Article

Stability Analysis of Company Co-Mention Network and Market Graph Over Time Using Graph Similarity Measures

Alexey Faizliev* , Vladimir Balash , Vladimir Petrov, Alexey Grigoriev, Dmitriy Melnichuk and Sergei Sidorov

Mathematics and Mechanics Department, Saratov State University, Saratov 410012, Russia
* Correspondence: faizlievar1983@mail.ru

check for updates

Abstract: The aim of the paper is to provide an analysis of news and financial data using their network representation. The formation of network structures from data sources is carried out using two different approaches: by building the so-called market graph in which nodes represent financial assets (e.g., stocks) and the edges between nodes stand for the correlation between the corresponding assets, by constructing a company co-mention network in which any two companies are connected by an edge if a news item mentioning both companies has been published in a certain period of time. Topological changes of the networks over the period 2005–2010 are investigated using the sliding window of six-month duration. We study the stability of the market graph and the company co-mention network over time and establish which of the two networks was more stable during the period. In addition, we examine the impact of the crisis of 2008 on the stability of the market graph as well as the company co-mention network. The networks that are considered in this paper and that are the objects of our study (the market graph and the company co-mention network) have a non-changing set of nodes (companies), and can change over time by adding/removing links between these nodes. Different graph similarity measures are used to evaluate these changes. If a network is stable over time, a measure of similarity between two graphs constructed for two different time windows should be close to zero. If there was a sharp change between the graphs constructed for two adjacent periods, then this should lead to a sharp increase in the value of the similarity measure between these two graphs. This paper uses the graph similarity measures which were proposed relatively recently. In addition, to estimate how the networks evolve over time we exploit QAP (Quadratic Assignment Procedure). While there is a sufficient amount of works studying the dynamics of graphs (including the use of graph similarity metrics), in this paper the company co-mention network dynamics is examined both individually and in comparison with the dynamics of market graphs for the first time.

Keywords: graph dynamics; social networks; market graph; graph similarity measures

1. Introduction

The modern economy is a complex system consisting of an enormous number of companies that interact with each other to achieve their own goals. Modeling the aggregate as well as local behavior of such systems is an extremely important, albeit complex problem. One of the modern approaches to building models of economic or financial systems is graph models that are based on transforming empirical data into a network representation using additional reasonable assumptions. In such graphs, nodes usually correspond to companies, and edges between nodes reflect the relations between them. The following may serve an example of such relationships:

- direct links between companies, for example, a supplier-consumer type relationship, i.e., one of the companies supplies goods or services to another company [1–5],
- relations between banks [6–10] (lenders and borrowers in the interbank loan market);
- the connections reflecting investments of one of the companies into another one [11–14].

Unfortunately, such information is often confidential and not always available to researchers. Therefore, for the construction of network models of economic interaction, not directly observable links between economic agents are often used:

- One of the possibilities in discovering connections between companies is to use correlations between the returns of companies' assets. In accordance with the efficient market hypothesis, it is assumed that stock prices of companies and their mutual behavior reflect all publicly available information about companies. Thus, economic and financial connections between companies may be reflected by the correlation of the log returns of company assets [15–32].
- Some researchers believed that a connection between companies arises, if both companies have common members of the board of directors [33–37].
- Several papers dealt with company co-mention networks in which connection between companies is reflected by the fact of mentioning the both companies in one news item [38–41].
- Also, in various applications it can be useful to build graphs of industrial or spatial affiliation of companies [40].

It should be noted that in the course of their activities, economic agents generate a large amount of publicly available information which includes news reports on companies published by news agencies. News flow also includes SEC reports, court documents, reports of various government agencies, business resources, company reports, announcements, industrial, and macroeconomic statistics. The news flow containing financial and economic news items is extremely intense (thousands of news items per second) and exhibits high unstructurability. The analysis of the characteristics of time series corresponding to the flow of financial and economic news is an important and interesting task. The study of such characteristics would allow a deeper understanding of the features of the news flow background and its dependence on the current situation on financial markets.

News analytics data providers such as Thompson Reuters and Raven Pack collect data from different sources including news agencies and social media (blogs, social networks, etc.) and process such data in real time [42,43]. The paper [44] studies structural characteristics of news flows generated by news agencies, enterprises, organizations, social networks, etc.

Following ideas of [38,39] the current paper presents a company co-mention network as a graph where the world's major companies mentioned in financial, business-related, and economic news flow are shown as nodes. If any two companies are mentioned in the same piece of news, the company co-mentions graph shows it as the edge between two nodes. Various important characteristics of company co-mention networks were studied with use of different SNA metrics, such as eigenvector centrality, degree of centrality, betweenness centrality, closeness centrality, frequency, etc. in papers [38,39], and key companies of networks were found. It was demonstrated that the degree distributions as well as the clustering-degree of the analyzed graphs follow a power law, but using a non-typical exponent value. The subgraph analysis in terms of industrial and spatial affiliation allowed for key company identification in the network and the possibility to research the analogous power-law distribution for company co-mention network subgraphs. The QAP analysis method was used in [40] to examine the correlations between the company co-mention network and graphs, describing sectoral (and spatial) affiliation. Papers [38,39] explore how the structural characteristics of the company co-mentions graph change over time, such as the distribution of the vertices degrees, the distribution of the average clustering coefficient, the edge density, the size of maximum clique, its connectivity. It was found that the power-law structure of the co-mention graph is quite stable. The degree exponent as well as the clustering-degree coefficient were at their lowest values during the

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

3 of 26

2007–2008 financial crisis. All maximum cliques comprise a big amount of companies from banking sector. The maximum independent set was the largest at the peak of the 2008 financial crisis.

The aim of the paper is to provide a joint analysis of news and financial data using their network representation. The formation of the network structures from different data sources is carried out using two different approaches:

- by building the so-called *market graphs* in which nodes represent financial assets (e.g., stocks) and the edges between nodes stand for the correlation between the corresponding assets.
- based on companies' co-mention in the news flow. The company co-mention network is constructed as follows: two companies are connected by an edge if a news item mentioning both companies has been published in a certain period of time.

The market graph as well as the company co-mention network are changing over time by means of adding or by removing links. Two co-mention networks built for two consecutive periods contain the same nodes (companies). However, the presence or absence of an edge (a link between two companies) depends on whether the news item mentioning these two companies was published during this period or not. The news generating process is random, so the presence of an edge in a specific period of time may vary from one period to another. We could not find any papers in which this behavior of the news flow is modeled. In this paper we construct the company co-mention network as well as the market graph using empirical data.

The main research questions that we would like to answer in this paper are the following ones:

1. Does the market graph remain stable over time? How significantly do the market graphs constructed for two consecutive 6-month windows differ? How did the crisis of 2008 change the stability of the market graph? Were the changes of the market graph during the crisis minor or noticeable?

2. Does the company co-mention network remain stable over time? How significantly do the company co-mention networks constructed for two consecutive windows differ? How did the crisis of 2008 change the stability of the company co-mention network? Were the changes of the company co-mention network during the crisis small or huge?

3. What of the two networks was more stable over time: the market graph or the company co-mention network?

4. How do the market graph and the company co-mention network constructed for the same time window differ?

To examine how the networks have actually evolved over time we will employ the approach described in the paper [45] based on the use of different graph similarity metrics. To avoid one-sided results associated with the wrong choice of the graph similarity metrics, in our study we will use four different measures:

- the Hamming distance (h) between graphs;
- a network similarity measure d proposed in [46] that quantifies how the set of central nodes (their ranking) has changed in a network;
- D-measure which is proposed in [47] and proved to be discriminative and computationally efficient to distinguish and quantify graph dissimilarities and which can identify and quantify topological differences between graphs;
- graph diffusion distance (GDD) [48] based on measuring the average similarity of heat diffusion on each graph.

Note the last three measures have appeared relatively recently, but have already demonstrated their advantages in several empirical studies.

In addition, we exploit QAP (Quadratic Assignment Procedure) which was introduced in [49] and developed in [50,51].

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

4 of 26

The paper has the following structure. Section 2 describes the procedures for constructing the market graph and the company co-mention network based on empirical data. In addition, in this section we describe a methodology for assessing the stability of graphs over time based on the use of graph similarity measures. Section 3 contains a brief description of the graph similarity measures, which we will use in our study. In Section 4, we describe the empirical data based on which we construct the market graph and the company co-mention network. Finally, we present our results on the analysis of the graph stability over time in Section 5. Section 5 of this paper presents new empirical study on the dynamics of company co-mention network both separately and in comparison with the dynamics of market graph, based on the estimation of graph similarity metrics of graphs constructed for successive time periods. For convenience of the reader, we list some notations in Abbreviations.

2. Data Transformation for Network Representation

2.1. Market Network Construction

The examination of properties for market networks has accentuated in the past few years. It seems that the notion of the market graph was firstly studied in the paper [15]. In his work Boginsky defined the market graph as a complete weighted graph in which the nodes (or vertices) serve as stocks and weights of edges match similarity between behavior of the stocks. The simplest way to quantify this similarity can be done with use of Pearson correlation coefficient. For this reason, Boginsky [15] suggests that an edge between two nodes (assets) is embedded in the market graph, if the corresponding value of Pearson correlation coefficient is bigger than a fixed threshold.

The market graph approach proposed in [15] has received much interest in the recent decade. In particular, many papers have obtained applicable empirical results using real market data while exploring various structural features and aspects of market graph such as maximum cliques, maximum independent sets, degree distribution [16–19], clustering in Pearson correlation [20], dynamics of the US market graphs [21], complexity of market graph [22]. The papers [17,23–26] examine distinct financial markets to find differences between them. Market graphs with measures of similarity diverse from correlation are under investigation in works [23,27–30]. An analysis for estimation of reliability of market graph approach results was presented in [31].

We construct the market network using the Pearson correlation; it means that two companies are connected if the value of the Pearson correlation between the two assets are above a given threshold (in this period of time).

More precisely, the market graph is constructed as follows. We denote by $P_i(t)$ the price of the asset i in day t. Then

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

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

Let n be the number of assets. We will suppose that random variable $R_i(t)$, $t = 1, 2, \ldots, N$, has a corresponding distribution R_i, $i = 1, 2, \ldots, N$, and the joint distribution of random $R_1, R_2, \ldots, R_N$ is not known.

The Pearson correlation coefficient between random variables R_i and R_j is defined by

$$r_{ij} = \frac{\sum (R_i(t) - \overline{R_i})(R_j(t) - \overline{R_j})}{\sqrt{\sum (R_i(t) - \overline{R_i})^2} \sqrt{(R_j(t) - \overline{R_j})^2}},$$

where $\overline{R_i} = \frac{1}{T} \sum_{i=1}^{T} R_i(t)$ denotes the mean value of R_i.

The Pearson correlation is the most popular measure exercised in the examination of the finance market. The main shortcoming of the Pearson correlation is weak robustness to deviations from the assumptions on identity distribution of the random variables in question.

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

The market graph constructed with use of the measure linearly dependent on the sign correlation was studied in [23,32]. The paper showed that the measure is capable for the analysis of the market graphs. As pointed out in [23], the sign correlation has a few important differences from the Pearson correlation, which makes it more applicable to our analysis than the classical correlation.

It is worth noting that for large graphs it would be more computationally effective to use a clustering algorithm to find clusters of highly connected nodes, and then compute similarity between pairs of nodes within each cluster and then between clusters [52]. Please note that the computation of node similarity between all pairs over the entire graph may be much more time costly for large networks. In our research the graph sizes do not exceed 1053 nodes and the construction of graphs may be made in reasonable time by calculating all pairwise correlations.

To expose the evolution of market structures we use the dynamic approach which is peculiarly useful for the comparisons of calm periods before the financial crisis of 2008 and crashes. For every stock pair, we take the log return time series in a time window with the length of six months, i.e., with price values included in the window. Using a sliding window approach, we can calculate the correlation matrices for each of the six-month sliding windows by shifting each subsequent window by one month. Thus, in the dynamic approach with sliding windows, we calculate a sequence of correlation matrices and corresponding market graphs. We chose the window length of six months for the following reasons. Too small a window length would lead to incorrect correlation dependencies, since the number of assets is more than 1000. On the other hand, choosing a larger window length would result in the effect of the shocks of one local period being reflected in the correlation matrix constructed for this long interval.

2.2. Network Representation of News Analytics Data

In the company co-mention network, the network "node" represents a company, while the nodes relationship is indicated by the edge. If a company was mentioned in the same news report with some other companies, it is connected with them. The company co-mention network can be viewed as an undirected weighted graph, and therefore, the company co-mention network can be treated as a social network.

We conduct the analysis of the companies' co-mention network according to the pattern outlined in [38]:

1. we assemble all economic, business-related and financial news published over six years (2005–2010);
2. we accomplish the process of data cleansing;
3. we chose companies cited in news reports during this time;
4. we divide the 6-year period into overlapping semiannual intervals. Each subsequent interval is obtained by shifting the previous one 1 month ahead. The result is 67 intervals of the same 6-month size (approximately 125 trading days).
5. we calculate the number of co-mentions (link weight) for every two companies cited together at least in one piece of news over each time interval. In case the companies are not co-mentioned in the given interval, the link weight is 0.
6. we used these weighted calculations of the collective companies' mentions to obtain symmetric co-mention matrices for each interval;
7. we explore the evolution of the co-mention matrices over the time, and the results of this study are being visualized and interpreted.

The first operation is executed by the news analytics providers among which are Raven Pack, Media Sentiments, and Thompson Reuters. They assemble news items in real time from various news providers and sources. They employ AI algorithm to accomplish analysis of each news item in real time for their subscribers. As a result, each news report is transformed into a set of metadata including time of publication, the name of company or asset, news relevance, novelty, etc. The comprehensive characterization of news analytics and its application in finance industry may be found in [42,43].

2.3. Methodology

Many real networks have evolved over time by adding/removing nodes or links between the nodes. The network at time *t* and the network at time *t* + 1 may differ from each other, even if the set of nodes has not changed. If these changes are neglectable, then the network remains stable over time. In addition, a network that has remained stable for a period of time may change sharply at some point in time due to some unexpected reasons.

The networks that are considered in this paper and that are the objects of our study (the market graph and the company co-mention network) have a non-changing set of nodes (companies), and can change over time by adding/removing links between these nodes:

- We construct the market graphs based on the correlations between assets for a 6-month window, moving the sliding window by one month ahead to construct the following subsequent graph.
- We construct the company co-mention network (for the same companies that form the market graph), adding an edge between two companies, if a news item mentioning both these companies was published during a 6-month window, shifting the sliding window by one month forward to construct the subsequent network.

To evaluate these changes, a graph similarity measure can be used. If a network is stable over time, the measure of similarity between two graphs constructed for two different time windows should be close to zero. If there was a sharp change between the graphs constructed for two adjacent periods, then this should lead to a sharp increase in the value of the similarity measure between these two graphs.

Currently, there are a large number of graph similarity measures. Each such measure has both positive characteristics and several drawbacks. Therefore, to avoid one-sided results caused by the wrong choice of the graph similarity measure, we will use different measures that have worked well in applied research and that evaluate the similarity of graphs in respect of various aspects and characteristics (topology, node ranking, etc.).

In this section, we describe two methods we will use to analyze the dynamics of graphs (both market graph and company co-mention network). Let $G_1, G_2, \ldots, G_T$ be the sequence of the graphs representing the states of a complex system at time slots $1, 2, \ldots, T$. Let $\rho(G_{t_1}, G_{t_2})$ be the value of a graph similarity measure calculated for two states G_{t_1} and G_{t_2}. As such a measure of similarity ρ, it can be possible to use various metrics that estimate the distance between the graphs (e.g., the difference in L_2-metric, graph-edit distance, or measures based on the presence of isomorphic subgraphs). Unfortunately, the use of these simple measures did not allow us to obtain interpretable results (both market graph and company co-mention network). Therefore, in our study we used the similarity measures described later in Section 3.

2.3.1. Dynamics Analysis Based on the Assessment of the Neighboring Graphs Similarity

The essence of this method is simple enough and consists of using two different similarity measures ρ_1 and ρ_2. It is desirable that these measures evaluate different types of graph dissimilarities (for example, topological and structural dissimilarities). As such measures, in the following sections we will use the Hamming distance and the d-measure defined below by Equations (2) and (3). The first measure measures the closeness of the local structural properties of graphs, while the second measures

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

7 of 26

the similarity of the centrality indices of vertices. Then we find the values of the measures ρ_1 and ρ_2 for all neighboring pairs of graphs. Thus, we get $T-1$ points

$$(\rho_1(G_1,G_2),\rho_2(G_1,G_2)),(\rho_1(G_2,G_3),\rho_2(G_2,G_3)),\dots,(\rho_1(G_{T-1},G_T),\rho_2(G_{T-1},G_T))$$

on the plane (ρ_1,ρ_2). Visualizing these points on the plane can help a researcher

- to find the periods in which the greatest changes occurred during the transition from one time interval to another;
- to find periods of stability in which there were no changes between adjacent graphs in terms of measures ρ_1 and ρ_2;
- to understand which characteristics of graphs have changed more: those that are evaluated by ρ_1 or those that are related to the measure ρ_2.

In particular, this approach was applied in [46] to analyze the dynamics of immigration flows between countries.

It should be noted that the proximity between pairs of points $(\rho_1(G_{i-1},G_i),\rho_2(G_{i-1},G_i))$ and $(\rho_1(G_{j-1},G_j),\rho_2(G_{j-1},G_j))$ as well as all points $i,i+1,\dots,j-1$ between them on the plane (ρ_1,ρ_2) does not guarantee the proximity of the initial G_{i-1} and the last G_j graphs. Therefore, despite the simplicity and clarity of the resulting visualizations, this approach has its limitations.

We apply this approach for visualization of the dynamics of both market graphs and company co-mention networks in Section 5.1.

2.3.2. Multidimensional Scaling Analysis Approach

Another idea that we also apply in our research is to use the multidimensional scaling analysis. First, we calculate the values of the distances between all pairs of graphs from the sequence $G_1,G_2,\dots,G_T$ using the measure ρ, and form a matrix of pairwise distances

$$\mathcal{A}=\begin{pmatrix} 0 & \rho_1(G_1,G_2) & \dots & \rho_1(G_1,G_T) \\ \rho_1(G_2,G_1) & 0 & \dots & \rho_1(G_2,G_T) \\ \dots & \dots & \dots & \dots \\ \rho_1(G_T,G_1) & \rho_1(G_T,G_2) & \dots & 0 \end{pmatrix}$$

Then applying the multidimensional scaling analysis (MSA) to matrix $\mathcal{A}$, we can derive underlying factors which influence the graph dynamics (with respect to the measure ρ). In particular, the MSA may expose essential underlying dimensions that help the researcher to interpret observed similarities or dissimilarities (distances) between the graphs.

This approach is applied to market graphs and company co-mention networks in Section 5.3.

3. Graph Similarity Measurement

The problem of finding adequate network stability and similarity measures has been the focus of research in the recent decades. The numerous algorithms, techniques and similarity measures can be grouped into several main categories: edit distance/graph isomorphism [53–55], common subgraphs [56–58], statistical methods (feature extraction) [59–63], and iterative methods [64–66]. Another simple way to quantify the similarity of two networks is to find Pearson correlation coefficient for two adjacency matrices corresponding to the networks.

Paper [46] pointed out that main shortcoming of many methods for graph similarity quantification is that they do not take into consideration topological structure of the networks. All edges are treated equally with no regard to the fact whether they link two disconnected components or two vertices in a dense network. To reflect and quantify topological similarities of the networks several approaches have been developed in papers [46,67,68].

In this section, we briefly describe the well-known Hamming distance and the graph similarity measures proposed in [46–48] that we will use in Section 5.

3.1. The Hamming Distance: Similarity of Local Structure

The Hamming distance is a special instance graph-edit distances and measures the number of edge deletions and insertions necessary to transform one graph into another. The Hamming distance can be used for a network dynamics analysis which shows how a network evolved over time in terms of its local structure. The brief description of this approach can be found in this subsection.

Let A^t denote the adjacency matrix of graph G at time t. The Hamming distance between networks at two time slots t_1 and t_2 is defined as follows:

$$h(G_1, G_2) = \frac{\sum_{i,j}^n \left| A_{ij}^{(1)} - A_{ij}^{(2)} \right|}{n(n-1)}. \tag{2}$$

The Hamming distance $h(G_1, G_2)$ is symmetric and varies from 0 to 1. If $h(G_1, G_2) = 1$ then the networks are completely different. If $h(G_1, G_2) = 0$ then these networks are identical.

3.2. d-Measure: Node Similarity Measure Based on Interval Orders

Paper [46] proposes a measure that describes the distance between two graphs G_1 and G_2. The measure $d(G_1, G_2)$ uses an interval order idea in a network theory by evaluating how the central nodes of network have changed.

Let G_1 and G_2 be two graphs which we would like to compare using the sets of their most important nodes. Let the graphs have the same number of vertices n and the same set of nodes. Let c_i^t be the centrality of node i in graph G_t, $t = 1, 2$. In our study we rank the nodes of the graphs based on the PageRank measure.

Let $\mathcal{R}^t = \left[\text{rank}_{ij}^t \right]$ represent our knowledge about comparable ranking of vertices in graph G_t formed by means of their centrality evaluation at time t:

$$\text{rank}_{ij}^t = \begin{cases} 1, & c_i^t - c_j^t > \varepsilon \\ 0, & \text{else.} \end{cases}$$

Paper [46] pointed out that the selection of parameter ε should be based on the problem under consideration. In our study we chose $\varepsilon = 0.00001$, so even relatively small changes in data would be taken into account.

Then d-measure between G_1 and G_2 (the distance between the two rankings for the networks G_1 and G_2) is defined in [46] using the Hamming distance formula:

$$d(G_1, G_2) = \frac{\sum_{i \neq j}^n \left| \text{rank}_{ij}^{(1)} - \text{rank}_{ij}^{(2)} \right|}{n \cdot (n-1)}. \tag{3}$$

The d-measure is symmetric and varies from 0 to 1.

3.3. D-Measure

The D-measure (dissimilarity measure) was proposed in [47].

Let the distance distribution in each node i of the graph G with n nodes, $P_i = \{p_i(j)\}$, is given, where $p_i(j)$ denotes the proportion of nodes which are connected to node i at distance j. Comprehensive information of the network topology in a compressed way is presented in the set of n node-distance distributions, $\{P_1, \ldots, P_n\}$.

For an N-nodes network, the set of n distance distributions $\{P_1, \ldots, P_n\}$, is normalized by $log(d+1)$, where d is the diameter of network.

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

9 of 26

NND is defined by the following equation:

$$NDD(G) = \frac{J(P_1, \ldots, P_n)}{\log(d+1)},$$ (4)

where

$$J(P_1, \ldots, P_n) = \frac{1}{n} \sum_{i,j} p_i(j) \log \left(\frac{p_i(j)}{\mu_j} \right)$$

is the Jensen–Shannon divergence of the N distributions and $\mu_j = \frac{1}{n} \sum_{i,j} p_i(j)$ is their average.

The D-measure was defined in [47] as follows:

$$D(G_1, G_2) = w_1 \sqrt{\frac{J(\mu_{G_1}, \mu_{G_2})}{\log 2}} + w_2 \left| \sqrt{NDD(G_1)} - \sqrt{NDD(G_2)} \right| + \frac{w_3}{2} \left(\sqrt{\frac{J(P_{\alpha G_1}, P_{\alpha G_2})}{\log 2}} + \sqrt{\frac{J(P_{\alpha G_1^c}, P_{\alpha G_2^c})}{\log 2}} \right),$$

where μ_{G_1}, μ_{G_2} are the graphs averaged node-distance distributions, NDD is defined in (4), G_1^c, G_2^c are complements of G_1 and G_2. The last term includes the comparison of α-centrality values of the graphs computed through the Jensen–Shannon divergence.

In our paper we use the weights $w_1 = w_2 = 0.45$, $w_3 = 0.1$, which was suggested in [47] as the most appropriate way to quantify structural dissimilarities in graphs.

3.4. Graph Diffusion Distance

Graph diffusion distance (GDD) was proposed in [48]. GDD is aimed at evaluating the dissimilarity between two graphs with the same number of nodes and is based on quantification of the average similarity of heat diffusion in the graphs. To compute the value of GDD it is necessary to find (for each graph) Laplacian exponential kernel matrices that arise in solving the heat diffusion problem with initial conditions restricted to single vertices. Then the value of GDD is defined in [48] as the Frobenius norm of the difference of the kernels, at the diffusion time in which the difference is achieved its maximum.

3.5. Combined Similarity Metric

It can be possible to visualize the overall changes in a graph as a point $(h(G_1, G_2), d(G_1, G_2))$ in two-dimensional space (h, d).

In case the obtained point is a null point, $(h(G_1, G_2), d(G_1, G_2)) = (0, 0)$, we treat the networks as identical. When $(h(G_1, G_2), d(G_1, G_2)) = (1, 0)$, the network structure differs entirely yet the central elements are same (i.e., complete or empty graph). If $(h(G_1, G_2), d(G_1, G_2)) = (1, 1)$ then two networks are different both in local structures and sets of key elements (for example, node chain compared to inverse chain). If $(h(G_1, G_2), d(G_1, G_2)) \rightarrow (0, 1)$ then there is a complete instability in terms of its central elements.

Paper [46] suggests transforming the two measures into one similarity measure

$$l_\alpha(G_1, G_2) = \alpha \cdot d(G_1, G_2) + (1 - \alpha) \cdot h(G_1, G_2),$$ (5)

where α is relative importance of the ranking distance.

When $\alpha = 0$, the similarity measure can be considered similar to classical measures as being based on the network structure. If $\alpha = 1$, then the networks are similar (in the case when network node-rankings disregarding the local structure). If $\alpha = 0.5$ then both measures are equal. The main trends in a network can be revealed by the two measures application, which can also be used to create a comparison of pairs of temporal networks with clustering procedure. This is used on a network to find its homogeneous periods or life cycles.

In our study, we use (5) with $\alpha = 0.1$.

J. Open Innov. Technol. Mark. Complex. **2019**, 5, 55

10 of 26

3.6. QAP Procedure

One of the methods for graph similarity estimation is the applied quadratic assignment procedures (QAP) regression. In our research we use QAP procedure to examine the stability of the market graph and company co-mention network over time. It should be noted that the application of the standard OLS regression would provide incorrect results due to the fact that this method relies on the assumption of independency of the observations and that they are identically distributed. Indeed, since many vertices of the network are connected by links, the directly or indirectly linked vertices have potentially dependent relation. Thus, the precondition for ordinary least squares method is not met.

For this reason, QAP regression proposed by D. Krackhardt in [49] uses nonparametric permutation. The QAP procedure permutes rows and columns of the graph matrices, and then correlation coefficient between independent adjacency matrices and the dependent adjacency matrix is calculated. The QAP procedure repeats permutations of rows and columns of the adjacency matrices many times to find a test statistic for testing the null hypothesis of the regression.

It was shown in [69] that in the case of high autocorrelation the QAP procedure leads to a much lower proportion of type 1 error than OLS regression.

In our research, we would like to find

- the dependence between the adjacency matrix of the market graph constructed in a given period and matrices constructed for other periods;
- the dependence between the adjacency matrix of the company co-mention network constructed in a given period and matrices constructed for other periods;
- the dependence between the adjacency matrix of the market graph constructed in a given period and the adjacency matrix of the company co-mention network constructed for the same period.

In such network matrices, the autocorrelation might occur. By this reason we employ the QAP regression procedure.

QAP method has proved to be successful in many applied problems: for identifying significant factors for predicting social relations [70], for finding important factors that influence web citation among universities [71], to study the job mobility of scientists [72], to recognize the patterns in patent network analysis [73].

4. Data

4.1. Financial Data

The database for constructing and analyzing the market graph was taken from the Yahoo Finance. The daily data were collected from Yahoo Finance database, which was used to retrieve historical prices of the companies traded in the largest stock exchanges for the period from 1 January 2005 to 31 December 2010 (i.e., 1500 trading days). To study the dynamics of the market graph, the 1500-day trading days interval was divided into 67 consecutive *overlapping* 125-day periods. The dates corresponding to each period are presented in Table 1.

Table 1. Periods ID's and their starting and ending dates.

Period	Start	End
1	01.01.2005	30.06.2005
2	01.02.2005	31.07.2005
3	01.03.2005	31.08.2005
…	…	…
…	…	…
65	01.05.2010	31.10.2010
66	01.06.20010	30.11.2010
67	01.07.20010	31.12.2010

Market network is formed based on correlation; it means that a company has connection with those companies which have the positive significant correlation of assets with it in this period of time. In our research market graphs were constructed as it is described in Section 2.

4.2. News Analytics Data

The paper analyzes the entire scope of financial, business-related and economic news published over six years (72 months) from 1 January 2005 to 31 December 2010. The news analytics data were cleared to eliminate all messages on the beginning and end of the exchange trading sessions and analytical reports with tabular data. Overall, the cleared data set contained over 8,550,000 messages for a six-year period. The intensity of the news flow remained rather stable over the time interval. The news count increased by an average of 2% per year. The monthly number's magnitude ranged from 90,000 to 145,000. The maximum points of co-mentions may correlate with the period of the early 2007–2008 financial crisis of (Figure 1).

The number of companies in which there was at least one mention of them in 5 years exceeded 24,000. Moreover, 18,500 enterprises had at least one joint mention in the same time interval.

Table 2 shows that 92.2% of the entire amount of news mentioned only one of the enterprises. 7.1% cited two companies, whereas 0.5% of all news items mentioned three companies. The number of news containing co-mentions (i.e., related to more than one firm) ranged between 5.5% and 11.4% in different months. Less than 0.05% of the messages contained the co-mention of four or more enterprises. News reports containing simultaneous mention of ten or more firms were fairly rare (fewer than 50 news items over a 6-year period). The highest number of enterprises cited in one piece of news was 14.

Table 3 shows thatthe total of co-mentioned pairs over five years was more than 1,757,000. Over 50% of news reports and 45% of co-mentions were associated with firms (stocks) traded in the United States. Over 90% of news and co-mentions were connected to companies (stocks) traded on the 15 largest exchanges. Table 4 shows the amount of news items mentioning a given number of companies in each year from 2005 to 2010.

For each of the 72 months the number of co-mentions of each pair of companies was calculated, then the corresponding adjacency matrices of co-mention graphs were created. At the next step, we ranked the companies by the average co-mention number per month. The leader (the most frequently co-mentioned company) was determined in the news stream along with other enterprises in 220 messages on average per month. For 4 years, over 4000 assets were cited with the leader. However, only about 200 companies had co-mentions together with the leader more than one time per year.

Table 2. The amount of news (Freq.) mentioning a given number of companies (K).

K	Freq.	Percent	Cum.	Co-mentions	Co-mentions Percent. Cum.
1	7,891,180	92.22	92.22	0	69.5
2	610,824	7.14	99.36	1,221,648	69.5
3	43,352	0.51	99.86	260,112	84.3
4	6887	0.08	99.94	82,644	89.0
5	1553	0.02	99.96	31,060	90.8
6	650	0.01	99.97	19,500	91.9
7	928	0.01	99.98	38,976	94.1
8	1611	0.02	100	90,216	99.2
9	126	0	100	9072	99.8
10	33	0	100	2970	99.9
11	11	0	100	1210	100.0
14	1	0	100	182	100.0
Total	8,557,156	100		1,757,590	100

Table 3. The amount of news items in each year from 2005 to 2010.

Year	Total Amount of News Items	News Items with Co-Mentions	The Share of News with Co-Mentions	Amount of Co-Mentions
2005	1,332,680	109,560	8.22	252,344
2006	1,351,598	117,933	8.73	269,538
2007	1,460,248	124,299	8.51	422,912
2008	1,451,137	116,103	8	312,870
2009	1,471,312	101,339	6.89	254,352
2010	1,490,181	96,742	6.49	245,574
Total	8,557,156	665,976	7.78	1,757,590

Table 4. The amount of news items mentioning a given number of companies (*K*) in each year from 2005 to 2010.

| K	2005	2006	2007	2008	2009	2010	Total
1	1,223,120	1,233,665	1,335,949	1,335,034	1,369,973	1,393,439	7,891,180
2	102,505	110,590	112,436	105,207	91,984	88,102	610,824
3	6395	6749	7408	8597	7655	6548	43,352
4	576	533	1384	1433	1335	1626	6887
5	65	45	449	309	300	385	1553
6	13	11	404	112	48	62	650
7	2	3	781	126	11	5	928
8	3	2	1314	276	3	13	1611
9	0	0	94	30	1	1	126
10	0	0	23	10	0	0	33
11	1	0	6	3	1	0	11
14	0	0	0	0	1	0	1
Total	1,332,680	1,351,598	1,460,248	1,451,137	1,471,312	1,490,181	8,557,156
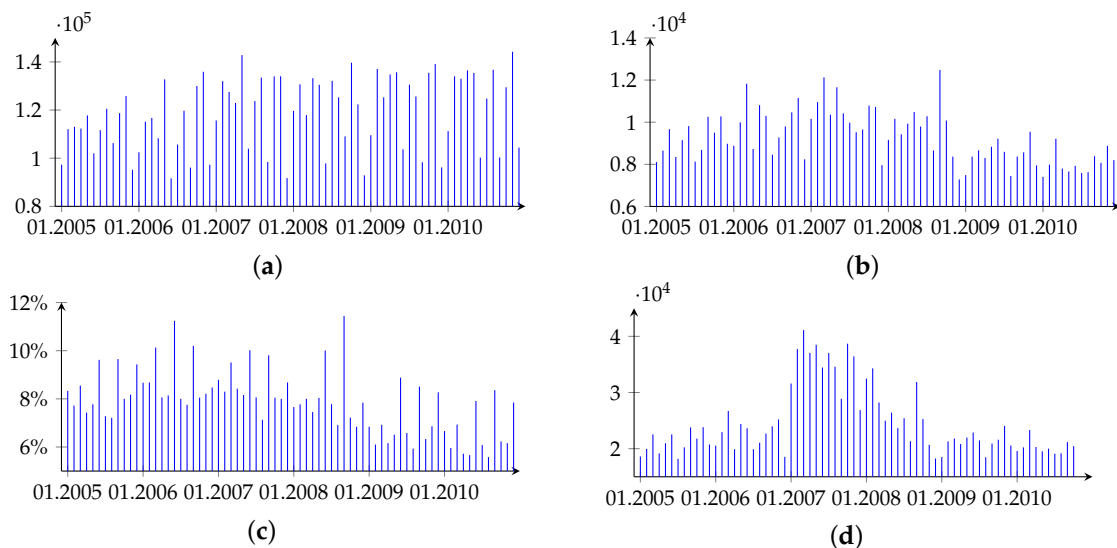
Figure 1. Monthly dynamics of (**a**) the amount of news; (**b**) the amount of news with co-mentions; (**c**) the share of news items with co-mentions; (**d**) the amount of co-mentions, from January 2005 to December 2010.

In our research company co-mention networks were constructed as it is described in Section 2.

5. Empirical Result

We divide 6-year interval into 67 half-years overlapping intervals and choose 1053 companies with highest density of news that mention them during the period under review. We excluded news with relevance under 80 (i.e., news with 80% or less probability of being connected with the company). Then for each time interval we check the amount of co-mentions for each pair of companies in one article (if two companies are both mentioned in one article during the period of time, the weight of the link is considered 1); if companies were not mentioned during the interval the weight of the link is considered 0. Then we form unweighted symmetric matrices of co-mentions for each time interval using these weighed calculations of the collective companies' mentions.

The market graph is based on correlations between 1053 shares chosen while forming the co-mention network. It means that a company has connection with those companies which have the positive significant correlation of assets with it in this period of time. In order to have market graphs similar to the co-mention graphs correlation threshold value is made 0.6. Figure 2 shows the dynamics of edge density of the resulting graphs for the chosen periods. It can be seen that the edge density had its highest values during the 2008 financial crisis (the dark fragment in the middle of the figure). It has to be mentioned that the density of the co-mention network has been reasonably stable and has been insignificantly rising before the 2008 financial crisis, while the market graph had noticeable edge density rise during the major events of the financial crisis.
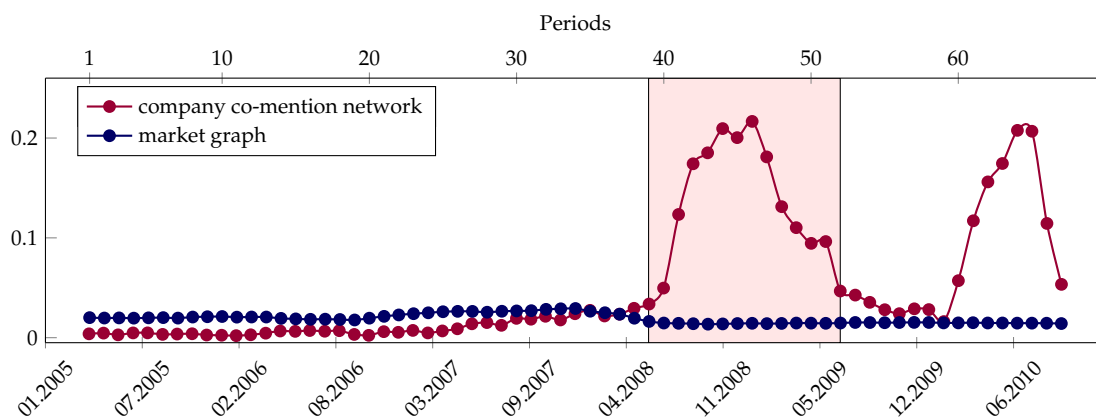

Figure 2. The dynamics of edge density for the company co-mention network (dark red) and the market graph (dark blue) from the first of 67 six-month periods (1 January 2005–30 June 2005)to the last six-month period (1 July 2010–31 December 2010). Points mark the middles of six-month periods. The outlier corresponds to 40–41 periods.

5.1. Similarity Analysis Using Measures h and d

We apply the proposed model to the co-mention network and to the market graph. The information about how the structure of the market graph changed over the adjacent half-years regarding ranking distance d and local structure distance h is shown in Figure 3.

For each six-month window (period) we constructed a market graph in accordance with the approach described in Section 2.1. The IDs of the periods and their starting and ending dates are given in Table 1. Thus, we obtained 67 market graphs $M_1, M_2, \ldots, M_{67}$ corresponding to the 67 six-month periods. Similarly, we obtained 67 company co-mention networks $C_1, C_2, \ldots, C_{67}$ corresponding to each of the 67 periods (see Table 1) using the methodology described in Section 2.2.

We found the values of d-metric for each pair of graphs constructed for all two consecutive 6-month periods, i.e., $d(M_1, M_2), d(M_2, M_3), \ldots, d(M_{66}, M_{67})$. In addition, we calculated the values of h-metrics for each pair of graphs constructed for all of two consecutive 6-month periods,i.e., $h(M_1, M_2), h(M_2, M_3), \ldots, h(M_{66}, M_{67})$.

Figure 3 shows the evolution of ranking and local structure distances between each pair of market graphs constructed for every pair of consecutive six-month periods, i.e., between 1 and 2, between 2 and 3,..., between 66 and 67. Thus, *i*-th point on the (h, d)-plane has coordinates $(h(M_i, M_{i+1}), d(M_i, M_{i+1}))$, $i = 1, \ldots, 66$. Each point on the plane characterizes the differences between the graphs at the current and previous time windows, evaluated by both the Hamming distance *h* and the *d*-measure. This visualization allows one to distinguish periods with higher or lower intensity of graph changes.

Figure 3 shows that the local structure of the market graph changed very little until the beginning of the 2008 crisis (blue points). However, during the crisis (red points), the values of the similarity measure *h* (i.e., the Hamming distance) between consecutive graphs increased sharply (more than ten-fold). Moreover, after the peak of the crisis was passed, the instability of the network local structure remained at the same high level (green points). On the other hand, the value of the measure *d*, which measures the proximity of the ranking of the vertices of two consecutive graphs, did not increase during the crisis.
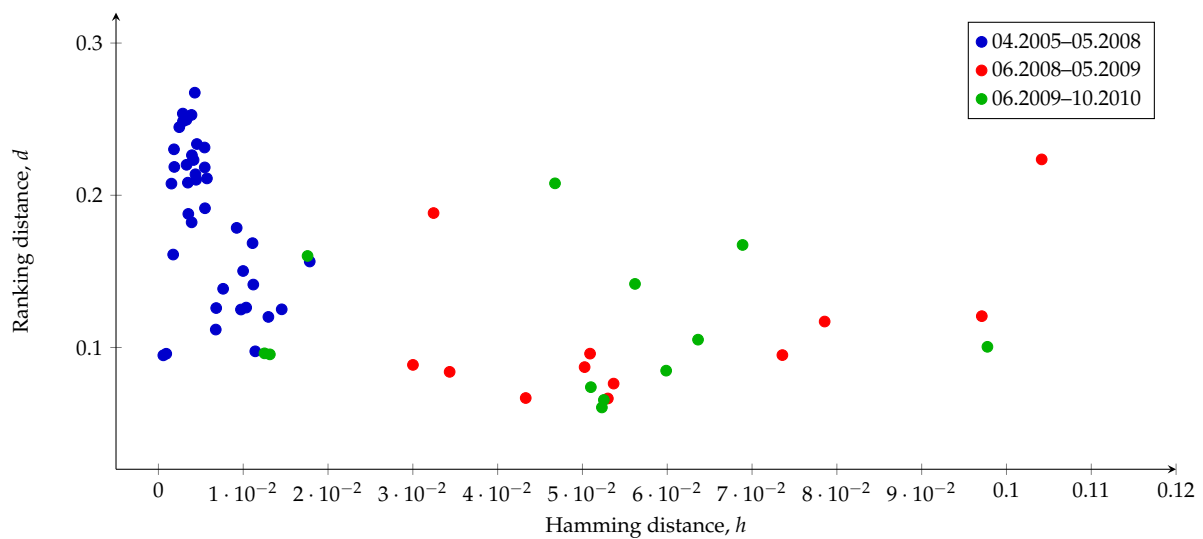

Figure 3. The evolution of ranking and the Hamming distances between each pair of market graphs constructed for every pair of consecutive six-month periods, i.e., 1–2, 2–3, . . . , 66–67.

The *i*-th point in Figure 3 show (h, d)-similarity of *i*-th and $(i + 1)$-th graphs constructed for the corresponding consecutive 6-month intervals defined in Table 1. Points $i = 1, \ldots, 38$ correspond to periods with midpoints from July 2005 to May 2008 and are colored in blue. Points $i = 39, \ldots, 50$ correspond to periods with midpoints from June 2008 to May 2009 and are colored in red. Points $i = 50, \ldots, 66$ correspond to periods with midpoints from June 2009 to October 2010 and are colored in green. It should be noted that the local structure of the market graph changed greatly at the beginning and during the financial crisis.

Figure 3 shows that structure of significant correlations between asset returns was slightly changing before the crisis, while turbulence in financial markets during the crisis was inducing the visible transformations of the market graphs. Structural changes slowed down for several periods and then they started again. The central vertices list of the market graphs was updating more intense before and after the crisis than during the crisis, i.e., the ranking order of the companies was more stable during the crisis. Perhaps, it was caused by the fact that during the crisis many vulnerable companies were from the same economic sectors that were exposed by risks.

It is well-known that if the edge densities of any two graphs are very different, then the Hamming distance between these graphs will be large. Thus, the main contribution to the change of the market graph structure was due to increase and decrease in the edge density of the graph which can be seen in Figure 2).

Please note that from the fact that the "blue" points are close to the "green" ones it does not follow that the corresponding graphs are (h, d)-close. To understand how much the graphs from the starting "blue" period differ from the "green" graphs, we conduct the multidimensional scaling analysis in Section 5.3.

Similarly, we found the values of h- and d-metrics for each pair of company co-mention networks built for all of two consecutive six-month periods, i.e., $h(C_1, C_2), h(C_2, C_3), \ldots, h(C_{66}, C_{67})$ and $d(C_1, C_2), d(C_2, C_3), \ldots, d(C_{66}, C_{67})$. Points with coordinates $(h(C_i, C_{i+1}), d(C_i, C_{i+1})), i = 1, 2, \ldots, 66$, are shown in Figure 4.

Unlike the market graph, the node ranking and the structure of co-mention networks did not change significantly over time. However, the network local structure had been changing in periods from April 2007 to March 2008 (Figure 4). This period occurs before and during the financial crisis of 2008.

Figure 4 shows that the co-mention network local structure changed slightly in 2007 (blue points). However, in the period before the crisis (red points), the values of the similarity measure h (i.e., the Hamming distance) between consecutive graphs increased by more than 1.5–2 times. Questions about what caused the changes in the local structure of the company co-mention network, as well as whether such changes in the characteristics of the news flow may be forerunner of crisis phenomena on the financial market, remain open. Surprisingly, at the very beginning of the crisis, the network local structure became more stable than in 2007, and remained stable in subsequent periods (green points). On the other hand, the value of the measure d, which measures the similarity in the ranking of the vertices of two consecutive graphs, did not increase during the crisis.
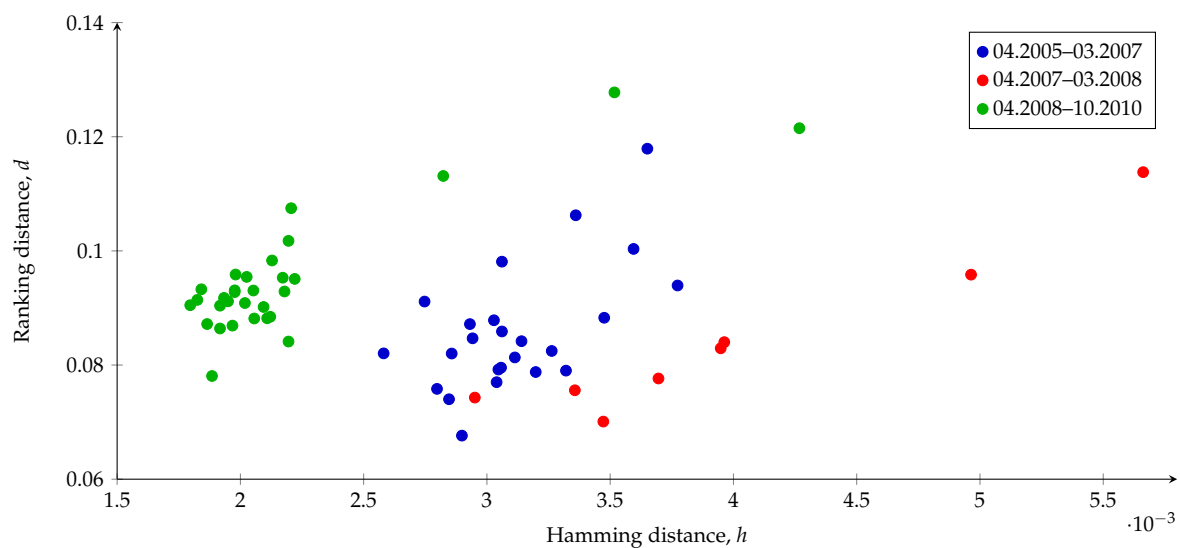

Figure 4. The evolution of the ranking and the Hamming distances between each pair of company co-mention networks constructed for each pair of consecutive six-month periods, i.e., 1–2, 2–3, …, 66–67.

The obtained values of the measures d and h for consecutive market graphs (Figure 3) significantly exceed the values of the measures d and h for consecutive company co-mention networks (Figure 4). Some values of measure d differ by more than 2 times, while the values of h-measure differ by an order of magnitude. In this sense, the company co-mention network is more stable than the market graph.

The information about how the structure of the market graph changed in the adjacent half-years regarding co-mention network is shown in Figure 5. The ranking distance has increased significantly while local structure distance has been stable and not high. So, from the local structure point of view the market graph and co-mention network are similar in many ways. The only exception are the

periods from 41 to 51 (with midpoints in August 2008–June 2009), when the United States subprime mortgage crisis started, and from 61 to 67 (with midpoints in April 2010–October 2010).
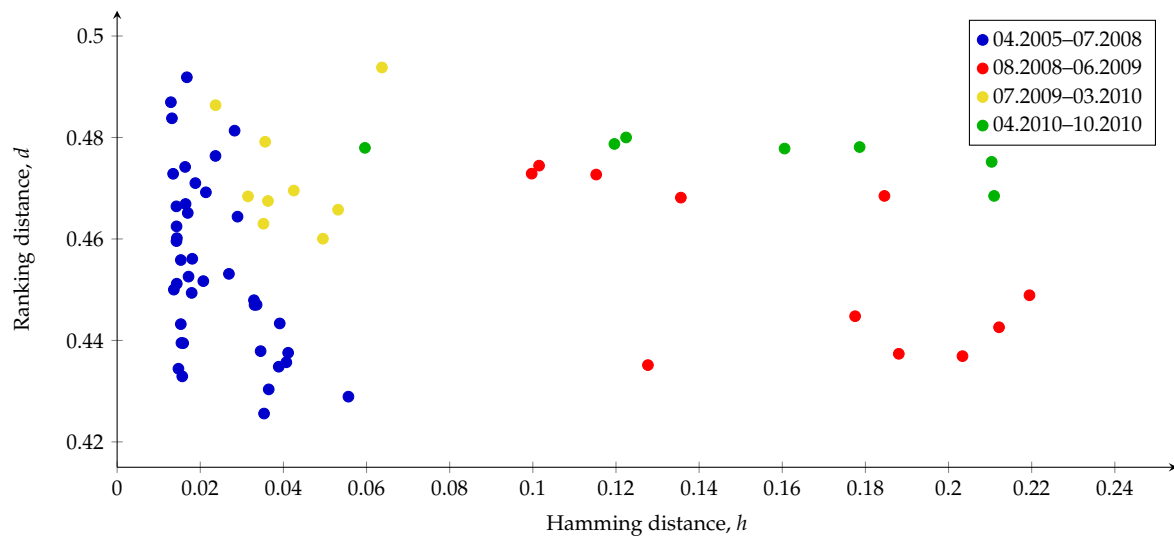

Figure 5. The evolution of the ranking and the Hamming distances between the co-mention networks and the market graphs constructed for the same period of time.

Financial and economic news which impacts an industry or a sector often mentions key companies of the industry or the sector. Therefore, the connection between companies reflected by their joint co-mention in a news item may be the result of their belonging to the same economic sector. It is known that correlations between returns on assets in the same sector are quite high. Therefore, it can be assumed that the market graph, constructed based on correlations between asset returns, and the company co-mention network, constructed on the basis of co-mentioning in the news, should be similar. However, as Figure 5 shows, this is not quite true: the differences are significant both with respect to network local structure (*h*), and with respect to node ranking (*d*).

5.2. QAP Correlation and Regression Analysis

Using networks of co-mentioning companies and market graphs, we carry out a QAP correlation analysis, since standard correlation analysis is not suitable for such data (as they are not independent from each other). This is contrary to one of the basic assumptions of linear regression analysis. QAP (Quadratic Assignment Procedure) was proposed and developed in [49–51,74]. We use QAP correlation analysis to determine the significance of correlations:

- for related networks of co-mention,
- for time-related market graphs,

When using the market graph as the main network, the corresponding cells of the matrix are compared to compute the Pearson correlation coefficient. Furthermore, this process is repeated, randomly rearranging the columns and rows to find a correlation. Lower Pearson correlation values for random permutations indicate a significant relationship between the respective matrices.

For the correlation analysis, we used the package R.

We apply QAP regression to find the factors which influence the market graph and the company co-mention network. For network presented in binary data, OLS should not be used when building regression, since this method requires observations to be independent and equally distributed. Connections between nodes in the network imply a potentially dependent relationship between either directly or indirectly connected nodes. Hence, the assumption is incorrect and the OLS method cannot be used. Rows and columns of network matrices in QAP are rearranged, thus the calculation of

correlations is done between the independent matrices and the dependent matrix. Test statistics can be obtained after several permutations, we use them to check the null regression hypothesis.

In our study, we wanted to find a connection between market graphs, company co-mention networks in adjacent periods of time. To investigate how the market graph is related to the company co-mention network, we used QAP regression, where $MarketGraph_t$ at time t is used as a dependent variable. Market graph matrices in previous periods and company co-mention networks in the current period were used as independent variables for QAP regression.

The results of the analysis are presented in Tables 5 and 6. Rows and columns of the dependent variable matrix were rearranged 1000 times. Matrices of independent variables are shown in Table 6. The QAP results showed that the market graph matrix is closely related to the market graph in the previous period of time. The exceptions are periods 37–43 (April 2008–October 2008)—the peak of the financial crisis. Company co-mention networks had a smaller impact on the market graph, though they are also significant for all models built.

Table 5. QAP Correlation Analysis.

Period	Correlation (Co-Mention)	Correlation (Market Graph)
1–7	0.1663800	0.4318455
7–13	0.1577212	0.4616943
13–19	0.1782792	0.4442250
19–25	0.1536598	0.3278332
25–31	0.1512387	0.3656817
31–37	0.1410627	0.3898995
37–43	0.1688668	0.2300743
43–49	0.1985533	0.3901301
49–55	0.2044689	0.3643512
55–61	0.2031029	0.3365267
61–67	0.2148604	0.4209299

QAP shows (Table 5) that there is a significant correlation both between adjacent co-mention networks and between adjacent market graphs. The estimated density of repeated launches of QAP shows that of all launches, correlations for random graphs turned out to be less than test statistics, and therefore the obtained correlation values can be considered statistically significant.

Estimated correlation coefficients are quite high. At the same time, the company co-mention network is stably reproduced from period to period. As for market graphs, the correlation values vary in wide ranges and it can be argued that it decreased during the beginning of the global financial crisis.

Since we have data for several types of graphs and periods of time, this also allows us to construct a linear regression on graphs. The market graph was taken as a dependent variable at the current time (period) of time (`Market graph`$_t$). The independent variables were the market graph at the previous point in time (`Market graph`$_{t-6}$) and the company co-mention network in the current period of time (`Co-mention`$_t$).

The QAP regression analysis of the dependence of the current market graph on the previous one, as well as on the current company co-mention graph, is given in Table 6. All coefficients of the models are statistically significant.

We also note that the coefficient for `Co-mention`$_t$ has its highest value for $t = 43$. This period corresponds exactly to the beginning of 2008 crisis. This indicates that during the crisis, the market graph had a special structure, which can be explained by the structure of a corresponding co-mention network.

Table 6. The results of the QAP regression analysis.

Period	Const	Market graph $_{t-6}$	Co-mention $_t$
7	0.001910899	0.4188506	0.01213292
13	0.002533768	0.5131000	0.01116052
19	0.001461219	0.3769753	0.01076794
25	0.005083621	0.4708504	0.01105582
31	0.014121804	0.6001313	0.02234221
37	0.015031197	0.4376609	0.03331619
43	0.170657423	0.5866308	0.10848345
49	0.051671309	0.3141328	0.05271308
55	0.006632613	0.1917882	0.03442195
61	0.098391587	0.6550072	0.05076811
67	0.018766982	0.2941414	0.02956882

5.3. Multidimensional Scaling

In this subsection we use the multidimensional scaling procedure to visually represent the matrix of pairwise distances between graphs (both market graphs and company co-mention networks). Multidimensional scaling was developed in [75] and aims in a graphical representation of distances between sets of objects [76]. Given a small number of dimensions, *k*, and for a given distance matrix with the distances between each pair of objects (graphs), multidimensional scaling algorithm is aimed in placing every object (graph) into *k*-dimensional Euclidian space in a way such that the between-object distances obtained by graph similarity measures would be preserved as close as possible.

The best-known methods of multidimensional scaling are metric, non-metric and generalized multidimensional scaling methods. Please note that metric multidimensional scaling algorithm finds a linear relationship, while non-metric multidimensional scaling algorithm is characterized by a set of nonparametric monotonic curves. Since we used quantitative rather than ordinal scales, the preference was given to the classical multidimensional scaling (MDS) which is also known as principal coordinates analysis [77].

Since we consider two sequence of graphs (market graphs and company co-mention networks) and use five measures for calculating distances between graphs, results are formed as the ten matrices of pairwise distances between graphs (five for market graphs, and five for co-mention graphs). Therefore, we apply the multidimensional scaling procedure to the ten distance matrices.

Let ρ be a similarity measure which finds the distance (similarity) $\rho(G_1, G_2)$ between two graphs G_1 and G_2. In our study we will use as ρ:

- the Hamming distance *h*;
- the network similarity measure *d* proposed in [46];
- *D*-measure [47];
- graph diffusion distance (GDD) [48].

Using the measure ρ, we can find the distance matrix (adjacency matrix) $(\rho(M_i, M_j))_{i,j=1}^{67}$ between all pairs of market graphs from our sequence $M_1, M_2, \ldots, M_{67}$. Also, using the measure ρ, we can calculate the distance matrix (adjacency matrix) $(\rho(C_i, C_j))_{i,j=1}^{67}$ between all pairs of company co-mention networks from the sequence $C_1, C_2, \ldots, C_{67}$.

Multidimensional scaling analysis allows us

- to visualize the dynamics of changes in the sequence of graphs;
- to find the number of components (factors) explaining the dynamics which is determined by adjacency matrices.

Therefore, the multidimensional scaling analysis can provide an important insight into the dynamics of both market graphs and company co-mention networks.

Figure 6a presents the results of multidimensional scaling applied to the distance matrix between market graphs which is calculated using *h*-measure defined in (2). Figure 6a shows that the local structure of the market graph is stable over time. During the financial crisis of 2008 (periods 38–50), the topological dissimilarity increases significantly and quickly returns to its previous level. Redundancy analysis shows that 56% of the variance is explained by the first principal component which is good enough.

Figure 6b presents the results of multidimensional scaling applied to the distance matrix between the co-mention graphs which is calculated using *h*-measure defined in (2). Figure 6b shows that the topological dissimilarity of co-mention graphs is largely decreased before the beginning of the crisis and quickly returns to its previous level after that. Only 20% of the variance is explained by the first principal component.

Figure 6c presents the results of multidimensional scaling to the market graph for distance matrix obtained using *d*-measure defined in (3). There can be seen a significant shift of the central nodes (companies) of the market graph during the crisis. 28% of the variance is explained by the first principal component.

Figure 6d presents the results of multidimensional scaling to the company co-mention graph for distance matrix obtained using *d*-measure defined in (3). Figure 6d shows that for the co-mention graph there is a monotone increase in the rank distance, which accelerates after the crisis. Thus, the crisis led to significant changes in the ranking order of the co-mention graph companies. Only 31% of the variance is explained by the first principal component.

The results of multidimensional scaling to the market graph and the co-mention graph based on the distance matrix obtained using the linear combination of *d* and *h* defined in (5) are presented in Figure 6e (with $\alpha = 0.5$) and Figure 6f (with $\alpha = 0.05$).

Figure 6g presents the results of multidimensional scaling applied to the market graph for distance matrix obtained using *D*-measure. It should be noted that the results are quite similar to the results shown in Figure 6c. 59% of the variance is explained by the first principal component.

Figure 6h presents the results of multidimensional scaling applied to the co-mention graph for distance matrix obtained using *D*-measure. There can be seen a significant decrease before the beginning of the crisis and an increase to higher level after that. 75% of the variance is explained by the first principal component.

Figure 6k,l present the results of multidimensional scaling applied to the market graph and to the co-mention graph respectively for distance matrix obtained using Graph Diffusion Distance. The results are similar to the results shown in Figure 6a. 39% of the variance is explained by the first principal component for the market graph and 13% for the co-mention graph.

The graph similarity measures (*D*-measure, Graph Diffusion Distance, *d*, *h*) showed similar results for the market graph in terms of the principal component method. In the case of the *D*-measure and *h*-metrics it suffices to use only the first principal component. For the co-mention network, there were obtained different results for different measures. Except for the *D*-measure, the first principal component explains less than 32% of the total variance.

However, it seems that the calculation of *D*-measure is the most time-consuming with comparison to other similarity measures. In our study, we used the corresponding R functions to estimate the similarity between graphs with 1053 nodes. The calculation of the similarity for each of the pairs using *D*-measure lasted about 5 times longer (and even more in case of increasing the edge density of the graphs) with comparison to *d*-, *h*-metrics and GDD.
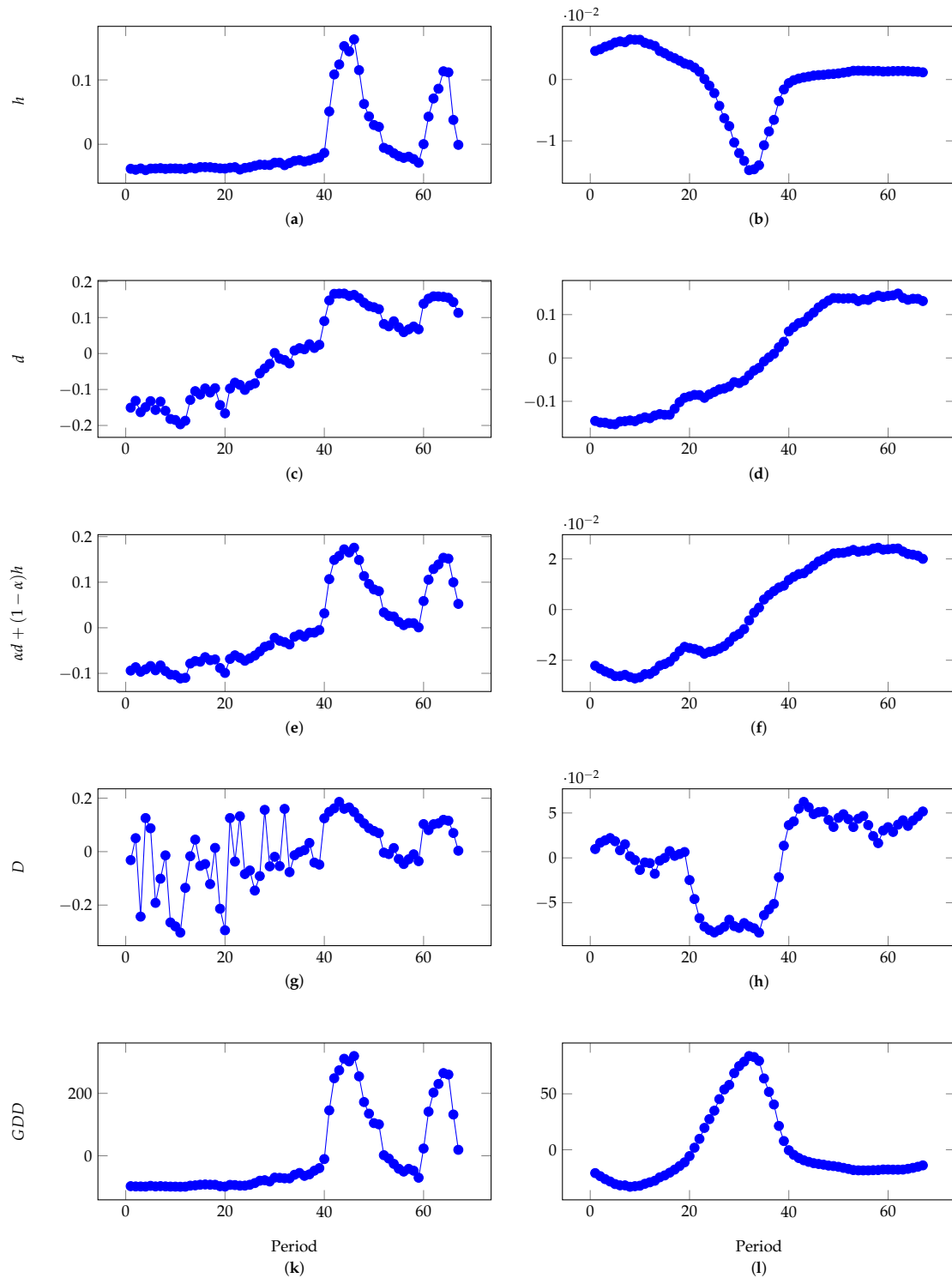
Figure 6. Graphic representation of the results of calculating the difference matrices between graphs. Figures (**a**,**c**,**e**,**g**,**k**) present the results of multidimensional scaling applied to the distance matrix between market graphs which is calculated using the Hamming distance, d-measure, l_α-measure with $\alpha = 0.05$, D-measure and GDD, respectively. Figures (**b**,**d**,**f**,**h**,**l**) present the results of multidimensional scaling applied to the distance matrix between company co-mention networks which is calculated using the Hamming distance, d-measure, l_α-measure with $\alpha = 0.05$, D-measure and GDD, respectively.

Below we draw some conclusions on the results of the multidimensional scaling (MDS).

We found that the one-factor model can explain a significant part of the change dynamics in the structure of both the market graph and the co-mention graph. However, the reliability of the conclusion essentially depends on the choice of a graph similarity measure.

One-factor estimates obtained by the MDS based on the distance matrix for the market graphs are turned out to be slightly diverse for different graph similarity measures. In particular, the use of h-measure and GDD metrics gives very similar results, which are different from the results obtained for d- and D- measures. The one-factor estimates obtained by the MDS for the co-mentioning graphs are more sensitive to the choice of the graph similarity measure.

We would like to note that visual representations of the evolution of the market graph constructed using the Hamming distance and GDD-measure (Figure 6a,k), show very similar temporal dynamics.

The visual representations of the evolution of the company co-mention network constructed using these two measures (Figure 6b,l) show also quite similar temporal dynamics, which differ only in sign.

The apparent similarity of the edge density dynamics (Figure 2) with the dynamics shown in Figure 6a,k indicates that the main factor, that has been identified by the MDS when using the Hamming distance or GDD-measure, is the graph edge density. In other words, the dynamics of graph changes obtained using the Hamming distance or GDD-measure can be easily explained by such a simple factor as the graph edge density.

On the other hand, the use of d-measures allowed us to identify almost identical dynamics for both the market graph and the co-mention network over time (Figure 6c,d). The figures show that these changes took place smoothly and continuously, while the ranking of the central nodes during the entire period under consideration changed quite significantly in both graphs.

The results obtained using the D-measure are more ambiguous. Figure 6g,h show that one factor is not sufficient to explain the dynamics of the market graph. It seems that the D-measure is a more adequate tool for network comparison, since it uses more factors to explain the differences between the graphs.

One method out of five (d-measure) shows a significant difference in the structure of graphs in the pre-crisis and after crisis periods. The dynamics of changes for the market graph are turned out to be not similar to the dynamics of the company co-mention network. However, we obtained the closest similarity when applying d-measure.

6. Conclusions

In this paper, we applied the methods of graph similarity analysis to study the network structures that describe the correlation relationship between the profitability of financial assets (market graphs) and the co-mentions of companies in the news flow (co-mention networks) during 2005–2010. In order to analyze the variability of the network structures over time, different methods were used to calculate the graphs similarity (graph diffusion distance, D-measure, node ranking similarity-based metric and the Hamming distance). In addition, QAP correlation and regression analysis were used to examine graphs similarity. The results of applying different methods for measuring differences in network structures turned out to be generally consistent with each other. The structures of graphs in adjacent periods are quite similar. However, the Hamming distance has shown great sensitivity to differences in market graphs, based on the data for the half year during the financial crisis of 2008, and the preceding and subsequent periods as well. On the other hand, nodes similarity-based metric better reflects the migration of the position of the central nodes in the co-mention graph. In addition, the use of the QAP procedure confirmed the presence of significant correlations between the adjacency matrices of the market graph and the company co-mention graph.

Our study analyzes changes in the graph properties corresponding to two parallel processes as well as similarities or differences in their dynamics. Moreover, we examine how stable the results of this analysis are regarding the choice of the graph similarity metrics. To do this, we calculated distance matrices of graphs constructed from data for successive periods, and analyzed the distance

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

22 of 26

matrices using the multidimensional scaling method (MSM). The results of applying five different graph similarity measures are compared. We can make the following conclusions:

- We found that the market graph constructed based on correlations between financial asset returns was significantly less stable over time than the company co-mention network in the period 2005–2010. In fact, the value of the Hamming distance between two consecutive market graphs reached the value around 0.1 in some periods, i.e., about 10% of links were added or removed in the graph when the six-month sliding window was shifted one month ahead. At the same time, the value of the Hamming distance between any two consecutive company co-mention networks did not exceed 0.06. In addition, the values of the d metric for the market graph were twice or triple as great as for the co-mention network.
- A common and quite intuitive point of view is that the changes in the news flow intensity and structure may be the cause of the volatility in financial markets. On the other hand, sharply increased volatility can cause a sharp surge in the amount of news items published by news agencies. According to these ideas, the structure of the news flow and the level of volatility should be correlated. However, as our results show, the structure and intensity of the news flow is extremely stable and cannot be either the cause or the result of changes in the volatility of the financial market.
- According to empirical data, the structure of the co-mention network slightly changed approximately one year before the crisis began. However, these changes are minor and cannot explain the appearance of the global financial crisis that broke out a year after.
- Please note that changes of the market graph structure are either related to the increase in volatility caused by the fall in financial asset prices during the crisis (the first peak in Figure 2), or to the volatility associated with the subsequent increase in asset prices (the second hump in Figure 2). These changes of the market graph structure are also well reflected in the Figure 6a,c,e,g,k. Perhaps, one could make the market graph more stable in time applying the dynamic formation of the threshold θ.

In this paper, we examined the evolution similarity of two network structures reflecting the same fundamental process, namely the pricing of financial assets. Obviously, company co-mentions is only a small part of the news flow background, but they are observable and available in real time, while correlations between asset prices are available with a delay. If the information contained in company co-mentions in financial and economic news flow is significant for stock market participants then it should be reflected in asset prices and similar trends should be present in the dynamic market graphs. Therefore, an interest for further research may include:

- the development of methods for joint analysis of trends in the evolution of two simultaneously formed networks;
- the development of models and methods for the detection of local mutual causality in the evolution of company co-mention network and market graphs.

Abbreviations

$P_i(t)$	the price of the asset i in day t
$R_i(t)$	the log return of asset i in day t defined by Equation (1)
n	the number of assets
r_{ij}	the Pearson correlation coefficient between random variables R_i and R_j defined in Section 2.1
ρ_1, ρ_2	any two different similarity measures
$G_1, G_2, \ldots, G_T$	the sequence of the graphs representing the states of a complex system at time slots $1, 2, \ldots, T$
$\mathcal{A}$	the matrix of pairwise distances between all pairs of graphs from the sequence $G_1, G_2, \ldots, G_T$ using the measure ρ
$h(G_1, G_2)$	the Hamming distance between networks G_1 and G_2 at two time slots t_1 and t_2 defined in Section 3.1
A^t	the adjacency matrix of graph G at time t
$\mathcal{R}^t = \left[\text{rank}_{ij}^t\right]$	the matrix representing information about relative ranking of nodes based on their centralities at time t
$d(G_1, G_2)$	the d-measure between G_1 and G_2 (the distance between the two rankings for the networks G_1 and G_2) defined by Equation (3)
$D(G_1, G_2)$	the D-measure (dissimilarity measure) between G_1 and G_2 defined by Equation (5)
$GDD(G_1, G_2)$	the graph diffusion distance between G_1 and G_2 described in Section 3.4
$l_\alpha(G_1, G_2)$	the linear combination of $d(G_1, G_2)$ and $h(G_1, G_2)$ defined by Equation (5)
$M_1, M_2, \ldots, M_{67}$	67 market graphs corresponding to the 67 six-month periods
$C_1, C_2, \ldots, C_{67}$	67 company co-mention networks corresponding to each of the 67 periods

References

1. Cheng, C.Y.; Chen, T.L.; Chen, Y.Y. An analysis of the structural complexity of supply chain networks. *Appl. Math. Model.* **2014**, *38*, 2328–2344. [CrossRef]
2. Bellamy, M.; Basole, R. Network Analysis of Supply Chain Systems: A Systematic Review and Future Research. *Syst. Eng.* **2013**, *16*, 235–249. [CrossRef]
3. Long, Q. Data-driven decision making for supply chain networks with agent-based computational experiment. *Knowl.-Based Syst.* **2018**, *141*, 55–66. [CrossRef]
4. Long, Q. A framework for data-driven computational experiments of inter-organizational collaborations in supply chain networks. *Inf. Sci.* **2017**, *399*, 43–63. [CrossRef]
5. Borgatti, S.P.; Li, X. On social network analysis in a supply chain context. *J. Supply Chain. Manag.* **2009**, *45*, 5–22. [CrossRef]
6. Boss, M.; Elsinger, H.; Summer, M.; Thurner, S., IV. Network topology of the interbank market. *Quant. Financ.* **2004**, *4*, 677–684. [CrossRef]
7. Affinito, M.; Pozzolo, A.F. *The Interbank Network across the Global Financial Crisis: Evidence from Italy*; Temi di discussione (Economic working papers) 1118, Bank of Italy, Economic Research and International Relations Area; Bank of Italy: Rome, Italy, 2017.
8. Stefano, B.; Guido, C.; Marco, D.; Stefano, G. Leveraging the network: A stress-test framework based on DebtRank. *Stat. Risk Model.* **2016**, *33*, 117–138.
9. Gofman, M. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *J. Financ. Econ.* **2017**, *124*, 113–146. [CrossRef]
10. Bundi, N.; Khashanah, K. Complex Interbank Network Estimation: Sparsity-Clustering Threshold. In *Complex Networks and Their Applications VII*; Aiello, L.M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L.M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 487–498.
11. Gorgoni, S.; Amighini, A.; Smith, M. (Eds.) *Networks of International Trade and Investment*; Vernon Press: Wilmington, DE, USA, 2018.
12. Hochberg, Y.V.; Lindsey, L.A.; Westerfield, M.M. Resource accumulation through economic ties: Evidence from venture capital. *J. Financ. Econ.* **2015**, *118*, 245–267. [CrossRef]

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

24 of 26

13. Bygrave, W.D. The structure of the investment networks of venture capital firms. *J. Bus. Ventur.* **1988**, *3*, 137–157. [CrossRef]

14. Xue, C.; Jiang, P.; Dang, X. The dynamics of network communities and venture capital performance: Evidence from China. *Financ. Res. Lett.* **2019**, *28*, 6–10. [CrossRef]

15. Boginsky, V.; Butenko, S.; Pardalos, P.M. *Innovations in Financial and Economic Networks*; Edward Elgar Publishing Inc.: Northampton, UK, 2003; pp. 29–45; Chapter on Structural Properties of the Market Graph.

16. Boginski, V.; Butenko, S.; Pardalos, P.M. Statistical analysis of financial networks. *Comput. Stat. Data Anal.* **2005**, *48*, 431–443. [CrossRef]

17. Huang, W.Q.; Zhuang, X.T.; Yao, S. A network analysis of the Chinese stock market. *Phys. Stat. Mech. Its Appl.* **2009**, *388*, 2956–2964. [CrossRef]

18. Tse, C.K.; Liu, J.; Lau, F.C.M. A network perspective of the stock market. *J. Empir. Financ.* **2010**, *17*, 659–667. [CrossRef]

19. Boginski, V.; Butenko, S.; Pardalos, P.M. Network Models of Massive Datasets. *Comput. Sci. Inf. Syst.* **2004**, *1*, 75–89. [CrossRef]

20. Onnela, J.P.; Kaski, K.; Kertész, J. Clustering and information in correlation based financial networks. *Eur. Phys. J. B* **2004**, *38*, 353–362. [CrossRef]

21. Boginski, V.; Butenko, S.; Pardalos, P.M. Mining market data: A network approach. *Comput. Oper. Res.* **2006**, *33*, 3171–3184; Part Special Issue: Operations Research and Data Mining. [CrossRef]

22. Emmert-Streib, F.; Dehmer, M. Identifying critical financial networks of the DJIA: Toward a network-based index. *Complexity* **2010**, *16*, 24–33. [CrossRef]

23. Bautin, G.A.; Kalyagin, V.A.; Koldanov, A.P.; Koldanov, P.A.; Pardalos, P.M. Simple measure of similarity for the market graph construction. *Comput. Manag. Sci.* **2013**, *10*, 105–124. [CrossRef]

24. Garas, A.; Argyrakis, P. Correlation study of the Athens Stock Exchange. *Phys. A Stat. Mech. Its Appl.* **2007**, *380*, 399–410. [CrossRef]

25. Vizgunov, A.; Goldengorin, B.; Kalyagin, V.; Koldanov, A.; Koldanov, P.; Pardalos, P.M. Network approach for the Russian stock market. *Comput. Manag. Sci.* **2014**, *11*, 45–55. [CrossRef]

26. Namaki, A.; Shirazi, A.H.; Raei, R.; Jafari, G.R. Network analysis of a financial market based on genuine correlation and threshold method. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 3835–3841. [CrossRef]

27. Bautin, G.A.; Kalyagin, V.A.; Koldanov, A.P. Comparative Analysis of Two Similarity Measures for the Market Graph Construction. In *Models, Algorithms, and Technologies for Network Analysis*; Goldengorin, B.I., Kalyagin, V.A., Pardalos, P.M., Eds.; Springer: New York, NY, USA, 2013; pp. 29–41.

28. Shirokikh, O.; Pastukhov, G.; Boginski, V.; Butenko, S. Computational study of the US stock market evolution: A rank correlation-based network model. *Comput. Manag. Sci.* **2013**, *10*, 81–103. [CrossRef]

29. Wang, G.J.; Xie, C.; Han, F.; Sun, B. Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 4136–4146. [CrossRef]

30. Kenett, D.Y.; Tumminello, M.; Madi, A.; Gur-Gershgoren, G.; Mantegna, R.N.; Ben-Jacob, E. Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *PLoS ONE* **2010**, *5*, e15032. [CrossRef] [PubMed]

31. Kalyagin, V.A.; Koldanov, A.P.; Koldanov, P.A.; Pardalos, P.M. Optimal decision for the market graph identification problem in a sign similarity network. *Ann. Oper. Res.* **2018**, *266*, 313–327. [CrossRef]

32. Faizliev, A.; Balash, V.; Vlasov, A.; Tryapkina, T.; Mironov, S.; Androsov, I.; Petrov, V. Analysis of the Dynamics of Market Graph Characteristics. In Proceedings of the Third Workshop on Computer Modelling in Decision Making (CMDM 2018), Saratov, Russia, 14–17 November 2018; Atlantis Press: Paris, France, 2019. [CrossRef]

33. Mahdi, K.; Almajid, A.; Safar, M.; Riquelme, H.; Torabi, S. Social Network Analysis of Kuwait Publicly-Held Corporations. *Procedia Comput. Sci.* **2012**, *10*, 272–281. [CrossRef]

34. Sankar, C.P.; Asokan, K.; Kumar, K.S. Exploratory social network analysis of affiliation networks of Indian listed companies. *Soc. Netw.* **2015**, *43*, 113–120. [CrossRef]

35. Battiston, S.; Catanzaro, M. Statistical properties of corporate board and director networks. *Eur. Phys. J. B* **2004**, *38*, 345–352. [CrossRef]

36. Vasques Filho, D.; O'Neale, D.R.J. Degree distributions of bipartite networks and their projections. *Phys. Rev. E* **2018**, *98*, 022307. [CrossRef]

37. Bargigli, L.; Giannetti, R. The Italian corporate system in a network perspective (1952–1983). *Phys. A Stat. Mech. Its Appl.* **2018**, *494*, 367–379. [CrossRef]

38. Sidorov, S.P.; Faizliev, A.R.; Balash, V.A.; Gudkov, A.A.; Chekmareva, A.Z.; Anikin, P.K. Company Co-mention Network Analysis. In *Computational Aspects and Applications in Large-Scale Networks*; Kalyagin, V.A., Pardalos, P.M., Prokopyev, O., Utkina, I., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 341–354. [CrossRef]

39. Balash, V.; Chekmareva, A.; Faizliev, A.; Sidorov, S.; Mironov, S.; Volkov, D. Analysis of News Flow Dynamics Based on the Company Co-mention Network Characteristics. In *Complex Networks and Their Applications VII*; Aiello, L.M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L.M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 521–533.

40. Sidorov, S.P.; Faizliev, A.R.; Balash, V.A.; Gudkov, A.A.; Chekmareva, A.Z.; Levshunov, M.; Mironov, S.V. QAP Analysis of Company Co-mention Network. In *Algorithms and Models for the Web Graph*; Bonato, A., Prałat, P., Raigorodskii, A., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 83–98.

41. Balash, V.A.; Faizliev, A.R.; Korotkovskaya, E.V.; Mironov, S.V.; Smolov, F.M.; Sidorov, S.P.; Volkov, D.A. The Evolution of Degree Distribution, Maximum Cliques and Maximum Independent Sets of Company Co-Mention Network over Time. *WSEAS Trans. Syst. Control.* **2019**, *14*, 97–103.

42. Mitra, G.; Mitra, L. (Eds.) *The Handbook of News Analytics in Finance*; John Wiley & Sons: Hoboken, NJ, USA, 2011.

43. Mitra, G.; Yu, X. (Eds.) *Handbook of Sentiment Analysis in Finance*; Albury Books: New York, NY, USA, 2016.

44. Sidorov, S.; Faizliev, A.; Balash, V. Measuring long-range correlations in news flow intensity time series. *Int. J. Mod. Phys. C* **2017**, *28*, 1750103. [CrossRef]

45. Donnat, C.; Holmes, S. Tracking network dynamics: A survey using graph distances. *Ann. Appl. Stat.* **2018**, *12*, 971–1012. [CrossRef]

46. Aleskerov, F.; Shvydun, S. Stability and Similarity in Networks Based on Topology and Nodes Importance. In *Complex Networks and Their Applications VII*; Aiello, L.M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L.M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 94–103.

47. Schieber, T.A.; Carpi, L.; Díaz-Guilera, A.; Pardalos, P.M.; Masoller, C.; Ravetti, M.G. Quantification of network structural dissimilarities. *Nat. Commun.* **2017**, *8*, 13928. [CrossRef] [PubMed]

48. Hammond, D.K.; Gur, Y.; Johnson, C.R. Graph diffusion distance: A difference measure for weighted graphs based on the graph Laplacian exponential kernel. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 419–422. [CrossRef]

49. Krackardt, D. QAP partialling as a test of spuriousness. *Soc. Netw.* **1987**, *9*, 171–186. [CrossRef]

50. Hubert, L. *Assignment Methods in Combinatorial Data Analysis*; Dekker: New York, NY, USA, 1987.

51. Dekker, D.; Krackhardt, D.; Snijders, T.A.B. Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika* **2007**, *72*, 563–581. [CrossRef] [PubMed]

52. Rossi, R.A.; Ahmed, N.K. Role Discovery in Networks. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 1112–1131. [CrossRef]

53. Bunke, H. Error correcting graph matching: on the influence of the underlying cost function. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 917–922. [CrossRef]

54. Messmer, B.T.; Bunke, H. A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 493–504. [CrossRef]

55. Bunke, H.; Dickinson, P.; Kraetzl, M.; Wallis, W. *A Graphtheoretic Approach to Enterprise Network Dynamics*; Birkhauser: Boston, MA, USA, 2007.

56. Fernández, M.L.; Valiente, G. A graph distance metric combining maximum common subgraph and minimum common supergraph. *Pattern Recognit. Lett.* **2001**, *22*, 753–758. [CrossRef]

57. Bunke, H.; Jiang, X.; Kandel, A. On the Minimum Common Supergraph of Two Graphs. *Computing* **2000**, *65*, 13–25. [CrossRef]

58. Gardiner, E.J.; Raymond, J.W.; Willett, P. RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs. *Comput. J.* **2002**, *45*, 631–644. [CrossRef]

59. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [CrossRef]

60. Dill, S.; Kumar, R.; Mccurley, K.S.; Rajagopalan, S.; Sivakumar, D.; Tomkins, A. Self-similarity in the Web. *ACM Trans. Internet Technol.* **2002**, *2*, 205–223. [CrossRef]

J. Open Innov. Technol. Mark. Complex. **2019**, *5*, 55

26 of 26

61. Borodin, A.; Roberts, G.O.; Rosenthal, J.S.; Tsaparas, P. Link Analysis Ranking: Algorithms, Theory, and Experiments. *ACM Trans. Internet Technol.* **2005**, *5*, 231–297. [CrossRef]

62. Papadimitriou, P.; Dasdan, A.; Garcia-Molina, H. Web graph similarity for anomaly detection. *J. Internet Serv. Appl.* **2010**, *1*, 19–30. [CrossRef]

63. Papadopoulos, A.; Manolopoulos, Y. Structure-Based Similarity Search with Graph Histograms. In Proceedings of the Tenth International Workshop on Database and Expert Systems Applications, DEXA 99, Florence, Italy, 3 September 1999.

64. Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment. *J. ACM* **1999**, *46*, 604–632. [CrossRef]

65. Blondel, V.D.; Gajardo, A.; Heymans, M.; Senellart, P.; Dooren, P.V. A Measure of Similarity Between Graph Vertices: Applications to Synonym Extraction and Web Searching. *SIAM Rev.* **2004**, *46*, 647–666. [CrossRef]

66. Heymans, M.; Singh, A.K. Deriving phylogenetic trees from the similarity analysis of metabolic pathways. *Bioinformatics* **2003**, *19*, i138–i146. [CrossRef]

67. Koutra, D.; Vogelstein, J.T.; Faloutsos, C. DELTACON: A Principled Massive-Graph Similarity Function. In Proceedings of the 2013 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, Austin, TX, USA, 2–4 May 2013; pp. 162–170. [CrossRef]

68. De Domenico, M.; Nicosia, V.; Arenas, A.; Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **2015**, *6*, 6864. [CrossRef] [PubMed]

69. Krackhardt, D. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Soc. Netw.* **1988**, *10*, 359–381. [CrossRef]

70. Rienties, B.; Héliot, Y.; Jindal-Snape, D. Understanding social learning relations of international students in a large classroom using social network analysis. *High. Educ.* **2013**, *66*, 489–504. [CrossRef]

71. Barnett, G.A.; Park, H.W.; Jiang, K.; Tang, C.; Aguillo, I.F. A multi-level network analysis of web-citations among the world's universities. *Scientometrics* **2014**, *99*, 5–26. [CrossRef]

72. Cantner, U.; Graf, H. The network of innovators in Jena: An application of social network analysis. *Res. Policy* **2006**, *35*, 463–480. [CrossRef]

73. Lee, W.J.; Lee, W.K.; Sohn, S.Y. Patent Network Analysis and Quadratic Assignment Procedures to Identify the Convergence of Robot Technologies. *PLoS ONE* **2016**, *11*, e0165091. [CrossRef] [PubMed]

74. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220. [CrossRef] [PubMed]

75. Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **1964**, *29*, 1–27. [CrossRef]

76. Borg, I.; Groenen, P.J.; Mair, P. *Applied Multidimensional Scaling and Unfolding*; Springer Briefs in Statistics; Springer International Publishing: Berlin/Heidelberg, Germany, 2018.

77. GOWER, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **1966**, *53*, 325–338. [CrossRef]