset graph which is based on correlation coefficients between stocks can be considered as signed graphs -that is, each edge between two vertices has positive or negative sign. The principles and rules about structure and dynamism of signed networks and related systems are chiefly based on balance theory concepts and corresponding theorems in graph theory.

The present study aims to propose a new framework for analyzing structural aspects of stock markets as signed networks in the perspective of balance theory. How far are the Stock markets structure from the complete balance and how is related this distance to the system situation? What is the role of positive and negative links in system function and which are important clusters and nodes dominating the network? These are some of questions could be investigated in this framework and lead to introduce some new concepts and indexes applicable in assessment, predict and control of stock markets.

This study has been performed chiefly on data derived from Tehran stock market. In order to compare, some data of Nasdaq are also studied. Although complete balance is almost never achieved in real networks, the results show the corresponding networks' structure fit well the balance condition. In addition, the results show the distance from a complete balanced structure increases during crisis periods and so it may be offered as an index for assessing overall system situation. Also it is observed that the links violate balance condition chiefly have positive sign implying the important role of this links in relation and connectivity thorough the system.

This methodology leads to useful applications in stock market analysis such as detecting structural patterns and main clusters, also determining important components and nodes (companies or industry groups) and relations and their effects on the system function, also can lead to networks' dynamics prediction. These applications could support decision making in either the micro level for example portfolio investment or the macro level such as regulating the market.

In the following, a brief review on main concepts is given in section 2. Then methods, algorithms and results are given in section 3. Results summary and discussion is presented in the section 4 and at last section 5 deals with conclusion and directions for future research.

2-Main concepts review

2-1- Structural balance in signed graphs

Graph theory provides a mathematical language for illustrating and modeling systems. In different cases the relations between components of a system are indicated by edges or arcs in the graph considering weights or not, or other types of network modeling. However, the interactions in many systems implies positive or negative effect, for example repulsive or attraction forces in electric or magnetic systems, activating or inhibiting effects in neural networks or gene networks, friendship or enmity in social networks and so on. So the interactions of such system can be represented by positive or negative links in a signed graph.

The basic idea underlying the balance theory in the context of signed graphs, rooted in Heider studies in phycology (1946), introduces balanced and imbalanced relationship: a relation of three components or triad is balanced if the product of three edges' signs is positive, else it would be a potential for imbalance and challenge in the system (figure 1).

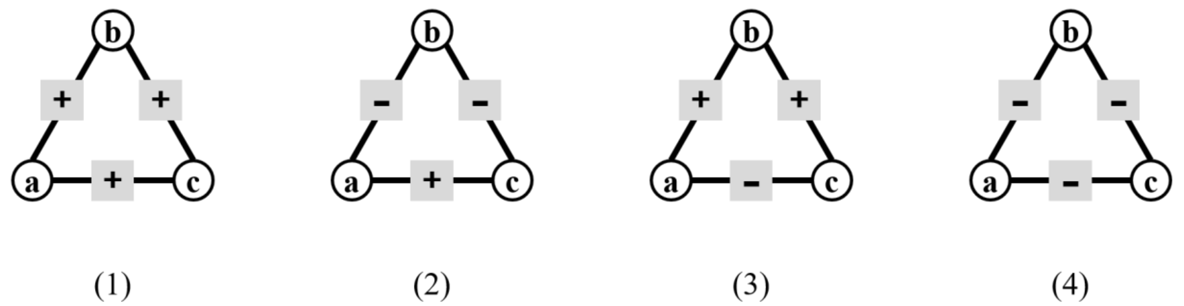

Fig 1. differnet combination of signed relations in a triad, (1) and (2) are balanced while (3) , (4) are not

This idea has been very fruitful to analyze system structure and dynamics in many different contexts such as biological, social and physical systems. The main concept implies the system changes toward more equilibrium, decreasing the imbalanced relations in align with decreasing the energy level. Some researches based on this concept in different fields are mentioned here:
Sontag (2007), Iacono and Altafini (2010), Lee et al. (2013), Tang et al. (2013), Singh et al. (2014), Su et al. (2017), He et al. (2019).

Mathematical theories about balance theory and structural properties of signed graph are chiefly based on studies by Catriwright and Harary (1956). A signed graph is defined structurally balanced if all triple cliques (triads) in it are balanced. They proved on this condition, either all the signs in the graph are positive or the set of vertices can be partitioned into two subsets, called plus sets, such that all of the edges between vertices of the same subset have positive sign and all of the edges between vertices of different subset have negative sign (figure 2. (a)). Also it is proved that this condition is equal to lack of any negative cycle in the graph.

Extending the concept of complete structural balance in order to adapt more real conditions, Davis (1967) proposed definition of weak structural balance in which there is no triads with exactly one negative edge. He proved this new condition is equal to the situation that there is no cycle with exactly one negative edge in the cycle and the set of vertices can be partitioned to k plus set (figure 2 (b)). Such a signed graph called k- balanced and this partitioning called balanced partitioning.

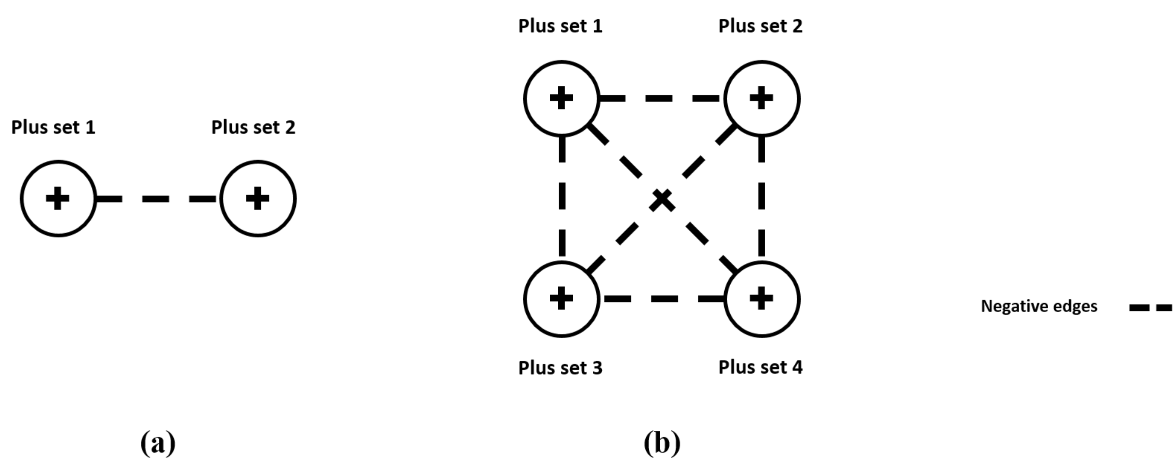

Fig 2. **(a)** On the condition of complete balance, a signed graph is partitioned into two plus set such that all of the edges inside each plus set are positive and all of the edges between them are negative, **(b)** Weak structural balance (Davis generalization), on this condition the signed graph could be partitioned into few plus sets.

The problem of finding the best k- partitioning for a signed graph such that the number of edges violating k- balanced partitions is minimized, is a NP-Hard problem and a reputable algorithm for it was presented by Doreian and Mrvar (2009).

2-2- Constructing stock networks

In stock markets, each company can be indicated by its stock price. Therefore, the interactions between stocks depict a complex network. In this paper, the stock networks are constructed based on the idea and definitions proposed at the first by Mantegna (1999). Considering time series derived by stock prices in the specific time window for each company, stock market can be modeled by a graph in which the vertices correspond to companies' stocks and the edges indicate the Pearson's correlation coefficient between two corresponding time series. Since the rate of changes in the prices are important and not just the absolute value of the variables, the logarithmic return of stock of company X at time t is defined by the formula (1), in which $p_X(t)$ denotes the price of stock X at time t.

$$r_x^t = \ln(p_X(t)/p_X(t-1)) \qquad (1)$$

and linear correlation coefficient between the two time series is computed as follows:

$$r_{xy} = \frac{\sum_{t=1}^{T} (r_x^t - \bar{r}_x)(r_y^t - \bar{r}_y)}{\sqrt{\sum_{t=1}^{T} (r_x^t - \bar{r}_x)^2} \sqrt{\sum_{t=1}^{T} (r_y^t - \bar{r}_y)^2}} \qquad (2)$$

In which r_X, r_Y denote time series corresponding to logarithmic return of stock company X and Y obtained by formula (1), respectively. T denotes to the length of time series equal to the number of time intervals in the intended time window. This simple model based on computing just pair to pair linear correlation and neglects nonlinearity and collinearity issues.

As we know, the value of a correlation coefficient ranges between -1, 1. So this value as the weight for the corresponding edge between company X and Y indicates a signed connection. Concluding all of the edges in the range [-1, 1] a complete graph is derived. Since the correlation coefficients with low absolute value don't imply linear relationship, a threshold equal to 0.7 is considered for excluding the edges with lower absolute value. The higher value of this threshold causes the higher number of graph components. However, for the threshold equal to 0.7, the largest connected component approximately covers all the vertices.

2-3- Data and software

The required data extracted from Tehran stock market in the time period 1394- 1398 (SH) and Nasdaq stock market in the time period 2014-2018. Data extraction and preprocessing was performed by Python software. Network analyses including component extraction, computing general network information, community detection and balance partitioning and so on are performed either by MATLAB or by Pajek 5.07 software.

3-Methods and main results

This paper modeling and analyzing some stock markets as signed graph from the point of view of balance theory, aims to answer some questions such as: how the corresponding networks are far from complete balance? Could be this distance considered as an index for the overall situation of the stock market? What is the sign distribution of the links violating balance condition?

For this purpose, some networks constructed based on data extracted from Tehran stock market during the time period 1394-1398. The companies were considered that their stocks were traded in at least 70% days during this period and the others excluded. For each company, there were single days or durations in the time period in which its stock wasn't traded. Such null data fulfilled and compensated by filtering and estimation techniques in data preprocessing phase. In order to compare the results with another stock market, data of Nasdaq stock market in the corresponding time period 2014-2018 was extracted and analyzed. Since the number of companies in Nasdaq stock market are about ten times more than Tehran stock market, 10% companies in Nasdaq were randomly selected and therefore about 250 companies in construction each network are considered. Data extraction and preprocessing steps were performed by Python software.

After data preprocessing, the networks were constructed for time periods 1394-1398, 1394-1396, 1396-1398 for Tehran stock market according to the formulas in the section 2-2. The time series for each company correspond to daily stock return based on closure prices in the mentioned time periods. Weight of each edge derives by computing linear correlation coefficient between the two time series, excluding the edges with lower absolute weight with threshold 0.7. The largest connected components in each network were extracted and named net94-98, net 94-96, net 96-98, respectively. For Nasdaq stock markets the networks constructed for the periods 2014-2018, 2014-2016, 2016-2018, then the largest connected component extracted and named net14-18, net14-16 and net 16-18, respectively.

The general information about structural properties of these networks are given in table1.

Table 1. General structural properties of the largest component in each network

Network name	Number of vertices	Number of isolated vertices	Number of vertices in the largest component	Average degree	Average density	Number of edges	Number of positive edges	Number of negative edges	Negative edges no./total edges no.
net94-96	266	30	236	32.9	0.13	3882	2830	1052	0.27
net96-98	283	29	254	65	0.25	8324	7956	368	0.04
net94-98	273	30	243	26.5	0.11	3220	2732	488	0.15
net14-16	250	23	227	51	0.22	5793	4038	1775	0.3
net16-18	250	28	222	43.9	0.17	5487	3852	1635	0.29
net14-18	250	18	232	71	0.3	8239	6154	2085	0.25

According to table1, there is a clear distinct in the ratio of negative edges no. for net96-98. As the next results show, this is the same network with the most distance from structural balance and the time period coincides with US attempts to leave Barjam (Joint Comprehensive Plan of Action) and its consequences on Iran' economy.

In order to analyzing the structural balance of these signed networks, Doriean- Mrvar k- balanced partitioning algorithm was run on each network for k=2, 3, 4 by Pajek 5.07 software. This algorithm starts from a random partitioning, swapping vertices between partitions, implements a heuristic search for minimizing the total number of error edges violating k- balance condition. This error number includes positive edges between k partitions and negative edges inside each partition. One of the parameters required for running this algorithm is the weight of importance positive and negative error links in the total error number and in this study this parameter was considered equal to 0.5 for both types of error links.

The algorithm was run once for the networks considering edges' weights and again for the networks with the edges indicating just the sign equal to 1 or -1. The results are the same for both cases and are given in table 2- table7.

Table 2. Results of k=2 balance partitioning for net94-96
Total number of edges violating balance condition (error number) = 37
Number of positive edges between 2 partitions = 35

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	95	1059	2	0.23
2	141	1738	0	017

Table 3. Results of k=2 balance partitioning for net96-98
Total number of edges violating balance condition (error number) = 131
Number of positive edges between 2 partitions = 46

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	8	6	1	0.21
2	246	7905	84	0.26

Table 4. Results of 2 balance partitioning for net94-98
Total number of edges violating balance condition (error number) = 54
Number of positive edges between 2 partitions = 53

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	99	1414	0	0.29
2	144	1265	1	0.12

Table 5. Results of 2 balance partitioning for net14-16
Total number of edges violating balance condition (error number) = 52
Number of positive edges between 2 partitions = 50

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	129	2612	1	0.31
2	98	1377	1	0.28

Table 6. Results of 2 balance partitioning for net16-18
Total number of edges violating balance condition (error number) = 25
Number of positive edges between 2 partitions = 18

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	154	3022	2	0.29
2	68	411	5	0.17

Table 7. Results of 2 balance partitioning for net14-18
Total number of edges violating balance condition (error number) = 13
Number of positive edges between 2 partitions = 12

Partition Number	Number of vertices	Number of positive edges	Number of negative edges	Average density
1	71	620	0	0.24
2	161	5522	1	0.42

As mentioned earlier, the k- balanced partitioning algorithm was run for k=2, 3, 4 and the minimum error was obtained versus k=2 for all of networks. This fact and also a very few number of total errors comparing to random sign distribution, show that the overall pattern of these networks complies with structural balance. In addition, the results show the error edges which violate the balance conditions are almost all positive edges between the two partitions and edges inside partitions are almost entirely positive.

It is remarkable that structural pattern of net96-98 differs from the others. The results show this network has higher distance from balance condition, noticeably more negative edges inside partition 2 and also fewer ratios of negative edges according to Table 1. Almost all the edges violating balance condition in the other networks are positive edges between two balanced partitions, while in net96-98, negative edges have important role in total error. As mentioned in the result explanations of Table1. , these differences as structural change toward imbalance implies the transition to a new state in the market and reshaping the network and its connections due to instability in Iran economics and industries condition, caused by US withdrawal from Barjam and sanction tightening. Therefore, the distance from balance condition and the network situation from this view point could introduce some indicators for system stability and prediction of crisis duration or abrupt changes in market that eventuate to transition and structural dynamism in the network.

So, this paper simply introduces the relative error number as an index for distance from balance condition defined as the ratio of total error number to the number of networks' edges, denoted by d(G) in formula (3).

$$d(G) = \text{total error no.} / \text{edges no.} \tag{3}$$

Due to importance of negative edges in imbalance condition, another index is defined as the ratio of negative edges violating the balance condition to total negative edges in the network and is denoted by

$$d(G) = \text{no. of negative edges inside balanced partitions} / \text{edges no. in the network} \tag{4}$$

Also, regarding the important role of positive edges between two balance partitions, the ratio of the positive edges to the total error introduced as another index.

$$d+ (G) = \text{no. of positive edges between partitions} / \text{total error no.} \tag{5}$$

Considering these definitions, table 8 shows the summary of results of 2- balanced partitioning for the networks.

Table 8. Summary results of balanced partitioning for the networks

Network name	Vertices no. (the largest component graph)	Edges no.	No. of negative edges	Total error	No. of the positive links Between two partitions	Balance distance index d(G)	d+(G)	d_(G)
net94-96	236	3882	1052	37	35	0.009	0.94	0.0019
net96-98	254	8324	368	131	46	0.015	0.35	0.2309
net94-98	243	3220	488	54	53	0.016	0.98	0.0020
net14-16	227	5793	1775	52	50	0.008	0.96	0.0011
net16-18	222	5487	1635	25	18	0.004	0.72	0.0042
net14-18	232	8239	2085	13	12	0.001	0.92	0.0004

It is observed that although all the networks are nearly balanced (low d(G)), net96-98 differs noticeably in indexes $d_+(G)$, d_(G) as indexes for the network instability and confirms the above explanations that mentioned.

4-Results discussion

In summary the results clarify that:

- From the point of view of balance theory in the signed networks context, the results confirm that the networks corresponding to stock markets are nearly balanced and the best balanced partitioning is achieved by k=2. This is a novel result could be useful in detecting the whole structural pattern and dividing the stock network into two balanced clusters, and determining the central nodes and important companies in each cluster based on centrality measures like vertex degree, also detecting important connections in the overall system function.

- The results show that the connections inside two balance partitions are almost entirely positive and the edges violating balance condition are chiefly positive connections between the two partitions. This is an important observation emphasis on the important role of these links in the network coherence and information flow across the system. Ehsani and Sepehri (2014) extended some information diffusion models for signed networks and clarified that adaptation between density based clusters and balanced partitions blocks noise and information flows across the whole system and positive links between balanced partitions facilitates the information flow.

- Although the results show that the structure of all the stock markets as signed networks approximately fit the balance pattern, the distance from complete balance is important. Some indicators were defined based on this distance and the share of positive links and negative links in structural imbalance. It is observed that these indexes are noticeably different for net 96-98.
  The time period (1396-1398) coincides with some important political events like US withdrawal from Barjam and sanctions tightening, increase in the exchange rate and as a consequence, severe transformations in Iran economics and reshaping the business activities of companies and their relations. These abrupt changes eventuate distance increase from balance condition and transition toward new state corresponding to new relations adjusting new condition. So these indicators (d+(g), d-(G) could be offered for assessment of system stability and applicable in system dynamics prediction and crisis durations.
- The results show the similarity between the networks derived from Tehran stock market and

Nasdaq stock market in general structural properties. Although, these differences are remarkable: the average density in Nasdaq networks are higher compared to Tehran stock networks. This fact shows higher cohesion and connectedness. Also, these networks are more balanced. In addition, the results generally show that the average density in the balanced partitions of Nasdaq networks is more than the average density in the corresponding network totally means that the balanced partitions are close to dense clusters. According to the results of Ehsani and Sepehri (2014), this fact clarifies more stable structure against perturbations and positive links between balanced partitions facilitate the information flow across the network.

5- Conclusions

This paper proposed a new methodology for analyzing stock markets as signed networks in the perspective of balance theory. For this purpose, stock markets modeled as signed networks based on some data of Tehran stock market and Nasdaq in several time periods. The results show that the whole structure of these networks approximately fit the complete structural balance. Also it was observed that the imbalance connections are chiefly positive connections between two balanced partitions and these connections have an important role in the system cohesion and information flow across the network.

Moreover, some indexes proposed based on the distance from structural balance for assessment overall system instability, could be useful in analysis and predict system dynamics and transitions during abrupt changes and economic crisis periods. This methodology can lead to many applications in analyze, assessment, perdition and control of finance markets such as dividing the market into two balanced clusters and detecting the important components, nodes and connections, indeed main companies and industries and their important relations and their effects in the system. These applications could support decision making either in the micro level for example portfolio investment based on balanced clusters minimizing the investment risk, or in the macro level such as regulating the market by controlling dominant companies in different clusters.

Finally, some directive points for future research are mentioned here. Firstly, constructing the networks in the present study was performed based on linear correlation coefficients between stock prices time series. However, this method neglects nonlinearity and interdependence between variables. So, it would be useful to define the edge weights based on the more complicated concepts like conditional entropy. As the next point, the relation between the sign and weight values of edges and its distribution is very important in structural patterns detection and could be investigated in another research. Also, dynamism of structural pattern and evolution of main partitions and components in time and its relation to important events and crises is suggested as an important research issue applicable in prediction and regulation of finance markets.

References

Allen, F., Babus, A. (2009). *Networks in finance*, in Kleindorfer P., Wind J. (Eds.), The network challenge, Wharton School Publishing, pp 367–382.

Altafini, C. (2012). Dynamics of opinion forming in structurally balanced social networks. *PloS one*, *7*(6), e38135.

Arcangelis, A.M., Rotundo, G. (2016). *Complex Networks in Finance,* in Commendatore P., Matilla-García M., Varela L., Cánovas J. (Eds.), Cánovas J. Complex Networks and Dynamics: Lecture Notes in Economics and Mathematical Systems, vol 683, Springer.

Battiston, S., Glattfelder, J.B., Garlaschelli, D., Lillo, F. and Caldarelli, G., 2010. *The structure of financial networks.* In: Estrada E., Fox M., Higham D., Oppo GL. (eds) Network Science Springer, London, (pp. 131-163).

Bekiros, S., Nguyen, D. K., Junior, L. S., & Uddin, G. S. (2017). Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research*, 256(3), 945-961.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological review*, 63(5), 277.

Davis, J. A. (1967). Clustering and structural balance in graphs, *Human relations*, 20(2), 181-187.
Doreian, P., & Mrvar, A. (2009). Partitioning signed social networks. *Social Networks*, *31*(1), 1-11.

Ehsani, M., & Sepehri, M. M. (2014). Balanced clusters and diffusion process in signed networks, *Journal of Industrial and Systems Engineering*, 7(1), 104-117.

Esfahanipour A., Zamanzadeh S.E. (2013). Stock Market Filtering Model Based on Minimum Spanning Tree in Financial Networks, *Amirkabir International Journal of Science & Research*, (Vol. 45), No.1, pp. 67 – 75.

Garas, A., Argyrakis, P., & Havlin, S. (2008). The structural role of weak and strong links in a financial market network, *The European Physical Journal B*, 63(2), 265-271.

He, X., Du, H., Feldman, M.W. and Li, G., 2019. Information diffusion in signed networks. *PloS one*, *14*(10).

Heider, F. (1946). Attitudes and cognitive organization, *Journal of Psychology* 21, 107–112.

Iacono, G., & Altafini, C. (2010). Monotonicity, frustration, and ordered response: an analysis of the energy landscape of perturbed large-scale biological networks. *BMC systems biology*, *4*(1), 83.

Korbel, J., Jiang, X. and Zheng, B., 2017. Transfer entropy between communities in complex networks. *arXiv preprint arXiv:1706.05543*.

Lin, C. C., Lee, C. H., Fuh, C. S., Juan, H. F., & Huang, H. C. (2013). Link clustering reveals structural characteristics and biological contexts in signed molecular networks. *PloS one*, *8*(6), e67089.

Mantegna, R.N. (1999). Hierarchical structure in financial markets, *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1), 193-197.

Onnela, J. P., Kaski, K., & Kertész, J. (2004). Clustering and information in correlation based financial networks, *The European Physical Journal B,* 38(2), 353-362.

Singh, R., Dasgupta, S., & Sinha, S. (2014). Extreme variability in convergence to structural balance in frustrated dynamical systems, *EPL (Europhysics Letters)*, *105*(1), 10003.

Smith, H. L. (2008). *Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems* (Vol. 41). American Mathematical Soc.

Sontag ED. (2007). Monotone and near-monotone biochemical networks. *Systems and Synthetic Biology*, 1:59-87.

Su, Y., Wang, B., Cheng, F., Zhang, L., Zhang, X. and Pan, L., 2017. An algorithm based on positive and negative links for community detection in signed networks. *Scientific reports*, *7*(1), pp.1-12.

Traag, V. A., Van Dooren, P., & De Leenheer, P. (2013). Dynamical models explaining social balance and evolution of cooperation. *PloS one*, *8*(4), e60063.

Tse CK. (2010). A network perspective of the stock market, *Journal of Empirical Finance*, (Volume 17), Issue 4, September, Pages 659-667.