



QAP Analysis of Company Co-mention Network

S. P. Sidorov, A. R. Faizliev^(✉), V. A. Balash, A. A. Gudkov,
A. Z. Chekmareva, M. Levshunov, and S. V. Mironov

Saratov State University, Saratov, Russian Federation
faizlievar1983@mail.ru
<http://www.sgu.ru>

Abstract. In our research we form a network called company co-mention network. News analytics data have been employed to collect the companies co-mentioning. Each company acquires a certain value based on the amount of news in which the company was mentioned. A matrix containing the number of co-mentioning news between pairs of companies has been created for network analysis. Each company is presented as a node, and news mentioning two companies establishes a link between them. The network is constructed quite similarly to social networks or co-citation networks. The networked map of the companies is used to visualize the dependence structure of the economy by identifying groups of companies that are more central than others. The analysis carried out in the context of sectors of economy and territorial affiliation made it possible to identify key companies and to explore the similarity of the power law of vertices within sectors. QAP analysis between the co-mention network and the sector affiliation network was carried out to examine the ability of the sector affiliation network to predict the structure of the co-mention network.

Keywords: Network analysis · News analytics · Degree distribution
SNA metrics · QAP analysis

1 Introduction

One of the examples of big data that has emerged relatively recently is information flows generated by news agencies, enterprises, organizations, social networks, etc. Such real time information flows feature a huge amount of news sources, unstructurability, a large amount of both sources and objects of news, and high frequency (thousands of news items per second). Therefore, one of the problems of information systems that process such data is to aggregate them into one or several indicators that allow to describe the intensity, stability, changing news flow structure, identify the most discussed news subjects, and so on. Note that

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

the characteristics and features of the news flow are currently not sufficiently studied, and the methods and algorithms for processing such data are not fully developed.

In this paper we use some methods of the graph theory and corresponding algorithms to examine news flow data. A social network analysis (SNA) approach allows to investigate (explain) structures in systems based on the relations among the system's components (nodes) rather than the attributes of individual cases [6]. Methods of social network analysis can be used to analyze the structure of relations in an organization [7, 8] or examine relationships among social entities [44]. The basic concepts of SNA are node and link, where a node refers to a unit (individual, object, item) and a link indicates a relationship between nodes.

News flow data can be easily converted into network data, since a company can be presented as a node and the act of mentioning two companies in one news item can be visualized as a link between them. Moreover, using this co-mention network we can study basic properties of networks, such as centrality and tie strength. Highly mentioned companies may be treated as key companies which are more significant to economy than other companies. Central nodes in this co-mention network are key companies in co-mention analysis. Moreover, we can treat the amount of co-mentions as the link weight. In this case, highly mentioned companies will have a higher value of link weight.

The paper [35] used various types of social network analysis metrics and citation indices to find key companies in the network. The focus of this paper is the analysis of different parts of co-mention network (or subnetworks) rather than the co-mention network as the whole. In our study, we use two ways to construct subnetworks:

- subnetworks represent different sectors of economy, such as products, resources, services, IT sectors.
- each subnetwork consists of companies with the same stock exchange affiliation.

The concept of link weight is crucial to our analysis. Through examination frequencies and other SNA metrics of each company within a subnet, co-mention analysis of companies can identify the key companies of the subnetwork. For co-mentioned companies, the weight of links may be useful for identification of the most frequently co-mentioned pairs of companies, as well as those who are poorly co-mentioned but useful in terms of providing diversity to the sector or subnetwork where they belong. We suppose that companies with weak links to other companies may play a valuable role in expanding the diversity of economic information within their subnetwork or clique.

In our research we would like to find answers to the following questions:

- Do the frequency, degree centrality, closeness centrality, betweenness centrality and eigenvector centrality of companies vary within subnetworks (clusters)? Within each group (or subnetwork) of companies, our research is going to find those which are more central than others. It is assumed that the more central a company is within a network, the more influential and more important the news about it must be.

- Does the type of the degree distribution vary within clusters? What type of the functional form has the clustering-degree relation for the clusters?
- Does the analysis of company co-mention network identify groups? Our hypothesis is that the network analysis of the company co-mention network reproduces the sector structure of the economy. The company co-mention network may be very sparse, but the graph of companies is expected to show which of them are often mentioned together or which belong to the same sector of economy. Each group or cluster of companies might be associated with a particular sector, for example, products, resources, services or information technology. Another hypothesis we would like to examine is that the structure of the company co-mention network reflects the clusterization of companies based on their stock exchange affiliation.

To provide answers to the first and the second questions we will use well-known SNA metrics and methods. The answer to the third question will rely on the quadratic assignment procedure (QAP) which was proposed and developed in [13, 18, 21, 27]. Since then, QAP has been widely used in social network analysis (see e.g. [5, 9, 10, 17, 34, 42], among many others). QAP is a peculiar type of permutation test which leaves the dyadic data structure under the permutations unchanged.

Our research questions are close to ones from the paper [20] which investigated whether the patterns of author co-citation can describe the structure of the field of communication.

Our study uses the data delivered by the news analytics providers. In our opinion, the data are quite typical and can be used for processing and analysis. Using this data, papers [36–39] examined different news flow characteristics, such as intensity, stability, volatility, long-term memory, fractality, etc.

2 Data

A huge amount of economic and financial news are generated in real time by news agencies, stocks exchanges, companies, magazines, papers, blogs and so on. Different companies are named in these news items. We accomplish company co-mention analysis in the five steps:

1. we assemble all economic and financial news items produced during one month of 2015 (February of 2015);
2. we collect a list of mentioned companies for each news item;
3. we calculate a weighted co-mention count for each pair of co-mentioned companies based on all set of available news items;
4. we produce a symmetric co-mention matrix using these weighted co-mention counts;
5. then we analyze this co-mention matrix statistically, and the results are visualized and interpreted.

Step 1 and 2 operations are executed by providers of news analytics. In our research, we manipulate these already processed data.

2.1 News Analytics Data

Two of the biggest providers of news analytics are Thompson Reuters and Raven Pack. They gather news items from diverse sources in real time. They collect data from different sources including news agencies and social media (blogs, social networks, etc.). They also use so-called pre-news, i.e. SEC reports, court documents, reports of various government agencies, business resources, company reports, announcements, industrial and macroeconomic statistics. Then news analytics providers handle preliminary analysis of each news item in real time. Using AI algorithms, they calculate news-related expectations (sentiments) based on the current market situation. As a rule, providers of news analytics provide to subscribers in real time the following attributes for each news item: time stamp, company name, company id, relevance of the news, event category, event sentiment, novelty of the news, novelty id, composite sentiment score of the news, among others. Subscribers of news analytics data may develop and exploit quantitative models or trading strategies based on both the news analytics data and financial time series data. The survey of applications for news analytics tools can be found in books [28,29].

2.2 Generating Company Co-mention Network

Methodology. Company co-mention network is formed based on co-mention; it means that a company has connection with those companies that have been mentioned in a news item together. A company co-mention network is a set of companies which have connections in pair to represent their co-mention relationship. Two companies are linked if there has been published a publicly available news item mentioning both of them. In such type of network, a company will be represented by 'node' or 'vertex' and the connection will be an 'edge'. Thus, we represent the company co-mention network as an undirected weighted graph. In some sense, company co-mention network can be considered as social network. Based on the available data of news analytics, we built an adjacency matrix which represents the relationship between companies in line with the approach described in [35].

Network. We deal with all the financial and economic news items released during one month period from February 1, 2015 to February 28, 2015 (i.e. 20 trading days). We eliminated all the news on the imbalance of supply and demand before both the opening and the closing of trading time of different stock exchanges. News of such type may amount to several hundreds of news coming out in a short time at the beginning and at the end of the trading sessions. During February 2015, there were published more than 230 thousand news items which mentioned more than 18,000 companies. We assembled the list of all companies which have at least one common news item with at least one other company. There are more than 7,000 such companies during February, 2015.

Table 1 presents the descriptive statistics of time series.

Table 1. The descriptive statistics, one day

	All news	Stock exchanges			Sectors			
		NYSE	LSE	Tokyo SE	Products	Resources	Services	IT
Mean	8383.43	2351.57	902.43	730.75	1772.21	1348.50	1896.46	505.61
St. dev.	5415.21	1569.98	596.61	524.85	1159.12	928.81	1235.39	356.55
Skewness	-0.60	-0.49	-0.63	-0.22	-0.54	-0.17	-0.56	-0.31
Kurtosis	1.75	1.64	1.88	1.89	1.74	2.13	1.76	1.76

After obtaining non-directional symmetric matrix with valued weights for the co-mention counts of each pair of companies, we use R packages for finding basic statistics and to visualize the network of companies.

2.3 Social Network Analysis

Social Network Analysis (SNA) describes social relations in terms of graph theory. SNA presents objects (e.g. individuals, groups, organizations, URLs, and other connected entities) within the network as nodes, and links represent relationships (e.g. friendship, co-authorship, organizations and sexual relationships) between the objects [1, 14, 22, 25, 40].

Social networks can be represented in the form of a diagram, where nodes are points and links are lines. SNA deals with measuring relationships between objects [12, 26, 30, 44].

The nodes in our network represent companies and the links represent co-mention relationships between the nodes. SNA provides both visual and mathematical analysis of relationships [33, 41].

Recent years have seen increased interest in the study of Social Media using SNA (see e.g. [3, 11, 23, 45, 47], among many others).

Key objects are those that are in relationships with many other objects. In the context of our analysis, a company with extensive links or co-mention with many other companies in the economy is considered more important than a company with relatively fewer links. Different types of SNA metrics can be used to find key companies in the network. In our analysis, we use the following well-known metrics: degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, frequency. A detailed description of these metrics can be found in the article [25].

2.4 Key Company Analysis According to Sectors of Economics

In addition to the co-mention matrix, we create a matrix describing which of ten economic sectors the companies belongs to (extraction of consumer discretionary, consumer staples, energy, financial, health care, industrials, information technology, raw materials, telecommunications services and utilities) and which stock exchange the companies is related to.

All the companies we considered were divided into four sectors of the economy in the following way:

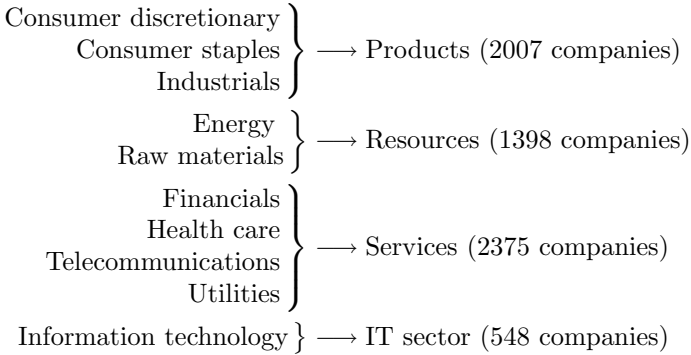


Table 2 lists the top-5 companies with the highest frequency of co-mention in each of the 4 sectors. The table also shows the number of company’s links and different standardized centrality indicators. As you can see from this table, General Motors with high frequency and with high degree is a key company in the products sector. It strongly dominates over the rest of the companies, so that it has the largest number of news co-mentions and links to other companies. We noted that in terms of proximity (closeness centrality), all considered companies are identical. This is typical for all sectors of the economy. Most of the largest companies in this sector belong to auto groups or aircraft manufactory.

Table 2 also shows the largest companies in the Resources sector. The leading companies in this sector are oil and gas companies, which is reasonable. At the same time American companies Apache Corporation and Continental Resources lead in co-mention frequency and number of links. The high of the Eigenvector centrality indicator proves that Apache Corporation and Continental Resources interact with other large companies in the sector. It acts as the bridge more often (it is a connecting link) connecting the companies of this sector.

In the services sector, there are several leaders which are the largest financial holdings: JPMorgan Chase, Citigroup, Bank of America. This group of companies are also leaders in all measures of centrality considered. Apple is the leader in the information technology sector for all of the indicators with a large margin. Earlier, in our article [35] it was shown that Apple is the leading (key) company among all the companies under consideration.

2.5 Key Company Analysis According to Stock Exchanges

Next, a network of co-mentions of companies belonging to different territorial zones was studied. We considered separately companies that are trading on the European, American and Asian stock exchanges. Precisely, we analyzed companies that traded on London SE (488 companies), New-York SE (1,715 companies) and Tokyo SE (586 companies).

Table 3 shows that the key companies of the London Stock Exchange in terms of the number of links are Barclays Bank PLC and Aviva PLC. Mining

Table 2. Companies with higher frequency for four sectors of the economy

Company	Frequency	Degree	Degree centrality $\times 10^1$	Closeness centrality $\times 10^3$	Betweenness centrality $\times 10^1$	Eigenvector centrality
<i>Products</i>						
General Motors Co	1051	143	0.71	2.0745	0.72	1.000
Ford Motor Co	691	90	0.45	2.0731	0.26	0.877
Volkswagen AG	668	101	0.50	2.0734	0.23	0.572
Boeing Co	663	103	0.51	2.0735	0.34	0.084
Bayerische Motoren Werke AG	652	108	0.54	2.0724	0.17	0.576
<i>Resources</i>						
Apache Corp	1583	118	0.84	2.5374	0.04	0.966
Continental Resources Inc	1538	101	0.72	2.5373	0.06	1.000
Pioneer Natural Resources Co	1422	70	0.50	2.5354	0.00	0.984
RSP Permian Inc	1369	67	0.48	2.5353	0.00	0.937
Devon Energy Corp	1282	92	0.66	2.5356	0.01	0.849
<i>Services</i>						
JPMorgan Chase & Co	882	137	0.58	1.6921	0.43	1.000
Citigroup Inc	837	128	0.54	1.6922	0.43	0.889
Bank of America Corp	757	142	0.60	1.6926	0.69	0.763
Goldman Sachs Group Inc	706	105	0.44	1.6915	0.13	0.868
American Express Co	622	126	0.53	1.6921	0.39	0.391
<i>IT sector</i>						
Apple Inc	805	127	2.32	6.1457	2.59	1.000
Alphabet Inc	472	67	1.22	6.1364	0.79	0.927
Twitter Inc	371	54	0.99	6.1293	0.27	0.640
Samsung Electronics Co Ltd	336	56	1.02	6.1364	0.79	0.618
Facebook Inc	307	47	0.86	6.1323	0.32	0.640

company BHP Billiton PLC has the largest number of co-mentioned news and interacts with the largest companies of this exchange. Further in the table the key companies of the New York Stock Exchange are given; all of them belong to the oil and gas industry. At the same time key companies of the Tokyo Stock Exchange relate to engineering and industrial sectors.

2.6 Degree Distribution Analysis

In this section, the distribution analysis for the sub-graphs on 4 sectors of the economy and for the 3 stock exchanges is provided. The description of this procedure can be found in the articles [2, 19].

Table 3. Companies with higher frequency for three stock exchanges

Company	Frequency	Degree	Degree centrality $\times 10^1$	Closeness centrality $\times 10^5$	Betweenness centrality $\times 10^2$	Eigenvector centrality
<i>London SE</i>						
Barclays PLC	484	42	1.72	3.2481	3.37	0.374
Aviva PLC	429	47	1.93	3.2517	5.20	0.308
BHP Billiton PLC	363	81	3.33	3.2576	4.29	1.000
BP PLC	313	53	2.18	3.2511	3.43	0.536
Associated British Foods PLC	303	68	2.79	3.2534	3.00	0.733
<i>New-York SE</i>						
Continental Resources Inc	1594	178	2.08	0.2767	0.50	0.720
Apache Corp	1588	184	2.15	0.2774	1.02	0.733
Anadarko Petroleum Corp	1386	217	2.53	0.2780	1.96	0.831
Basic Energy Services Inc	1336	252	2.94	0.2778	1.05	1.000
Halliburton Co	1281	247	2.88	0.2778	1.52	0.924
<i>Tokyo SE</i>						
Bombardier Inc	515	154	5.26	2.0368	6.32	0.994
Calfrac Well Services Ltd	489	105	3.59	2.0299	0.59	0.710
Alara Resources Ltd	464	120	4.10	2.0320	1.08	0.883
Newalta Corp	428	147	5.03	2.0345	1.86	1.000
Strad Energy Services Ltd	414	101	3.45	2.0305	0.68	0.633

A degree distribution is called power-law distribution if

$$n(k) \sim Ak^{-\gamma},$$

where γ is degree exponent.

For all stock exchanges under review (Table 5) and economic sectors (Table 4) the distribution of degrees follows the power law. Figure 1 shows the dependence of the number of companies $n(k)$ on node degree k for NYSE.

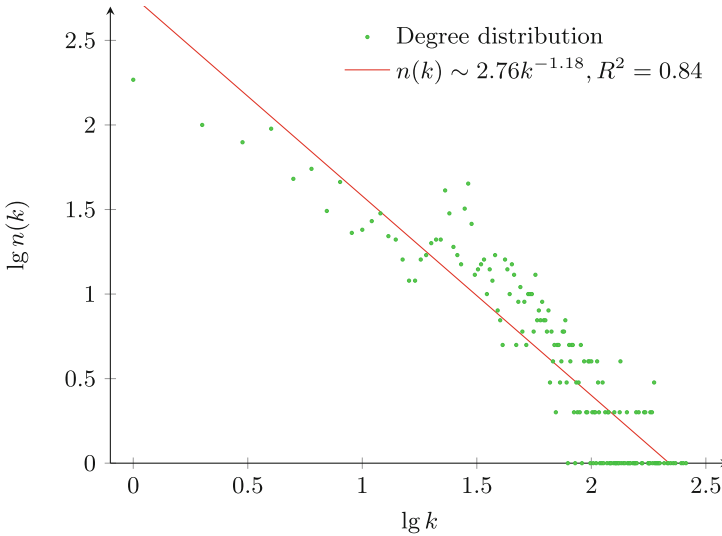
The resulting models are statistically significant at alpha level of 0.01. However, the degree exponent for all stock exchanges and sectors of the economy does not fall in the interval (2, 3). Thus, the decrease in the degrees of vertices is slower than for typical social networks [16, 24]. It is also notable that the fewer companies are in a subgraph, the slower the degree of vertices decreases and the less the coefficient of determination is.

Table 4. Degree distribution for four sectors of the economy

Subgraphs	Degree exponent γ	R^2
Products	1.06	0.82
Resources	0.91	0.79
Services	1.14	0.84
Information technology	0.64	0.65

Table 5. Degree distribution for three stock exchanges

Subgraphs	Degree exponent γ	R^2
London SE	1.10	0.81
New-York SE	1.18	0.84
Tokyo SE	0.68	0.44

**Fig. 1.** The degree distribution of the New-York's companies co-mention network

2.7 Clustering Coefficient Distribution

In this section, clustering coefficient distributions analysis will be carried out for sub-graphs in 4 sectors economy and for 3 stock exchanges described above. Description of these procedures can be found in articles [4, 15, 43].

The average clustering coefficient of nodes $C(k)$ with degree k has been found:

$$C(k) \sim Bk^{-\beta},$$

where the exponent β usually lies between 1 and 2 [31, 32, 46].

For the given networks, the clustering-degree distribution relation follows the power law.

The resulting models are statistically significant at alpha level of 0.01. Herewith, the exponent β is turned out less than 1 for all the subgraphs under consideration (Tables 4 and 5).

It should also be noted that the power dependence between the local clustering coefficient and the degree is manifested for sufficiently large degrees of the vertices of k . There is no dependence for relatively small k . This fact is typical for all subgraphs under consideration and it can be well observed in the example of a subgraph (Fig. 2). Herewith, the exponent β is turned out less than 1 for all stock exchanges and sectors of the economy (Tables 6 and 7).

Sometimes the flow contains news which mention a large number of actors (companies). For example, about 0.5% of news reports allude to 10 or more companies. In such cases, the procedure we used to fill the incidence matrix generated a “pleiad” - a subgraph that included all possible verges between the actors mentioned in the report, which led to deviations of the vertex degree and clustering coefficient values from the general pattern. Note that exclusion of a large number of co-occurrences from the news analysis eliminated this problem. However, this method leads to a significant loss of information. It seems more accurate to use a modified procedure for clustering coefficient calculation that takes into account the peculiarities of the co-occurrences flow. We are planning to do this in the future.

In further research, we are taking into account this feature of the news flow, particularly in procedures for the news flow filtering and decomposing. We propose to allocate and consider repetitive co-mentions (a stable part of the graph in time), co-mentions caused by certain events in politics or economy (eventual co-mentions), and, finally, random perturbations separately.

Table 6. Local clustering coefficient for four sectors of the economy

Subgraphs	Exponent β	R^2
Products	0.70	0.58
Resources	0.44	0.46
Services	0.70	0.61
Information technology	0.61	0.54

Table 7. Local clustering coefficient for three stock exchanges

Subgraphs	Exponent β	R^2
London SE	0.68	0.55
New-York SE	0.68	0.74
Tokyo SE	0.78	0.74

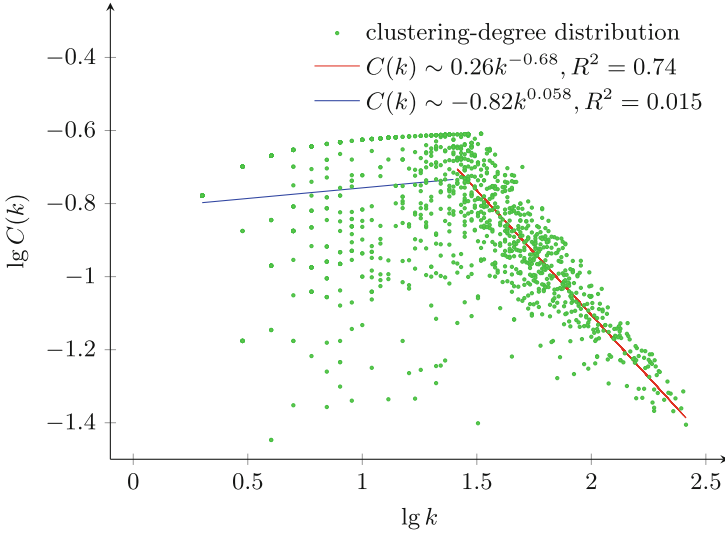


Fig. 2. The clustering-degree distribution of the New-York's companies co-mention network

3 QAP Correlation and Regression Analysis

Using the co-mention matrix and the companies' sector affiliation matrix (as well as the stock exchange affiliation matrix) we conduct QAP Correlation Analysis. QAP (Quadratic Assignment Procedure) was proposed and developed in [13, 18, 21, 27]. We use QAP Correlation Analysis to identify correlations

- between the co-mention network and companies' sector affiliation,
- between the co-mention network and stock exchange affiliation.

With the co-mention network as a prime network, corresponding cells of the sector affiliation matrix (as well as the stock exchange affiliation matrix) are compared to compute the value of Pearson's correlation. We repeat the process randomly permuting columns and rows to find the correlation. A lower value of Pearson's correlation means a stronger relationship between the matrices.

The first research hypothesis states that the graph of companies resulting from a network analysis identify sectors. First, QAP analyses between the co-mention network and the sector affiliation network was carried out to examine whether the sector affiliation network can predict the structure of the co-mention network. We calculate the value of Pearson correlation using R package.

The QAP correlation analysis shows a significant correlation between the co-mention network and the stock exchange affiliation ($r = 0.053$, $p = 0.000$) and the sector affiliation network ($r = 0.020$, $p = 0.000$).

We exploited the QAP procedure for testing the significance of the correlation coefficients. Estimated density of QAP replications for the sector affiliation

network is shown in Fig. 3. Similar results were also obtained for the network of co-mentions and the Stock exchange affiliation. The observed values of the correlation coefficients were higher than the model values in all simulated 500 samples. Thus, the observed correlation coefficients are statistically significant, while they are close to zero. This can be explained by the fact that the adjacency matrices were of large dimension and were sufficiently discharged. This could contribute to the underestimation of the correlation coefficient.

We evaluated the linear regression between the elements of the co-mention network matrix and the stock exchange affiliation matrix, as well as elements of the co-mention network matrix and the sector affiliation matrix. The values of the parameters found by the least squares method and p -value obtained by the QAP procedure are given in Table 8.

The QAP regression analysis shows (Table 8), that only 0.2% variance is predicted by the model ($R^2 = 0.002$). This value is relatively low, and indicates an insufficient inclusion of explanatory variables. The coefficients of the model are statistically significant.

Table 8. QAP regression analysis

Coefficients	Estimate	P-value
Intercept	0.003	0.000
Sector affiliation	0.015	0.000
Stock exchange affiliation	0.033	0.000

The co-mention network of companies has similar structure to their territorial connections. We visualize the co-mention matrix using R package. We identify clusters using the level of link weights which is derived from co-mention frequencies. We use the sector affiliation matrix as attribute data to see if companies of one cluster have the attribute in common. Figure 4 shows a small part of network map of companies. Nasdaq companies with New York Stock Exchange and Dax companies with London Stock Exchange make two clusters. London companies along with GM (General Motors) are more likely to be a bridge for the US connection with Europe.

Big companies generate a much bigger news flow than small companies. For this reason, the co-mention matrix is dense for the largest companies and is much sparser for small companies. However, the overall spatial and sector affiliation of companies significantly affects the probability and frequency of co-mentions. In our opinion, the widening of the range of analyzed companies induces the sparsity of the co-mention matrix, which leads to a drop in the share of the explained variance.

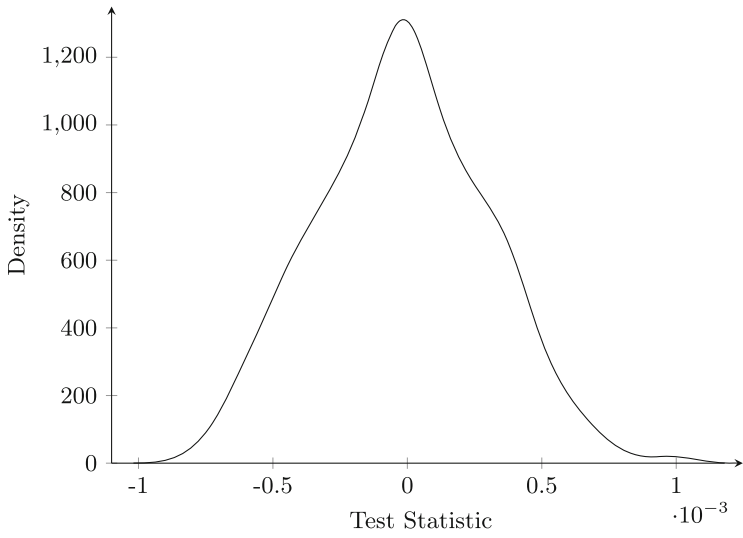


Fig. 3. Estimated density of QAP replications for the sector affiliation network

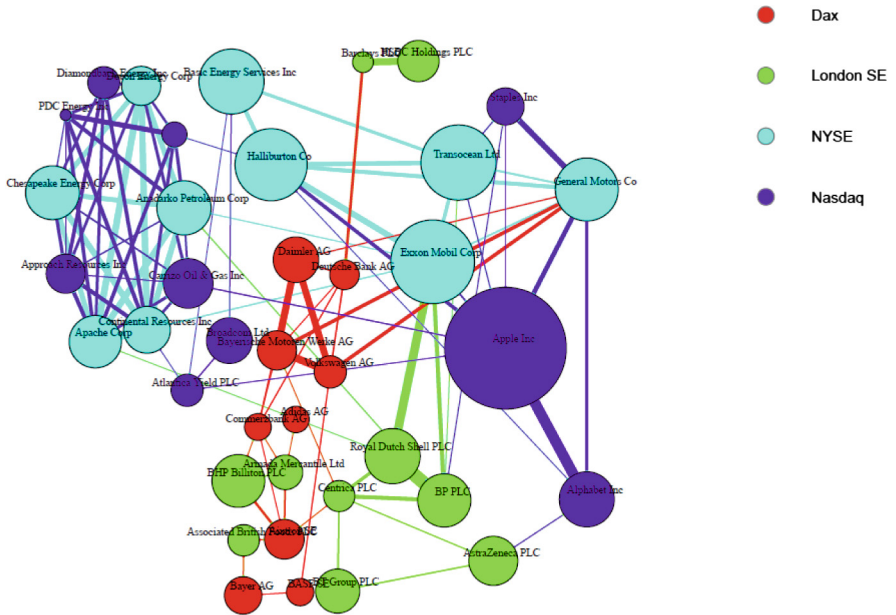


Fig. 4. Networked Map of companies

4 Conclusion

In this article, we investigated the relationship between company co-mention network and the sector affiliation matrix. Moreover, we identified key companies in different sectors of the economy using various indicators of network analysis, such as frequency, normalized degree of centrality, closeness centrality, betweenness centrality and eigenvector centrality. We discovered that different network analysis indicators show different values for different companies. But some of the companies have high significance for all indicators considered. At the same time, the majority of leading (key) companies belong to the New York Stock exchange. It was shown that the distribution of degrees and clustering-degree relations for our network adheres to the power law, although with nonstandard indicators of exponent. QAP analysis showed the presence of significant positive correlation between company co-mention network and stock exchange affiliation, and between company co-mention network and sector affiliation network.

References

1. Abbasi, A., Altmann, J.: On the correlation between research performance and social network analysis measures applied to research collaboration networks. In: 44th Hawaii International Conference on System Sciences (HICSS), pp. 1–10. IEEE (2011)
2. Albert, R., Barabasi, A.L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002)
3. An, J., Kwak, H.: What gets media attention and how media attention evolves over time: large-scale empirical evidence from 196 countries. In: Proceedings of the Eleventh International Conference on Web and Social Media, pp. 464–467. The AAAI Press, Palo Alto, Montreal, May 2017
4. Anthonisse, J.M.: The rush in a directed graph. Technical (1971)
5. Atzmueller, M., Schmidt, A., Kloepper, B., Arnu, D.: HYPGRAPHS: an approach for analysis and assessment of graph-based and sequential hypotheses. In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Raś, Z.W. (eds.) NFMCP 2016. LNCS (LNAI), vol. 10312, pp. 231–247. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61461-8_15
6. Barnett, G.: A longitudinal analysis of the international telecommunication network, 1978–1996. *Am. Behav. Sci.* **44**, 1638–1655 (2001)
7. Barnett, G., Danowski, J.: The structure of communication: a network analysis of the international communication association. *Hum. Commun. Res.* **19**(2), 264–285 (1992)
8. Barnett, G., Salisbury, J.: Communication and globalization: a longitudinal analysis of the international telecommunication network. *J. World Syst. Res.* **2**(16), 1–17 (1996)
9. Basov, N., Lee, J.S., Antoniuk, A.: Social networks and construction of culture: a socio-semantic analysis of art groups. In: Cherifi, H., Gaito, S., Quattrociocchi, W., Sala, A. (eds.) *Complex Networks & Their Applications V. COMPLEX NETWORKS 2016*, vol. 693, pp. 785–796. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50901-3_62

10. Choi, E., Lee, K.C.: Relationship between social network structure dynamics and innovation: micro-level analyses of virtual cross-functional teams in a multinational B2B firm. *Comput. Hum. Behav.* **65**, 151–162 (2016)
11. Coletto, M., Garimella, K., Gionis, A., Lucchese, C.: A motif-based approach for identifying controversy. In: *Proceedings of the Eleventh International Conference on Web and Social Media*, pp. 496–499. The AAAI Press, Palo Alto, Montreal, May 2017
12. Correa, C., Crnovrsanin, T., Ma, K.L.: Visual reasoning about social networks using centrality sensitivity. *IEEE Trans. Vis. Comput. Graph.* **18**(1), 106–120 (2012)
13. Dekker, D., Krackhardt, D., Snijders, T.A.B.: Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika* **72**(4), 563–581 (2007)
14. Deville, P., Song, C., Eagle, N., Blondel, V.D., Barabási, A.L., Wang, D.: Scaling identity connects human mobility and social interactions. *Proc. Nat. Acad. Sci.* **113**(26), 7047–7052 (2016). <http://www.pnas.org/content/113/26/7047>
15. Granovetter, M.: The strength of weak ties. *Am. J. Sociol.* **78**, 1360 (1973)
16. Guelzim, N., Bottani, S., Bourgine, P., Kepes, F.: Topological and causal structure of the yeast transcriptional network. *Nat. Genet.* **31**, 60–63 (2002)
17. Haishu, Q., Ying, L., Xin, O.: Industrial association, common information spill out and industry stock indexes co-movement. *Syst. Eng. Theory Pract.* **36**(11), 2737 (2016)
18. Hubert, L.: *Assignment Methods in Combinatorial Data Analysis*. Dekker, New York (1987)
19. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabasi, A.L.: The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000)
20. Kim, H., Barnett, G.A.: Social network analysis using author co-citation data. In: *AMCIS 2008 Proceedings*, Paper 172, pp. 1–9 (2008)
21. Krackardt, D.: Qap partialling as a test of spuriousness. *Soc. Netw.* **9**(2), 171–186 (1987)
22. Landherr, A., Friedl, B., Heidemann, J.: A critical review of centrality measures in social networks. *Bus. Inf. Syst. Eng.* **2**(6), 371–385 (2010)
23. Le, H., Shafiq, Z., Srinivasan, P.: Scalable news slant measurement using twitter. In: *Proceedings of the Eleventh International Conference on Web and Social Media*, pp. 584–587. The AAAI Press, Palo Alto, Montreal, May 2017
24. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A.: Transcriptional regulatory networks in *saccharomyces cerevisiae*. *Science* **298**(5594), 799–804 (2002)
25. Liu, B.: *Web Data Mining*. Springer, Heidelberg (2007). <https://doi.org/10.1007/978-3-540-37882-2>
26. Liu, X., Bollen, J., Nelson, M.L., de Sompel, V.: Co-authorship networks in the digital library research community. *Inf. Process. Manag.* **41**(6), 1462–1480 (2005)
27. Mantel, N.: The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**(2), 209–220 (1967)
28. Mitra, G., Mitra, L. (eds.): *The Handbook of News Analytics in Finance*. Wiley, Hoboken (2011)
29. Mitra, G., Yu, X. (eds.): *Handbook of Sentiment Analysis in Finance* (2016)
30. Onnela, J.P., Arbesman, S., Gonzalez, M.C., Barabasi, A.L., Christakis, N.A.: Geographic constraints on social network groups. *PLoS ONE* **6**(4), 1–7 (2011). <https://doi.org/10.1371/journal.pone.0016939>

31. Ravasz, R., Barabasi, A.L.: Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003)
32. Ravasz, R., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabasi, A.L.: Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002)
33. Said, Y.H., Wegman, E., Sharabati, W.K., Rigsby, J.: Social networks of author-coauthor relationships. *Comput. Stat. Data Anal.* **52**(4), 2177–2184 (2008)
34. Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., Strohmaier, M.: Linguistic neighbourhoods: explaining cultural borders on wikipedia through multilingual co-editing activity. *EPJ Data Sci.* **5**, 1–20 (2016)
35. Sidorov, S.P., Faizliev, A.R., Balash, V.A., Gudkov, A.A., Chekmareva, A.Z., Anikin, P.K.: Company co-mention network analysis. *Springer Proceedings in Mathematics and Statistics* (2018, in press)
36. Sidorov, S., Faizliev, A., Balash, V.: Measuring long-range correlations in news flow intensity time series. *Int. J. Mod. Phys. C* **28**(08), 1750103 (2017)
37. Sidorov, S., Faizliev, A., Balash, V.: Scale invariance of news flow intensity time series. *Nonlinear Phenom. Complex Syst.* **19**(4), 368–377 (2016)
38. Sidorov, S., Faizliev, A., Balash, V.: Fractality and multifractality analysis of news sentiments time series. *IAENG Int. J. Appl. Math.* **48**(1), 90–97 (2018)
39. Sidorov, S., Faizliev, A., Balash, V., Korobov, E.: Long-range correlation analysis of economic news flow intensity. *Phys. A* **444**, 205–212 (2016)
40. Sinatra, R., Wang, D., Deville, P., Song, C., Barabási, A.L.: Quantifying the evolution of individual scientific impact. *Science* **354**(6312), aaf5239 (2016). <http://science.sciencemag.org/content/354/6312/aaf5239>
41. Tang, J., Zhang, D., Yao, L.: Social network extraction of academic researchers. In: *Seventh IEEE International Conference on Data Mining*, pp. 292–301. IEEE (2007)
42. Vahtera, P., Buckley, P.J., Aliyev, M., Clegg, J., Cross, A.R.: Influence of social identity on negative perceptions in global virtual teams. *J. Int. Manag.* **23**(4), 367–381 (2017)
43. Wagner, A., Fell, D.A.: The small world inside large metabolic networks. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 1803–1810 (2001)
44. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge (1994)
45. West, R., Pfeffer, J.: Armed conflicts in online news: a multilingual study. In: *Proceedings of the Eleventh International Conference on Web and Social Media*, pp. 309–318. The AAAI Press, Palo Alto, Montreal, May 2017
46. Yook, S.H., Oltvai, Z.N., Barabasi, A.L.: Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928–942 (2004)
47. Zhang, A., Culbertson, B., Paritosh, P.: Characterizing online communities using coarse discourse structures. In: *Proceedings of the Eleventh International Conference on Web and Social Media*, pp. 357–366. The AAAI Press, Palo Alto, Montreal, May 2017