GOSPODARKA NARODOWA

3(319)2024, 56–69
DOI: 10.33119/GN/187564
gnpje.sgh.waw.pl
SGH

The Polish Journal of Economics
Istnieje od / Published since 1931

Artur F. Tomeczek
Collegium of World Economy,
SGH Warsaw School of Economics,
Poland
(*corresponding author*)

Tomasz M. Napiórkowski
Collegium of World Economy,
SGH Warsaw School of Economics,
Poland

PageRank and Regression as a Two-Step Approach to Analysing a Network of Nasdaq Firms During a Recession: Insights from Minimum Spanning Tree Topology

Wykorzystanie PageRank oraz regresji jako dwuetapowej analizy sieci firm Nasdaq w czasie recesji. Wnioski z topologii minimalnego drzewa rozpinającego

Abstract

The presence of focal firms driving entire stock markets has been proven by a series of existing studies that relied on the topological properties of minimum spanning trees. Historically, central firms have been identified primarily based on the degree centrality of nodes. This article proposes an alternative selection method, combining PageRank scores and modularity classes, which does away with the problem of ties in rankings when selecting a specific number of nodes. We use PageRank-based network analysis along with regression analysis to identify focal firms in the Nasdaq-100 index during the three most significant recent recessions in the United States. This approach validates and robustly supports our two-step method, showing that the combination of minimum spanning trees and our selection method explains over 90% of the Nasdaq-100 index's dynamics. The analysis identified significant topological changes during the global financial crisis (with CSCO emerging as the star firm) and the COVID-19 pandemic (exhibiting strong market co-movements).

Keywords:
regression, minimum spanning tree, recession, crisis, stock market

JEL classification codes:
D85, G10, G11, G32, L14

Article history:
submitted: November 28, 2023
revised: March 19, 2024
accepted: April 18, 2024

Streszczenie

Występowanie wpływowych firm oddziałujących na cały rynek akcji zostało potwierdzone w wielu badaniach bazujących na topologii minimalnych drzew rozpinających. Historycznie centralne firmy były identyfikowane przede wszystkim na podstawie centralności stopniowej wierzchołków. Niniejszy artykuł przedstawia alternatywną metodę selekcji, która stanowi połączenie wyników PageRank i klas modularności oraz pozwala wyeliminować remisy w rankingach podczas selekcji określonej liczby wierzchołków. Wykorzystano analizę sieciową na podstawie centralności PageRank połączoną z analizą regresji, aby zidentyfikować wpływowe firmy w indeksie Nasdaq-100 podczas trzech ostatnich recesji w Stanach Zjednoczonych. Wykazano zasadność oraz odporność zaproponowanego dwuetapowego podejścia łączącego minimalne drzewa rozpinające z autorską metodą selekcji,

Słowa kluczowe:
regresja, minimalne drzewo rozpinające, recesja, kryzys, rynek akcji

Kody klasyfikacji JEL:
D85, G10, G11, G32, L14

Historia artykułu:
nadesłany: 28 listopada 2023 r.
poprawiony: 19 marca 2024 r.
zaakceptowany: 18 kwietnia 2024 r.

które tłumaczy ponad 90% dynamiki indeksu Nasdaq-100. Analiza zidentyfikowała istotne zmiany w topologii podczas globalnego kryzysu finansowego (rola CSCO jako firmy centralnej) oraz pandemii COVID-19 (wspólne ruchy akcji).

Introduction

Markets are driven by the price dynamics of focal (central) firms, as shown by minimum spanning tree (MST) analyses of financial data [**Bonanno et al., 2004**; **Mantegna, 1999**; **Onnela et al., 2003**]. Because of the topological properties of an MST, the existing methodology can successfully identify central firms based on their degree centrality. However, using node degree to select a specific number of leading firms creates a problem of ties in rankings, i.e., the inability to establish a pecking order. To address this issue, our study – as the first, to our knowledge – uses PageRank (a network centrality measure) and modularity class (network community detection) to identify key drivers of the Nasdaq index during recent economic and financial crises. To determine if the use of PageRank is appropriate, we adopted a two-stage approach where values of stocks selected as major market drivers are regressed on the value of the studied market index. This novel approach in the network analysis field allows for cross-validity of results obtained by individual methods.

Therefore, the aim of this study is to examine if PageRank can be used to identify influential Nasdaq firms and measure their network position during the three most recent recessions in the United States: the dot-com bubble, the global financial crisis, and the COVID-19 pandemic.

Using network methodology, we attempt to answer the following research questions: [RQ1] Is PageRank a suitable measure to identify central firms? [RQ2] Can central firms be used to explain the performance of a large share (90% or more) of Nasdaq-100 index dynamics? [RQ3] Is there any overlap among the firms identified as central to NDX across the studied economic crises? [RQ4] Which recession had the most severe impact on the network topology?

PageRank is a network centrality measure, originally designed as Google's website ranking algorithm [**Brin, Page, 1998**]. It has since become widely used in network analysis [**Brandes, Erlebach, 2005**; **Gleich, 2015**; **Lü et al., 2016**] to identify influential nodes. The applications of PageRank can also be found in financial network analysis [**Pereira et al., 2019**; **Tomeczek, 2021**; **Yun et al., 2019**].

The text is organised as follows. First, we present a literature review on network analysis applications in financial market studies. Second, network methodology, data, and network analysis results are presented. Third, the econometric modelling procedure is described, and its results are analysed. Lastly, we present the obtained results and conclude the study.

The empirical analysis shows that PageRank is a suitable measure to identify central firms in an MST network as nodes with the highest PageRank score can be used to explain the performance of at least 90% of Nasdaq-100 index dynamics. Interestingly, there is very little overlap among the firms identified as central across the studied economic crises: only CSCO and INTC were identified as focal firms during two of the three crises. The analysis finds significant topology changes during the global financial crisis (with CSCO emerging as the star firm) and the COVID-19 pandemic crisis (exhibiting strong market co-movements). The impact of the dot-com bubble crisis is relatively mild compared to other analysed crises.

Literature review

A minimum spanning tree (MST) of an undirected weighted network (graph) is a subgraph that minimises the sum of its edge weights. The seminal work for the application of MSTs to the stock market data is **Mantegna** [**1999**]. This concept was later expanded in a series of articles exploring the interdependence between stock market indices [**Bonanno et al., 2000**], the impact of frequency on network topology [**Bonanno et al., 2001**], and the differences between MSTs based on historical stock prices and on simulated data [**Bonanno**

et al., 2003]. The results of these studies are summarised in **Bonanno et al.** [**2004**]. The authors conclude that "equity time series are then carrying economic information which can be detected by using specialized filtering procedures. (…) Filtering procedures, like the one we are proposing, are able to undress the signals from the noise and reveal the more relevant information" [**Bonanno et al., 2004**: 365]. The authors also note that "the study of correlation based financial networks is a fruitful method able to filter out economic information from the correlation coefficient matrix of a set of financial time series" [**Bonanno et al., 2004**: 370].

MSTs quickly became of great interest to financial analysis. Their topologies can be used for portfolio selection, akin to the modern portfolio theory, and exhibit shrinkage during a crisis, which is quantified by the normalised tree length [**Onnela, Chakraborti, Kaski, Kertész, 2003**; **Onnela et al., 2003**]. Stocks in a portfolio can be selected based on their location in the network, as nodes on the peripheries of the MST tend to maximise risk diversification [**Danko, Šoltés, 2018**; **Onnela et al., 2003**]. These stocks are represented by leaves (nodes with a degree centrality of one), meaning they have only a single connection to other firms. Over the years, MST networks have been utilised in the analysis of international bond markets [**Dias, 2012**; **Gilmore et al., 2010**], international currency markets [**Wang, Xie, 2016**; **Wang et al., 2012**], international stock markets [**Coelho et al., 2007**; **Kwon, Yang, 2008**; **Wang et al., 2018**], and individual national stock markets [**Huang et al., 2022**; **Jung et al., 2006**; **Tomeczek, 2022**].

Crises and recessions are two of the most persistent problems in economics. The global financial crisis had a severe impact on the international banking network [**Claessens, Van Horen, 2015**; **Legenzova et al., 2019**; **Minoiu, Reyes, 2013**]. Financial MST networks can be used for the quantification of the interdependence of stocks and the identification of influential firms on the market [**Bonanno et al., 2001**; **Huang et al., 2022**; **Tomeczek, 2022**]. MST weight can also be interpreted as a measure of the severity of a financial crisis, while its topology visualises significant changes that can occur in the stock market during severe economic downturns [**Onnela, Chakraborti, Kaski, Kertész, 2003**; **Onnela et al., 2003**]. The topological properties of correlation networks are useful in the identification of influential nodes and the quantification of the co-movements between *markets, which are especially important during a crisis period* [**Coelho et al., 2007**; **Gilmore et al., 2010**].

The primary characteristic of most correlation networks is that they are represented by undirected graphs with no parallel edges. A crucial limitation of such an approach is the inability to show the direction of influence between firms [**Tang et al., 2019**; **Tomeczek, 2021**]. This can be mended by the application of a filtering algorithm to an undirected network. MST algorithms can eliminate excess information from unfiltered correlation networks. The earliest MST algorithm is **Borůvka's** [**1926**], but the two most popular are **Kruskal's** [**1956**] and **Prim's** [**1957**]. These algorithms are widely used in network science and have a long history in graph theory. Their practical applications include computer networks, transportation networks, and the travelling salesman problem [**Graham, Hell, 1985**]. The MST method has also been successfully adapted to study many issues relevant to financial markets.

MSTs based on stock price data have interesting qualities that are crucial to the analysis of financial crises as a sudden drop in the normalised tree length (or total weight of a tree) can be a good indicator of a market disruption. Recent examples have explored the crises in the stock markets of China [**Huang et al., 2022**; **Nie, Song, 2023**; **Zhao et al., 2022**], Japan [**Kanno, 2021**], Morocco [**Bouhlal, Brahim Sedra, 2022**], Pakistan [**Memon et al., 2020**], Poland [**Tomeczek, 2022**], South Africa [**Mbatha, Alovokpinhou, 2022**], and the United Kingdom [**Balcı et al., 2021**].

This study is focused on the MST approach to correlation networks, but other methods can also be used. An alternative to an undirected MST network is a directed Granger causality network. Such networks have mostly been used in the analysis of international interdependence of stock market indices [**Papana et al., 2017**; **Tang et al., 2019**; **Výrost et al., 2015**] or interdependence of individual financial firms [**Wang, Si et al., 2021**; **Wang, Yi et al., 2021**; **Yun et al., 2019**]. Another case is the directed value migration network based on S&P 500 firms [**Siudak, 2022a**, **2022b**], which extends the traditional MST-based approach by including the changes in market capitalisation.

Since 1999, MST-based undirected correlation networks have become a popular and important tool for financial analysis. They are relatively easy to calculate, interpret and visualise. Most crucially, they do not require arbitrary thresholds for edge filtering [Tomeczek, 2022].

Research methods

This section of the paper outlines the procedure selected to test our research hypothesis that PageRank is a suitable measure to identify central firms and that those central firms can be used to explain the performance of a large share (90% or more) of Nasdaq-100 index dynamics. This research hypothesis is a direct answer to RQ1 and RQ2, while the latter two research questions are exploratory in nature.

The MST methodology used in this article is based on a procedure described in numerous previous articles [Coelho et al., 2007; Dias, 2012; Gałązka, 2011; Gilmore et al., 2010; Huang et al., 2022; Jung et al., 2006; Mantegna, 1999; Onnela, Chakraborti, Kaski, Kertész, 2003; Onnela et al., 2003; Tomeczek, 2022]. Every network (graph) consists of a set of n nodes (vertices) and a set of m edges (links). Correlation networks are represented by a symmetric matrix $n \times n$.

For this analysis, three MST networks are constructed, related to the most important recessions that took place in the United States in the 21st century. The periods of recession are identified based on a GDP-based recession indicator [Hamilton, 2022]. The MST networks correspond to the dot-com bubble (DOTCOM, three quarters, 2001/Q1–2001/Q3), global financial crisis (GLOBAL, seven quarters, 2007/Q4–2009/Q2), and the COVID-19 pandemic (COVID, two quarters, 2020/Q1–2020/Q2). The analysis focuses on the Nasdaq-100 constituents (as of November 28, 2022) taken from the official website [Nasdaq, 2022]. The daily adjusted close stock price data are taken from the Yahoo Finance database, accessed on November 28, 2022 [Yahoo Finance, 2022].

The financial time series data must be processed in several steps. Starting with $R_i(t)$, which is the daily log-return defined as:

$$R_i(t) = \ln X_i(t) - \ln X_i(t - \Delta t) \tag{1}$$

where $X_i(t)$ is the daily close stock price of firm i at time t and Δt is the one-day time interval. The vectors of daily log-returns are utilised in Pearson correlation calculations which result in a symmetric matrix with coefficients ranging from –1 to 1. The matrix then needs to be converted so that lower values indicate a stronger correlation:

$$d_{ij} = \sqrt{2(1 - r_{ij})} \tag{2}$$

where d_{ij} is the distance between firms i and j and r_{ij} is the Pearson correlation coefficient between the daily log-returns of firms i and j. This formula results in a symmetric distance matrix ($d_{ij} = d_{ji}$) with values ranging from 0 (perfect correlation) to 2 (perfect anticorrelation). This matrix can be represented by an undirected weighted graph.

For network analysis, we use the open-source Gephi software [Bastian et al., 2009]. The MST algorithm is provided by a plugin based on Kruskal [1956]. Communities in networks are identified using modularity [Blondel et al., 2008; Lambiotte et al., 2014]. Degree centrality, closeness centrality, betweenness centrality, and PageRank are measures of centrality widely used in network analysis [Brandes, 2001; Brandes, Erlebach, 2005; Brin, Page, 1998]. The degree centrality of node v is defined as the number of edges connecting to node v.

Closeness centrality is defined as [Brandes, Erlebach, 2005]:

$$c_C(v) = \frac{1}{\sum_{u \in V} d(v,u)} \tag{3}$$

where $c_C(v)$ is the closeness centrality of node v and $d(v,u)$ is the shortest path (distance) between nodes v and u. Closeness centrality is the reciprocal of the sum of the shortest paths from a given node to all other nodes in the network.

Betweenness centrality is defined as [**Brandes and Erlebach, 2005**]:

$$c_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v) \tag{4}$$

where $c_B(v)$ is the betweenness centrality of node v and $\delta_{st}(v)$ is the ratio of the shortest paths between nodes s and t that pass through node v to all the shortest paths between nodes s and t. The betweenness centrality of a node measures its access to the information flow between other nodes. Both closeness and betweenness indicate the systemic importance of a node. However, high centrality does not guarantee high betweenness, as exemplified by nodes with a degree centrality of one.

PageRank is defined as [**Brin and Page, 1998**]:

$$PR(v) = (1-d) + d\left(\frac{PR(u_1)}{C(u_1)} + \ldots + \frac{PR(u_n)}{C(u_n)}\right) \tag{5}$$

where $PR(v)$ is the PageRank of node v, $PR(u_i)$ is the PageRank of node u_i, $C(u_i)$ is the out-degree of node u_i, node u_i points to node v, and d is the damping factor. The value of d can be set between 0 and 1, but most commonly it is set at 0.85. This is also the value used in this study. PageRank is widely used in modern network analyses to measure the importance of nodes. It can also be defined as the probability that a random walker in an iterative Markov chain will visit a node. The sum of the PageRank values of all the nodes in a network is equal to one (or 100%). PageRank can be used in directed and undirected networks (our network is an example of the latter).

Network analysis will designate key stocks that served as major network nodes for the Nasdaq index during the DOTCOM, GLOBAL and COVID economic and financial crises. The aim of the econometric analysis is to examine what share of this index's variability can be explained by the selected stocks. Despite its issues, such as inflationary bias [**Napiórkowski, 2022**], this share will be measured by the R-squared coefficient of determination calculated for each of the three models. To ensure the validity of the obtained statistic, each estimated model will be tested for the correct specification of its structural form, followed by a series of residuals' tests.

Equations for three models are listed as (6) for DOTCOM, (7) for GLOBAL, and (8) for COVID, where the value of the Nasdaq (NDX) index is modelled as a function of previously established key stocks and an error term (ε_t). The models' parameters (β_i, where i=0, 1 … 6) were estimated with the ordinary least squares method on 184 (6), 441 (7), and 125 (8) observations. A 5% level of statistical significance was selected as a benchmark for further analysis.

$$\ln(NDX_t) = \beta_0 + \beta_1 \ln(ADI_t) + \beta_2 \ln(CSCO_t) + \beta_3 \ln(EBAY_t) + \beta_4 \ln(HON_t) + \\ + \beta_5 \ln(INTC_t) + \beta_6 \ln(PAYX_t) + \varepsilon_t \tag{6}$$

$$\ln(NDX_t) = \beta_0 + \beta_1 \ln(CSCO_t) + \beta_2 \ln(CTAS_t) + \beta_3 \ln(GOOG_t) + \beta_4 \ln(INTC_t) + \\ + \beta_5 \ln(KLAC_t) + \beta_6 \ln(PCAR_t) + \varepsilon_t \tag{7}$$

$$\ln(NDX_t) = \beta_0 + \beta_1 \ln(ADBE_t) + \beta_2 \ln(AMAT_t) + \beta_3 \ln(AMGN_t) + \beta_4 \ln(AVGO_t) + \\ + \beta_5 \ln(CSX_t) + \beta_6 \ln(MSFT_t) + \varepsilon_t \tag{8}$$

To overcome the problem of model under- or over-specification, the number of explanatory variables was established with the following formula: $Modularity\ class + 1$, where the COVID and DOTCOM models each have five modularity classes. The GLOBAL model had four modularity classes due to the presence of a strongly dominant firm (CSCO); hence, the number of independent variables in the GLOBAL model was established with the $Modularity\ class + 2$ formula.

Results

Table 1 presents the network statistics of each MST, where the total weight is the sum of the weights of all edges included in the MST; the average weight (also known as normalised tree length) is the total weight divided by the number of edges; the average path length is the average length of the shortest paths connecting nodes; and the diameter is the length of the longest shortest path. Table 2 shows the six most central firms of each MST, sorted according to the PageRank value. The PageRank ranking list is similar to the degree ranking list, but the PageRank list has fewer ties. The same resolution (4.00) is used in all networks for modularity class calculations.

Table 1. Network statistics

MST	Nodes	Edges	Diameter	Modularity classes	Average path length	Total weight	Average weight
DOTCOM	65	64	15	5	6.572	62.022	0.969
GLOBAL	76	75	10	4	4.247	63.517	0.847
COVID	99	98	16	5	6.902	61.089	0.623

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Table 2. Central firms

Ticker	Stock name	PageRank	Modularity class	Degree	Closeness centrality	Betweenness centrality
DOTCOM firms						
ADI	Analog Devices, Inc.	0.0427	3	6	0.1798	0.2847
HON	Honeywell International Inc.	0.0397	4	6	0.2344	0.6270
EBAY	eBay Inc.	0.0311	2	4	0.1368	0.1215
CSCO	Cisco Systems, Inc.	0.0304	3	4	0.1557	0.1215
PAYX	Paychex, Inc.	0.0289	0	4	0.1803	0.2034
INTC	Intel Corporation	0.0287	2	4	0.1778	0.2287
GLOBAL firms						
CSCO	Cisco Systems, Inc.	0.1321	0	23	0.4286	0.8721
PCAR	PACCAR Inc.	0.0470	3	8	0.3623	0.4591
INTC	Intel Corporation	0.0298	2	5	0.3363	0.3052
KLAC	KLA Corporation	0.0260	2	4	0.2161	0.1041
GOOG	Alphabet Inc.	0.0254	1	4	0.3112	0.1041
CTAS	Cintas Corporation	0.0238	3	4	0.2896	0.2663
COVID firms						
MSFT	Microsoft Corporation	0.0393	1	9	0.2163	0.4374
ADBE	Adobe Inc.	0.0336	1	7	0.1828	0.1382
AMAT	Applied Materials, Inc.	0.0305	0	7	0.2168	0.5500
CSX	CSX Corporation	0.0297	0	6	0.1418	0.1191
AVGO	Broadcom Inc.	0.0284	4	6	0.1461	0.2096
AMGN	Amgen Inc.	0.0243	2	5	0.1341	0.1189

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

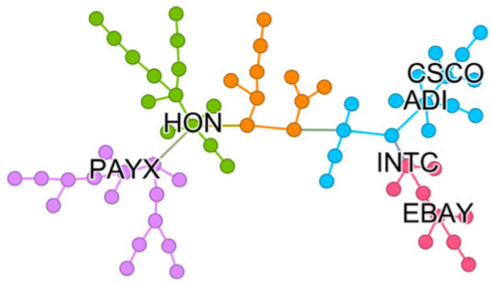
The DOTCOM network (Figure 1) has 65 nodes and 64 edges. According to PageRank, the top nodes are ADI (0.0420), HON (0.0397), EBAY (0.0311), CSCO (0.0304), PAYX (0.0289), and INTC (0.0287). HON has the highest betweenness centrality (0.6270). The average weight (0.969) is the highest of all analysed networks. This network has five communities (modularity classes).

The GLOBAL network (Figure 2) has 76 nodes and 75 edges. Unlike the other two MSTs, this network has a dominant firm at its core. CSCO leads in every measure, including PageRank (0.1321) and betweenness centrality (0.8721), and is at the centre of the star topology with 23 direct connections. Because of the large number of nodes adjacent to CSCO, this network has the smallest diameter (10) and the fewest communities (4). At the same time, it has the second-highest average weight (0.847). Notably, this is the longest recession (seven quarters), which makes CSCO's role even more noteworthy.

The COVID network (Figure 3) has 99 nodes and 98 edges. While COVID has the most nodes (99) and the largest diameter (16), it also has the lowest average weight (0.623), indicating a significant level of price co-movements during this crisis. The top nodes with the highest PageRank score are MSFT (0.0393), ADBE (0.0336), AMAT (0.0305), CSX (0.0297), AVGO (0.0284), and AMGN (0.0243). The node with the highest betweenness centrality is SNPS (0.6815), which is noteworthy as this is the only network where such a node falls outside of the top six in the PageRank score ranking.

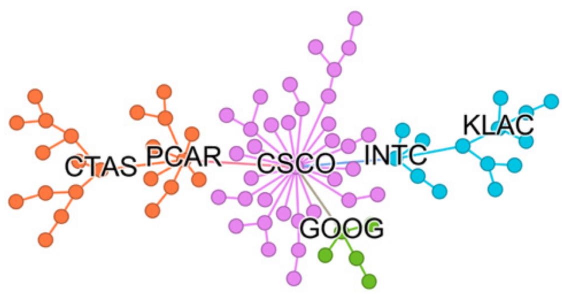
The results of the three MST networks allow us to identify the central firms for each period. Each MST has a different firm as the most central. The GLOBAL crisis, despite being the longest period studied, was the only case where the central firm (CSCO) was so dominant that it formed a star topology. The impact of crises on the network topology is two-fold. First, it can lower the diameter by shortening the paths between the nodes, which was the most evident during the GLOBAL crisis. Second, it lowers the average edge weight by decreasing the distance between the nodes (increasing the correlation coefficients), which was observed during the COVID crisis. As such, we have two different examples of significant topology changes: the emergence of the star firm (GLOBAL network) and the strong co-movements of the entire market (COVID network). The DOTCOM crisis, meanwhile, was relatively unremarkable.

Figure 1. MST DOTCOM network



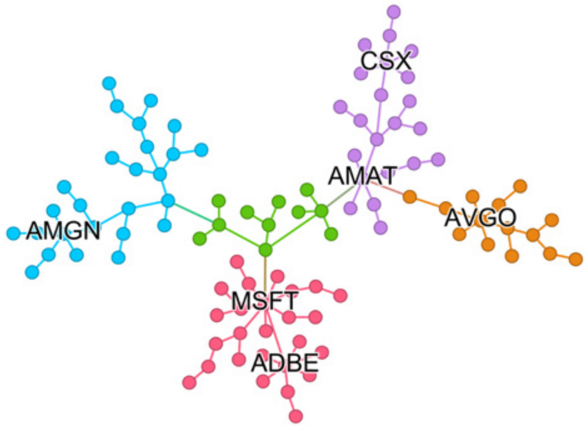
Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Figure 2. MST GLOBAL network



Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Figure 3. MST COVID network



Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Moving to the second step, i.e., econometric modelling, first variables were tested for stationarity using the Augmented Dickey-Fuller test (Table 3). For the DOTCOM and GLOBAL models, all variables were stationary after first differencing – *d(1)*. In the case of the COVID model, the NDX, AMGN, and MSFT variables reached stationarity after second-order differencing. Initial estimation using *d(2)* variables yielded models that were not econometrically sound. Hence, a decision was made to estimate the COVID model's parameters using *d(1)* variables. To accommodate the risk of spurious regression, residuals of all the models were checked for the presence of unit root [Wooldridge, 2009].

Table 3. Results of the Augmented Dickey-Fuller tests for used variables (p-values)

Variable	d(0)	d(1)	d(2)
DOTCOM			
log_NDX	0.3614	1.364e-26	–
log_ADI	0.3313	1.307e-21	–
log_CSCO	0.5879	1.272e-39	–
log_EBAY	0.9506	1.811e-30	–
log_HON	0.7002	1.157e-20	–
log_INTC	0.2922	2.692e-25	–
log_PAYX	0.302	3.884e-43	–
GLOBAL			
log_NDX	0.947	1.581e-60	–
log_CSCO	0.7036	1.131e-60	–
log_CTAS	0.1312	9.206e-82	-
log_GOOG	0.7804	3.098e-81	-
log_INTC	0.7672	4.139e-25	-
log_KLAC	0.8771	2.142e-83	-
log_PCAR	0.2296	5.037e-51	-
COVID			
log_NDX	0.7818	0.1193	4.144e-09
log_ADBE	0.8843	0.0041	–
log_AMAT	0.9179	1.134e-05	–
log_AMGN	0.5091	0.1186	1.784e-14
log_AVGO	0.8941	0.0001	–
log_CSX	0.7225	0.0487	–
log_MSFT	0.5663	0.1646	2.394e-09

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

The results of structural tests (Table 4) indicate that the GLOBAL model needs to include the squares of all explanatory variables, while the squares of AMAT and AMGN were added to the COVID model.

Table 4. Structural form tests (p-values)

Test/Model	DOTCOM	GLOBAL	COVID
Non-linearity test (squares)	0.276	0.310	0.000
RESET (squares and cubes) test for specification	0.087	0.000	0.014

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

The residuals for each model have been checked for heteroskedasticity (with three tests: White, Breusch-Pagan, and Breusch-Pagan Koenker robust variant), autocorrelation (Breusch-Godfrey LM), normal distribution (Doornik-Hansen, Shapiro-Wilk, Lilliefors and Jarque-Bera), and stationarity (Augmented Dickey-Fuller). Where possible we used multiple tests to account for individual issues, e.g., normal distribution requirement for the Breusch-Pagan test [Pindyck, Rubinfeld, 1998] or the large-sample bias of certain distribution tests [Filed, 2013; Mooi, Sarstedt, 2011]. Heteroskedasticity (tests 1–3) and autocorrelation (test 4) issues (Table 5) were accounted for with the Heteroskedasticity and Autocorrelation Consistent (HAC) estimators correction [Andrews, 1991; Zeileis, 2004]. The residuals do not follow normal distribution only in the GLOBAL model (testes 5–8), which may be connected with the presence of a strongly dominant firm (CSCO) that acts as an outlier. However, given the number of observations (441), this result is not expected to impact the validity of the obtained results, especially the interpretation of the coefficient of determination, which we are chiefly interested in. Lastly, the residuals for all models were stationary (test 9), eliminating the possibility of spurious regression.

Table 5. Residuals' tests (p-values)

Test number	Test/Model	DOTCOM	GLOBAL	COVID
1	White's test for heteroskedasticity	0.006	2.11E-21	0.004
2	Breusch-Pagan test for heteroskedasticity	0.035	4.64E-27	0.252
3	Breusch-Pagan test for heteroskedasticity (robust variant)	0.066	4.29E-14	0.203
4	LM test for autocorrelation up to order 5	0.010	0.004	0.211
5	Doornik-Hansen test	0.406	2.30E-07	0.818
6	Shapiro-Wilk W	0.287	6.50E-05	0.640
7	Lilliefors test (~)	0.220	0.000	0.330
8	Jarque-Bera test	0.572	4.41E-09	0.950
9	Augmented Dickey-Fuller	1.37E-05	8.20E-59	5.57E-09

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

The P-values for the F statistics for all models allow for rejecting the test's null hypothesis that all estimated coefficients are equal to each other and zero (Table 6).

Table 6. P-value for F-stat. and R-squared

Statistic/Model	DOTCOM	GLOBAL	COVID
R-squared	0.9087	0.9200	0.9771
P-value (F)	3.31E-89	4.6E-226	2.59E-94

Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Having established the models' statistical soundness, it is possible to comment on the obtained R2 values, which represent the share of changes in NDX explained by the right hand of the equation. Stocks selected through the network analysis proved to be very strong nodes in the studied networks as changes in selected stocks explain upwards of 90% of the Nasdaq-100 index's dynamics. Their coefficients were found to be statistically significant at least at the 10% level of statistical significance (Table 7).

Table 7. Significance of estimated coefficients across three models

DOTCOM			GLOBAL			COVID		
Coefficient	Value	Sig.	Coefficient	Value	Sig.	Coefficient	Value	Sig.
const	−0.001	**	const	0.001		const	0.001	
d_l_ADI	0.148	***	d_l_CSCO	0.219	***	d_l_ADBE	0.132	***
d_l_CSCO	0.229	***	d_l_CTAS	0.129	***	d_l_AMAT	0.087	***
d_l_EBAY	0.182	***	d_l_GOOG	0.231	***	d_l_AMGN	0.116	***
d_l_HON	0.066	**	d_l_INTC	0.178	***	d_l_AVGO	0.052	*
d_l_INTC	0.187	***	d_l_KLAC	0.047	**	d_l_CSX	0.097	***
d_l_PAYX	0.134	***	d_l_PCAR	0.084	***	d_l_MSFT	0.397	***
			sq_d_l_CSCO	0.504	^	sq_d_l_AMAT	0.332	***
			sq_d_l_CTAS	0.708	*	sq_d_l_AMGN	−1.452	***
			sq_d_l_GOOG	−0.728	***			
			sq_d_l_INTC	0.620	*			
			sq_d_l_KLAC	−0.511	**			
			sq_d_l_PCAR	−0.373	**			

Where: "d_l_x" – differenced natural logarithm of variable x; "sq_x" – square of variable x; "***" – statistical significance at alpha = 1%, "**" at alpha = 5% and "*" at alpha = 10%.
^ – coefficient of the sq_d_l_CSCO was not found to be statistically significant after applying HAC (p-value = 0.1549); before its p-value = 0.0937.
Source: Authors' own elaboration based on data from Yahoo Finance [2022].

Discussion

Our results have practical implications for future financial networks. Starting with financial analysis, the identification of influential firms is beneficial for macroprudential policy monitoring and risk assessment. Instead of monitoring and calculating risk for a large number of firms, identifying and then analysing only a few key movers potentially lowers monitoring and assessment costs and leads to higher profits.

Further savings can be achieved in the process of Exchange-Traded Fund (ETF) creation. This especially applies to market- and industry-wide ETFs, i.e., instruments representing markets (e.g., ETFs for S&P500 or Nasdaq) or industries (e.g., ETFs for high-tech) and comprising many firms. In the case of such instruments, physical (1-for-1) replication is expensive or impossible. Instead of investing in a wide range of firms, ETF creators can simply identify key movers and invest only in those stocks. Such an approach will lead to lower costs (lower total expense ratio, TER). A higher tracking difference is a noted potential disadvantage of such an approach. However, our research shows that proper selection of key firms leads to a high degree of market representation (high R-squared values). Admittedly, our research shows that key firms change period-to-period, but a low-cost, repeatable two-step network regression analysis presented in this study should decrease this disadvantage. We focus on nodes with the highest PageRank score, regardless of their neighbouring nodes.

MSTs can be powerful tools for financial analysis. There is some ambiguity as to what exactly constitutes a central firm. The most common way of identifying central firms is degree centrality, which measures the local influence of nodes. Alternatively, shortest path-based centralities can be used to measure systemic importance [Tomeczek, 2022]. As mentioned before, this study uses PageRank. Node degree centrality and PageRank scores are usually strongly correlated, but PageRank has an advantage in tie-breaking in financial networks. While PageRank ties are possible, they only become a problem in very large complex networks, and the number of nodes in correlation modelling is limited to the number of firms listed on the stock market.

The use of modularity classes is beneficial because they can be utilised to select the number of top firms for analysis, and they also provide visual clarity to the graphs.

While the emergence of a star firm during a severe financial crisis is to be expected, CSCO might have seemed like an unlikely candidate for this role. Historically, MST studies of the entire American stock market have identified General Electric as the star firm [**Bonanno et al., 2004**; **Mantegna, 1999**; **Onnela et al., 2003**]. Due to data availability requirements, most MST analyses focus on daily log-returns. As shown by **Bonanno et al.** [**2001**], data frequency has an impact on network topology. A future study could explore recent crises and recessions in the intraday context. The intraday data can allow researchers to better pinpoint and quantify short-term market disruptions, for example those caused by formal communications among monetary bodies or by individual pandemic lockdowns. One of the many possible factors to consider when creating a market portfolio is the CSR performance of firms [**Okafor et al., 2021**; **Tsai, Wu, 2022**; **Valls Martínez et al., 2022**]. The exploration of the relationship between CSR policies and the MST network position might be a worthwhile research endeavour.

Conclusions

The aim of our study was to examine if PageRank values can be used to identify influential Nasdaq firms and measure their network position during the three most recent recessions in the United States. To achieve this goal, we developed a two-step self-testing procedure. In the first step, focal firms for the NDX index were identified using network analysis. In the second step, these identified firms were used as independent variables in an econometric regression analysis, with the NDX index as the dependent variable. The method and its results can be beneficial for financial analysis, macroprudential policy monitoring and risk assessment, and portfolio creation (e.g., construction of ETFs). While both these measures tend to be highly correlated, the practical benefit of using PageRank over degree centrality in correlation-based financial networks is its advantage in tie-breaking.

Our empirics have shown that PageRank is a suitable measure to identify central firms [RQ1] as such firms can be used to explain the performance of at least 90% of Nasdaq-100 index dynamics [RQ2]. Hence, our research hypothesis has been confirmed. Interestingly, there is very little overlap in terms of which firms were found as central for NDX across studied economic crises [RQ3]. Only two firms (CSCO and INTC) were found to be focal firms during two (DOTCOM and GLOBAL) of the three studied events. Our analysis identified significant topology changes during the GLOBAL crisis (the emergence of CSCO as the star firm) and the COVID crisis (strong market co-movements). On the other hand, the impact of the DOTCOM crisis was mild in comparison [RQ4].

Our study, like any other, is not without its limitations. First, we studied only times of crisis, which have their own specificity in investors' behaviour. As much as this was planned, further studies employing the PageRank-Regression two-step approach shown in this text should also consider periods of booms and stagnations (i.e., market horizontal movements). Second, since the aim of our study was to examine if a specific method can be used, we did not explain the specificity of focal firms, i.e., their match to specific economic crises. Therefore, as this is beyond the scope of this research, we recommend further investigation of this matching. Third and lastly, we treat each of the three crises homogenously, and therefore disregard potential differences between them. Our results, especially the statistics of the GLOBAL model, show that identifying and adding individual, crisis-specific characteristics is another area for further exploration.

References

Andrews D.W.K. [1991], Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59 (3): 817–858, https://doi.org/10.2307/2938229.

Balcı M.A., Akgüller Ö., Can Güzel S. [2021], Hierarchies in communities of UK stock market from the perspective of Brexit, *Journal of Applied Statistics*, 48 (13–15): 2607–2625, https://doi.org/10.1080/02664763.2020.1796942.

Bastian M., Heymann S., Jacomy M. [2009], Gephi: An open source software for exploring and manipulating networks, *Proceedings of the International AAAI Conference on Web and Social Media*, 3 (1): 361–362, https://doi.org/10.1609/icwsm.v3i1.13937.

Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E. [2008], Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10): P10008, https://doi.org/10.1088/1742-5468/2008/10/P10008.

Bonanno G., Caldarelli G., Lillo F., Mantegna R.N. [2003], Topology of correlation-based minimal spanning trees in real and model markets, *Physical Review E*, 68 (4): 046130, https://doi.org/10.1103/PhysRevE.68.046130.

Bonanno G., Caldarelli G., Lillo F., Miccichè S., Vandewalle N., Mantegna R.N. [2004], Networks of equities in financial markets, *The European Physical Journal B – Condensed Matter*, 38 (2): 363–371, https://doi.org/10.1140/epjb/e2004-00129-6.

Bonanno G., Lillo F., Mantegna R.N. [2001], High-frequency cross-correlation in a set of stocks, *Quantitative Finance*, 1 (1): 96–104, https://doi.org/10.1080/713665554.

Bonanno G., Vandewalle N., Mantegna R.N. [2000], Taxonomy of stock market indices, *Physical Review E*, 62 (6): R7615–R7618, https://doi.org/10.1103/PhysRevE.62. R7615.

Borůvka O. [1926], O jistém problému minimálním, *Práce Moravské Přírodovědecké Společnosti*, 3: 37–58.

Bouhlal F., Brahim Sedra M. [2022], The impact of COVID-19 on the topological properties of the Moroccan stock market network, *Investment Management and Financial Innovations*, 19 (2): 238–249, https://doi.org/10.21511/imfi.19 (2).2022.21

Brandes U. [2001], A faster algorithm for betweenness centrality, *The Journal of Mathematical Sociology*, 25 (2): 163–177, https://doi.org/10.1080/0022250X.2001.9990249.

Brandes U., Erlebach T. (eds.) [2005], *Network analysis: Methodological foundations*, https://doi.org/10.1007/b106453.

Brin S., Page L. [1998], The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, 30 (1–7): 107–117, https://doi.org/10.1016/S0169-7552(98)00110-X.

Claessens S., Van Horen N. [2015], The impact of the global financial crisis on banking globalization, *IMF Economic Review*, 63 (4): 868–918, https://doi.org/10.1057/imfer.2015.38.

Coelho R., Gilmore C.G., Lucey B., Richmond P., Hutzler S. [2007], The evolution of interdependence in world equity markets – Evidence from minimum spanning trees, *Physica A: Statistical Mechanics and Its Applications*, 376: 455–466, https://doi.org/10.1016/j.physa.2006.10.045.

Danko J., Šoltés V. [2018], Portfolio creation using graph characteristics, *Investment Management and Financial Innovations*, 15 (1): 180–189, https://doi.org/10.21511/imfi.15 (1).2018.16.

Dias J. [2012], Sovereign debt crisis in the European Union: A minimum spanning tree approach, *Physica A: Statistical Mechanics and Its Applications*, 391 (5): 2046–2055, https://doi.org/10.1016/j.physa.2011.11.004.

Filed A. [2013], *Discovering statistics using IBM SPSS Statistics*, SAGE Publications, London.

Gałązka M. [2011], Characteristics of the Polish Stock Market Correlations, *International Review of Financial Analysis*, 20 (1): 1–5, https://doi.org/10.1016/j.irfa.2010.11.002.

Gilmore C.G., Lucey B.M., Boscia M.W. [2010], Comovements in government bond markets: A minimum spanning tree analysis, *Physica A: Statistical Mechanics and Its Applications*, 389 (21): 4875–4886, https://doi.org/10.1016/j.physa.2010.06.057.

Gleich D.F. [2015], PageRank beyond the web, *SIAM Review*, 57 (3): 321–363, https://doi.org/10.1137/140976649.

Graham R.L., Hell, P. [1985], On the history of the minimum spanning tree problem, *IEEE Annals of the History of Computing*, 7 (1): 43–57, https://doi.org/10.1109/MAHC.1985.10011.

Hamilton J.D. [2022], *Dates of U.S. recessions as inferred by GDP-based recession indicator*, https://fred.stlouisfed.org/series/JHDUSRGDPBR (accessed on 4.12.2022).

Huang C., Zhao X., Su R., Yang X., Yang X. [2022], Dynamic network topology and market performance: A case of the Chinese stock market, *International Journal of Finance & Economics*, 27 (2): 1962–1978, https://doi.org/10.1002/ijfe.2253.

Jung W.-S., Chae S., Yang J.-S., Moon H.-T. [2006], Characteristics of the Korean stock market correlations, *Physica A: Statistical Mechanics and Its Applications*, 361 (1): 263–271, https://doi.org/10.1016/j.physa.2005.06.081.

Kanno M. [2021], Risk contagion of COVID-19 in Japanese firms: A network approach, *Research in International Business and Finance*, 58: 101491, https://doi.org/10.1016/j.ribaf.2021.101491.

Kruskal J. B. [1956], On the shortest spanning subtree of a graph and the traveling salesman problem, *Proceedings of the American Mathematical Society*, 7 (1): 48–48, https://doi.org/10.1090/S0002-9939-1956-0078686-7.

Kwon O., Yang J.-S. [2008], Information flow between stock indices, *EPL (Europhysics Letters)*, 82 (6): 68003, https://doi.org/10.1209/0295-5075/82/68003.

Lambiotte R., Delvenne J.-C., Barahona M. [2014], Random walks, Markov processes and the multiscale modular organization of complex networks, *IEEE Transactions on Network Science and Engineering*, 1 (2): 76–90, https://doi.org/10.1109/TNSE.2015.2391998.

Legenzova R., Gaigalienė A., Jurakovaitė O. [2019], Evaluation of the post-crisis EU banking network connectedness in the global context, *Oeconomia Copernicana*, 10 (1): 37–53, https://doi.org/10.24136/oc.2019.002.

Lü L., Chen D., Ren X.-L., Zhang Q.-M., Zhang Y.-C., Zhou T. [2016], Vital nodes identification in complex networks, *Physics Reports*, 650: 1–63, https://doi.org/10.1016/j.physrep.2016.06.007.

Mantegna R. N. [1999], Hierarchical structure in financial markets, *The European Physical Journal B*, 11 (1): 193–197, https://doi.org/10.1007/s100510050929.

Mbatha V. M., Alovokpinhou S. A. [2022], The structure of the South African stock market network during COVID-19 hard lockdown, *Physica A: Statistical Mechanics and Its Applications*, 590: 126770, https://doi.org/10.1016/j.physa.2021.126770.

Memon B. A., Yao H., Tahir R. [2020], General election effect on the network topology of Pakistan's stock market: network-based study of a political event, *Financial Innovation*, 6 (1): 2, https://doi.org/10.1186/s40854-019-0165-x.

Minoiu C., Reyes J. A. [2013], A network analysis of global banking: 1978–2010, Journal of Financial Stability, 9 (2): 168– 184, https://doi.org/10.1016/j.jfs.2013.03.001.

Mooi E., Sarstedt M. [2011], *A concise guide to market research. The process, data, and methods using IBM SPSS Statistics*, Springer, Berlin–Heidelberg.

Napiórkowski T. [2022], *Praktyczna analiza danych za pomocą metod ilościowych*, Oficyna Wydawnicza SGH, Warszawa.

Nasdaq [2022], *Quotes for Nasdaq-100 index*, https://www.nasdaq.com (accessed on 28.11.2022).

Nie C.-X., Song F.-T. [2023], Stable versus fragile community structures in the correlation dynamics of Chinese industry indices, *Chaos, Solitons & Fractals*, 167: 113044, https://doi.org/10.1016/j.chaos.2022.113044.

Okafor A., Adeleye B. N., Adusei M. [2021], Corporate social responsibility and financial performance: Evidence from U. S tech firms, *Journal of Cleaner Production*, 292: 126078, https://doi.org/10.1016/j.jclepro.2021.126078.

Onnela J.-P., Chakraborti A., Kaski K., Kertész J. [2003], Dynamic asset trees and Black Monday, *Physica A: Statistical Mechanics and Its Applications*, 324 (1–2): 247–252, https://doi.org/10.1016/S0378-4371(02)01882-4.

Onnela J.-P., Chakraborti A., Kaski K., Kertész J., Kanto A. [2003], Dynamics of market correlations: Taxonomy and portfolio analysis, *Physical Review E*, 68 (5): 056110, https://doi.org/10.1103/PhysRevE.68.056110.

Papana A., Kyrtsou C., Kugiumtzis D., Diks C. [2017], Financial networks based on Granger causality: A case study, *Physica A: Statistical Mechanics and Its Applications*, 482: 65–73, https://doi.org/10.1016/j.physa.2017.04.046.

Pereira E. J. D. A. L., Ferreira P. J. S., Silva M. F. D., Miranda J. G. V., Pereira H. B. B. [2019], Multiscale network for 20 stock markets using DCCA, *Physica A: Statistical Mechanics and Its Applications*, 529: 121542, https://doi.org/10.1016/j.physa.2019.121542.

Pindyck R. S., Rubinfeld D. L. [1998], *Econometric models and econometric forecasts*, Irwin/McGraw-Hill, New York.

Prim R. C. [1957], Shortest connection networks and some generalizations, *Bell System Technical Journal*, 36 (6): 1389–1401, https://doi.org/10.1002/j.1538-7305.1957.tb01515.x.

Siudak D. [2022a], A network analysis of the value migration process on the financial market. The effect of value migration network structure on stock returns, *Expert Systems with Applications*, 191: 116129, https://doi.org/10.1016/j.eswa.2021.116129.

Siudak D. [2022b], The effect of self-organizing map architecture based on the value migration network centrality measures on stock return, Evidence from the US market, *PLOS ONE*, 17 (11): e0276567, https://doi.org/10.1371/journal.pone.0276567.

Tang Y., Xiong J. J., Luo Y., Zhang Y.-C. [2019], How do the global stock markets influence one another? Evidence from finance big data and Granger causality directed network, *International Journal of Electronic Commerce*, 23 (1): 85–109, https://doi.org/10. 1080/10864415.2018.1512283.

Tomeczek A. F. [2021], A financial network analysis of the equity linkages in Poland, *Research Papers of Wrocław University of Economics*, 65 (4): 129–143, https://doi.org/10.15611/pn.2021.4.08.

Tomeczek A. F. [2022], A minimum spanning tree analysis of the Polish stock market, *Journal of Economics and Management*, 44: 420–445, https://doi.org/10.22367/jem.2022.44.17.

Tsai H.-J., Wu Y. [2022], Changes in corporate social responsibility and stock performance, *Journal of Business Ethics*, 178 (3): 735–755, https://doi.org/10.1007/s10551-021-04772-w.

Valls Martínez M. D. C., Soriano Román R., Mart′ín-Cervantes P. A. [2022], Should risk-averse investors target the portfolios of socially responsible companies?, *Oeconomia Copernicana*, 13 (2): 439–474, https://doi.org/10.24136/oc.2022.014.

Výrost T., Lyócsa Š., Baumöhl E. [2015], Granger causality stock market networks: Temporal proximity and preferential attachment, *Physica A: Statistical Mechanics and Its Applications*, 427: 262–276, https://doi.org/10.1016/j.physa.2015.02.017.

Wang G.-J., Si H.-B., Chen Y.-Y., Xie C., Chevallier J. [2021], Time domain and frequency domain Granger causality networks: Application to China's financial institutions, *Finance Research Letters*, 39: 101662, https://doi.org/10.1016/j.frl.2020.101662.

Wang G.-J., Xie C. [2016], Tail dependence structure of the foreign exchange market: A network view, *Expert Systems with Applications*, 46: 164–179, https://doi.org/10.1016/j.eswa.2015.10.037.

Wang G.-J., Xie C., Han F., Sun B. [2012], 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*, 391 (16): 4136–4146, https://doi.org/10.1016/j.physa.2012.03.036.

Wang G.-J., Xie C., Stanley H. E. [2018], Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks, *Computational Economics*, 51 (3): 607–635, https://doi.org/10.1007/s10614-016-9627-7.

Wang G.-J., Yi S., Xie C., Stanley H. E. [2021], Multilayer information spillover networks: Measuring interconnectedness of financial institutions, *Quantitative Finance*, 21 (7): 1163–1185, https://doi.org/10.1080/14697688.2020.1831047.

Wooldridge J. M. [2009], *Introduction to econometrics. A modern approach*, CENGAGE Learning, Mason, South-Western.

Yahoo Finance [2022], *Yahoo Finance database*, https://finance.yahoo.com (accessed on 28.11.2022).

Yun T.-S., Jeong D., Park S. [2019], "Too central to fail" systemic risk measure using PageRank algorithm, *Journal of Economic Behavior & Organization*, 162: 251–272, https://doi.org/10.1016/j.jebo.2018.12.021.

Zeileis A. [2004], Econometric Computing with HC and HAC Covariance Matrix Estimators, *Journal of Statistical Software*, 11 (10): 1–17, https://doi.org/10.18637/jss.v011.i10.

Zhao B., Yang W., Wen J., Zhang W. [2022], The Financial Market in China under the COVID-19, *Emerging Markets Finance and Trade*, 58 (13): 3726–3738, https://doi.org/10.1080/1540496X.2022.2070472.