# A NOTE ON THE CONSTRUCTION OF STOCK MARKET NETWORKS

**Tomáš Výrost**


**Abstract**

The paper investigates stock market networks within CEE-3 countries (the Czech Republic, Hungary and Poland) and Germany. The usual approach in constructing such networks is based on the calculation of stock return correlations, their transformation to distances and creation of a graph with edge weights set to the calculated distances. To reduce the complexity of the structure, a reduction in the number of edges is conducted, mostly by utilizing the minimum spanning tree (MST) of the graph. This paper considers alternative approaches to the subgraph selection problem and compares the topological properties of the ensuing graph structures. Specifically, the paper considers three approaches: the widely used minimum spanning tree (MST), planar maximally filtered graph proposed by Tumminello, Lillo & Mantegna (2005), and the "winner-take-all" approach suggested by Tse, Liu & Lau (2010). After constructing graphs using the three approaches, the paper then focuses on centrality and graph centralization. It is shown that although the MST does not have clear economic justification, not much information is lost by its use when studying graph centrality.



**Key words:** stock market networks, minimum spanning trees, planar graphs

**JEL Code:** L14, G1


## Introduction

During the last decades, much research has been published on the subject of networks in many different fields. The use of graph theory to capture the structure of the underlying network has traditionally been used in operations research to analyze network flows and transportation problems. Over time, the methodology was used in other areas as well, such as social networks (Wasserman & Faust, 1994), the World Wide Web, the internet, neurological studies of the human brain, networks of critical infrastructure such as power distribution lines and others.

This paper focuses on the networks based on stock markets returns in three Central and Eastern European countries (CEE-3 – the Czech republic, Hungary and Poland), which can be described as emerging markets and Germany as a representative of a developed market. The problem we are addressing deals with the choice of the algorithm that can be used to construct the network. The usual approach to the construction of stock market networks is to assign each stock a vertex and use correlation between the return series to assign edge weights. By doing so, we obtain a complete graph, as correlation is defined for all pairs of stocks. To reduce the number of edges, and thus make the subsequent analysis more tractable, the usual choice is to reduce the network to a suitable subgraph. The subgraph that is used most often in the literature is the minimum spanning tree (MST), which results in a severe reduction in the retained number of edges. The purpose of this paper is to consider alternative subgraphs that may be used and describe their properties in an empirical analysis of stock market networks of CEE-3 and Germany.

# 1 The subgraphs for the construction of stock market networks

In this section, we consider several structures that can be used to describe a stock market network. We start by the most frequently used minimum spanning trees (MST), and then introduce the alternatives – planar maximal filtered graphs (PMFG) of Tumminello, Lillo & Mantegna (2005) and the "winner-take-all" approach of Tse, Liu & Lau (2010).

## 1.1 Minimum spanning trees

The first type of network we use is the one based on the minimum spanning tree, as described in the seminal work of Mantegna (1999). To allow for the analysis of time-varying properties of the network, rolling correlation coefficients are calculated over the subsamples of the return series, producing a series of correlation matrices. These are in turn transformed into distances by the formula (Mantegna, 1999):

$$d_t(i,j) = \sqrt{2(1-\rho_{ijt})} \tag{1}$$

where $d_t(i,j)$ is the calculated distance between stocks $i$ and $j$ ($i, j = 1, 2, \ldots, N$; where $N \in \mathbb{N}$ is the number of stocks) and $\rho_{ijt}$ is the rolling correlation coefficient between the returns series of stock $i$ and $j$ at time $t$. The minimum spanning tree is than constructed from the calculated distances $d_t(i,j)$, which are used as edge weights. As a spanning tree, the resulting network represents an acyclic connected graph on $N$ vertices, having $N - 1$ edges. As the complete graph on $N$ vertices has $N(N-1)/2$ edges, the reduction is quite significant.

Despite the dominant use of MSTs in the literature, their use raises some potentially important questions. First, the issue of the reduction in *number of edges* needs to be addressed. As the MST retains only a small number of the original edges, does the MST still capture the relevant properties of the network?

Second, the *kind of the edges* retained might be questioned as well. Even though the MST is constructed in a way that prefers the edges with small weights (and thus high correlations), the requirement of creating a tree (a connected graph) ultimately may lead to the inclusion of edges with small correlation. For example, if we had a vertex with extremely low correlations to all its neighbors, the connectedness of the MST forces one of such edges to be incorporated into the MST. As the MST always has $N - 1$ edges, this might lead to the retention of small correlations instead of larger ones (with other vertices). This property is an immediate consequence of the choice of using an MST.

Third, the choice of MST lacks easy *economic justification.* Although creating a spanning tree is straightforward in graph theory, its choice (particularly with respect to the preceding remarks on the nature of correlations in the MST) does not follow any prior economic reasoning. Retaining only a small number of edges, not even the largest ones, the necessity to satisfy a connectedness requirement that has no economic counterpart might call the whole analysis to question, as the choice of the network creation algorithm clearly has implications on further analysis.

### 1.2 Planar maximally filtered graph

To present an alternative to the use of MSTs and allow for a larger number of edges in the resulting network, Tumminello, Lillo & Mantegna (2005) proposed the planar maximally filtered graph (PMFG).

The construction of a PMFG follows a process, in which one starts with the number of vertices, corresponding to the number of stocks and an ordered list of correlations between them. Edges are added one by one in the decreasing order of correlation, satisfying a condition that the edge that is to be added does not break planarity of the graph. Similarly as in a MST, the PMFG omits some edges based on a graph property – the MST had to by acyclic, the PMFG is planar. As PMFG is a simple graph, the number of edges can be obtained as a consequence of Euler's formula as $3N - 6$.

Although the PMFG retains a larger number of edges than a MST, its design principle is similar, and thus a similar question arises regarding the justification for its choice – it is not immediately clear, why planarity presents a desirable economic property in a stock market network.

### 1.3 The "winner-take-all" approach

Another simple approach was suggested by Tse, Liu & Lau (2010). Instead of selecting edges based on a graph property (tree structure, planarity), the edges are selected based on the criterion of correlations exceeding a fixed threshold (hence the abbreviation THR for such graphs). The advantage of this approach lies in the simplicity of the rule, and in the fact that it does not prescribe a structure to the network – conversely, it allows for the explorative study of the structure induced by the data. However, the approach raises a new issue, that needs to be addressed – the choice of the threshold level.

## 2 Data and methodology

The dataset consisted of 50 stocks traded on the prime stock exchange in the Czech Republic, Poland and Hungary (CEE-3), as well as Germany[1]. The CEE-3 countries have many economic, as well as historical links. The addition of Germany to the sample has the advantage of allowing for the analysis of interaction of CEE-3 markets with a large developed market. Germany has also the closest major market geographically. The economic ties to the CEE-3 also make this a natural choice.

Our sample includes the constituents of the leading local stock indices, traded between January 2003 and October 2012. Daily closing prices were obtained from the Thomson Reuters Datastream. Most studies on stock markets networks construct rolling correlations from stock returns that are converted to distances by the equation given in (1). As financial time series usually exhibit autocorrelation and heteroscedasticity, the analysis in this paper is based on standardized residuals from models dealing with these effects. The data and the modeling stage are essentially the same as in Výrost, Lyócsa & Baumöhl (2013). To create the models, we have first checked the stationarity of each weekly return series with a version of the KPSS stationarity test as proposed by Hobijn, Franses & Ooms (2004). At significance level of 0.1 we were unable to reject the null of mean stationarity for all of the analyzed

[1] For the *Czech Republic*, the stocks used included Erste group bank (ERSTE), Philip morris CR (PM), ČEZ (CEZ), Komercni banka (KB), Unipetrol (UNI), Telefónica CR (O2), for *Hungary* Egis pharmaceuticals (EGIS), Est media (EST), MOL (MOL), Magyar telekom (MTK), OTP bank (OTP), PannErgy (PAE), Richter Gedeon (REG), Synergon (SYN), for *Poland* KGHM (KGHM), Bank Polska Kasa Opieki (PEO), Polski Koncern Naftowy Orlen (PKN), Telekomunikacja Polska (TPS), Asseco Poland (ACP), Bank Handlowy w Warszawie (BHW), BRE Bank (BRE), Boryszew (BRS) and for *Germany* Adidas (ADS), Allianz (ALV), BASF (BAS), Bayerische Motoren Werke (BMW), Bayer (BAYN), Beiersdorf (BEI), Commerzbank (CBK), Continental (CON), Daimler (DAI), Deutsche Bank (DBK), Deutsche Boerse (DB1), Deutsche Post (DPW), Deutsche Telekom (DTE), E.ON (EOAN), Fresenius Medical Care (FME), Fresenius SE & Co KGaA (FRE), HEICO Corporation (HEI), Henkel AG & Co. (HEN3), Infineon Technologies (IFX), K+S Aktiengesellschaft (SDF), Linde Aktiengesellschaft (LIN), Deutsche Lufthansa (LHA), Merck KGaA (MRK), Munich RE (MUV2), SAP (SAP), Siemens Aktiengesellschaft (SIE), ThyssenKrupp AG (TKA), Volkswagen (VOW3).

series. We then went on to fit a suitable ARMA($p$, $q$) – GARCH(1, 1) models, to account for both effects in each of the returns series. The ARMA parameters were evaluated up to 5 lags. As for the specification of the GARCH part, the following models were considered: GARCH, AVGARCH, NGARCH, GJR-GARCH, APARCH, NAGARCH, TGARCH, FGARCH and CSGARCH. All selected models were required to show no significant autocorrelation and conditional heteroscedasticity in standardized residuals in up to 25 lags (about 5% of the sample), tested by the Ljung-Box test. Of all models matching the above mentioned conditions, the optimal model was chosen according to the Bayesian information criterion[2]. The models were used to obtain standardized residuals, which were then used to calculate 52-week Pearson product moment rolling correlation coefficients, which were then turned into distances by using the equation (1).

The creation of MSTs is well documented, see e.g. Prim (1957). As for the PMFGs, they were created using the procedure described in the previous section by adding edges according to the magnitude of correlation, making sure the graph remains planar. To verify planarity, we used an implementation of the Boyer & Myrvold (2004) planarity test.

The creation of THR graphs requires the choice of a threshold parameter. For the purposes of this paper, the threshold was chosen as the value of the correlation coefficient that would be significant at 5% on 52 observations[3], which is the window size for the calculation of rolling correlations.

## 3 Results

Before analyzing time-varying properties of the graphs, we have first calculated all three graphs from the whole sample, that is, based on the correlation for the entire available period. The resulting graphs are shown in Fig. 1.

As the vertex sets of all graphs remains the same, the only difference is presented by the number of edges. The MST is shown to have the least edges, but is therefore also easier to interpret. A strong clustering by country is noticeable, particularly among German stocks.

The PMFG is a planar graph, although it is not shown as such, to make the graphs easier to compare. One may note that there is an increase of edges not only between stocks traded within the same country, but also between countries. Lastly, the THR is a graph that retains most edges. This is a consequence of a rather low threshold we have chosen. Again, the density of edges between the German stocks is clearly visible from the graph.

[2] Full details on selected models can be found in Výrost, Lyócsa & Baumöhl (2013).
[3] The value for the correlation coefficient to be significant was 0.27324.

It is also clear, that the THR will contain vertices with higher degrees – as the graph has more edges, the vertex degrees will go up. In fact, the THR contains vertices of degree more than 40 (among 50 vertices).

We next turn our attention to another graph property, expressed in terms of centrality. Centrality has several measures, which measure the way a vertex belongs to the "center" of the graph.

In our paper, we use three measures: the vertex degree (number of incident edges of a vertex), closeness (the inverse of the average length of the shortest paths to all the other vertices in the graph) and betweenness (the number of geodesics going through a given vertex). Tab. 1 shows the vertices in each graph that have the highest centrality measures (only the first three highest ranking groups of vertices are shown[4]).

**Tab. 1: The groups of vertices with highest centrality**

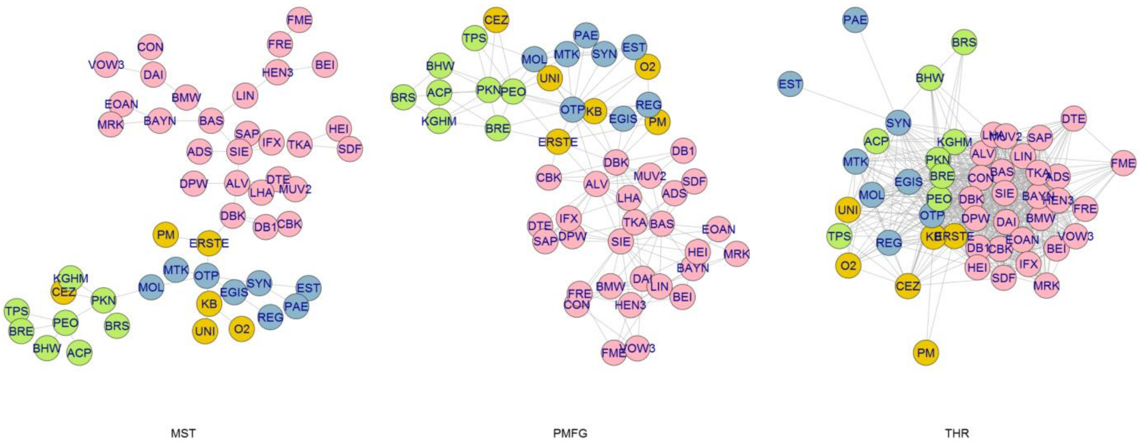
| Centrality | Graph | 1. | 2. | 3. |
|---|---|---|---|---|
| Vertex Degree | MST | **ALV OTP** SIE | **PEO** PKN | BAS **DBK** |
| | PMFG | SIE | **OTP** | **ALV** |
| | THR | **PEO** | **BRE DBK ERSTE KB OTP** PKN | BAS DPW SIE |
| Closeness | MST | **ALV** | **DBK** | **ERSTE** |
| | PMFG | **ALV** | **DBK** | **OTP** |
| | THR | **DBK** | BAS | **PEO** |
| Between-ness | MST | **OTP** | **ALV** | SIE |
| | PMFG | **OTP** | SIE | **ALV** |
| | THR | SYN | SIE | BAS |

Source: Own calculations.

First, one might see that when measured by the vertex degree, there are numerous ties in the graphs. This is understandable for MST, as the vertex degrees take only small number of values. Another noteworthy fact shown in Tab. 1 is that all stocks marked in bold are either banks or financial companies (such as insurance company Allianz – ALV). The result holds even if we use other centrality measures. Hence, these results hint of the importance of financial institutions in the linkages between markets – the stocks in the table come from all countries in the sample.

[4] We have also conducted a study for graphs based on rolling windows, which exhibited similar matches between the top three central vertices among MST, PMFG and THR.

**Fig. 1: The minimum spanning tree (left), planar maximal filtered graph (middle) and a threshold-based graph of the whole sample (right)**




Source: own calculation

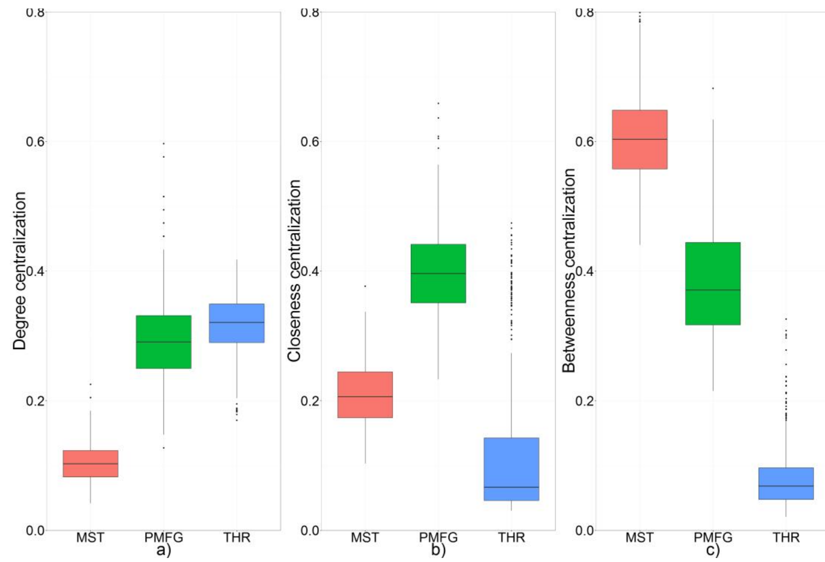
To pursue the analysis of centrality in a time-varying context, we have analyzed the centrality measures using graphs created from rolling correlations. The centrality measures we have discussed so far were all calculated for single vertices in a graph. In order to characterize a graph as a whole, we have used the so called centralization measures. Centralizations are graph-level scores that are based on the differences of the centrality measure of a central vertex and all other vertices, normalized by the maximum such value obtainable on a graph with similar structure. We were thus been able to calculate centralizations for all graphs created using rolling correlations. The statistical properties of centralization measures are shown in Fig 2.

The results indicate that vertex degree centralization is much smaller for MST than in both other graphs (a consequence of small number of edges, and low vertex degrees).

When compared with closeness and betweenness, we see that the centralizations for THR are getting much smaller. As betweenness is based on the number of geodesics passing through a vertex, a graph with many edges may have multiple paths between any two vertices. The opposite is the case with MST – betweenness centralization in a tree structure is quite high.

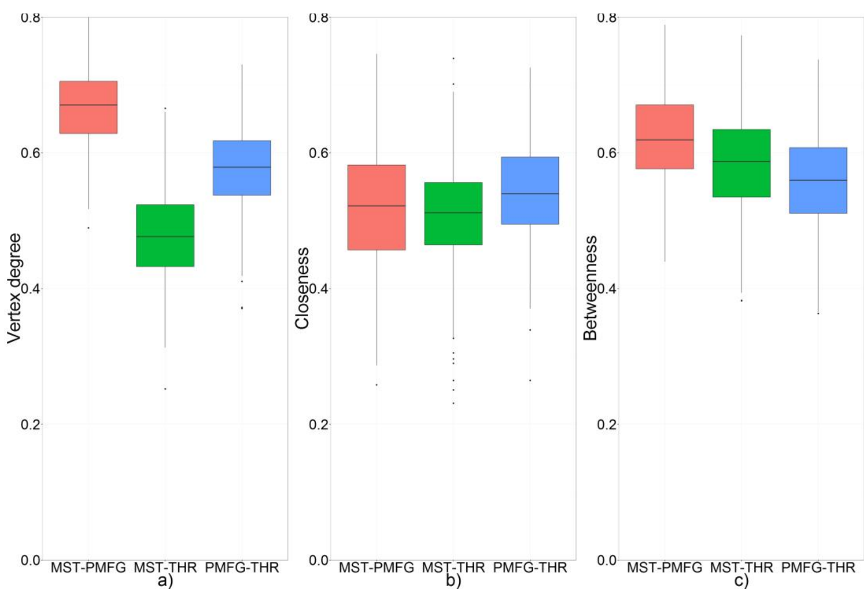
Lastly, we were interested in the relations of the ranking based on centrality of vertices in analyzed graphs. To analyze this aspect, we have calculated the centrality measures for each vertex in every graph type and each time. For example, we created a MST and PMFG at time *t*, and calculated the centrality measures for all vertices. To test whether the different graph structures influenced the centrality results, we wanted to compare the centrality scores assigned to each vertex (effectively creating vertex ranks based on centrality). To analyze whether the scores correspond to each other, we have calculated Kendal τ nonparametric correlation coefficient. The statistical properties of these correlation coefficients for all graph types are shown in Fig. 3. The results are interesting, because closeness and betweenness centrality shows remarkable similarities across graph types. In all cases the correlations are fairly high.

**Fig. 2: Boxplots of the centralization measures**




Source: own calculation

**Fig. 3: Boxplots of the Kendall correlation coefficients**




Source: own calculation

## Conclusion

In this paper we have dealt with the alternative ways to create a stock market network based on weekly returns. As the structure most frequently used (MST) raises some concerns, we have used an empirical study to compare it to two alternatives.

We have shown that the alternative structures, the planar maximal filtered graph and a threshold-based approach lead to structures with more edges than is the case with MST, possibly retaining more information in the network. The analysis we conducted on all networks dealt primarily with centrality, as it allows for the identification of the most "important" vertices in a graph (and hence companies).

First, we looked at the vertices with highest centrality measures. The results were very similar, both over the whole sample as well as over rolling windows. On the other hand, the overall character of the graph itself is different, as is expressed by the graph centralization.

Second, not much information is lost when using an MST to calculate rankings based on centrality – the rankings are quite similar for all structures.

To conclude, although the MST does not have clear economic justification, our findings suggest that not much information is lost by its use when studying centrality. Thus, the MST remains a viable alternative to other structures.

## Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11. The author would also like to thank Štefan Lyócsa and Eduard Baumöhl for their insightful comments.

## References

Baumöhl, E., & Výrost, T. (2000). Stock Market Integration: Granger Causality Testing with respect to Nonsynchronous Trading Effects. *Czech Journal of Economics and Finance*, *60*(5), 414-425.

Boyer, J. M., & Myrvold, W. J. (2004). On the cutting edge: Simplified O (n) planarity by edge addition. *Journal of Graph Algorithms and Applications*, *8*(3), 241-273.

Coelho, R., Gilmore, C. G., Lucey, B., Richmond, P., & Hutzler, S. (2007). The evolution of interdependence in world equity markets—Evidence from minimum spanning trees. *Physica A: Statistical Mechanics and its Applications*, *376*, 455-466.

Dias, J. (2012). Sovereign debt crisis in the European Union: A minimum spanning tree approach. *Physica A: Statistical Mechanics and its Applications*, *391*(5), 2046-2055.

Hobijn, B., Franses, P. H., & Ooms, M. (2004). Generalizations of the KPSS-test for stationarity. *Statistica Neerlandica*, *58*(4), 483-502.

Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., & Ben-Jacob, E. (2010). Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS one*, *5*(12), e15032.

Lyócsa, Š., Výrost, T., & Baumöhl, E. (2012). Stock market networks: The dynamic conditional correlation approach. *Physica A: Statistical Mechanics and its Applications*, *391*(16), 4147-4158.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, *11*(1), 193-197.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell system technical journal*, *36*(6), 1389-1401.

Tse, C. K., Liu, J., & Lau, F. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, *17*(4), 659-667.

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10421-10426.

Tumminello, M., Lillo, F., & Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, *75*(1), 40-58.

Výrost, T., & Lyócsa, Š., & Baumöhl, E. (2013). Time-varying Network Structure of Cross-correlations Among Stock Markets of CEE-3 and Germany. Forthcomming.

## Contact

Tomáš Výrost

Faculty of Business Economics in Košice

University of Economics in Bratislava

Tajovského 13

041 30 Košice, Slovak republic

E-mail: tomas.vyrost[at]euke.sk