Physica Scripta. Vol. T106, 48–54, 2003

Asset Trees and Asset Graphs in Financial Markets

J.-P. Onnela[1], A. Chakraborti[1], K. Kaski[1,*] J. Kertész[2] and A. Kanto[3]

[1] Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FIN-02015 HUT, Finland
[2] Department of Theoretical Physics, Budapest University of Technology & Economics, Budafoki út 8, H-1111, Budapest, Hungary, and Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FIN-02015 HUT, Finland
[3] Department of Quantitative Methods in Economics and Management Science, Helsinki School of Economics, P.O.Box 1210, FIN-00101 Helsinki, Finland

Received March 18, 2003; accepted April 2, 2003

PACS Ref: 89.65.−s, 89.75.−k, 89.90.+n

Abstract

This paper introduces a new methodology for constructing a network of companies called a dynamic asset graph. This is similar to the dynamic asset tree studied recently, as both are based on correlations between asset returns. However, the new modified methodology does not, in general, lead to a tree but a disconnected graph. The asset tree, due to the minimum spanning tree criterion, is forced to "accept" edge lengths that are far less optimal (longer) than the asset graph, thus resulting in higher overall length for the tree. The same criterion also causes asset trees to be more fragile in structure when measured by the single-step survival ratio. Over longer time periods, in the beginning the asset graph decays more slowly than the asset tree, but in the long run the situation is reversed. The vertex degree distributions indicate that the possible scale free behavior of the asset graph is not as evident as it is in the case of the asset tree.

1. Introduction

In a recent paper Mantegna suggested to study the clustering of companies using the correlation matrix of asset returns [1], transforming correlations into distances, and selecting a subset of them with the minimum spanning tree (MST) criterion. In the resulting tree, the distances are the edges connecting the nodes, or companies, and thus a taxonomy of the financial market is formed. This method was later studied by Bonanno *et al.* [2], while other studies on clustering in the financial market are [3–7], and those specifically on market crashes [8, 9].

Recently, we have studied the properties of a set of asset trees created using the methodology introduced by Mantegna in [10–12], and applied it in the crash context in [13]. In these studies, the obtained multitude of trees was interpreted as a sequence of evolutionary steps of a single "dynamic asset tree", and different measures were used to characterize the system, which were found to reflect the state of the market. The economic meaningfulness of the emerging clustering was also discussed and the dynamic asset trees were found to have a strong connection to portfolio optimization.

In this paper, we introduce a modified methodology which, in general, does not not lead to a tree but a graph, or possibly even several graphs that do not need to be interconnected. Here we limit ourselves to studying only one type of "dynamic asset graph", which in terms of its size is compatible and thus comparable with the dynamic asset tree. Although in graph theory a tree is defined as a type of graph, the terms asset graph and asset tree are used here to refer to the two different approaches and as concepts are mutually exclusive and noninterchangeable. The aims of this paper are to introduce this modified approach and to demonstrate some of its similarities and differences with our previous approach. Further considerations of topology and taxonomy of the financial market obtained using dynamic asset graphs are to be presented later.

2. Constructing asset trees and asset graphs

In this paper, the term financial market refers to a set of asset price data commercially available from the Center for Research in Security Prices (CRSP) of the University of Chicago Graduate School of Business. Here we will study the split-adjusted daily closure prices for a total of $N = 477$ stocks traded at the New York Stock Exchange (NYSE) over the period of 20 years, from January 2, 1980 to December 31, 1999. This amounts a total of 5056 price quotes per stock, indexed by time variable $\tau = 1, 2, \ldots, 5056$. For analysis and smoothing purposes, the data is divided time-wise into M *windows* $t = 1, 2, \ldots, M$ of width T, where T corresponds to the number of daily returns included in the window. Several consecutive windows overlap with each other, the extent of which is dictated by the window step length parameter δT, which describes the displacement of the window and is also measured in trading days. The choice of window width is a trade-off between too noisy and too smooth data for small and large window widths, respectively. The results presented in this paper were calculated from monthly stepped four-year windows. Assuming 250 trading days a year gives $\delta T = 250/12 \approx 21$ days and $T = 1000$ days, which we found optimal from a wide set of values for both parameters [10]. With these choices, the overall number of windows is $M = 195$.

In order to investigate correlations between stocks we first denote the closure price of stock i at time τ by $P_i(\tau)$ (Note that τ refers to a date, not a time window). We focus our attention to the logarithmic return of stock i, given by $r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1)$ which for a sequence of consecutive trading days, i.e., those encompassing the given window t, form the return vector $\boldsymbol{r}_i^t$. In order to characterize the synchronous time evolution of assets, we use the equal time correlation coefficients between assets i and j defined as

$$\rho_{ij}^t = \frac{\langle \boldsymbol{r}_i^t \boldsymbol{r}_j^t \rangle - \langle \boldsymbol{r}_i^t \rangle \langle \boldsymbol{r}_j^t \rangle}{\sqrt{[\langle \boldsymbol{r}_i^{t2} \rangle - \langle \boldsymbol{r}_i^t \rangle^2][\langle \boldsymbol{r}_j^{t2} \rangle - \langle \boldsymbol{r}_j^t \rangle^2]}}, \tag{1}$$

* e-mail: kimmo.kaski@hut.fi

where $\langle \ldots \rangle$ indicates a time average over the consecutive trading days included in the return vectors. Due to Cauchy–Schwarz inequality, these correlation coefficients fulfill the condition $-1 \leq \rho_{ij} \leq 1$ and form an $N \times N$ correlation matrix $\mathbf{C}^t$, which serves as the basis for graphs and trees to be discussed in this paper.

For the purpose of constructing asset graphs and asset trees, we define a distance between a pair of stocks. This distance is associated with the edge connecting the stocks and it reflects the level at which the stocks are correlated. We use a simple non-linear transformation $d_{ij}^t = \sqrt{2(1-\rho_{ij}^t)}$ to obtain distances with the property $2 \geq d_{ij} \geq 0$, forming an $N \times N$ symmetric distance matrix $\mathbf{D}^t$. Now two alternative approaches may be adopted. The first one leads to asset trees, and the second one to asset graphs. In both approaches the trees (or graphs) for different time windows are not independent of each other, but form a series through time. Consequently, this multitude of trees or graphs is interpreted as a sequence of evolutionary steps of a single *dynamic asset tree* or *dynamic asset graph*.

In the first approach we construct an asset tree according to the methodology by Mantegna [1]. This approach requires an additional hypothesis about the topology of the metric space, namely, the so-called ultrametricity hypothesis. In practice, it leads to determining the minimum spanning tree (MST) of the distances, denoted $\mathbf{T}^t$. The spanning tree is a simply connected acyclic (no cycles) graph that connects all N nodes (stocks) with $N-1$ edges such that the sum of all edge weights, $\sum_{d_{ij}^t \in \mathbf{T}^t} d_{ij}^t$, is minimum. We refer to the minimum spanning tree at time t by the notation $\mathbf{T}^t = (V, E^t)$, where V is a set of vertices and E^t is a corresponding set of unordered pairs of vertices, or edges. Since the spanning tree criterion requires all N nodes to be always present, the set of vertices V is time independent, which is why the time superscript has been dropped from the notation. The set of edges E^t, however, does depend on time, as it is expected that edge lengths in the matrix $\mathbf{D}^t$ evolve over time, and thus different edges get selected into the tree at different times.

In the second approach we construct asset graphs. This consists of extracting the $N(N-1)/2$ distinct distance elements from the upper (or lower) triangular part of the distance matrix $\mathbf{D}^t$, and obtaining a sequence of edges $d_1^t, d_2^t, \ldots, d_{N(N-1)/2}^t$, where we have used a single index notation. The edges are then sorted in an ascending order and form an ordered sequence $d_{(1)}^t, d_{(2)}^t, \ldots, d_{(N(N-1)/2)}^t$. Since we require the graph to be representative of the market, it is natural to build the graph by including only the strongest connections in it. The number of edges to include is, of course, arbitrary. Here we include only $N-1$ shortest edges in the graph, thus giving $E^t = \{d_{(1)}^t, d_{(2)}^t, \ldots, d_{(N-1)}^t\}$. This is motivated by the fact that the asset tree also consists of $N-1$ edges, and this choice renders the two methodologies comparable, and possibly even similar to one another. The presented mechanism for constructing graphs defines them uniquely and, consequently, no additional hypotheses about graph topology are required. It is important to note that both the set of vertices V^t and the set of edges E^t are time dependent, and thus we denote the graph with $\mathbf{G}^t = (V^t, E^t)$. The choice to include only the $N-1$ shortest edges in the graph means that the *size* of the graph, defined as the number of its edges, is fixed at $N-1$. However, the *order* of the graph, defined as the number of its vertices, is not fixed for the graph but varies as a function of time. This is due to the fact that even a small set of vertices may be strongly inter-connected, and thus may use up many of the available edges. This may also lead to the formation of cycles in the graph. These aspects are clearly different from the tree approach, where the order is always fixed at N and no cycles are allowed by definition.

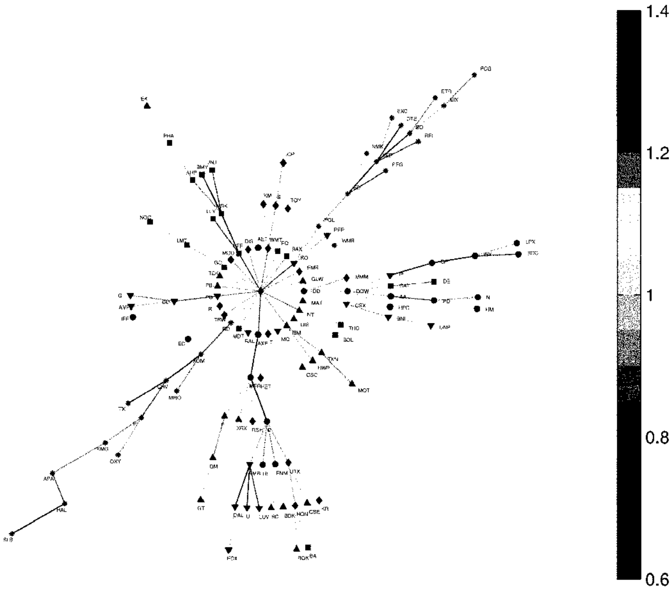

Fig. 1. A sample asset tree $\mathbf{T}^t$ for $t = 168$, where General Electric (GE) is used as the central vertex.

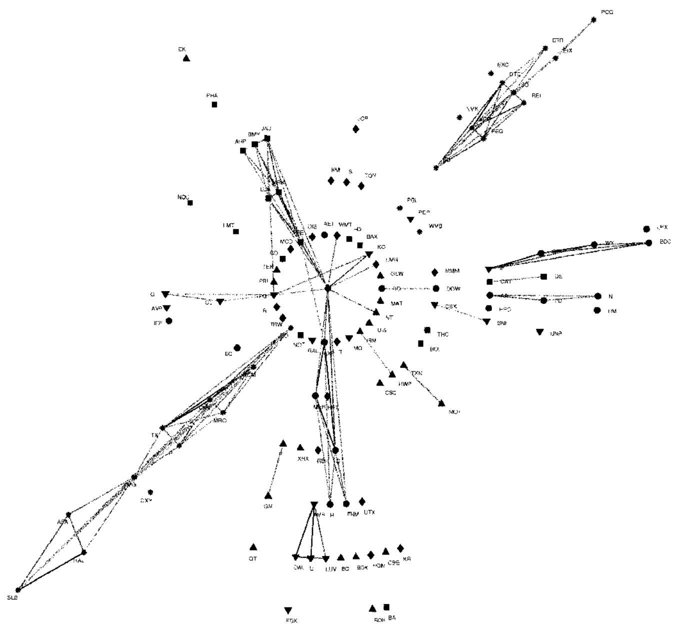

Fig. 2. A sample asset graph $\mathbf{G}^t$ for $t = 168$.

In order to compare the two methodologies visually, Figs. 1 and 2 show a sample plot of the asset tree and the asset graph, obtained using the first and the second approach, respectively. Here a smaller dataset is used, which consists of 116 S&P 500 stocks, extending from the beginning of 1982 to the end of 2000 [14]. The window width was set at $T = 1000$ trading days, and the shown sample tree is located time-wise at $t = 168$, corresponding to January 1, 1998. Distance between a pair of vertices is indicated by the color of the incident edge, as given by the color bar in Fig. 1. The sample plots show how the asset tree spans all the nodes, whereas the asset graph contains only their subset. The isolated vertices in Fig. 2 are included for purposes of comparison only, but are not included in the vertex set V^t of the graph $\mathbf{G}^t = (V^t, E^t)$. This figure also shows how the asset graph is disconnected and consists of several components, and it demonstrates the presence of cycles in it, which are not allowed for the asset tree. The lengths of the edges in the asset graph are clearly shorter than those in the asset tree, as indicated by the differences in their color. The different markers in Figs. 1 and 2 correspond to different business sectors of the studied companies as defined by Forbes, *http://www.forbes.com*. These definitions can be used to form economically meaningful clusters in the tree, as studied in detail in [12]. Also the asset graph tends to link together stocks that belong to the same business sector, a property to be studied further in some other context.

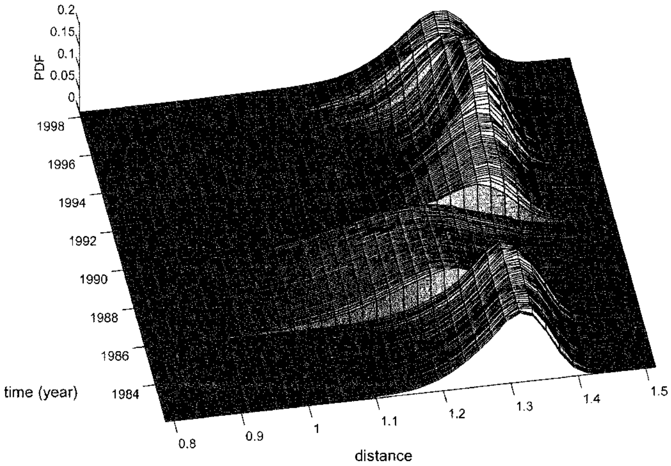

Fig. 3. Distribution of all $N(N-1)/2$ distance elements d_{ij} contained in the distance matrix $\mathbf{D}^t$ as a function of time.

3. Market characterization

For the market characterization let us start by comparing the two approaches, i.e., asset tree and asset graph, by examining visually their edge length distributions. We present three distribution plots of (i) distance elements d_{ij} contained in the distance matrix $\mathbf{D}^t$ (Fig. 3), (ii) distance elements d_{ij} contained in the asset (minimum spanning) tree $\mathbf{T}^t$ (Fig. 4), and (iii) distance elements d_{ij} contained in the asset graph $\mathbf{G}^t$ (Fig. 5). In all these plots, but most prominently in Fig. 3, there appears to be a discontinuity in the distribution from roughly 1986 to 1990, such that a part has been cut out, pushed to the left and made flatter. This anomaly is a manifestation of Black Monday (October 19, 1987), and its length along the time axis is related to the choice of window width T [12, 13].

We can now reconsider the tree and graph construction mechanism described earlier. Starting from the distribution of the $N(N-1)/2$ distance matrix elements in Fig. 3, for the asset tree we pick the shortest $N-1$ of them, subject to the constraint that all vertices are spanned by the chosen edges. For the purpose of building the graph, however, this constraint is dropped and we pick the shortest elements from the distribution in Fig. 3. Therefore, the distribution of graph edges in Fig. 5 is simply the left tail of the distribution of distance elements in Fig. 3, and it seems that the asset graph rarely contains edges longer than about 1.1, the largest distance element being $d_{\max} = 1.1375$. In contrast, in the distribution of tree edges in Fig. 4 most edges included in the tree seem to come from the area to the right of the value 1.1, and the largest distance element is now $d_{\max} = 1.3549$.

Instead of using the edge length distributions as such, we can characterize the market by studying the location (mean) of the edge length distribution by defining a simple measure, the *mean distance*, as

$$\bar{d}(t) = \frac{1}{N(N-1)/2} \sum_{d^t_{ij} \in \mathbf{D}^t} d^t_{ij}, \tag{2}$$

where t denotes the time at which the tree is constructed, and the denominator is the number of distinct elements in

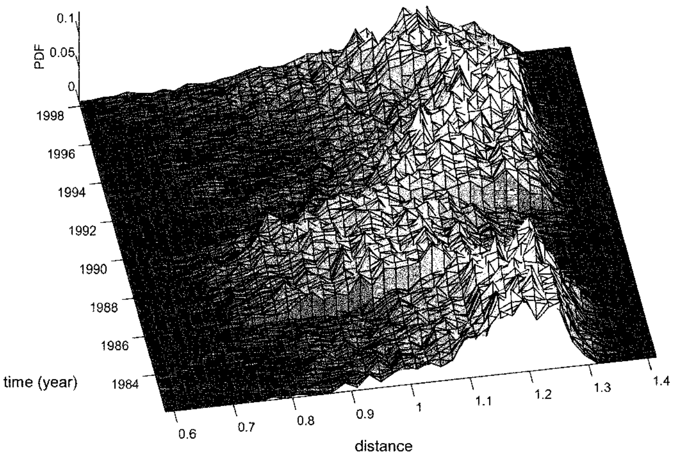

Fig. 4. Distribution of the $(N-1)$ distance elements d_{ij} contained in the asset (minimum spanning) tree $\mathbf{T}^t$ as a function of time.

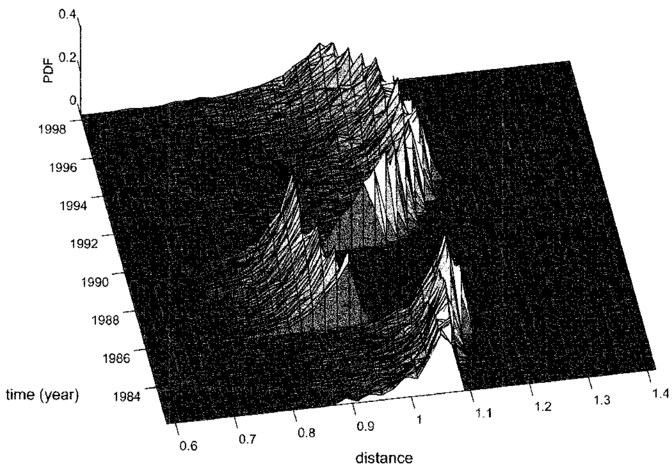

Fig. 5. Distribution of the $(N-1)$ distance elements d_{ij} contained in the asset graph $\mathbf{G}^t$ as a function of time.

the matrix. It is noted that one could instead use the *mean correlation coefficient*, defined as

$$\bar{\rho}(t) = \frac{1}{N(N-1)/2} \sum_{\rho_{ij}^t \in \mathbf{C}^t} \rho_{ij}^t, \tag{3}$$

which would lead to the same conclusions, as the mean distance and mean correlation coefficient are mirror images of one another and, consequently, it suffices to examine either one of them. In a similar manner, we can characterize the asset tree and the asset graph, which are both simplified networks representing the market, but use only $N-1$ distance elements d_{ij}^t out of the available $N(N-1)/2$ in the distance matrix $\mathbf{D}^t$. Thus we define the *normalized tree length* for the asset tree as

$$L_{\mathrm{mst}}(t) = \frac{1}{N-1} \sum_{d_{ij}^t \in \mathbf{T}^t} d_{ij}^t, \tag{4}$$

and the *normalized graph length* for the asset graph as

$$L_{\mathrm{graph}}(t) = \frac{1}{N-1} \sum_{d_{ij}^t \in \mathbf{G}^t} d_{ij}^t, \tag{5}$$

where t again denotes the time at which the tree or graph is constructed, and $N-1$ is the number of edges present. These three measures are depicted in Fig. 6, from which the following observations are made. First, all three measures behave very similarly, which is also reflected by the level of mutual correlations. Pearson's linear and Spearman's rank-order correlation coefficients between the mean distance and normalized tree length are 0.98 and 0.92, respectively, while between the mean distance and the normalized graph length they turned out to be 0.96 and 0.87. Thus, the normalized tree length seems to track the market slightly better. Second, the average values of these measures decrease in moving from the mean distance, via the normalized tree length, to the normalized graph length, being 1.29, 1.12 and 1.00, respectively. Also, the normalized tree length is always higher than the the normalized graph length, differing on average by 0.13. Thus it seems that the asset tee, due to the minimum spanning tree criterion, is forced to "accept" edge lengths that are far less optimal (longer) than the asset graph, resulting in a higher average value for the normalized tree length than for the normalized graph length. Third, the normalized graph length tends to exaggerate the depression caused by the crash, which can be traced back to the graph construction mechanism. We have earlier studied just one of these measures, namely, the normalized tree length and found it to be descriptive of the overall market state. Furthermore it turned out to be closely related to market diversification potential, i.e., the scope of the market to eliminate specific risk of the minimum risk Markowitz portfolio [10, 11]. The fact that the normalized distance and normalized tree length behave so similarly suggests that they are, at least to some extent, interchangeable measures.

4. Evolution of asset trees and asset graphs

The robustness of asset graph and asset tree topology can be studied through the concept of *single-step survival ratio*, defined as the fraction of edges found common in two consecutive graphs or trees at times t and $t-1$:

$$\sigma(t) = \frac{1}{N-1} |E^t \cap E^{t-1}|. \tag{6}$$

As before, E^t refers to the set of edges of the graph or the tree at time t, $\cap$ is the intersection operator and $|\ldots|$ gives the number of elements in the set. Although it has not been explicitly indicated in the definition, $\sigma(t)$ is dependent on the two parameters, namely, the window width T and the step length δT. Figure 7 shows the plots of single-step survival for both the graph (upper curve) and the tree (lower curve) for $\delta T = 250/12 \approx 21$ days and $T = 1000$ days.

The most evident observation is that the graph seems to have a higher survival ratio than the tree. For the graph, on average, some 94.8% of connections survive, whereas the corresponding number for the tree is 82.6%. In addition,

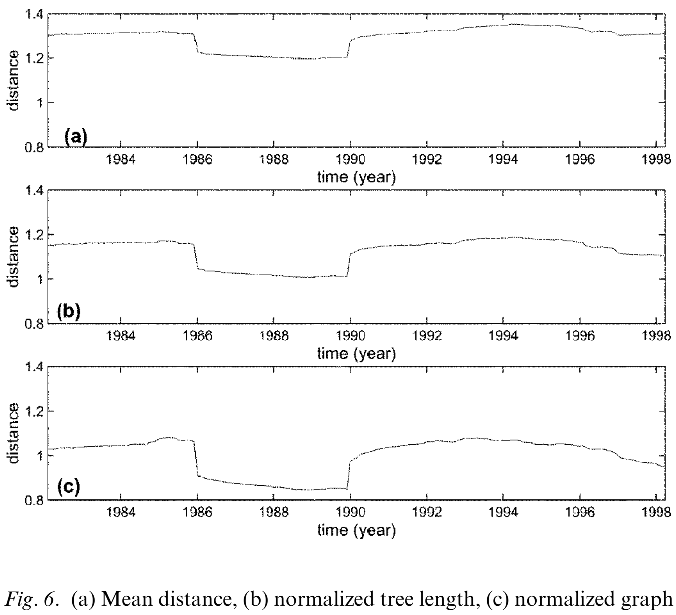

Fig. 6. (a) Mean distance, (b) normalized tree length, (c) normalized graph length as functions of time.

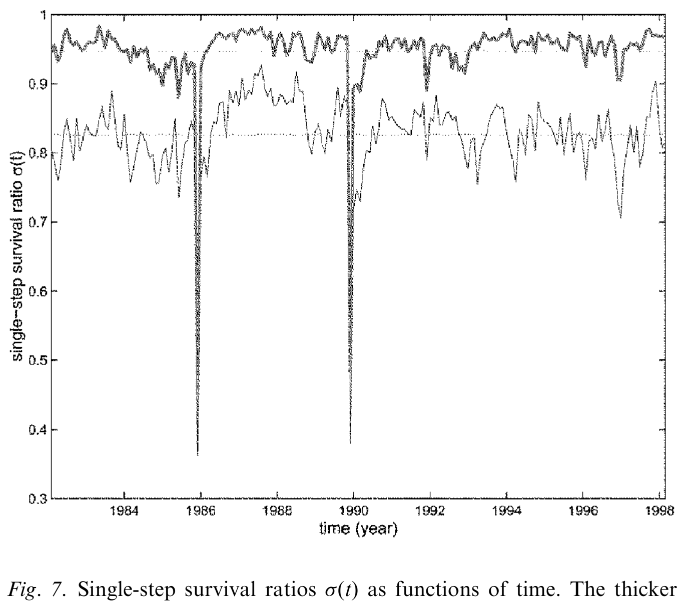

Fig. 7. Single-step survival ratios $\sigma(t)$ as functions of time. The thicker (upper) curve is for the graph and the thinner (lower) for the tree. The dashed lines indicate the corresponding average values.

the fluctuation of the single-step survival ratio, as measured by its standard deviation, is smaller for the graph at 5.3% than for the tree at 6.2%. In general, both curves fluctuate together, meaning that the market events causing re-wirings in the graph also cause re-wirings in the tree. This is very clear in the two sudden dips in both curves, which result from the re-wirings related to Black Monday [13]. Although both curves fall drastically, the one for the asset graph falls less, indicating that the graph is more stable than the tree also under extreme circumstances, such as market crashes. The higher survival ratio for the asset graph is not caused by any particular choice of parameters, but is reproduced for all examined parameter values. Some indication of the sensitivity of the single-step survival ratio on the window width parameter is given in Table I. The fact that asset graphs are more stable than asset trees is related to their construction mechanism. The spanning tree constraint practically never allows choosing the shortest available edges for the tree and, consequently, the ensuing structure is more fragile. A short edge between a pair of stocks corresponds to a very high correlation between their returns. This may result from the companies having developed a cooperative relationship, such as a joint venture, or it may be simply incidental. In the first case the created bond between the two stocks is likely to be longer lasting than if the MST criterion forced us to include a weaker bond between two companies.

We can expand the concept of single-step survival ratio to cover survival over several consecutive time steps δT. Whereas the single-step survival ratio was used to study short-term survival, or robustness, of graphs and trees, the *multi-step survival ratio* is used to study their long-term survival. It is defined as

$$\sigma(t, n) = \frac{1}{N-1} |E^t \cap E^{t-1} \cdots E^{t-n+1} \cap E^{t-n}|, \qquad (7)$$

where only those connections that have persisted for the whole time period of length $n\delta T$ without any interruptions are taken into account. According to this formula, when a bond between two vertices breaks even once within n steps and then reappears, it is not counted as a survived connection. A closely related concept is that of graph or tree *half-life* $t_{1/2}$, defined as the time in which half the number of initial connections have decayed, i.e., $\sigma(t, t_{1/2}\delta T) = 0.5$. The multi-step survival ratio is plotted in Fig. 8, where the half-life threshold is indicated by the dashed horizontal line.

The time axis can be divided into two regions based on the nature of the decay process, and these regions are located somewhat differently for the graph and the tree. The precise locations of the regions are, of course, subject

Table I: *Mean and standard deviation of the single-step survival ratio $\sigma(t)$ for the asset graph and the asset tree for different values of window width T, given in days.*

		$T = 500$	$T = 1000$	$T = 1500$
mean $\sigma(t)$	tree	72.3%	82.6%	86.9%
	graph	90.1%	94.8%	96.0%
std $\sigma(t)$	tree	7.5%	6.2%	5.4%
	graph	6.3%	5.3%	4.5%

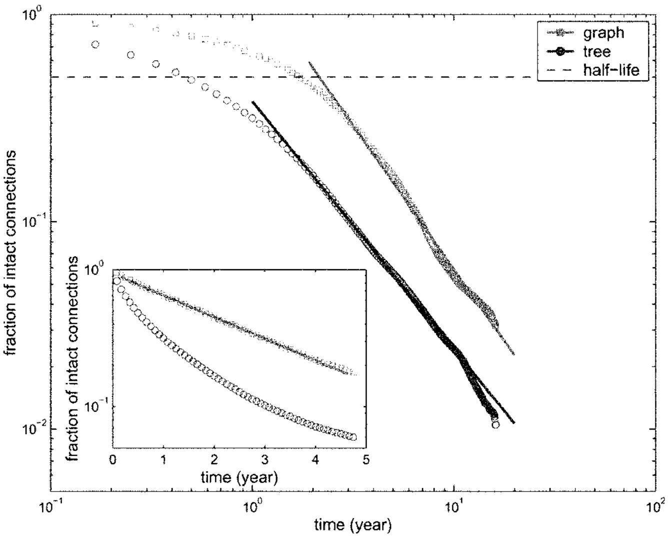

Fig. 8. Multi-step survival ratio $\sigma(t, k)$ for asset graph and tree as a function of survival time k, averaged over the time domain t.

to speculation, but for the purpose carrying out fits and analysis they need to be fixed. For the asset graph the first and second regions, discretized according to $\delta T = \frac{1}{12}$ year as mentioned before, are defined on the intervals $(\frac{1}{2}, 4)$ and $(4\frac{1}{12}, 16\frac{1}{6})$, respectively, both given in years. Within the first region, the graph exhibits clean exponential decay, as witnessed by the fitted straight line on lin-log scale in the inset of Fig. 8. Somewhere in between the two regions there is a cross-over to power-law behavior, which is evident within the second region, resulting in a straight line on the log-log plot of the same figure. For the asset tree the regions are defined on $(\frac{1}{12}, 1\frac{1}{2})$ and $(1\frac{7}{12}, 11\frac{1}{4})$. Within the first region the asset tree decays faster than exponentially, as can be verified by comparison with the straight line decay of the graph in the inset. Similarly to the graph, there is a cross-over to a power-law, although the slope is faster than for the graph. If we write the power law decay as $\langle\sigma(t, n)\rangle_t \sim n^{-\gamma}$, the fits yield for the asset graph $\gamma \approx 1.39$, whereas for the asset tree we have $\gamma \approx 1.19$.

The finding concerning the slower decay of the asset graph within the first region is fully compatible with the results obtained with single-step survival ratio. Since the graph shows higher survival ratio over a single-step, it is to be expected that, at least in the early time horizon (within the first region), graphs should decay more slowly than trees. The half-lives for both the graph and the tree occur within the first region, and thus it is not surprising that the graph half-life is much longer than the tree half-life. For the graph, we obtained $t_{1/2} \approx 1.71$ years, and for the tree $t_{1/2} \approx 0.47$. Although the half-lives depend on the value of window width T, the differences between them persist for different parameter values. When measured in years, for window widths of $T = 2$, $T = 4$ and $T = 6$, the corresponding half-lives for the asset tree are 0.22, 0.47 and 0.75 years, whereas for the asset graph they are 0.88, 1.71 and 2.51 years, respectively.

Interestingly, the situation seems to be reversed within the second region, where both decay as power-law. Here the higher exponent γ for the asset graph indicates that it actually decays faster than the asset tree. This finding could, of course, be influenced by our choice of the window width T. Explorations with that parameter revealed another interesting phenomenon; the slope for the asset tree seems to be independent of window width, as discussed in [12], but for the asset graphs this is not the case. For $T = 500$, $T = 1000$ and $T = 1500$, given in days, we obtained for the asset tree the exponents $\gamma = 1.15$, $\gamma = 1.19$ and $\gamma = 1.17$, respectively, which, within the error bars, are to be considered equal [12]. For the asset graph, however, we obtain the values of $\gamma = 2.07$, $\gamma = 1.39$ and $\gamma = 1.55$. Although no clear trend can be detected in these values, a matter that calls for further exploration, it is clear that the value of γ is higher for the asset graph than for the asset tree. Therefore, the asset graph decays more slowly than the asset tree within the first region, while within the second region the situation is just the opposite.

5. Distribution of vertex degrees in asset trees and asset graphs

As the asset graph and asset tree are representative of the financial market, studying their structure can enhance our understanding of the market itself. Recently Vandewalle *et al.* [15] found scale free behavior for the asset tree in a limited time window. They proposed the distribution of the vertex degrees $f(k)$ to follow a power law of the following form

$$f(k) \sim k^{-\alpha}, \quad (9)$$

with the exponent $\alpha \approx 2.2$. Later, we studied this phenomenon further with a focus on asset tree dynamics [12]. We found that the asset tree exhibits, most of the time, scale free properties with a rather robust exponent $\alpha \approx 2.1 \pm 0.1$ during times of normal stock market operation. In addition, within the error limits, the exponent was found to be constant over time. However, during crash periods when the asset tree topology is drastically affected, the exponent changes to $\alpha \approx 1.8 \pm 0.1$, but nevertheless the asset tree maintains its scale free character. The interesting question is whether asset graphs also display similar scale free behavior and if so, are there are differences in the value of the exponent. As Figs. 9 and 10 make clear, the observed data does not fit as well with scale free behavior for the asset graph as it does for the asset tree. The obtained average value for the exponent of the asset graph is significantly lower, i.e. $\alpha \approx 0.9 \pm 0.1$. In addition, the exponent for the asset graph varies less as a function of time and does not show distinctively different behavior between normal and crash markets.

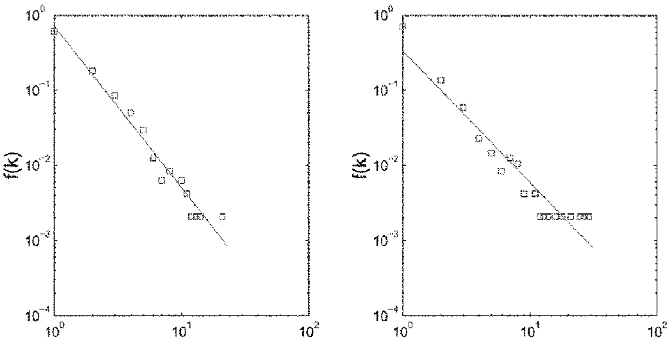

Fig. 9. Typical plots of vertex degree distributions for normal (left) and crash topology (right) for the asset tree. The exponents and goodness of fit for them are are $\alpha \approx 2.15$, $R^2 \approx 0.96$ and $\alpha \approx 1.75$, $R^2 \approx 0.92$.

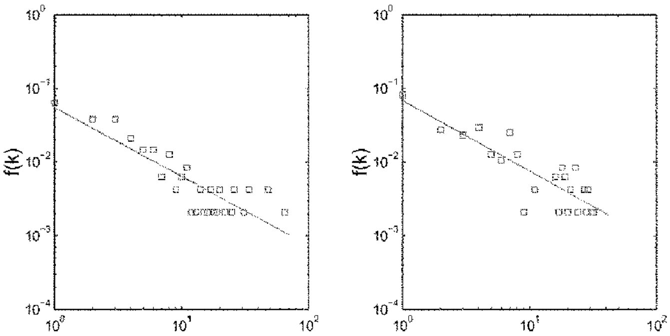

Fig. 10. Typical plots of vertex degree distributions for the asset graph. The exponents and goodness of fit for them are $\alpha \approx 0.93$, $R^2 \approx 0.74$ and $\alpha \approx 0.96$, $R^2 \approx 0.74$, respectively.

In the case of the asset tree, there were sometimes clear outliers, as one node typically had a considerably higher vertex degree than the power-law scaling would predict. This outlier was used as a *central node*, a reference node against which some tree properties were measured. However, the fact that one node often had "too high" vertex degree provided further support for using one of the nodes as the center of the tree, as discussed in detail in [12]. In case of the dynamic asset graph, these types of outliers are not present. This observation merely reflects upon the differences between the topologies produced by the two different methodologies but does not, as such, rule out the possibility of using one of the nodes as a central node[1].

In [12], we estimated the overall goodness of power-law fits for the asset trees by calculating the R^2 *coefficient of determination*, a measure which indicates the fraction of the total variation explained by the least-squares regression line. Averaged over all the time windows, we obtained the values $R^2 \approx 0.93$ and $R^2 \approx 0.86$, with and without outliers excluded, respectively. Since there were no outliers in the data for asset graphs, it was used as such to give an average of $R^2 \approx 0.75$. This indicates that the scale free behavior is not evident in this case.

6. Summary and conclusion

In summary, we have introduced the concept of dynamic asset graph and compared some of its properties to the dynamic asset tree, which we have studied recently. Comparisons between edge length distributions reveal that the asset tree, due to the minimum spanning tree criterion, is forced to "accept" edge lengths that are far less optimal (longer) than the asset graph. This results in a higher average value for the normalized tree length than for the normalized graph length although, in general, they behave very similarly. However, the latter tends to exaggerate market anomalies and, consequently, the normalized tree length seems to track the market better. The asset graph was also found to exhibit clearly higher single-step survival ratio than the asset tree. This is understandable, as the spanning tree criterion does not allow the shortest connections to be included in the tree

[1] The reason why the concept of central node is not applicable to dynamic asset graphs is that nothing guarantees that we have just one graph. We may have many that are not connected, and thus we cannot determine how far from each other these graphs are.

and their omission leads to a more fragile structure. This is also witnessed by studying the multi-step survival ratio, where it was found that in the early time horizon the asset graph shows exponential decay, but the asset tree decays faster than exponential. Later on, however, both decay as a power-law, but here the situation is reversed and the asset tree decays more slowly than the asset graph. We also studied the vertex degree distributions produced by the two alternative approaches. Earlier we have found asset trees to exhibit clear scale free behavior, but for the asset graph scale free behavior is not so evident. Further, the values obtained for the scaling exponent are very different from our earlier studies with asset trees.

Acknowledgements

J.-P. O. is grateful to European Science Foundation for a REACTOR grant to visit Hungary, the Budapest University of Technology and Economics for the warm hospitality. Further, the role of Harri Toivonen at the Department of Accounting, Helsinki School of Economics, is acknowledged for carrying out CRSP database extractions. J.-P. O. is also grateful to the Graduate School in Computational Methods of Information Technology (ComMIT), Finland. The authors are also grateful to R. N. Mantegna for very useful discussions and suggestions. This research was partially supported by the Academy of Finland, Research Center for Computational Science and Engineering, project no. 44897 (Finnish Center of Excellence Program 2000-2005) and OTKA (T029985).

References

1. Mantegna, R. N., Eur. Phys. J. B **11**, 193 (1999).
2. Bonanno, G., Lillo, F. and Mantegna, R. N., Quantitative Finance **1**, 96 (2001).
3. Kullmann, L., Kertész, J. and Mantegna, R. N., Physica A **287**, 412 (2000).
4. Bonanno, G., Vandewalle, N. and Mantegna, R. N., Phys. Rev. E **62**, R7615 (2000).
5. Kullmann, L., Kertész, J. and Kaski, K., preprint available at cond-mat/0203278 (2002).
6. Laloux, L. *et al.*, Phys. Rev. Lett. **83**, 1467 (1999); V. Plerou *et al.*, preprint available at cond-mat/9902283 (1999).
7. Giada, L. and Marsili, M., preprint available at cond-mat/0204202 (2002).
8. Lillo, F. and Mantegna, R. N., preprint available at cond-mat/0209685 (2002).
9. Lillo, F. and Mantegna, R. N., preprint available at cond-mat/0111257 (2001).
10. Onnela, J.-P., "Taxonomy of Financial Assets," M.Sc. Thesis, Helsinki University of Technology, Finland (2002).
11. Onnela, J.-P., Chakraborti, A., Kaski, K. and Kertész, J., Eur. Phys. J. B **30**, 285 (2002).
12. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J. and Kanto, A., Phys. Rev. E., in press (2003).
13. Onnela, J.-P., Chakraborti, A., Kaski, K. and Kertész, J., Physica A **324**, 247 (2003).
14. Supplementary material on the dataset is available at http://www.lce.hut.fi/~jonnela/.
15. Vandewalle, N., Brisbois, F. and Tordoir, X., Quantitative Finance **1**, 372 (2001).