

Mutual Information–Based Hierarchies on Warsaw Stock Exchange

P. FIEDOR*

Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland

A popular method for network analysis of financial markets is a notable part of econophysics research. The networks created in such efforts are focused exclusively on linear correlations between stocks. While Pearson's correlation is the obvious starting point, it would be useful to look at its alternatives as to whether they provide improvements to this methodology, particularly given Pearson's correlation coefficient considers only a limited class of association patterns. We propose to use mutual information–based hierarchical networks, as mutual information is a natural generalisation of Pearson's correlation. We estimate mutual information using naive plug-in estimator as consistent bias is not harmful to this application, however other methods may also be used. We then transform the mutual information into an Euclidean metric and create minimal spanning trees and maximally filtered planar graphs. We apply this methodology to Warsaw Stock Exchange for years between 2000 and 2013, and comment on the differences with the standard methodology, as well as the structural changes on Warsaw Stock Exchange which the study reveals.

DOI: [10.12693/APhysPolA.127.A-33](https://doi.org/10.12693/APhysPolA.127.A-33)

PACS: 05.10.–a, 64.60.aq, 89.65.–s, 89.70.–a

1. Introduction

It is a common assumption in economics that prices change randomly [1]. This assumption is naturally followed by an investigation into whether the stochastic processes for different financial instruments are independent, or whether there are dependencies driven by hidden common economic factors driving these processes. Econophysicists [2–4] created a popular method for the construction of hierarchical networks based on single linkage hierarchical clustering [5–7], to investigate this question.

Different applications of hierarchical clustering use various similarity measures, but virtually all financial research uses Pearson's correlation coefficient and its derivatives. This is the simplest and most natural method, and it gives useful results [5, 8, 9]. However, while these measures of dependence are computationally efficient and their properties are well understood, they consider only a limited class of association patterns, such as linear and monotonically increasing functions. While Pearson's rho may be statistically insignificant even when there is a strong association pattern (for example, when $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = X^2$, then $\rho(X, Y) = 0$ even though they clearly not independent), from Jensen's inequality we know that the mutual information is equal to zero if and only if the two random variables under consideration are independent. Thus it satisfies one of the fundamental properties desired in a measure of dependence proposed by Rényi [10], while Pearson's rho does not. Since we know that financial data exhibits non-linear patterns (there exists a large body of literature presenting

non-linear behaviour of stock markets [11], currency exchange rates [12], and market indices [13]), we want to check whether using a non-linear dependence measure enhances the results. This leads us to proposing a method based on information theory to account for non-linear behaviour in the financial markets. From a more pragmatic perspective, we have compared Pearson's rho and mutual information for all pairs in the studied data (described below), and found many instances where Pearson's correlation coefficient is equal to zero, while mutual information is statistically significant (for example looking at data for the year 2000, around 30% of the pairs for which $\rho = 0 \pm 0.01$ (where median deviation from 0 equals 0.1 for the whole dataset) are characterised by mutual information above the median for the whole dataset).

To account for the non-linear patterns we have proposed a methodology based on Lempel–Ziv algorithm for data compression which asymptotically approximates mutual information rate [14]. In this study we use a simpler methodology, and extend the standard methodology by exchanging correlation with mutual information, which is a more general measure due to its information theoretic roots [15]. Clustering based on mutual information has indeed been successfully performed in other applications [16–18]. Mutual information can be interpreted as a measure of how much information the two studied processes share. Its estimation involves some problems [19, 20], which are not severe in our application however [14]. Particularly our method is not sensitive to consistent bias (thus we use naive plug-in estimator for simplicity), nonetheless there exist methods to account for the bias in mutual information estimation [20, 21].

In this paper we simplify the methodology presented in [14] and use hierarchical clustering based on mutual information to study interdependencies between financial instruments. Thus this study modifies the canonical

*e-mail: Pawel.F.Fiedor@ieee.org

methodology of hierarchical clustering for financial data by exchanging Pearson's correlation with mutual information (estimated using plug-in method). We create minimal spanning trees and planar maximally filtered graphs for Warsaw Stock Exchange for the years between 2000 and 2013. We show how this changes the results by comparison with correlation based methodology, and briefly discuss the resulting structural picture of the Warsaw Stock Exchange over the years.

2. Similarity measure and networks

The topological structure of financial networks is usually based on correlation coefficient of log price changes. To extend the similarity measure to account for non-linear dependencies we propose to use mutual information instead. Mutual information is most often defined in the context of Shannon's entropy [22], which is a measure of uncertainty of a random variable X :

$$H(X) = - \sum_i p(x_i) \log_2 p(x_i) \quad (1)$$

summed over all possible outcomes $\{x_i\}$ with respective probabilities of $p(x_i)$.

We define mutual information for two discrete random variables X and Y as:

$$I_S(X, Y) = H(X) + H(Y) - H(X, Y). \quad (2)$$

Mutual information measures information shared between the two variables, therefore both linear and non-linear dependencies, hence using it to describe dependencies on financial markets is natural. Mutual information is non-negative and $I_S(X, X) = H(X)$.

There is a large number of estimators of entropy and mutual information, a presentation of which can be found in Refs. [19, 23–26]. In this application we only compare the measure between pairs, thus we need not worry about consistent bias, which is the main problem with simple estimators. Therefore we use the plug-in estimator of entropy for the sake of simplicity. Such estimator is the entropy of the empirical distribution [19]:

$$\hat{H}_{emp}(X) = - \sum_{x \in X} \frac{\Lambda(x)}{n} \log \frac{\Lambda(x)}{n}, \quad (3)$$

where $\Lambda(x)$ is the number of data points having value x , and n is the sample size. Such entropy estimators are consistently biased downward.

Further, we convert mutual information to an Euclidean metric to have a distance matrix of a fully connected graph. We use the metric proposed in [7]. The quantity

$$d(X, Y) = H(X|Y) + H(Y|X) = H(X) + H(Y) - 2I_S(X, Y) \quad (4)$$

satisfies the Euclidean axioms [7]. But $d(X, Y)$ is not appropriate for all purposes as $I_S(X, Y)$ and $d(X, Y)$ are biased with regards to the size of clusters [27], since mutual information depends on the size of the studied sequences. Hence another, relative measure is defined by

being normalised through dividing by the total entropy. Then the quantity:

$$D(X, Y) = 1 - \frac{I_S(X, Y)}{H(X, Y)} = \frac{d(X, Y)}{H(X, Y)} \quad (5)$$

is a metric, with $D(X, X) = 0$ and $D(X, Y) \leq 1$ for all pairs (X, Y) .

We have introduced the similarity measure and the distance measure for fully connected graphs, but these graphs need to be filtered. Here we briefly discuss methods for producing such filtered networks, which allow for easier analysis of the most important information within the system. The distance matrix \mathcal{D} containing $D(X, Y)$ for all studied pairs is defined as above. From \mathcal{D} we create an ordered list \mathcal{S} , in which the distances are listed in decreasing order. The minimal spanning tree (MST) is created using the ordered list \mathcal{S} . Starting from the pair with the largest similarity measure D an edge is added to the graph between elements X and Y if and only if the graph obtained this way is still a forest or a tree [28]. After all appropriate links are added such graph is always reduced into a tree [28, 29]. Less constrained graphs can also be constructed, where the genus is fixed: $g = k$. Such graphs are created similarly: from the ordered list \mathcal{S} , starting from the pair with the largest similarity measure, we add an edge between that pair if and only if the resulting graph can still be embedded on a surface of genus $g \leq k$. Such graphs are less topologically restrictive than MST, always contains the relevant MST and also additional loops and cliques [28]. If $g = 0$ the resulting graph is planar [30]. Such graph is the simplest extension of the MST, and is called the planar maximally filtered graph (PMFG).

For our study we also need a measure of centrality of a node in a hierarchical network. We use the Markov centrality, the choice and nature of which is explained in detail in Ref. [31]. Here we only mention that centrality is one of the more important metrics in any network, as central nodes are considered important to the evolution of the system as a whole.

3. Empirical Application

To apply mutual information-based hierarchical networks in practice we use daily log price changes ($r_t = \ln(p_t/p_{t-1})$) for Warsaw Stock Exchange (GPW)[†] for stocks listed between the 3rd of January 2000 and the 5th of July 2013. Out of this dataset we choose 357 stocks which were listed during at least 1000 consecutive days. Those stocks are then divided into 26 sectors according to the Warsaw market classification. We divide these time series into yearly sub-series, and for each year disregard the stocks with incomplete data (not listed all throughout given year). Therefore for every year we have a varying number of studied stocks.

[†]<http://bossa.pl/notowania/metastock/>

For the purpose of estimating the entropy and mutual information these time series are discretised into four distinct states. The states represent quartiles, therefore each state is assigned the same number of data points. This design leaves out any unnecessary arbitrary parameters, which could affect the results and conclusions reached while using the data. This and similar experimental setups have been used in similar studies [14, 32, 33] and proved to be very efficient [34]. Alternatively one could use permutation entropy on real valued data. In our experience this approach does not work well with financial data however.

We estimate the entropy and mutual information for time series for every company in every studied year (using `infotheo` package for R), and on this basis we calculate the similarity measure D , thus allowing us to create both minimal spanning trees and planar maximally filtered graphs. We also calculate Markov centrality for each node in each created network. For comparison, we also create the same networks using standard correlation-based measure, and calculate Markov centrality on this basis as well.

All the results have been calculated for specific companies for specific years between 2000 and 2013. The results are then grouped by sector to allow for easy understanding of the structural changes happening on the Warsaw market during the studied years.

4. Discussion

As the nature of such large networks is that they are not easily analysed visually, we have calculated Markov centrality for each node in each calculated network to analyse the differences between the studied structures. In Table we present, in respective columns, Pearson’s correlation coefficients between correlation–based and mutual information–based (i) minimal spanning trees and (ii) planar maximally filtered graphs, as well as between minimal spanning trees and planar maximally filtered graphs based on (iii) correlation coefficient and (iv) mutual information. Minimal spanning trees and planar maximally filtered graphs are very similar (ρ around 0.8), which is to be expected since PMFG contains MST and adds to it only most relevant additional information. But the correlation– and mutual information–based structures are also strongly correlated at around 0.6 and 0.7 for MST and PMFG respectively. This shows that the presented methodology is not completely abandoning the preceding methodology, but instead it adds the non-linear dependencies, which has been the intention. It is worth noting that mutual information has been shown to perform better at recreating sector structure of the market from prices [35] (important feature of this method, as it cannot be reproduced by simulating a market) than the approach based on correlation [14]. In this study the stocks are divided into as many as 26 sectors, which means that such analysis would run into under-sampling issues with the number of studied stocks and be unreliable. It is therefore not presented, but our studies

of the Warsaw market show that the results would be similar to these reported for New York’s market [14].

TABLE

This table is based on Markov centrality calculated for each node in each constructed network. Below we present, in respective columns, Pearson’s correlation coefficients between correlation–based and mutual information–based (i) minimal spanning trees and (ii) planar maximally filtered graphs, as well as between minimal spanning trees and planar maximally filtered graphs based on (iii) correlation coefficient and (iv) mutual information. Minimal spanning trees and planar maximally filtered graphs are very similar (ρ around 0.8), which is to be expected since PMFG contains MST and adds to it only most relevant additional information. But the correlation– and mutual information–based structures are also strongly correlated at around 0.6 and 0.7 for MST and PMFG respectively. This shows that the presented methodology is not completely abandoning the preceding methodology, but instead it adds the non-linear dependencies.

Year	$\rho \sim I_S$		MST \sim PMFG	
	MST	PMFG	ρ	I_S
2000	0.495	0.559	0.793	0.796
2001	0.754	0.873	0.888	0.891
2002	0.680	0.704	0.825	0.878
2003	0.534	0.718	0.808	0.818
2004	0.523	0.692	0.855	0.853
2005	0.472	0.625	0.817	0.763
2006	0.555	0.727	0.886	0.816
2007	0.773	0.640	0.888	0.913
2008	0.606	0.760	0.883	0.925
2009	0.715	0.696	0.895	0.863
2010	0.666	0.732	0.867	0.886
2011	0.422	0.490	0.901	0.819
2012	0.615	0.669	0.823	0.799
2013	0.382	0.433	0.811	0.718
Average	0.585	0.666	0.853	0.838

We also look at the results aggregated into sectors. Warsaw Stock Exchange groups its stocks into 26 specific sectors. We present the same analysis for minimal spanning trees and planar maximally filtered graphs, but ignore the results for Pearson’s correlation for the sake of compact presentation. As following from the above, these would not be drastically different. We analyse average and aggregate Markov centrality by sector in the studied years. The average presents the changes relating to the whole sector, while the aggregate shows more clearly the impact of a given sector to the whole market, but loses sight of the changes happening to the whole sector, while being more dependent on the important stocks.

In Fig. 1a the average Markov centrality by sector for mutual information-based MST for all studied years is presented, while in Fig. 1b we have presented the aggregate Markov centrality by sector for the same. The average is calculated as the mean of Markov centralities for the stocks belonging to a given sector, while the aggregate is calculated as the sum of Markov centralities for the stocks belonging to a given sector. Both are presented

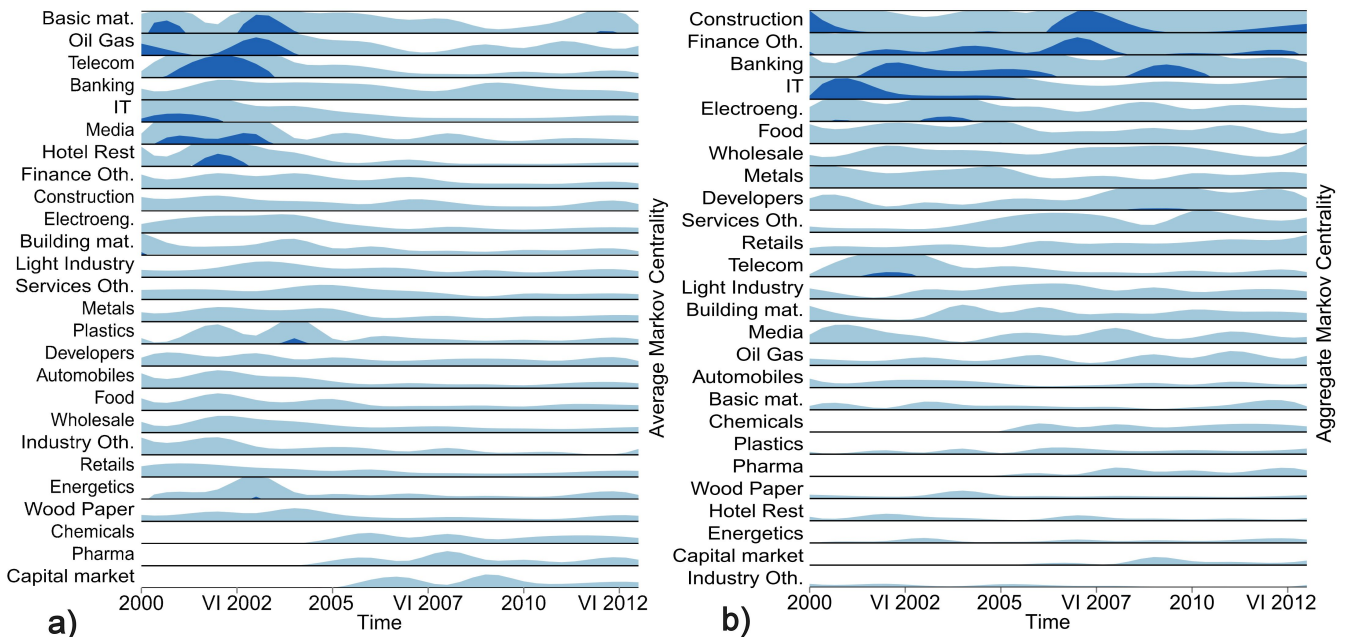


Fig. 1. Average (a) and aggregate (b) Markov centrality by sector for MST based on mutual information. The importance of various sectors clearly changes over time, in accordance with the market history, particularly the history of problems within these sectors. Smoothing was performed using Forsythe, Malcolm and Moler’s cubic spline interpolation.

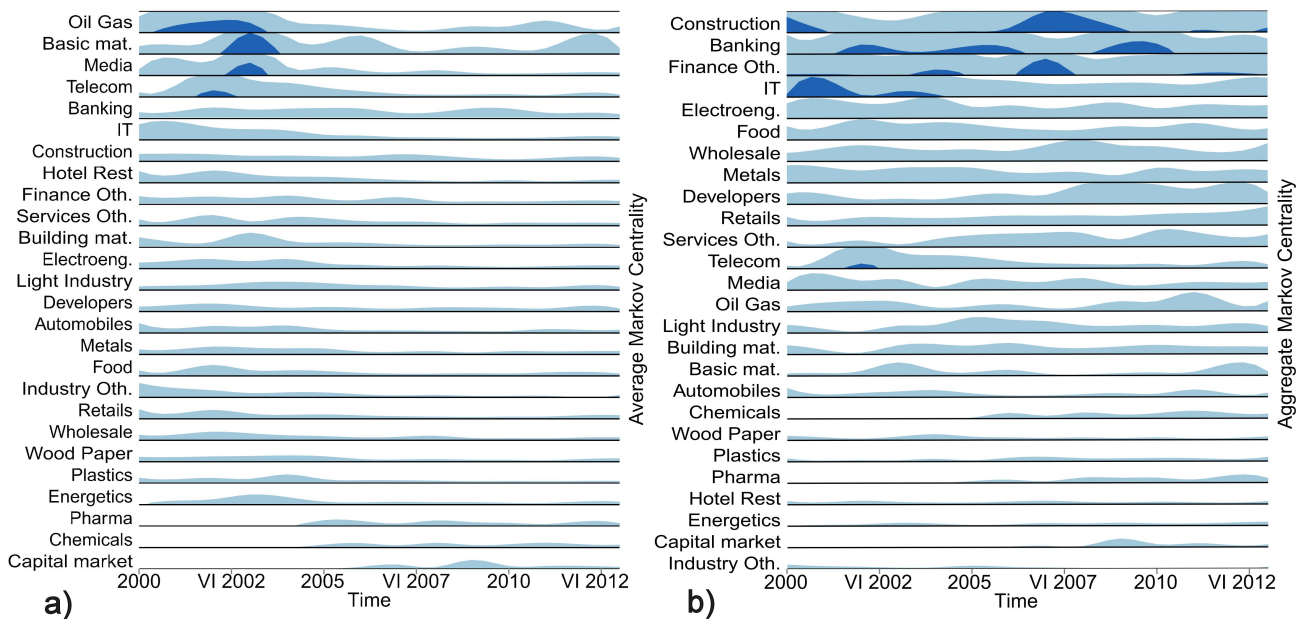


Fig. 2. Average (a) and aggregate (b) Markov centrality by sector for PMFG based on mutual information. These results are very similar to the ones obtained using minimally spanning trees, hinting that it may not be necessary to use PMFG in such analyses. Smoothing was performed using Forsythe, Malcolm and Moler’s cubic spline interpolation.

as horizon graphs [36] for compact presentation. Particularly interesting are the aggregate Markov centrality results. We can see banking sector’s significance rising around the time of the financial crisis of 2008–2010, rising significance of the construction sector around the 2007 subprime mortgage crisis (even though the crisis has been in the US the financial markets are linked well enough for such conditions to be affecting markets globally),

and rising significance of the IT sector around the dot-com bubble of 2001. We thus conclude that such analysis presents valuable informations about the financial markets. Interestingly, the significance of the banking sectors has fallen steadily hinting that the crisis lead by the banks is at an end. The significance of construction sector is on the rise, which may potentially hint what the analysts’ attention should be aimed at.

In Fig. 2a the average Markov centrality by sector for mutual information–based PMFG for the studied years is presented, while in Fig. 2b we have presented the aggregate Markov centrality by sector for the same. The results are not significantly different from the above, hence we will not discuss them in detail. It does seem apparent that it may be sufficient to use the MST given these results.

5. Conclusions

In this paper we have simplified a methodology for creating hierarchical networks based on information theoretic approach (incorporating non-linear dependencies) in financial markets presented in [14]. We have obtained results which are close to the results obtained using standard methodology based on Pearson’s correlation. As there are no computational or other problems with the presented methodology, while the results are similar or better than those obtained using Pearson’s correlation, we propose that this method should be used in network analyses of financial markets. We have also commented on the apparent structural changes on the Warsaw Stock Exchange as revealed by the networks created for the years between 2000 and 2013. It is of particular importance that the banking crisis seems to be at an end, but the construction sector is becoming a more significant part of the market in the last couple of years. Further studies should look into other measures of non-linear dependency, particularly ones which do not require discretisation of the financial data.

Acknowledgments

The author would like to thank Thomas Kern for his help with reproducing horizon graphs with ggplot2 in R.

References

- [1] P.A. Samuelson, *Ind. Manag. Rev.* **6**, 41 (1965).
- [2] B.B. Mandelbrot, *J. Bus.* **36**, 394 (1963).
- [3] L.P. Kadanoff, *Simulation* **16**, 261 (1971).
- [4] R.N. Mantegna, *Physica A* **179**, 232 (1991).
- [5] R.N. Mantegna, *Eur. Phys. J. B* **11**, 193 (1999).
- [6] A.K. Jain, R.C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs 2000.
- [7] A. Kraskov, H. Stogbauer, R.G. Andrzejak, P. Grassberger, *Europhys. Lett.* **70**, 278 (2005).
- [8] M. Tumminello, T. Aste, T. Di Matteo, R. N. Mantegna, *Eur. Phys. J. B* **55**, 209 (2007).
- [9] S. Drożdż, F. Grummer, F. Ruf, J. Speth, *Physica A* **294**, 226 (2001).
- [10] A. Rényi, *Acta Math. Acad. Sci. H.* **10**, 441 (1959).
- [11] D. Hsieh, *J. Financ.* **46**, 1839 (1991).
- [12] D. Hsieh *J. Bus.* **62**, 339 (1989).
- [13] J.A. Scheinkman, B. Lebaron, *J. Bus.* **62**, 311 (1989).
- [14] P. Fiedor, *Phys. Rev. E* **89**, 052801 (2014).
- [15] T. Cover, J. Thomas, *Elements of Information Theory*, John Wiley & Sons 1991.
- [16] Fang Zhou, Jieyue He, Wei Zhong, in: *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, IEEE 2007, p. 1051.
- [17] Fang Zhou, Jieyue He, Wei Zhong, Y. Pan, *Int. J. Comput. Biol. Drug Des.* **2**, 187 (2009).
- [18] A.C. Muller, S. Nowozin, C.H. Lampert, *Pattern Recognition, Lect. Notes Comp. Sci.*, Springer, Berlin 2012, Chap. “Information Theoretic Clustering Using Minimum Spanning Trees”.
- [19] L. Paninski, *Neural Comput.* **15**, 1191 (2003).
- [20] M. Palus, V. Komarek, T. Prochazka, Z. Hrnčir, K. Sterbova, *IEEE Eng. Med. Biol.* **20**, 65 (2001).
- [21] A. Papana, D. Kugiumtzis, *Int. J. Bifurcat. Chaos* **18**, 4197 (2009).
- [22] C.E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [23] J. Beirlant, E.J. Dudewicz, L. Györfi, E.C. van der Meulen, *Int. J. Math. Stat. Sci.* **6**, 17 (1997).
- [24] G. Darbellay, I. Vajda, *IEEE T. Inform. Theory* **45**, 1315 (1999).
- [25] C.O. Daub, R. Steuer, J. Selbig, S. Kloska, *BCM Bioinformatics* **5**, 118 (2004).
- [26] W. Nemenman, W. Bialek, R. de Ruyter van Steveninck, *Phys. Rev. E* **69**, 056111 (2004).
- [27] M. Li, J.H. Badger, X. Chen, S. Kwong, P. Kearney, H. Zhang, *Bioinformatics* **17**, 149 (2001).
- [28] M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, *PNAS* **102**, 10421 (2005).
- [29] M. Tumminello, F. Lillo, R.N. Mantegna, *J. Econ. Behav. Organ.* **75**, 40 (2010).
- [30] G.L. Miller, *J. Comb. Theory* **43**, 25 (1987).
- [31] P. Fiedor, *Physica A* **413**, 180 (2014).
- [32] N. Navet, Shu-Heng Chen, in: *Natural Computing in Computational Finance*, Eds. T. Brabazon, M. O’Neill, Vol. 100 of *Studies in Computational Intelligence*, Springer 2008.
- [33] P. Fiedor, in: *Proceedings of the IEEE Computational Intelligence for Financial Engineering & Economics 2014*, Eds. A. Serguieva, D. Maringer, V. Palade, R.J. Almeida, IEEE, London 2014, p. 247.
- [34] R. Steuer, L. Molgedey, W. Ebeling, M. Jiménez-Montaño, *Eur. Phys. J. B* **19**, 265 (2001).
- [35] G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, *Phys. Rev. E* **68**, 046130 (2003).
- [36] J. Heer, N. Kong, M. Agrawala, in: *ACM Human Factors in Computing Systems*, ACM, Boston 2009, p. 1303.