On Stock Market Dynamics through Ultrametricity of Minimum Spanning Tree

Hokky Situngkir*
(hs@compsoc.bandungfe.net)
Dept. Computational Sociology
Bandung Fe Institute

Yohanes Surya
(yohaness@centrin.net.id)
Surya Research International

(May 10, 2005)

Abstract

We analyze the evolving price fluctuations by using ultrametric distance of minimally spanning financial tree of stocks traded in Jakarta Stock Exchange 2000-2004. Ultrametricity is derived from transformation of correlation coefficients into the distances among stocks. Our analysis evaluates the performance of ups and downs of stock prices and discovers the evolution towards the financial and economic stabilization in Indonesia. This is partly recognized by mapping the hierarchical trees upon the realization of liquid and illiquid stocks. We remind that the methodology is useful in two terms: the evaluation of spectral market movements and intuitively understanding for portfolio management purposes.

Keywords:
ultrametricity, minimum spanning tree, liquidity, Jakarta Stock Exchange.

*Corresponding author; homepage: http://www.bandungfe.net/hs

1 Introduction

It is always interesting to see many objects correlated in networks - in a space where we have the ability to visualize the whole by terms of each. Obviously, the visualization can give us a new perspective on the system as presented by the idea of ultrametric space in which our objects laid upon. Directly or indirectly motivated by the spin glass theory - discovering the ultrametric distances among configurations of states exhibited by statistical cluster analysis [10] - there have been some previous works on wide range of fields analyzing in the similar way, e.g.: the hierarchical taxonomy of set of stocks [9] that can intuitively help the portfolio management [13], network exhibited by several Gross National Product time correlations [11], the taxonomy of political parties in multiparty system [14], biology inspired innovation theory [5].

Some analyses of the ultrametric space in financial data (e.g.: some stocks in DJIA) have given extraction of interesting information regarding the clustering among same business sectors of traded stocks indicating that most of the same sectors depicts similar fluctuations over time[9]. However, further comprehensive analysis showed that the exhibited tree of financial data does not contain merely standard economic factors, but also psychological factor of investors [12]. Apparently, this does not at all reduce the advantage of the methodology to build investment portfolio, moreover it gives us an objective relations among traded stocks comprised not merely by intuitive business sectors but any information implicitly present in time series data of each stock. The paper presented here is intended to show another information gathered when the methodology is implemented to see some Indonesian stocks traded in Jakarta Stock Exchange (JSX) over the time period of 2000 to 2004.

The research on comparing any hierarchical trees evolving over each trading years is specifically introduced to develop understanding about the the fluctuations of stock prices over particular periods in JSX. It is clear that this can give useful input for investment society trying to build portfolio or even for the bourse company to have feedback about the spectrum of traded stocks as evaluation about the market.

First of first, the paper presents several notions related to ultrametricity and hierarchical clustering tree of traded stocks that is followed by some implementations to empirical data. In advance, we elaborate a discussion about how we use the results to have a look into the liquidity signatures among stocks. There is an expectation that this can improve broad uses of the methodology in further financial analysis and practical usage in our endeavor to reveal some levels of complexity in financial market [3].

2 The Ultrametricity in Financial Analysis

The concept of ultrametricity is rooted in physics as described in [8]. In accordance with the analytical steps derived in [13], we can see how financial analysis adopt the ideas of ultrametric space as described in this section.

If we denote the price of stock i as $p_i(t, \delta t)$, then the logarithmic return can be written,

$$r_i(t) = \ln p_i(t + \delta t) - \ln p_i(t) \tag{1}$$

and the correlation coefficients can be calculated as

$$c_{ij} = \frac{\langle r_i(t,\delta t) r_j(t,\delta t)\rangle_T - \langle r_i(t,\delta t)\rangle_T \langle r_j(t,\delta t)\rangle_T}{\sqrt{\langle r_i^2(t,\delta t)\rangle_T - r_i(t,\delta t)\rangle_T^2}\sqrt{\langle r_j^2(t,\delta t)\rangle_T - r_j(t,\delta t)\rangle_T^2}} \tag{2}$$

with values of the coefficient indicating correlations as

$$c_{ij} = \begin{cases} 1, & \text{completely correlated} \\ 0, & \text{uncorrelated} \\ -1, & \text{completely uncorrelated} \end{cases} \tag{3}$$

Furthermore, in terms of geometric representation, we can normalize the return

$$\hat{r_i}(t, st) = \frac{r_i(t,\delta t) - \langle r_i(t,\delta t)\rangle_T}{\sqrt{\langle r_i^2(t,\delta t)\rangle_T - \langle r_i(t,\delta t)\rangle_T^2}} \tag{4}$$

that gives us the correlation coefficient of

$$c_{ij} = \langle \hat{r_i}(t,\delta t)\hat{r_i}(t,\delta t)\rangle_T \tag{5}$$

In advance, we combine the N records of normalized price return into N-dimensional vector of

$$\overrightarrow{r_i} = \frac{1}{\sqrt{N}}[\hat{r_i}(t_0,\delta t), \hat{r_i}(t_1,\delta t), \ldots \hat{r_i}(t_N,\delta t)] \tag{6}$$

with

$$N = \frac{T}{\delta t} + 1, \text{and } t_k = t - k\delta t \tag{7}$$

Here, we can write the correlation coefficient as

$$c_{ij} = \overrightarrow{r}_i \overrightarrow{r_j} \tag{8}$$

Eventually, we can write the Phytagorean equation in order to have the distance between stock i and j

$$d_{ij}^2 = \overrightarrow{r_i}^2 - \overrightarrow{r_j}^2 \tag{9}$$

Since we have the unit length as seen in equation (8), $\overrightarrow{r_i}^2 = c_{ii} = 1$, then

$$d_{ij} = \sqrt{2(1 - \overrightarrow{r_i}\overrightarrow{r_j}} \quad = \quad \sqrt{2(1 - c_{ij})} \tag{10}$$

From eq.(10), we can directly have the distance between two stocks from their correlation coefficient in the euclidean space. In this euclidean space, it has three properties, i.e.:

$$\begin{array}{ll} (a) & d_{ij} = 0 \Leftrightarrow i = j \\ (b) & d_{ij} = d_j i \\ (c) & d_{ij} \leq d_{ik} + d_{kj} \end{array} \tag{11}$$

The first property in eq.(11) shows that if we have two stocks completely correlated ($c_{ij} = 1$), they are not separated in any distance ($d_{ij} = 0$), and vice versa. Furthermore, the second property shows that stock i correlated to j in the same value as stock j correlated to the i - implying that the distance between two stocks is the same as from each any of the two stocks ($c_{ij} = c_{ji}$ and $d_{ij} = d_{ji}$). The third one is well known as triangular inequality relies on the Phytagorean equation of (9).

By having the euclidean space of the stock prices, we are moving to extract the ultrametric space from it. An ultrametric space is the space where all distances within is ultrametric. The ultrametric distance is understood as distance with the first and second properties described in eq.(11) and the third one replaced by stronger inequality of

$$(d) \quad d_{ij}^{ult} \leq max(d_{ik}^{ult}, d_{kj}^{ult}) \tag{12}$$

From the eq.(10) we have n objects linked together by $n-1$ edges based on their correlation coefficients. Thus, there are some ultrametric spaces can be taken from it. But from all of those, we have a simple subdominant ultrametric space defined as the one with minimally spanning tree. Here, the minimum spanning tree is a tree with the smallest sum of all of distances.

A simple way to have the locally minimum spanning tree is by applying the Kruskal's Algorithm [6]. The Kruskal's algorithm is simply described by the following algorithm:

```
begin
    do while (all vertex in the graph)
        Find the least edge in the graph;
        Mark it with any given color, e.g.: blue;
        Find the least unmarked (uncolored) edge in the graph
                that doesn't close a colored or blue circuit;
        Mark this edge red;
```

```
    end;
The blue edges form the desired minimum spanning tree;
end;
```

3 Ultrametric Distances of Stocks in JSX

We choose 91 stocks traded in JSX of highest market capitalization average in the period of October 1st - December 31st 2004. They come from various business sectors of: (a) agriculture sector, (b) miscellaneous industry, (c) basic industry and chemicals, (d) mining, (e) financial sector, (f) property, real estate, and building construction, (g) infrastructure, utilities, and transportation, (h) trade, services, and investment, and (i) consumer goods industry.

On price return matrix of $T \times M$ of M stocks over T hourly time interval from January 2000 to December 2004, we calculate the correlation coefficients among them. Then the yielded correlation matrix is used to calculate the distance matrix among the stocks by the eq.(10). If we assume that 1 year is equivalent to 250 trading days and one day consists of 6 trading hours, then a month is equivalent with 126 trading hours. In order to see how the correlation coefficients and the distances among stocks evolve through time, we make month-wise (of $\tau = 126$ hours) time windows and calculate the probability distribution functions of each iteratively. Thus, approximately we have 46 overall number of windows to investigate how the correlations among stocks evolved over time.

Our correlation coefficients forms an $m \times M$ correlation matrix (C) and we have a sequence of $M(M-1)/2$ correlation coefficients and we calculate the four moments of distribution in each respective time window. The first moment of distribution is known as the *mean correlation coefficient* that is calculated as

$$\mu(1) = \frac{1}{N(N-1)/2} \sum_{i,j} c_{ij} \tag{13}$$

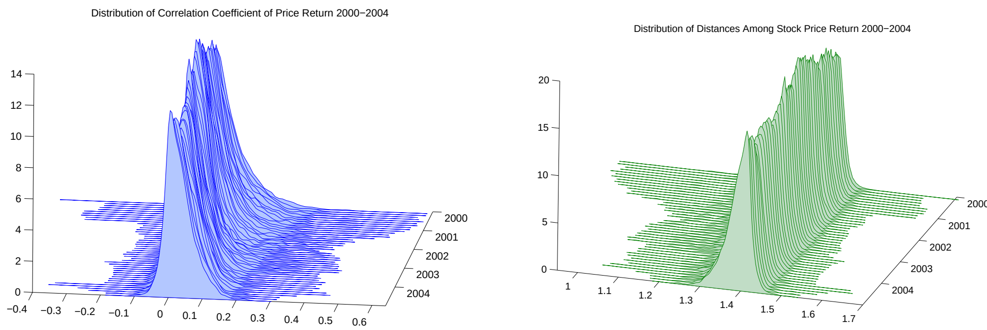

Figure 1: The distribution of correlation coefficients of traded stock price return in JSX 2000-2004 (*left*) and the distribution of distances among stocks in the same period (*right*).

The second moment known as variance correlation coefficient is calculated as

$$\mu(2) = \frac{1}{N(N-1)/2}\sum_{i,j}(c_{ij} - \mu(1))^2 \tag{14}$$

the skewness as

$$\mu(3) = \frac{1}{N(N-1)/2}\sum_{i,j}\frac{(c_{ij} - \mu(1))^3}{\sqrt{\mu(2)^3}} \tag{15}$$

and the kurtosis

$$\mu(4) = \frac{1}{N(N-1)/2}\sum_{i,j}\frac{(c_{ij} - \mu(1))^4}{\mu(2)^2} \tag{16}$$

Similarly, we do the investigation with the distances among stocks. The distance matrix is also in the form of $M \times M$ matrix and we can calculate the four moments of distance in the similar way. The distribution of the correlation coefficients and distances over (monthly) time window and the moments of distribution are showed in figure 1 and figure 2 respectively.

From the figures we can clearly see that over time windows in the period of 2000-2004, the distances among stocks relatively grow bigger as the mean

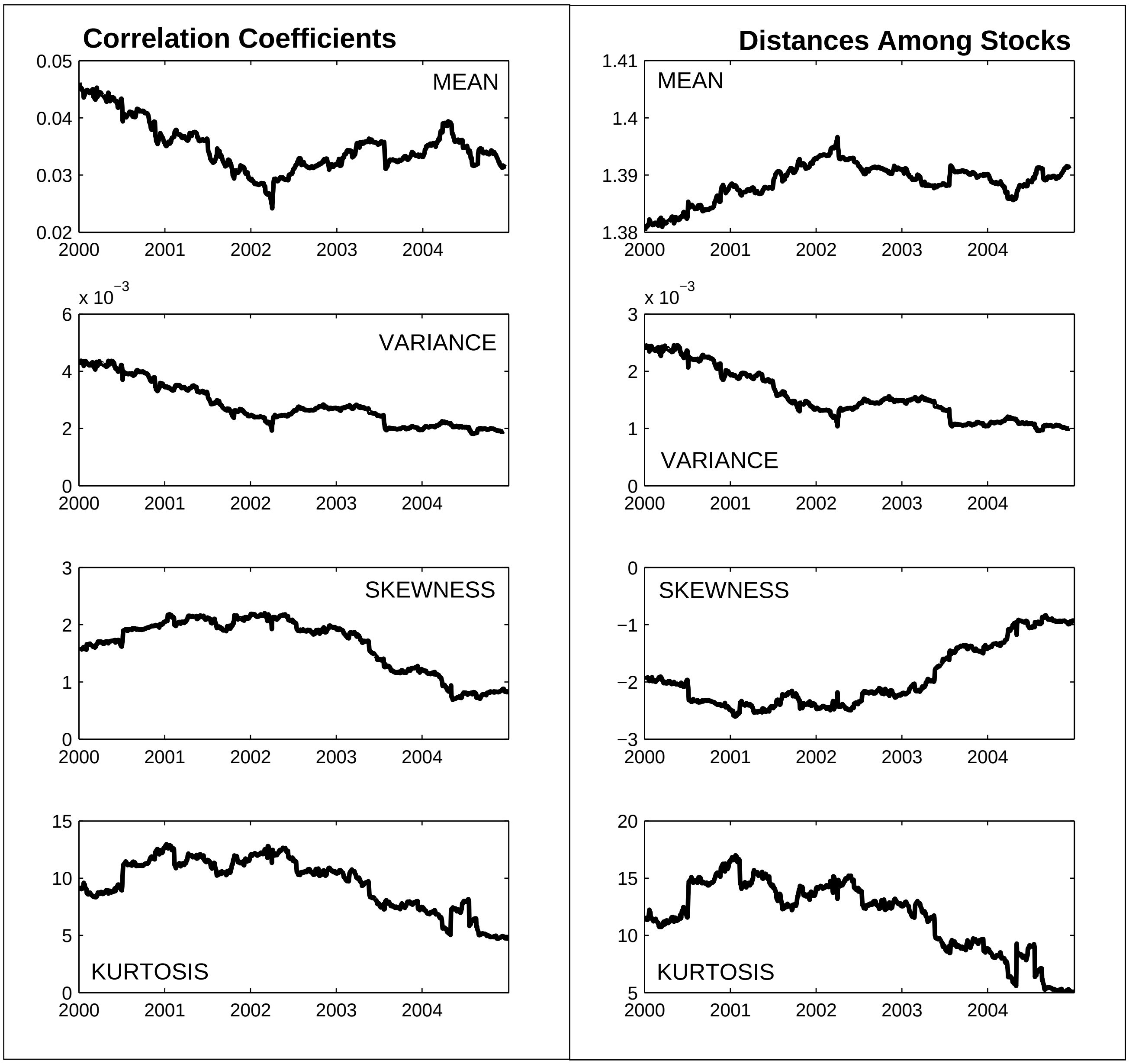

Figure 2: The moments of correlation coefficients (*left*) and distances among stocks (*right*) over time.

correlation coefficient get smaller. As described in [12], the large value of the mean correlation coefficient is related to the occurance of large crashes. By comparing it to our findings, we can simply understood the declining of the mean correlation coefficient as the economic recovery after the monetary crisis several years before the period of our data. Interestingly, as time grows in our time period, this fact is also showed by the higher moments of distribution. Here, we recognize (again) the high correlation among moments of correlation coefficients and distances among stocks traded in Indonesia.

We move now to the analysis of the ultrametricity space of stocks. We do the previously elaborated Kruskal's algorithm to the distance matrix in order to have the simple locally minimum spanning tree of the stocks. Figure 3 shows the yielded ultrametric spaces of four trading years. The figure is produced by using Pajek, the software for analysis and visualization of large networks [1]. Geometrically we can see that in 2001 *TLKM* and *ASII* are highly connected to other stocks showing that the fluctuations of both stocks is frequently referred by some particular stocks. This pattern is relatively persistent in the following year (2002). However, in the year 2003, the pattern is changed since the both stocks are no longer highly connected to other stocks. In stead, there are apparently some connections among some sectors: consumer goods industries and the financial sector. Eventually, in 2004, the previous geometric pattern seems to dwindle. There seems no single dominant fluctuation reference and the stock connects to each other arbitrarily; nevertheless, several stocks seems still to be connected with the same sectors.

In advance, figure 4 shows the ultrametric distances among the stocks. Obviously, the pattern we can see in this single linkage hierarchical tree is the closely connection between two or more stocks based upon respective distance. We can see that in the trading period of 2001, the highly capitalized stocks are clustered in a very short distance (the shortest cluster in left of 2001 tree is well-known cigarette industries: *GGRM* and *HMSP*, electronic

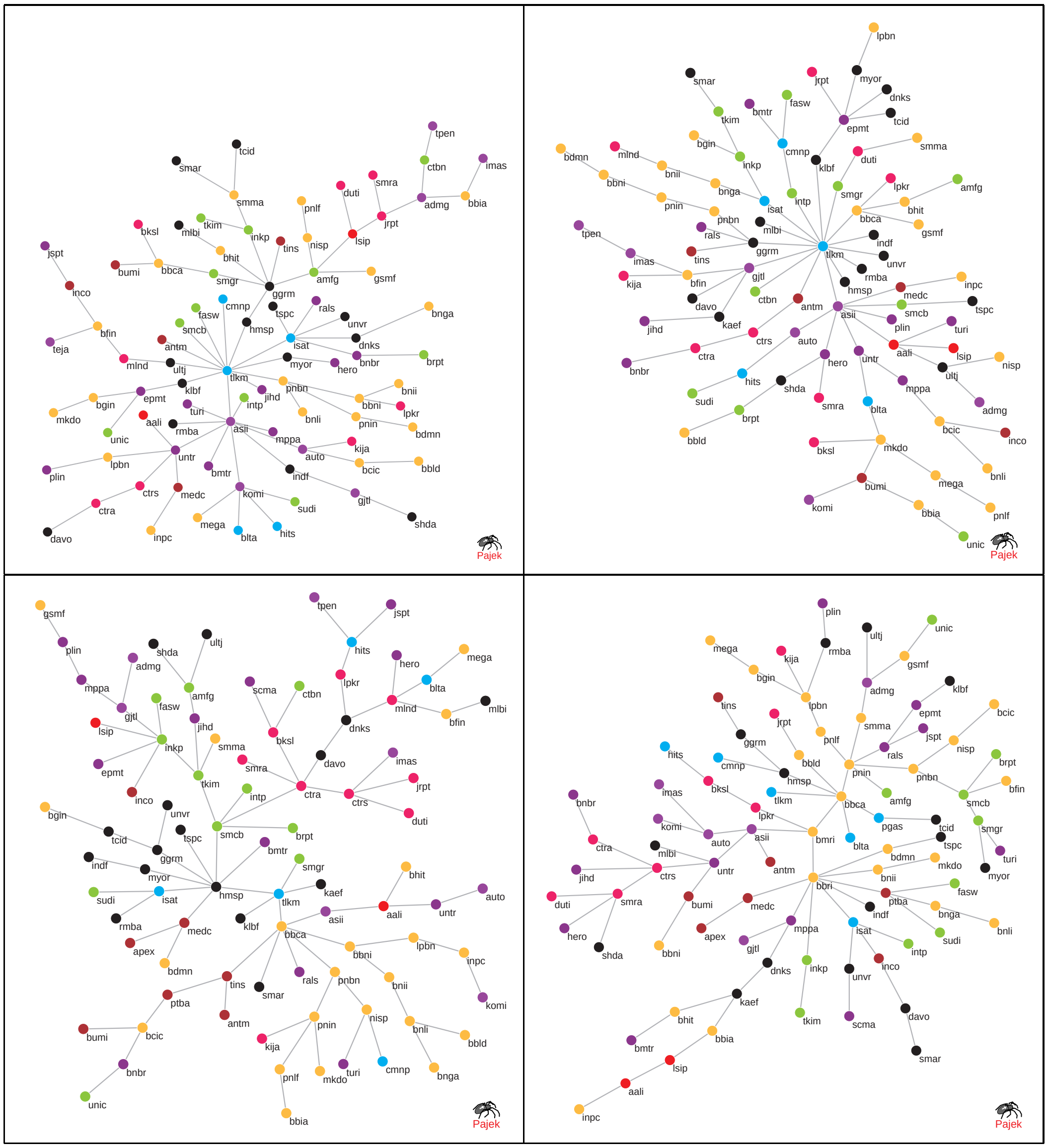

Figure 3: The minimum spanning tree of traded stocks in JSX in trading period 2001 (*top-left*), 2002 (*top-right*), 2003 (*bottom-left*), and 2004 (*bottom-right*). The colors of vertex represents the business sectors: agriculture (*red*), mining (*maroon*), basic industry (*green*), miscellaneous industry (*purple*), consumer goods industry (*black*), property and building construction (*pink*), infrastructure and transportation (*cyan*), finance (*orange*), trade and investment (*blue*).

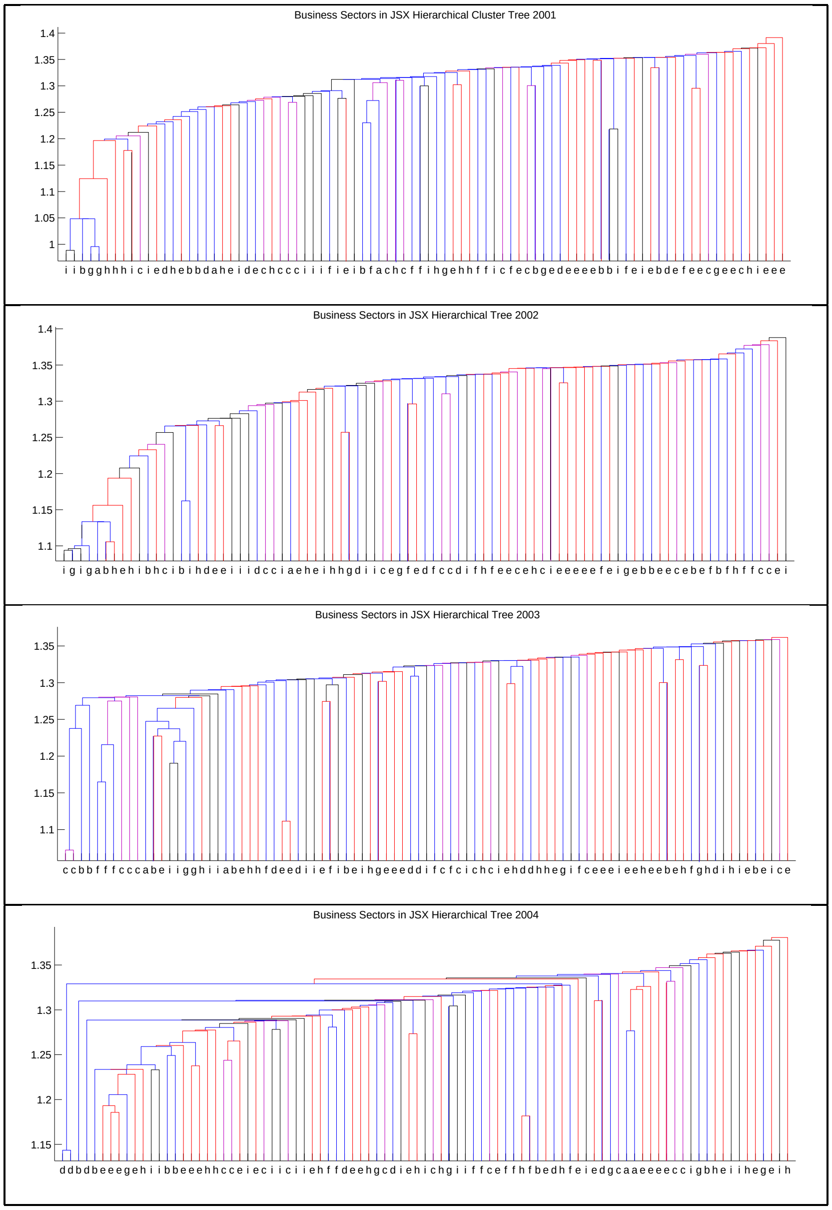

Figure 4: The hierarchical tree of clustered traded stocks in JSX in 2001-2004. The letters in axis of each graph represents the respective business sector: (a) agriculture sector, (b) miscellaneous industry, (c) basic industry and chemicals, (d) mining, (e) financial sector, (f) property, real estate, and building construction, (g) infrastructure, utilities, and transportation, (h) trade, services, and investment, and (i) consumer goods industry. *Purple lines* indicate basic industry and chemicals, *black ones* indicate consumer goods industry, and *red* ones financial and investment sector.

company: *ASII*, and telecommunication companies: *ISAT* and *TLKM* (with shortest distance < 1). This pattern is persistent over the next trading period with further respective distances (closest distance > 1) but disappear in the following year. Overall look of the hierarchical tree is that some sectors, e.g.: financial and investment sectors, basic good industries, are clustered together with various distances on each year showing their dominance of the whole fluctuations. In fact, intuitively the aggregate index of all trading stocks in JSX is frequently recognized to be too much influenced by several sectors only.

4 Discussions: On Liquid and Illiquid Stocks

As it has been uttered previously, the reference stocks of overall fluctuations are seemingly popular stocks. Furthermore, as popularity of a stock is often correlated to its market capitalization or roughly liquidity, the dominant stocks for each period are mostly the most liquid ones. To investigate this liquidity signature, we draw the minimal spanning tree by coloring the vertices based on their market capitalization (figure 5). From the figure, it is obvious that in all of the year segments, the most liquid stocks are laid in the center of the drawings and become the reference of many other stocks that are relatively more illiquid.

The hierarchical tree seems to present the similar properties (figure 6). In the figure, thirty of the most liquid stocks play highly correlated fluctuations throughout the years. In fact, when the stock market is evolving from monetary crisis to the more stable economic system (showed by the smaller mean correlation function in section 2), the robust pattern in Indonesia stock market data is the signature of liquid stocks in the hierarchical tree.

This fact becomes interesting when we relate it to the fact of investor's behavior. When only several stocks with high market capitalization are highly correlated over the years, it gives us a symptom that investor's decision is

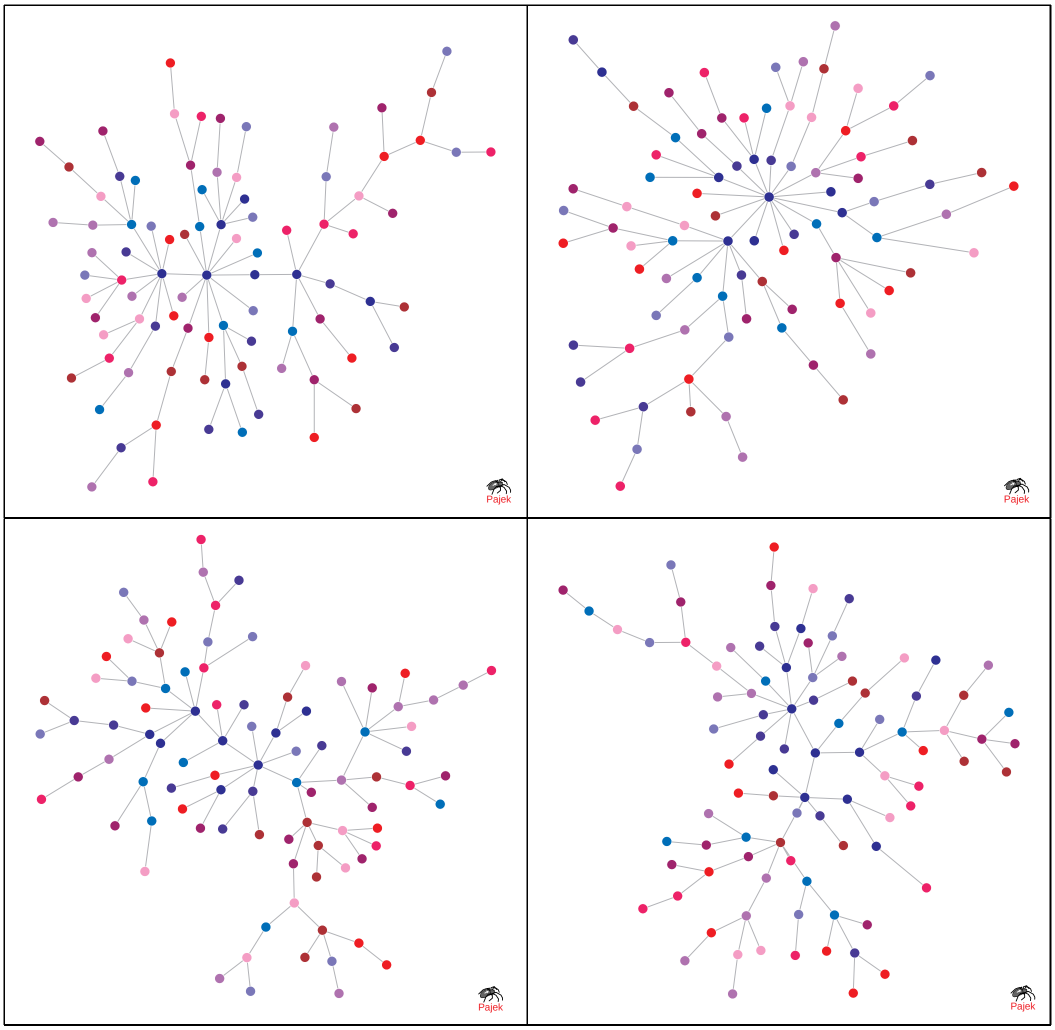

Figure 5: Similar with figure 3, but each vertex colored respect to the liquidity of the stock from the most liquid (*bluest one*) to the most illiquid (*the most red one*) of the chosen stocks.

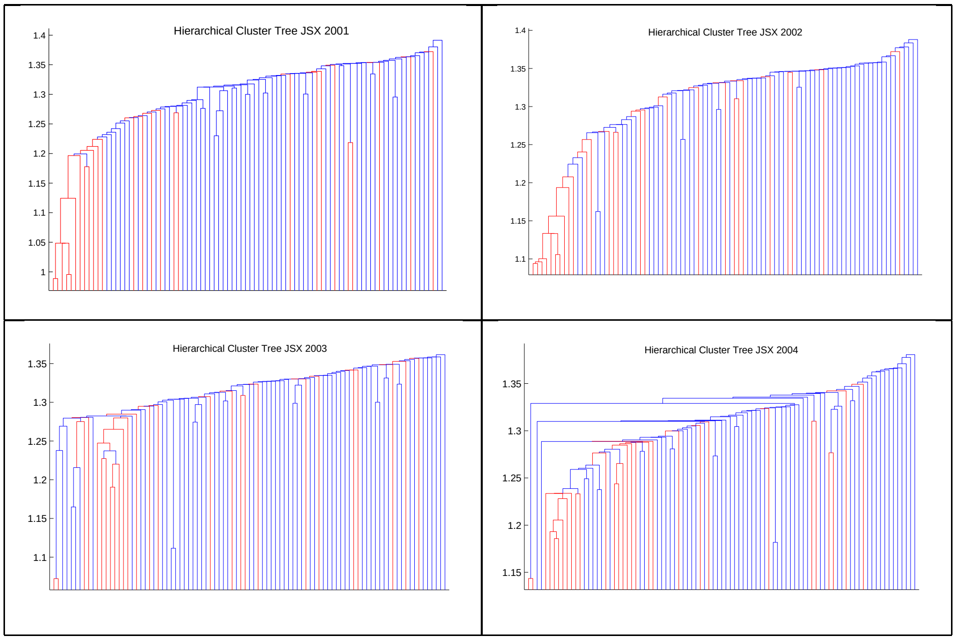

Figure 6: Similar to figure 4 showing 30 of the most liquid stocks (*colored by red line*).

mostly referred to the fluctuations of the most liquid stocks. This is also discovered in our analysis on the statistical aspects of Indonesian stock market [15]. Most investors are very sensitive to fundamental rumors and market rush frequently occurs. Referring to this fact, it is not an exaggeration if we concur that the investor's decisions on illiquid stocks are highly correlated to the way the most liquid stocks fluctuate. This means that the fluctuations of most illiquid stocks will very much depends upon the more liquid ones.

The market behaves like this can simply stated to be organized by the herding behavior among investors [4] - but this effect is stronger since investors are not only herd by the fundamental rumors but also highly related to the fluctuations of the liquid stocks. Referring to [7] we can see it as an emergent factor in inefficient market.

Furthermore, by investigating the total length of MST yielded on many time windows in our data periods (figure 7), it is clear that the total distances

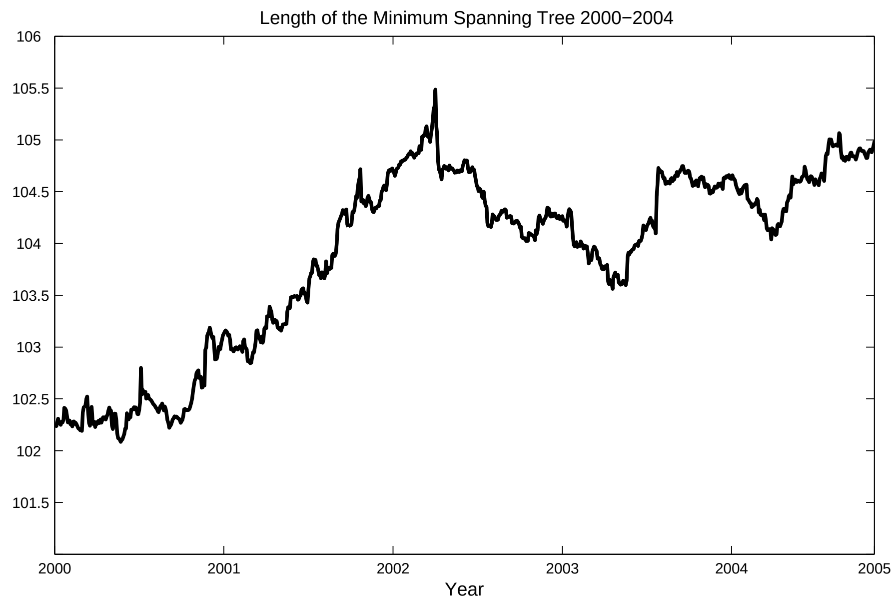

Figure 7: The evolution of the total length of minimum spanning tree of stocks traded in JSX in the period 2000-2004.

among stocks are hiking in the trading years of 2000-2004. Considering the investor's portfolio management, we can say that after the economic system stabilized over time, investors tend to recognize more stocks to invest with. This fact explains the dwindling dominance of several stocks in first years of our minimally spanning financial trees while at the same time, the total subdominant ultrametric lengths are higher over time. We can say that the more stable economic system in Indonesia has brought together positive impact on the evolution of overall stock prices fluctuations.

5 Concluding Remarks

The ultrametricity in locally minimal spanning tree that is constructed based on stock price fluctuations can help us extract the information concealed in the correlation coefficients of stock price returns. Geometrically, by visualizing the ultrametric space of financial data, we can see the appearance of patterns of fluctuations depicting the spectral movement of the market. In the other hand, from the hierarchical diagram of the financial ultrametric space we can visualize more clearly how a stock specifically correlated to one another.

From the implementation of the methodology into Indonesian stock market data in the period of 2000-2004, we can see that stock market is in the phase of the market stabilization after previous economic shock of monetary crisis. This is indicated by higher length of minimally spanning trees over time by means of the lower correlation among stock fluctuations. In other words, the investors seems to vary their portfolios not only considering investment in liquid stocks.

However, another possible explanation to the higher total distances of MST among stock fluctuations may also regard the parameter used by investors to measure the investment climate, namely the aggregate index of Jakarta Composite Index (*IHSG*). From the hierarchical tree we can see the dominance of several stocks in the composite index that also intuitively recognized by commonsensical thought. In advance, this can propose a further research within analysis of composite index evaluation in order to formalize a new possible index describing overall ups and downs of stock prices in Indonesian market more justly by eliminating the effect of several dominant sectors traded in Indonesian bourse.

Eventually, we can conclude the paper by recognizing a powerful methodology to - at least - reduce the complexity of the stock market that is realized to have so many levels of complexity since long time ago. The methodology has two sides of analytical knife: of the advantage on evaluating the overall market and in the other hand an intuitive tool to build portfolio based on relations among stocks.

6 Acknowledgement

The paper presents a post-research analysis previously funded by the cooperation of Surya Research International and PT Bursa Efek Jakarta (February-April 2005). Authors want to acknowledge helpful data correspondence with Vitri Herma Susanti, as supervised by Kandi Sofia S. Dahlan (Head, Research and Development Division, the Jakarta Stock Exchange) and BFI colleagues for insightful discussions. All possible errors remain authors'.

References

[1] Batagelj, V. and Mrvar, A. (2005). *Pajek: Program for Large Network Analysis*. Homepage: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[2] Bonanno, G., Lillo, F., and Mantegna, R. N. (2001). "High Frequency Cross-correlation in a Set of Stocks". *Quantitative Finance* 1: 96-104.

[3] Bonanno, G., Lillo, F., and Mantegna, R. N. (2001). *Levels of Complexity in Financial Markets*. arxiv:cond-mat/0104369.

[4] Cont, R. and Bouchaud, J. (2000). "Herd Behavior and Aggregate Fluctuations in Financial Markets. *Macroeconomic Dynamics* 4:170-96.

[5] Khanafiah, D. and Situngkir, H. (2004). *Innovation as Evolution: Case Study Phylomemetic Tree of Cellphone Designs*. Working Paper WP2004 Bandung Fe Institute. Pre-print: arxiv:nlin.AO/0412043.

[6] Kruskal Jr., J.B. (1956). "On the shortest spanning subtree of a graph and the travelling salesman problem". *Proceeding of American Mathematical Society* 7: 48-50.

[7] Makowiec, D. (2004). "On Modeling of Inefficient Market". *Physica A* 344:36-40.

[8] Mantegna, R. N. (1999). "Hierarchical Structure in Financial Market". *The European Physical Journal B* 11:193-7.

[9] Mantegna, R. N. and Stanley, H. E. (2002). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge UP.

[10] Mezard, M., Parisi, G., and Virasoro, M. A. (eds.). (2002). *Spin Glass Theory and Beyond*. World Scientific.

[11] Miśkiewicz, J. and Ausloos, M. (2005). *G7 Country Gross Domestic Product (GDP) Time Correlations: A Graph Network Analysis*. Pre-print: arxiv:physics/0504099.

[12] Onnela, J-P., Chakraborti, A., Kaski, K., Kertész, J., and Kanto, A. (2003). "Dynamics of Market Correlations: Taxonomy and Portfolio Analysis". *Physical Review E* 68:056110.

[13] Schulz, M. (2003). *Statistical Physics and Economics: Concepts, Tools, and Applications*. Springer.

[14] Situngkir, H. and Surya, Y. (2004). *Hierarchical Taxonomy in Multi-Party System*. Working Paper WPM2004 Bandung Fe Institute. Preprint: arxiv:nlin.PS/0405005.

[15] Situngkir, H., Surya, Y., and Hariadi, Y. (2005). *Membandingkan Continuous dan Continuous Interval Trading System dalam Perspektif Likuiditas*. Research Report February 1 - April 30, 2005. Surya Research International.