Dynamics of market correlations: Taxonomy and portfolio analysis

Onnela, J-P., et al. (2003)

경상국립대학교 수학과 석사과정 오서영

1. Introduction

Hierarchical arrangement of stocks through studying the clustering of companies by using correlations of asset returns (Mantegna, 1999)
- In this paper, we study the **time dependent** properties of the **minimum spanning tree** and call it a '**dynamic asset tree**'.
- **Dynamic asset trees** can be used to simplify this complexity in order to grasp the essence of the market without drowning in the abundance of information.

The robustness of tree topology and the consequences of the market events on its structure.
- The minimum spanning tree, as a strongly pruned representative of asset correlations, is found to be robust and descriptive of stock market events.

Apply dynamic asset trees in the field of **portfolio optimization**
- expect a connection between dynamic asset trees and the Markowitz portfolio optimization scheme
- although the topological structure of the tree changes with time, the companies of the minimum risk Markowitz portfolio are always located on the outer leaves of the tree.

2. Return correlations and dynamic asset trees

Dataset

The financial market refers to a set of data commercially available from the Center for Research in Security Prices (CRSP) of the University of Chicago Graduate School of Business.

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 02-Jan-1980 to 31-Dec-1999
5056 price quotes per stock.

The data is divided time-wise into M windows $t = 1, 2, \ldots, M$ of width T corresponding to the number of daily returns included in the windows.
Several consecutive windows overlap with each other, the extent of which is dictated by the window step length parameter δT , describing the displacement of the window, measured also in trading days.
($\delta T \approx 20.8$ days and $T = 1000$ days, the overall number of windows is $M = 195$)

2. Return correlations and dynamic asset trees

Closure price of stock i at time τ by $P_i(\tau)$, τ refers to a date
Logarithmic return of stock i:

$$r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1)$$

Those encompassing the given window t, form the return vector ${r^t}_i$.
In order to characterize the synchronous time evolution of assets,
we use the equal time correlation coefficients between assets i and j defined as:

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

<…> indicates a time average over the consecutive trading days

2. Return correlations and dynamic asset trees

Characterize the correlation coefficient distribution by its first four moments and their correlations with one another.
The first moment is the mean correlation coefficient:

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

The non-diagonal elements $p_{ij}{}^t$ of the upper (or lower) triangular matrix

We evaluate the higher order moments for the correlation coefficients, so that the variance is:

$$\lambda_2(t) = \frac{1}{N(N-1)/2} \sum_{(i,j)} (\rho_{ij}^t - \bar{\rho}^t)^2,$$

2. Return correlations and dynamic asset trees

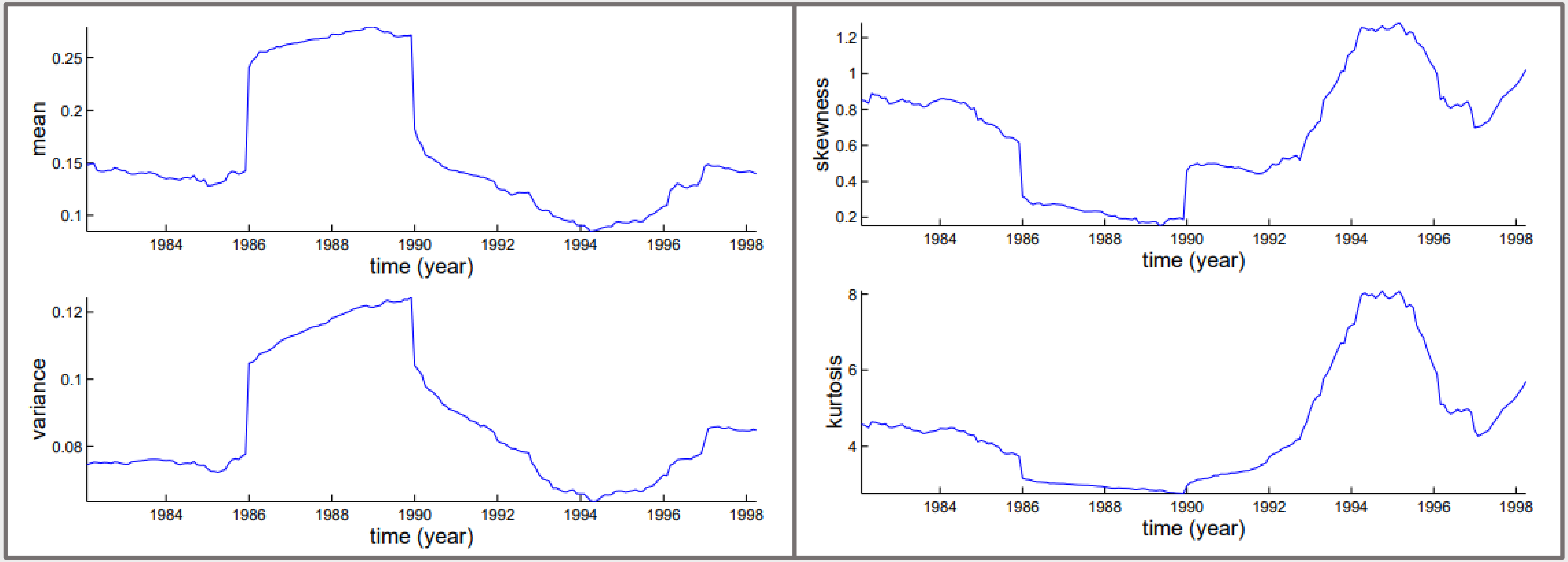
The skewness is:

$$\lambda_3(t) = \frac{1}{N(N-1)/2} \sum_{(i,j)} (\rho_{ij}^t - \bar{\rho}^t)^3 / \lambda_2^{3/2}(t),$$

The kurtosis is:

$$\lambda_4(t) = \frac{1}{N(N-1)/2} \sum_{(i,j)} (\rho_{ij}^t - \bar{\rho}^t)^4 / \lambda_2^2(t).$$

2. Return correlations and dynamic asset trees



The effect and repercussions of Black Monday (October 19, 1987)

2. Return correlations and dynamic asset trees

Whether these four different measures are correlated
we determined the Pearson's linear and Spearman's rank-order correlation coefficients,
which between the mean and variance turned out to be 0.97 and 0.90
Between skewness and kurtosis 0.93 and 0.96, respectively.
→ Thus the first two and the last two measures are very strongly correlated.

Construct an asset tree.

- Non linear transformation $d_{ij} = \sqrt{2(1-\rho_{ij})}$
- Ultrametric space

→ This hypothesis is motivated a posteriori by the finding that the associated taxonomy is meaningful from an economic point of view. (Mantegna, 1999)

- MST : simply connected graph that connects all N nodes of the graph P 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.

2. Return correlations and dynamic asset trees

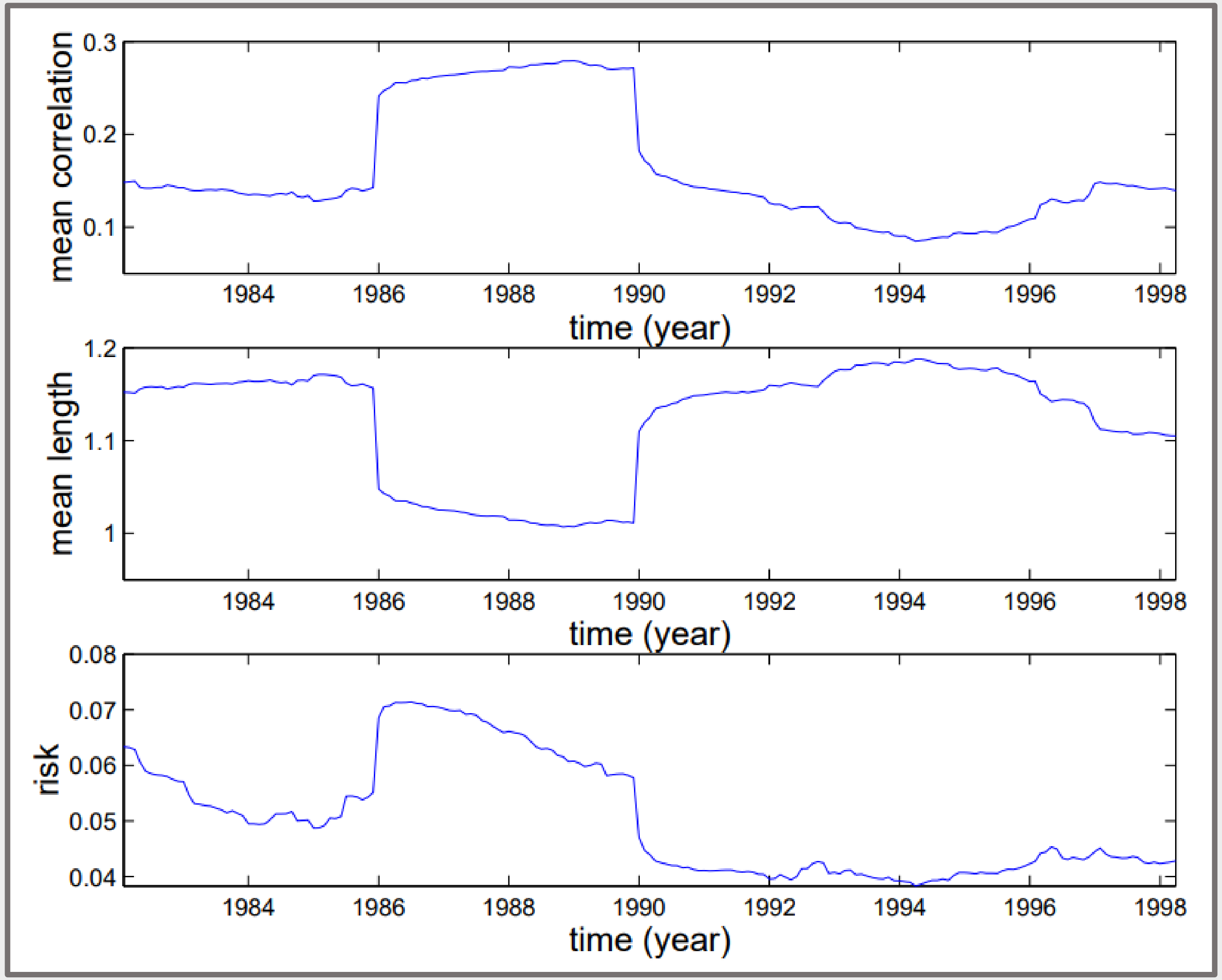
Asset trees constructed for different time windows are not independent from each other, but form a series through time.
→ Consequently, this multitude of trees is interpreted as a sequence of evolutionary steps of a single dynamic asset tree.

As a simple measure of the temporal state of the market (the asset tree), we define the **normalized tree length** as:

$$L(t) = \frac{1}{N-1} \sum_{d_{ij}^t \in \mathbf{T}^t} d_{ij}^t,$$

where t again denotes the time at which the tree is constructed, and N−1 is the number of edges present in the MST.

2. Return correlations and dynamic asset trees



(a) the mean correlation coefficient $\bar{\rho}(t)$,
(b) the normalized tree length L(t),
(c) the risk of the minimum risk portfolio
,as functions of time.

2. Return correlations and dynamic asset trees

MST effectively reduce the information space from $N(N-1)/2$ separate correlation coefficients to $N-1$ tree edges.

- Whether essential information is lost in the reduction.

1. The two measures are strongly anti-correlated testifies to the success of the pruning process.

→ One is justified to contemplate the MST as a strongly reduced representative
of the whole correlation matrix, which bears the essential information about asset correlations.

2. The MST retains the salient features of the stock market (1987 market crash)

- The market ,during crash, is moving together is thus manifested in two ways.
    1) The ridge in the plot of the mean correlation coefficients indicates that
    the whole market is exceptionally strongly correlated. (Fig 2. (a))
    2) The corresponding well in the plot of the normalized tree length shows
    how this is reflected in considerably shorter than average length of the tree so that the tree, on average, is very tightly packed.
    Upon letting the window width $T \to 0$, the two sides of the ridge converge to a single date, which coincides with Black Monday. (Fig 2. (b))

3. Tree occupation and central vertex

Characterizing the spread of nodes on the tree

- The quantity of mean occupation layer:

$$l(t, v_c) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{lev}(v_i^t),$$

where $lev(v_i)$ denotes the level of vertex v_i .

The levels, not to be confused with the distances d_{ij} between nodes, are measured in natural numbers in relation to the central vertex v_c ,whose level is taken to be zero.

→ The mean occupation layer indicates the layer on which the mass of the tree, on average, is conceived to be located.

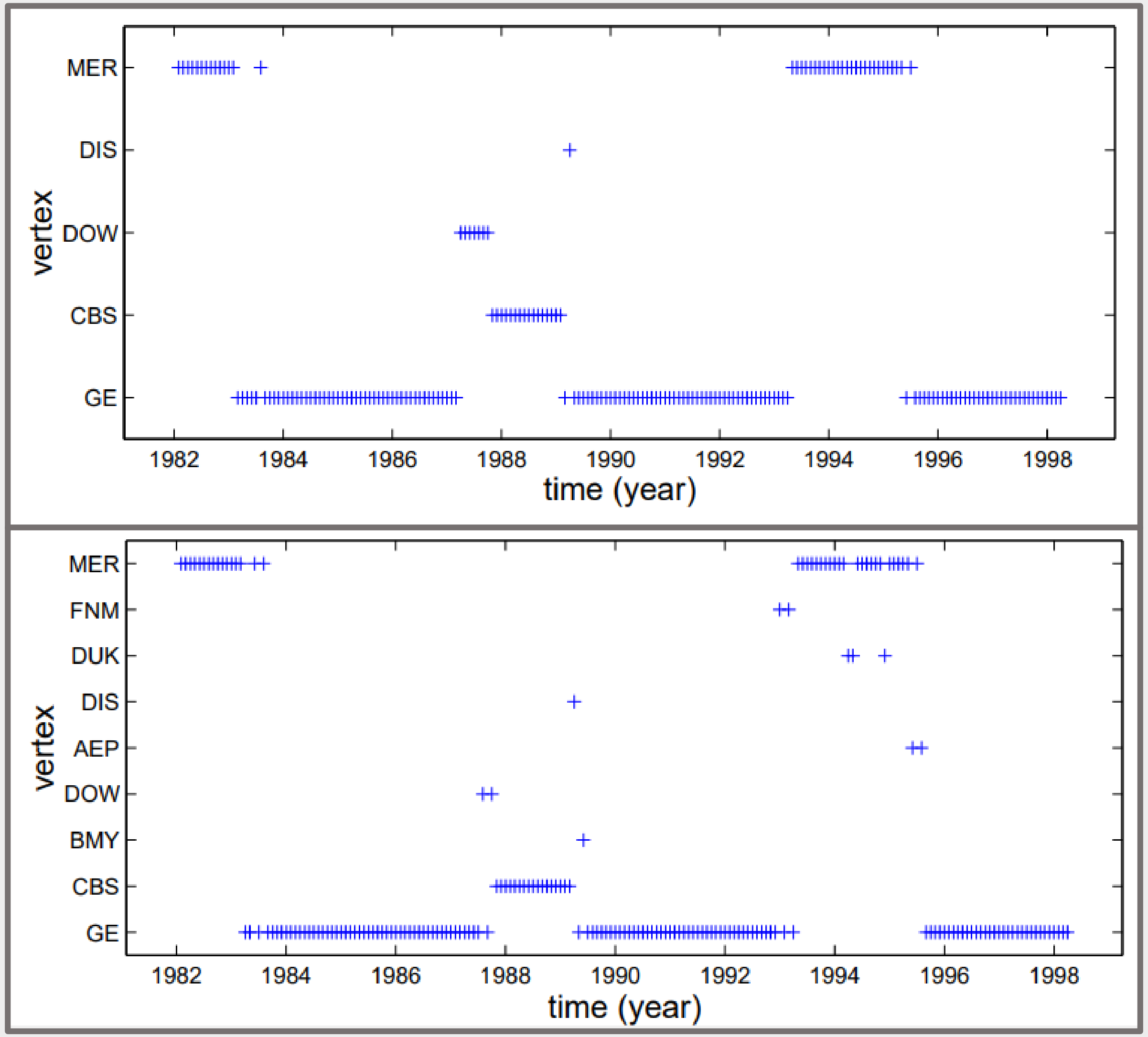
→ Root of the tree

3. Tree occupation and central vertex

Three alternative definitions have emerged for the **central vertex** in studies:

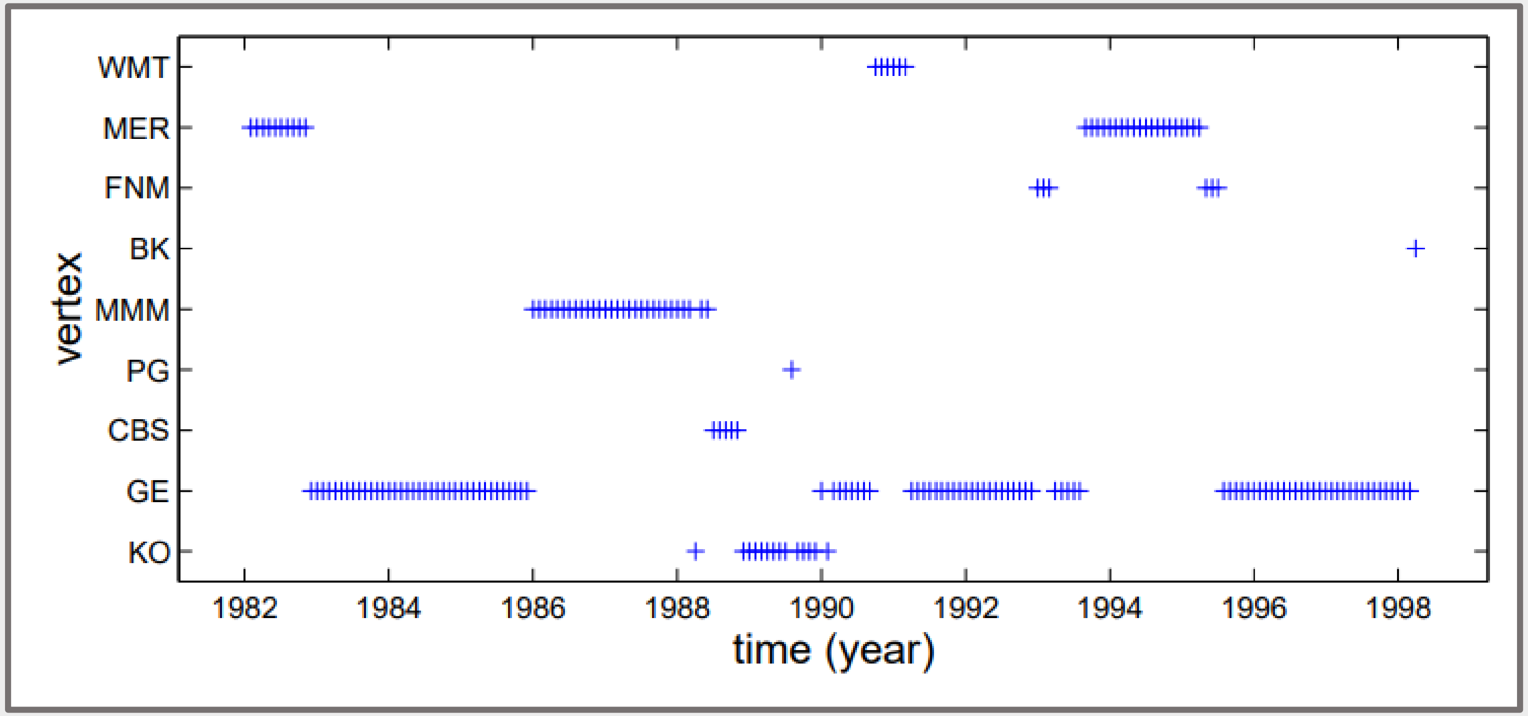
1. (Vertex degree criterion) The node with the highest vertex degree, i.e. the number of edges which are incident with (neighbor of) the vertex.

2. (Weighted vertex degree criterion) Defines the central vertex as the one with the highest sum of those correlation coefficients that are associated with the incident edges of the vertex.
-> gives more weight to short edges, since a high value of p_{ij} corresponds to a low value of d_{ij}.
-> This is reasonable, as short connections link the vertex more tightly to its neighborhood than long ones.



3. Tree occupation and central vertex

3. **(Center of mass criterion)** In considering a tree T_t at time t, the vertex v_i that produces the lowest value for mean occupation layer $l(t, v_i)$ is the center of mass, given that all nodes are assigned an equal weight and consecutive layers (levels) are at equidistance from one another.



3. Tree occupation and central vertex

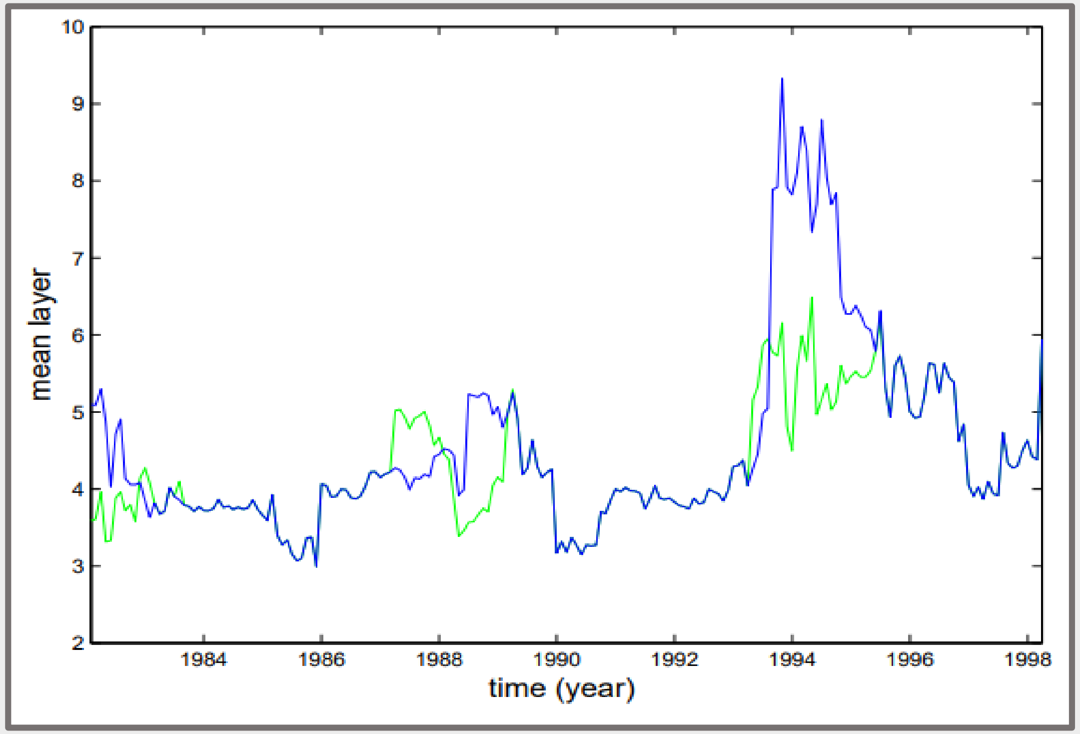
Three alternative definitions for the central vertex lead to very similar results.
The vertex degree and the weighted vertex degree criteria coincide 91.8% of the time.
Overall, the three criteria yield the same central vertex in 63.6% of the cases.

A vertex with a high vertex degree carries a lot of weight around it, which in turn may be highly connected to others.

Two different interpretations may be given to these results.

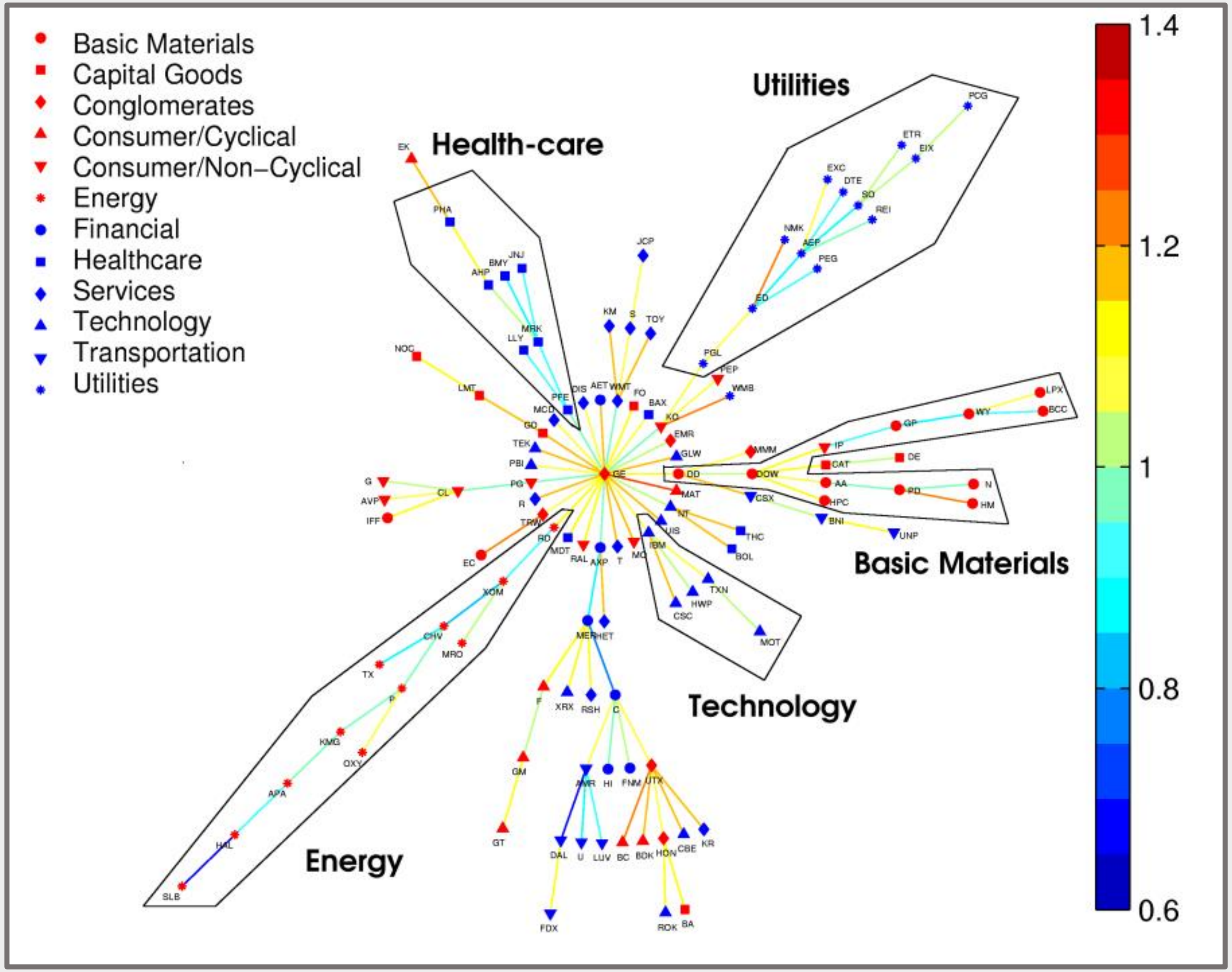
(i) static (fixed at all times) central vertex

(ii) dynamic (updated at each time step) central vertex.



Mean occupation layer $l(t, v_i)$ as a function of time, with static and dynamic central vertices

4. Tree clusters and their economic meaningfulness



Snapshot of a dynamic asset tree connecting the examined 116 stocks of the S&P 500 index.
The tree was produced using four-year window width and it is entered on January 1, 1998.

General Electric (GE) was used as the central vertex and eight layers can be identified.

4. Tree clusters and their economic meaningfulness

The term **branch** refers to a subset of the tree, to all the nodes that share the specified common parent.

- There are some branches in the tree, in which most of the stocks belong to just one sector, indicating that the branch is fairly homogeneous with respect to business sectors
- Since the grouping of stocks is not perfect at the branch level, we define a smaller **subset** whose members are more homogeneous.

The term **cluster** is defined as a subset of a branch, but a more accurate definition is based on the following four rules.

(i) A cluster is named after the cluster parent, which is the node in the cluster closest to the central vertex and it is the starting node of the cluster.

(ii) If there are more than one potential cluster parent, the one resulting in the most complete cluster is chosen as the cluster parent. The nodes that are left outside the formed cluster are considered outliers.

(iii) Only those edges that are required to connect the cluster are included.

(iv) If there are nodes in a cluster which do not belong there, and they do not have children that belong to the cluster either, they are not included.

4. Tree clusters and their economic meaningfulness

Complete and **incomplete**

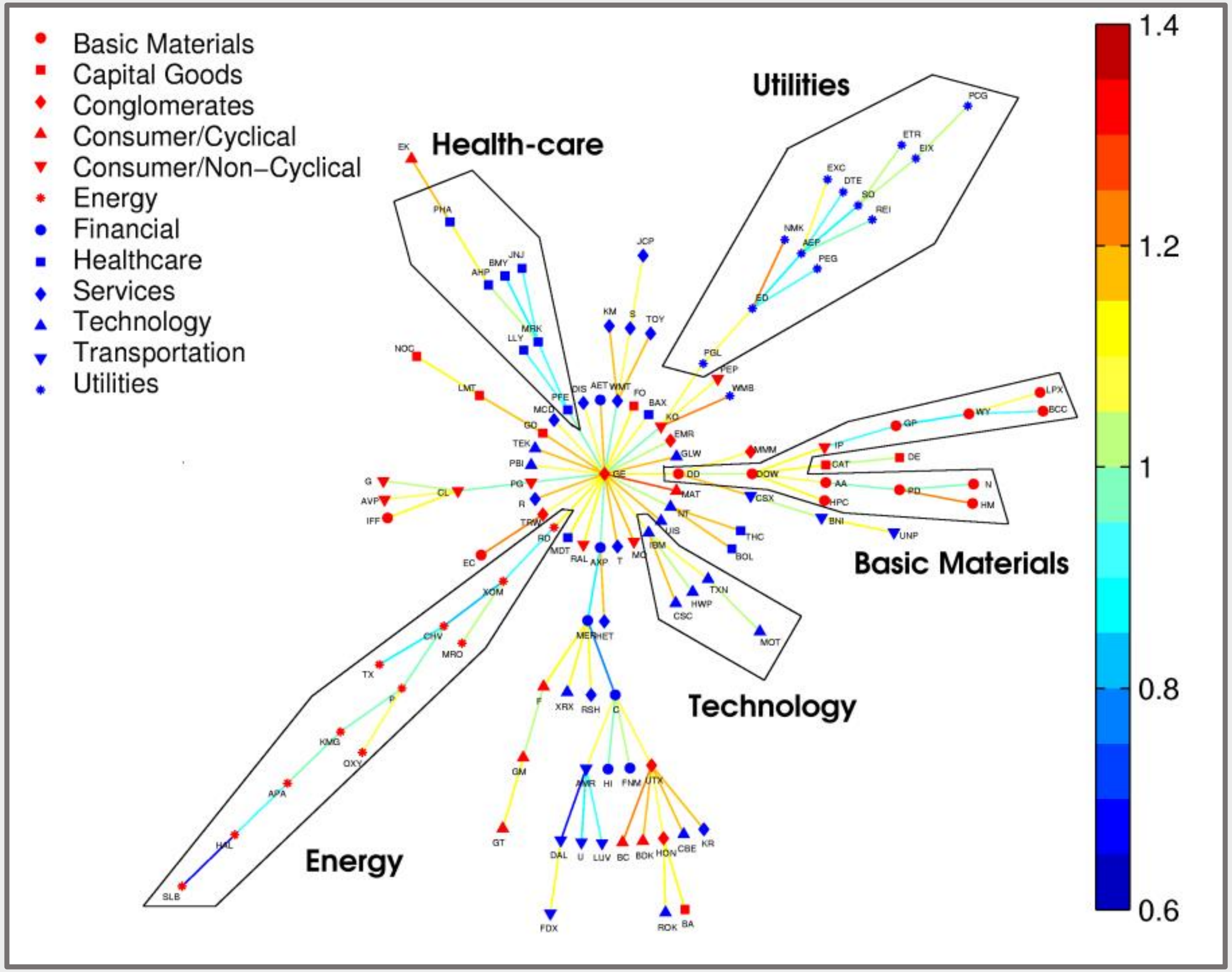
- A complete cluster contains all the companies of the studied set belonging to the corresponding business sector, so that none are left outside the cluster.
- In practice, however, clusters are mostly incomplete, containing most, but not all, of the companies of the given business sector, and the rest are to be found somewhere else in the tree.
- Only the Energy cluster was found complete.

These clusters, whether complete or incomplete, are characterized by the normalized cluster length, defined for a cluster c as follows:

$$L_c(t) = \frac{1}{N_c} \sum_{d_{ij}^t \in c} d_{ij}^t,$$

where Nc is the number of stocks in the cluster

4. Tree clusters and their economic meaningfulness



Compared with the normalized tree length, which for the sample tree at time $t*$ is $L(t*) \approx 1.05$.

$L_{Energy}(t*) \approx 0.92$,

$L_{Health\text{-}care}(t*) \approx 0.98$.

$L_{Utilities}(t*) \approx 1.01$.

$L_{Basic\ materials}(t*) \approx 1.03$.

$L_{Tehnology}(t*) \approx 1.07$.

-> Most clusters seem to be more tightly packed than the tree on average

4. Tree clusters and their economic meaningfulness

MST seems to provide a taxonomy that is well compatible with the sector classification provided by an outside institution.

- There are, however, some observed deviations to the classification ,which all for an explanation.

(i) Uncertainty in asset prices in the minds of investors causes some seemingly random price fluctuations to take place, and this introduces “noise” in the correlation matrix.

Therefore, it is not reasonable to expect a one-to-one mapping between business sectors and MST clusters.

(ii) Business sector definitions are not unique, but vary by the organization issuing them.

(iii) Historical price time series is old. Therefore, one should use contemporary definitions for business sectors etc., as those most accurately characterize the company.

(iv) In many classification systems, companies engaged in substantially different business activities are classified according to where the majority of revenues and profits comes from.

(v) Some cluster outliers can be explained through the MST clustering mechanism, which is based on correlations between asset returns.

5. Scale free structure of the asset tree

Scale free behavior for the asset tree in a limited time window

- Vandewalle et al proposed the distribution of the vertex degrees f(n) to follow a power law behavior:

$$f(n) \sim n^{-\alpha},$$

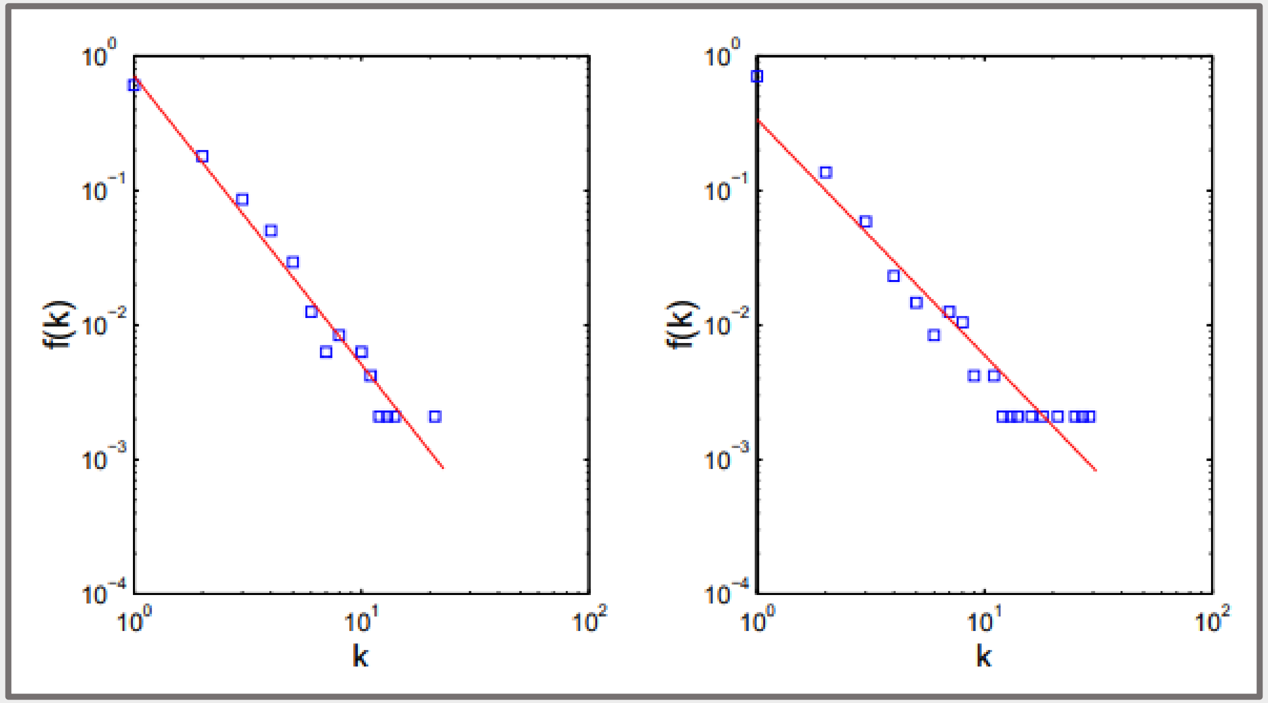
with the exponent $\alpha \approx 2.2$. This exponent implies that the second moment of the distribution would diverge in the infinite market limit.

- The second moment of the distribution is always dominated by the rare but extremely highly connected vertices

5. Scale free structure of the asset tree

Our aim is to study the property of scale freeness in the light of asset tree dynamics:

- The asset tree has, most of the time, scale free properties with a rather robust exponent $\alpha \approx -2.1 \pm 0.1$ for normal topology.
- For most of the time the distribution behaves in a universal manner, meaning that the exponent α is a constant within the error limits.
- However, when the behavior of the market is not 'business as usual' the exponent also changes.



Typical plots of vertex degree for normal (left) Crash topology (right)

6. Asset tree evolution

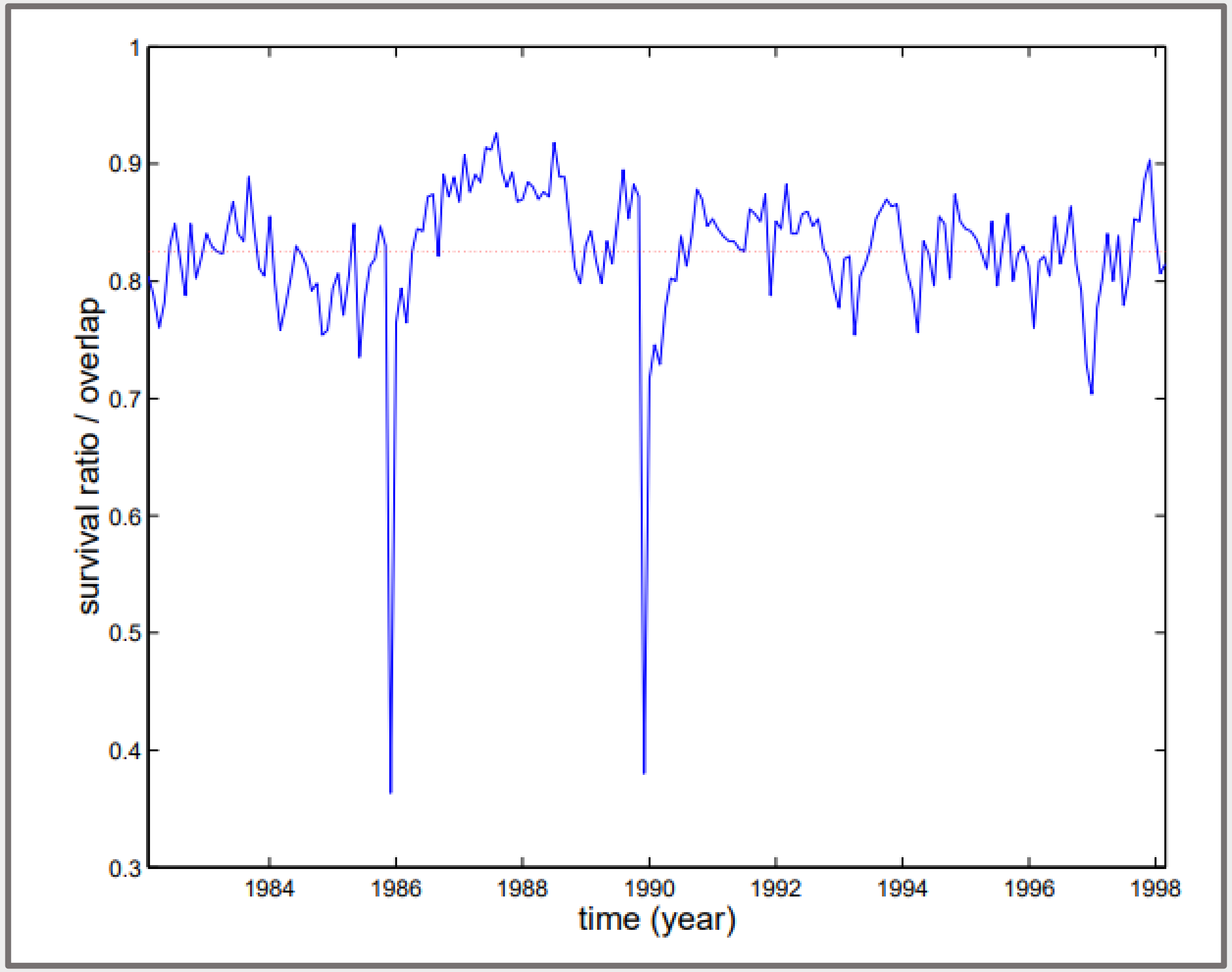
In order to investigate the robustness of asset tree topology, we define the **single-step survival ratio** of tree edges as the fraction of edges found common in two consecutive trees at times t and t−1 as:

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

E(t) refers to the set of edges of the tree at time t, ∩ is the intersection operator and |...| gives the number of elements in the set.
Under normal circumstances, the tree for two consecutive time steps should look very similar, at least for small values of window step length parameter δT.

While some of the differences can reflect real changes in the asset taxonomy, others may simply be due to noise.

6. Asset tree evolution



Single-step survival ratio for T=1000 and δT ≈ 20.8

(i) A large majority of connections survives from one time window to the next.

(ii) The two prominent dips indicate a strong tree reconfiguration taking place, and they are window width T apart, positioned symmetrically around Black Monday, and thus imply topological reorganization of the tree during the market crash

(iii) Single-step survival ratio σ(t) increases as the window width T increases while δT is kept constant. Thus an increase in window width renders the trees more stable with respect to single-step survival of connections.

(iv) Variance of fluctuation around the mean is constant over time, except for the extreme events and the interim period, and it gets less as the window width increases.

6. Asset tree evolution

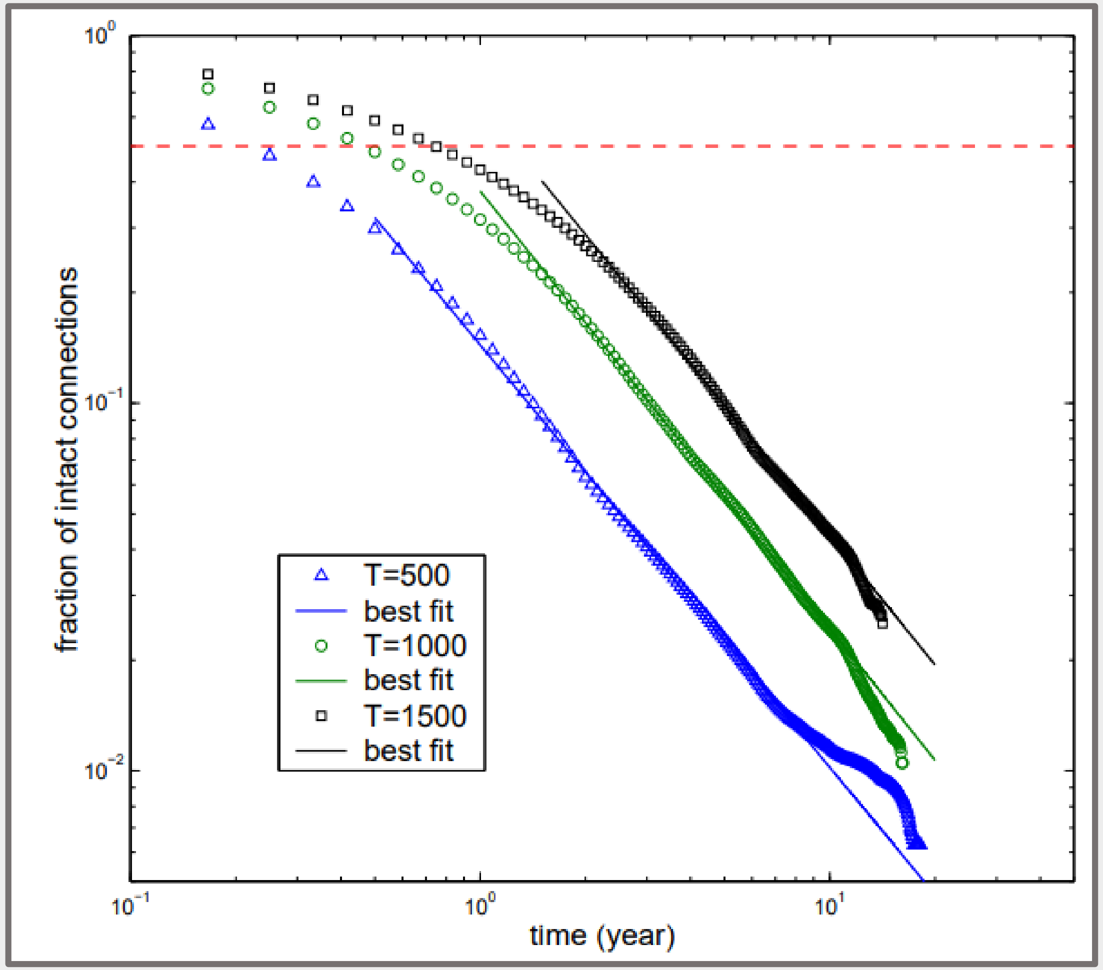
The long term evolution of the trees, we introduce the **multi-step survival ratio** at time t as:

$$\sigma(t,k) = \frac{1}{N-1}|E(t) \cap E(t-1) ... E(t-k+1) \cap E(t-k)|,$$

where only those connections that have persisted for the whole time period without any interruptions are taken into account.

Many connections in the asset trees evaporate quite rapidly in the early time horizon.
However, this rate decreases significantly with time, and even after several years there are some connections that are left intact.
→ some companies remain closely bonded for times longer that a decade.

6. Asset tree evolution



The horizontal axis an be divided into two regions.

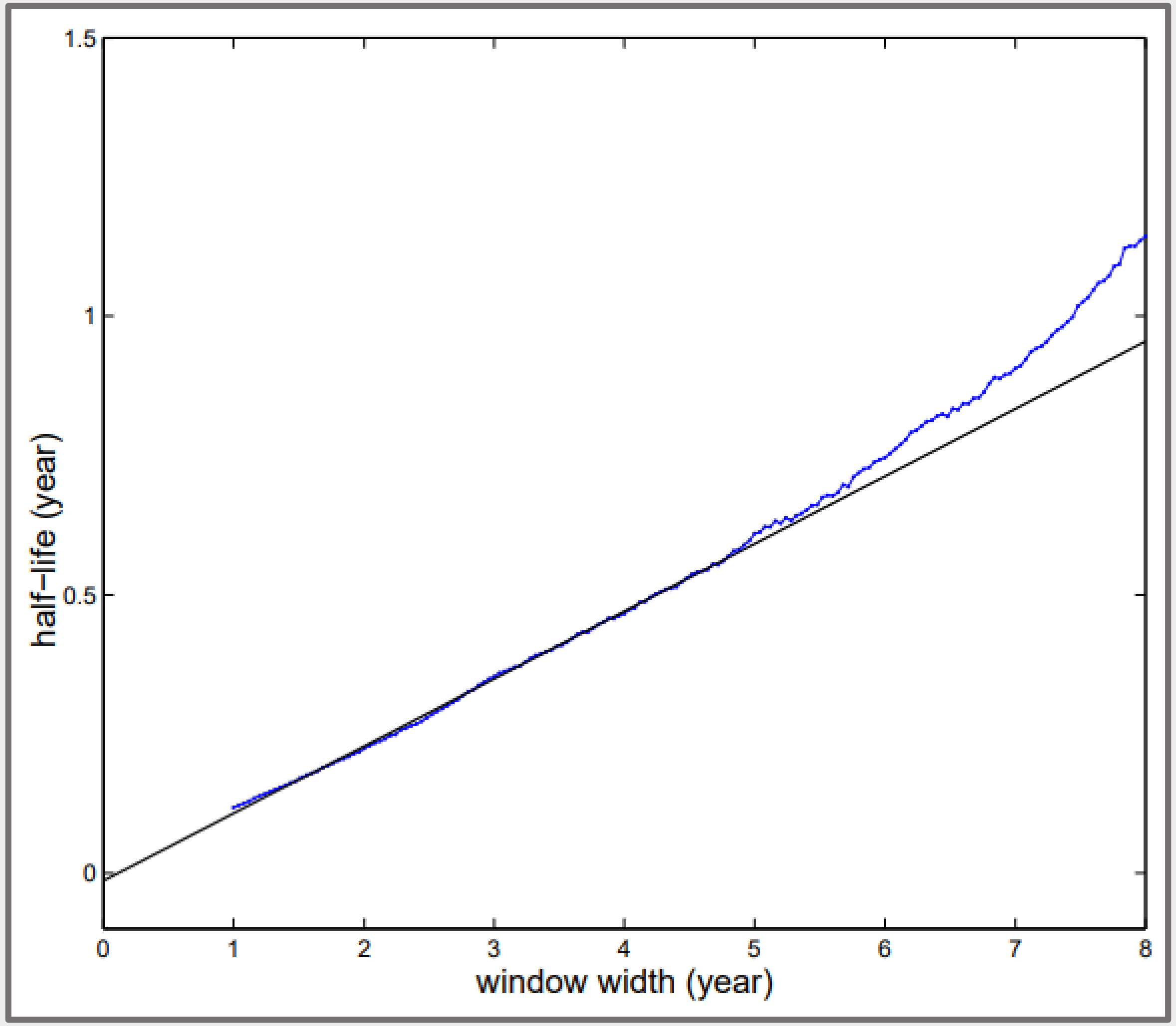
1. Within the first region, decaying of connections is roughly exponential, and takes place at different rates for different values of the window width.

2. Within the second region, when most connections have decayed and only some 20%-30% remain, there is a cross-over to power law behavior.

-> the power law decay in the second region seems independent of the window width.

Multi-step survival ratio for three different values of window width

6. Asset tree evolution



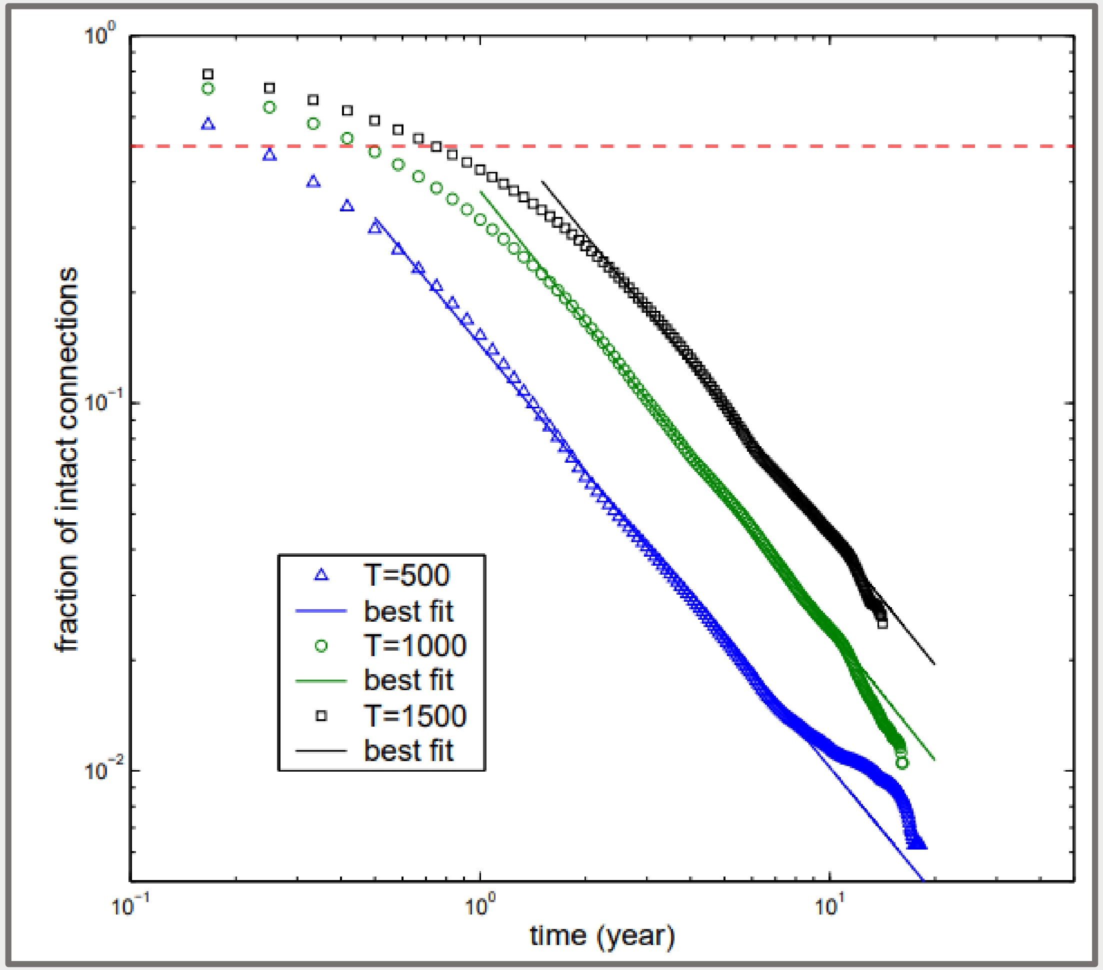
Plots of half-life $t_{1/2}$ as a function of window width T

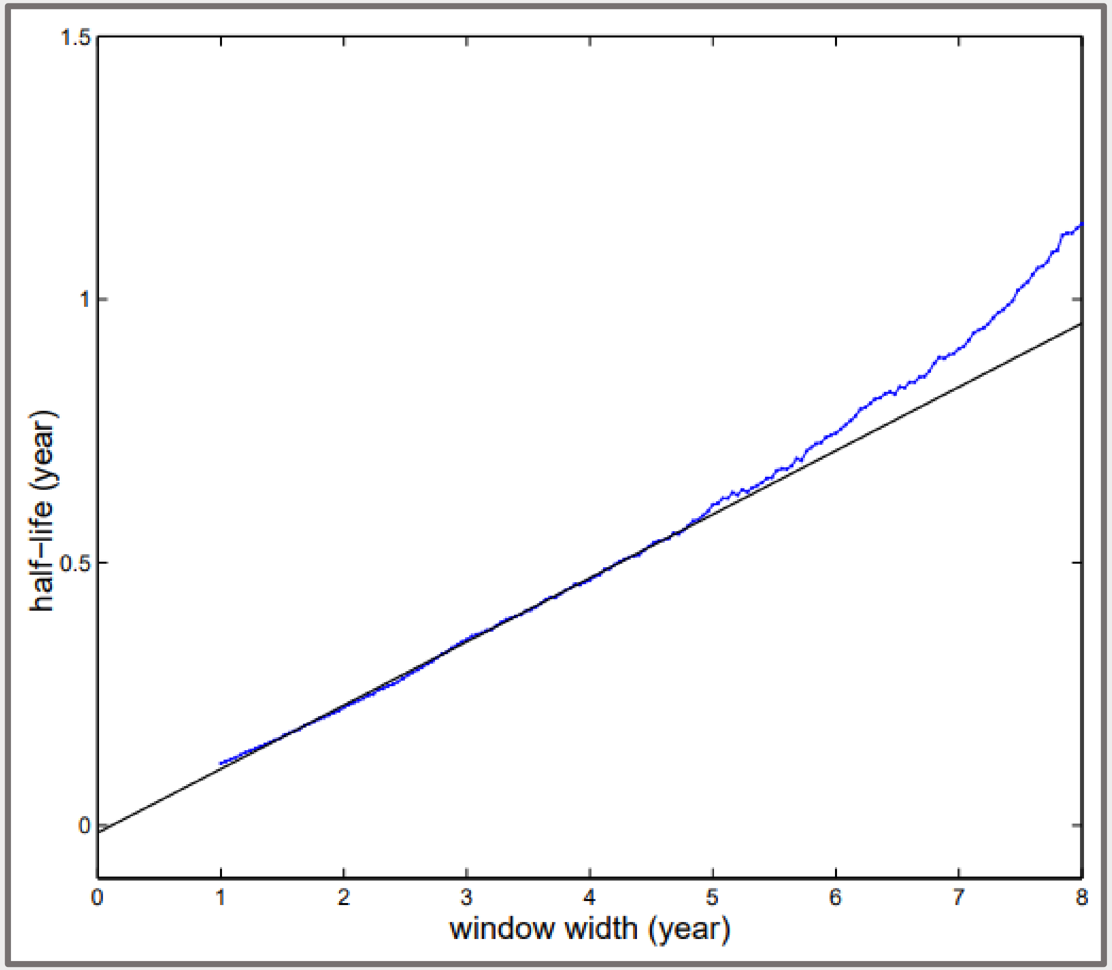
half-life of the survival ratio $t_{1/2}$,
: the time interval in which half the number of initial connections have decayed.

$$\sigma(t, t_{1/2}) = 0.5.$$

Clean linear dependence on for values of T being between 1 and 5 years, after which it begins to grow faster than a linear function.
For the linear region, the tree half-life exhibits $t_{1/2} \approx 0.12T$ dependence.

6. Asset tree evolution





7. Portfolio analysis

Portfolio optimization problem

- the task is to determine how the assets are located with respect to the central vertex.
- Let r_m and r_M denote the returns of the minimum and maximum return portfolios, respectively. The expected portfolio return varies between these two extremes, and can be expressed as:

$$r_{\mathbf{P},\theta} = (1-\theta)r_m + \theta r_M;$$

- where θ is a fraction between 0 and 1. Hence, when θ = 0, we have the minimum risk portfolio, and when θ = 1, we have the maximum return (maximum risk) portfolio.

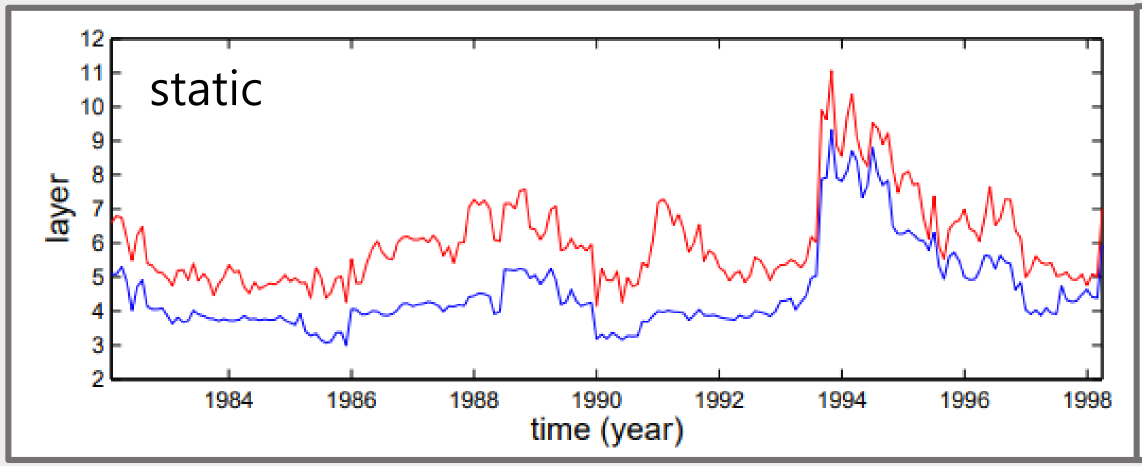
- The higher the value of θ, the higher the expected portfolio return $r_{P,\theta}$.
- Consequently, the higher the risk the investor is willing to absorb.

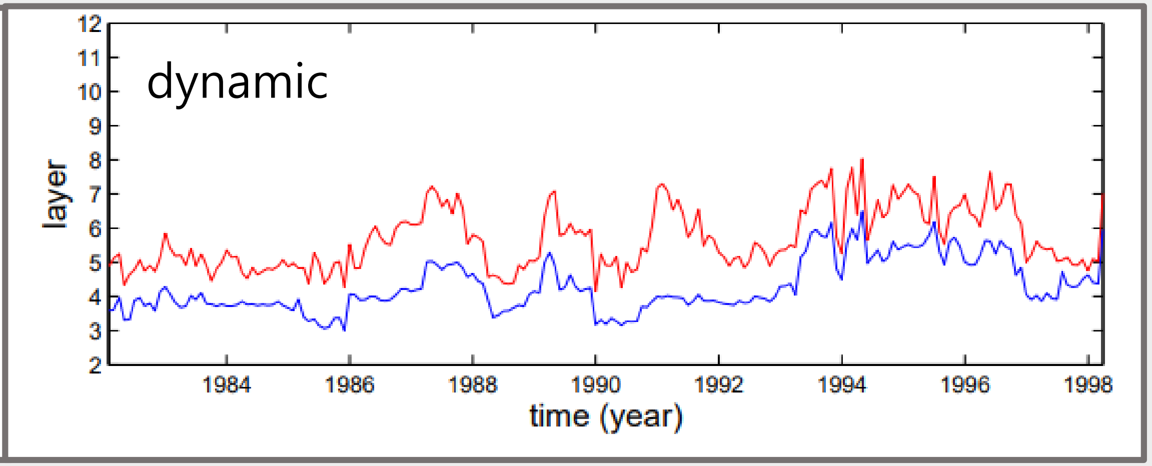
7. Portfolio analysis

- We define a single measure, the **weighted portfolio layer** as:

$$l_{\mathbf{P}}(t,\theta) = \sum_{i \in \mathbf{P}(t,\theta)} w_i \,\mathrm{lev}(v_i^t),$$

where $\sum_{i=1}^{N} w_i = 1$ and further, as a starting point, the constraint $w_i \geq 0$ for all i,
which is equivalent to assuming that there is no short-selling.
→ The purpose of this constraint is to prevent negative values for $l_p(t)$,
which would not have a meaningful interpretation in our framework of trees with central vertex.



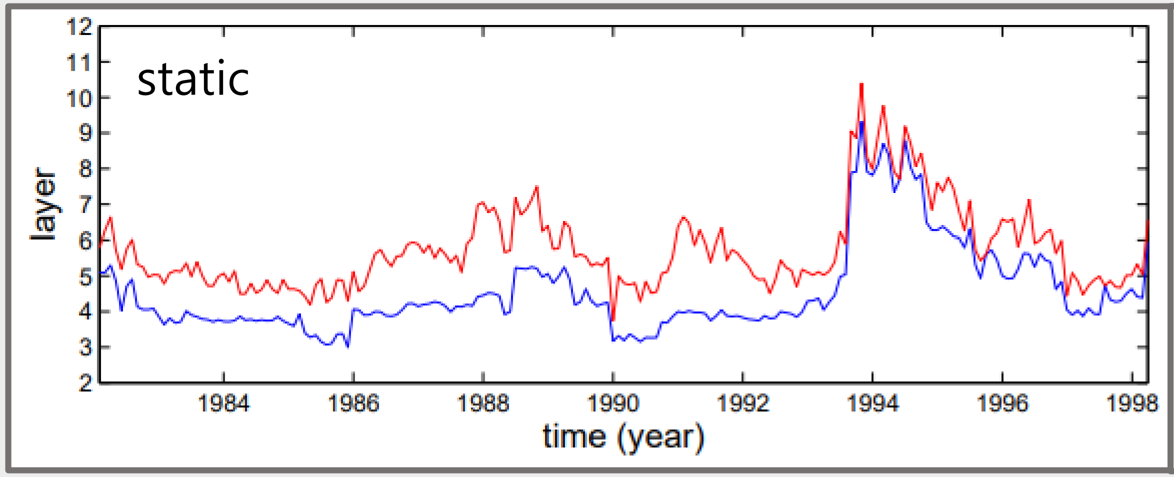


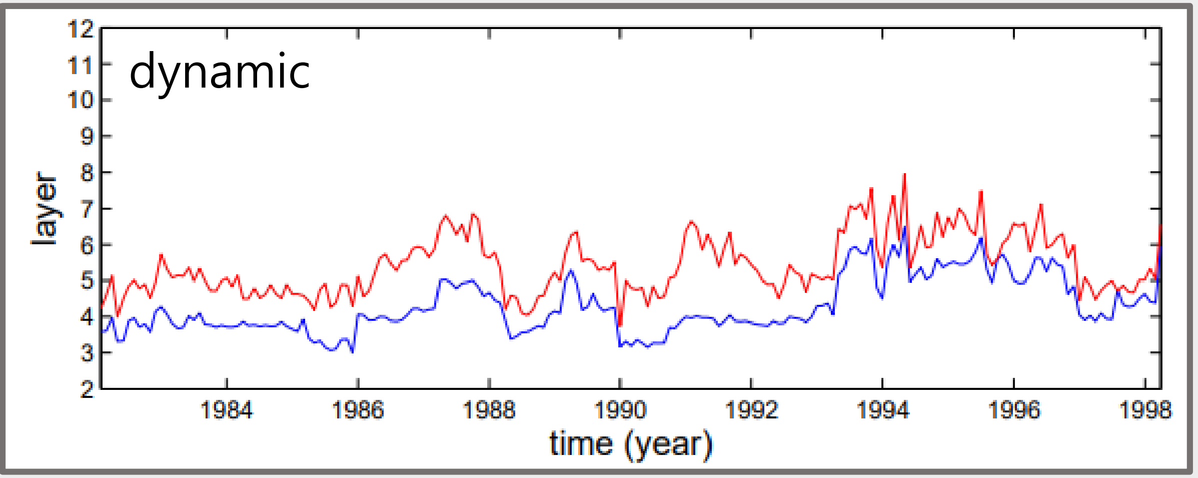
Weighted minimum risk portfolio layer l_p(t, θ=0) with no short-selling and mean occupation layer $l(t, v_i)$ against time.

7. Portfolio analysis

The weighted portfolio layer never assumes negative values and the short selling condition, in fact, is not necessary → allowing for short-selling
The weighted portfolio layer is now 99.5% of the time higher than the mean occupation layer and, with the same central vertex configuration as before, the difference between the two is 1.18 and 1.14 in the upper and lower plots, respectively.

The difference in layers is also slightly larger for static than dynamic central vertex, although not by much. As the stocks of the minimum risk portfolio are found on the outskirts of the tree, we expect larger trees to have greater diversification potential
→ The scope of the stock market to eliminate specific risk of the minimum risk portfolio.





Conclusion

1. The tree evolves over time and have found that the normalized tree length decreases and remains low during a crash, thus implying the shrinking of the asset tree particularly strongly during a stock market crisis.
2. We have also found that the mean occupation layer fluctuates as a function of time, and experiences a downfall at the time of market crisis due to topological changes in the asset tree.
3. Our studies of the scale free structure of the MST show that this graph is not only hierarchic al in the sense of a tree but there are special, highly connected nodes and the hierarchical structure is built up from these.
4. As for the portfolio analysis, it was found that the stocks included in the minimum risk portfolio tend to lie on the outskirts of the asset tree: on average the weighted portfolio layer an be almost one and a half levels higher, or further away from the central vertex, than the mean occupation layer for window width of four years.
5. Correlation between the risk and the normalized tree length was found to be strong, though not as strong as the correlation between the risk and the mean correlation coefficient. Thus we conclude that the diversification potential of the market is very closely related to the behavior of the normalized tree length.