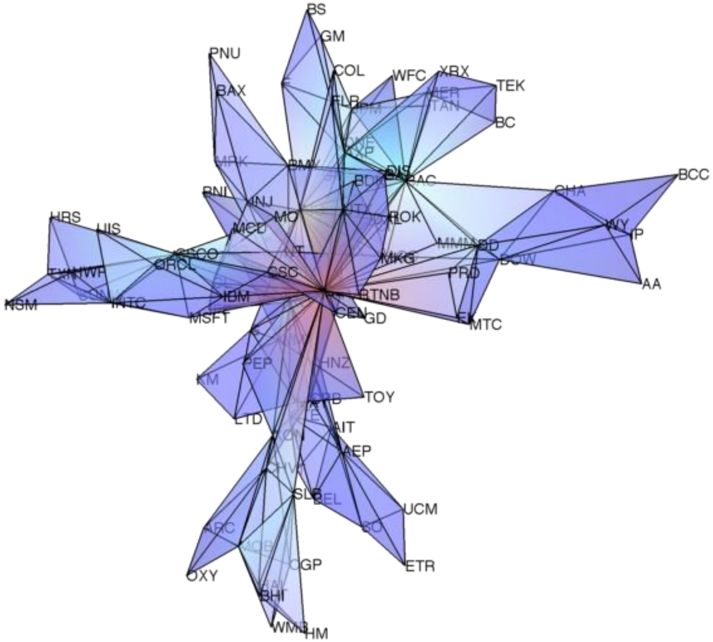
Market structure discovery with Clique Forests



Tomaso Aste

Financial Computing and Analytics Group


Tomaso Aste


Philip Treleaven


Denise Gorse

UCL FINANCIAL COMPUTING

- **Centre for Doctoral Training in Financial Computing & Analytics**
- **Financial Risk Management MSc**
- **Computational Finance MSc**

- **Centre for Bolckchain Technologies**

http://fincomp.cs.ucl.ac.uk/

UCL CBT


Guido Germano

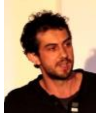

Fabio Caccioli


Christopher Clack


Giacomo Livan


Paolo Barucca


Simone Righi


Daniel Hulme


Robert E Smith


Paolo Tasca


Carolyn Phelan


Geoff Goodell


Nikhil Vadgama


Jessica James


Nick Firoozye


Ariane Chapelle

EPSRC
Pioneering research and skills

Research Councils UK
Digital Economy
Transforming Business and Society

European Commission

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Begin of "coding session"

Download MFCF from: http://www0.cs.ucl.ac.uk/staff/T.Aste/software/

- Open with R studio: "experiments_UCL_FINTECH_(1,2,3).R"
- Select all
- Press RUN

End of "coding session"

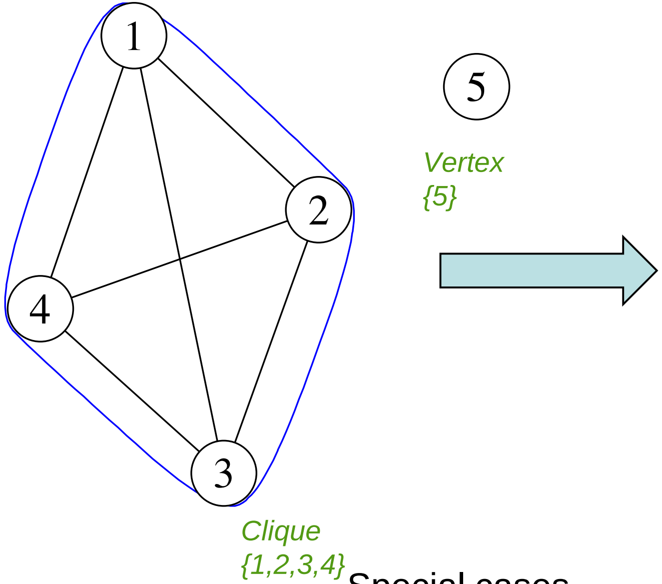
(Software available in a Python wrapper and Matlab available as well)

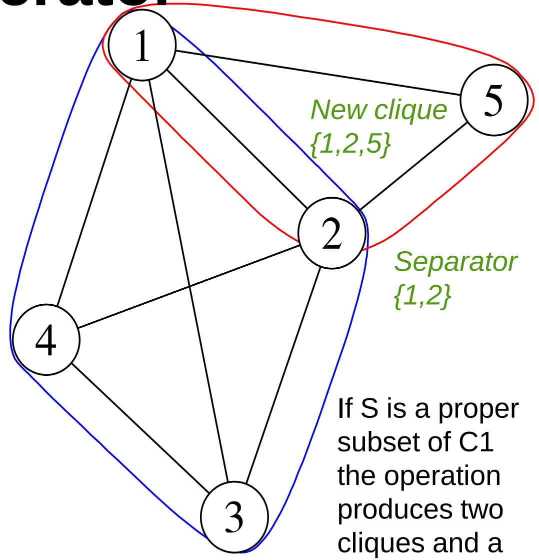
What the algorithm does?

Growing Clique Forests

Sparse (chordal) networks are built by connecting local configurations (cliques) accordingly with a gain function

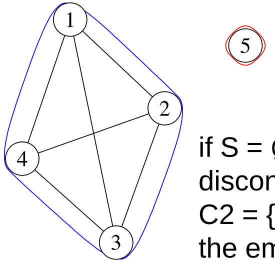
Clique Expansion Operator



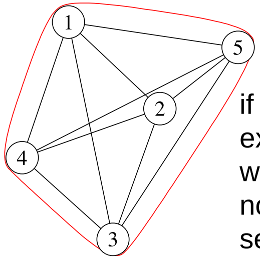


If S is a proper subset of C1 the operation produces two cliques and a separator

Special cases



if S = Ø the result produces two disconnected cliques C1 and C2 = {5} and the separator is the empty set



if S = C1 the operation purely expands the original clique with the new vertex and does not introduce a new clique or separator.

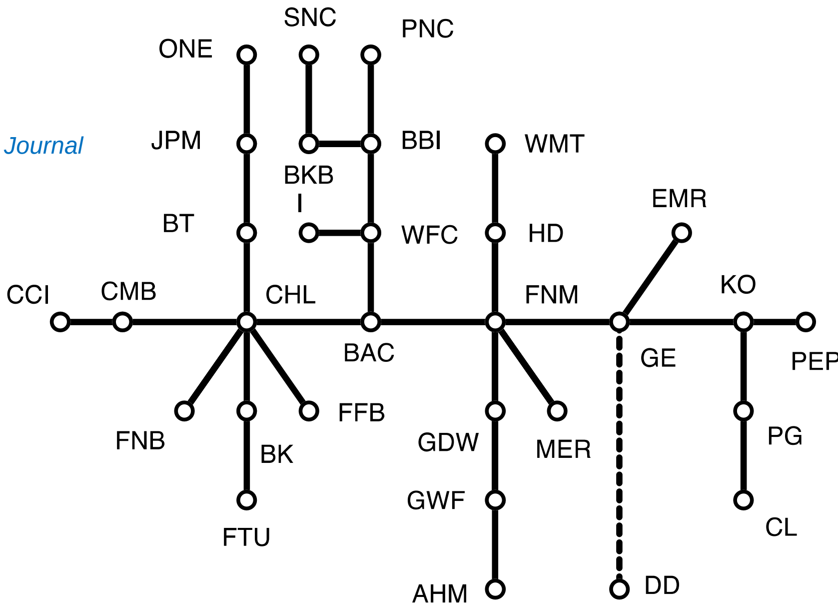
Example: Correlation Networks

Minimum Spanning Tree (MST)

Otakar Boruvka (1926)

Mantegna, Rosario N. "Hierarchical structure in financial markets." *The European Physical Journal B-Condensed Matter and Complex Systems* 11.1 (1999): 193-197.

Clique size 2



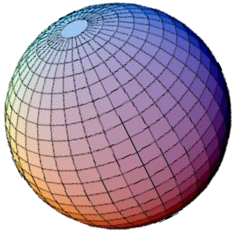
Generalization: Information Filtering Networks

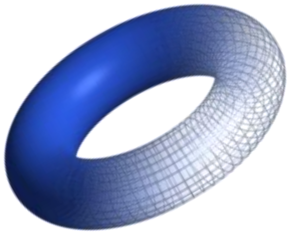
M. Tumminello, TA, T. Di Matteo, R.N. Mantegna, “A tool for filtering information in complex systems”, *PNAS* 102 (2005) 10421-10426.
Guido Previde Massara, Tiziana Di Matteo and Aste, Tomaso Network filtering for big data: Triangulated maximally filtered graph Journal of complex Networks, 5 (2016) 161--178
W.M. Song, T. Di Matteo and T. Aste, “Hierarchical information clustering by means of topologically embedded graphs”, *PLoS ONE*, 7 (2012) e31929
Won-Min Song, T. Di Matteo, TA, Nested hierarchies in planar graphs, Discrete Applied Mathematics 159 (2011) 2135-2146.

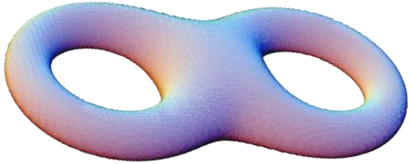
The minimum spanning tree is an example with gain function given by the correlation

Information Filtering Networks

Embedding networks on surfaces: PMFG







Any network can be embedded on a surface: the embedding of K_N is possible on an orientable surface S_g of genus

$$g \geq g^* = \left\lceil \frac{(N-3)(N-4)}{12} \right\rceil$$

TA, T. Di Matteo and S. T. Hyde, *Physica A* 346 (2005) 20-26.
M. Tumminello, TA, T. Di Matteo, R.N. Mantegna, “A tool for filtering information in complex systems”, *PNAS* 102 (2005) 10421-10426.
TA, R Gramatica, T. Di Matteo, Philosophical Magazine 92 (2012) 246-254
T Aste, R Gramatica and T. Di Matteo, Phys. Rev. E., 86 (2012) 036109.

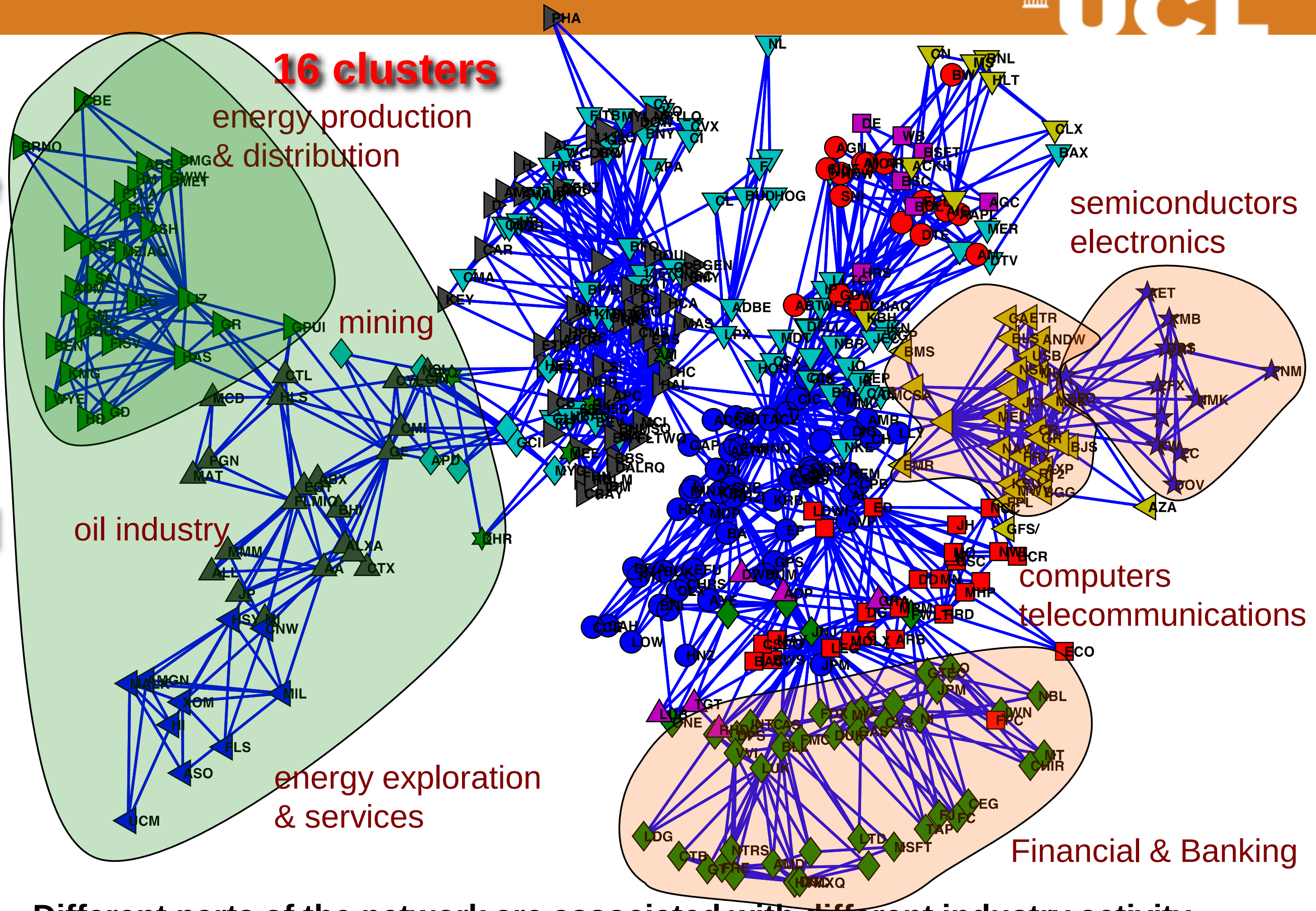
any Γ_N is a sub-graph of K_N and can be embedded on S_g

G. Ringel, Map Color Theorem, Springer-Verlag, Berlin, (1974) cap. 4; P. J. Gilbin, Graphs, Surfaces and Homology, Chapman and Hall, 2nd edition (1981); G. Ringel and J. W. T. Youngs, Proc. Nat. Acad. Sci. USA 60 (1968) 438-445.

The surface constraints the complexity of the network

(the degree of interwoveness)

The structure of Information Filtering Network is Meaningful

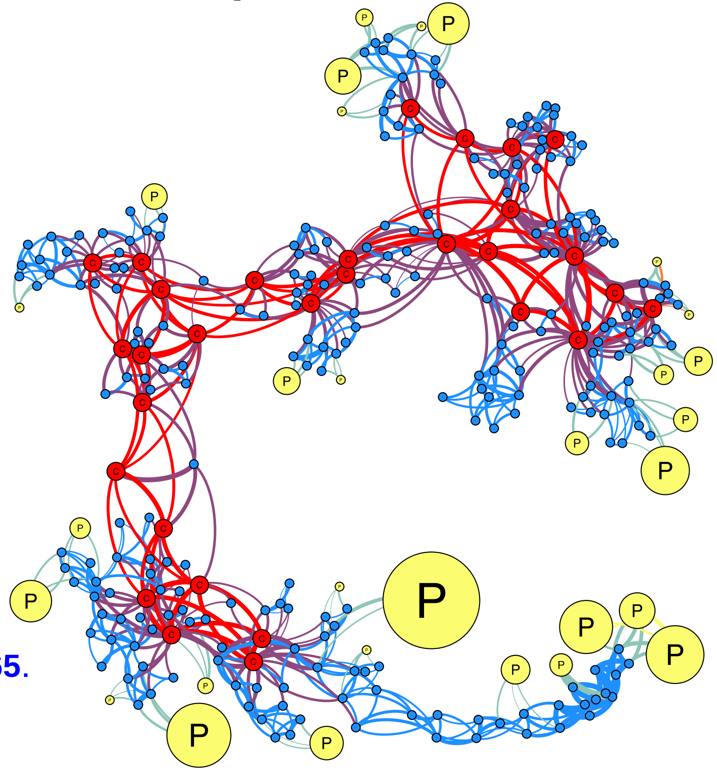


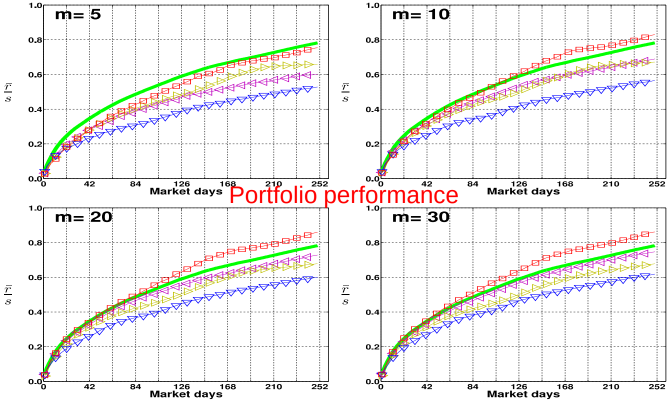
Different parts of the network are associated with different industry activity

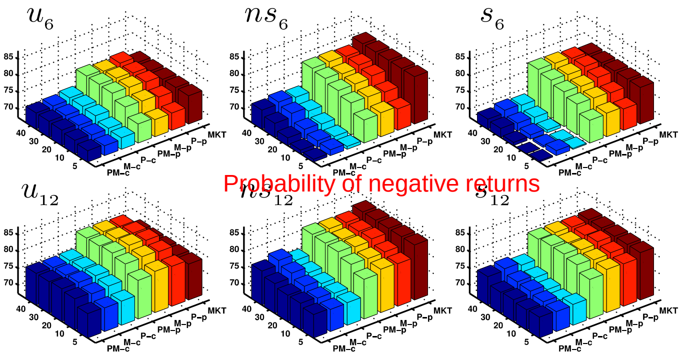
Information Filtering Networks can be used for efficient portfolio differentiation by selecting stocks from the periphery of the PMFG

Different parts of the network carry different risk factors

The structure of Information Filtering Network is Useful



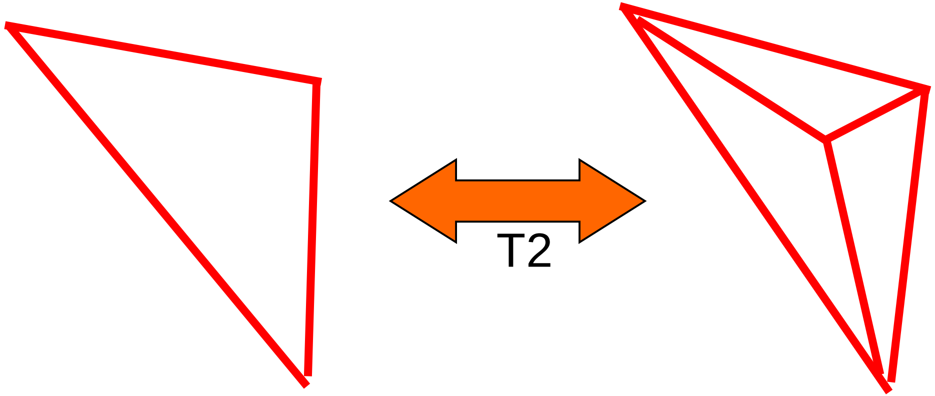


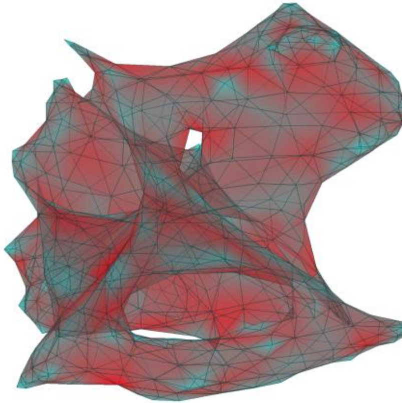


F. Pozzi, T. Di Matteo, and TA , “Spread of risk across financial markets: better to invest in the peripheries”, **Scientific Reports 3 (2013) 1665**.

TMFG

Triangulated
Maximal
Filtering
Graph





Chordal planar graphs can be constructed through one elementary move: the insertion of an edge within a triangle

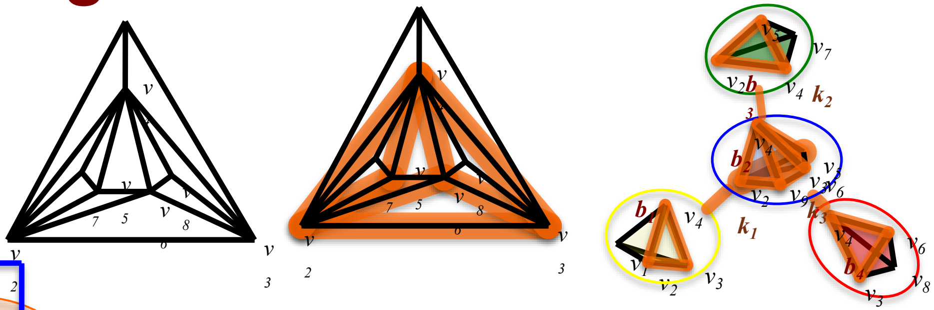
GP Massara, T Di Matteo, T Aste Network filtering for big data: Triangulated maximally filtered graph" Journal of complex Networks 5 (2) (2016) 161-178

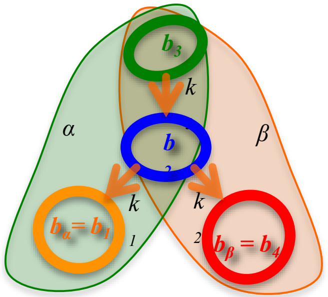
TA and D. Sherrington, "Glass transition in self organizing cellular patterns", *J. Phys. A* 32 (1999) 7049-56.
TA, Ruggero Gramatica and T. Di Matteo, Exploring complex networks via topological embedding on surfaces , Phys. Rev. E., 86 (2012) 036109 arXiv:1107.3456v1 (2011)
TA, Ruggero Gramatica and T. Di Matteo, Random and frozen states in complex triangulations, Philosophical Magazine (2011).

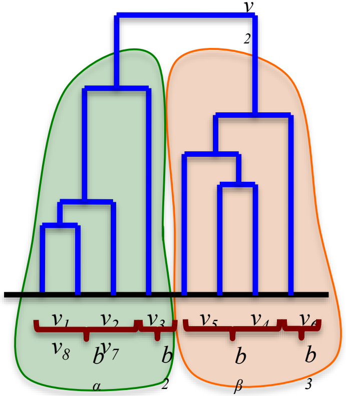
Hierarchy and clustering

Information Filtering Networks are Partial Order Sets (poset)

Minimal Cycles (3-cliques) in Maximal Planar Graphs contain other cliques inside or/and they are contained inside the other cliques providing a natural hierarchy







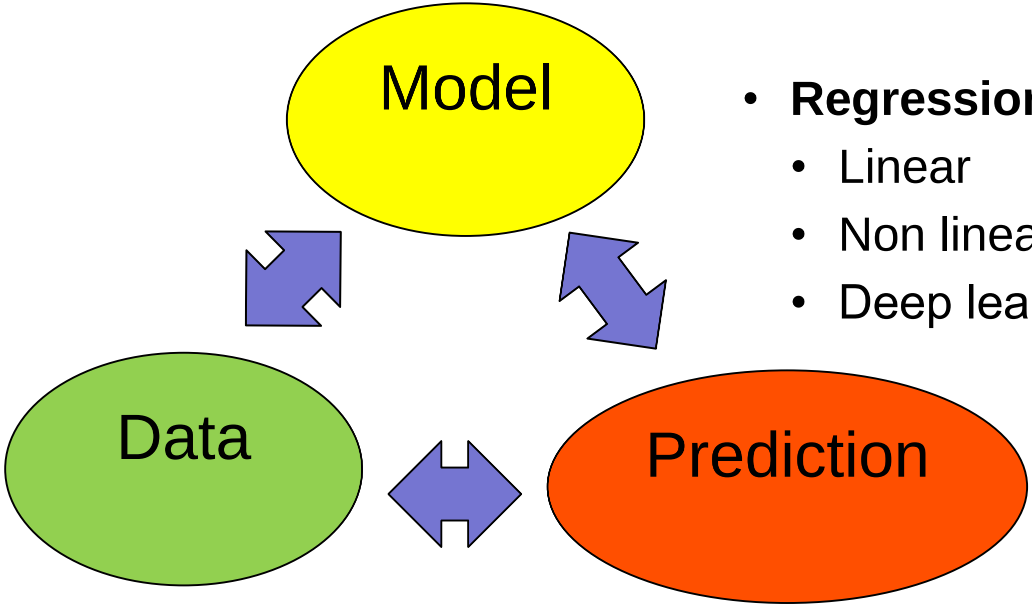
Directed **B**ubble **H**ierarchical **T**ree

A deterministic method to capture both local clustering and global hierarchical organization without introducing any characteristic scale

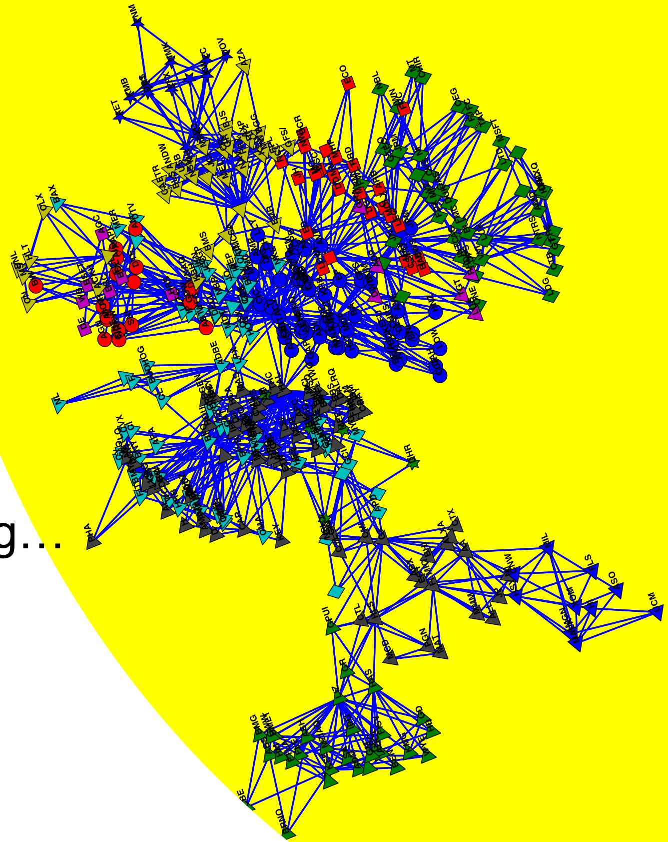
W.M. Song, T. Di Matteo and T. Aste, “Hierarchical information clustering by means of topologically embedded graphs”, *PLoS ONE*, 7 (2012) e31929
Won-Min Song, T. Di Matteo, TA, Nested hierarchies in planar graphs, Discrete Applied Mathematics 159 (2011) 2135-2146.

Using
Information Filtering Networks
quantitatively

From data to prediction: solving an inverse problem



- **Regression**
  - Linear
  - Non linear
  - Deep learning…



First difficulty

The more you look
The less you see

(Curse of dimensionality)

To model a system with *N* interacting variables we have

$$\frac{N^2}{2} \text{ £ Quantities to Estimate £ } N!$$

- For each variable we can perform a set of observations
- **Information increases <u>linearly</u> with the number of variables observed**
- But model-parameters increase <u>at least</u> with the <u>square</u>!

The weakest link

(Error propagation)

Data come with their own uncertainty

Heterogeneous data sources have big differences in reliability

Model accuracy is determined by the most uncertain data source

A solution: parsimonious modeling

Predictive models built on networks solve the three main problems:

1. **Curse of dimensionality**: *Sparsity reduces the number of model parameters to O(N)*
2. **Error propagation**: *model construction can be done locally*
3. **Interpretability**: *sparse networks are easy to interpret*

Barfuss, Wolfram, et al.
Parsimonious modeling with information filtering networks
Physical Review E 94.6 (2016): 062306.

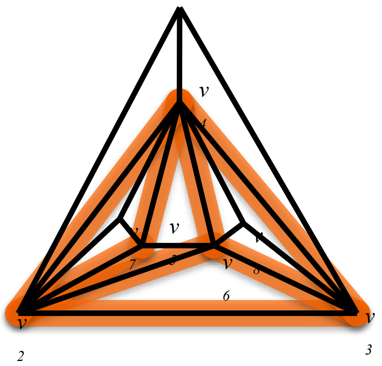
Probabilistic modeling with Information Filtering Networks

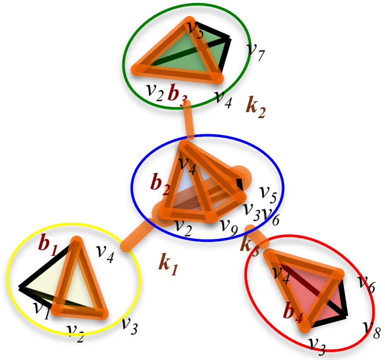
(clique forests)

Predictive modeling with Information Filtering Networks

Given a conditional probability structure organized in a **chordal graph**, then

$$p(X_{All}) = \frac{\prod_{cliques} p(X_{cliques})}{\prod_{separators} p(X_{separators})^{k_s - 1}}$$





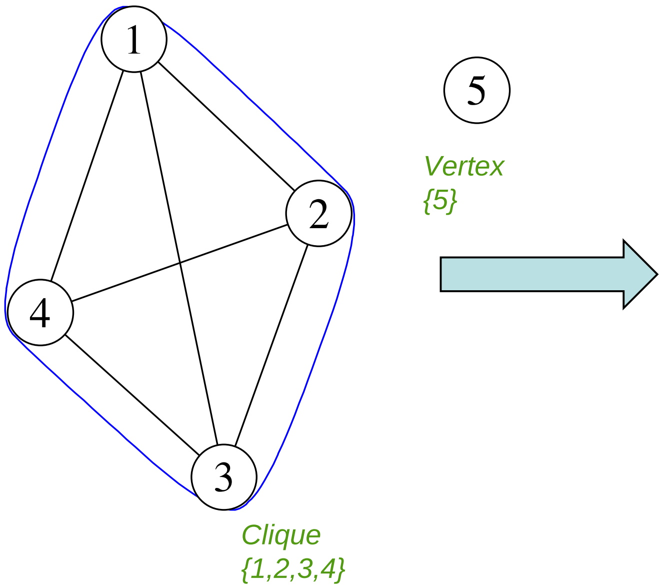
Theorem:

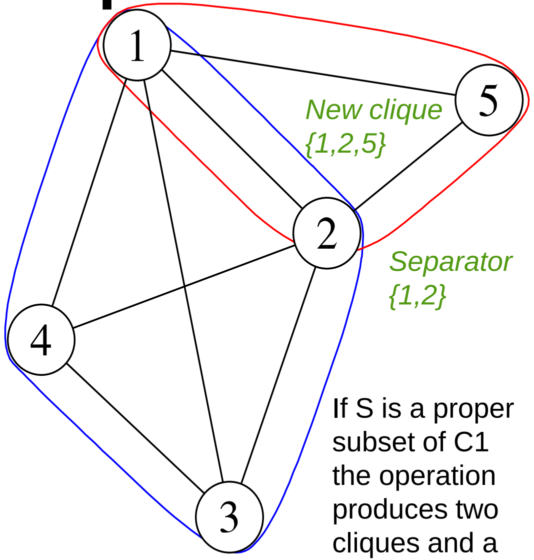
A graph G is **chordal** iff it has a **clique forest** T(C, S)

Parsimonious modeling with information filtering networks." Barfuss, Wolfram, et al. Physical Review E 94.6 (2016): 062306.

The algorithm MCFC

Clique Expansion Operator

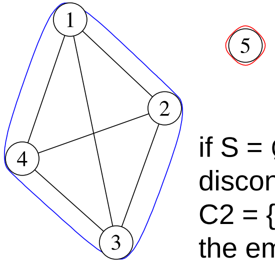




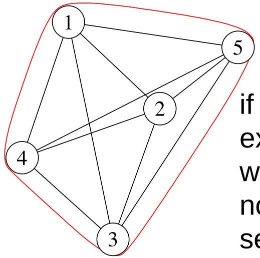
If S is a proper subset of C1 the operation produces two cliques and a separator

Sparse (chordal) networks are built by connecting local configurations (cliques) accordingly with a gain function

Special cases



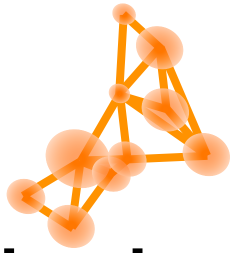
if S = ∅ the result produces two disconnected cliques C1 and C2 = {5} and the separator is the empty set



if S = C1 the operation purely expands the original clique with the new vertex and does not introduce a new clique or separator.

Predictive modeling with Information Filtering Networks

Information Filtering Network (chordal)



Multivariate probability

$$p(X_{All}) = \frac{\prod_{cliques} p(X_{cliques})}{\prod_{separators} p(X_{separators})^{k_s - 1}}$$

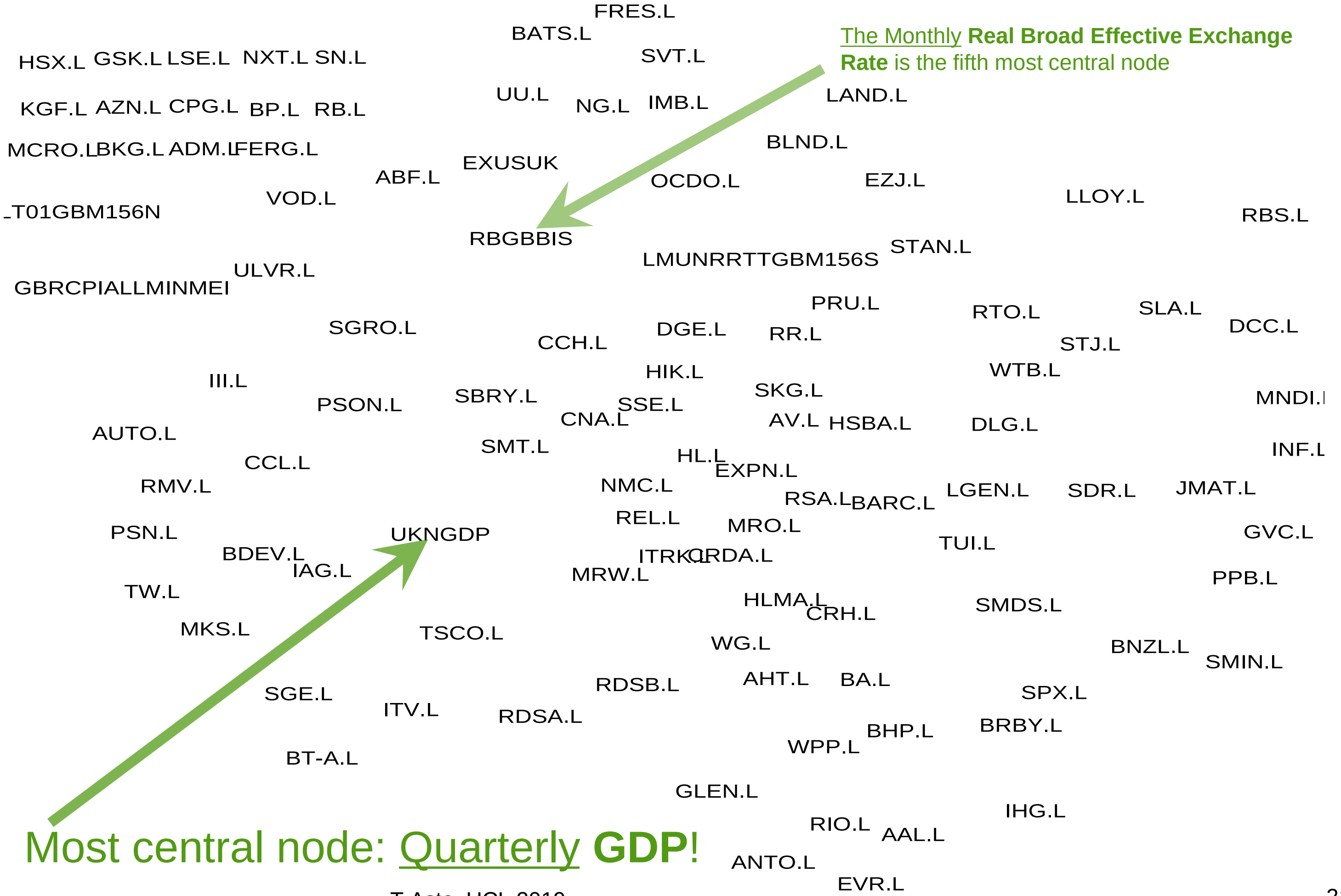
Modelling, Prediction & Forecasting

Applications

Mixing daily monthly and quarterly data

Gain function:
R^2 from local support vector regression (non-linear)

The Monthly **Real Broad Effective Exchange Rate** is the fifth most central node



Most central node: Quarterly **GDP**!

Networks & investment strategies

Inverse covariance estimated from the clique forest compared with the sample one

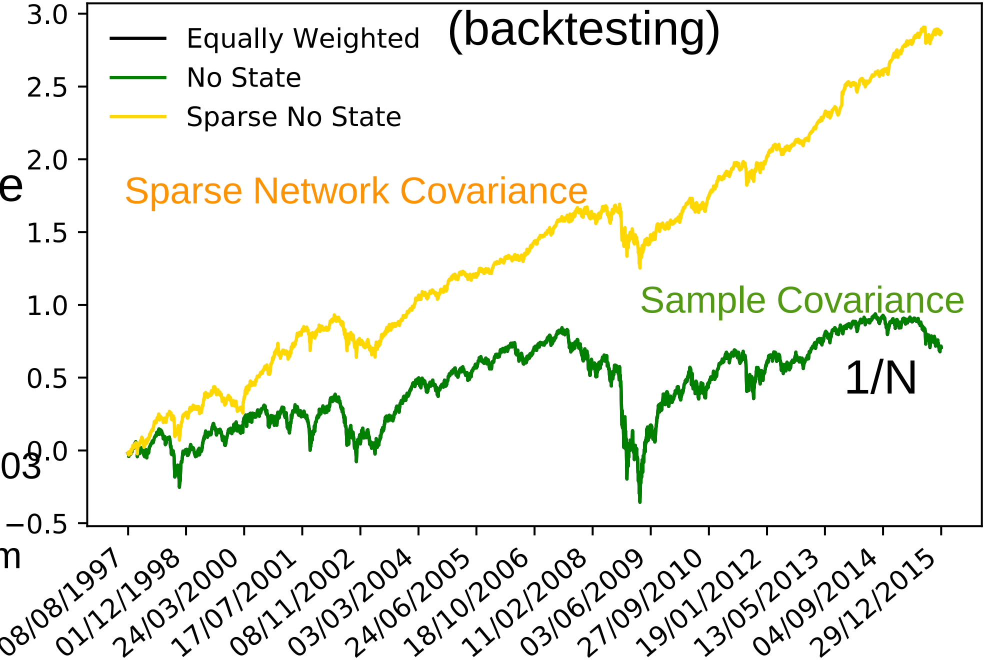
Experiment

Dataset: 100 stocks chosen at random among Russel 1000 index (RIY index) traded between 01/02/1995 and 12/31/2015

Two states from Euclidean distance

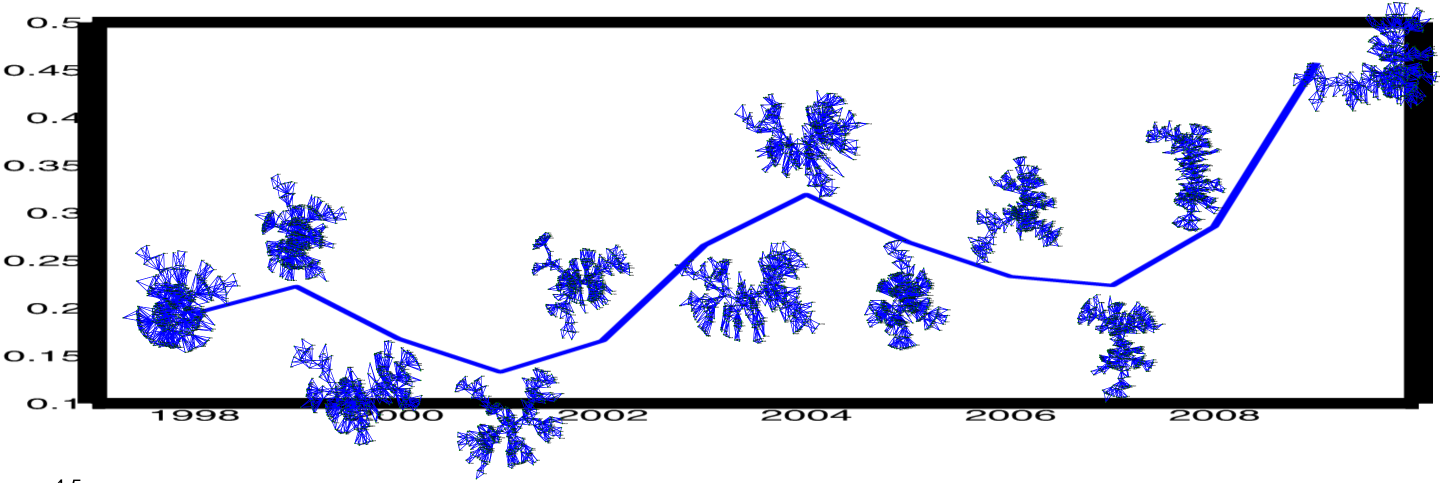
Black–Litterman asset allocation

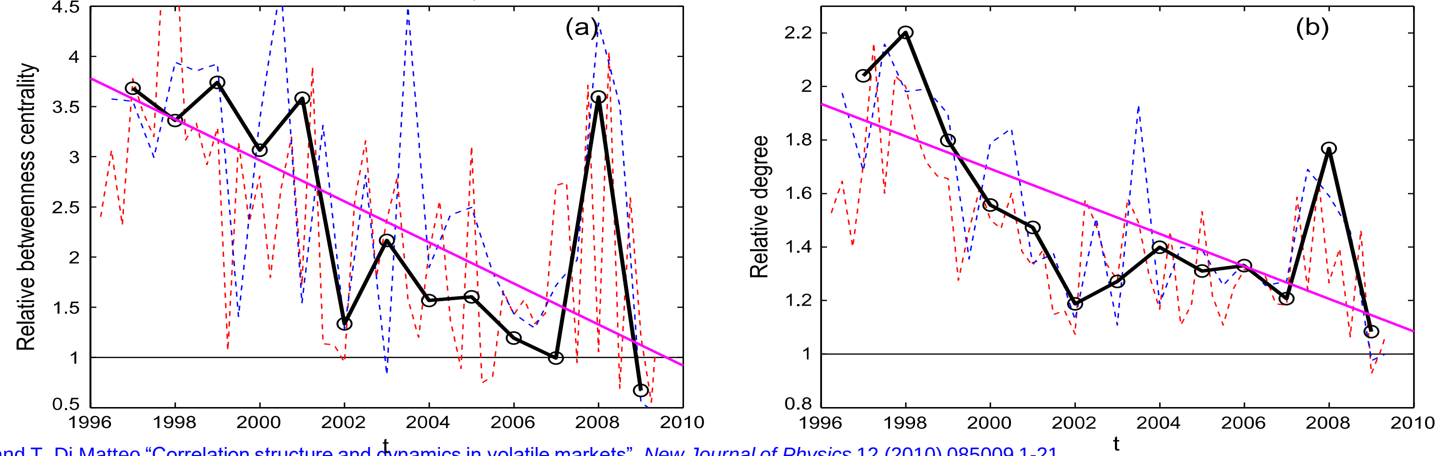
Confidence level c=0.003
Risk aversions π=3
Selection matrix uniform



PF Procacci & T Aste, Forecasting Market States, *preprint arXiv:1807.05836 (2018)*

Dynamical evolution





T. Aste, W. Shaw and T. Di Matteo "Correlation structure and dynamics in volatile markets", *New Journal of Physics* 12 (2010) 085009 1-21.
T. Di Matteo, F. Pozzi, T. Aste, "The use of dynamical networks to detect the hierarchical organization of financial market sectors", *Eur. Phys. J. B* 73 (2010) 3–11.

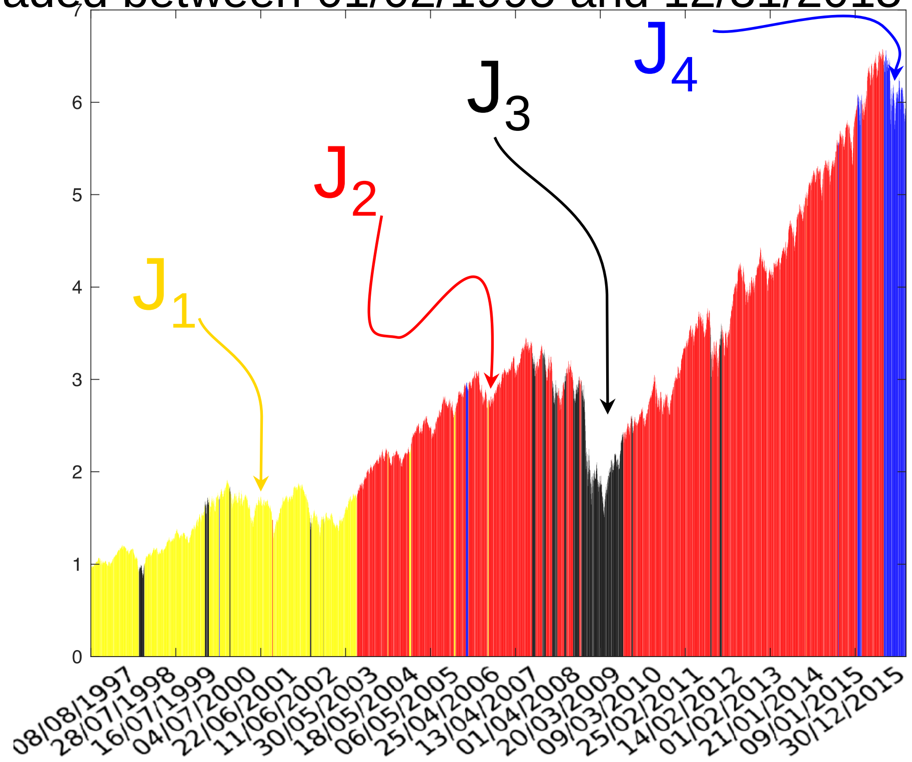
Market states switch

ICC clustering method with penalized likelihood

Experiment

Dataset: 100 stocks chosen at random among Russel 1000 index (RIY index) traded between 01/02/1995 and 12/31/2015

Four states from Likelihood



PF Procacci & T Aste, Forecasting Market States, *preprint arXiv:1807.05836 (2018)*

Market states & investment strategies

ICC clustering method with Euclidean distance

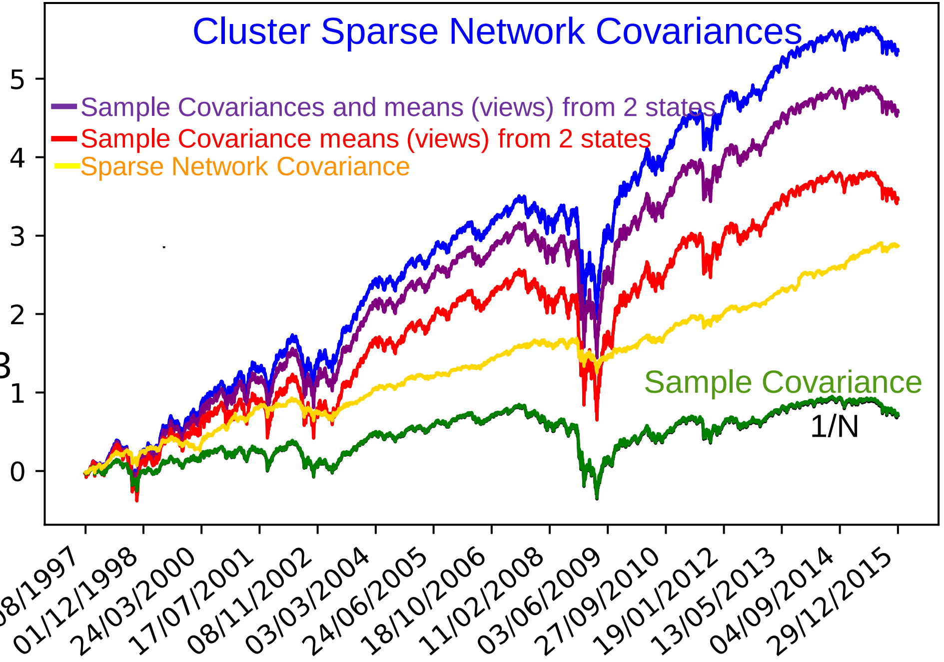
Experiment

Dataset: 100 stocks chosen at random among Russel 1000 index (RIY index) traded between 01/02/1995 and 12/31/2015 (backtesting)

Two states from Euclidean distance

Black–Litterman asset allocation

Confidence level c=0.003
Risk aversions π=3
Selection matrix uniform



PF Procacci & T Aste, Forecasting Market States, *preprint arXiv:1807.05836 (2018)*

Conclusions and take-home message

- We are solving an inverse problem
- More data do not necessary provide more accurate models
- Networks are great tool to navigate complexity
- Models build on networks provide powerful quantitative tools for risk management, asset allocation and investment strategies
- A lot more can be done! *Collaborations welcome*

Links and references

LINKS

FCA Group Page:
http://fincomp.cs.ucl.ac.uk/introduction/

My articles:
https://scholar.google.co.uk/citations?user=27pUbTUAAAAJ&hl=en

Software:
TMFG & Clique Forests
https://github.com/cvborkulo/NetworkComparisonTest/pull/5
https://uk.mathworks.com/matlabcentral/fileexchange/56444-tmfg

RELEVANT PAPERS

"Learning Clique Forests." Massara, Guido Previde, and Tomaso Aste. arXiv preprint arXiv:1905.02266 (2019).

"Predicting future stock market structure by combining social and financial network information." Souza, Thársis TP, and Tomaso Aste. arXiv preprint arXiv:1812.01103 (2018).

"Forecasting market states." Procacci, Pier Francesco, and Tomaso Aste. arXiv preprint arXiv:1807.05836 (2018).

"Sparse causality network retrieval from short time series." Aste, Tomaso, and Tiziana Di Matteo. Complexity 2017 (2017).

"Parsimonious modeling with information filtering networks." Barfuss, Wolfram, et al. Physical Review E 94.6 (2016): 062306.

"Network filtering for big data: Triangulated maximally filtered graph" GP Massara, T Di Matteo, T Aste Journal of complex Networks 5 (2) (2016) 161-178

"Relation between financial market structure and the real economy: comparison between clustering methods" N Musmeci, T Aste, T Di Matteo PloS one 10 (3), e0116201

"Risk diversification: a study of persistence with a filtered correlation-network approach." Musmeci, N., T. Aste, and T. D. Matteo. Journal of Network Theory in Finance 1.1 (2015): 77-98.