OCS

Correlation, hierarchies, and networks in financial markets

Rosario Nunzio Mantegna

Palermo University, Italy

***O**bservatory of **C**omplex **S**ystems*

OCS

Work done in collaboration with

- Claudia Coronnello
- **Fabrizio Lillo**
- Salvatore Micciché
- **Michele Tumminello**
- Gabriella Vaglica

Some of the work also done in collaboration with

- Tomaso Aste (ANU)
- Tiziana Di Matteo (ANU)
- Esteban Moro (Carlos III, Madrid)

OCS

Overview

Quantifying and modeling information present in a correlation matrix

- Filtering the most stable information of the correlation matrix;

-Hierarchical trees and correlation based trees from correlation matrices;

-Evaluating the statistical robustness of a filtered matrix and with a correlation based tree with a bootstrap approach;

- Modeling hierarchies;

- Quantitatively comparing filtered correlation matrices

OCS

Financial markets as complex systems

A financial market can be considered as a `model’ complex system.

In a financial market there are many heterogeneous agents interacting to perform the collective task of finding the best price for financial assets.

OCS

A basic paradigm: Arbitrage opportunity

One of the main paradigms used for the modeling of a financial market is the absence of **arbitrage opportunity.**

An *arbitrage opportunity* is present in a market when an economic agent can devise a trading strategy which is able to provide her or him a financial gain continuously and without risk.

OCS

St. Louis

Miami

An example

At a given time 1 kg of wheat costs 1.30 USD in St. Louis and 1.45 USD in Miami.

The cost of transporting and storing 1 kg of wheat from St. Louis to Miami is 0.05 USD

By buying 10,000 kg in St. Louis and selling them immediately after in Miami it is possible to make a risk-free profit

10000 (1.45-1.30-0.05)=1000 USD

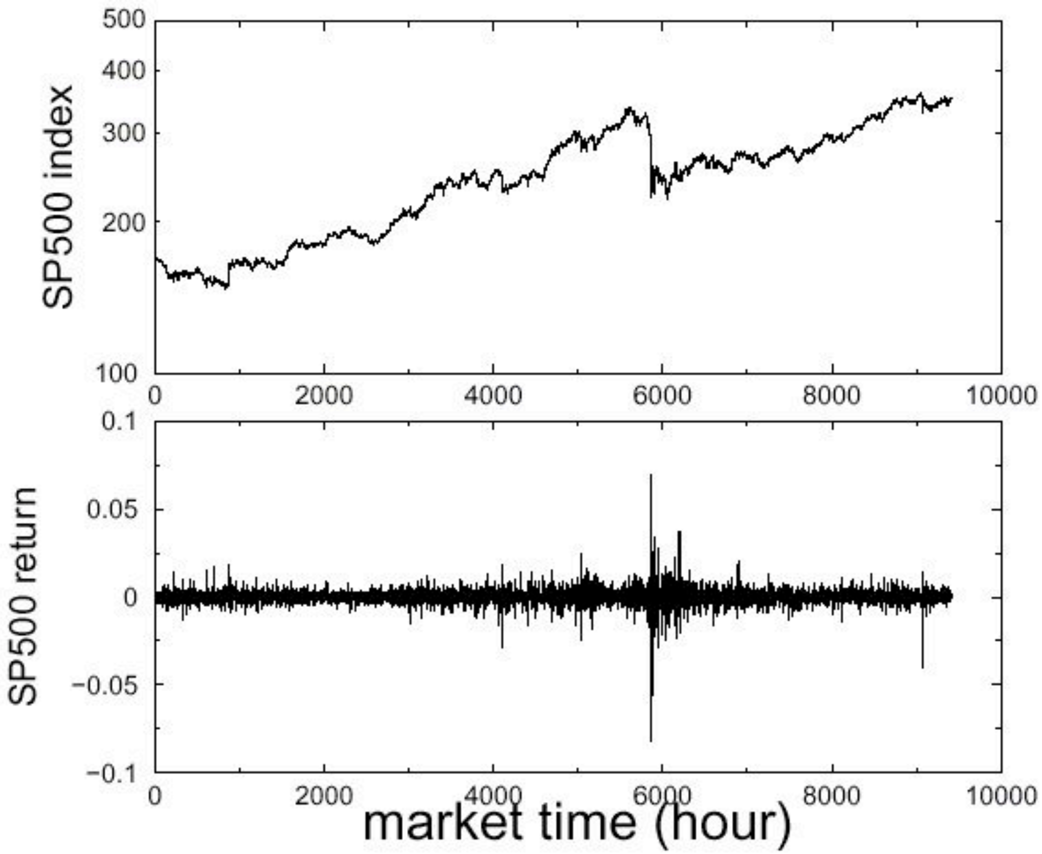
If this action is repeated this implies that the price in St. Louis increases (where the demand increases) and in Miami decreases (where the supply increases).

OCS

Financial assets are unpredictable

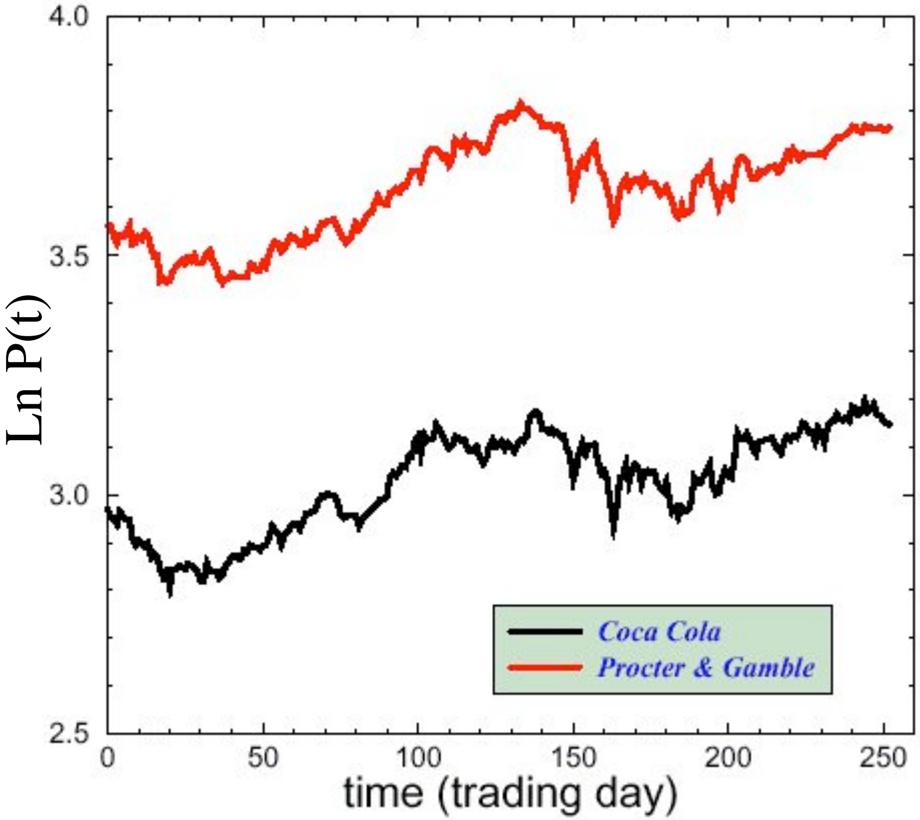
In an efficient market, the continuous exploiting of an arbitrage opportunity implies its disappearance after a (usually) short time period.

The absence of arbitrage opportunities implies that the price dynamics of a financial asset must be **unpredictable.**



OCS

Cross-correlation between stock returns are well-known



They may be quantified by the correlation coefficient ρ_{ij}

OCS

Cross Correlation

N data series of length T

$$r_i(t_j), \quad j = 1,\ldots,T; \quad i = 1,\ldots,N$$

Example:
Log-return of stock price

$$r_i(t) \equiv \ln P_i(t) - \ln P_i(t-\tau)$$

Pearson's correlation coefficient:

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left\langle r_i^2 - \langle r_i \rangle^2 \right\rangle \left\langle r_j^2 - \langle r_j \rangle^2 \right\rangle}}$$

Correlation Matrix

$$C = \left(\rho_{ij}\right)$$

Other correlation estimators:

-Fourier estimator
-Maximum Likelihood
 correlation estimator
-…

OCS

Statistical reliability of cross correlation coefficients

$N\,T$ data $\longrightarrow \sim N^2$ correlation coefficients:

Statistical uncertainty is unavoidably associated with the estimation of the correlation coefficient obtained from a finite number of records.

It is therefore important to device methods to

- Filter statistically reliable information;

- Quantitatively assess the stability of the filtered information;

- Model the filtered information.

How to analyze the complexity of a correlation matrix?

Clustering e.g. Hierarchical Clustering
Super Paramagnetic Clustering
Maximum Likelihood Clustering
Sorting Point Into Neighbors

Correlation Based Networks e.g. Minimum Spanning Tree (MST)
Average Linkage Minimum Spanning Tree
Planar Maximally Filtered Graph (PMFG)

OCS

Hierarchical clustering

By starting from a correlation matrix (which is a similarity measure)

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.413	0.518	0.543	0.529	0.341	0.271	0.231	0.412	0.294
IBM		1	0.471	0.537	0.617	0.552	0.298	0.475	0.373	0.270
BAC			1	0.547	0.591	0.400	0.258	0.349	0.370	0.276
AXP				1	0.664	0.422	0.347	0.351	0.414	0.269
MER					1	0.533	0.344	0.462	0.440	0.318
TXN						1	0.305	0.582	0.355	0.245
SLB							1	0.193	0.533	0.592
MOT								1	0.258	0.166
RD									1	0.590
OXY										1

AXP	MER	0.664
IBM	MER	0.617
SLB	OXY	0.592
BAC	MER	0.591
RD	OXY	0.590
TXN	MOT	0.582
IBM	TXN	0.552
AIG	AXP	0.543
MER	RD	0.440

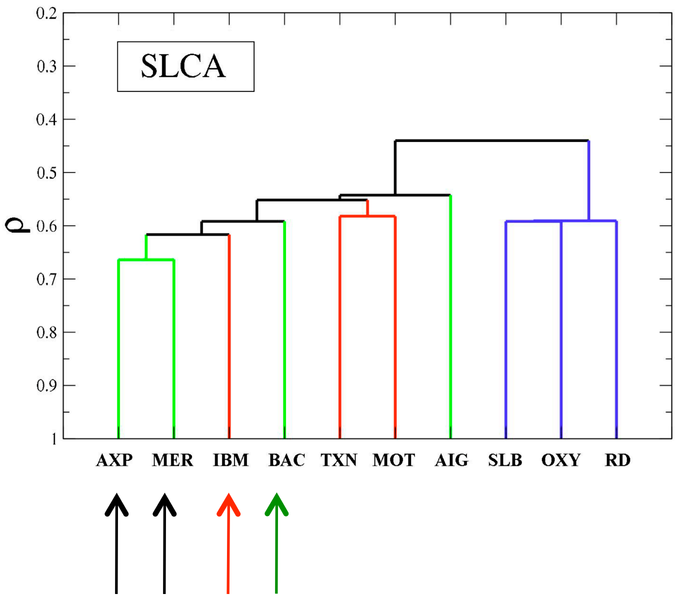
OCS

Hierarchical clustering

One may obtain a simplified matrix by using classical clustering methods such us the single linkage clustering

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.543	0.543	0.543	0.543	0.543	0.440	0.543	0.440	0.440
IBM		1	0.591	0.617	0.617	0.552	0.440	0.552	0.440	0.440
BAC			1	0.591	0.591	0.552	0.440	0.552	0.440	0.440
AXP				1	0.664	0.552	0.440	0.552	0.440	0.440
MER					1	0.552	0.440	0.552	0.440	0.440
TXN						1	0.440	0.582	0.440	0.440
SLB							1	0.440	0.590	0.592
MOT								1	0.440	0.440
RD									1	0.590
OXY										1

$C^{<}_{SL}$



From n(n-1)/2 matrix elements to n-1 matrix elements

OCS

Hierarchical clustering

By starting from a correlation matrix (which is a similarity measure)

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.413	0.518	0.543	0.529	0.341	0.271	0.231	0.412	0.294
IBM		1	0.471	0.537	0.617	0.552	0.298	0.475	0.373	0.270
BAC			1	0.547	0.591	0.400	0.258	0.349	0.370	0.276
AXP				1	0.664	0.422	0.347	0.351	0.414	0.269
MER					1	0.533	0.344	0.462	0.440	0.318
TXN						1	0.305	0.582	0.355	0.245
SLB							1	0.193	0.533	0.592
MOT								1	0.258	0.166
RD									1	0.590
OXY										1

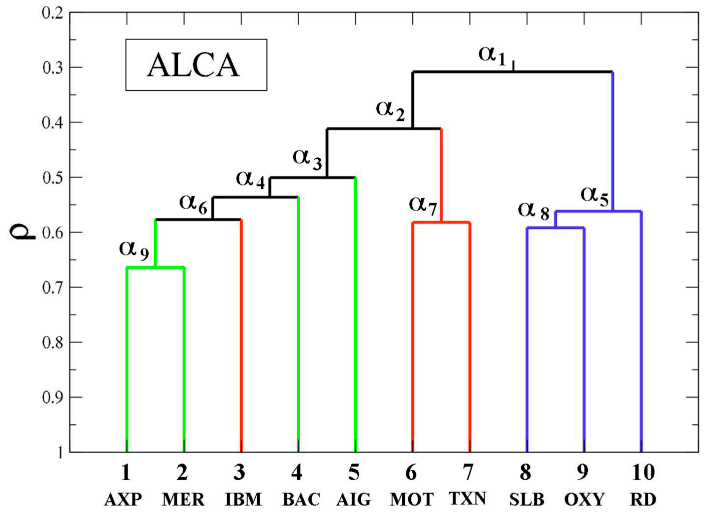
AXP	MER	0.664
IBM	MER	0.617
SLB	OXY	0.592
BAC	MER	0.591
RD	OXY	0.590
TXN	MOT	0.582
IBM	TXN	0.552
AIG	AXP	0.543
MER	RD	0.440

OCS

Hierarchical clustering

Or, for example, the average linkage clustering

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.501	0.501	0.501	0.501	0.412	0.308	0.412	0.308	0.308
IBM		1	0.536	0.577	0.577	0.412	0.308	0.412	0.308	0.308
BAC			1	0.536	0.536	0.412	0.308	0.412	0.308	0.308
AXP				1	0.664	0.412	0.308	0.412	0.308	0.308
MER					1	0.412	0.308	0.412	0.308	0.308
TXN						1	0.308	0.582	0.308	0.308
SLB							1	0.308	0.562	0.591
MOT								1	0.308	0.308
RD									1	0.562
OXY										1



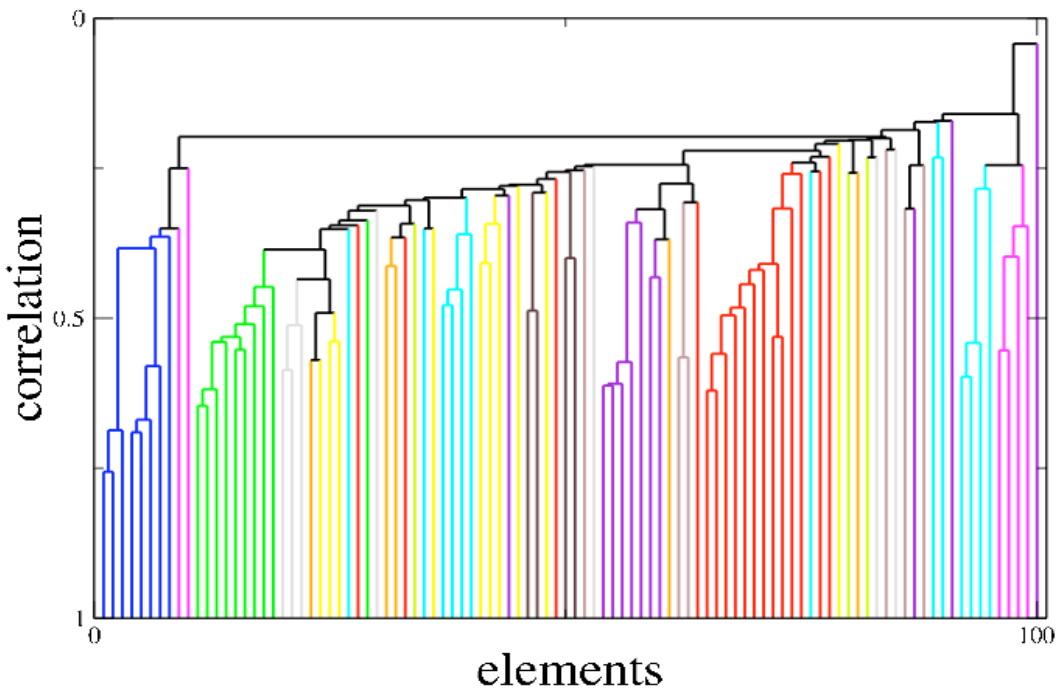
$C^{<}_{AL}$

From n(n-1)/2 matrix elements to n-1 matrix elements

Hierarchical clustering output in a typical case

OCS

$N = 100$ (NYSE) daily returns 1995 - 1998



Average Linkage Cluster Analysis

$$C^{<} = (\rho_{ij}^{<})$$

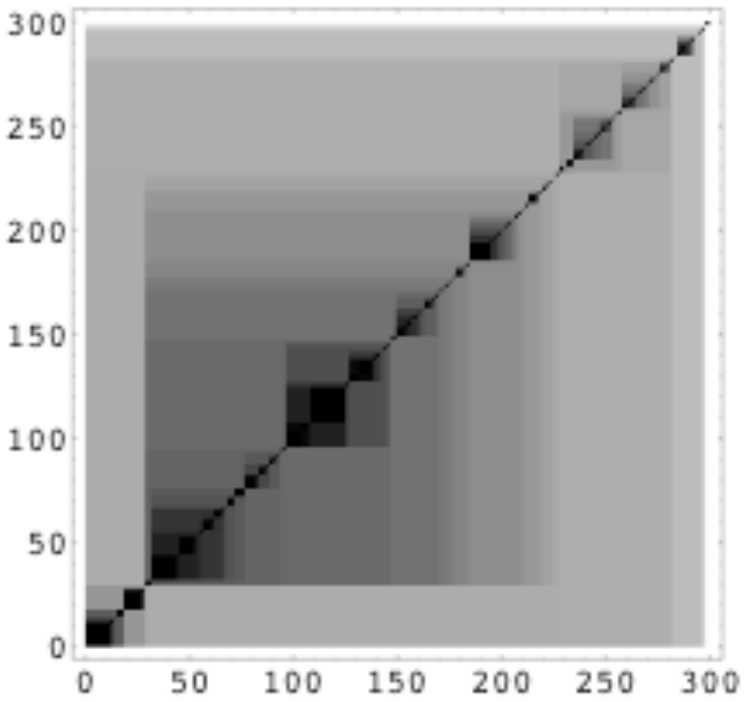
$$\rho_{ij}^{<} = \rho_{\alpha_k}$$

where
α_k
is the first node where elements *i* and *j* merge together

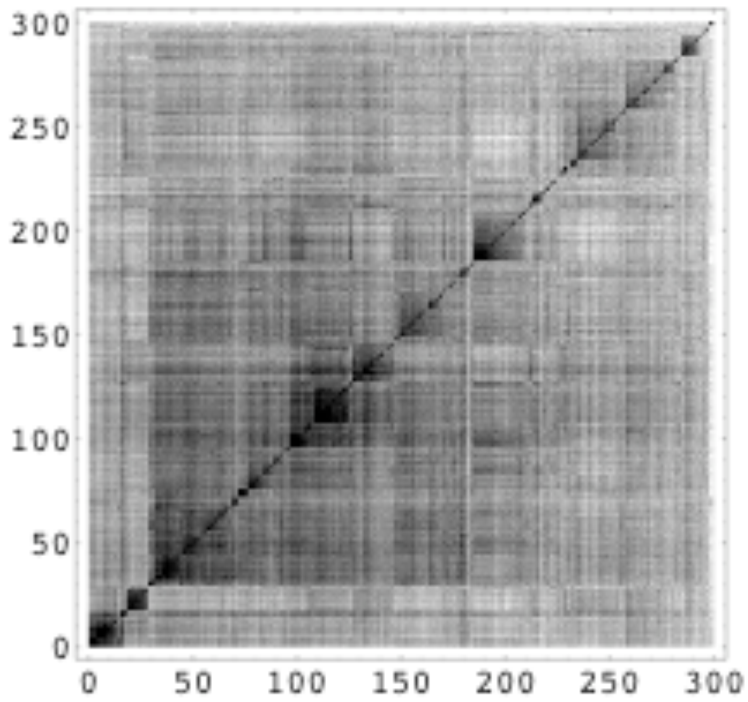
OCS

Filtered matrix

$N = 300$ (NYSE); daily returns 2001 - 2003

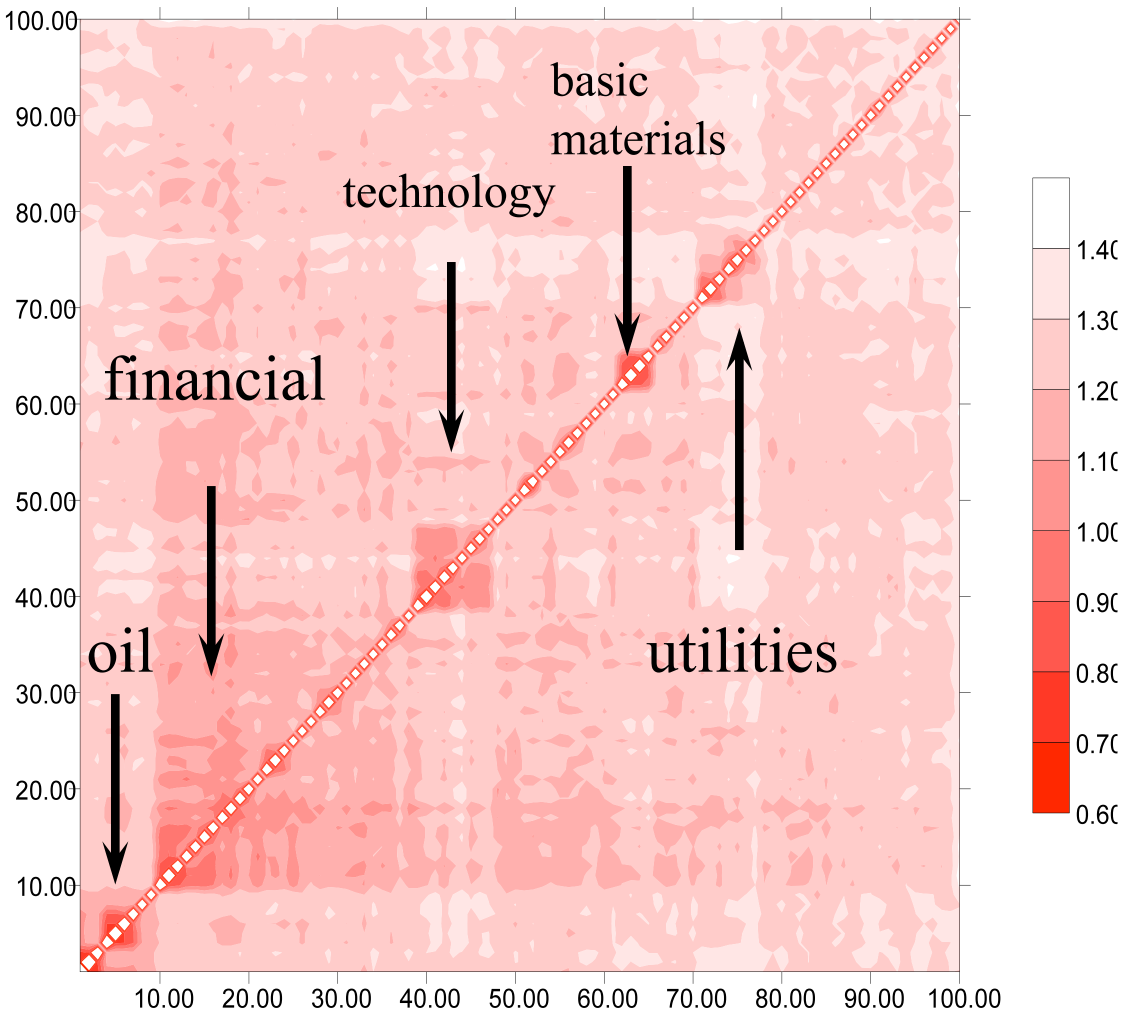


$C^<$ from ALCA



C

The complete matrix is richer of information



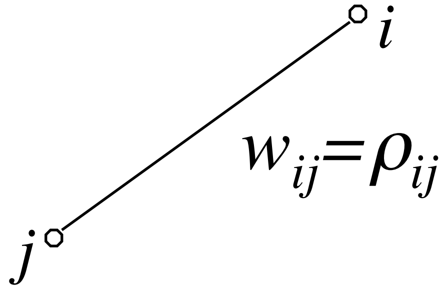
When one uses the stock order of the hierarchical tree the correlation matrix assumes a better readability

n=100 stocks NYSE (1995-1998)

OCS

Correlation based networks

$$(i, j, \rho_{ij}) \longrightarrow$$

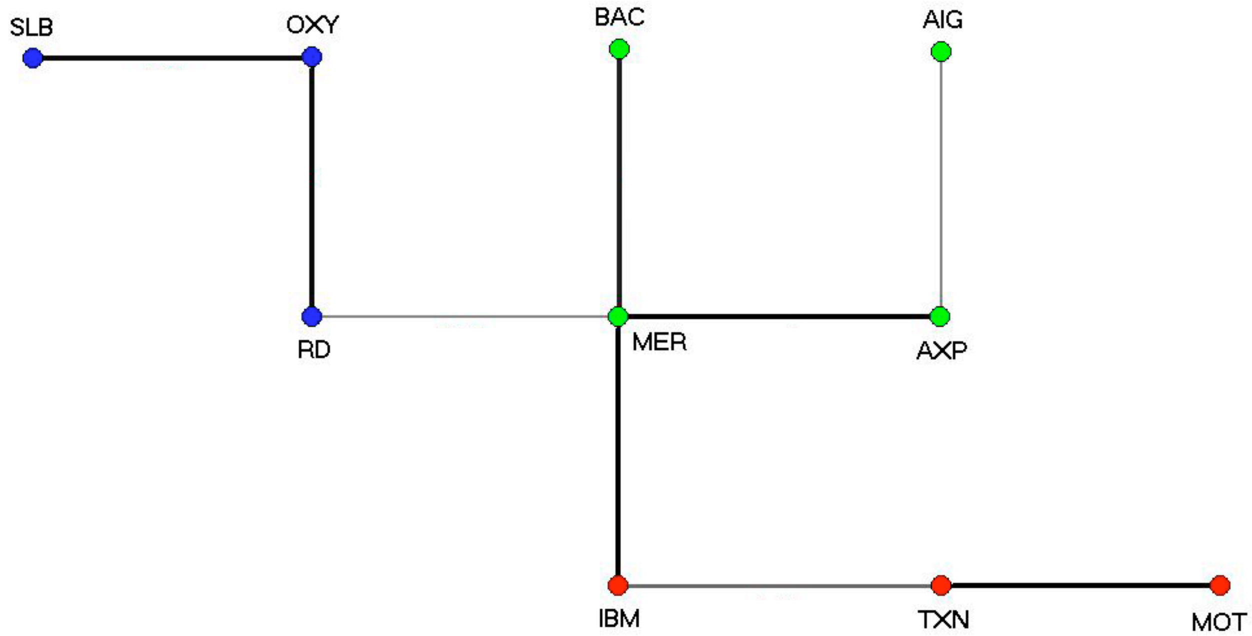


$$C = \begin{pmatrix} 1 & 0.13 & 0.90 & 0.81 \\ 0.13 & 1 & 0.57 & 0.34 \\ 0.90 & 0.57 & 1 & 0.71 \\ 0.81 & 0.34 & 0.71 & 1 \end{pmatrix} \quad \rightarrow \quad S = \begin{pmatrix} 1 & 3 & 0.90 \\ 1 & 4 & 081 \\ 3 & 4 & 0.71 \\ 2 & 3 & 0.57 \\ 2 & 4 & 0.34 \\ 1 & 2 & 0.13 \end{pmatrix}$$

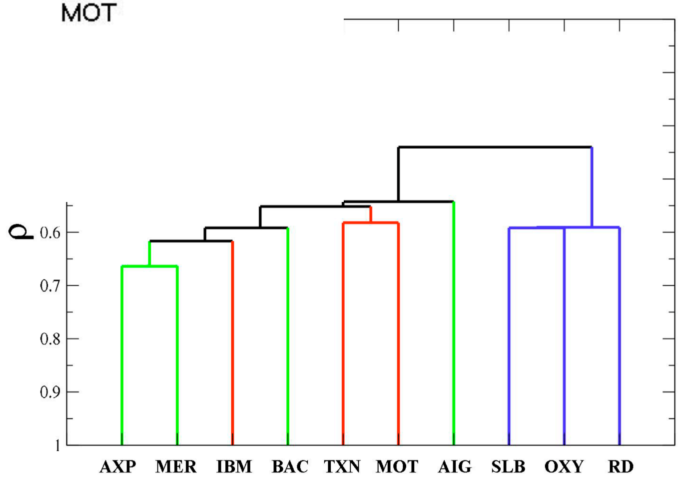
Correlation Matrix (*C*)

Sorted List of Links *(S)*

Correlation based tree(s)



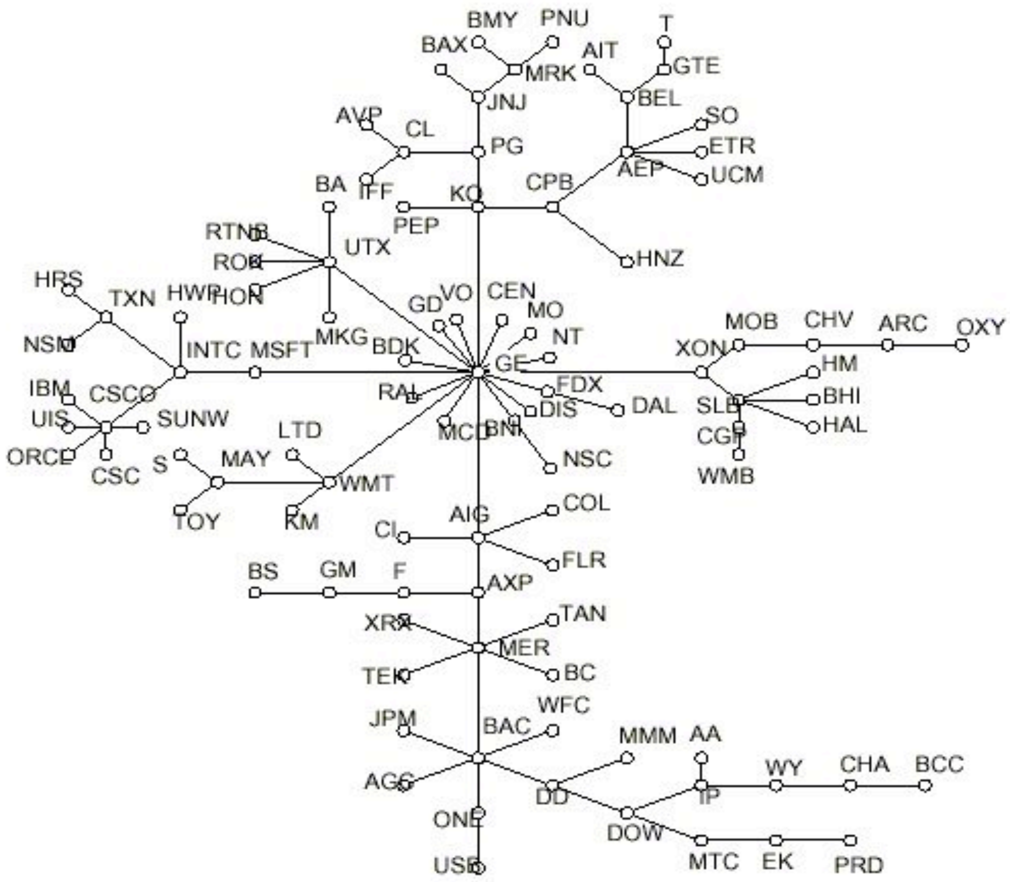
For the single linkage clustering procedure the correlation based tree is the minimum spanning tree



Correlation based trees and hierarchical trees do not carry the same amount of information.

OCS

A typical minimum spanning tree



$N = 100$ (NYSE)

daily returns

1995 - 1998

$T = 1011$

OCS

Minimum spanning tree and Planar maximally filtered graph

Define a **similarity measure between the elements of the system**

Construct the list *S* **by ordering similarities in decreasing order**

Starting from the first element of *S*, **add** the corresponding **link** **if** and only if **the graph is still a Forest or a Tree**

Minimum Spannig Tree
MST

Starting from the first element of *S*, **add** the corresponding **link** **if** and only if **the graph is still Planar (g=0)**

Planar Maximally Filtered Graph
PMFG

R.N.M., Eur. Phys. J. B 11, 193. (1999).

M. Tumminello, T. Di Matteo, T. Aste and R.N.M., PNAS USA 102, 10421 (2005)

OCS

The Planar Maximally Filtered Graph

The Planar Maximally Filtered Graph is

- a topologically planar graph;

- connecting all elements of the graph by keeping the shortest links and allowing at least 3 links for each element;

- topologically embedded in a surface of genus 0;

- a graph allowing loops.

OCS

Graph Genus

The genus of a graph is the minimum number of handles that must be added to the plane to embed the graph without any crossings.

A planar graph therefore has graph genus 0.

The complete graph has genus:

$$g(K_n) = \left[\frac{(n-3)(n-4)}{12} \right]$$

OCS

Number of elements and properties

N = number of vertices (different elements)

M = number of links

PMFG:

- **M = 3 (N-2)** corresponding to complete triangulations on the sphere.
- Graph with a **genus 0** embedding.

MST:

- **M = N-1**.
- absence of loops.

OCS

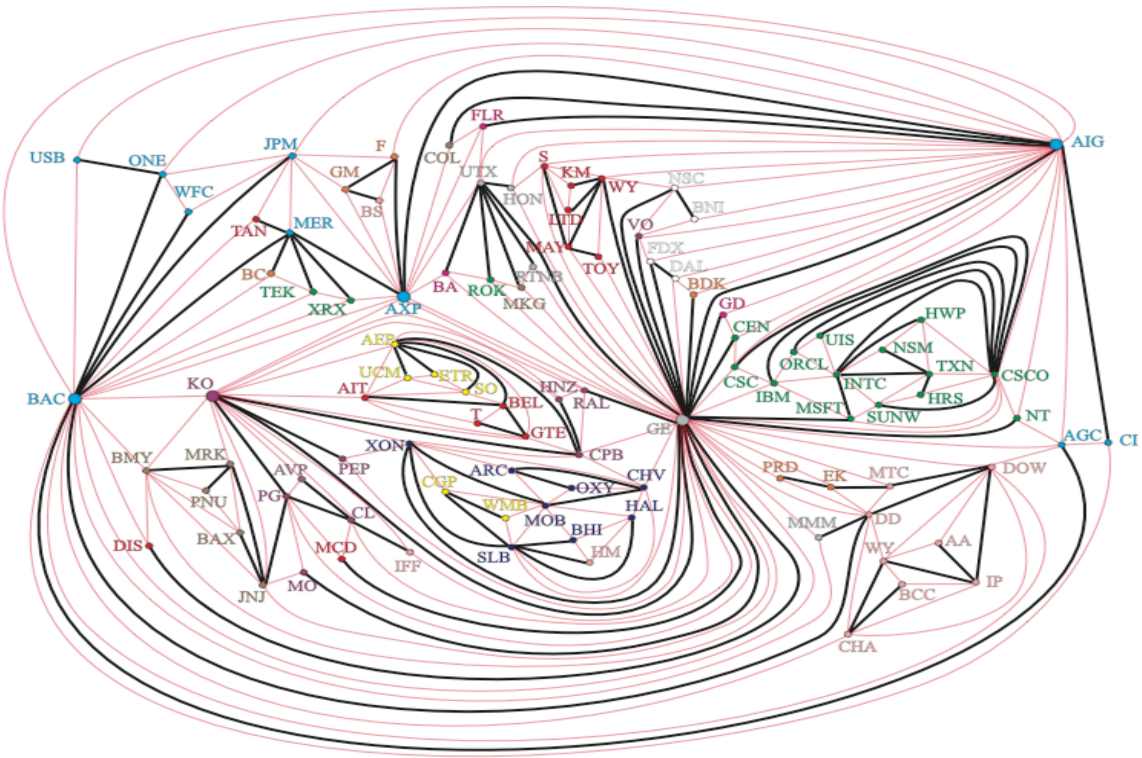
Hierarchical structure

We have proved that the Minimum Spanning Tree is always included into the Planar Maximally Filtered Graph or in any graph embedded in a surface of genus *g* and selected with a constructing algorithm similar to the one used for minimum spanning tree and planar maximally filtered graph.

The hierarchical tree of the graphs obtained with this constructing algorithm are the same as the one of the minimum spanning tree (they are characterized by the same clusters).

OCS

The Planar Maximally Filtered Graph



$N = 100$ (NYSE)

daily returns

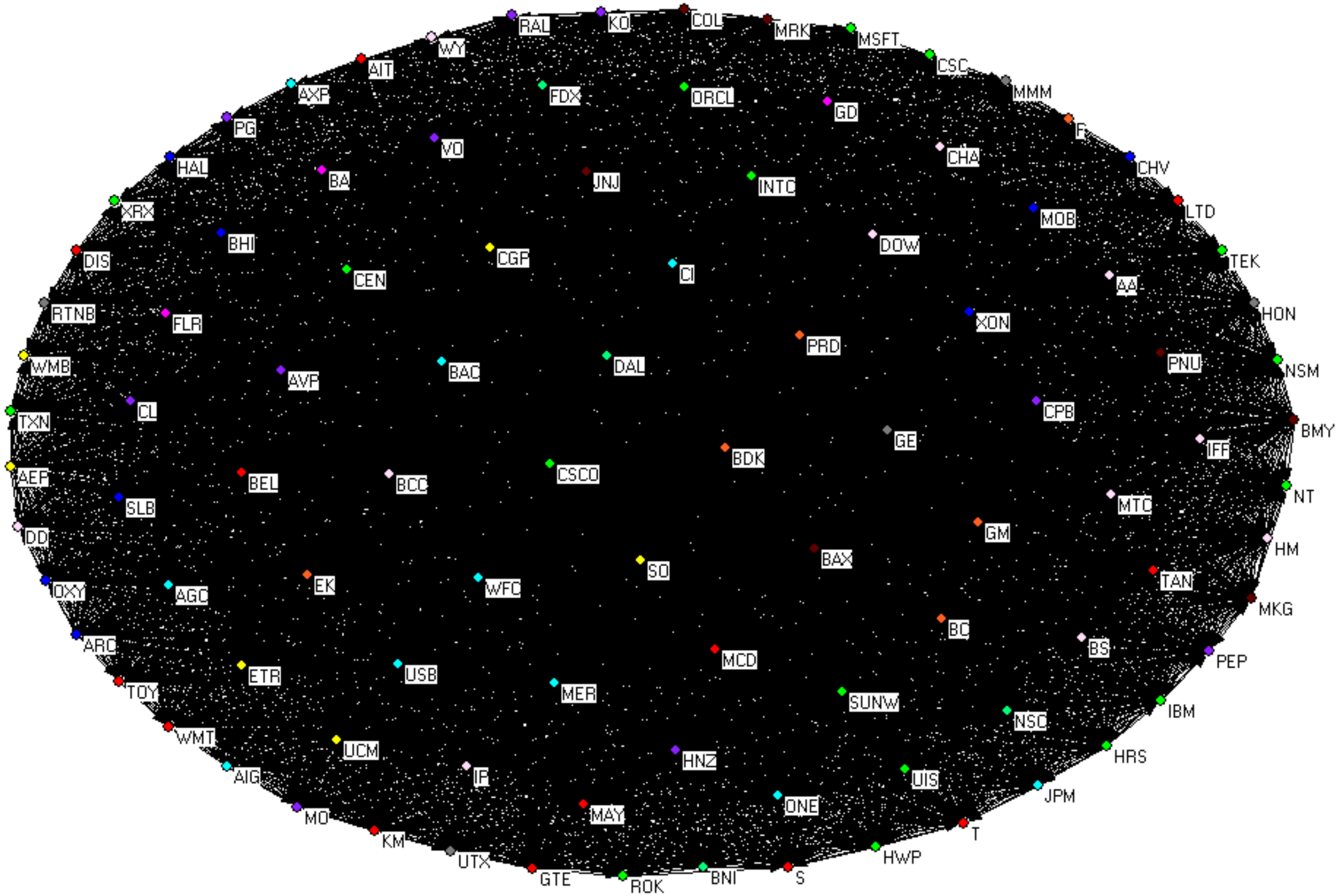
1995 - 1998

$T = 1011$

M. Tumminello, T. Di Matteo, T. Aste and R.N. M., PNAS USA 102, 10421 (2005)

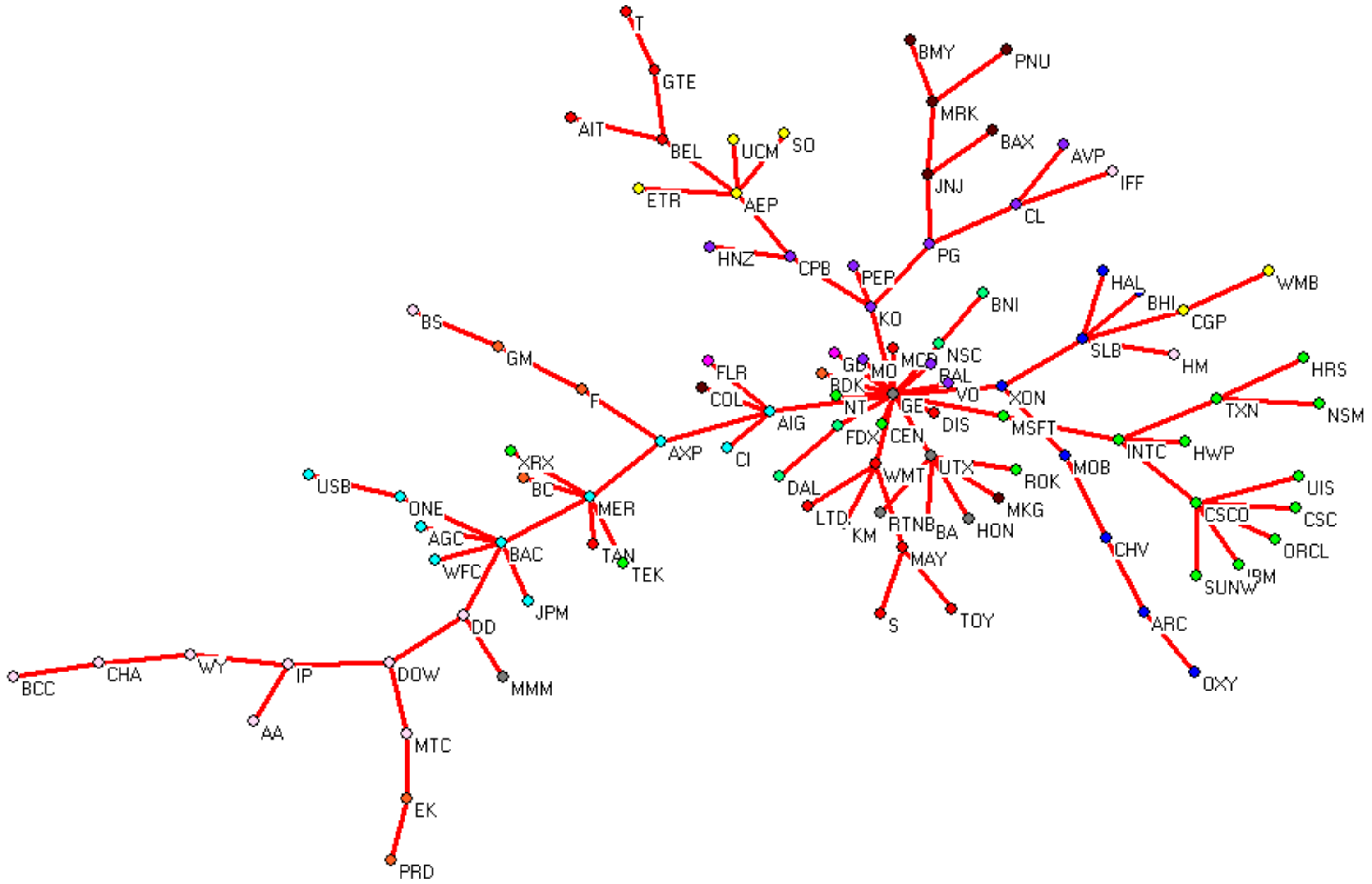
OCS

It is still much less than the complete network!!!!



OC

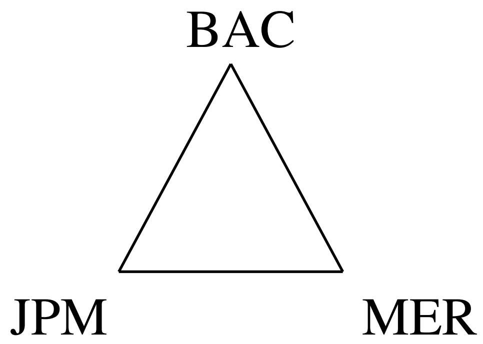
But more than the minimum spanning tree

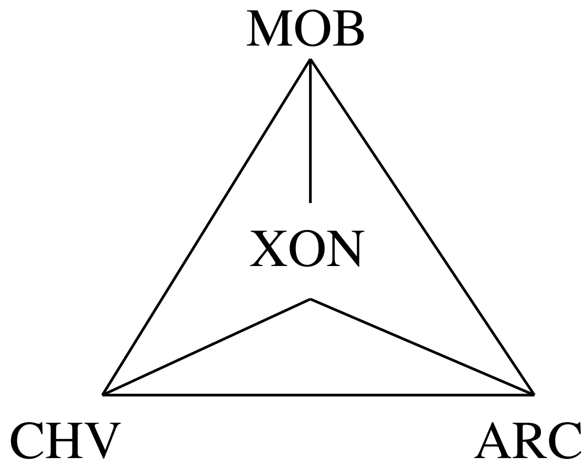


OCS

Loops are present in the PMFG

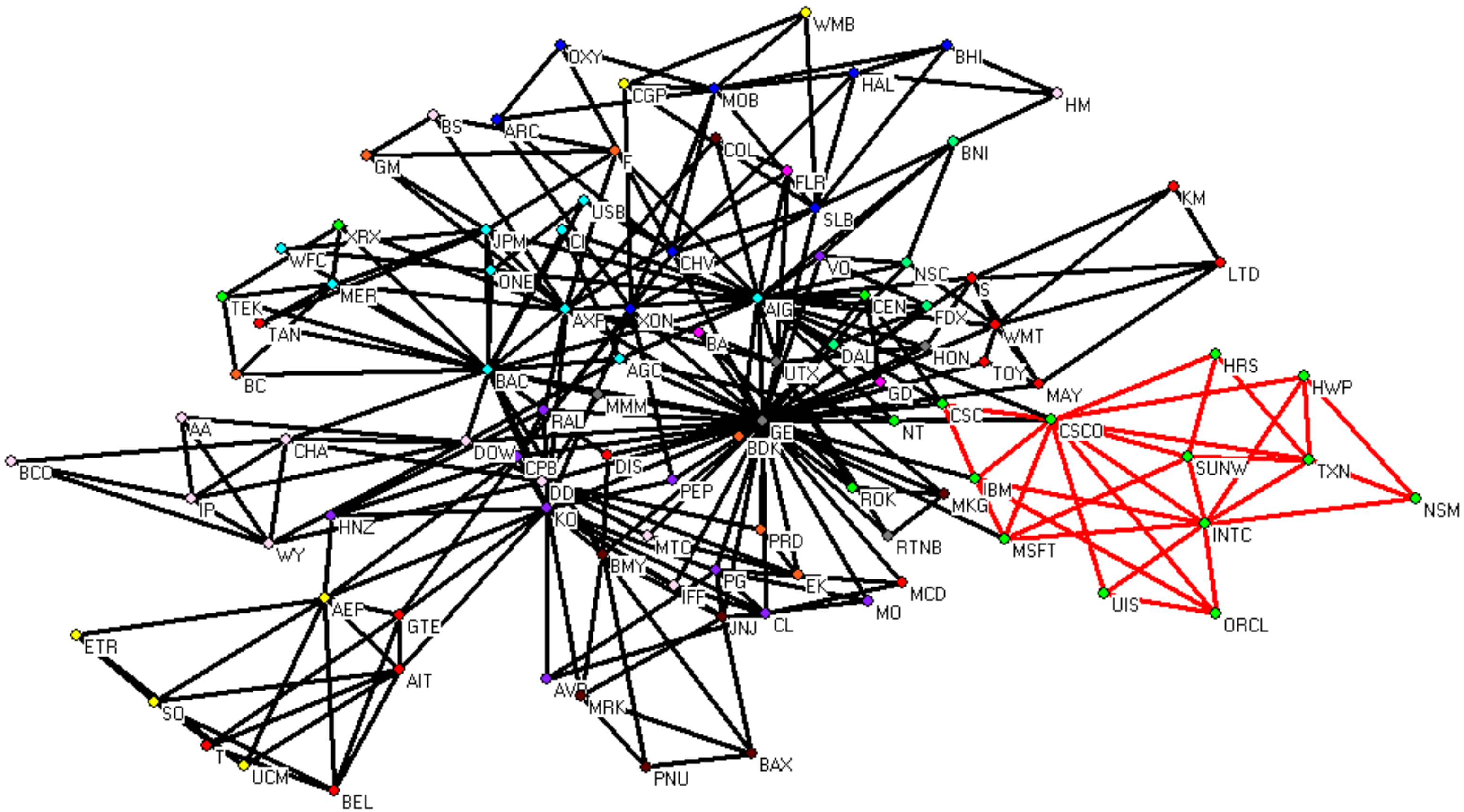
When g=0, the topological constraints allows the observation of **cliques** of 3 and 4 vertices.





OCS

Focusing on the technology cluster



OCS

How to assess the stability of the information filtered out?

OCS

A validation based on bootstrap

Data Set

	e_1	e_2	e_3	…	e_n
t_1	0.113	1.123	-0.002	…	0.198
t_2	1.567	0.789	0.842	…	-0.234
t_3	1.065	-1.962	0.567	…	1.785
t_4	1.112	0.998	-0.424	…	2.735
t_5	-0.211	0.312	-0217	…	0.587
…	…	…	…	…	…
T	0.479	-1.828	-2.041	…	-0.193

Pseudo-replicate Data Set

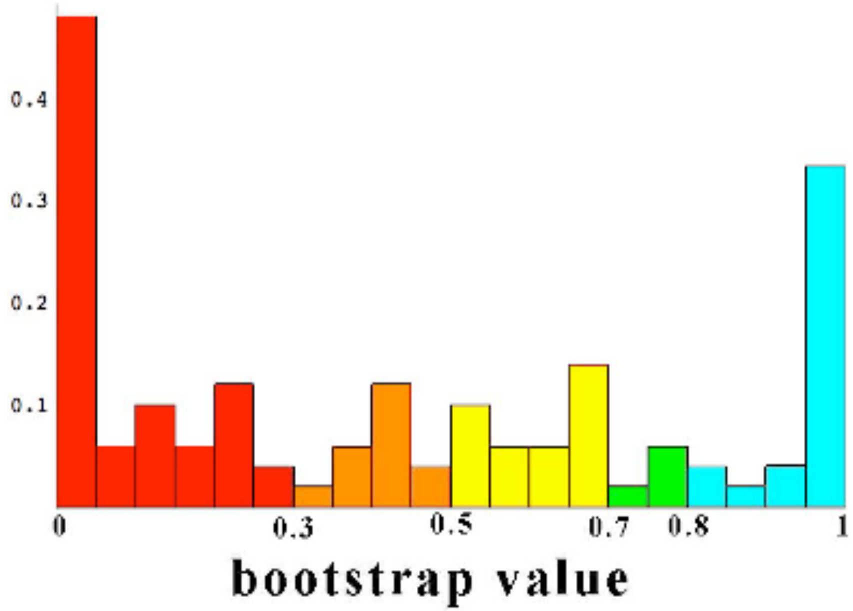
e_1	e_2	e_3	…	e_n
1.567	0.789	0.842	…	-0.234
0.113	1.123	-0.002	…	0.198
1.065	-1.962	0.567	…	1.785
0.113	1.123	-0.002	…	0.198
0.479	-1.828	-2.041	…	-0.193
…	…	…	…	…
0.479	-1.828	-2.041	…	-0.193

M surrogated data matrices are constructed, e.g. M=1000.

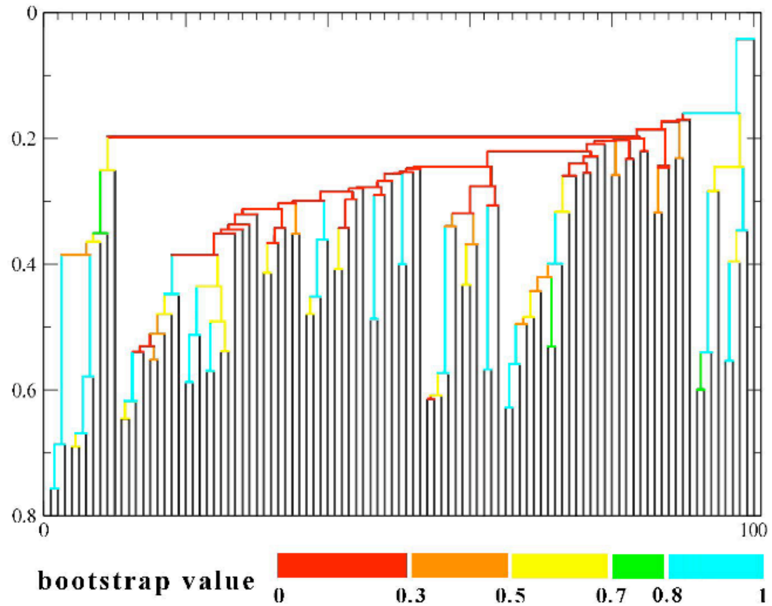
OCS

Bootstrap value of nodes of hierarchical trees

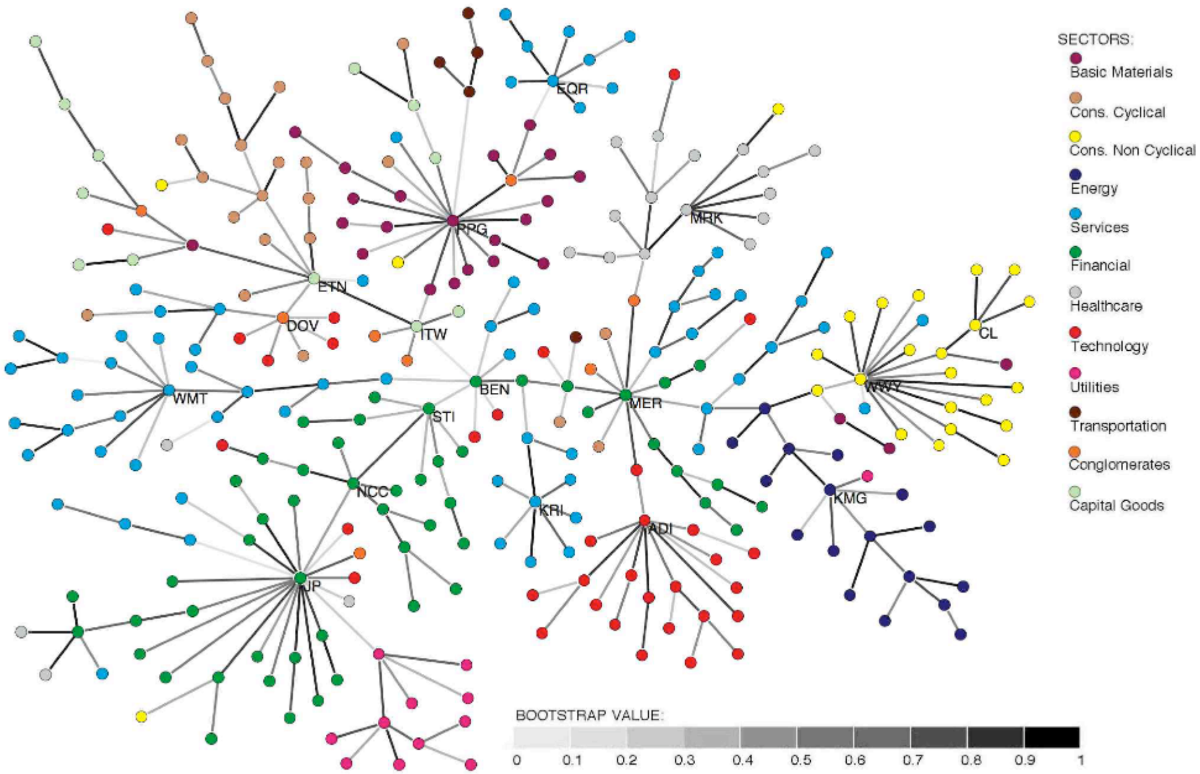
bootstrap value distribution



ALCA



Statistical reliability of the minimum spanning tree



N = 300 (NYSE)

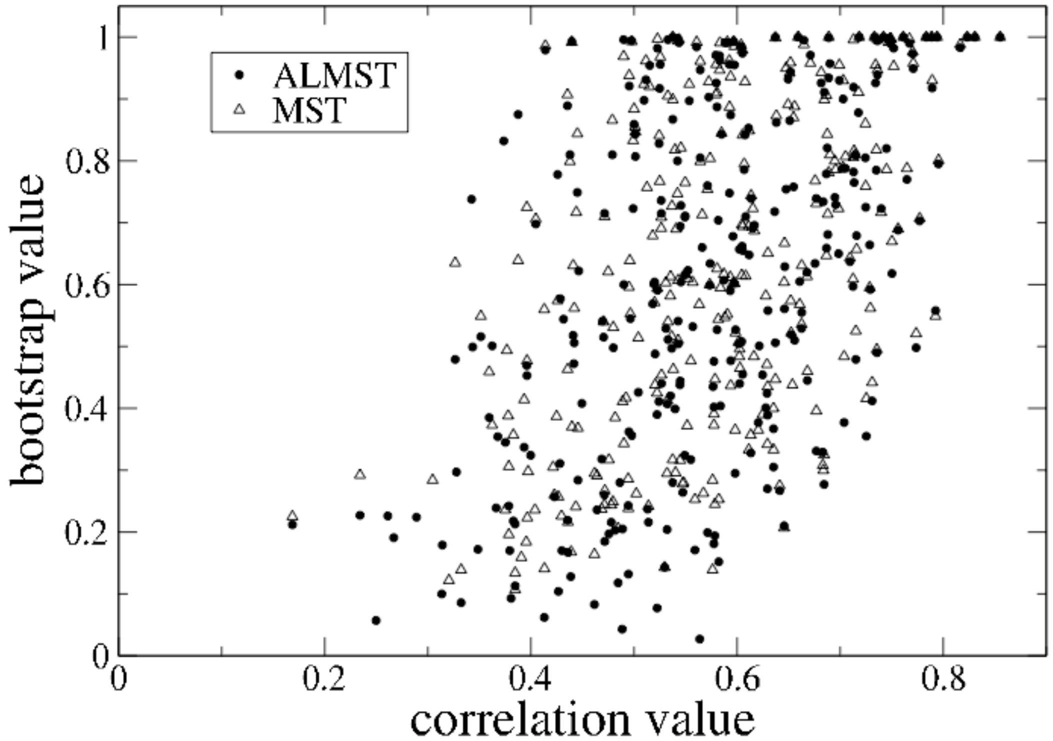
daily returns

2001 - 2003

T = 748

M. Tumminello, C. Coronnello, S. Miccichè, F. Lillo and R.N.M., Int. J. Bifurcation Chaos **17,** 2319-2329 (2007).

Bootstrap vs correlation



$N = 300$ (NYSE)

daily returns

2001 - 2003

$T = 748$

For Gaussian series: $\sigma_\rho = \dfrac{1-\rho^2}{\sqrt{T-3}}$

OCS

The Hierarchically Nested Factor Model (HNFM)

A factor is associated to each node

$$x_i(t) = \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h}^{\;2}}\; \varepsilon_i(t)$$

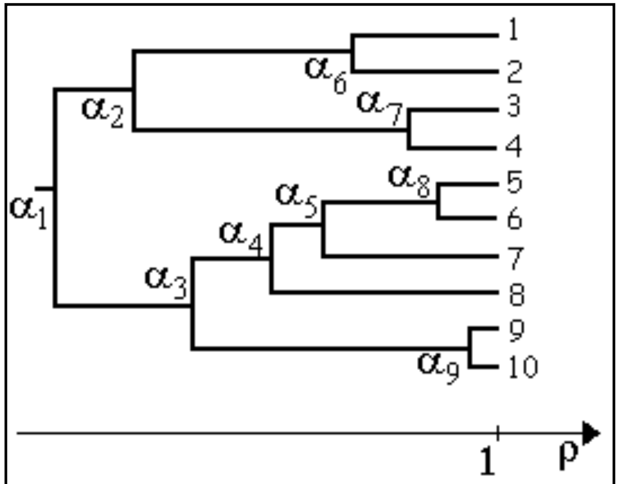
α_h-th factor   Idiosyncratic term

$$\gamma_{\alpha_h} = \sqrt{\rho_{\alpha_h} - \rho_{g(\alpha_h)}}\,;\; \gamma_{\alpha_1} = \sqrt{\rho_{\alpha_1}}$$

The model explains $C^{<} = (\rho_{ij}^{<})$

$$\langle x_i \cdot x_j \rangle = \sum_{\alpha_h \in G(i) \cap G(j)} \gamma_{\alpha_h}^{\;2} = \rho_{\alpha_k} = \rho_{ij}^{<}$$

e.g. $\langle x_1 \cdot x_4 \rangle = \gamma_{\alpha_2}^{\;2} + \gamma_{\alpha_1}^{\;2} = \rho_{\alpha_2} - \rho_{\alpha_1} + \rho_{\alpha_1} = \rho_{\alpha_2}$



$G(i) =$ Pedigree of element i,

e.g. $G(9) = \{\alpha_1, \alpha_3, \alpha_9\}$

$g(\alpha_h) =$ Parent of node α_h,

e.g. $g(\alpha_7) = \alpha_2$

M. Tumminello, F. Lillo, R.N. Mantegna, Hierarchically nested factor model from multivariate data, EPL 78 (3), Art. No. 30006 (2007).

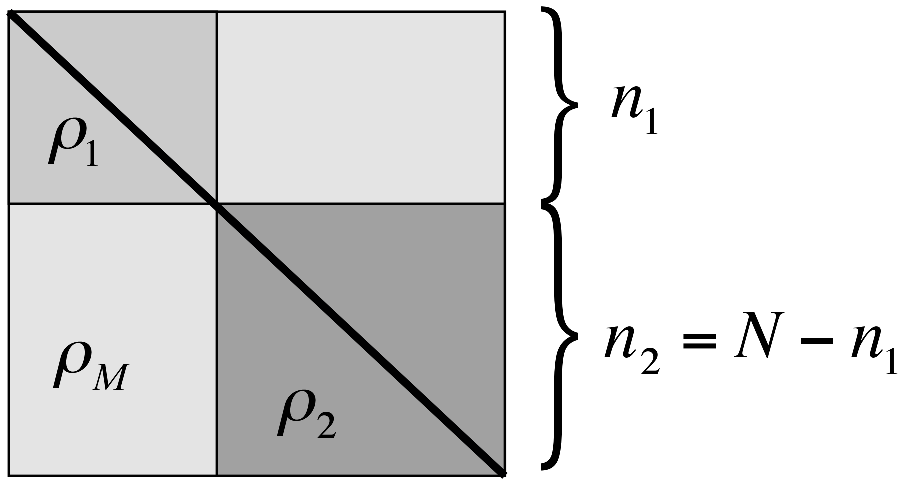
OCS

A simple hierarchically nested model

$$x_i(t) = \gamma_0 f^{(0)}(t) + \gamma_1 f^{(1)}(t) + \sqrt{1-\gamma_0^{\,2}-\gamma_1^{\,2}}\,\varepsilon_i(t) \quad \text{for } i \le n_1$$

$$x_i(t) = \gamma_0 f^{(0)}(t) + \gamma_2 f^{(2)}(t) + \sqrt{1-\gamma_0^{\,2}-\gamma_2^{\,2}}\,\varepsilon_i(t) \quad \text{for } n_1 < i \le N$$

$\boldsymbol{C} =$



$$\rho_1 = \gamma_0^{\,2} + \gamma_1^{\,2}$$

$$\rho_2 = \gamma_0^{\,2} + \gamma_2^{\,2}$$

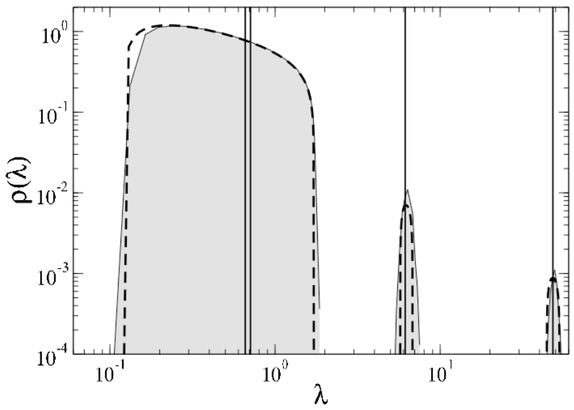
$$\rho_M = \gamma_0^{\,2}$$

Spectral Analysis

OCS

2 large eigenvalues

$$\lambda^{\pm} = \frac{1}{2}\left[2 + q_{+} \pm (q_{-}^{2} + 4\, n_1\, n_2\, \rho_M^{2})^{1/2}\right]$$



2 corresponding eigenvectors

$$\vec{z}^{\pm} = (u^{\pm} \ldots u^{\pm}, v^{\pm}, \ldots v^{\pm})$$

where $u^{\pm} = 1/\sqrt{2 n_1\left[1 + y^2 \mp y\sqrt{1+y^2}\right]}$, $v^{\pm} = \pm 1/\sqrt{2 n_2\left[1 + y^2 \pm y\sqrt{1+y^2}\right]}$,

$q_{\pm} = (n_1 - 1)\rho_1 \pm (n_2 - 1)\rho_2$ and $y = q_{-}/(4\, n_1\, n_2\, \rho_M^{2})^{1/2}$

Principal Component Analysis is not able to reconstruct the true model and/or to give insights about its hierarchical features

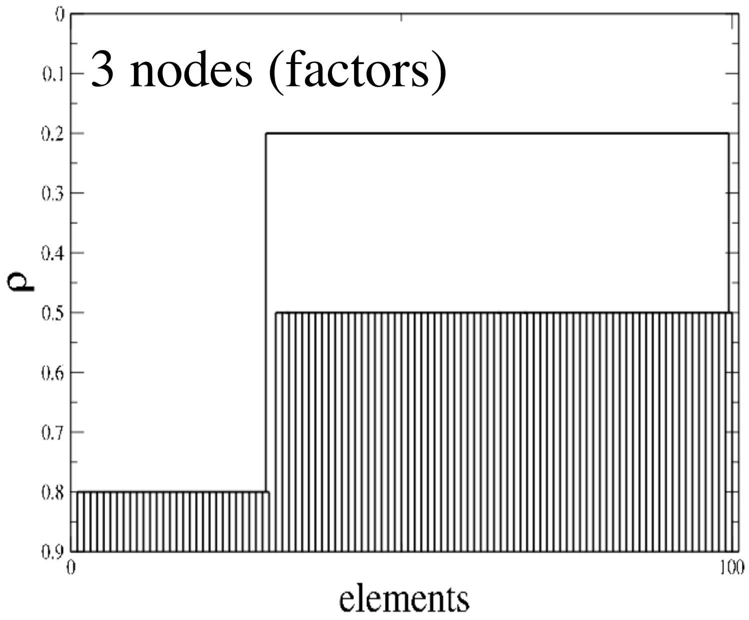
OCS

The HNFM allows to simulate the system. We use hierarchical clustering to investigate the simulations so that we can estimate the ability of hierarchical clustering to detect a hierarchically nested system.

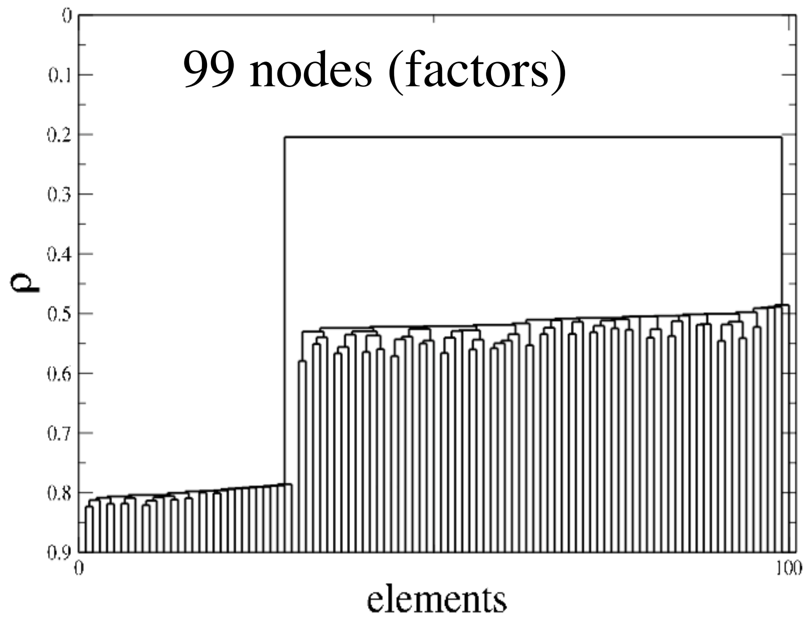
A problem of the HC method: HNFM by hierarchical clustering always detects n-1 factor

A solution: Evaluation of node statistical uncertainty and node reduction

Hierarchical tree of the model



Hierarchical tree reconstruction



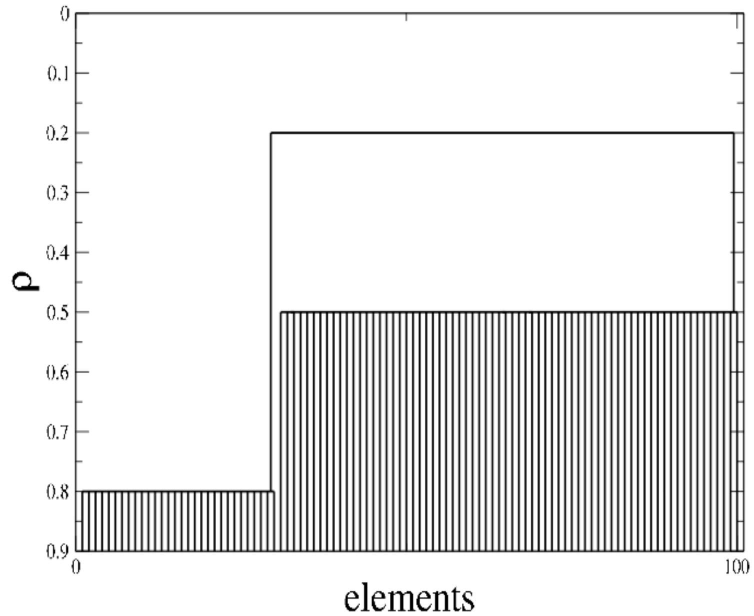
Self-consistent node-factor reduction

OCS

- Select a bootstrap value threshold *bt* .

- For each node α_k :

  If $b(\alpha_k) < bt$ then merge the node α_k with his first ancestor α_q (in the path to the root) such that $b(\alpha_q) \geq bt$.



- How to chose *bt* ?

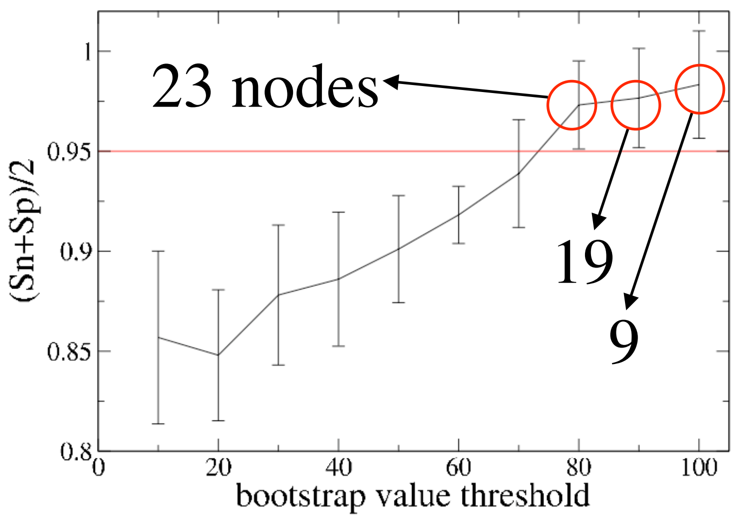
In a self-consistent way!

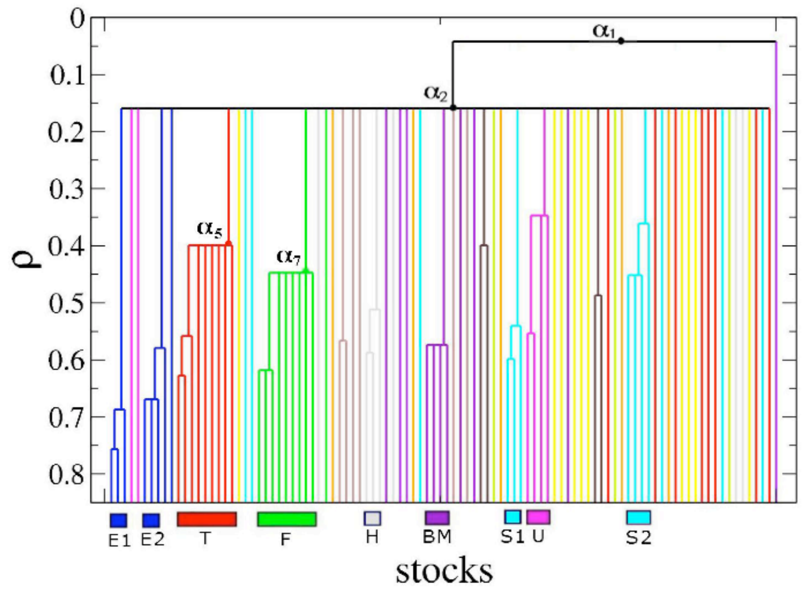
HNFM correctly detects the model when *bt*>0.70

OCS

Node reduction for an empirical system

Daily return of 100 stocks traded at NYSE in the time period 1/1995-12/1998 (*T*=1011)





Sn = sensitivity; Sp = specificity

$$S_n = \frac{TP}{TP + FN} \qquad S_p = \frac{TN}{TN + FP}$$

OCS

Interpretation of factors

HNFM associated to the reduced dendrogram with 23 nodes. Equations for stocks belonging to the Technology and Financial Sectors.

Financial Factor

$$x_i^F(t) = \gamma_{\alpha_7} f^{(\alpha_7)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_7}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_i(t),$$

$$x_j^T(t) = \gamma_{\alpha_5} f^{(\alpha_5)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_5}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_j(t).$$

Technology Factor

OCS

$C^<$ is a correlation matrix

$$\boxed{\begin{array}{l} C^< = (\rho_{ij}^<) \\ \rho_{ij}^< = \rho_{\alpha_k} \end{array}}$$

where α_k is the first node where elements i and j merge together.

$$\boxed{\text{If } \rho_{ij}^< \geq 0 \ \forall \ i,j \text{ then } C^< \text{ is positive definite.}}$$

Indeed $C^<$ is the correlation matrix of a suitable factor model named Hierarchically Nested Factor Model.

M. Tumminello, F. Lillo and R.N.M., EPL 78, 30006 (2007).

OCS

Filtered correlation matrices

We consider **two filtered correlation matrices**, ,
obtained by applying the **A**verage **L**inkage **C**luster **A**nalysis and the **S**ingle **L**inkage **C**luster **A**nalysis to the empirical correlation matrix respectively. $\mathbf{C}^{<}_{\mathbf{ALCA}}$ and $\mathbf{C}^{<}_{\mathbf{SLCA}}$

For comparison we also consider filtered correlation matrices obtained with Random Matrix Theory (RMT) and shrinkage technique.

The filtered matrix obtained with the shrinkage technique is defined as

$$\mathbf{C}^{SHR}(\alpha)= \alpha\ \mathbf{T} + (1-\alpha)\ \mathbf{C}$$

OCS

How to quantify the amount of information filtered from the correlation matrix?

How to quantify the stability of the filtered information?

OCS

Kullback-Leibler distance

$$K(p,q) = E_p\left[\log\left(\frac{p}{q}\right)\right]$$, where p and q are pdf's.

For multivariate Gaussian distributed random variables we have[1]:

$$K(P(\Sigma_1, X), P(\Sigma_2, X)) = \frac{1}{2}\left[\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) - n\right] = K(\Sigma_1, \Sigma_2)$$

Minimizing the Kullback-Leibler distance is equivalent to maximize the likelihood in the maximum likelihood factor analysis.

[1]M. Tumminello, F. Lillo and R.N.M., PRE 76, 031123 (2007).

OCS

Expectation values

$$E\left[K(\Sigma, S_1)\right] = \frac{1}{2}\left\{ n\log\left(\frac{2}{T}\right) + \sum_{p=T-n+1}^{T}\left[\frac{\Gamma^I(p/2)}{\Gamma(p/2)}\right] + \frac{n(n+1)}{T-n-1}\right\}$$

$$E\left[K(S_1, \Sigma)\right] = \frac{1}{2}\left\{ n\log\left(\frac{T}{2}\right) - \sum_{p=T-n+1}^{T}\left[\frac{\Gamma^I(p/2)}{\Gamma(p/2)}\right]\right\}$$

$$E\left[K(S_1, S_2)\right] = \frac{1}{2}\frac{n(n+1)}{T-n-1}$$

where Σ is the true correlation matrix of the system while S_1 and S_2 are sample matrices of Σ from two independent realizations of length T.

The three expectation values are independent from Σ,
i.e they do not depend from the underlying model

OCS

Kullback vs Frobenius

- The expectation values of Frobenius distance are model dependent, e.g. for a system of *n=2* Gaussian random variables with correlation coefficient ρ it is

$$\mathrm{E}\left[F(\Sigma,\mathbf{S})\right] = \mathrm{E}\left[\sqrt{\mathrm{tr}\left[(\Sigma-\mathbf{S})(\Sigma-\mathbf{S})^{\mathrm{T}}\right]}\right] = \frac{2}{\sqrt{\pi T}}\left(1-\rho^2\right)$$

where Σ is the model correlation matrix of the system while $\mathbf{S}$ is a sample correlation matrix obtained from a realization of length T.

Kullback-Leibler distance

OCS

The Kullback-Leibler distance can also be analytically calculated random variables following a **multivariate Student's t-distribution**[1]:

$$P_{\Sigma}(\mathbf{x},\mu)=\frac{\Gamma\left(\frac{n+\mu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)\sqrt{(\mu\,\pi)^{n}\,|\Sigma|}}\,\frac{1}{\left[1+\frac{1}{\mu}\,\tilde{\mathbf{x}}\,\Sigma^{-1}\mathbf{x}\right]^{\frac{n+\mu}{2}}}$$

If $\dfrac{\mu}{n} << 1$ then :

$$K(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)=\frac{1}{2}\left[\log\left(\frac{|\mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|}\right)+n\,\log\left(\frac{\mathrm{tr}\left(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\right)}{n}\right)\right]$$

[1]G. Biroli, J.-P. Bouchaud, M. Potters, Acta Phys. Pol. B 38, 4009 (2007),

OCS

Gaussian vs Student

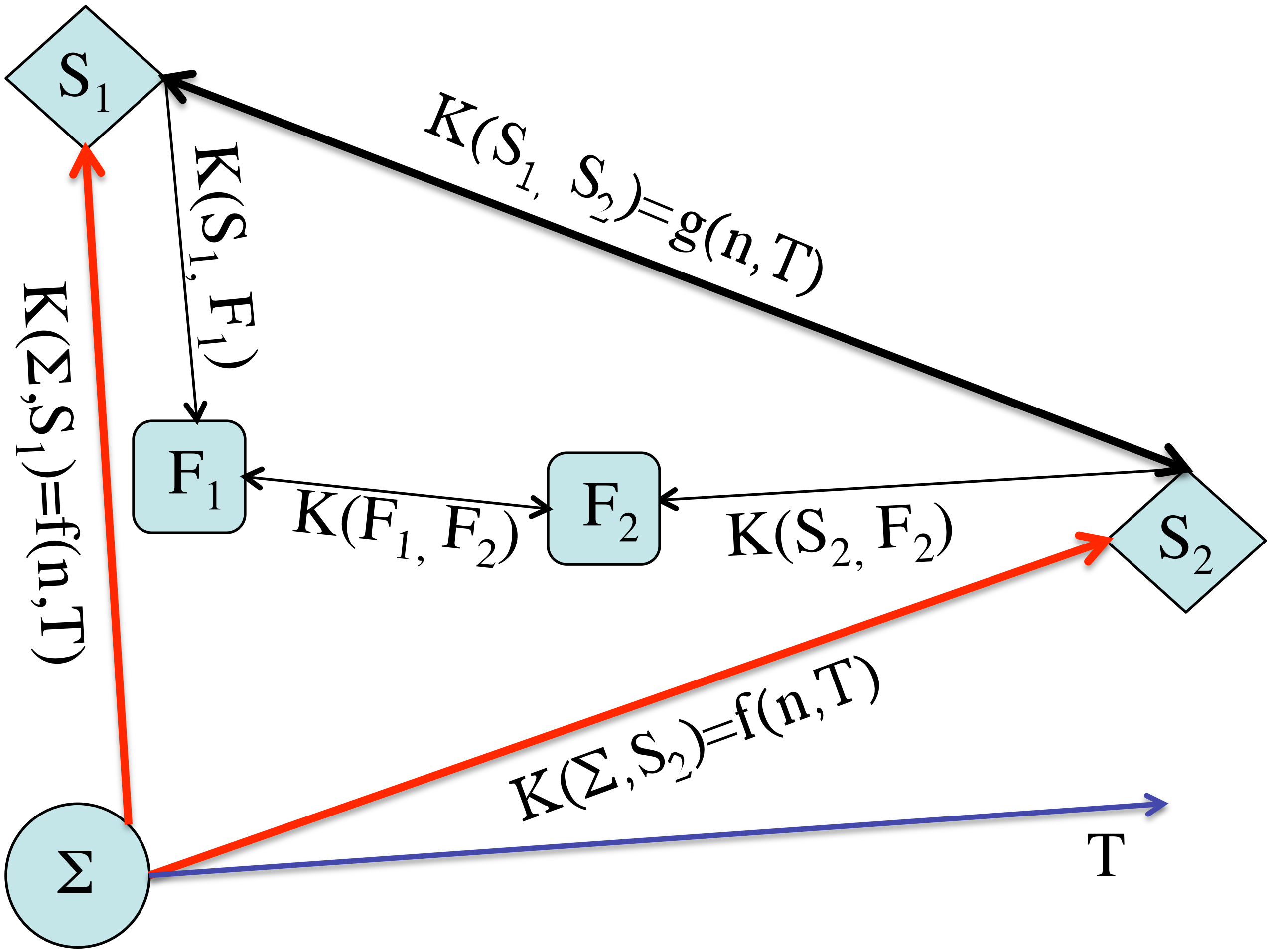
$$K_G(\Sigma_1,\Sigma_2) = \frac{1}{2}\left[\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) - n\right]$$

$$K_S(\Sigma_1,\Sigma_2) = \frac{1}{2}\left\{\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + n\log\left[\frac{\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right)}{n}\right]\right\}$$

$$\text{If } \Sigma_1 \cong \Sigma_2 \Rightarrow K_G(\Sigma_1,\Sigma_2) \cong K_S(\Sigma_1,\Sigma_2)$$

OCS

Comparison of filtering procedures



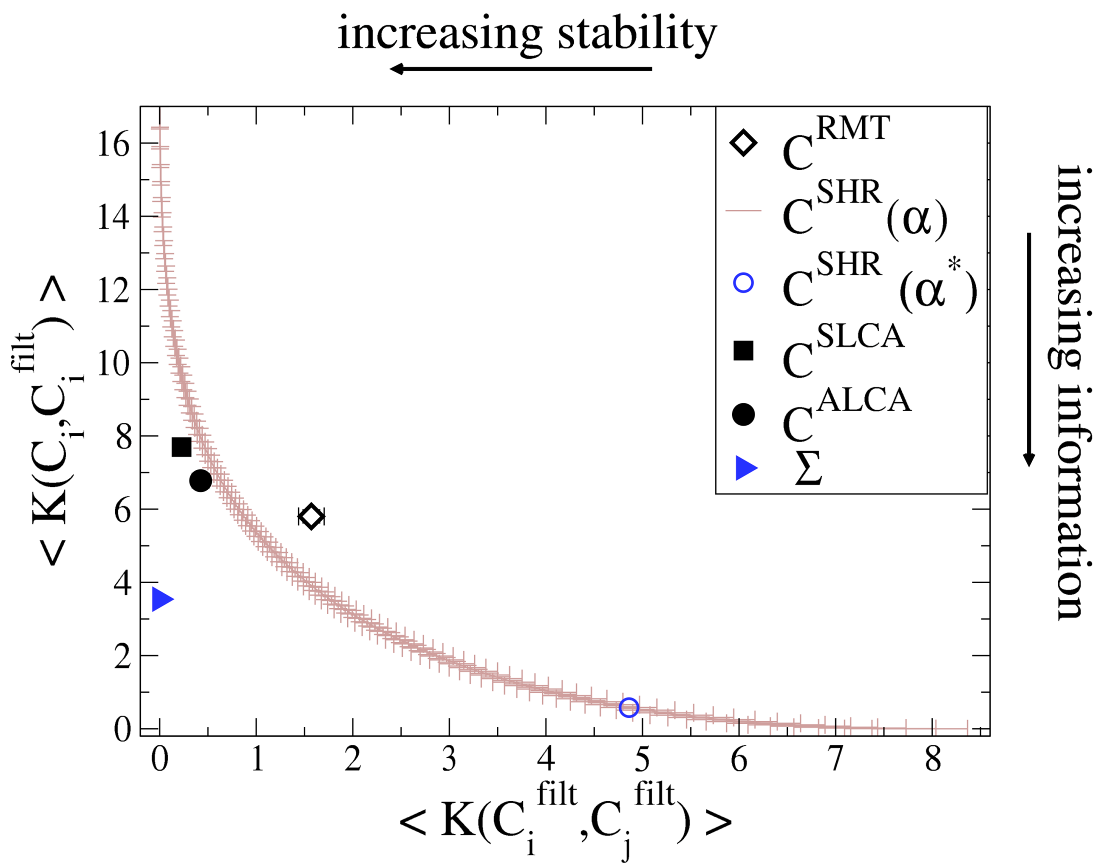
S_1 and **S_2** are sample correlation matrices estimated from independent realizations/bootstrap-replicas of the system.

F_1 and **F_2** are matrices filtered from S_1 and S_2 respectively.

Σ is the true correlation matrix of the system.

Comparison of filtered correlation matrices (block model)

OCS



Block diagonal model with 12 factors.

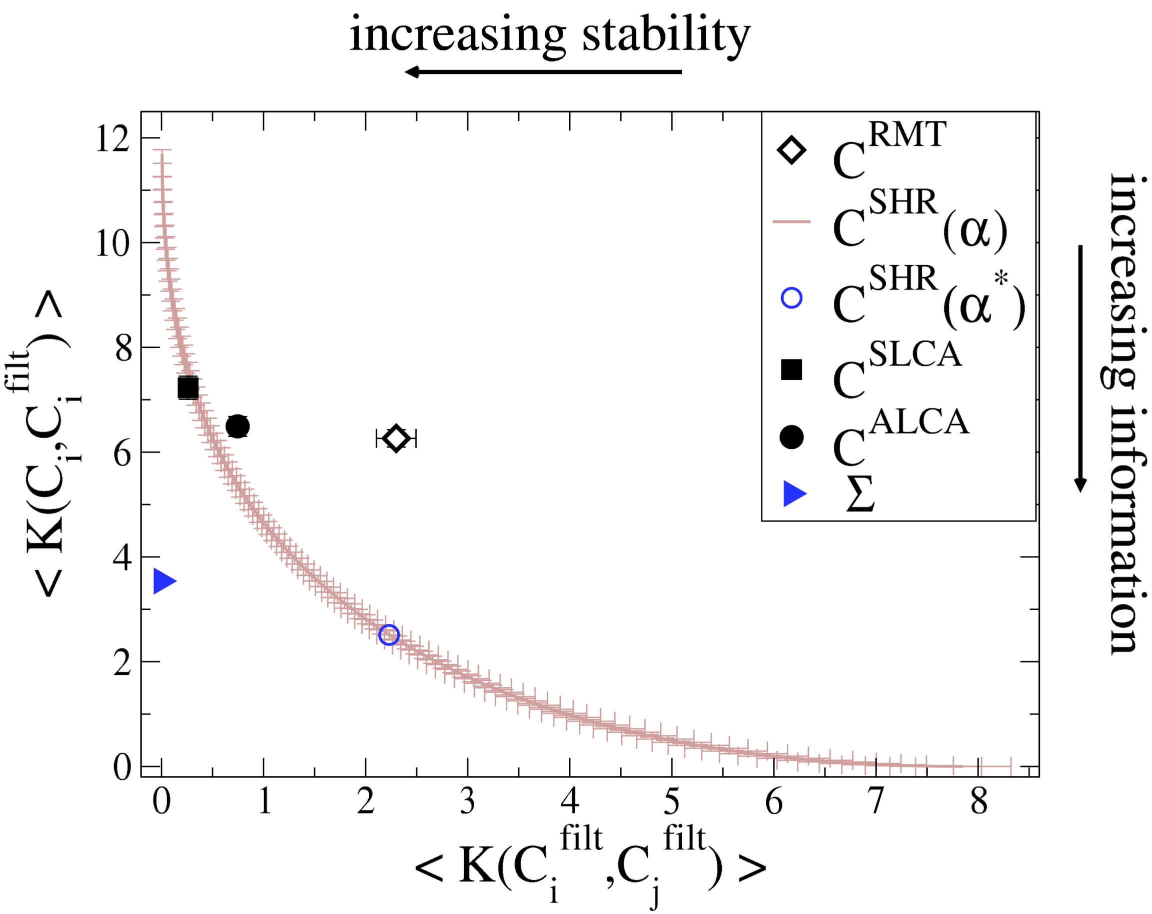
N=100,
T=748.

Gaussian random Variables.

M. Tumminello, F. Lillo and R.N.M., Acta Physica Polonica B 38, 4009-4026 (2007).

Comparison of filtered correlation matrices (HNFM model)

OCS



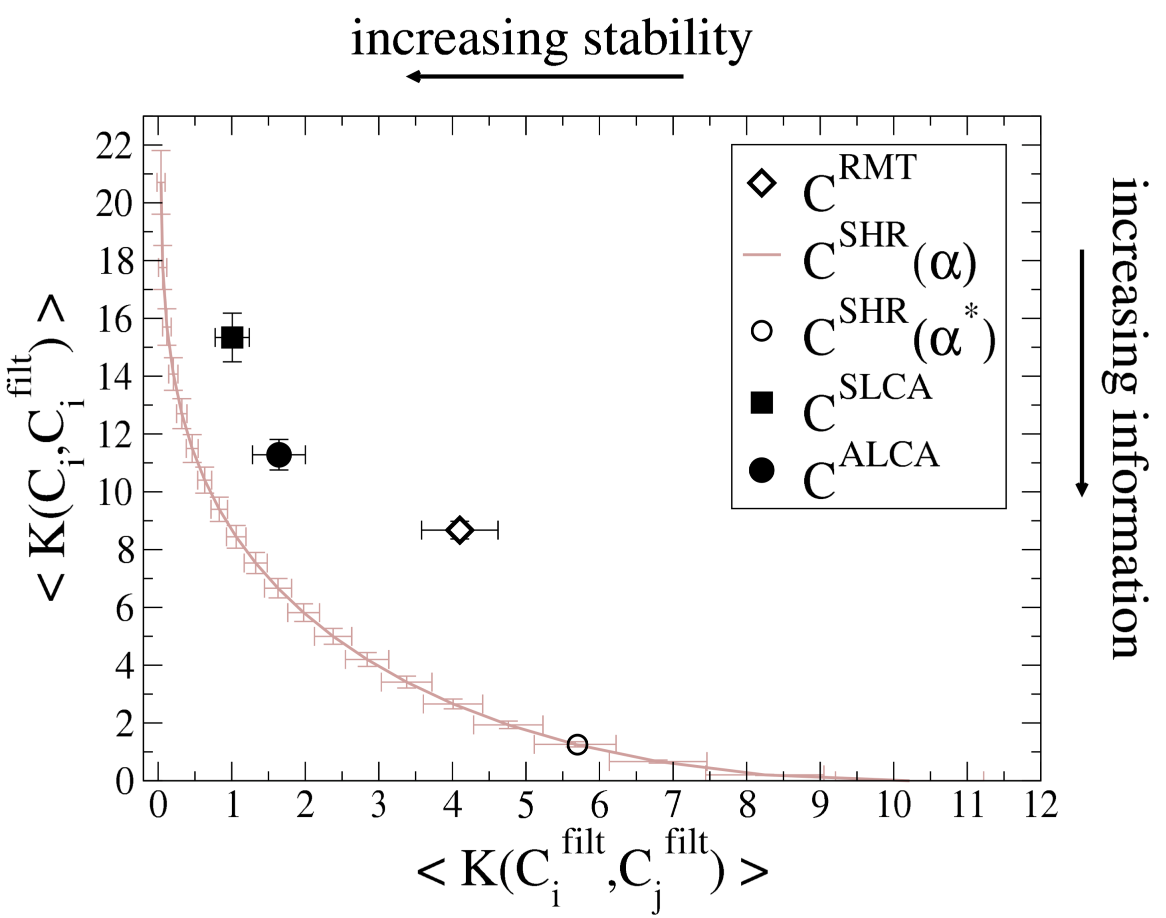
HNFM
with 23 factors.

N=100,
T=748.

Gaussian random variables.

OCS

Comparison of filtered correlation matrices (empirical data)



$N = 300$ (NYSE)

daily returns

2001 - 2003

$T = 748$

Another empirical system[1,2]

Inventory variation of market members trading an asset at the Spanish Stock Market

[1] Vaglica G, Lillo F, Moro E, R.N.M., PHYSICAL REVIEW E 77, 036110 (2008)

[2] Lillo F, Moro E, Vaglica G, R.N.M., NEW JOURNAL OF PHYSICS 10, 043019 (2008)

OCS

Investigated variable

- **Inventory variation** = the value (i.e. price times volume) of an asset exchanged as a buyer minus the value exchanged as a seller in a given time interval.

$$v_i(t) \equiv \sum_{s=t}^{t+\tau} \varepsilon_i(s) p_i(s) V_i(s)$$

sign
+1 for buys
-1 for sells

price

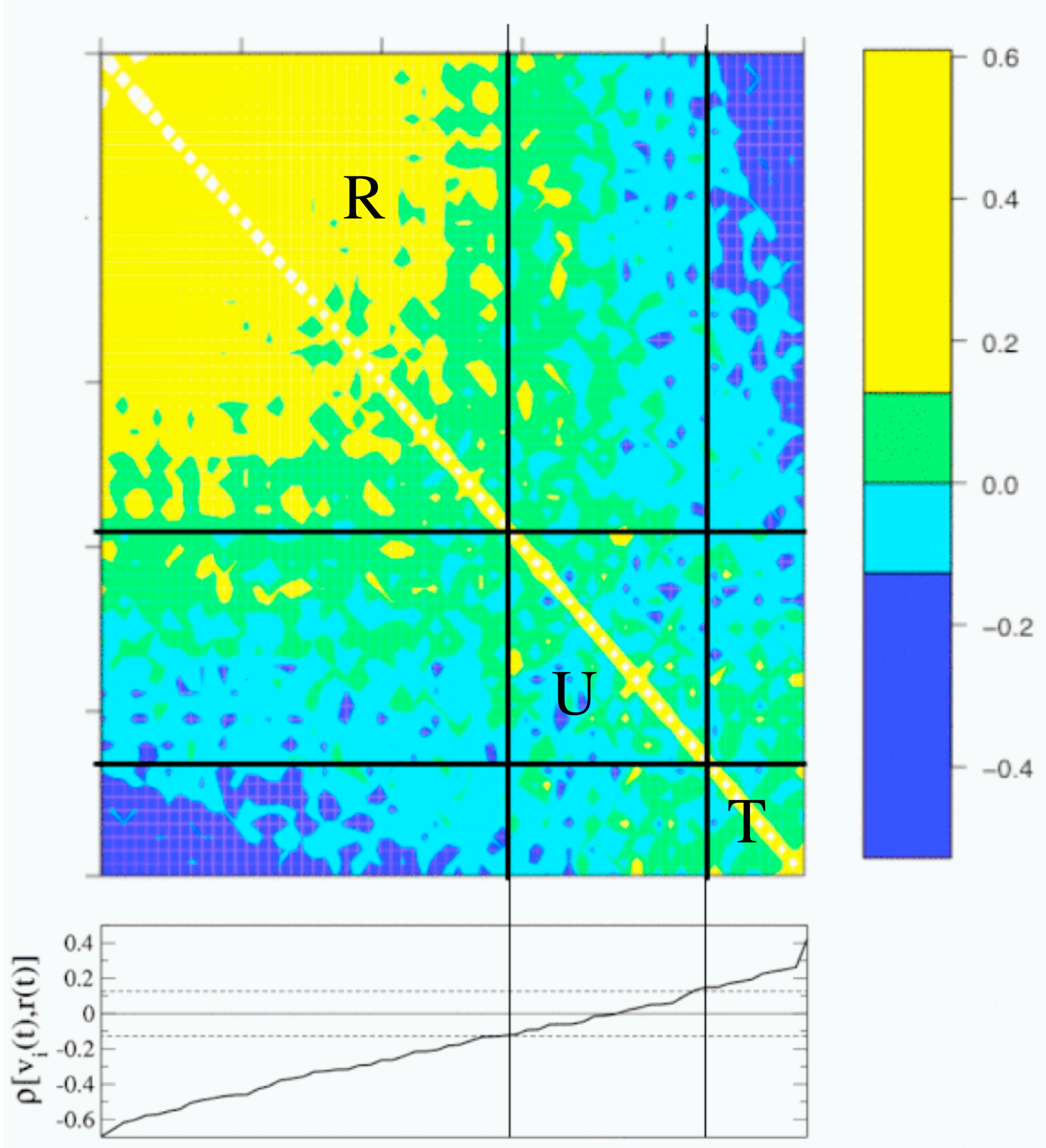
volume

In this talk, we investigate the $\tau = 1$ trading day

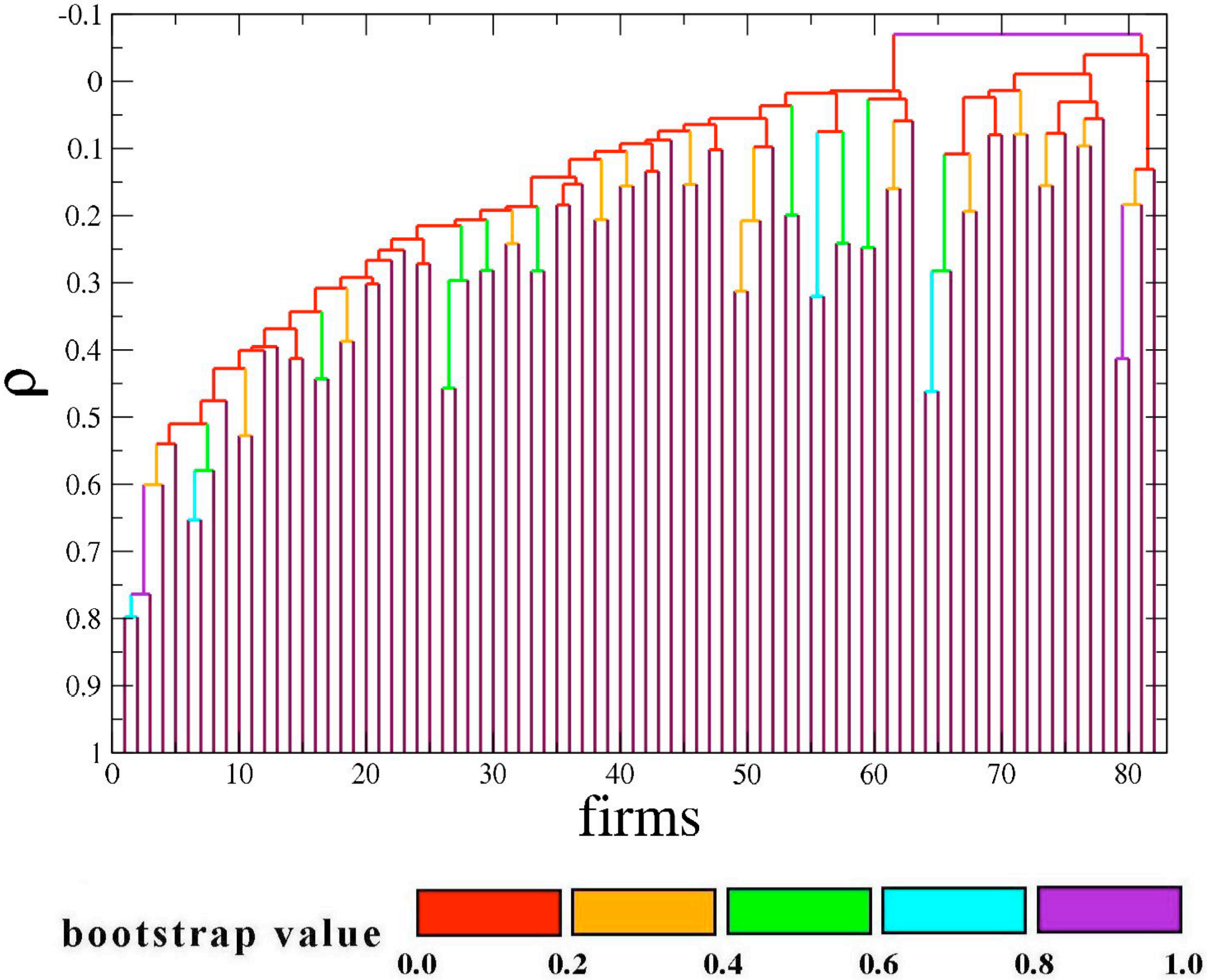
OCS

BBVA 2003

Inventory variation correlation matrix obtained by sorting the market members in the rows and columns according to their correlation of inventory variation with price return

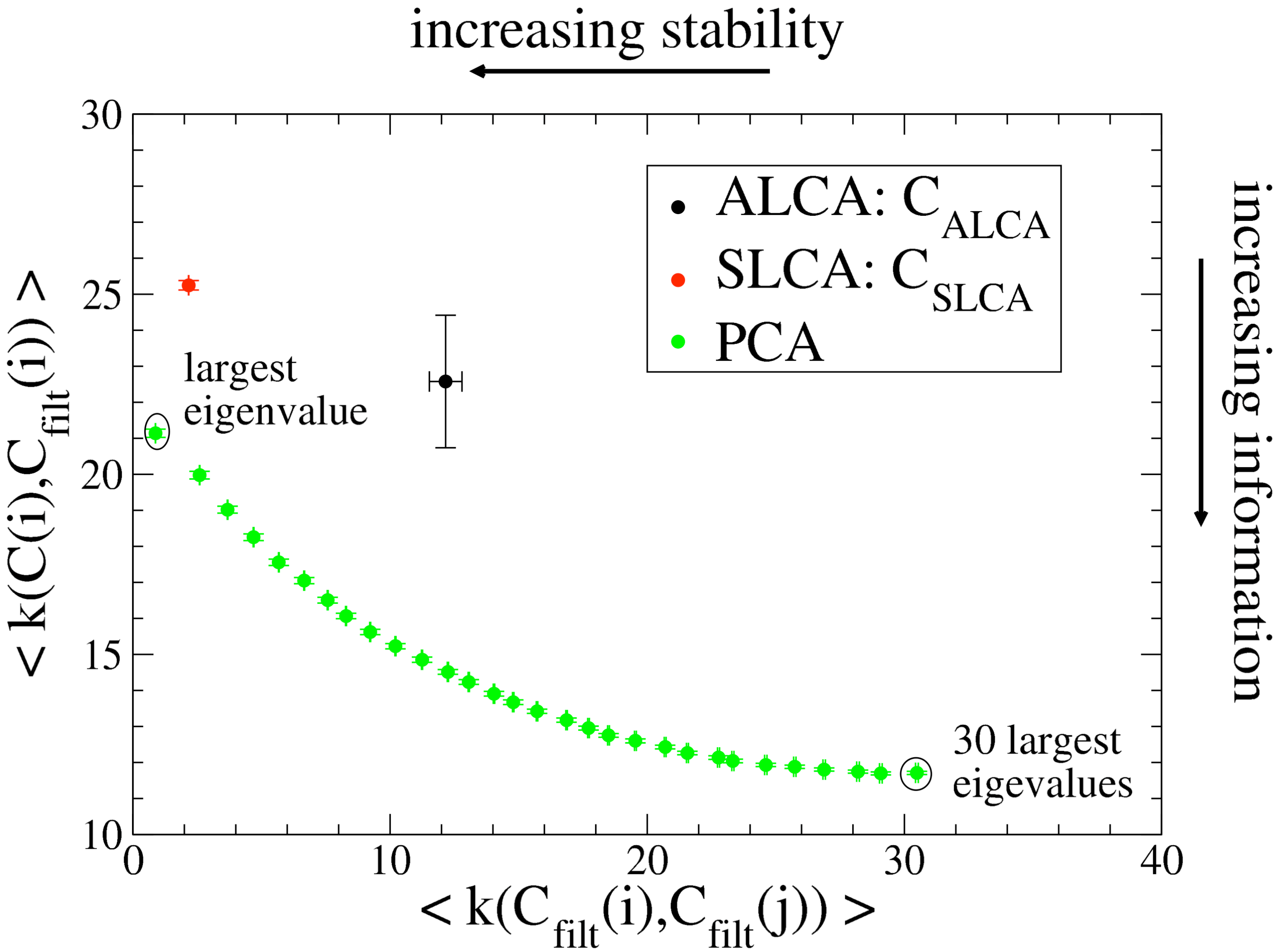


The hierarchical tree



OCS

The best filtering procedure we find is the one from principal component analysis



OCS

Empirical data are compatible with a one-factor model of inventory variation dynamics

The empirical findings on the daily data suggest the following agent (market member) based model

$$v_i(t) = \gamma_i r(t) + \epsilon_i(t)$$

price return

idiosyncratic noise

$\gamma_i > 0$ trending market members (ex: momentum strategies);

$\gamma_i < 0$ reversing market members (ex: contrarians' strategies);

$\gamma_i \approx 0$ uncategorized market members.

see also, Lillo and R.N.M., Phys. Rev. E **72**, 016219 (2005)

Conclusions

OCS

We describe the structure of an empirical correlation matrix by using hierarchical trees and correlation based networks.

We estimate the statistical reliability of links in hierarchical trees and correlation based networks by using a bootstrap based approach.

We show how to model hierarchies detected by hierarchical clustering in terms of a factor model, i.e. the hierarchically nested factor model.

We use the Kullback-Leibler distance in order to compare different techniques used to filter the most stable information of correlation matrices.

OCS

Thank you!

OCS website: http://ocs.unipa.it