**Электронный научный журнал "Математическое моделирование, компьютерный и натурный эксперимент в естественных науках" http://mathmod.esrae.ru/**
URL статьи: mathmod.esrae.ru/19-72
**Ссылка для цитирования этой статьи:**
Faizliev A.R., Levshunov M.A, Glazov R.V., Tryapkina T.S, Mironov S.V., Androsov I.A, Petrov V.S. Preliminary Analysis of the Evolution of Market Graph Characteristics // Математическое моделирование, компьютерный и натурный эксперимент в естественных науках. 2018. №3
***The work was supported by the Russian Fund for Basic Research, project 18-37-00060.***

УДК 51-77, 519.17

# PRELIMINARY ANALYSIS OF THE EVOLUTION OF MARKET GRAPH CHARACTERISTICS

Файзлиев А.Р.[1], Левшунов М.А, Глазов Р.В., Тряпкина Т.С, Миронов С.В.[2], Андросов И.А., Петров В.С.
Саратовский государственный университет имени Н.Г.Чернышевского, Россия, Саратов
[1]faizlievar1983@mail.ru, [2]mironovsv@info.sgu.ru

# ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ЭВОЛЮЦИИ ХАРАКТЕРИСТИК РЫНОЧНОГО ГРАФА

Faizliev A.R.[1], Levshunov M.A., Glazov R.V., Tryapkina T., Mironov S.V.[2], Androsov I.A, Petrov V. S.
Saratov State University, Russia, Saratov
[1]faizlievar1983@mail.ru, [2]mironovsv@info.sgu.ru

**Аннотация.** В работе формируются и исследуются рыночные графы. Сети, представленные такими графами, достаточно похожи по строению на социальные сети или сети совместного цитирования. Каждая компания является узлом, и положительная значимая корреляция между активами двух компаний устанавливает связь между ними. Матрица, содержащая связи между парами компаний, создана для сетевого анализа компаний, акции которых торгуются на финансовых рынках США. Было показано, что распределение степеней и коэффициент кластеризации для нашей сети подчиняются степенному закону. Для построения графов использовались реальные рыночные данные. Алгоритмы для формирования и анализа сети и для визуализации результатов реализованы с использованием языка C++.

Ключевые слова: анализ сетей, рыночный граф, распределение степеней, максимальная клика

**Abstract.** In our research we form a network which is called a market graph. The network is constructed quite similar to social networks or co-citation networks. Each company is a node and the positive significant correlation between assets of the two companies establishes a link between them. A matrix containing links between pairs of companies is created for network analysis of companies whose shares are traded on financial markets of the USA. It was shown that distribution of degrees and clustering coefficient for our network follows the power law. Market data have been employed to constructed graph, and C++ has been used for network analysis as well as network visualization.

Keywords: network analysis, market graph, degree distribution, maximum clique

**Introduction.** One of the most important problems in modern finance is the search for effective ways to generalize and visualize the stock market data. It can provide researchers and practitioners with useful information about the behavior of the market. Currently, a large number of shares are traded on the stock markets and their number is steadily increasing. The huge amount of data is being generated by the stock market every day. This data is usually visualized by thousands of charts reflecting the price of each asset for a certain period of time. The analysis of such data is becoming increasingly difficult as the number of shares increases.

One of the key aspects of modern economic systems is that they behave as complex systems with a huge amount of interdependent parts and connections. Analysis of the properties of the market network has attracted increasing attention in the last decade. The concept of a market graph was considered in [1], in which the market network is defined as a full weighted graph where the nodes represent the assets and the weights of the arcs reflect the similarity between the behavior of assets. In the article [1], the edge between two vertices is inserted into the market graph if the corresponding value of the correlation coefficient is higher than the specified threshold. In recent years, there has been an increased interest to applying and developing an approach based on the market graph. These research papers include empirical studies based on real market data and examine the various structural properties and attributes of the market graph, such as maximum clicks, maximum independent sets, the distribution of powers [2-5], clustering of the Pearson correlation [6], the dynamics of the market graphs of the US market [7], the complexity of the market graph [8]. The articles [3, 9-12] study the distinctive features of individual financial markets. Market graphs with similarity measures that differ from the correlation are studied in [9, 13-17].

Social network analysis (SNA) allows us to analyze the structure of relations in an organization [18, 19]). The paper [20] considers SNA as a method of examining relationships among social entities. The fundamental concepts of SNA are node and link. A node is the unit (individual, object, item) and a link serves as the relationship between nodes.

Data of financial market can be easily transformed into network data. A market network is a set of companies, which have connections in pair to represent their relationship. Two companies are considered in a relationship if there has been positive significant correlation between their assets. In such type of network, a company will be called as “node” or “vertex” and the connection will be an “edge”. Market network will be represented by undirected unweighted graph. Market network is similar to social networks. Different type of social network analysis metrics can be used for finding edge density, degree distribution, maximum clique and maximum independent set in the network.

This methodology allows you to visualize a set of data representing its elements in the form of vertices and observe certain relationships between them. The study of the structure of the graph representing the data set is important for understanding the internal properties of the market that it represents, as well as for improving the organization of storage and retrieval of information.

In our research we would like to find the type of the degree distribution, the type of the clustering-degree distribution exhibited by the market network. Moreover, we would like to estimate the size of the maximum clique in the market graph.

Note that the last two decades have seen extensive research in the area of degree distribution analysis of complex networks arisen in sociology, physics, and biology. It has been shown that many networks have similar degree distributions [21-26]. It turned out that most of real networks have degree distributions that are scale-free [21]. In other words, their degree distributions are power-law.

The main purpose of this paper is to identify the dynamics of changes in the structural properties of the market graph over time. The paper deals with graphs based on stock prices data for different periods of time during 2013–2017 to study the evolution of some characteristics of these graphs.

## 1. Data

The database for constructing and analyzing the market graph was taken from the resource [27]. The daily data were collected from Thomson Reuters database, which was used to retrieve historical prices of the companies traded in the NYSE and NASDAQ for the period from November 22, 2013 to November 10, 2017 (i.e. 1000 trading days). The daily closing prices have been adjusted for dividends and splits. Our analysis includes only stocks only stocks that had been traded without gaps and omissions during this period (3736 different stocks remained, and only 15 stocks from S&P500 except 15 were eliminated).

To study the dynamics of the market graph, the 1000-day trading days interval was divided into 10 consecutive 500-day periods. Each period except the first is obtained by shifting the previous one by 50 days. Thus, two neighboring periods have 450 common days. The dates corresponding to each period are presented in Table 1.

Time periods Table 1

| Period | Start | End |
|---|---|---|
| 1 | 22.11.2013 | 13.11.2015 |
| 2 | 04.02.2014 | 26.01.2016 |
| 3 | 17.04.2014 | 07.04.2016 |
| 4 | 30.06.2014 | 20.06.2016 |
| 5 | 10.09.2014 | 31.08.2016 |
| 6 | 21.11.2014 | 11.11.2016 |
| 7 | 03.02.2015 | 24.01.2017 |
| 8 | 16.04.2015 | 06.04.2017 |

| | | |
|---|---|---|
| 9 | 29.06.2015 | 19.06.2017 |
| 10 | 09.09.2015 | 30.08.2017 |
| 11 | 20.11.2015 | 10.11.2017 |

Market network is formed based on correlation; it means that a company has connection with those companies which have the positive significant correlation of assets with it in this period of time.

The formal procedure for constructing the market graph is as follows. We denote by $P_i(t)$ the price of the asset $i$ in day $t$. Then

$$R_i(t) = \ln \frac{P_i(t)}{P_i(t-1)} \tag{1}$$

is the logarithm of the ratio of the price of the asset $i$ in day $t$ to the price in the previous day $t-1$. Let

$$C_{ij} = PCC\big(R_i(1), R_i(2), \ldots, R_i(k), R_j(1), R_j(2), \ldots, R_j(k)\big), \tag{2}$$

where $PCC$ is the Pearson correlation coefficient.

The edge between the vertices $i$ and $j$ is added to the graph if $C_{ij} \geq \theta$, which means that the prices for these two assets behave identically over time, and the degree of this similarity is determined by the corresponding value of the Pearson correlation coefficient.

## 2. Network Analysis

### 2.1. Edge Density

The edge density of a simple undirected graph $G$ is defined as the ratio of the number of edges of a graph to the maximum possible number of edges in it [28]:

$$D = \frac{2|E|}{|V|(|V|-1)}, \tag{3}$$

where $V$ is the number of vertices of the graph and $E$ is the number of edges of a graph.

The edge density is an important characteristic of the market graph. The increase in the edge density indicates a certain “globalization” of the stock market, i.e. that more and more assets significantly affect each other and the change in prices of one asset entails a change in the prices of other stock assets.

### 2.2. Degree Distribution

The graph $G = (V, E)$ is connected if there is a path from any vertex to any vertex in the set $V$. If the graph is disconnected, it can be decomposed into several connected subgraphs, which are referred to as the connected components of $G$.

The degree of a vertex is the number of edges emanating from it. For every integer number k one can calculate the number of vertices $n(k)$ with the degree equal to $k$, and then get the probability that a vertex has the degree $k$ as $P(k) = n(k)/n$, where n is the total number of vertices. The function $P(k)$ is referred to as the degree distribution of the graph. The degree distribution is an important characteristic of a graph representing a dataset.

It should be noted that real graphs that arise in different fields (economics, Internet, telecommunications, finance, medicine, biology, sociology) exhibit the degree distribution that follows the power-law model [21-26]. According to this model, the probability that a vertex has degree $k$ (that is, there exist $k$ edges originating from it) asymptotically follows

$$P(k) \propto k^{-\gamma} \text{ or } \log P(k) \propto -\gamma \log k,$$

which shows that this function has a linear dependence in the logarithmic scale.

An important characteristic of this model is its scale-free property. It implies that the fractal structure of a network remains constant despite its development and growth over time [29].

### 2.3. Clustering Analysis

The local clustering coefficient for node $i$ is defined by

$$C_i = \frac{E_i}{k_i(k_i - 1)},$$

where $E_i$ is the number of links connecting the immediate neighbors of node $i$, and $k_i$ is the degree of node $i$. The average value of clustering coefficients of all nodes in a network is called the average clustering coefficient. The value of the average clustering coefficient quantifies the strength of connectivity within the network. The paper [30] examines protein-protein interaction networks and metabolic networks, which have to demonstrate large average clustering coefficients. The analogues result has been established for collaboration networks in academia and the entertainment industry in papers [31, 32]. Let $C(k)$ denote the average clustering coefficient of nodes with degree $k$. It has been found that for most of real networks $C(k)$ follows

$$C(k) \sim \frac{B}{k^{\beta}},$$

where the exponent $\beta$ usually lies between 1 and 2 [33-35].

### 2.4. Maximum Cliques

Given a subset $S \subseteq V$, by $G(S)$ we denote the subgraph induced by $S$. A subset $C \subseteq V$ is a clique if $G(C)$ is a complete graph, i.e. it has all possible edges. The maximum clique problem is to find the largest clique in a graph.

The clique is a set of vertices, which are fully interconnected. That is why any financial asset, which belongs to the click, is strongly correlated with all other financial assets in this click. Because of this fact, the asset is bound to a specific click only in case when its behavior is similar to all other assets in this group. It is clear that one of the main characteristics of stock market is the maximum size of clique, because it shows the largest possible group of similar objects (financial assets, which are cross correlated to each other).

The maximum clique problem (as well as the maximum independent set problem) is known to be NP-hard [36]. Moreover, it turns out that these problems are difficult to approximate [37, 38]. This makes these problems especially challenging in large graphs. However, as we will see later, a special structure of the co-mention graph allows us to get the exact solution of the maximum clique problem.

The variant of Bron – Kerbosch algorithm is used in order to calculate an accurate maximum click. Bron – Kerbosch algorithm is the algorithm which allows to find maximal cliques in the undirected graph [39]. Dutch scientists Bron Conradomi and Jupe Kerbosh developed this algorithm and published it in 1973. There some other algorithm, which can solve the problem of maximum clique and works better in some graphs with a little quantity of vertexes. Actually, Bron – Kerbosch algorithm and its improvements work effectively.

The main form of Bron – Kerbosch algorithm is recursive search algorithm with return. It founds all maximum cliques in the graph $G$. The algorithm is linear relative to the number of cliques in the graph. The working time of this algorithm with some extra tests $O\left(n^{n/3}\right)$. But the algorithm is more effective for random graphs.

### 3. Evolution of Market Network

The density of distribution of correlation coefficients for the US stock market is almost symmetrical and has a form which is similar to the normal with the mean around 0.2 (Fig. 1, 2). The comparison of densities for different periods of time shows that distributions are similar to each other. The proportion of present edges to all possible edges in the network are shown in Table 2. The density of edges increases over the time, its peak is reached during 5 and 6 periods and after that it goes down. (Table 2). A positive mean implies that financial assets of the USA market are related to each other on average. The correlation in case of negative mean is rather rare. Because of that, it is more difficult to form a diversified portfolio of shares whose yields move in different directions. The hypothesis on power-low degree distribution of vertexes' degree is confirmed. It means that degree distribution of vertexes is approximated by a power-law model. At the same time the power coefficient $\gamma$ is less than 1 for all periods of time. For the given networks, the clustering-degree distribution relation also follows the power law (Fig. 3). The resulting models is statistically significant at any significance level. Herewith, the exponent $\beta$ turns out less 1 for all the subgraphs under consideration (Table 2). The plot of the clustering-degree relation, i.e. $C(k)$ as a function of node degree $k$, is shown in Fig. 4.

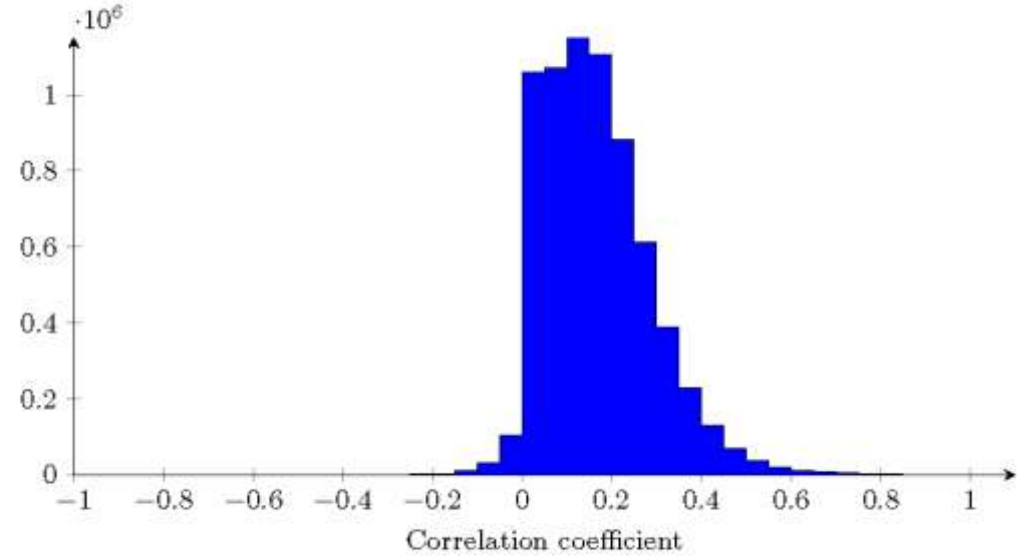


Fig. 1. Distribution of correlation coefficients (1st period)

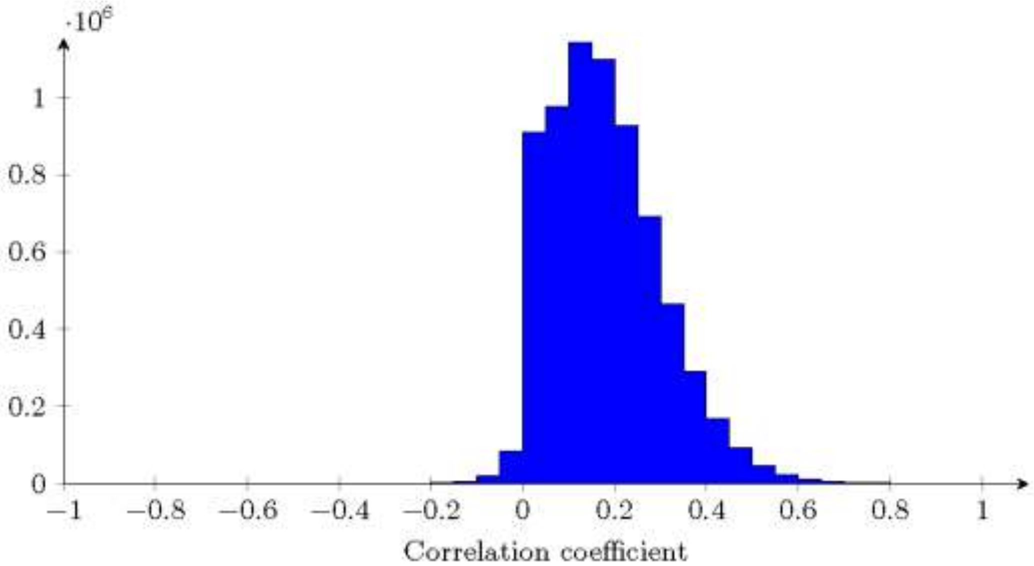


Fig. 2. Distribution of correlation coefficients (11th period)

Table 2

Characteristics of graphs

| Period | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Density | 0.014 | 0.018 | 0.019 | 0.02 | 0.023 | 0.023 | 0.02 | 0.02 | 0.019 | 0.014 | 0.012 |
| Coefficient $\gamma$ | 0.86 | 0.84 | 0.82 | 0.82 | 0.79 | 0.81 | 0.84 | 0.84 | 0.85 | 0.92 | 0.82 |
| Coefficient $\beta$ | 0.25 | 0.25 | 0.24 | 0.24 | 0.22 | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.22 |
| Clique size | 116 | 127 | 139 | 147 | 166 | 159 | 149 | 149 | 149 | 141 | 137 |

**ISSN 2541-9269**

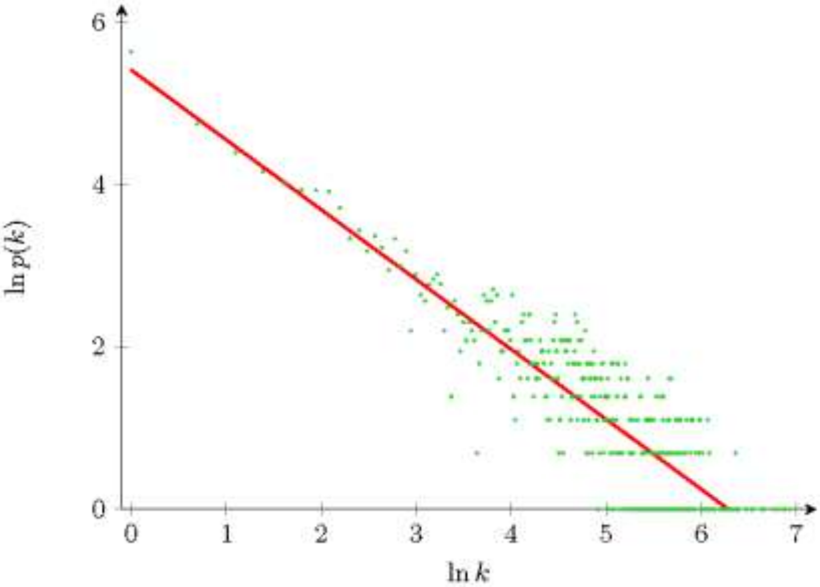

Fig. 3. The degree distribution of the market network

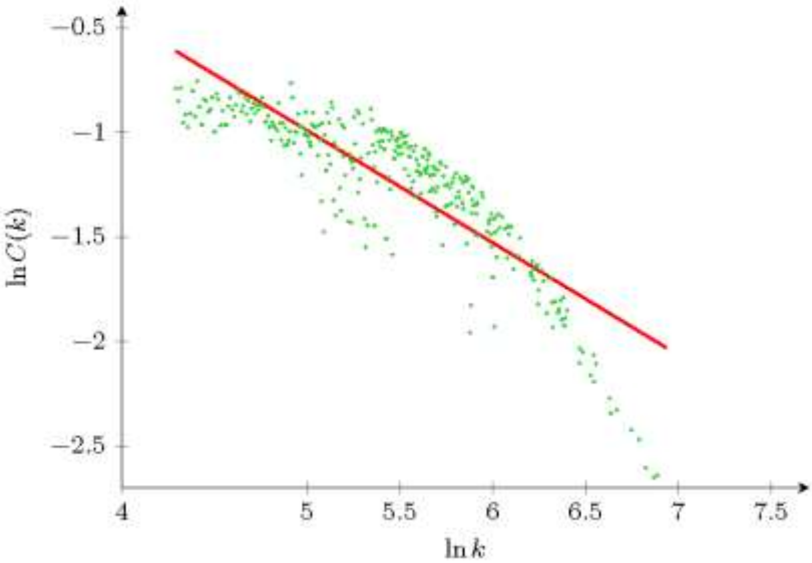

Fig. 4. The clustering-degree relation of the market network

The sizes of the maximum cliques of the market of the USA big enough (Table 2), and the peak is account form September 2014 till August 2016. A clique is a set of fully interconnected vertices. That is why any asset owned by the clique is strongly associated with all other assets in this clique. In this way, an increase of the maximum clique may mean an intensity increase of markets' globalization in this period of time.

**Conclusion.** In this paper we transform financial data into the market graph. The examination of graph properties gives new understanding of the financial internal structure. We investigated the dynamics and changes of the market graph structural properties over time. As a result, we came to several interesting conclusions based on our research. It was shown that the power-law structure of the market graph is fairly stable. Unlike real social graphs, the market graph displays power-low distribution of degrees with non-typical indicators of degree exponent. Therefore it can be outlined that the concept of 'self-organized network' may be employed for the market graph, and the financial market can be viewed as a 'self-organized' system. The sizes of the maximum cliques on the market of the USA are big enough and the peak is account from September 2014 till August 2016.

*This work was supported by the Russian Fund for Basic Research, project 18-37-00060.*

**References**

1. Boginsky V., Butenko S., Pardalos P.M. On structural properties of the market graph // Innovations in Financial and economic networks / Ed. by A. Nagurney. Northampton: Edward Elgar Publishing Inc., 2003. P. 29-45.
2. Boginski V., Butenko S., Pardalos P.M. Statistical analysis of financial networks // Computational Statistics & Data Analysis. 2005. V. 48. №2. P. 431-443.
3. Huang W.-Q., Zhuang X.-T., Yao S.A network analysis of the Chinese stock market // Physica A: Statistical Mechanics and its Applications. 2009. V. 388. №14. P. 2956-2964.
4. Tse C.K., Liu J., Lau F.C.M. A network perspective of the stock market // Journal of Empirical Finance. 2010. V. 17. №4. P. 659-667.
5. Boginski V., Butenko S., Pardalos P.M. Network models of massive datasets // Computer Science and Information Systems. 2004. V. 1. №1. P. 75-89.
6. Onnela J.-P., Kaski K., Kertész J. Clustering and information in correlation based financial networks // The European Physical Journal B. 2004. V. 38. №2. P. 353-362.
7. Boginski V., Butenko S., Pardalos P.M. Mining market data: A network approach // Computers & Operations Research. 2006. V. 33. №11. P. 3171-3184. Special Issue: Operations Research and Data Mining.
8. Emmert-Streib F., Dehmer M. Identifying critical financial networks of the DJIA: Toward a network-based index // Complexity. 2010. V. 16. №1. P. 24-33.
9. Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. Simple measure of similarity for the market graph construction // Computational Management Science. 2013. V. 10. №2. P. 105-124.
10. Garas A., Argyrakis P. Correlation study of the Athens stock exchange // Physica A: Statistical Mechanics and its Applications. 2007. V. 380. №C. P. 399-410.
11. Vizgunov A., Goldengorin B., Kalyagin V., Koldanov A., Koldanov P., Pardalos P.M. Network approach for the Russian stock market // Computational Management Science. 2014. V. 11. №1. P. 45-55.
12. Namaki A., Shirazi A.H., Raei R., Jafari G.R. Network analysis of a financial market based on genuine correlation and threshold method // Physica A: Statistical Mechanics and its Applications. 2011. V. 390. №21. P. 3835-3841.

13.Bautin G.A., Kalyagin V.A., Koldanov A.P. Comparative analysis of two similarity measures for the market graph construction // Models, Algorithms, and Technologies for Network Analysis / Ed. by B.I. Goldengorin, V.A. Kalyagin, P.M. Pardalos. New York, NY: Springer New York, 2013. P. 29-41.
14.Shirokikh O., Pastukhov G., Boginski V., Butenko S. Computational study of the US stock market evolution: a rank correlation-based network model // Computational Management Science. 2013. V. 10. №2-3. P. 81-103.
15.Wang G.-J., Xie C., Han F., Sun B. Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree // Physica A: Statistical Mechanics and its Applications. 2012. V. 391. №16. P. 4136-4146.
16.Kenett D.Y., Tumminello M., Madi A., Gur-Gershgoren G., Mantegna R.N., Ben-Jacob E. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market // PLoS ONE. 2010. V. 5. №12. P. e15032.
17.Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. Optimal decision for the market graph identification problem in a sign similarity network // Annals of Operations Research. 2017. V. 266. №1-2. P. 313-327.
18.Barnett G.A., Danowski J.A. The structure of communication: A network analysis of the international communication association // Human Communication Research. 1992. V. 19. №2. P. 264-285.
19.Barnett G.A., Salisbury J.G.T. Communication and globalization: A longitudinal analysis of the international telecommunication network // Journal of World System Research. 1996. V. 2. №16. P. 1-17.
20.Wasserman S., Faust K. Social Network Analysis. Cambridge University Press, 1994.
21.Albert R., Barabasi A.-L. Statistical mechanics of complex networks // Reviews of Modern Physics. 2002. V. 74. P. 47-97.
22.Dorogovtsev S.N., Mendes J.F.F. Evolution of networks // Advances in Physics. 2002. V. 51. №4. P. 1079-1187.
23.Newman M.E.J. The structure and function of complex networks // SIAM Review. 2003. V. 45. №2. P. 167-256.
24.Albert R. Scale-free networks in cell biology // Journal of Cell Science. 2005. V. 118. P. 4947-4957.
25.Boccaletti S., Latora V., Moreno Y. Chavez M., Hwang D.U. Complex networks: Structure and dynamics // Physics Reports. 2006. V. 424. P. 175-308.
26.Lofdahl C., Stickgold E., Skarin B., Stewart I. Extending generative models of large scale networks // Procedia Manufacturing. 2015. V. 3, Supplement C. P. 3868-3875. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.
27.Kaggle. URL: https://www.kaggle.com.
28.Diestel R. Graph theory. Springer, 2017.
29.Barabási A.-L., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. №5439. P. 509-512.

30. Wagner A., Fell D.A. The small world inside large metabolic networks // Proceedings of the Royal Society of London Series B Biological Sciences. 2001. V. 268. P. 1803-1810.
31. Anthonisse J.M. The rush in a directed graph: Rep. Amsterdam: Stichting Mathematisch Centrum, 1971. URL: http://oai.cwi.nl/oai/asset/9791/9791A.pdf.
32. Granovetter M. The strength of weak ties // American Journal of Sociology. 1973. V. 78. P. 1360-1380.
33. Ravasz E., Somera A.L., Mongru D.A., Oltvai Z.N., Barabasi A.L. Hierarchical organization of modularity in metabolic networks // Science. 2002. V. 297. P. 1551-1555.
34. Ravasz E., Barabási A.-L. Hierarchical organization in complex networks // Physical Review E. 2003. V. 67. №2. P. 026112.
35. Yook S.H., Oltvai Z.N., Barabasi A.L. Functional and topological characterization of protein interaction networks // Proteomics. 2004. V. 4. P. 928-942.
36. Garey M.R., Johnson D.S. Computers and Intractability; A Guide to the Theory of NP-Completeness. New York, NY: W.H. Freeman & Co., 1990.
37. Arora S., Safra S. Probabilistic checking of proofs: a new characterization of NP // Proceedings of 33rd Annual Symposium on Foundations of Computer Science. IEEE, 1992. P. 2-13.
38. Hástad J. Clique is hard to approximate within $n^{(1-\varepsilon)}$ // Proceedings of 37th Conference on Foundations of Computer Science. IEEE Comput. Soc. Press, 1996. P. 627-636.
39. Cohn R., Russell J. Bron – Kerbosch algorithm. VSD, 2013.