Comparative Analysis of Financial Network Topology for the Russian, Chinese and US Stock Markets

VLADIMIR BALASH  
Saratov State University  
Faculty of Mechanics and Mathematics  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
vladimirbalash@yandex.ru

SERGEI SIDOROV  
Saratov State University  
Risk Institute  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
sidorovsp@yahoo.com

ALEXEY FAIZLIEV  
Saratov State University  
Risk Institute  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
faizlievar1983@mail.ru

ALFIA CHEKMAREVA  
Saratov State University  
Risk Institute  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
chekmareva.alfiya@mail.ru

ALEXEY GRIGORIEV  
Saratov State University  
Faculty of Mechanics and Mathematics  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
alexprgrigoriev@gmail.com

DMITRIY MELNICHUK  
Saratov State University  
Faculty of Mechanics and Mathematics  
83 Astrakhanskaya Street  
RUSSIAN FEDERATION  
melnichukdv@sgu.ru

Abstract: This paper studies the properties of the Russian stock market by employing the data-driven science and network approaches. The theory of complex networks allows us to build and examine topological network structures of the market with the further identification of relationships between stocks and the analysis of hidden information and market dynamics. In this paper we will present an analysis of structural and topological properties of the networks constructed for the US and China stock markets with the properties of corresponding networks constructed for the Russian stock market using a dataset spanning over eight years.

Key–Words: Network analysis, Market graph, Degree distribution, Hierarchical tree, Minimum spanning tree


1 Introduction

To represent the interaction framework between agents in economic or financial systems, one can use network models. The definition of a financial network was given in [1], in which the nodes are various organizations, including banks, firms and investors, which are linked through a network of financial interdependencies, for example, cross-ownership of financial assets, inter-organizational debts / liabilities, social relations between board members and so on. A classic example of a banking (financial) network is built in [2], where it is illustrated how systemic risk can arise as a result of financial interdependencies between banks. Most of real-world networks are constantly changing over time. The links between economic agents are usually activated or deactivated in dependence of the irregular interactions between the elements of the economic system. These activation templates may be sporadic and represent a huge role in the dynamics of network processes. One of examples of this situation is information spreading in social networks. The topology of network (heterogeneity in the connections, abundance of loops) plays an important role in the economic system dynamics behavior.

Financial markets and corresponding networks may have a huge amount of assets (nodes) and inter-
connections (edges). It is useful to simplify initial networks by filtering the noises. Some of the approaches employed in filtering of financial networks are based on hierarchical tree (HT) [3, 4], planar maximally filtered graphs (PMFG) [5], asset graph (AG) [6], and partial correlation network [7, 8, 9]. Another successful method of financial network simplification was developed by Mantegna [3] and applied in [10]. The approach is based on the minimum spanning tree method for examining of hierarchical structures of financial networks.

In recent years there has been an increased interest in the application and development of an approach based on market graphs. These papers include empirical studies based on real market data and explore various structural properties and attributes of market graphs, such as maximum clicks, maximum independent sets, degree distribution [11, 12, 13, 14], clustering, Pearson correlations [15], market graphs of the US market [16], the complexity of the market graph [17]. The concept of a market graph was first considered in [18], in which a market network is defined as a full weighted graph, where nodes represent assets and arc weights reflect similarities between asset behavior (which can be measured, for example, using correlation). In article [18], the edge between two vertices is inserted into the market graph, if the corresponding value of the correlation coefficient is above a given threshold. The articles [19, 20, 21, 12, 22] studied the distinctive features of individual financial markets. Market graphs with similarity measures other than correlation were studied in [19, 23, 24, 25, 8, 26].

In this paper, we use some methods of complex network theory to study Russian financial market network. Complex network analysis and SNA methods allow us to investigate the structure of Russian financial market network that are based on the correlation relationship between the components of the system (assets).

Network analysis methods have proved to be successful in studying various individual major financial markets such as US market [27, 28, 29], Germany [30, 4], EU [31, 32, 33], Italian [34, 35]. Some research papers have been investigate developing markets such as China [29, 36, 12, 37], Brazil [38, 39], Korea [40], Russia [21], and Mexico [41, 42]. Moreover, some papers employed network approach to examine global markets [43, 44].

It turned out that the network analysis is capable to examine the importance of companies using different centrality measures [45, 45] and to analyze the systemic risks and financial market stability based on the topological properties of networks [46, 47, 48]. Systemic risk contagion has been analyzed in [49, 50] to find major influencers based on banking behaviors in the global environment. The paper [51] studies European markets and shows that EU markets are exposed to systemic risks. It should be noted that systemic risk analysis based on network approach may provide useful implications for market regulations.

The paper [29] employs various methods of network analysis to investigate financial networks of both in China and the United States and to study how the two markets behave differently. In this paper, we will present a similar analysis of Russian stock market. One of the main research questions of this paper is to find how the US and China stock markets differ in the structures and topological properties from Russian stock market using

- market graph approach proposed in [18, 11],
- hierarchical tree (HT) method [3],
- minimum spanning tree (MST) approach [3].

In this paper we compare topological properties of the networks constructed for the US and China stock markets in [29] with the properties of corresponding networks constructed for the Russian stock market using a dataset spanning over nine years.

The paper is organized as follows. First, Section 2 describes the data and method used to construct networks. Section 3 presents the market network properties using market graph approach proposed in [18, 11]. Section 4 examines the hierarchical structures of the Russian stock network using HT, and MST approaches and compares them with findings of the paper [29]. Finally, conclusions are presented in Section 5.

2 Data and Methodology

2.1 Data Description

Russia has a developing economy which is the sixth among countries in the world in terms of GDP (PPP). The RTS index and the Moscow Stock Exchange index are the main indicators of the Russian stock market. The difference between them is that the RTS is calculated in dollars, and the MICEX in rubles. Historical data of daily prices for each company were taken from open sources on Yahoo Finance, for the period from 10/01/2012 to 09/04/2019 (1664 trading days). Similarly, data on the MICEX index were taken.
from the Moscow Stock Exchange for the same period. A total of 194 companies are considered, among which 32 companies are part of the MICEX index.

The descriptive statistics for log returns of the IMOEX index over the period between 10/01/2012 and 09/04/2019 are presented in Table 2.2. The result of Garch(1,1) modelling for log returns of the IMOEX index over the period between 10/01/2012 and 09/04/2019 is presented in Fig. 1.

Some companies (CBOM, DSKY, FIVE, LNTA, MOEX, POLY, RNFT, RUAL, SFIN) that are part of the MICEX index were not included in our dataset. This is due to the fact that they have a very short trading history (only 2-3 years), which is insufficient to build a relevant model. We also did not include companies that had more than 250 missed trading days in a row in our dataset. Thus, only 194 issuers remained in our dataset from the initial list of 278 issuers traded on the Moscow stock exchange. The YNDX company has started to be traded on the Moscow stock exchange from 02/06/2014, but it has been traded on the NASDAQ from 02/05/2011, from which the missing data were taken. After cleaning and the selection process, we include 194 stocks traded on IMOEX in our dataset, as shown in Table 1. We place these 194 stocks among all ten industry sectors. Table 2.2 presents basic information about datasets.

### 2.2 Market Network Construction

To calculate the correlation $\rho_{ij}$ for a pair of stocks, it is necessary to use time series of prices (Adj Close) $P_i(t)$ for each company $s_i$ at the time point $t$. To smooth the oscillations we use the log returns $Y_i(t)$ of the company $s_i$ in the time period $[t - \Delta t, t]$ defined by

$$Y_i(t) = \ln P_i(t) - \ln P_i(t - \Delta t), \quad (1)$$

where $\Delta t = 1$ for daily prices. Then we find the Pearson correlation coefficient for each pair of companies $s_i$ and $s_j$ as follows

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{\langle (Y_i^2) \rangle - \langle Y_i \rangle^2} \sqrt{\langle (Y_j^2) \rangle - \langle Y_j \rangle^2}} \quad (2)$$

where $\langle \cdot \rangle$ denotes the average value.

Based on the values of the Pearson correlation coefficient for each pair of companies we can construct the $N \times N = 194 \times 194$ distance matrix using the equation

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \quad (3)$$

### Table 1: 194 component stocks of IMOEX are included in our dataset. In this table, we list sector name, and numbers of stocks for each industry sector of these 194 stocks. All 10 industry sectors are represented

<table>
<thead>
<tr>
<th>Sector</th>
<th>Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Materials</td>
<td>36</td>
</tr>
<tr>
<td>Consumer Cyclicals</td>
<td>11</td>
</tr>
<tr>
<td>Consumer Non-Cyclicals</td>
<td>9</td>
</tr>
<tr>
<td>Energy</td>
<td>22</td>
</tr>
<tr>
<td>Financials</td>
<td>16</td>
</tr>
<tr>
<td>Healthcare</td>
<td>3</td>
</tr>
<tr>
<td>Industrials</td>
<td>24</td>
</tr>
<tr>
<td>Technology</td>
<td>3</td>
</tr>
<tr>
<td>Telecommunications Services</td>
<td>11</td>
</tr>
<tr>
<td>Utilities</td>
<td>59</td>
</tr>
</tbody>
</table>

Figure 1: GARCH(1,1) time series of the volatility of the IMOEX index daily log returns

### 3 Basic network properties

In this section, we examine the topological properties of the IMOEX stock network using the dynamic approach applied in [29]. We take the sliding window size $L = 250$ to meet the requirement of $L/N > 1$, where the number of stocks is $N = 194$. In total, there are 1413 windows for IMOEX. We calculated the log returns for all IMOEX stocks by using Eq. (1) and then we found the Pearson correlation coefficient matrix for each window over the period between 10/01/2012 and 09/04/2019 using Eq. (2).

We compare the results obtained for the Russian stock market with the results obtained in the paper [29] for the Chinese and US stock markets. The dataset S&P468 has been described in [29] and consists of 468 component stocks of S&P500. The
dataset CSI163 has been described in [29] and includes of 163 component stocks of CSI300.

Thus, for IMOEX we obtain a network with 194 vertices. Using Eq. (3) we calculate the distance matrices for each window. It should be noted that all negative correlation values are transformed into positive distances. All vertices in the network are fixed and correspond to stocks. However, the edges vary in each sliding window as the distances change. Then we average all the distance matrices over all sliding windows of the study periods. The statistical properties of the IMOEX stock network are examined in this section.

The maximum number of edges in average distance matrix is equal to $N(N-1)/2 = 18,721$ and it is a huge number. We can simplify the network by filtering edges with small weights. In this section the market graph is constructed based on threshold approach: for a fixed threshold value of $\theta$ we delete those edges whose weight are bigger than $\theta$ and preserve the remaining edges. In other words, we filter the network edges as follows:

$$e_{ij} = \begin{cases} 
1, & d_{ij} < \theta, \\
0, & d_{ij} \geq \theta.
\end{cases}$$

Table 3 presents the basic properties of the IMOEX network for different values of $\theta$ taken from 0.5 to 1.5. It can be seen that the number of edges is zero if the threshold value is $\theta = 0.8$, while the graph becomes complete at $\theta = 1.5$. The figure 2 shows edge densities for different thresholds from 0.5 to 1.5. In the interval from 0.5 to 1.1, network densities are close to 0, which means that all edges are filtered. In the range from 1.43 to 1.5, the densities are close to 1, i.e. all edges are present at the network. For charts based on Chinese and American stock markets [29] we see that the two curves have a similar shape with a slopes that lie between 0.6 and 1. It should be noted that the transformation interval for the IMOEX network is from 1.1 to 1.43, which is atypical for stock networks. As a rule, stock networks show the transformation interval from 0.6 to 1. The study shows that Russian stock network has a different form of transforming interval than the USA and China. This is due to much lower correlations between assets traded in the Russian stock market, which could be explained by e.g. low liquidity of most stocks in IMOEX. Moreover, it can be easily verified that even highest liquidity shares provide low correlations. Network density for a graph constructed using these stocks maintains similar dependency from $\theta$ as for the whole IMOEX network.

Table 2: Basic information of IMOEX datasets including the number of stocks, the number of trading days, the average log returns $\langle \hat{r} \rangle$, and the standard deviation $\sigma_r$ of the log returns is presented.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Stocks</th>
<th>Days</th>
<th>$\langle \hat{Y} \rangle$</th>
<th>$\sigma_Y$</th>
<th>$\langle Y_{\text{min}} \rangle$</th>
<th>$\langle Y_{\text{max}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMOEX</td>
<td>194</td>
<td>1664</td>
<td>2.66e-4</td>
<td>0.0125</td>
<td>-0.1141</td>
<td>0.0551</td>
</tr>
<tr>
<td>CSI163</td>
<td>163</td>
<td>2149</td>
<td>1.48e-04</td>
<td>0.0340</td>
<td>-0.4832</td>
<td>0.1002</td>
</tr>
<tr>
<td>S&amp;P468</td>
<td>468</td>
<td>2228</td>
<td>1.47e-04</td>
<td>0.0252</td>
<td>-0.3413</td>
<td>0.2160</td>
</tr>
</tbody>
</table>

Figure 2: Edge densities of CSI163, S&P468 and IMOEX for different thresholds $\theta$ from 0.5 to 1.5. It shows that the densities for IMOEX increase sharply from 0 to 1 in the interval of $\theta$ from 1.05 to 1.45.

### 3.1 Degree distribution
We investigate the degree distributions for the IMOEX networks filtered with different values of $\theta$. The degree distributions are noisy for both small values of $\theta$ and large ones. For a sufficiently small interval $\theta = (1.15; 1.2)$, the degree distributions follow closely to the power law distribution. After processing the data, we build the degree distribution with the typical power law shape in Fig. 3. The log-binning procedure was used to build the regression and calculate the power law exponent. The value of the power law exponent is $\gamma = 1.14$ with $R^2 = 0.85$. Since the use of linear approximation (linear-binning) significantly underestimates the exponent $\gamma$. A negative tilt angle indicates that the network is scale-free, in which a small part of
the vertices has greater degrees, and most of the vertices have smaller degrees.

Fig. 3 shows that the degree distributions of IMOEX (with $\theta = 1.2$), CSI163 (with $\theta = 0.68$) and S&P468 (with $\theta = 0.75$) networks follow the power law in the form of $p_k \sim k^{-\gamma}$ with $\gamma = 1.14$, $\gamma = 0.9935$ and $\gamma = 1.2323$ respectively. It means that the networks are scale free in which a small portion of vertices have larger degrees, while a large portion of vertices have small degrees. At the first sight, the IMOEX network has the same scale-free behaviour as the CSI163 and S&P468 networks. However, if we take the same level of $\theta$ as it was taken for the CSI163 and S&P468 networks (i.e. $\theta = 0.65 \pm 0.1$) then the IMOEX network will be very sparse and almost all vertices will be isolated.

### Table 3: The max possible number of edges $|e|_{max}$ for $N = 194$ vertices, the existing number of edges $|e|$, the edge density $|e|_{density}$, the average degree $\langle d \rangle$, the average distance $\langle d_{ij} \rangle$, the maximum distance $d_{ij}^{max}$ are presented for different $\theta$ from 0.5 to 1.5 in a step of 0.1, the standard deviation $\sigma$

| $\theta$ | $|e|_{max}$ | $|e|$ | $|e|_{density}$ | $\langle d \rangle$ | $\langle d_{ij} \rangle$ | $d_{ij}^{max}$ | $\sigma$ |
|-----|----------|------|---------------|--------|---------------|---------------|------|
| 0.5  | 18721    | 0    | 0.0000        | 0.0000 | 0.0000        | 0.0000        | 0.0000 |
| 0.6  | 18721    | 1    | 0.0001        | 0.0103 | 0.5458        | 0.5458        | 0.0000 |
| 0.7  | 18721    | 2    | 0.0001        | 0.0206 | 0.5848        | 0.6238        | 0.0552 |
| 0.8  | 18721    | 2    | 0.0001        | 0.0206 | 0.5848        | 0.6238        | 0.0552 |
| 0.9  | 18721    | 4    | 0.0002        | 0.0412 | 0.7201        | 0.8969        | 0.1631 |
| 1    | 18721    | 17   | 0.0009        | 0.1753 | 0.8957        | 0.9994        | 0.1263 |
| 1.1  | 18721    | 80   | 0.0043        | 0.8247 | 1.0235        | 1.0992        | 0.0915 |
| 1.2  | 18721    | 376  | 0.0201        | 3.8763 | 1.1351        | 1.1999        | 0.0756 |
| 1.3  | 18721    | 1996 | 0.1066        | 20.5773| 1.2393        | 1.2998        | 0.0648 |
| 1.4  | 18721    | 14260| 0.7617        | 147.0103| 1.3477        | 1.4000        | 0.0551 |
| 1.5  | 18721    | 18721| 1.0000        | 193.0000| 1.3638        | 1.4748        | 0.0563 |

### 3.2 Average clustering coefficient

The average clustering coefficient (ACC) is the average of all clustering coefficients over all vertices. The clustering coefficient means transitivity for a single vertex, while the average clustering coefficient is an indicator of the transitivity and density for the entire network. Fig. 4 presents the average clustering coefficient for the IMOEX network and its comparison with a random network. Compared to the random network of the same size, the ACC for the IMOEX network is significantly larger than that of the random network. It shows that the IMOEX network is well connected with better transitivity. For comparison, Fig. 4 presents the average clustering coefficients for China and the USA networks. Fig. 4 shows that the IMOEX network does not behave in the similar way: ACC values are always lesser than for CSI163 and S&P468 networks, i.e. the transitivity and density for the IMOEX network are much lower than for China and the USA networks.

### 3.3 Average path length

Fig. 5 shows the average path length for the IMOEX network compared to a random network of the same size. It can be seen that the IMOEX network is significantly different from the random network. For comparison, Fig. 5 also shows the average clustering coefficients for China and the USA networks. Fig. 5 indicates that the IMOEX network behaves differently: the average path length has a peak at $\theta = 1.1$ while the CSI163 and S&P468 networks have peaks at $\theta = 0.6$ and 0.65 respectively. Moreover, the average path length for the IMOEX network is much bigger than for the CSI163 and S&P468 networks for $\theta > 1$. Thus, the China and the USA networks are more con-

![Figure 3: Log-Log degree distributions of IMOEX, CSI163 and S&P468 networks. By using different $\theta$, we can filter out edges with larger distances. This figure shows the log-Log degree distribution of IMOEX (with $\theta = 1.2$), CSI163 (with $\theta = 0.68$) and S&P468 (with $\theta = 0.75$) networks.](image-url)
connected in the sense that economic shocks and important financial and economic news can transfer along networks much quicker.

3.4 Betweenness centrality

Betweenness of a vertex is defined as the number of shortest paths passing through it. Thus, this indicator demonstrates the contribution of each vertex to global connectivity. Averaging over individual betweenness coefficients over all vertices, we can compare the betweenness values for any two vertices. Fig. 6 shows that the peak of the average betweenness for the network is achieved at the threshold value $\theta = 1.35$. It can be seen (Fig. 5) that the CSI163 and S&P468 networks have peaks at $\theta = 0.62$ and $0.81$ respectively. Moreover, the maximum values of the average betweenness for the IMOEX network are much lower than for the CSI163 and S&P468 networks. Thus, the China and the USA networks have a greater number of nodes with higher betweenness centrality that have more control over the network (in comparison with the IMOEX network).

We visualize the stock network of IMOEX with different values of $\theta = 1.1, 1.2$ in Fig. 7. The figures show that the network can be greatly simplified by using small values of $\theta$. Table 3 indicates that the edge density for the IMOEX network increases dramatically from $\theta = 1.1$ to $\theta = 1.4$.

3.5 Components

A component of a network is defined as a subnetwork with connected vertices. If a network has $N$ vertices, then the size of component can vary from 1 (for an isolated node) to the maximum value $N$ (if all vertices are connected). The number of components, the max component size, and the average component size are plotted in Fig. 8, 9 and 10 for the IMOEX, CSI163 and S&P468 networks filtered with different threshold values. The figure shows transitions of the networks from a large number of small isolated components into a connected giant network. At the same time, for the average component size there is a sharp transition to the giant component at $\theta \approx 1.35$.

To examine how industry sectors are linked in the IMOEX network, we list the properties of the network with $\theta = 1.4$ in Table 3.5. The table shows that the Russian industry sectors has approximately the same average degree as well as the average clustering coefficient. However, the values of average betweenness coefficient differ for different sectors assuredly. It should be noted that the same results were obtained for the CSI163 and S&P468 networks in [29].
Table 4: In this table, we list the sector name and the numbers of stocks, the average degree \( \langle d \rangle \), the average clustering coefficient \( \langle C \rangle \), and the average betweenness coefficient \( \langle B \rangle \) for each industry sector of these 194 stocks. The values are calculated from the IMOEX network with \( \theta = 1.4 \).

Table 3.5 presents the top 10 stocks of IMOEX network with the biggest values of degree \( d_i \) and betweenness \( b_i \) arranged in descending order in the top part and the bottom part, respectively. The stock code, company name, industry name, and values of degree \( d_i \), clustering coefficient \( c_i \) and betweenness \( b_i \) are presented. SBERBANK, which is a leading financial company in Russia, has the largest degree of 136 and JSC "OGK-2" is Russia’s largest heat generating company, has the largest betweenness coefficient of 1551.

4 Hierarchical Structure of Stock Networks

Graphs constructed by certain rules may reveal its properties. In this section we create and study hierarchical trees and minimal spanning tree to extract valuable fragments from the graph based on the average distance matrix, which we construct by averaging sliding windows over the study period.

4.1 Hierarchical Tree

Clustering algorithms can be used on average distance matrix to group stocks which prices correlate and to
create four hierarchical trees each using its own normalised distance definition (Fig. 11). Dendogram from Scipy package is used to draw them on a plot from Matplotlib package. In the trees, each leaf node indicates the stock of a particular company with its ticket as a label. The height of where two branches are very likely to be from the same economy is defined as minimum, distance between their centroids: $d_{a,b} = \frac{1}{N_a N_b} \sum_i \sum_j d_{i,j}^a - d_{i,j}^b$, $d_{a,b} = \min(d_{a,i}, d_{b,j})$, $d_{a,b} = d(o_a, o_b)$. We create four hierarchical trees each using its own normalised distance definition (Fig. 11). Dendogram from Scipy package is used to draw them on a plot from Matplotlib package. In the trees, each leaf node indicates the stock of a particular company with its ticket as a label. The height of where two branches are very likely to be from the same economy.

**Table 5:** Top stocks with highest values of degree $d_i$, clustering coefficient $c_i$ and betweenness $b_i$ ranked in descending order of $d_i$ and $b_i$ for IMOEX network with the $\theta = 1.35$. Stock tickers in bold indicate the stocks appear at both top 10 stocks. Cap denotes capitalization in millions of RUB.

### Table 5: Top stocks with highest values of degree $d_i$, clustering coefficient $c_i$ and betweenness $b_i$ ranked in descending order of $d_i$ and $b_i$ for IMOEX network with the $\theta = 1.35$. Stock tickers in bold indicate the stocks appear at both top 10 stocks. Cap denotes capitalization in millions of RUB.

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Name</th>
<th>Industry</th>
<th>$d_i$</th>
<th>$c_i$</th>
<th>$b_i$</th>
<th>Cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBER</td>
<td>SBERBANK</td>
<td>Financials</td>
<td>136</td>
<td>0.5082</td>
<td>1407</td>
<td>4.92 - 10^4</td>
</tr>
<tr>
<td>OGBK</td>
<td>OGK-2</td>
<td>Utilities</td>
<td>136</td>
<td>0.5031</td>
<td>1551</td>
<td>4.25 - 10^4</td>
</tr>
<tr>
<td>RASP</td>
<td>RASPADSKAYA</td>
<td>Energy</td>
<td>132</td>
<td>0.5191</td>
<td>1218</td>
<td>9.60 - 10^4</td>
</tr>
<tr>
<td>RSTI</td>
<td>ROSSETI</td>
<td>Utilities</td>
<td>131</td>
<td>0.5400</td>
<td>719</td>
<td>2.16 - 10^5</td>
</tr>
<tr>
<td>SBERP</td>
<td>SBERBANK OF RUSSIA</td>
<td>Financials</td>
<td>128</td>
<td>0.5519</td>
<td>556</td>
<td>4.90 - 10^6</td>
</tr>
<tr>
<td>FEES</td>
<td>FSK EES</td>
<td>Utilities</td>
<td>128</td>
<td>0.5560</td>
<td>249</td>
<td>2.08 - 10^5</td>
</tr>
<tr>
<td>VTBR</td>
<td>VTB</td>
<td></td>
<td>127</td>
<td>0.5668</td>
<td>131</td>
<td>2.29 - 10^2</td>
</tr>
<tr>
<td>MSNG</td>
<td>MOSENERGO</td>
<td>Utilities</td>
<td>125</td>
<td>0.5735</td>
<td>144</td>
<td>8.96 - 10^4</td>
</tr>
<tr>
<td>BSPB</td>
<td>BANK ST PETERSBURG</td>
<td>Financials</td>
<td>125</td>
<td>0.5472</td>
<td>499</td>
<td>2.80 - 10^4</td>
</tr>
<tr>
<td>RSTP</td>
<td>ROSSETI</td>
<td></td>
<td>122</td>
<td>0.5780</td>
<td>215</td>
<td>2.14 - 10^2</td>
</tr>
</tbody>
</table>

We used all four definitions. The distance algorithm depends on how distance value is calculated, we used all four definitions. The distance $d_{a,b}$ between two clusters $c_a$ and $c_b$ is defined as minimum, maximum and average for all pairs $(o_a, o_b)$ from $c_a$ and $c_b$ respectively or as distance between their centroids: $d_{a,b} = \max(d_{a,i} - d_{b,j})$, $d_{a,b} = \min(d_{a,i} - d_{b,j})$, $d_{a,b} = 1/N_a N_b \sum_i \sum_j d_{i,j}^a - d_{i,j}^b$, $d_{a,b} = d(o_a, o_b)$. We create four hierarchical trees each using its own normalised distance definition (Fig. 11). Dendogram from Scipy package is used to draw them on a plot from Matplotlib package. In the trees, each leaf node indicates the stock of a particular company with its ticket as a label. The height of where two branches connect shows how they correlate. The lower the branches of similar clusters or stocks merge the more positive correlation there is between them. We set color threshold of $0.7 \times \max(d^*)$ to highlight clusters below this value. The companies in same colored branches are very likely to be from the same economy.

**Figure 8:** The number of components for IMOEX, CSI163 and S&P468 networks.

**Figure 9:** The max component size for the IMOEX, CSI163 and S&P468 networks.
Figure 10: The average component size for the IMOEX, CSI163 and S&P468 networks.

sectors which means that hierarchical trees can show economy sectors from price correlation matrix alone.

Figure 11: Hierarchical clustering trees based on average distance for stocks in IMOEX using: (a) complete-linkage, (b) single-linkage, (c) average linkage, (d) centroid linkage

The maximum-based linkage criterion showed better clustering results if we compare four trees generated by the hierarchical clustering algorithms.

4.2 Minimum Spanning Tree
IMOEX network, based on average distance matrix, can be simplified by extracting edges with minimal distance while maintaining graph connectivity. Connected graph with $N$ vertices can be converted into the graph with $N - 1$ edges without loops with minimum total length of edges or minimum spanning tree (MST). This graph is called minimum spanning tree. It brings huge advantages to the study of networks of stocks by reducing noises and simplifying computations. We use Python Scipy minimum spanning tree algorithm on an average distance matrix. Networkx python package is used to get its graphic representation on a Matplotlib plot as shown in Fig. 12.

After the edge filtering process, some nodes are still well-connected, and they may be considered the key stocks in the market. The most connected stocks are from utilities, financials and energy sectors. These results let us extract the most influential companies from the given data. It should be noted that Markowitz optimal portfolio consists of stocks corresponding to leaf nodes of MST. Therefore, we can use MST approach for portfolio optimisation.

Top 10 stocks are shown in the table 4.2. As expected, financials, energy and utilities sectors are dominant in the Russian stock market which is different from the US and Chinese stock markets. Further studies of minimum spanning tree over time may reveal important changes in the core stocks and sectors of the market. Moreover, the result of the algorithm application differs between study periods and this is helpful when studying network topological dynamics.

4.3 Hierarchical tree for a pooled market graph (IMOEX and S&P468)
The obtained results indicate a significant difference in the density of correlation between stock returns and the characteristics of market graphs constructed for the financial markets of the US, China and the Russian Federation.

Such differences can be explained by the relatively less developed financial market of the Russian Federation as a whole, and industry affiliation of companies, greater uncertainty and a high share of risks specific to each of the companies, etc. It is known that a significant share in the total capitalization of the financial market of the Russian Federation falls
on the shares of companies focused on the export of raw materials. It is interesting to check whether industry or country is more important for the correlation dependence. To do this, we apply the hierarchical tree method for a pooled market graph (IMOEX and S&P468). If industry factors would be more important, then Russian stocks should be placed to clusters according to their industry affiliation. If cross-country differences are more important, then clusters should be placed to clusters according to their country affiliation.

The results of applying the hierarchical classification method are presented in the Fig. 13. All Russian companies fall into one cluster (in the figure it is marked in blue at the left side of the dendrogram). S&P500 companies are located in other clusters, which are well explained by their industry affiliation.
5 Conclusion

In this paper, using the daily close prices we calculated the correlation matrices for 194 assets trading on the Russian stock exchange during the study period from 2012 to 2019 for all sliding windows. Based on these correlation matrices we constructed the market networks with stocks as the nodes and correlation relationships as the edges. Our goal was to study the networks properties using the complex network theory from data science perspective. We employed such network filtering methods as threshold, hierarchical tree, minimum spanning tree for simplification of the networks. Then we examine the network properties for the filtered networks. We compare topological properties of the networks constructed for the US and China stock markets with the properties of corresponding networks constructed for the Russian stock market using a dataset spanning over eight years. It turns out that the IMOEX network share similar topological properties in some case, but it also differ from the CSI163 and S&P468 networks in many aspects. The study shows that the IMOEX network share similar topological properties of the Russian stock network has a different from the USA and China dependence on $\theta$, which is the consequence of a much lower correlations between assets traded in the Russian stock market. One the one hand, it may be explained by low liquidity of most stocks traded in IMOEX. On the other hand, even shares with high liquidity have low correlations.

Acknowledgements: This work was supported by the Russian Science Foundation, project 19-18-00199.

References:


