Statistical Standard, Methodology and Application in Data Management and Usage

Optimal Filter to Analyze the Topological Network of Stocks Market

Maman Abdurachman Djauhari

Message to the audiences

This talk is to introduce a new theory to analyze the behavior of stocks market. It is an improvement of our previous theory published in **Djauhari and Gan (2015)** and appeared in the journal of Econophysics "*Physica A: Statistical Mechanics and its applications*" **(Q2, IF: 2.243, H Index: 126, More than 200 reads in RG)**

Two examples (NYSE and KLSE) will be presented to illustrate its advantages.

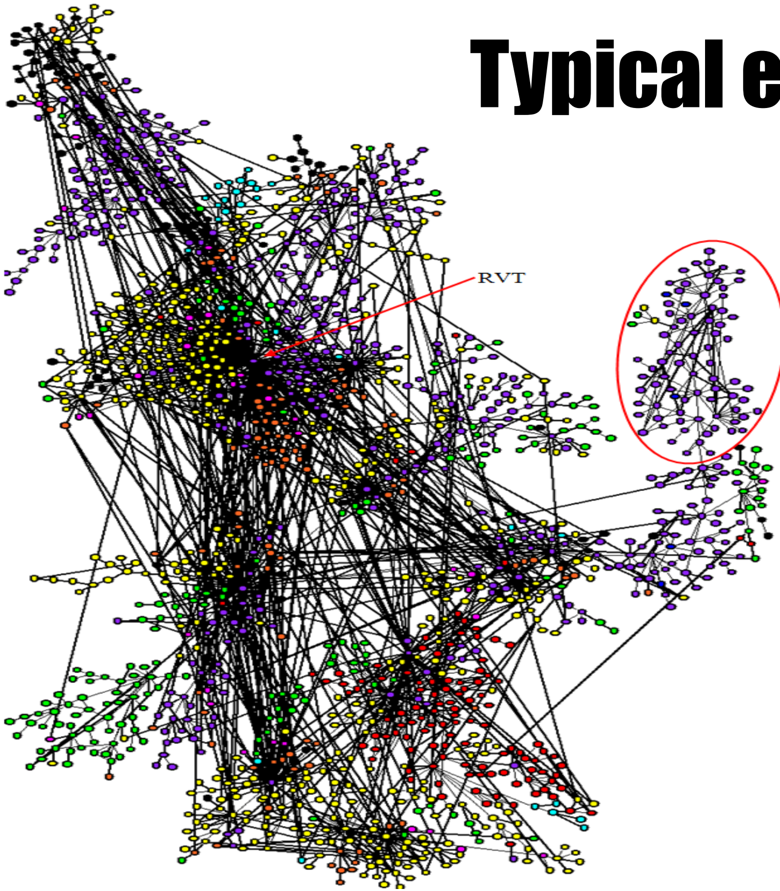
Highlights

- Stocks market is a complex system (consisting of the stocks and their relationships/interactions). By using GTA, stocks market behavior is represented in the form of a complex network.
- The larger the number of stocks, the higher the complexity level of market behavior. To reduce the complexity level, the network is filtered.
- The most adopted filter is not optimal to quantify and visualize the most important transmission of stock's behavior to the others'.
- A theory of optimal filter is then introduced and a method to find it is developed.
- Two examples will be presented; one comes from NYSE and the other one from KLSE.

Problem & Idea

- We show that the most adopted filter, MST, is not optimal. Therefore, MST-based network topology provides a misleading economic interpretation.
- In this talk, we introduce a theory of optimal MST and then develop a method to find it.
- To implement this theory, we use the notion of the forest of all MSTs (Forest, in short).
- Once an optimal MST has been found, the economic interpretation of the corresponding network topology is derived by means of usual tools such as network centralities and the power-law of degree distribution.

Typical example



Complex network of 1515 stocks traded at NYSE from 2005 – 2014. The links represent their inter-correlations (data were collected from www.nasdaq.com).

How to find an optimal network topology?

Main advantages of Forest

Why we use Forest? As we will show,

- It simplifies the process to find an optimal MST and lead us to a uniqueness theorem of MST and to a robust filter.
- It leads to an efficient algorithm to find MST; more efficient than Kruskal's algorithm and Prim's that usually used in stocks network analysis.
- Forest provides a direct computation of SDU.

NOTE: If MST is to visualize the most important transmission of a stock's behavior to the others', SDU is to represent the hierarchical taxonomy of the evolution of the most interacted stocks.

Unpleasant property of MST-based filter

Almost certain that, in practice, a degeneracy of correlation coefficients is present. This makes,

- MST not unique. It is then hard to believe that the MST issued from the most widely used algorithm is optimal in representing the market behavior.
- It depends on nodes ordering.

To handle this problem, in the next slides we introduce a theory of optimal MST.

A theory of optimal MST

I. Criteria

Let F denote the Forest in our study. We say that M^*, an MST in F, is optimal if it satisfies these three conditions,

1. $D(M^*) = \min_{M \text{ in } F} \{D(M)\}$ where $D(M)$ is the diameter of M.
2. $SPL_G(M^*) = \min_{M \text{ in } F} \{SPL_G(M)\}$ where

   $$SPL_G(M) = \sum_{i \text{ in } M} \sum_{j \text{ in } M} PL_{ij} ,$$

   and PL_{ij} is the length of the path between stocks i and j in M. In other words, $SPL_G(M)$ is the total path lengths in M.
3. $SPL_S(M^*) = \min_{M \text{ in } F} \{SPL_S(M)\}$ where

   $$SPL_S(M) = \sum_{k=1,\ldots,K} \sum_{i \text{ in } B(k)} \sum_{j \text{ in } B(k)} PL_{ij}$$

   and K is the number of business sectors, and $B(k)$ is the smallest sub-tree of M consisting of all stocks in the k-th business sector.

NOTE:

These criteria are simpler than those proposed in **Djauhari and Gan (2015)** because we omit the condition on the number of leaves.

We can show that the number of leaves in *F* and that in any MST are the same.

NOTE:

For computation purpose, we can show that

$$SPL_G(M) = (n-1)\sum_{i \text{ in } M} 1/cc(i)$$

where $cc(i)$ is the closeness centrality score of the stock i (Borgatti (2005)).

As the transmission of stock's behavior to the others is fast in an optimal MST, the value of $SPL_G(M^*)$ is then minimum.

This criterion reflects the connectedness of stocks in the same business sector; they are strongly connected. From economic point of view, they are homogeneous and tend to be clustered. Therefore, in an optimal MST, its value is minimal.

II. Selection Procedure

- Construction of the Forest
- Construction of Optimal MST

Construction of the Forest

Let D be the distance matrix related to the correlation matrix as defined in Mantegna and Stanley (2000). We construct the sequence $D, D^2, D^4, D^8, \ldots$ where matrix multiplication is defined in the usual sense but multiplication and summation of two real numbers a and b are defined as $\max\{a,b\}$ and $\min\{a,b\}$.

This sequence has a limit that can be achieved in less than $1.4427\ln(n)$ iterations (Djauhari (2012)).

Let $d^+(i,j)$ be the (i,j)-th element of this limit. If Δ is the adjacency matrix of the Forest, then its (i,j)-th element is,

$$\delta(i,j) = \begin{cases} 1; d(i,j) - d^+(i,j) = 0 \text{ and } i \neq j \\ 0; d(i,j) - d^+(i,j) \neq 0 \text{ or } i = j \end{cases}$$

where $d(i,j)$ is the (i,j)-th element of D.

NOTE (Uniqueness theorem of MST):

As an immediate result of this construction, we have a uniqueness theorem of MST: "*If N is the sum of all elements of* Δ, *then MST is unique if and only if* $N = 2(n-1)$." This theorem is important to test whether MST in the network is unique or not.

Construction of Optimal MST

- An important property of *F* is that an MST is optimal in the market if and only if it is optimal in *F*.
- Let us remove all leaves from *F* and denote *H* the remaining sub-graph. Evidently, *H* could contain no leaf or at least one. In the latter case, we repeat the process of leaves removal until we get the sub-graph *H** consisting no leaf.
- If we need *m* repetitions and *M* is an MST in *H**, then the union of *M* and all removed leaves in the *m*-th repetition and all removed leaves in the -th repetition and … all removed leaves in the 1st removal, is an MST of the stocks network.

- As a result, an optimal MST can simply be selected heuristically from *H**.

Case of NYSE

- The 100 most traded stocks at NYSE during the whole year 2012 are analyzed based on Closing Price. Only n = 98 stocks are further analyzed due to the completeness of data.
- The daily data were downloaded from http://finance.yahoo.com. and www.nyse.com/about/listed/nyid_components.shtml.
- The correlation matrix is defined in such a way that the *i*-th row (and column) corresponds to the *i*-th stock listed in these links (accessed on July 25, 2013).
- MST issued from Kruskal's algorithm (Kruskal network, in brief) and the Forest are presented in Figure 1 and 2.

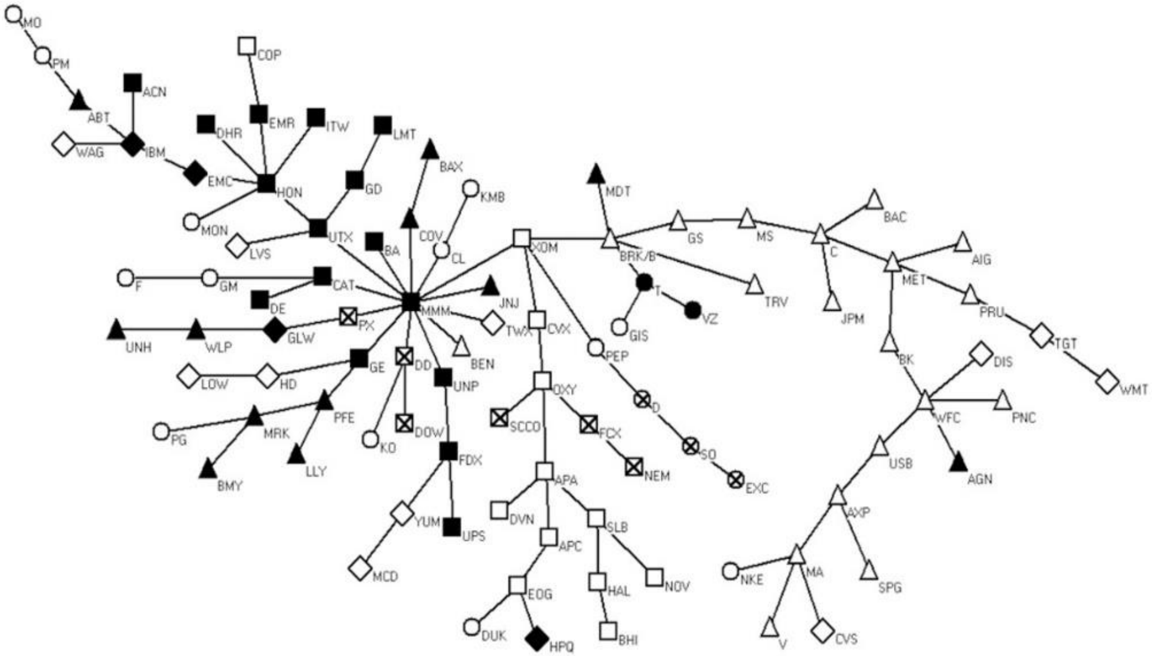


Figure 1: Kruskal's network topology of 98 most traded stocks at NYSE

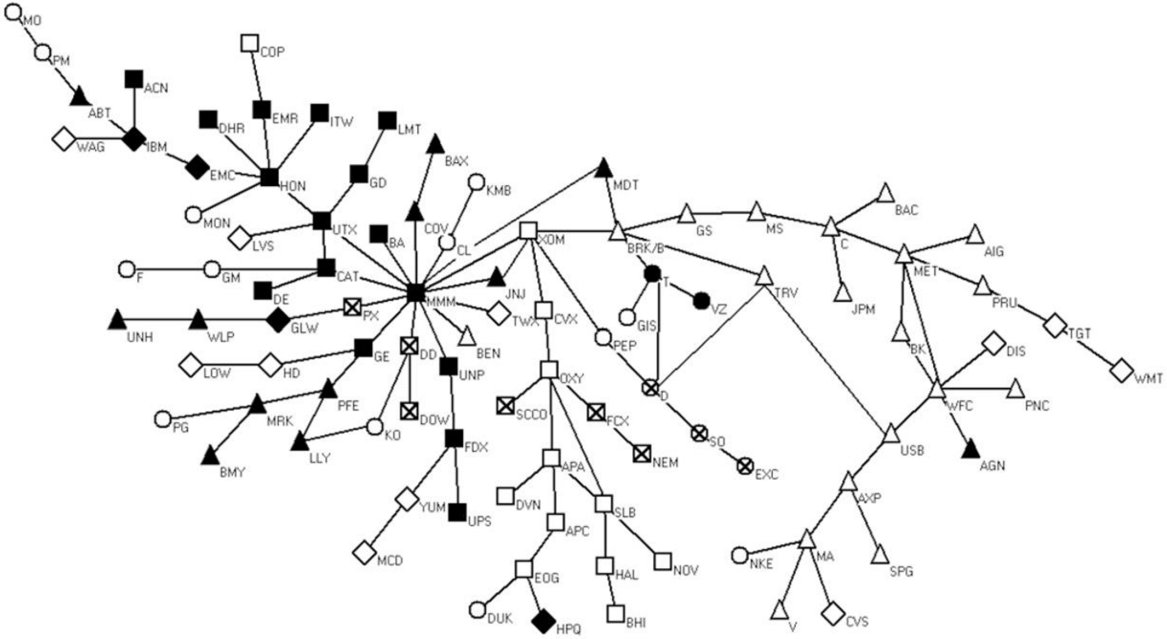


Figure 2: Forest of 98 most traded stocks at NYSE

NOTE:

Figure 2 indicates that it is hard to believe that Kruskal's network (MST-based filter) is optimal. In fact, as we will show in Figure 3, the sub-graph H^* of Forest contains many circuits. Therefore, there are so many MSTs therein.

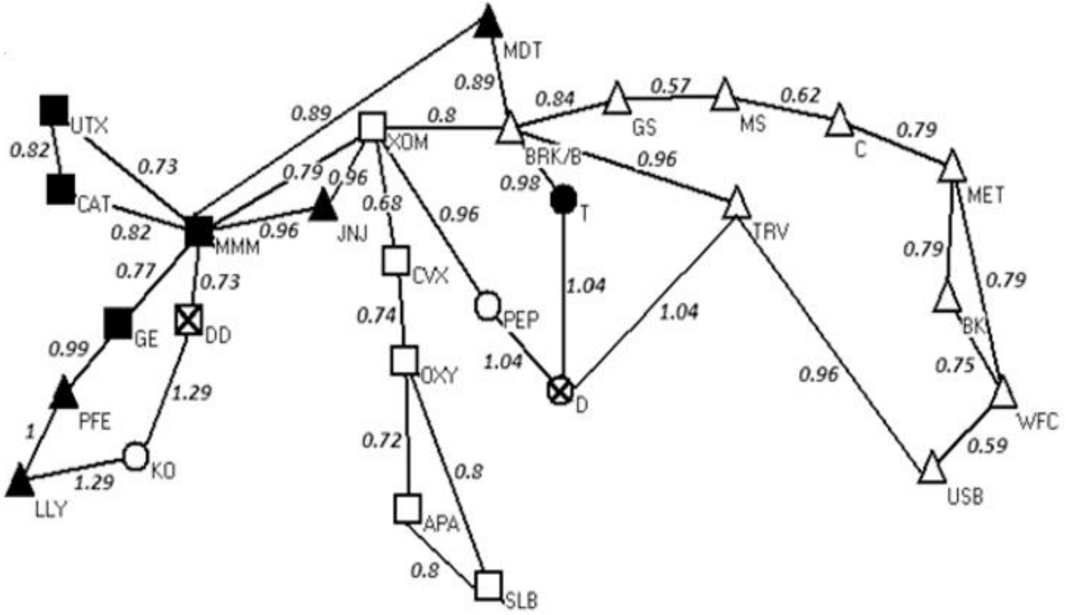


Figure 3: H* - The largest sub-graph of Forest without leaf

NOTE:

After a deep study on H^*, we find that the Kruskal's MST in H^* which leads to an optimal MST in the network is as showed in Figure 4.

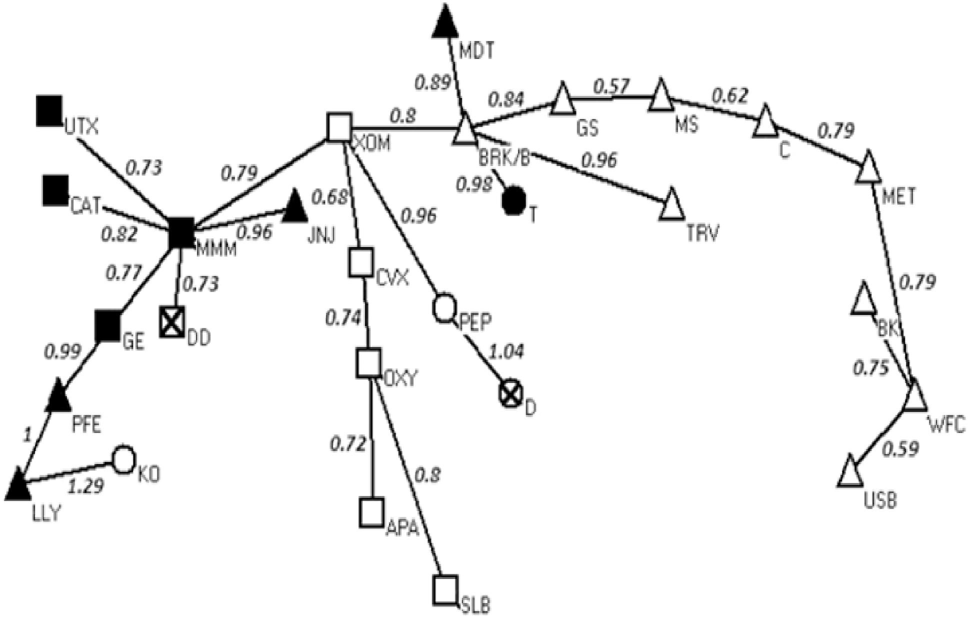


Figure 4: Kruskal's MST of H* leading to optimal network topology

Optimal network topology of NYSE

- The union of MST of H^* in Figure 4 and all removed leaves in the last repetition and all removed leaves in the second last repetition and … all removed leaves in the 1^{st} removal, is an optimal MST that we are looking for.

- The result is in Figure 5.

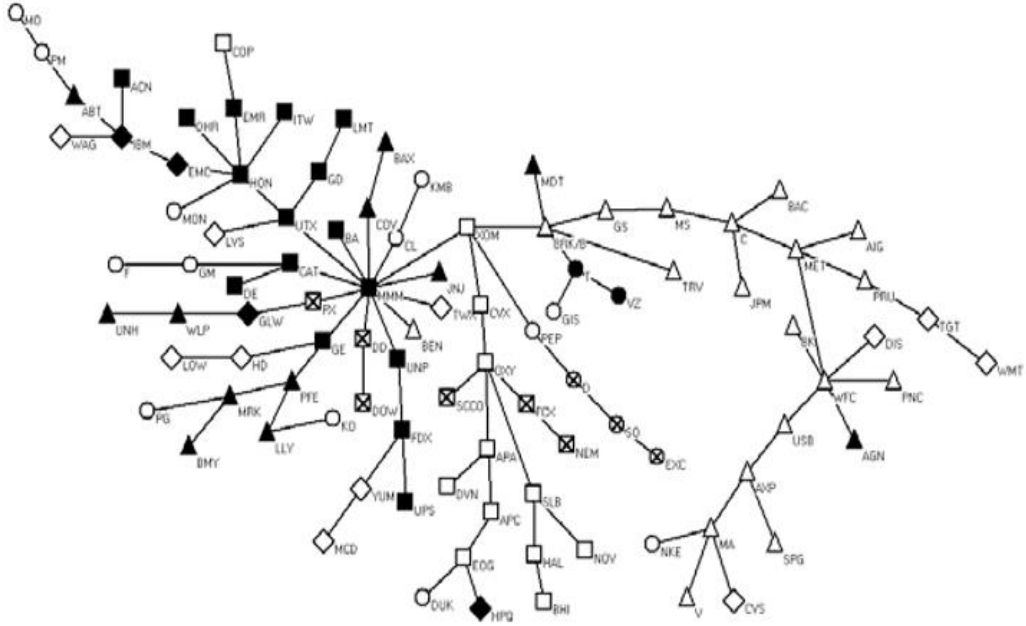


Figure 5: Optimal network topology

NOTE:

For comparison study, we also present the **worst network** in Figure 6.

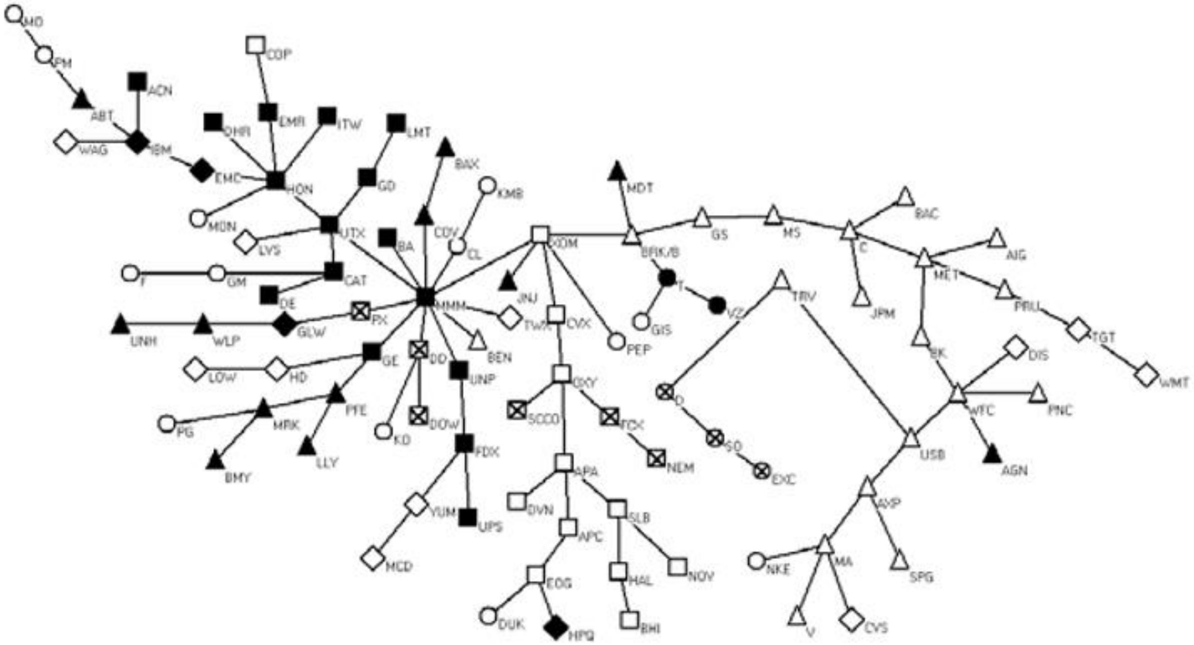


Figure 6: Worst network topology

Network topology property

Table 1: The characteristics of network topology

Characteristic	Network topology		
	Optimal	Kruskal	Worst
Diameter	18	19	20
SPL_G	67058	69094	73314
SPL_S	5780	5956	5974

Table 2: The characteristics of power-law

Characteristic	Network topology		
	Optimal	Kruskal	Worst
Constant c	0.5852	0.6035	0.6715
Exponent γ	1.7102	1.7822	1.8619

NOTE:

These two tables show that the theory of optimal MST introduced here finds its justification. All characteristics (diameter, SPL_G, SPL_S, c and γ) of optimal network have smallest value.

Case of KLSE

- We analyze 100 most traded stocks at KLSE from January 2007 to January 2009. Here also, the data are about Closing Price.
- Only 90 stocks are further analyzed. The data for the other 10 stocks are not complete.
- MST issued from Kruskal's algorithm is in Figure 7. The nodes are colored according to their business sector; blue (*trading and services*), black (*technology*), brown (*properties*), cyan (*industrial products*), olive green (*finance*), pink (*plantation*), purple (*IPC*), red (*construction*), and yellow (*consumer products*).

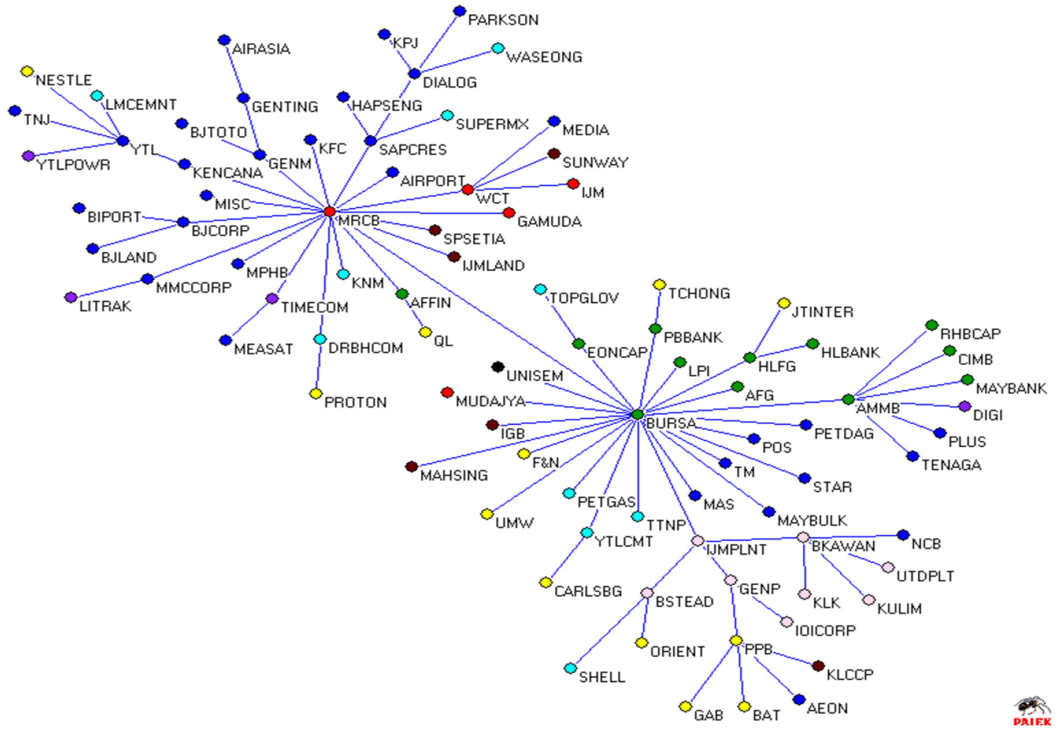


Figure 7: Kruskal's network topology of 90 most traded stocks at KLSE

NOTE:

- At a glance, in that figure, two big clusters reveal. The upper cluster is led by MRCB (Malaysian Resources Corporation Berhad) in the *Construction* sector.
- Meanwhile, the lower one is led by BURSA (Bursa Malaysia Berhad) in the *Finance* sector.
- Unfortunately, it can be shown (Show!) a degeneracy of the correlation coefficients is present which means that MST is not unique. Thus, it is questionable that the network in this figure is optimal.
- By using the same analysis as in NYSE case, the readers are invited to find an optimal network for KLSE.

References

- Bonanno, G., Caldarelli, G., Lillo, F. and Mantegna, R.N. (2003) "Topology of correlation-based minimal spanning trees in real and model markets," *Physical Review E*, 68, 046130-046133.
- Borgatti, S.P. (2005) "Centrality and network flow," *Social Networks*, 27, 55-71.
- Brida, J.G. and Risso, W.A. (2008) "Multidimensional minimal spanning tree: The Dow Jones case," *Physica A*, 387, 205-5210.
- Djauhari, M.A. (2012) "A robust filter i stock networks analysis," *Physica A*, 391, 5049-5057.
- Djauhari, M.A. and Gan, S.L. (2013) "Minimal spanning tree problem in stock network analysis: An efficient algorithm," *Physica A*, 392, 2226-2234.
- Djauhari, M.A. and Gan, S.L. (2015) "Optimality problem of network topology in stocks market analysis," *Physica A*, 419, 108-114.
- Mantegna, R.N. (1999) "Hierarchical structure in financial markets," *European Physical Journal B*, 11, 193-197.
- Mantegna, R.N. and Stanley, H.E. (2000) *An Introduction to Econophysics: Correlation and Complexity in Finance*, Cambridge University Press, London.
- Oh, G. (2014) "Grouping characteristics of industry sectors in financial markets," *Physica A*, 395, 261-268.
- Ulusoy, T., Keskin, M., Shirvani, A., Deviren, B., Kantar, E. and Dönmez, C.C. (2012) "Complexity of major UK companies between 2006 and 2010: Hierarchical structure method approach," *Physica A*, 391, 5121-5131.

THANK YOU FOR YOUR INTEREST

For further discussion please contact:
maman_djauhari@yahoo.com
Or visit:
Centre for Research in Statistics and Data Analysis
(The Bandung Statistical Research Institute)
Tjahaja Bina Statistika Indonesia, Ltd,P, (Commanditaire Vennootschap)
Jl. Kanayakan A-15, Bandung 40135, Indonesia
http://www.tbs-i.com/