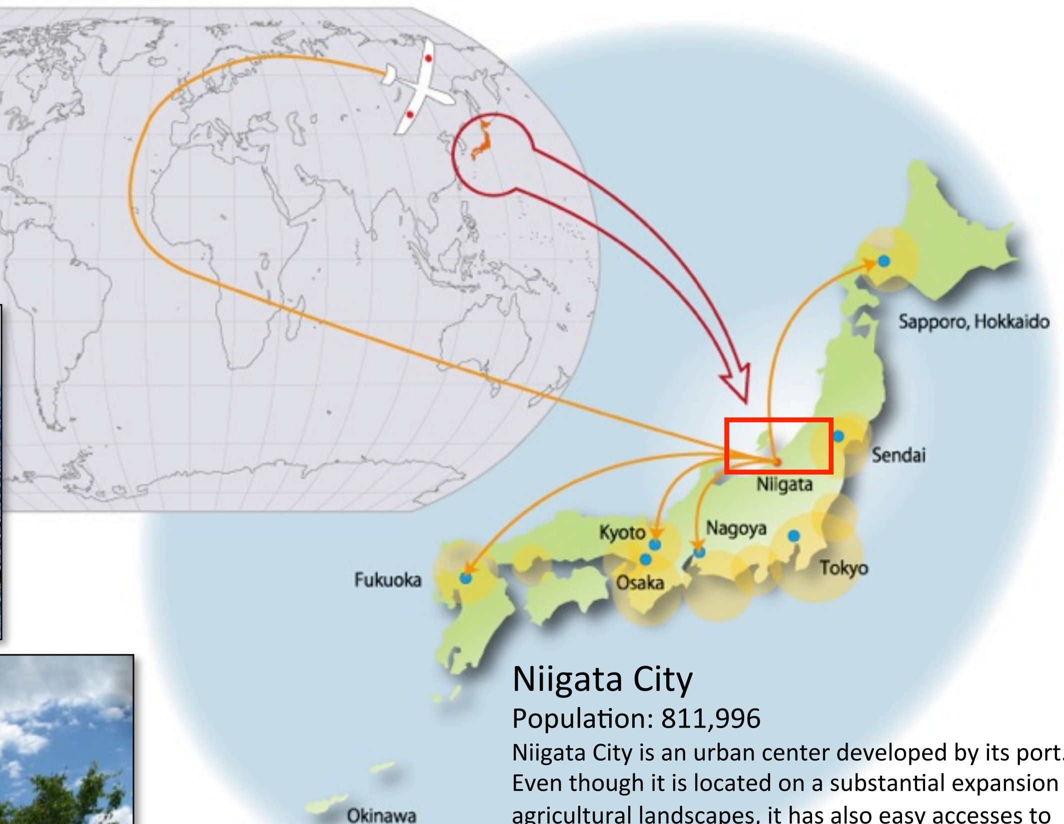
Frustration in Financial Markets

Hiroshi Iyetomi
Department of Mathematics, Niigata University

Collaborators (My PhD students): Takeo Yoshikawa and Yuta Arai

Where am I from?



Niigata City

Population: 811,996

Niigata City is an urban center developed by its port. Even though it is located on a substantial expansion of agricultural landscapes, it has also easy accesses to major cities by airplanes, express omnibuses, and bullet trains. Also from its international airport, there are regular flights to Harbin, Shanghai, Seoul, Vladivostok , Khabarovsk, Guam. Niigata aspires to be a gateway to the East Asia.

Econophysics Approach to Financial Markets

- Laloux, L., Cizeau, P., Bouchaud, J.-P., Potters, M., Phys. Rev. Lett. **83**, 1467 (1999).
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Stanley, H., E., Phys. Rev. Lett. **83**, 1471 (1999).

Noise elimination using the random matrix theory (RMT)

Statistically meaningful information

- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., Stanley, H., E., Phys. Rev. E **65**, 066126 (2002).
- Utsugi, A., Ino, K., Oshikawa, M., Phys. Rev. E **70**, 026110 (2004).
- Kim, D., H., Jeong, H., Phys. Rev. E 72, 046133 (2005).
- Pan, R., K., Shinha, S., Phys. Rev. E 76, 046116 (2007).
- And more

> Insight into correlation structure in stock exchange markets
> Identification of collective motion of business groups

Objective: Stock Market as a Network

We revisit correlation structures in the Tokyo Stock Exchange (TSE) and S&P 500 markets from a network theoretic point of view.

- Correlation matrix is regarded as an **adjacency matrix** for a network
- The network thus constructed has links with **negative** weights as well as links with **positive** weights.

Communities are then detected through minimizing frustration due to anticorrelations inside communities of nodes (stock prices).

Community detection for networks with links of both signs:

- S. Gómez, P. Jensen and A. Arenas, A., Phys. Rev. E 80, 036115 (2009).
- V.A. Traag and J. Bruggeman, Phys. Rev. E **80**, 016114 (2009).

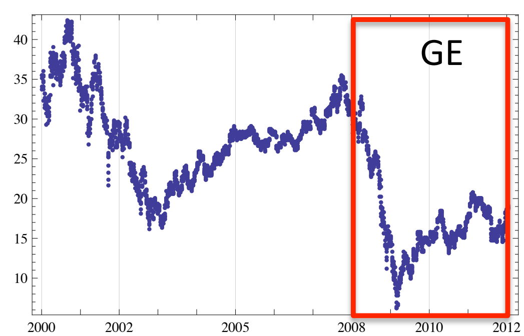
Market Data Used in the Present Study

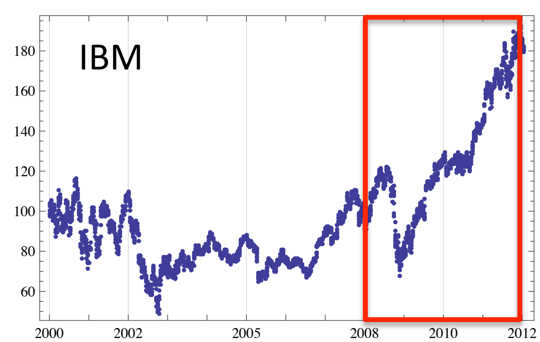
Tokyo Stock Exchange

> 557 stocks
> 2707 days (Jan. 4, 1996 through Dec. 29, 2006)
> 33 industrial groups (SIC)

S&P 500

> 483 stocks
> 1009 days (Jan.2, 2008 through Dec. 30, 2011)
> 24 industrial groups (GICS)





Standard Industrial Classification (SIC) System --- TSE

Main classification	No.	Sub classification	No. of Stocks
Fishery, Agriculture & Forestry	1	Fishery, Agriculture & Forestry	4
Minig	2	Mining	2
Construction	3	Construction	31
Manufacturing	4	Foods	32
	5	Textiles & Apparels	17
	6	Pulp & Paper	5
	7	Chemicals	60
	8	Pharmaceutical	13
	9	Oil & Coal Products	3
	10	Rubber Products	7
	11	Glass & Ceramics Products	15
	12	Iron & Steel	20
	13	Nonferrous Metals	14
	14	Metal Products	8
	15	Machinery	55
	16	Electric Appliances	72
	17	Transportation Equipment	30
	18	Precision Instruments	13
	19	Other Products	14
Electric Power & Gas	20	Electric Power & Gas	13
Transportation, Information & Communication	21	Land Transportation	17
	22	Marine Transportation	5
	23	Air Transportation	1
	24	Warehousing & Harbor Transportation Services	5
	25	Information & Communication	10
Trade	26	Wholesale Trade	21
	27	Retail Trade	17
Finance & Insurance	28	Banks	20
	29	Securities & Commodity Futures	6
	30	Insurance	6
	31	Other Financing Business	9
Real Estate	32	Real Estate	6
Services	33	Services	6

Global Industry Classification Standard (GICS) --- S&P 500

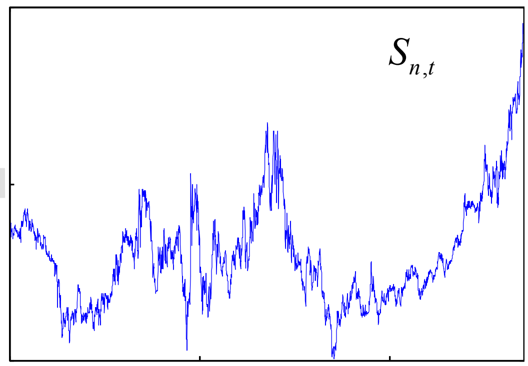
Sector	No.	Industry	No. of Stocks
Energy	1	Energy	39
Materials	2	Materials	29
Industrials	3	Capital Goods	40
	4	Commercial & Professional Services	11
	5	Transportatin	9
Consumer Discretionary	6	Automobiles & Components	4
	7	Consumer Durables & Aparel	14
	8	Consumer Services	14
	9	Media	15
	10	Retailing	32
Consumer Staples	11	Food & Staples Retailing	9
	12	Food Beverage & Tobacco	21
	13	Household & Personal Products	6
Health Care	14	Health Care Equipment & Services	31
	15	Pharmaceuticals, Biotechnology & LifeSciences	19
Financials	16	Banks	15
	17	Diversified Financials	27
	18	Insurance	21
	19	Real Estate	16
Information Technology	20	Software & Services	28
	21	Technology Hardware & Equipment	26
	22	Semiconductors & Semiconductor Equipment	16
Telecommunication Services	23	Telecommunication Services	8
Utilities	24	Utilities	33

Preprocessing of the Stock Price Data

Log return:

$$R_{n,t} \equiv \ln S_{n,t+1} - \ln S_{n,t}$$

$S_{n,t}$: Price of stock *n* at time *t*

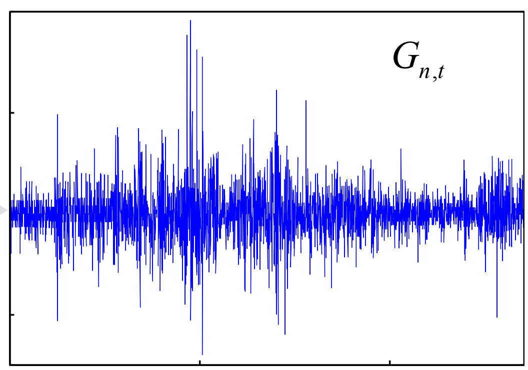


Standardization:

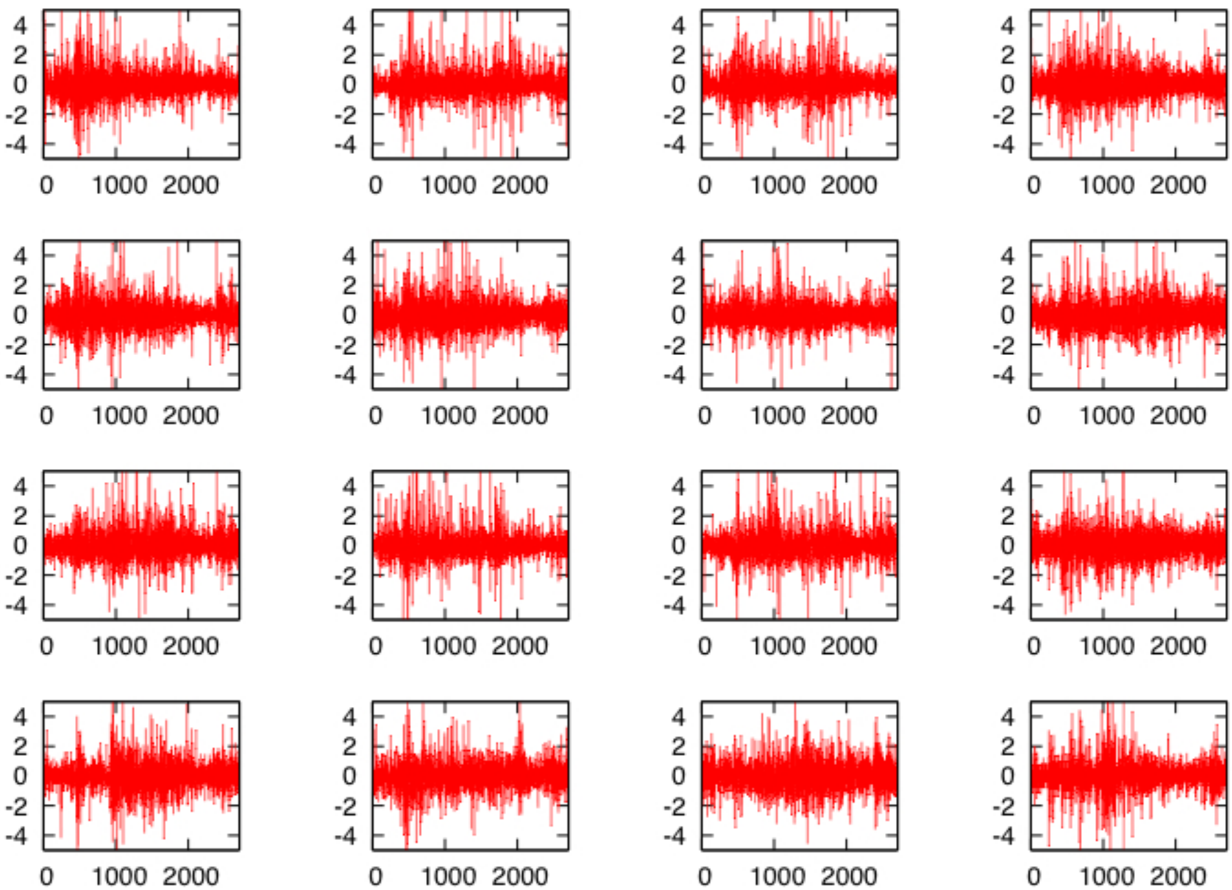
$$G_{n,t} \equiv \frac{R_{n,t} - \langle R_{n,t} \rangle}{\sigma_n}$$

$$\langle R_n \rangle \equiv \frac{1}{T} \sum_{t=1}^{T} R_{n,t}$$

$$\sigma_n^2 \equiv \frac{1}{T} \sum_{t=1}^{T} \left(R_{n,t} - \langle R_n \rangle \right)^2$$

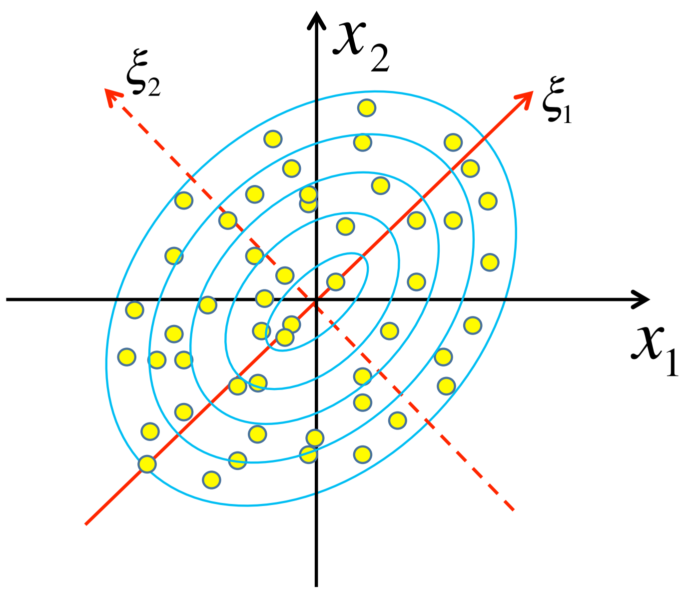


Examples of the Preprocessed TSE Data



Principal Component Analysis (PCA)

$$\mathbf{C}\mathbf{V}_i = \lambda_i \mathbf{V}_i$$ ⟹ Principal Components (normal coordinates in physics) ξ_1, ξ_2



$$\mathbf{C} = \begin{pmatrix} \langle x_1^2 \rangle = 1 & \langle x_1 x_2 \rangle \\ \langle x_2 x_1 \rangle & \langle x_2^2 \rangle = 1 \end{pmatrix} \Rightarrow \mathbf{C}' = \mathbf{U}\mathbf{C}\mathbf{U}^{+} = \begin{pmatrix} \langle \xi_1^2 \rangle = \lambda_1 & 0 \\ 0 & \langle \xi_2^2 \rangle = \lambda_2 \end{pmatrix}$$

Spectral Decomposition of Correlation Matrix

Data matrix G

$$\mathbf{G} \equiv \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1T} \\ G_{21} & G_{22} & \cdots & G_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1} & G_{N2} & \cdots & G_{NT} \end{pmatrix}$$

N time series of length T

Correlation matrix C

$$\mathrm{C} \equiv \frac{1}{T}\mathrm{G}\mathrm{G}^{\mathrm{T}} \qquad C_{i,j} = \frac{1}{T}\sum_{t=1}^{T} G_{i,t}G_{j,t}$$

The correlation matrix can be decomposed into

$$\mathrm{C} = \sum_{i=1}^{N} \lambda_i \mathrm{u}_i \mathrm{u}_i^{\mathrm{T}}$$

How many principal components should we take into account as statistically significant ones?

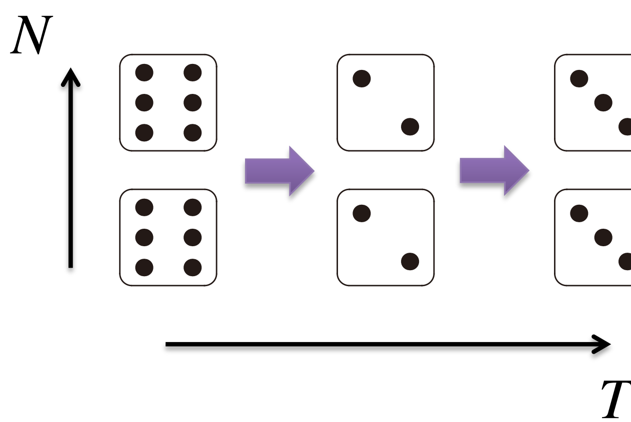
λ_i: the eigenvalues of C, $\lambda_1 > \lambda_2 > \cdots > \lambda_N$

u_i: the eigenvectors corresponding to λ_i $\qquad \mathrm{u}_i \cdot \mathrm{u}_i = 1$

Illusion in Randomness

Finite length of time series data brings about spurious correlations, which may be observed even in completely random data.

The RMT is a clue to this problem!



$$T \to \infty \qquad \Rightarrow \qquad C_{ij} = \delta_{ij}$$

RMT

$$T \to \infty, \quad N \to \infty \quad \text{with fixed } Q \equiv \frac{T}{N} \quad \Rightarrow \quad \text{nontrivial results for } C_{ij}$$

Random Matrix Theory (RMT)

$\{h_{ij}\}$ are random variables following N(0,1) and hence mutually independent.

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,T} \\ \vdots & \ddots & \vdots \\ h_{N,1} & \cdots & h_{N,T} \end{pmatrix}$$

The random correlation matrix is calculated by

$$\mathbf{C} = \frac{1}{T}\mathbf{H}\mathbf{H}^{\mathrm{T}}$$

In the limit *N,T*→∞ with $Q \equiv T/N$ fixed,

> Eigenvalue distribution

> Eigenvector components distribution

> Nearest-neighbor eigenvalue distribution, etc.

have universal properties.

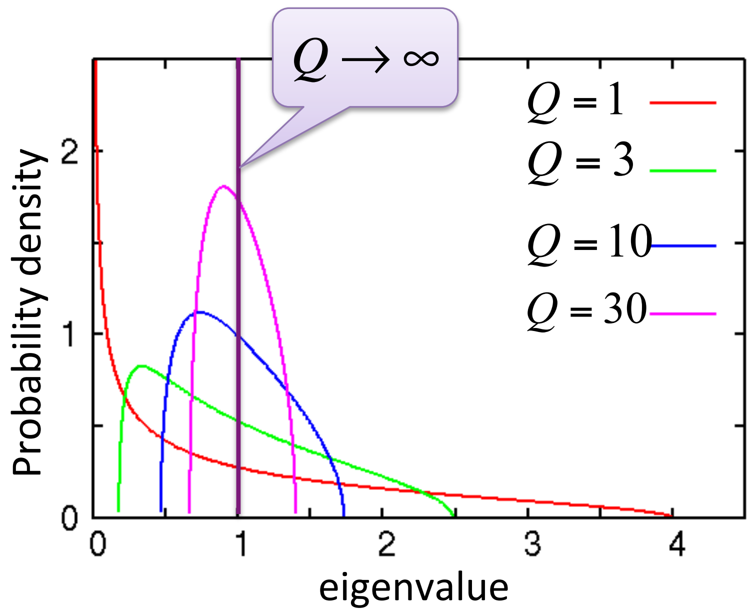
RMT: Eigenvalue Distribution

The eigenvalue distribution for the random matrix of infinite size with a given aspect ratio *Q*:

$$\rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

where

$$\lambda_\pm = \left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right).$$



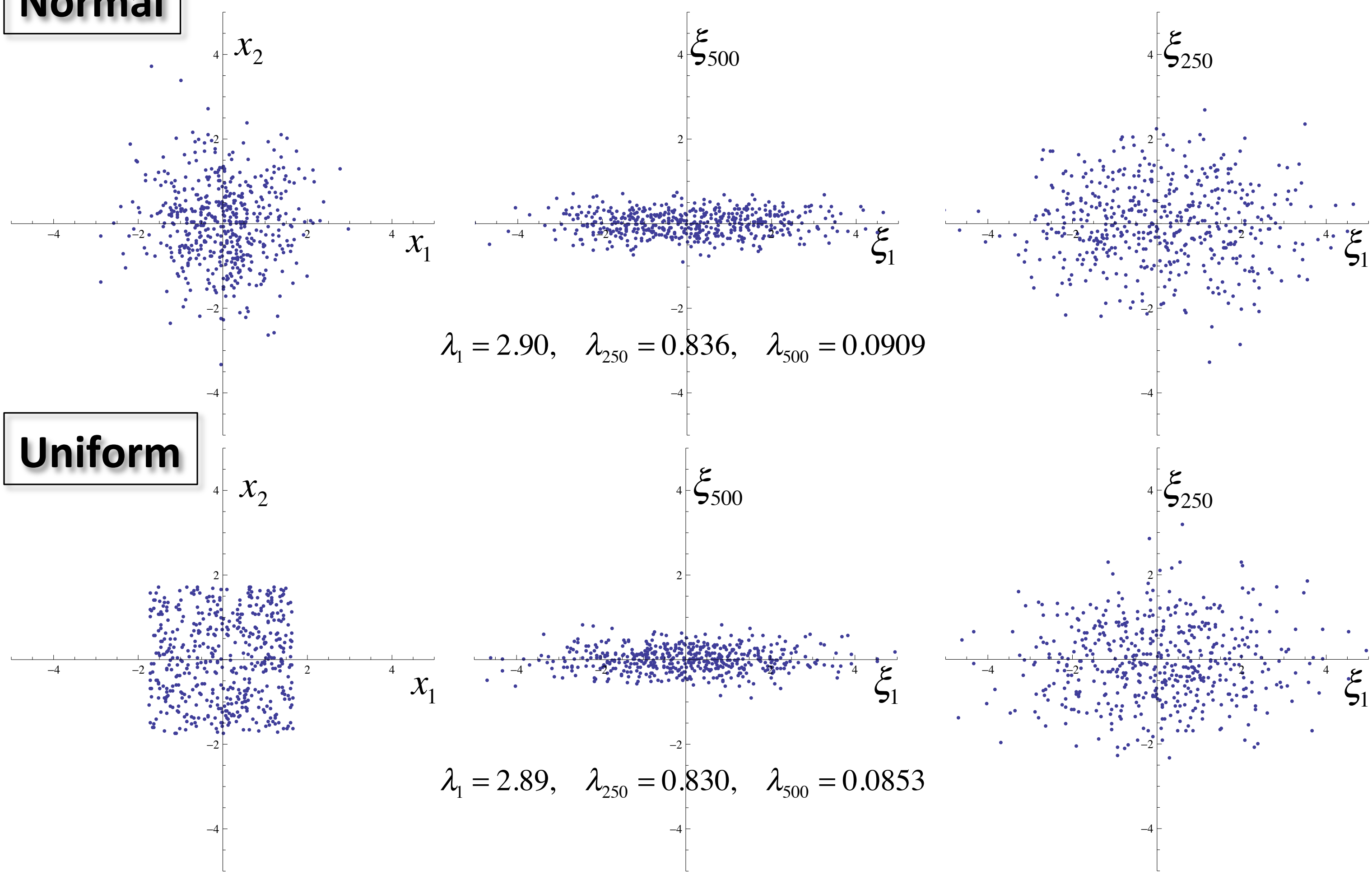
$$\lambda > \lambda_+$$ Appearance of statistically meaningful correlations

Numerical Demonstration of RMT

$$Q=\frac{T}{N}=\frac{1000}{500}=2$$

Normal

Uniform



$$\lambda_1=2.90,\quad \lambda_{250}=0.836,\quad \lambda_{500}=0.0909$$

$$\lambda_1=2.89,\quad \lambda_{250}=0.830,\quad \lambda_{500}=0.0853$$

My First Encounter with RMT

VOLUME 46, NUMBER 6 PHYSICAL REVIEW LETTERS 9 FEBRUARY 1981

Exact Results for the Two-Dimensional One-Component Plasma

B. Jancovici
Laboratoire de Physique Théorique et Hautes Energies, Université de Paris-Sud, F-91405 Orsay, France
(Received 5 November 1980)

At some special temperature T_0, the distribution functions of a two-dimensional one-component plasma are explicitly computed up to the four-body one. The correlations have a Gaussian falloff. The distribution functions at T_0 are used for building a temperature expansion around T_0.

VOLUME 46, NUMBER 6 PHYSICAL REVIEW LETTERS 9 FEBRUARY 1981

or, equivalently, the internal energy U and the specific heat c at $\Gamma=2$. Their excess parts per particle are found to be

$$U_{\text{exc}}/N = -\tfrac{1}{4}e^2\ln(\pi\rho L^2) - \tfrac{1}{4}e^2 C, \qquad (18)$$

and

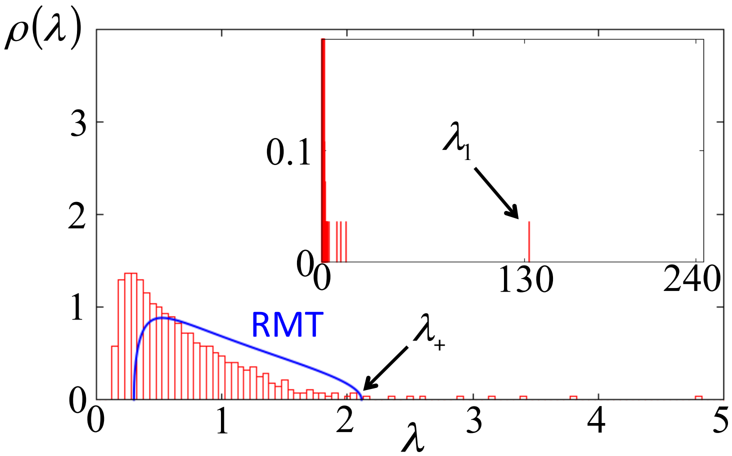
$$c_{\text{exc}}/N = k_B(\ln 2 - \pi^2/24). \qquad (19)$$

The author is indebted to C. Deutsch, H. DeWitt, J. Ginibre, and J. L. Lebowitz, for helpful re-

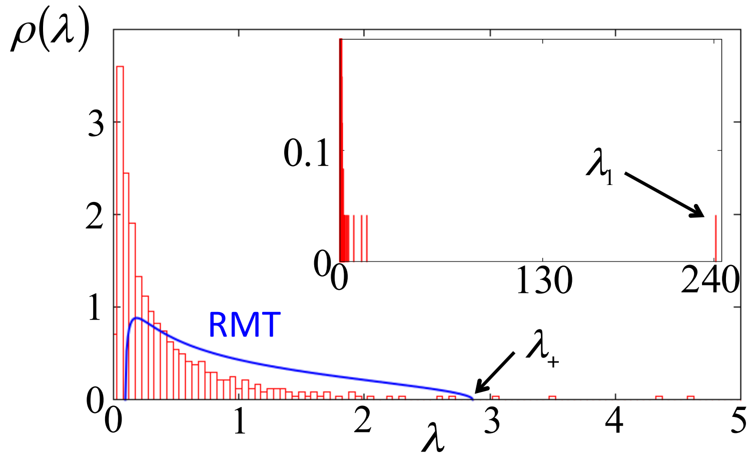
Gen. Phys. 9, 1539 (1976).
[2] R. R. Sari and D. Merlini, J. Stat. Phys. 14, 91 (1976).
[3] E. H. Hauge and P. C. Hemmer, Phys. Norv. 5, 109 (1971), and references quoted therein.
[4] A. Alastuey and B. Jancovici, to be published.
[5] J. Ginibre, J. Math. Phys. 6, 440 (1965).
[6] C. L. Mehta, *Random Matrices* (Academic, New York, 1967).
[7] D. R. Nelson, Phys. Rev. B 18, 2318 (1978), and references quoted therein.

Distribution of Eigenvalues for Correlation Matrix

TSE



S&P 500



$$Q = T / N = 2706 / 557 \cong 4.86$$

$$\lambda_{13} \cong 2.17 > \lambda_{+} \cong 2.11 > \lambda_{14} \cong 2.07$$

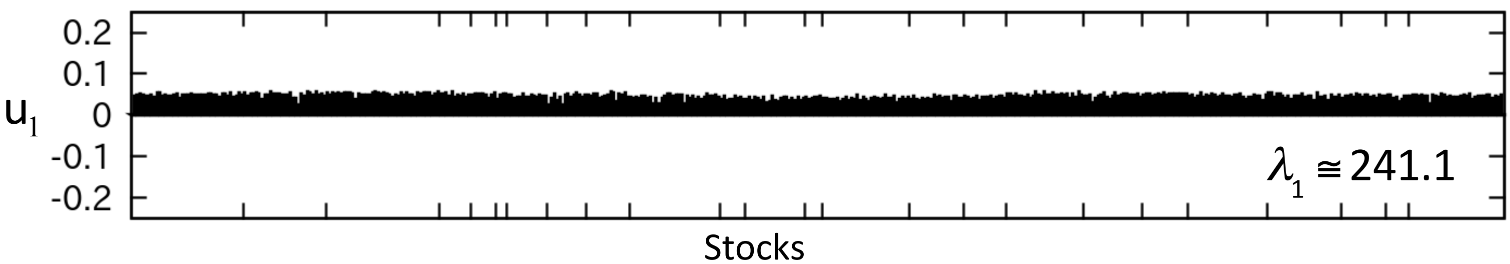
$$\mathbf{C} \approx \sum_{i=1}^{13} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

$$Q = T / N = 1008 / 483 \cong 2.01$$

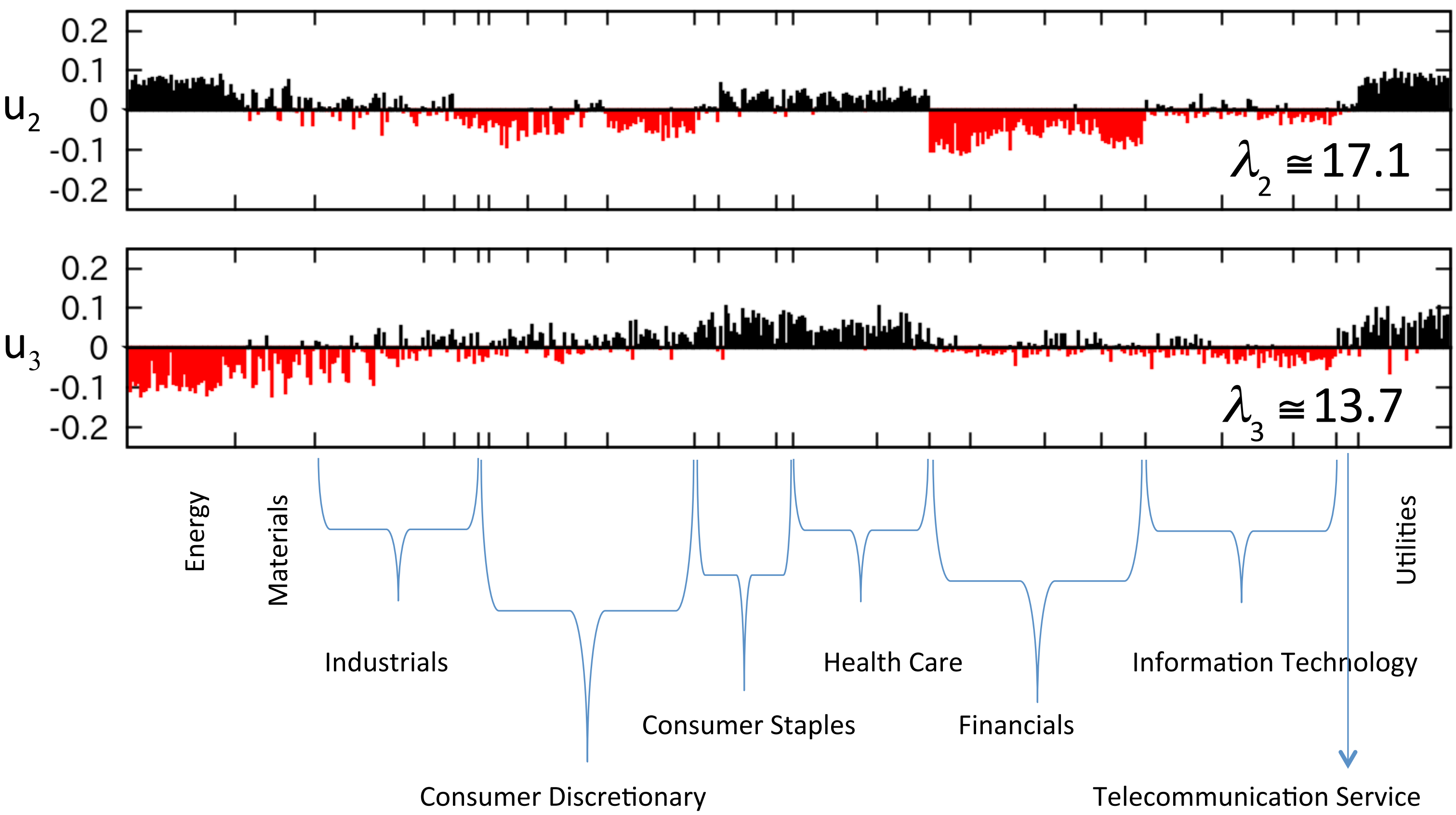
$$\lambda_{9} \cong 3.07 > \lambda_{+} \cong 2.86 > \lambda_{10} \cong 2.71$$

$$\mathbf{C} \approx \sum_{i=1}^{9} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

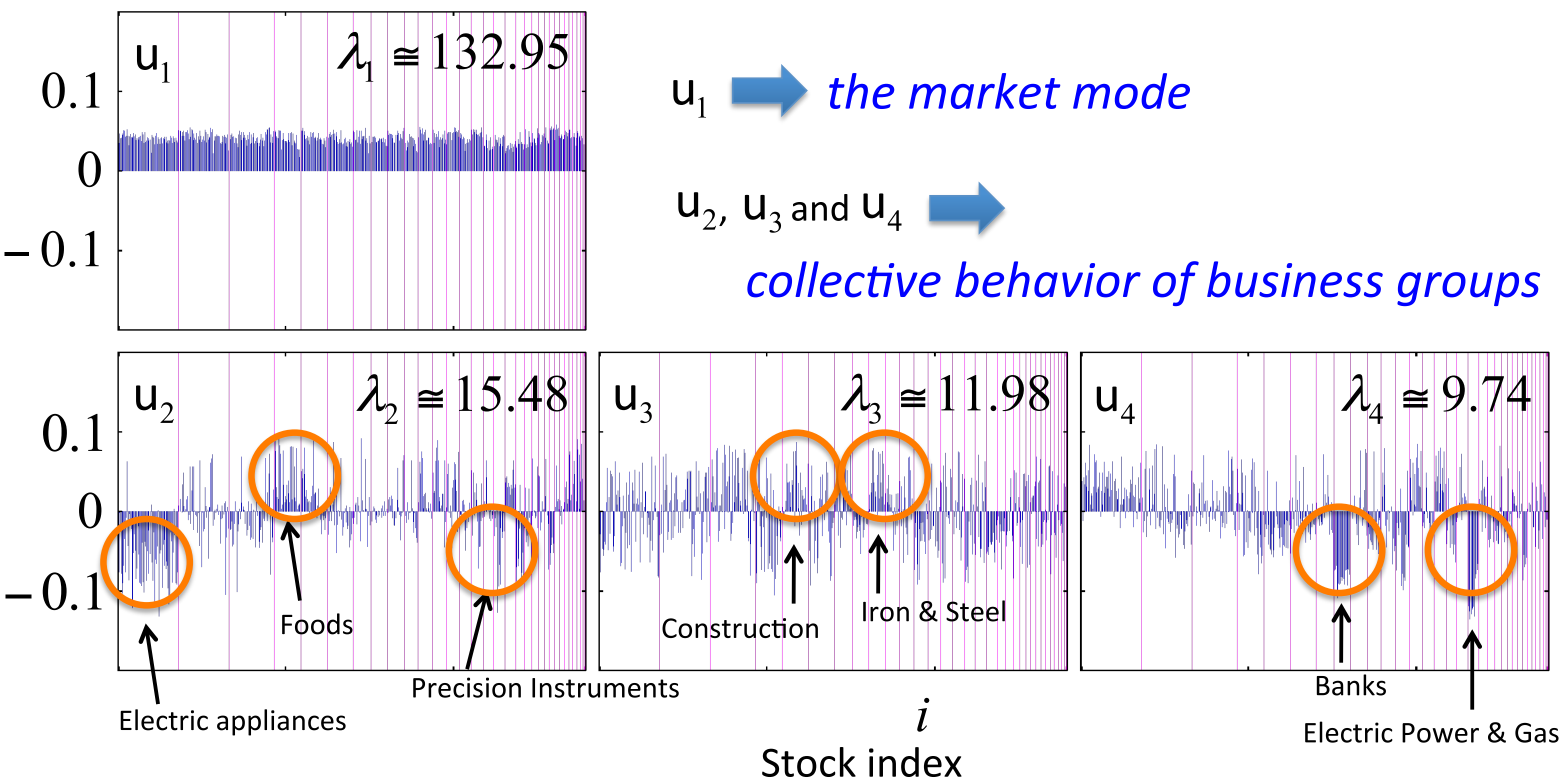
Market Mode in S&P 500



Business Group Behavior in S&P 500



Collective Behavior in the Tokyo Stock Exchange



Utsugi, A., Ino, K., Oshikawa, M., Phys. Rev. E **70**, 026110 (2004).

Group Correlations

The correlation matrix may be decomposed into

$$\mathbf{C} = \mathbf{C}_{\text{market}} + \mathbf{C}_{\text{group}} + \mathbf{C}_{\text{random}}$$

$$= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathrm{T}} + \sum_{i=2}^{N_{\mathrm{C}}} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} + \sum_{i=N_{\mathrm{C}}}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

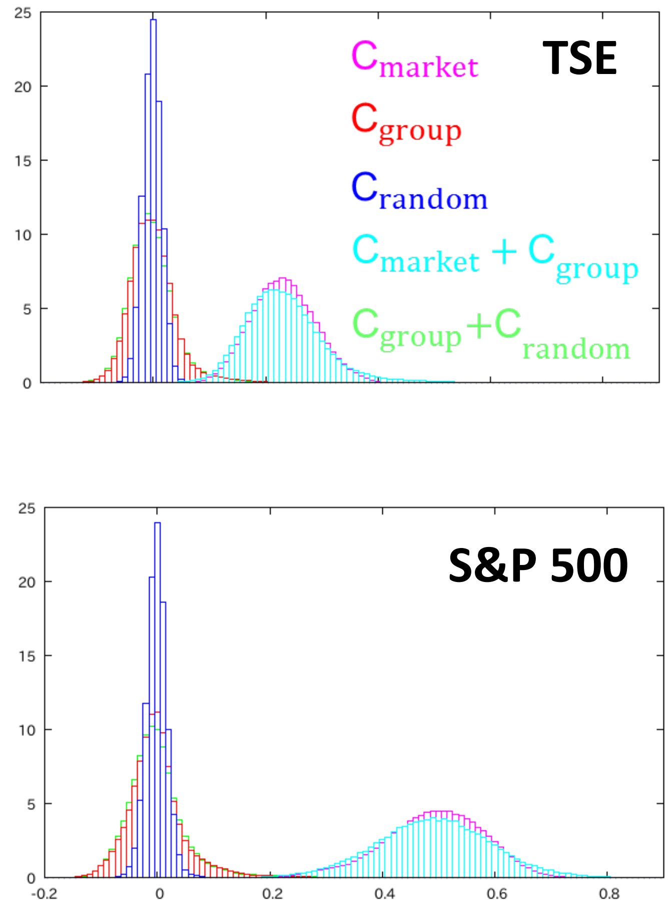
$$\lambda_{N_{\mathrm{C}}} > \lambda_+ > \lambda_{N_{\mathrm{C}}+1}$$

N_{C} : number of the meaningful principal components

The intrinsic market structure may be reflected in

$$\mathbf{C}_{\text{group}} = \sum_{i=2}^{N_{\mathrm{C}}} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

Distributions of Correlation Coeffcients



Network Approach to Financial Markets

Hierarchical structure in financial markets

R.N. Mantegna[a]

Eur. Phys. J. B **11**, 193–197 (1999)

Istituto Nazionale per la Fisica della Materia, Unità di Palermo, 90128, Palermo, Italy
Dipartimento di Energetica ed Applicazioni di Fisica, Università di Palermo, Viale delle Scienze, 90128, Palermo, Italy

Received 24 March 1999 and Received in final form 28 June 1999

Abstract. I find a hierarchical arrangement of stocks traded in a financial market by investigating the daily time series of the logarithm of stock price. The topological space is a subdominant ultrametric space associated with a graph connecting the stocks of the portfolio analyzed. The graph is obtained starting from the matrix of correlation coefficient computed between all pairs of stocks of the portfolio by considering the synchronous time evolution of the difference of the logarithm of daily stock price. The hierarchical tree of the subdominant ultrametric space associated with the graph provides a meaningful economic taxonomy.

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}$$

$$d(i,j) = \sqrt{2\,(1-\rho_{ij})}$$

$$-1 \leqslant \rho_{ij} \leqslant 1 \quad \Rightarrow \quad 2 \geqslant d(i,j) \geqslant 0$$

Network Approach to Financial Markets

Hierarchical structure in financial markets

R.N. Mantegna[a]

Eur. Phys. J. B **11**, 193–197 (1999)

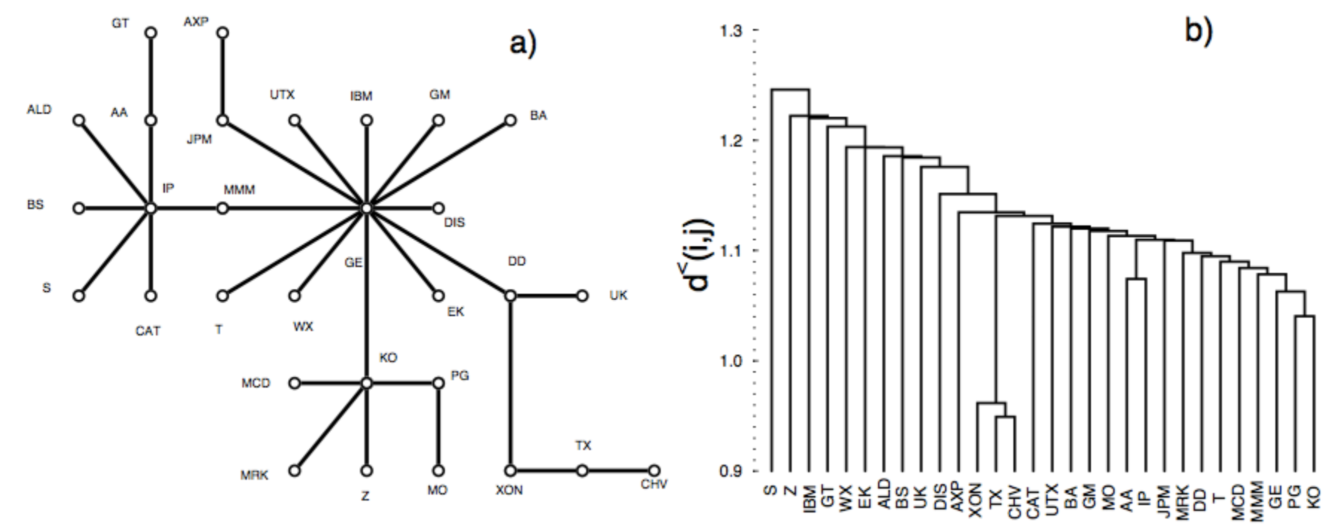
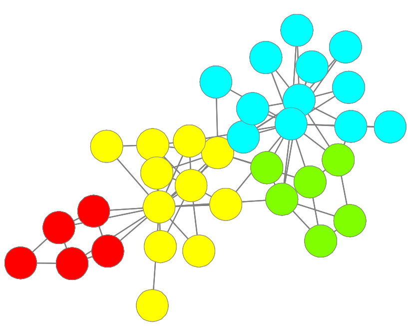


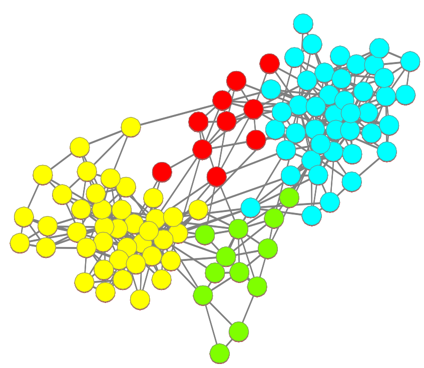
Fig. 1. (a) Minimal spanning tree connecting the 30 stocks used to compute the Dow Jones Industrial Average. The 30 stocks are labeled by their tick symbols. The distance between the stocks is bounded as: CHV-TX $0.90 < d(i,j) \le 0.95$; XON–TX $0.95 < d(i,j) \le 1.00$; KO-PG $1.00 < d(i,j) \le 1.05$; MMM–GE–KO, DD–GE–T, AA–IP and MRK–KO–MCD $1.05 < d(i,j) \le 1.10$; CAT–IP–MMM, AXP–JPM–GE–GM, BA–GE–UTX, DD–XON and MO–PG $1.10 < d(i,j) \le 1.15$; DIS–GE–EK, DD–UK, BS–IP–ALD and GE–WX $1.15 < d(i,j) \le 1.20$; AA–GT, GE–IBM, KO–Z and IP–S $1.20 < d(i,j) \le 1.25$. (b) Hierarchical tree of the subdominant ultrametric space associated with the minimal spanning tree of a). In the hierarchical tree, several groups of stocks homogeneous with respect to the economic activities of the companies are detected: (i) oil companies (Exxon (XON), Texaco (TX) and Chevron (CHV)); (ii) raw material companies (Alcoa (AA) and International paper (IP)) and (iii) companies working in the sectors of consumer nondurable products (Procter & Gamble (PG)) and food and drinks (Coca Cola (KO)). The ultrametric distance at which individual stocks are branching from the tree is given by the y axis.

Communities in Networks

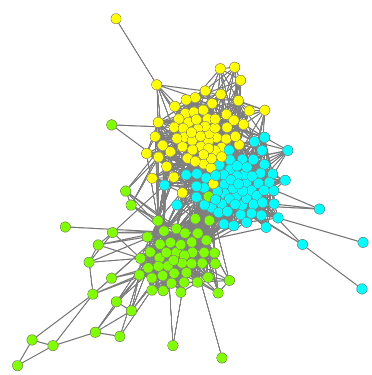
Community: a group of nodes which are tightly connected to each other through links.



Karate club

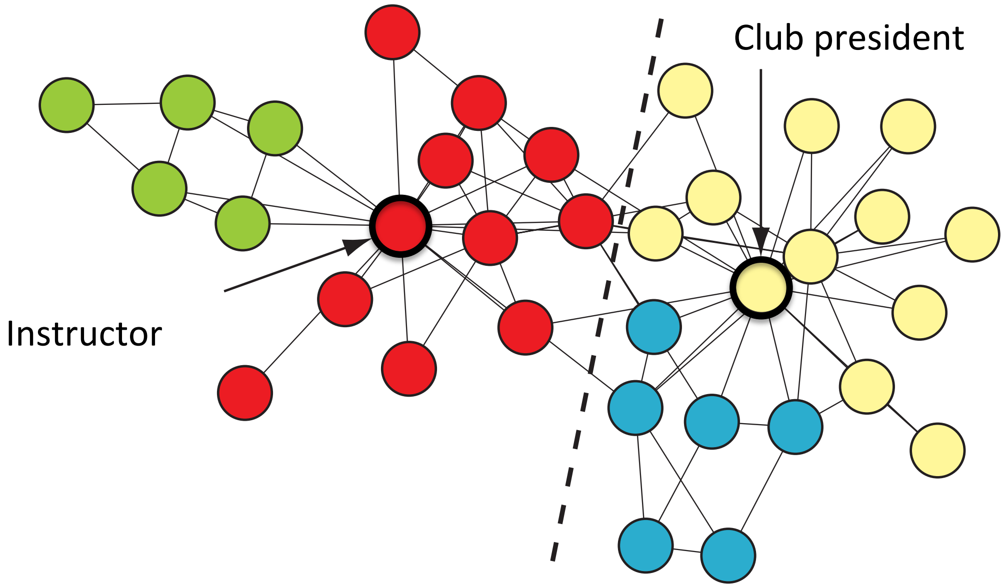


Amazon.com



Jazz club

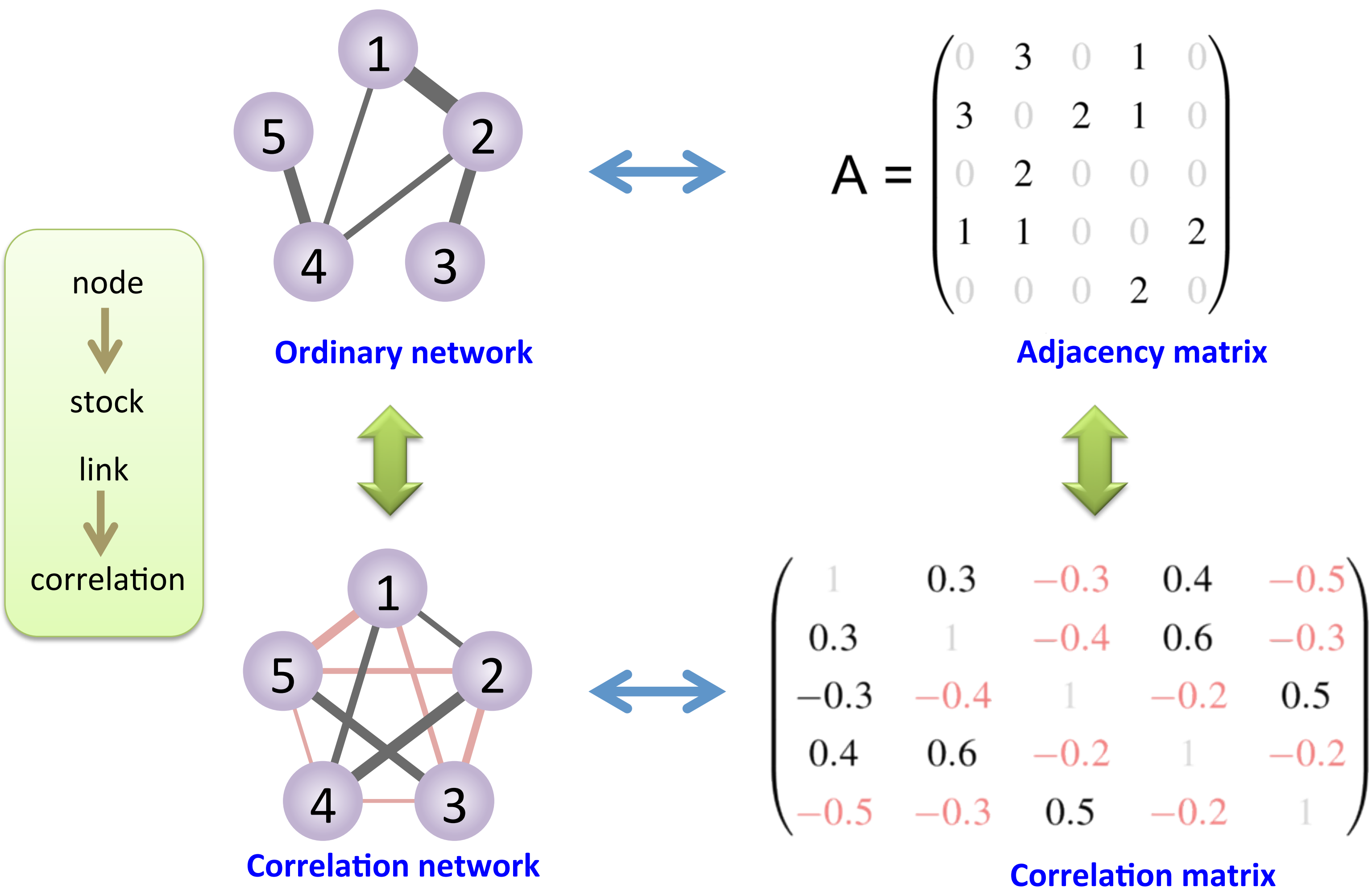
Power of the Community Detection



Four communities are obtained through the modularity maximization for the Karate Club network [1], and the dashed line depicts its actual split into two groups.

[1] W. W. Zachary: An information flow model for conflict and fission in small groups, J. Anthropological Res., Vol. 33, pp. 452-473, 1977.

Construction of a Stock Correlation Network



Detection of Stock Communities

◎ How to define a community in a stock correlation network?

Group of nodes (stocks) which are (positively) correlated to each other

◎ How to detect the optimized community structure

Maximization of the total sum of weights (minimization of the degree of frustration) inside communities

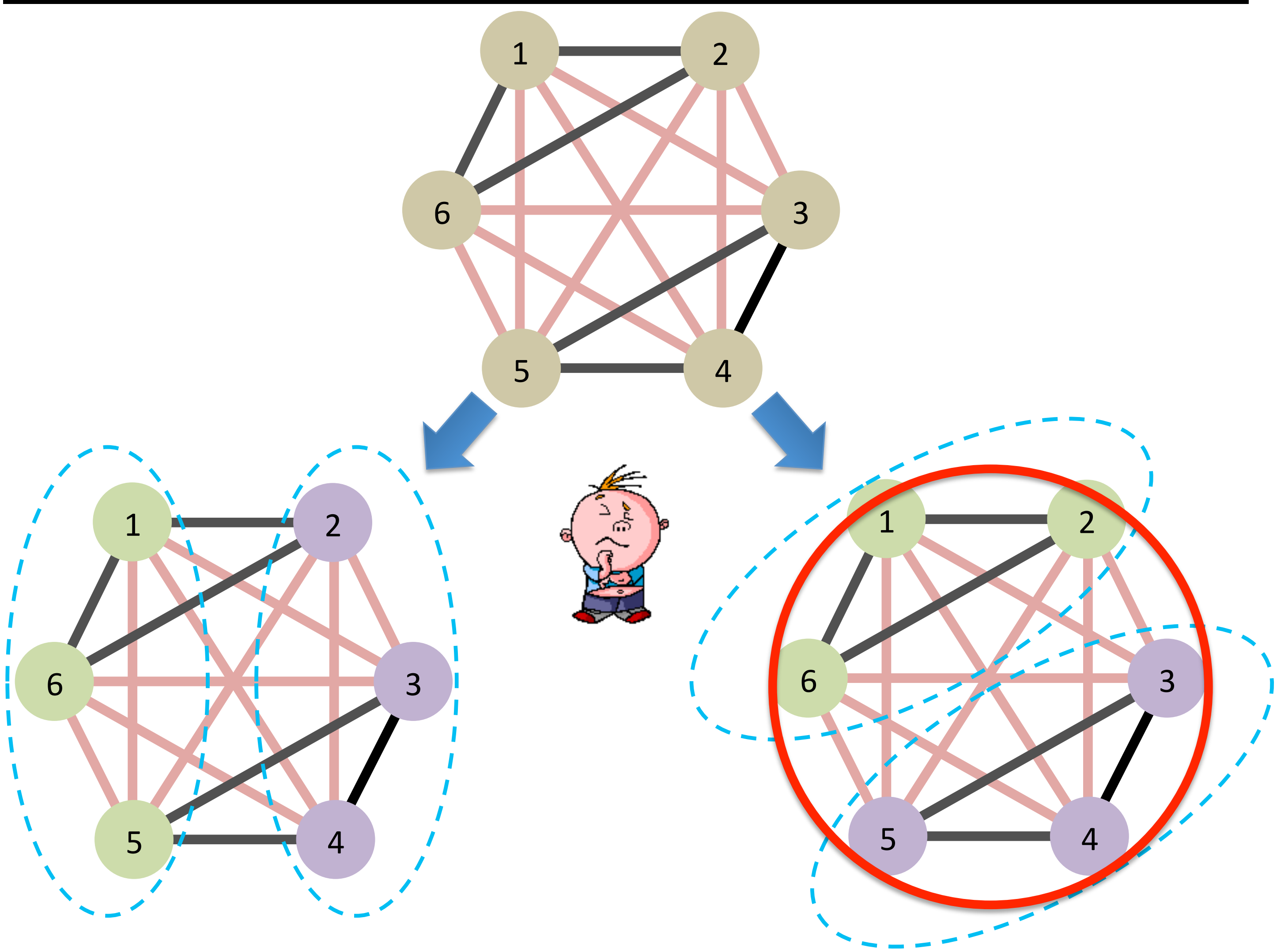
simulated annealing

Identification of comoving groups, which are anticorrelated to each other

◎ Measurement of the contribution of each principal component

$$\mathbf{C}_{\mathrm{group}}^{(l)} \equiv \sum_{i=2}^{l} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \qquad \begin{cases} l = 2,3,\ldots,13 & \text{TSE} \\ l = 2,3,\ldots,9 & \text{S\&P500} \end{cases}$$

Illustrative Example of the Community Detection



Detected Community Structure for the TSE

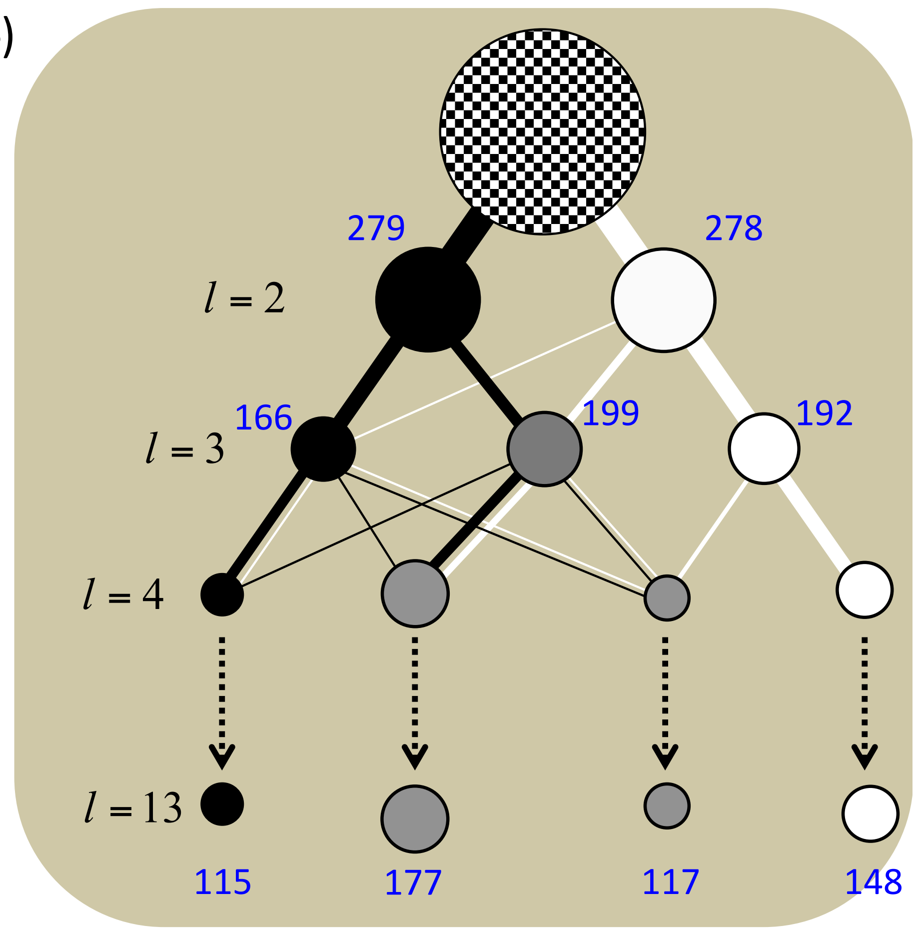
$$C_{\text{group}}^{(l)} = \sum_{i=2}^{l} \lambda_i u_i u_i^{T} \quad (l = 2,3,\ldots,13)$$

$l = 2$
perfectly separated into two groups

$l = 3$
$\lambda_3 u_3 u_3^T$ decomposes the two groups into three

$l = 4$
inclusion of $\lambda_4 u_4 u_4^T$ further decomposes the three groups into four

$l \geq 4$
four communities remain in the almost same way

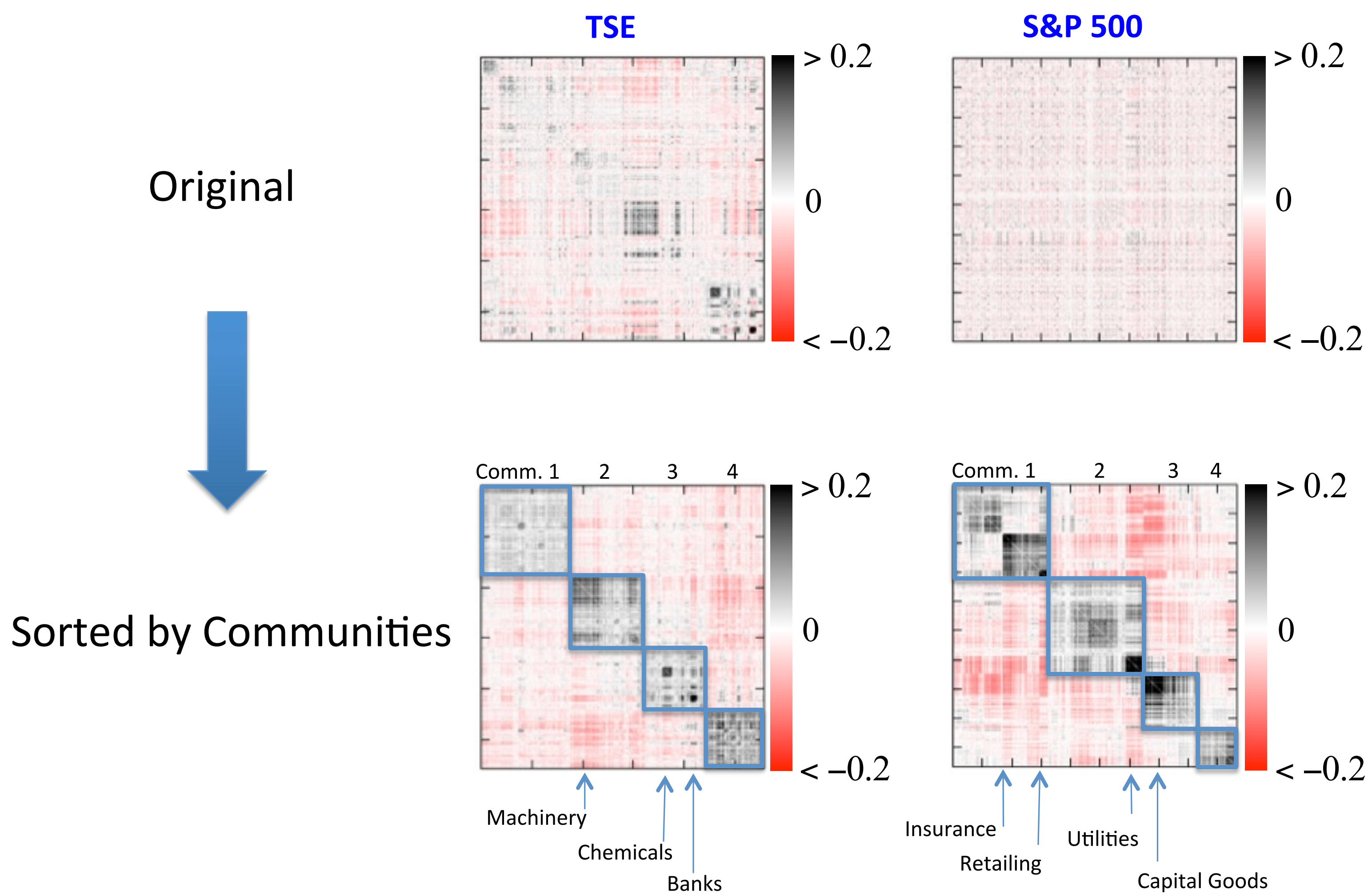


Detected Community Structure for S&P 500

l	Comm. 1	Comm. 2	Comm. 3	Comm. 4
2	247	236	—	—
3	197	161	125	—
4	184	159	140	—
5	183	162	138	—
6	171	161	151	—
7	170	161	152	—
8	168	163	152	—
9	166	157	93	67
$C_{\mathrm{group}} + C_{\mathrm{random}}$	166	157	93	67

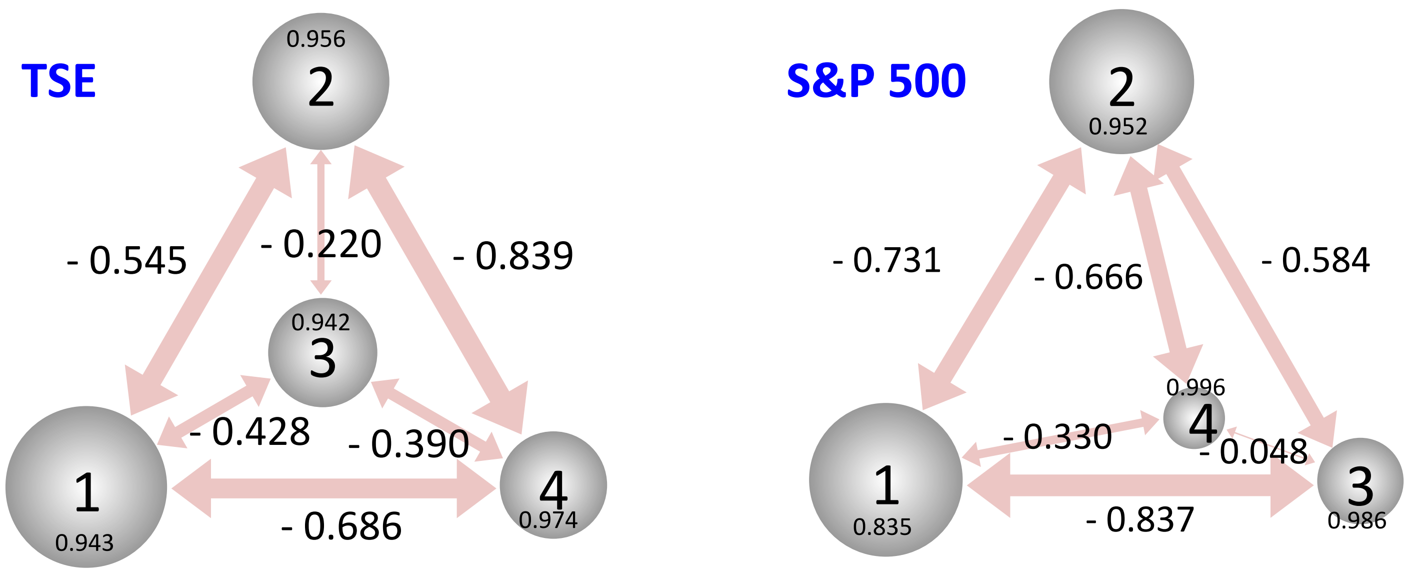
Almost identical community structure

Group Correlation Matrix Sorted by Communities



Polarization of Links

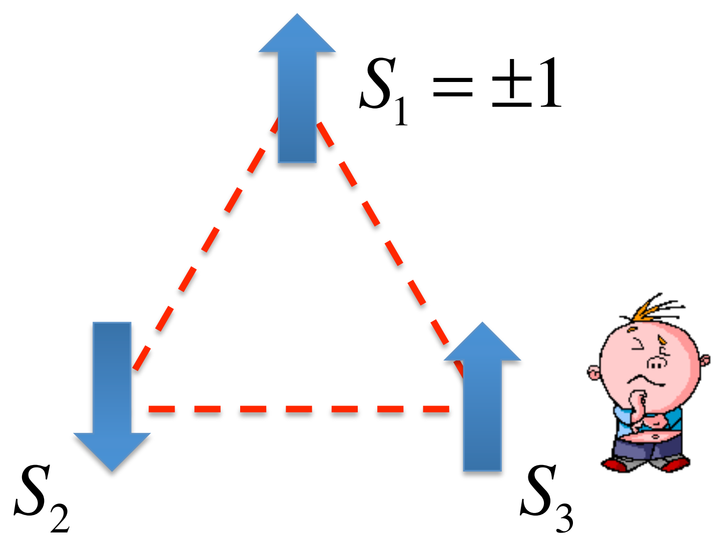
$$\text{Polarization ratio: } P_{\text{corr}} \equiv \frac{\sum_{\text{intra/inter}} A_{ij}}{\sum_{\text{intra/inter}} \left| A_{ij} \right|}$$



- TSE: three groups strongly anticorrelated to each other (comm1, comm2, comm4) + one group relatively neutral to the others (comm3)
- S&P 500: two tripolar frustration structures (comm1, comm2, comm3/comm4)

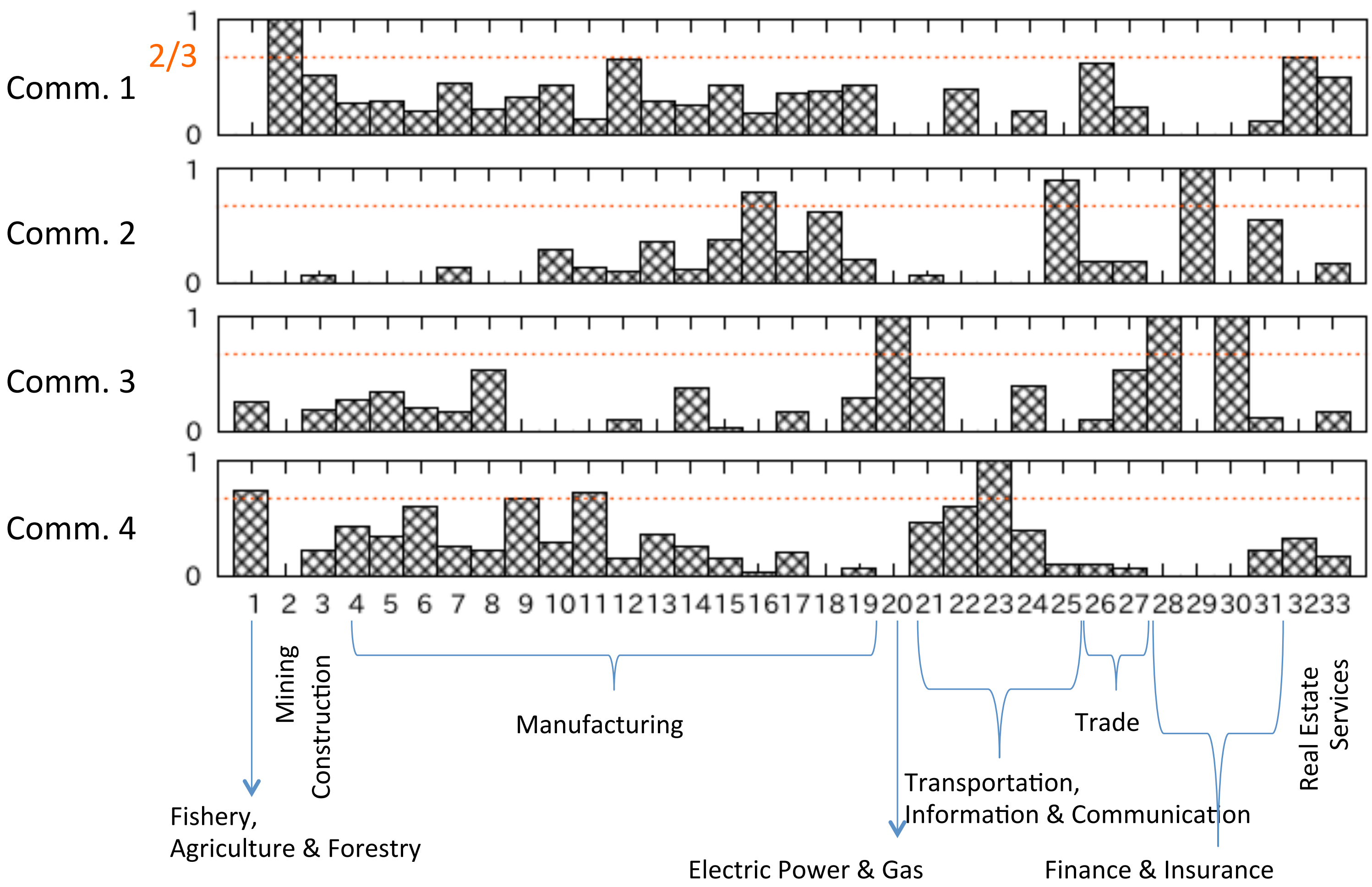
Antiferromagnetic Ternary Spin System

$$H = -J\left(S_1 S_2 + S_2 S_3 + S_3 S_1\right), \quad J < 0$$

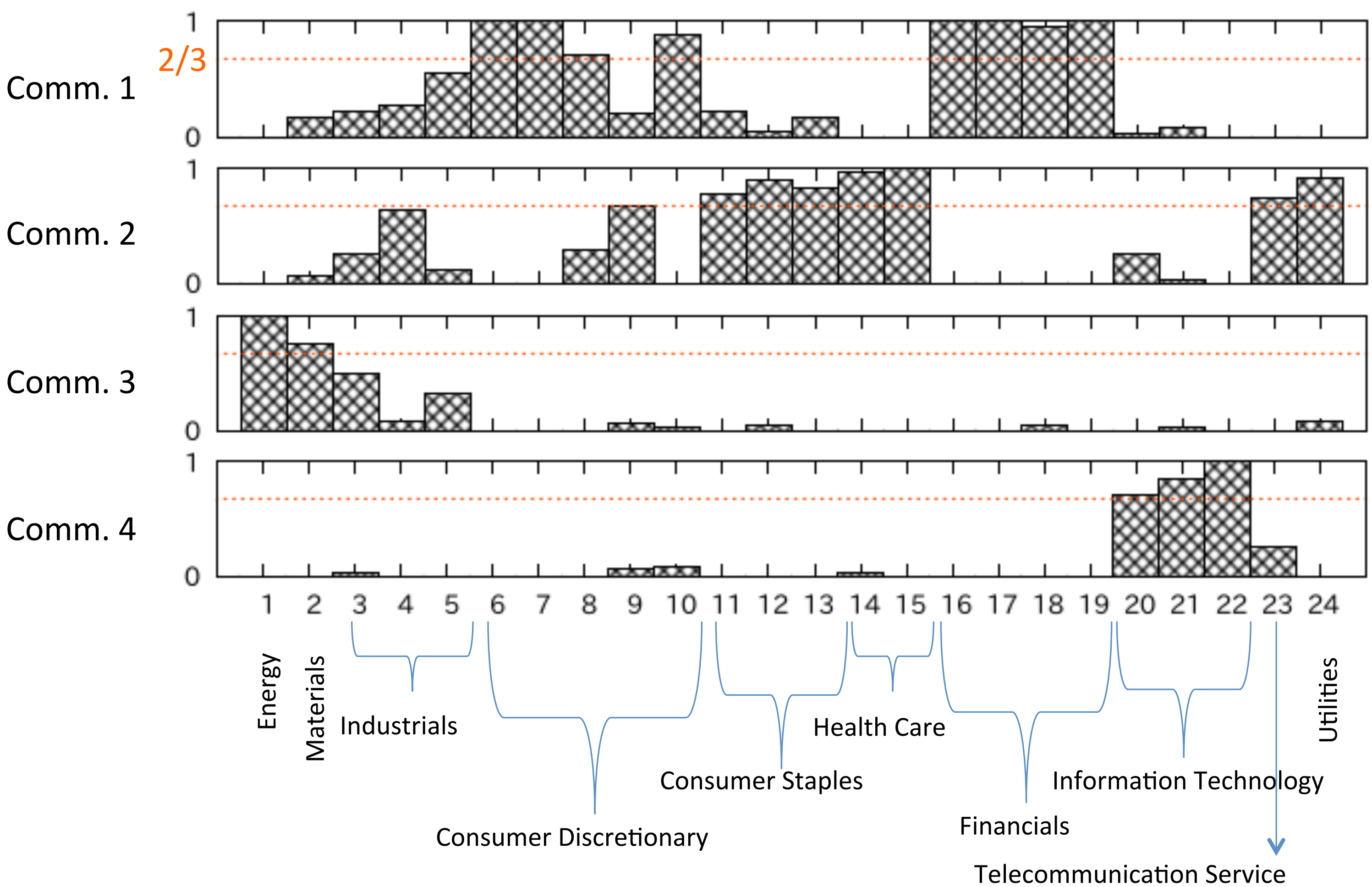


$$\left\langle S_1 S_2 \right\rangle = \frac{e^{\beta J} - e^{-3\beta J}}{3e^{\beta J} + e^{-3\beta J}} \quad \underset{T \to 0}{\Rightarrow} \quad -\frac{1}{3}$$

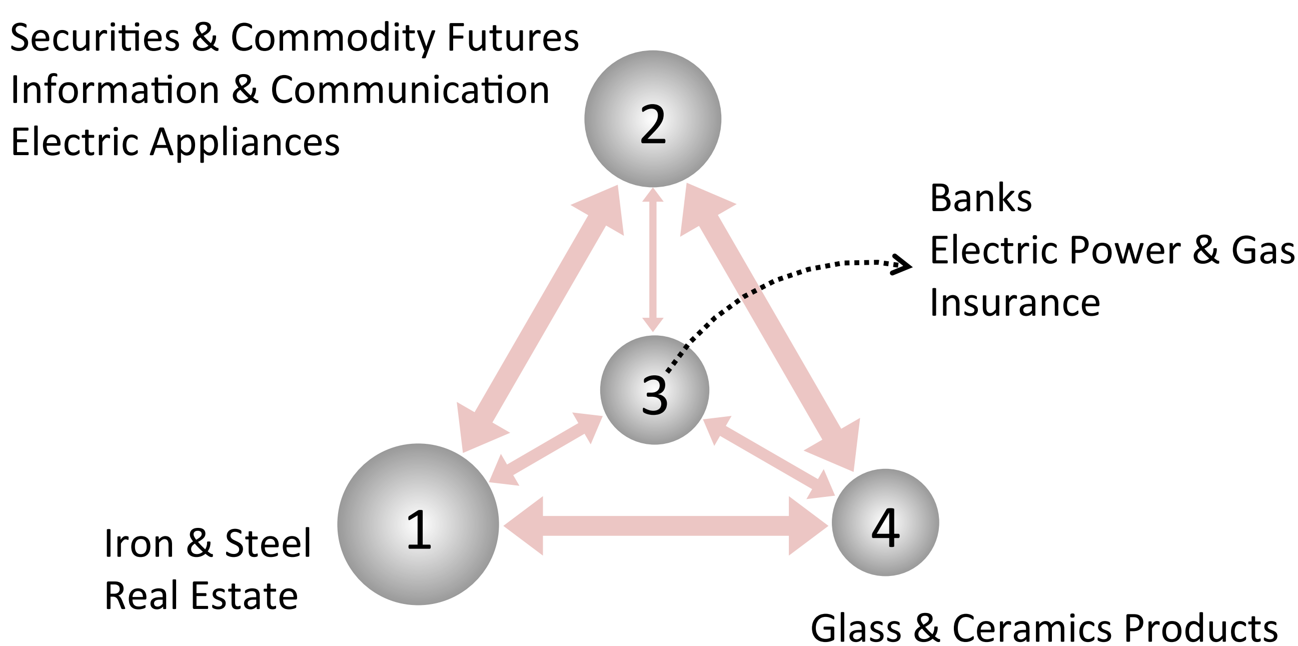
Industrial Groups in Communities: TSE



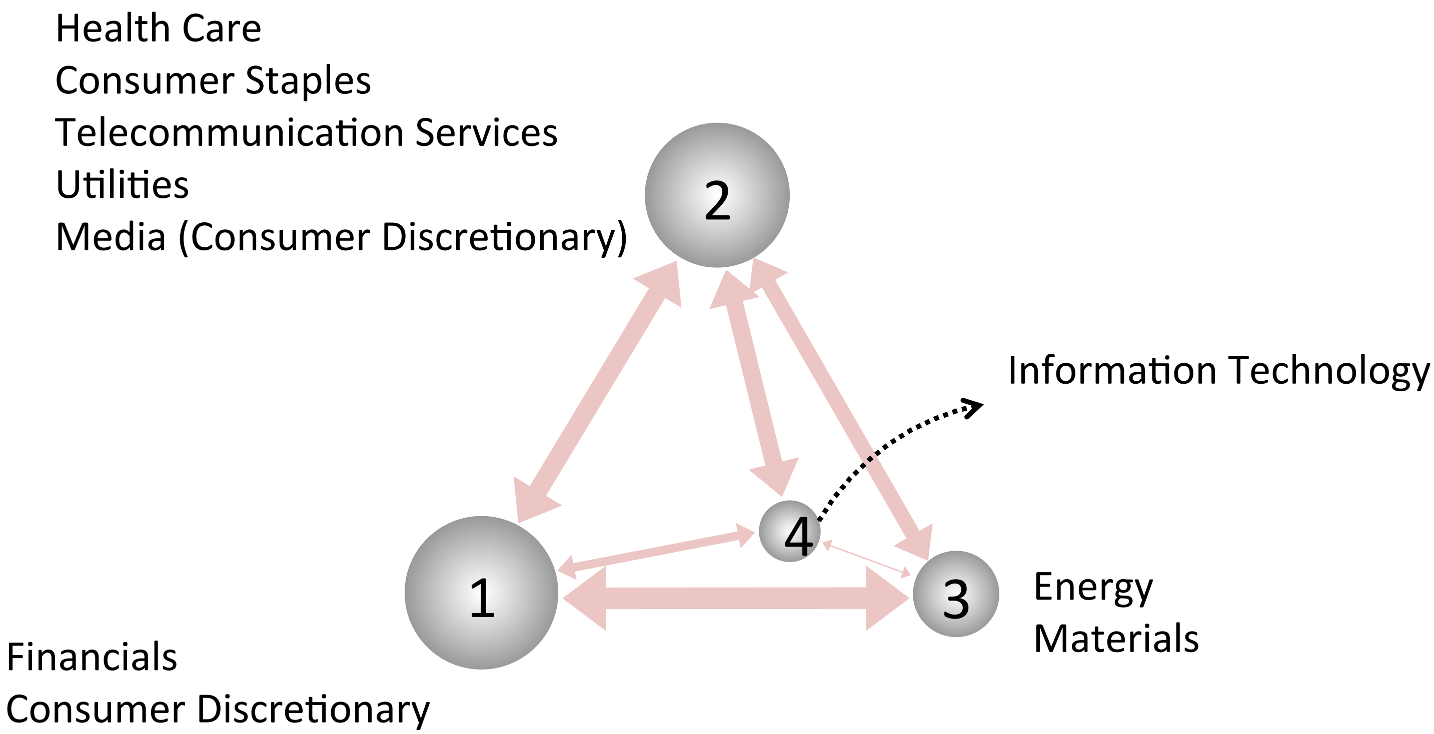
Industrial Groups in Communities: S&P 500



Frustration Structure Embedded in TSE

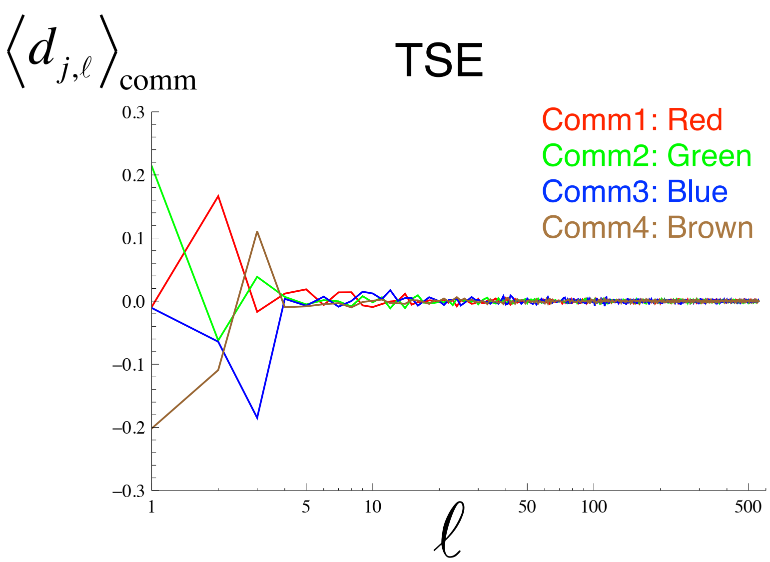


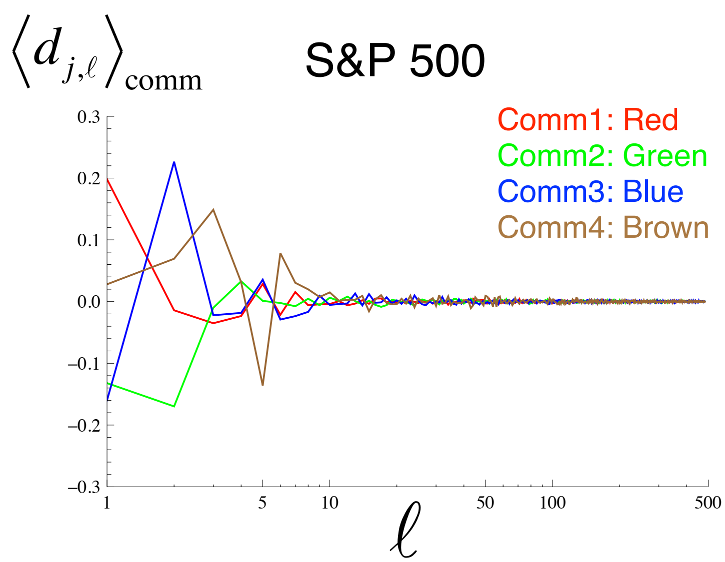
Frustration Structure Embedded in S&P 500



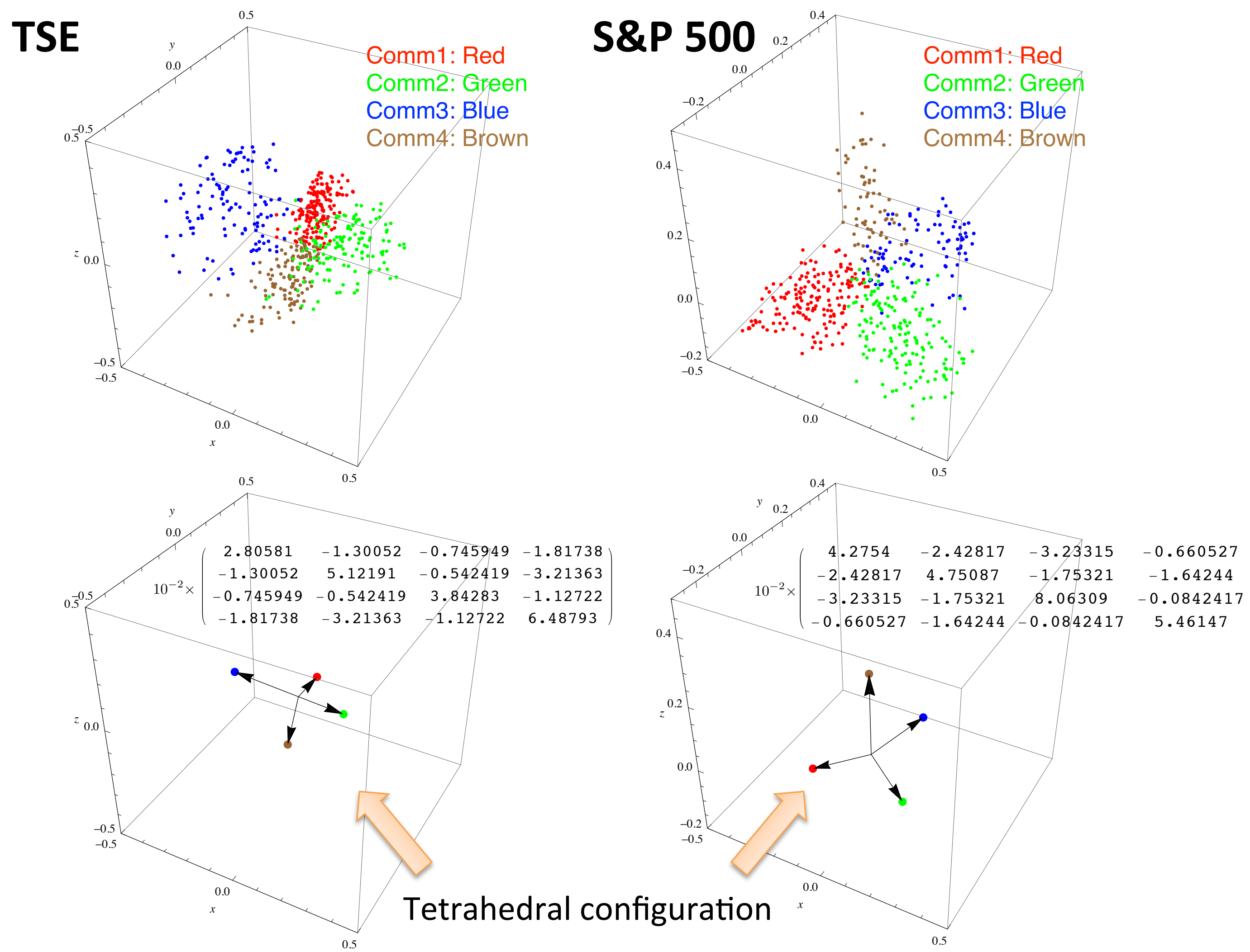
Correlation State Space

$$\mathbf{d}_j \equiv \begin{bmatrix} \sqrt{\lambda_2}\, u_{2,j} \\ \sqrt{\lambda_3}\, u_{3,j} \\ \vdots \\ \sqrt{\lambda_{N_C}}\, u_{N_C,j} \end{bmatrix} \Rightarrow \left[\mathbf{C}_{\text{group}}\right]_{ij} = \mathbf{d}_i \cdot \mathbf{d}_j$$





Distribution of Stocks in 3D Correlation State Space



$$10^{-2}\times\begin{pmatrix} 2.80581 & -1.30052 & -0.745949 & -1.81738 \\ -1.30052 & 5.12191 & -0.542419 & -3.21363 \\ -0.745949 & -0.542419 & 3.84283 & -1.12722 \\ -1.81738 & -3.21363 & -1.12722 & 6.48793 \end{pmatrix}$$

$$10^{-2}\times\begin{pmatrix} 4.2754 & -2.42817 & -3.23315 & -0.660527 \\ -2.42817 & 4.75087 & -1.75321 & -1.64244 \\ -3.23315 & -1.75321 & 8.06309 & -0.0842417 \\ -0.660527 & -1.64244 & -0.0842417 & 5.46147 \end{pmatrix}$$

Summary

We revisited correlation structure in the TSE and S&P 500 markets from a network theoretic point of view.

- The correlation matrix of stock price changes, purified by the RMT, was used to construct a network
- Communities were detected through minimization of the frustration inside communities.
- We found frustrated multipolar structures embedded in the financial markets

Such a hidden structure may give rise to complicated market behavior.

> More comparative study: application to Korean and other markets
> Temporal variation of the frustration structure; relationship to the market mode, especially market crush
> Relationship to real economy?

Thank you for your attention!