OCS

Heterogeneity and specialization of investors in financial markets

Rosario Nunzio Mantegna

Palermo University, Italy

Observatory of complex systems

OCS

Representative agent in economic theory

The most basic idealized assumption used in economic theory.

In mainstream economics, the economic actors are described in terms of a representative agent, which:

- is fully rational;

- has access to all available information;

- is able to process all information instantly and without errors.

Full individual rationality is observed in so-called Marshallian models

OCS

The representative agent

Journal of Economic Perspectives—Volume 6, Number 2—Spring 1992—Pages 117–136

Whom or What Does the Representative Individual Represent?

Alan P. Kirman

Paradoxically, the sort of macroeconomic models which claim to give a picture of economic reality (albeit a simplified picture) have almost no activity which needs such coordination. This is because typically they assume that the choices of all the diverse agents in one sector—consumers for example—can be considered as the choices of one "representative" standard utility maximizing individual whose choices coincide with the aggregate choices of the heterogeneous individuals.

My basic point in this paper is to explain that this reduction of the behavior of a group of heterogeneous agents *even if they are all themselves utility maximizers*, is not simply an analytical convenience as often explained, but is both unjustified and leads to conclusions which are usually misleading and often wrong.

OCS

5. THE INFORMATIONAL ROLE OF PRICES; CONCLUSIONS

Prices are often referred to as signals. However, in non-stochastic economies, they clearly play no formal role in transferring information. No one learns anything from prices. People are constrained by prices (often in just the right way so that individual rationality is transformed into collective rationality); however, they are not informed by prices in the classical Walrasian or Marshallian models. Nevertheless, it is an old idea that prices contain information. Perhaps the clearest statement appears in Hayek (1945, p. 527):

> "We must look at the price system as . . . as mechanism for communicating information if we want to understand its real function . . . The most significant fact about this system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action . . . by a kind of symbol, only the most essential information is passed on . . . "

Hayek wrote the above in criticism of the planning literature of the 1940's. That literature, taking the mathematical models of Walras, and the welfare theorems of Lange, literally, assumed that the State could set prices in such a way as to induce an efficient allocation and also the income distribution desired by the Leaders of the State. (The

Sanford J Grossman, An introduction to the Theory of Rational Expectations Under Asymmetric Information, Review of Economic Studies (1981) XLVIII 541-559

OCS

On the Impossibility of Informationally Efficient Markets

By SANFORD J. GROSSMAN AND JOSEPH E. STIGLITZ*

If competitive equilibrium is defined as a situation in which prices are such that all arbitrage profits are eliminated, is it possible that a competitive economy always be in equilibrium? Clearly not, for then those who arbitrage make no (private) return from their (privately) costly activity. Hence the assumptions that all markets, including that for information, are always in equilibrium and always perfectly arbitraged are inconsistent when arbitrage is costly.

We propose here a model in which there is an equilibrium degree of disequilibrium: prices reflect the information of informed individuals (arbitrageurs) but only partially, so that those who expend resources to obtain information do receive compensation. How informative the price system is depends on the number of individuals who are informed; but the number of individuals who are informed is itself an endogenous variable in the model.

jectures concerning certain properties of the equilibrium. The remaining analytic sections of the paper are devoted to analyzing in detail an important example of our general model, in which our conjectures concerning the nature of the equilibrium can be shown to be correct. We conclude with a discussion of the implications of our approach and results, with particular emphasis on the relationship of our results to the literature on "efficient capital markets."

I. The Model

Our model can be viewed as an extension of the noisy rational expectations model introduced by Robert Lucas and applied to the study of information flows between traders by Jerry Green (1973); Grossman (1975, 1976, 1978); and Richard Kihlstrom and Leonard Mirman. There are two assets: a safe asset yielding a return R, and a risky

The American Economic Review 70, 393-408 (1980)

OCS

Conceptual challenges

Heterogeneity at the micro level

James J. Heckman

University of Chicago and American Bar Foundation

Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture

This paper summarizes the contributions of microeconometrics to economic knowledge. Four main themes are developed. (1) Microeconometricians developed new tools to respond to econometric problems raised by the analysis of the new sources of micro data produced after the Second World War. (2) Microeconometrics improved on aggregate time-series methods by building models that linked economic models for individuals to data on individual behavior. (3) An important empirical regularity detected by the field is the diversity and heterogeneity of behavior. This heterogeneity has profound consequences for economic theory and for econometric practice. (4) Microeconometrics has contributed substantially to the scientific evaluation of public policy.

The Journal of Political Economy, Vol. 109, No. 4 (Aug., 2001), pp. 673-74

OCS

Empirical evidence of market heterogeneity

OCS

Trading profiles in financial markets

Heterogeneity of traders is observed and hypothesized in different financial studies.

In microstructure studies at least three kind of traders are considered:

- informed traders,
- uninformed traders,
- dealer or market maker.

OCS

Institutional and individual investors

Finance studies also distinguish between

- institutional investors
and
-individual investors.

OCS

Momentum and contrarian strategies

Momentum investors are buying stocks that were past winners.

A contrarian strategy consists of buying stocks that have been losers (or selling short stocks that have been winners).

The contrarian strategy is formulated on the assumption that the stock market overreacts and a contrarian investor can exploit the inefficiency related to market overreaction by reverting stock prices to fundamental values.

Empirical investigations performed in different markets have shown that institutional investors are preferentially momentum investors whereas individual investors usually prefer a contrarian strategy.

OCS

The trading profile of institutional investors

They are large institutions moving a large amount of the market.

At the US equity markets in the eighties[¶] they were by large majority (77%) characterized by a momentum resulting strategies.

The size of the institutions imposes to trade orders as "packages" or "hidden orders" to minimize market impact[§].

[¶]M.Grinblatt et al, American Economic Review 85, 1088-1105 (1995)

[§]L.K.C. Chan et al, J. of Finance LI, 1681-1713 (1996)

OCS

Grinblatt, Titman and Wermers (1995)

They investigated the trading pattern of fund managers by examining the quarterly holdings of 155 mutual funds (information from CDA Investment Technologies and CRSP data) over the 1975-1984 period.

The large majority of funds (77%) had a momentum investment profile.

Authors found relatively weak evidence that funds tended to buy and sell the same stocks at the same time (herding).

¶M.Grinblatt et al, American Economic Review 85, 1088-1105 (1995)

OCS

Chan and Lakonishok (1995)

The trading of "packages" or "hidden orders" executed by institutional investors were studied by investigating all trades executed by 37 large investment management firms from July 1986 to December 1988. These data were collected by SEI Corp., a consulting organization in the area of financial services.

It is worth noting that they use manager's trading history to reconstruct the manager's trading packages in each stock. The markets considered are NYSE and AMEX and trades cover about 5 percent of the total value of trading of the two exchanges over this period.

[§]L.K.C. Chan et al, J. of Finance LI, 1681-1713 (1996)

OCS

A more recent empirical study on “packages”

The trading of “packages” or “hidden orders” executed by institutional investors has also been studied in the Australian Stock Exchange¶.

Authors have investigated the daily investment of 34 active Australian equity managers. Data were provided directly by the managers and includes transaction in equity stocks, futures contracts and options securities. The sample period is from Jan 2, 1995 to Dec 31, 2001. From the information provided authors extract information about “packages” traded by the fund managers.

¶D.R.Gallagher and A. Looi, Accounting and Finance 46, 125-147 (2006).

OCS

Foreign and domestic investors

A study¶ about positive feedback trading and herding by foreign investors was performed by considering Korean market in the period before and at the Korea's economic crisis of 1997.

Researchers observe strong evidence of positive feedback trading and herding by foreign investors before the period of Korea's economic crises. They also find that there is no evidence that herding is more important during the crisis period.

They classify investors in three groups: (i) Foreign investors; (ii) Korean institutional investors and (iii) Korean individual investors.

¶H. Choe et al, J. of Financial Economics 54, 227-264 (1999).

OCS

Choe, Kho and Stulz

Foreign investors are mainly positive feedback traders (momentum) and positive feedback is driven by individual stock returns.

Korean individual investors are mainly contrarian with respect to individual stock returns (but perhaps positive feedback traders with respect to market return).

Korean institutional investors are momentum traders for individual stocks and contrarian with respect to the market.

Authors use daily and intradaily data from
Dec 2, 1996 to Dec 27, 1997 of 414 stocks listed at the KSE

¶H. Choe et al, J. of Financial Economics 54, 227-264 (1999).

OCS

Data from Finland

A similar study was done by Grinblatt and Keloharju¶. They investigated the central register of shareholdings for Finnish Central Securities Depository, a comprehensive data source. This data set reports individual and institutional holdings and stock trades on a daily basis.

We have previously seen that holding of U.S. mutual funds and U.S. pension funds are typically performed by analyzing quarterly data.

Data consists of each owner's stock exchange trades from Dec 27, 1994 through Dec 30, 1996.

¶M.Grinblatt and M.Keloharju, J. of Financial Economics 55, 43-67 (2000)

OCS

Grinblatt and Keloharju

They find that:
- foreign investors tend to be momentum investors;
- individual investors tend to be contrarian;
- domestic institutional investor tend to present a mixed behavior.

A resulting strategy can therefore be associated with the investment profile of these three groups of investors.

OCS

Analysis of the Taiwan market

Studies performed by Barber, Lee, Liu and Odean¶ have the performance of individual and institutional investors at the Taiwan Stock Exchange. Both with respect to individual day traders and to portfolio selection.

Data allow authors to identify trades made by individuals and by institutions, which fall into one of four categories (corporations, dealers, foreigners, or mutual funds).

To analyze who gains and loses from trade, they construct portfolios that mimic the purchases and sales of each investor group during the time period 1995 to 1999.

¶ B.M.Barber, Y.-T.Lee, Y.-J.Liu and T.Odean, Do Individual Day Traders Make Money? Evidence from Taiwan (http://papers.ssrn.com 2004).

¶ B.M.Barber, Y.-T.Lee, Y.-J.Liu and T.Odean, Just How Much Do Individual Investors Lose by Trading? (http://papers.ssrn.com 2005).

OCS

Intraday data: a Nasdaq study

Griffin et al[§] study daily and intradaily cross-sectional relation between stock returns and the trading of individual and institutional investors in Nasdaq 100 securities.

They observe that most brokerage houses specialize in dealing with either institutional or individual clients.

Data consists of all the trades and quotes in Nasdaq 100 stocks from May 1, 2000 to February 28, 2001.

They have information to assign trading volume to brokerage houses:
-primarily dealing with individual investors;
-primarily dealing with institutional investors;
-acting as market makers.

[§]J.M.Griffin et al, J. of Finance LVIII, 2285-2320 (2003)

OCS

Griffin, Harris and Topaloglu

They observe that institutional trading largely follows past stock returns both at a daily and at an intradaily time horizon.

The reverse is not observed. Inventory variation does not predict stock return.

OCS

Conditional predictability

Empirical investigations performed by using proprietary trade data information obtained from the Korean Stock Exchange (Choe et al 1999) and from NASDAQ (Griffin et al 2003) shown that stock returns have some ability to forecast inventory variation of groups of investors whereas the evidence of return predictability on the basis of investor inventory variation is negligible both at a daily and intradaily time horizon.

OCS

An empirical study covering an entire market

OCS

Motivations

- Most of the studies performed in the econophysics research follow one of the two approaches:
  - Empirical studies of aggregated quantities, such as prices, volumes, volatility, etc.
  - Theoretical and/or numerical studies of agent based models attempting to reproduce the stylized facts of aggregated variables.

- Only in few cases an empirically based agent based investigation is possible due to difficulties in accessing data with transparent or coded agent's identity.

OCS

We present an empirical research providing a first base for an empirically grounded agent based model

This is possible by investigating the database of the Spanish Stock Market which is containing the transparent information on trading market members.

OCS

The Spanish Stock Market (BME)

Madrid 1831
Bilbao 1890
Barcelona 1915
Valencia 1970

⇒

CATS
Computer Assisted
Trading System
1980

⇒

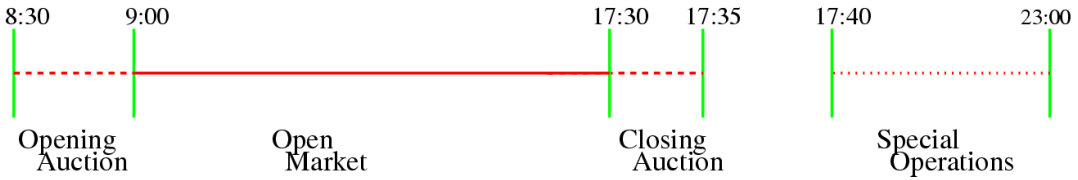
SIBE
Stock Exchange
Interconnection System
1995

⇒

IBEX 35

Trading Hours

- Market orders
- Market to limit orders
- Limit orders



Information dissemination (SIBE)

- Trades, with price, volume and counterparties of the trade
- Order book, with the five best buy/sell positions
- Index (IBEX 35, LATIBEX, Nuevo Mercado) information

In 2004 the BME was the eight in the world in market capitalization.

OCS

Market members

- Market members are credit entities and investment firms which are members of the stock exchange and are entitled to trade in the market.
- Approx 200 market members at the BME (350/250 at the NYSE)
- We only study approximately 80 because:
  - Not all the members trade during the whole period
  - We have only chosen those members whose activity is continuous

Snapshot of our database

VALOR	VOLUMEN	PRECIO	SOCCOM	SOCVEN	HORA	FECHA
TEF	236	2187	9405	9858	90108	01/06/2000
TEF	1764	2187	9405	9487	90108	01/06/2000
ANA	110	3800	9839	9855	90109	01/06/2000
CAN	37	2194	9839	9578	90109	01/06/2000
CAN	151	2200	9839	9412	90109	01/06/2000
VIS	214	700	9821	9561	90109	01/06/2000
SOL	286	1299	9839	9838	90110	01/06/2000
ALB	104	2710	9839	9843	90110	01/06/2000
ALB	29	2719	9839	9419	90110	01/06/2000
ACX	97	3689	9839	9843	90111	01/06/2000
AGS	120	1445	9839	9487	90111	01/06/2000
AGS	110	1448	9839	9485	90111	01/06/2000
ACS	107	2930	9839	9863	90111	01/06/2000
SCH	11226	1045	9858	9880	90112	01/06/2000
CTE	96	1935	9839	9832	90112	01/06/2000
CTE	50	1955	9839	9872	90112	01/06/2000
CTE	14	1958	9839	9426	90112	01/06/2000
FER	237	1296	9839	9560	90112	01/06/2000
SGC	50	3980	9820	9560	90113	01/06/2000
ACR	161	1139	9839	9487	90113	01/06/2000
ACR	47	1140	9839	9845	90113	01/06/2000
DRC	20	803	9839	9573	90114	01/06/2000
DRC	267	805	9839	9484	90114	01/06/2000
AUM	111	1649	9839	9474	90114	01/06/2000

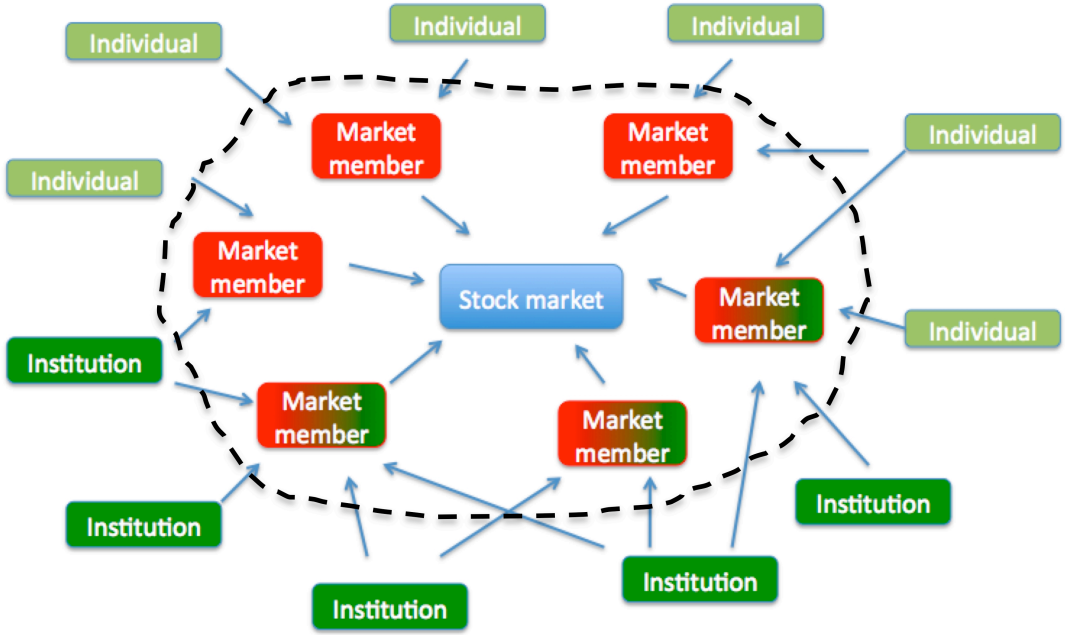
OCS

Market members vs agents

Market members (MMs) are not agents. A market member may act on behalf of many different agents.

This could be due either because a

MM acts as an intermediary or because a MM is doing client trading.



OCS

Data

We investigate 4 highly capitalized stocks:
Telefonica (TEF), Banco Bilbao Vizcaya Argentaria (BBVA), Banco Santander Central Hispano (SAN) and Repsol (REP)

The investigated period is 2001-2004

We investigate market dynamics by focusing on the trading of each selected stock separately for each available calendar year.

By doing so we have up to 4x4 distinct sets of results

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.

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

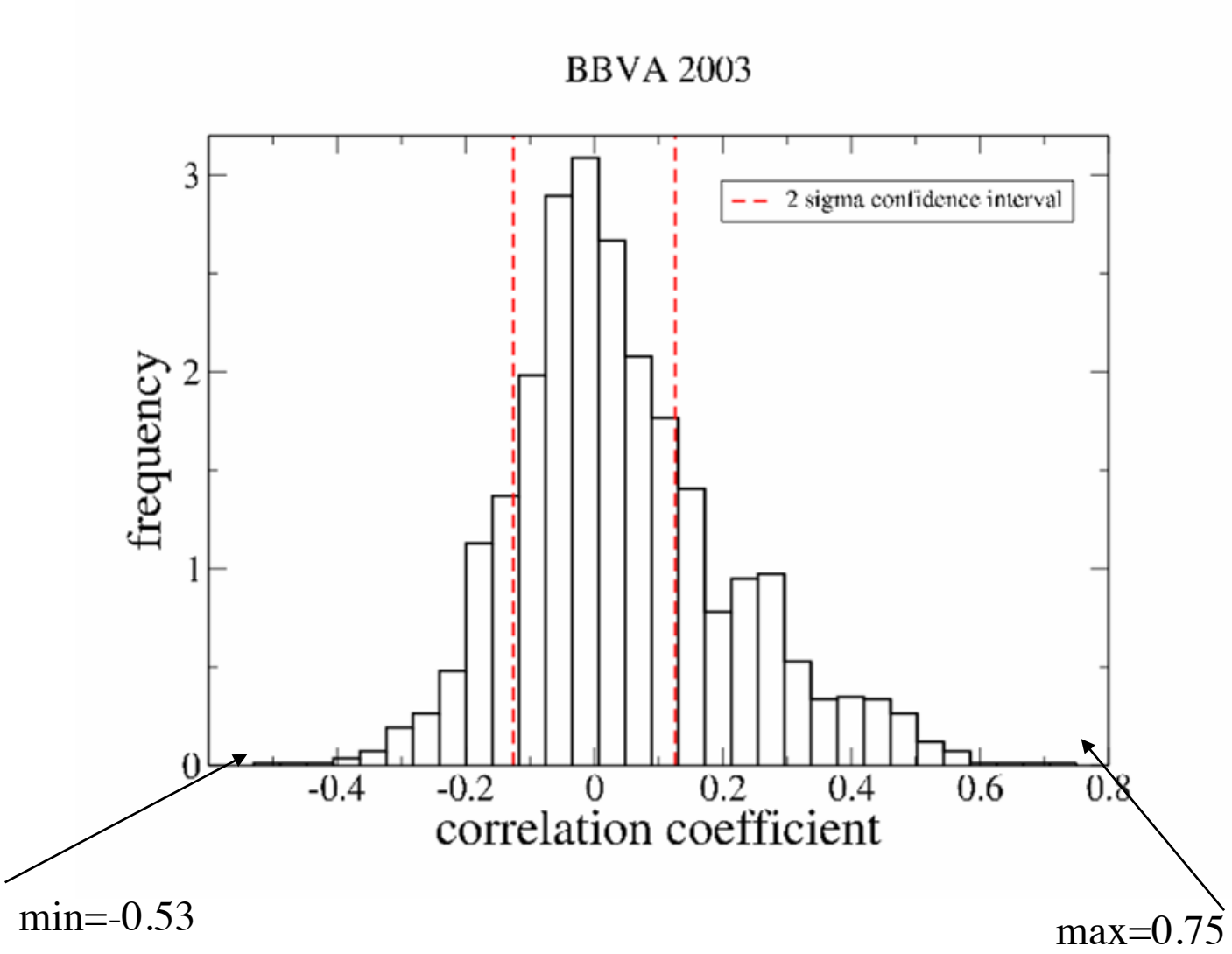
sign
+1 for buys
-1 for sells

price

volume

OCS

Correlation between MMs' inventory variation



OCS

Correlation matrix of MM inventory variation

Is the cross correlation matrix of MM inventory variation carrying information about the market dynamics?

Random effects can be tested by using Random Matrix Theory

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}$$

$$\lambda_{min}^{max} = \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}) \quad Q = T/N$$

OCS

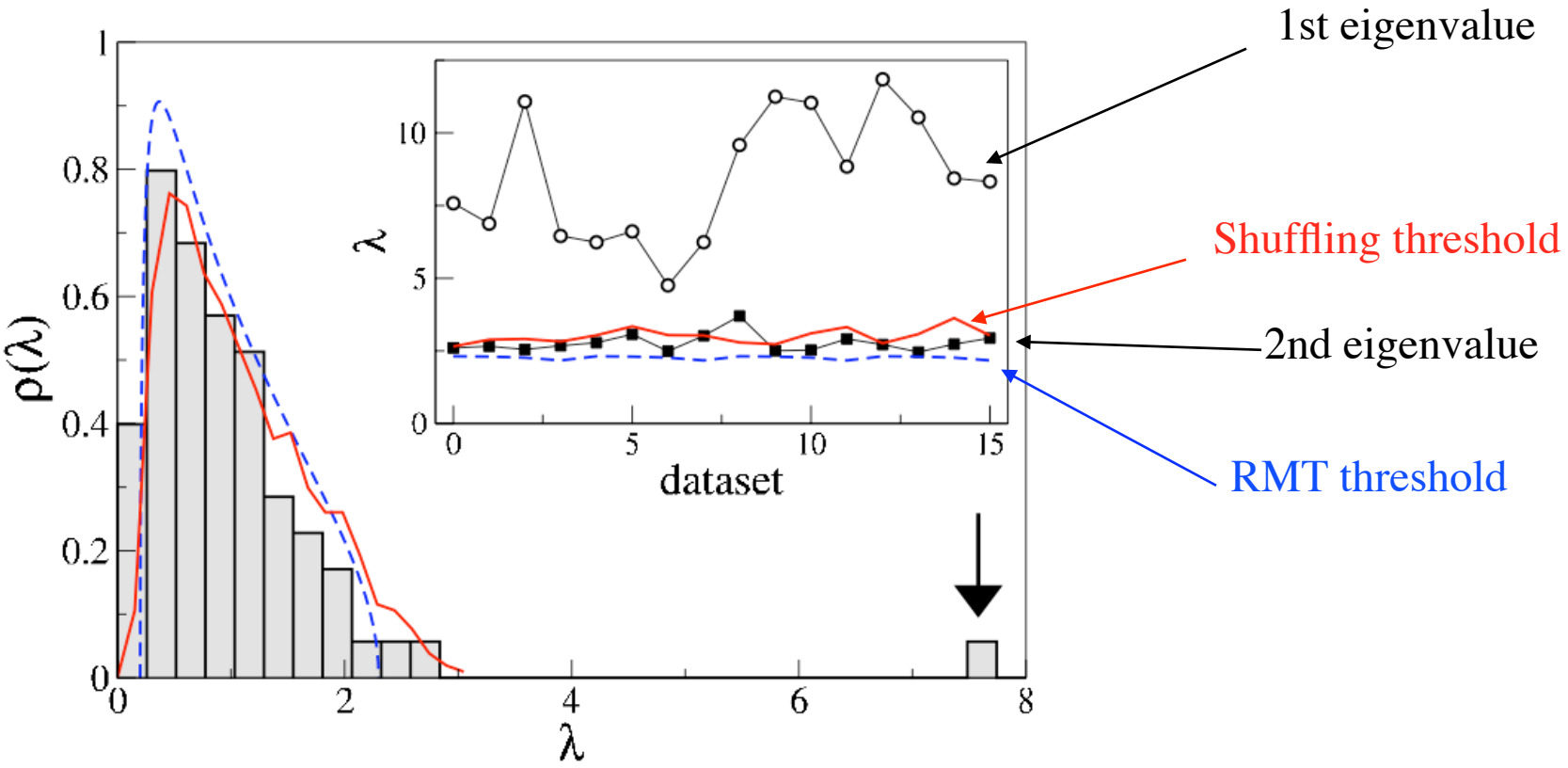
Potential spurious correlation

Other spurious correlation could come from the constraints posed by the trading itself.

In a trade the buyer MM increases the inventory and the seller MM decreases it of an equal amount leading to potential spurious correlation

To control for this effect we perform a series of shuffling experiments.

Eigenvalue spectrum



The first eigenvalue is not compatible with random trading and is therefore carrying information about the collective dynamics of firms.

OCS

Origin of collective behavior

- Which is the meaning of the largest eigenvalue of the correlation matrix of inventory variation?

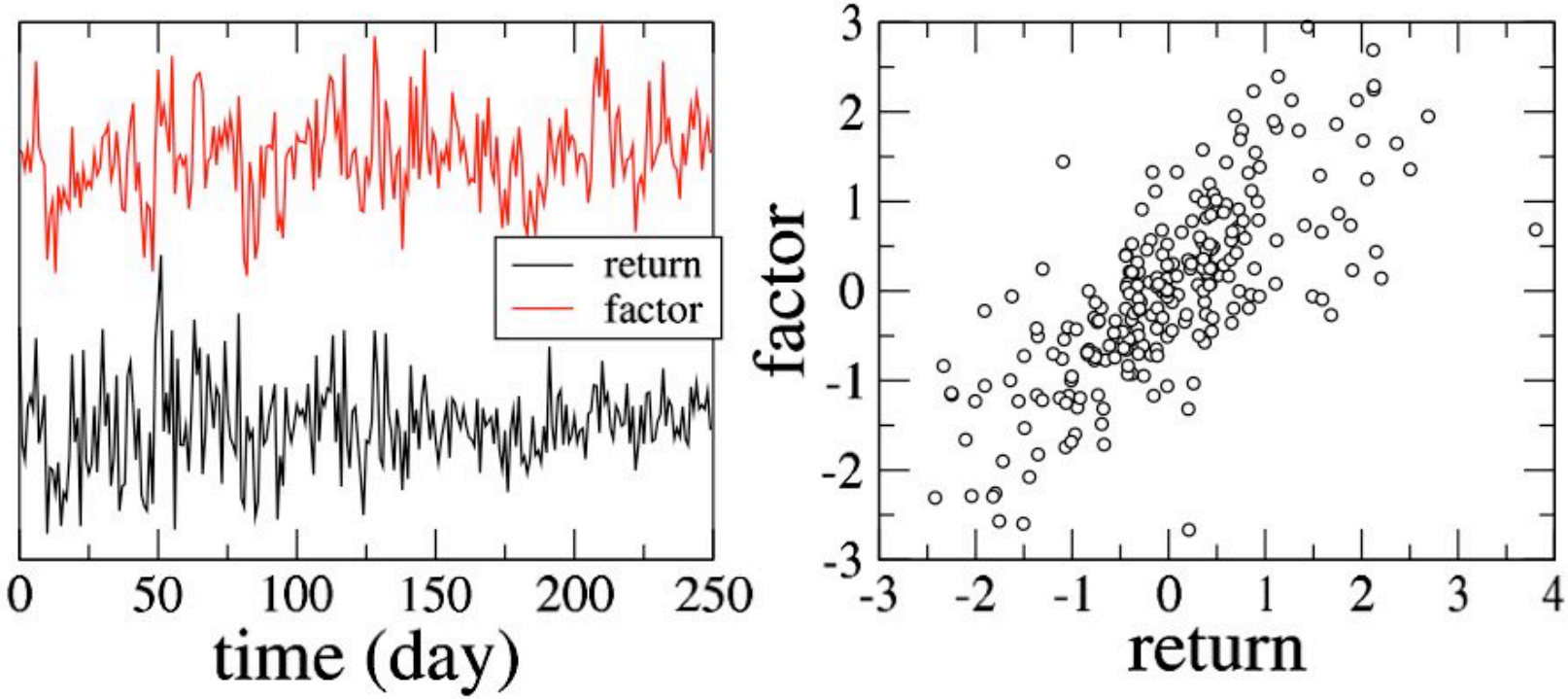
- Principal Factor Analysis suggests that there is a factor which is driving the inventory variation of many firms

- The presence of the collective behavior is not due to the fact that some firms are buying and other are selling (shuffling experiment)

- Rather it suggests that there are groups of firms having systematically the same position in the market as the other members of the group they belong to.

OCS

The factor driving inventory variation



Correlation between the factor and price return ranges between 0.47 and 0.74, being statistically significant at 99% confidence in all 16 sets

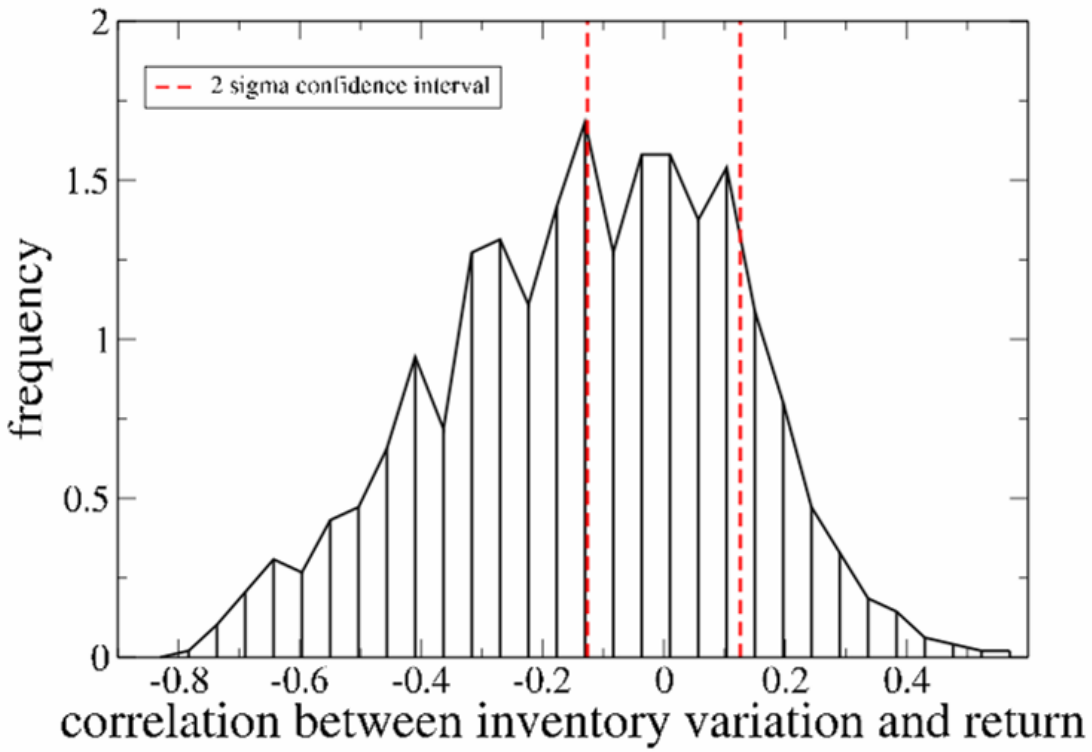
OCS

A taxonomy of market members

"reversing" MMs (ex: contrarian traders)

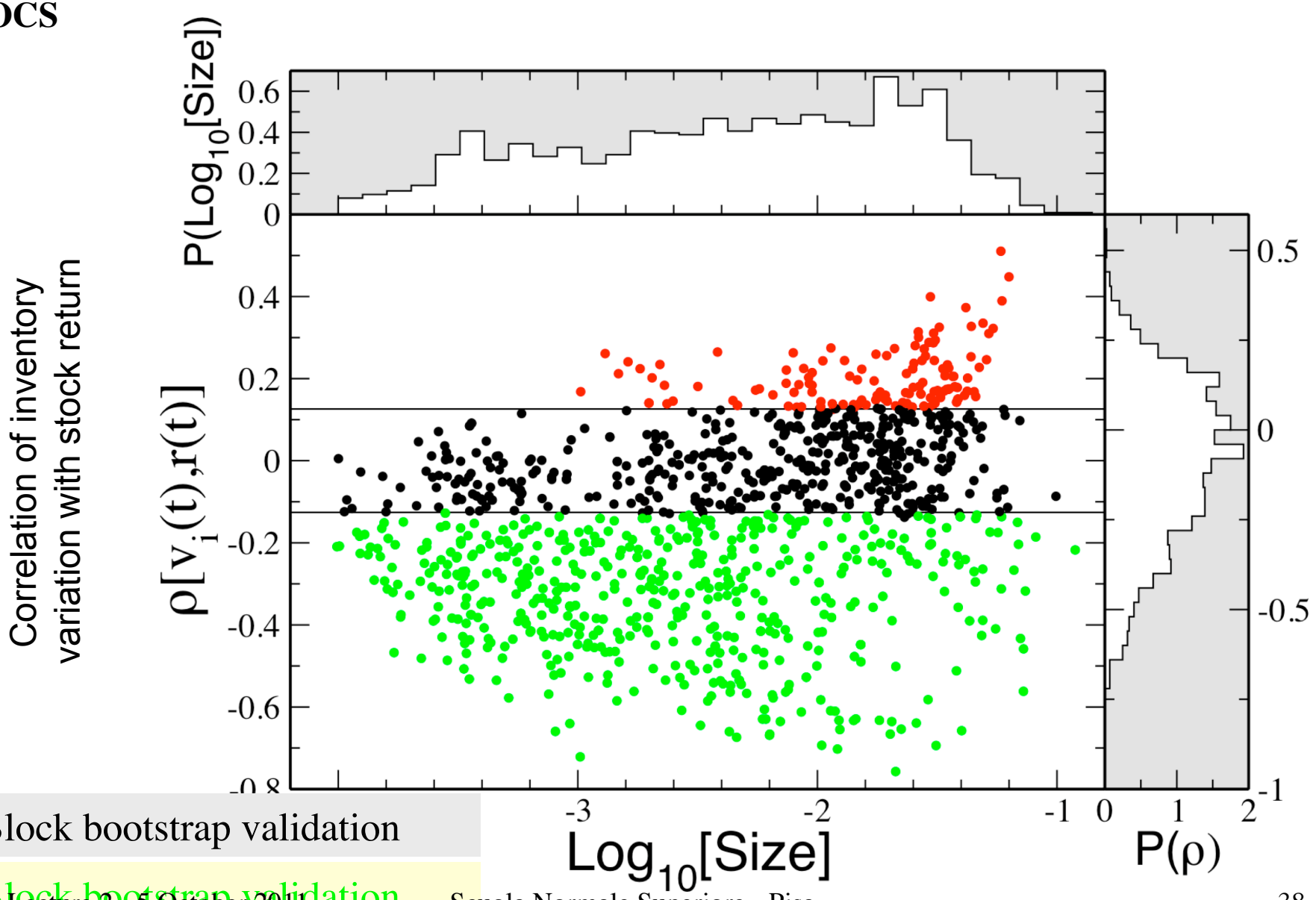
Uncategorized "noisy" firms

"trending" MMs (ex: momentum traders)



OCS

A closer look



- Block bootstrap validation

- Block bootstrap validation

OCS

A one-factor model of inventory variation dynamics

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

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

price return

idiosyncratic noise

$\gamma_i > 0$ trending MMs (ex: momentum strategies)

$\gamma_i < 0$ reversing MMs (ex: contrarians strategies)

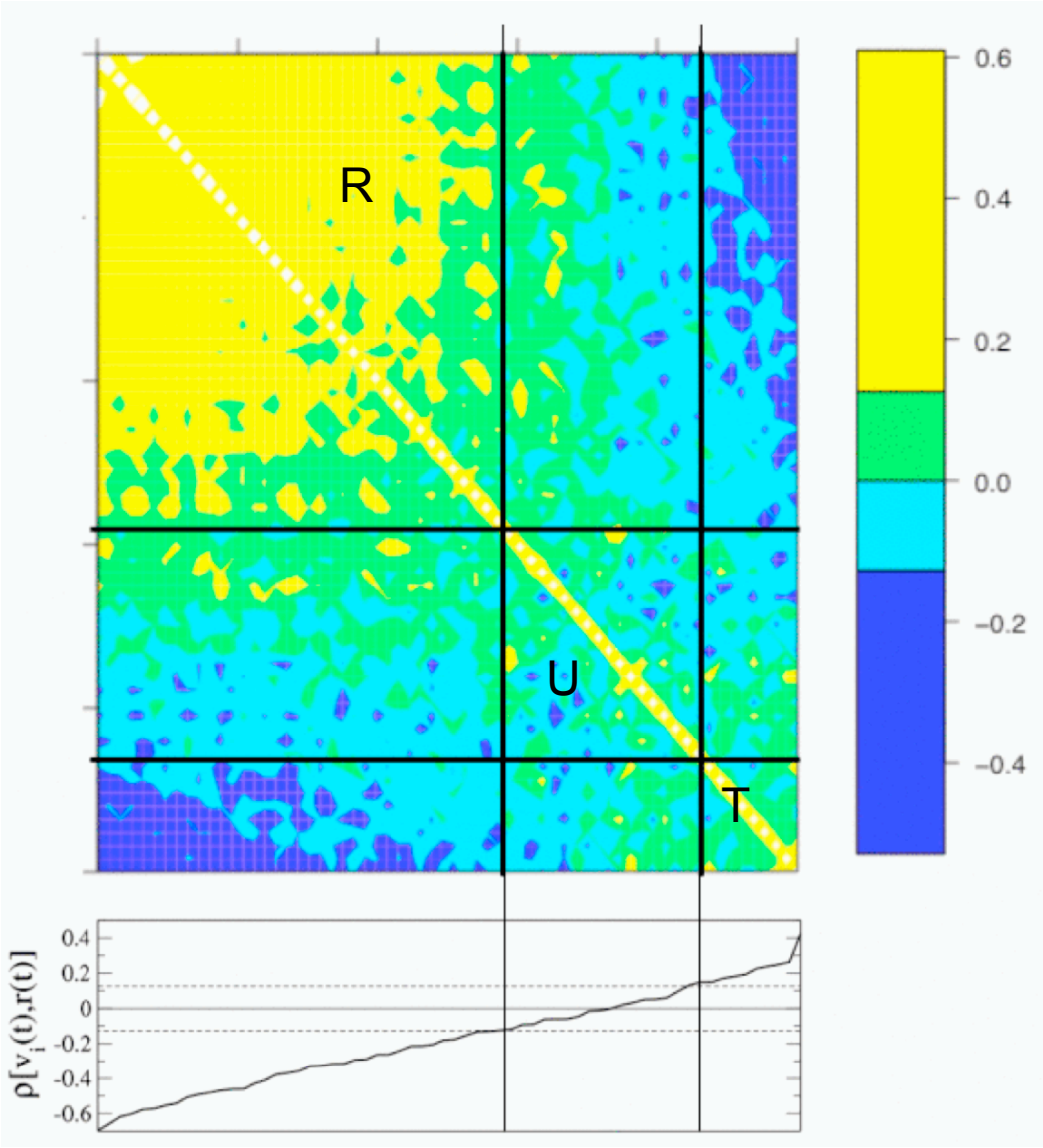
$\gamma_i \approx 0$ (effective) noisy firms

see, Lillo and Mantegna, Phys. Rev. E **72**, 016219 (2005)

OCS

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

BBVA 2003



OCS

The taxonomy is rather stable over the years

Categorization of active MMs for the Telefonica Stock

TEF	2001	2002	2003	2004
Reversing	43	39	42	37
Uncategorized	28	31	31	29
Trending	11	10	8	6
Total	82	80	81	72

OCS

Specialization is stable over the years

P(Y|X) is the probability that a MM of the group X switches to group Y in the next year (data for Telefonica stock averaged over 3 years)

X

Y	Reversing	Uncategorized	Trending
Reversing	0.71	0.19	0.03
Uncategorized	0.16	0.62	0.35
Trending	0.02	0.07	0.44
Exited	0.11	0.12	0.18

OCS

Trading profile of MMs of different groups

Is there a statistical evidence of a different herding behavior between the two characterized groups?

Such information might be helpful to develop an agent based model taking into account the different resulting strategies observed in the market.

OCS

A simple measure of univariate herding

For each group and for each day, by considering the inventory variation of the day, we measure the herding ratio h

$$h = \frac{\#\,of \cdot buying \cdot firms}{\#\,of \cdot active \cdot firms \cdot in \cdot the \cdot group}$$

OCS

Moreover, for each group we are able to detect buy herding days and sell herding days.

These are days when the number of buying MMs is not consistent with a binomial null hypothesis at a 95% confidence level.

OCS

Percentage of herding intervals observed for the groups of reversing, uncategorized and trending market members. One day time horizon.

	2001			2002			2003			2004		
	ALL	BH	SH	ALL	BH	SH	ALL	BH	SH	ALL	BH	SH
Reversing	66.8	34.8	32.0	65.2	34.8	30.4	64.8	31.2	33.6	59.6	27.2	32.4
Uncategorized	22.4	11.2	11.2	16.4	7.2	9.2	21.2	10.8	10.4	19.2	10.4	8.8
Trending	10.4	7.2	3.2	6.4	2.4	4.0	6.0	2.0	4.0	2.4	1.2	1.2

Herding is much more frequent in reversing MMs, less frequent in trending MMs and intermediate for uncategorized MMs.

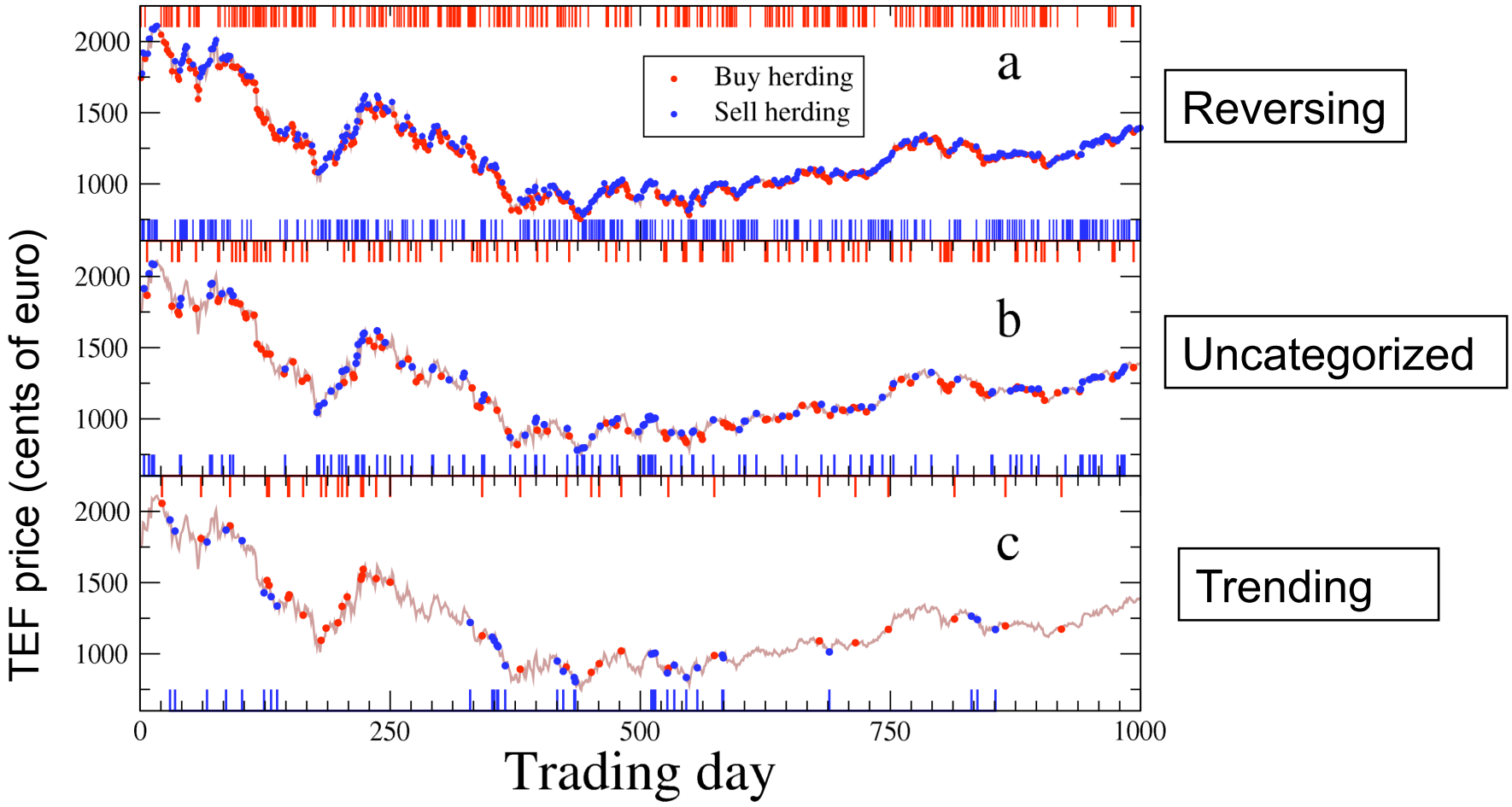
OCS

Percentage of herding intervals observed for the groups of reversing, uncategorized and trending market members. 15 min intraday time horizon.

	2001			2002			2003			2004		
	ALL	BH	SH	ALL	BH	SH	ALL	BH	SH	ALL	BH	SH
Reversing	35.1	17.4	17.7	34.5	17.3	17.2	29.2	14.7	14.5	26.6	13.3	13.3
Uncategorized	10.1	5.3	4.8	11.6	5.7	5.9	10.2	5.3	4.9	11.5	6.3	5.2
Trending	3.7	2.1	1.6	6.7	3.4	3.3	3.9	1.7	2.2	3.3	1.7	1.6

Herding is less pronounced at intraday time horizons. Due to the limited number of trending MMs we cannot prove that trending MMs are trying to avoid herding.

Herding days for the Telefonica stock



OCS

Different firms can buy or sell quite a different amount of money. Is this observation invalidating our herding observations?

We complement our herding investigation with a study considering an indicator of the net flow of value outcoming (or incoming) from each group. This indicator is a simplified version of the buy ratio used by Grinblatt et al, J. Finance Econ. 55, 43 (2000).

$$b = \frac{\sum_{i \in buying\_MMs} v_i}{\sum_{i \in all\_group\_MMs} |v_i|}$$

OCS

Mean value of the buy ratio b of MMs of a specific group which are active in a given daily time interval. (MMs trading Telefonica in 2001).

	$\langle b \rangle_{Shuf}$	$\langle b \rangle$	$\langle b \rangle_{BH}$	$\langle b \rangle_{SH}$
Reversing 43 (# of intervals)	0.538	0.52±0.28	0.77±0.15	0.22±0.16
		(250)	(87)	(80)
Uncategorized 28 (# of intervals)	0.418	0.48±0.16	0.55±0.16	0.43±0.15
	★	□ (250)	□ (28)	★□ (28)
Trending 11 (# of intervals)	0.556	0.51±0.25	0.81±0.20	0.22±0.19
		(250)	(18)	(8)

★□ indicates that mean values come from the same distribution according to a *t*-test with a 99% confidence level.

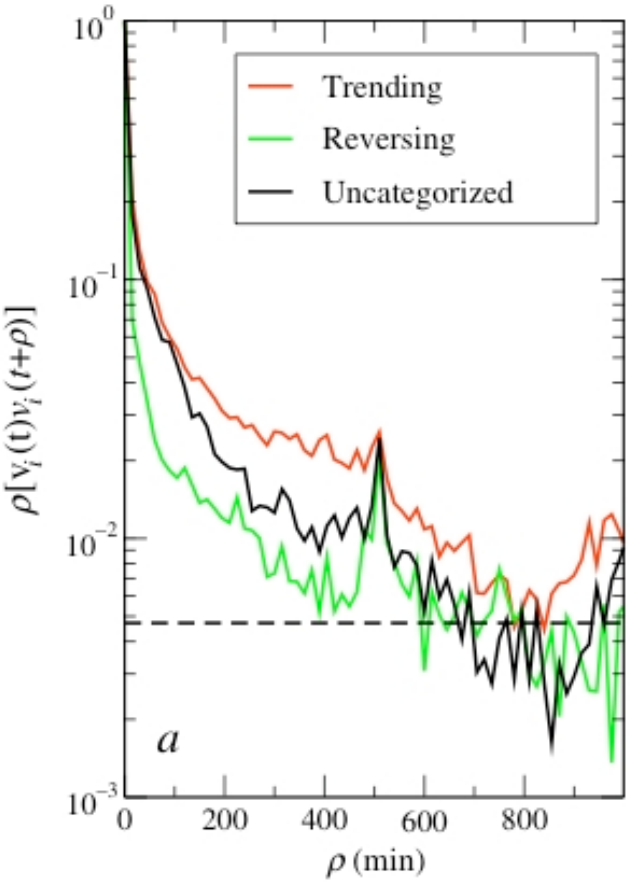
OCS

Granger causality

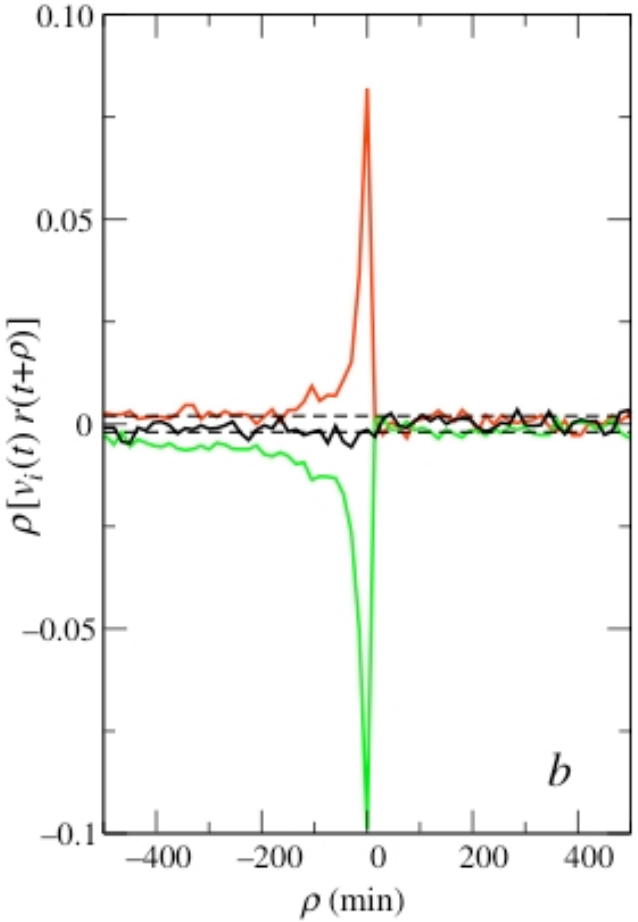
Is return Granger-causing inventory variation or vice versa?

OCS

Average autocorrelation is different for different groups of stocks



Lagged correlation between return and inventory variation suggests that return drives inventory variation

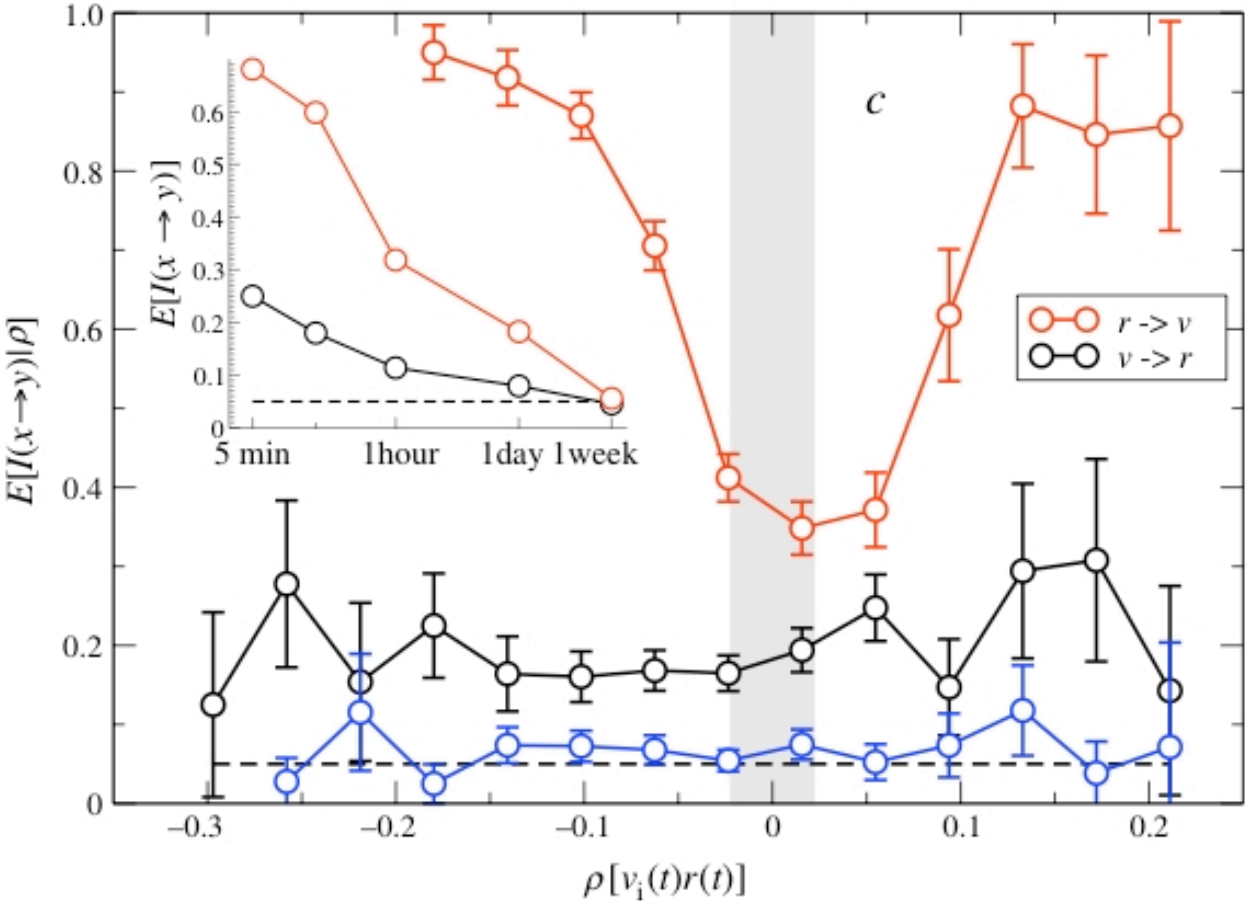


OCS

Granger causality test

We have studied the amount of time X Granger-causes Y with a 95% confidence threshold.
Main results are obtained for a 15 minutes time horizon.

Blue symbols are a test with shuffled data.

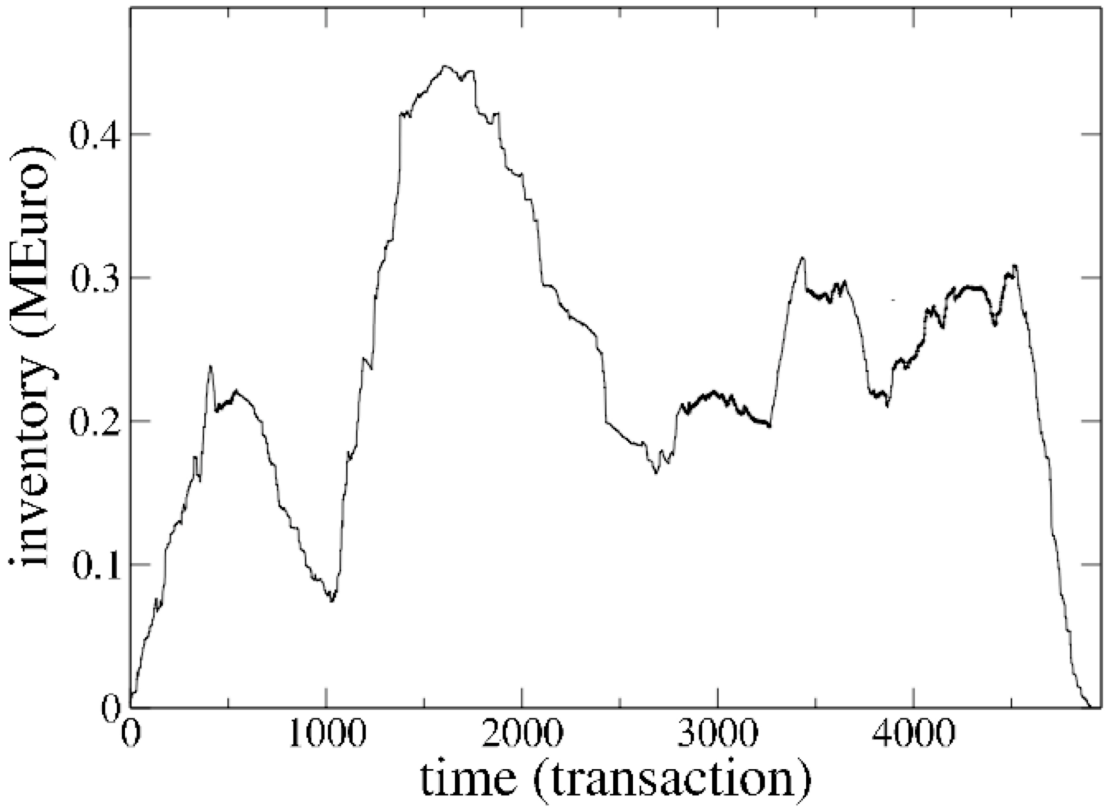


OCS

Submission of large orders in packages

OCS

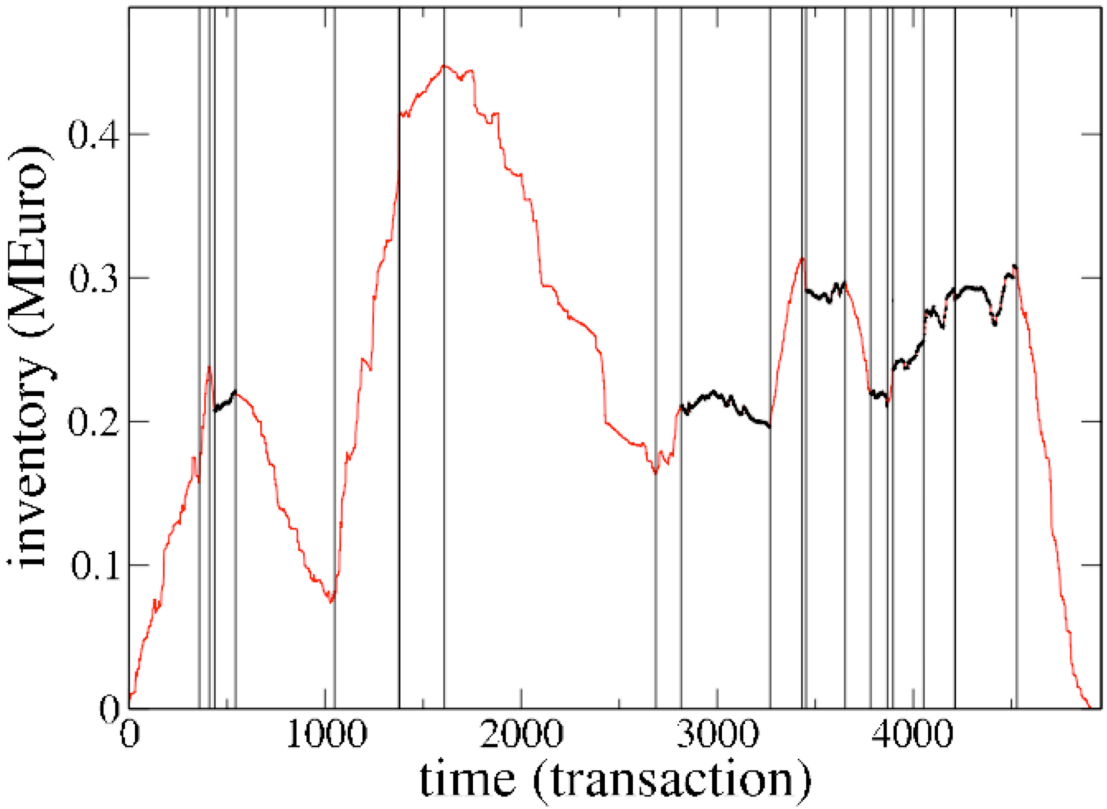
An example of inventory profile



Credit Agricole trading Santander

OCS

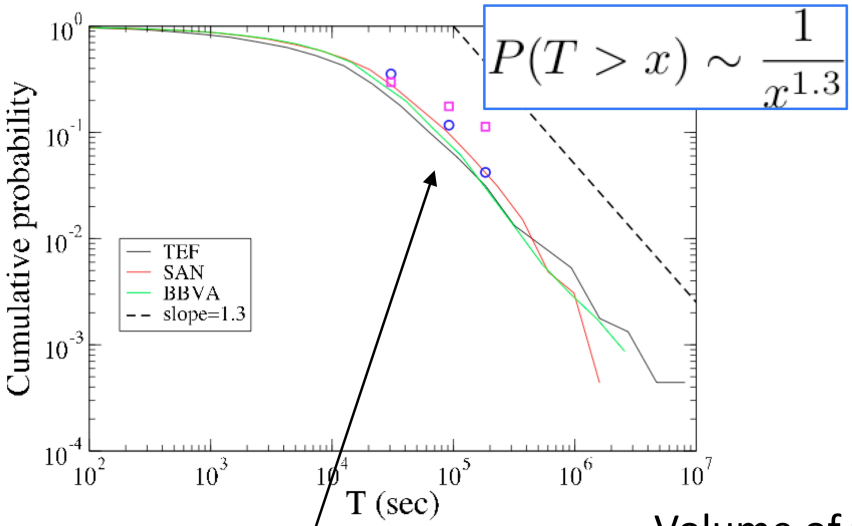
Detecting packages of large orders



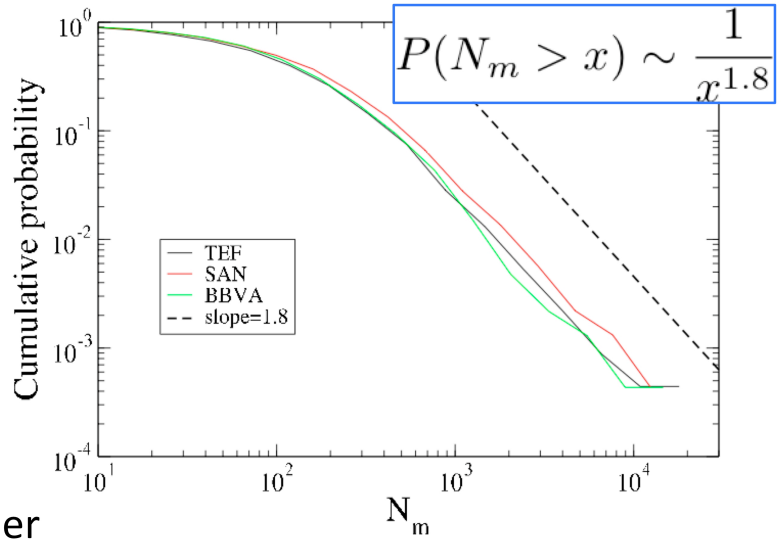
G. Vaglica, F. Lillo, E. Moro, R.N. Mantegna, Scaling laws of strategic behavior and size heterogeneity in agent dynamics, Physical Review E 77, 036110 (2008).

Scaling relations of packages

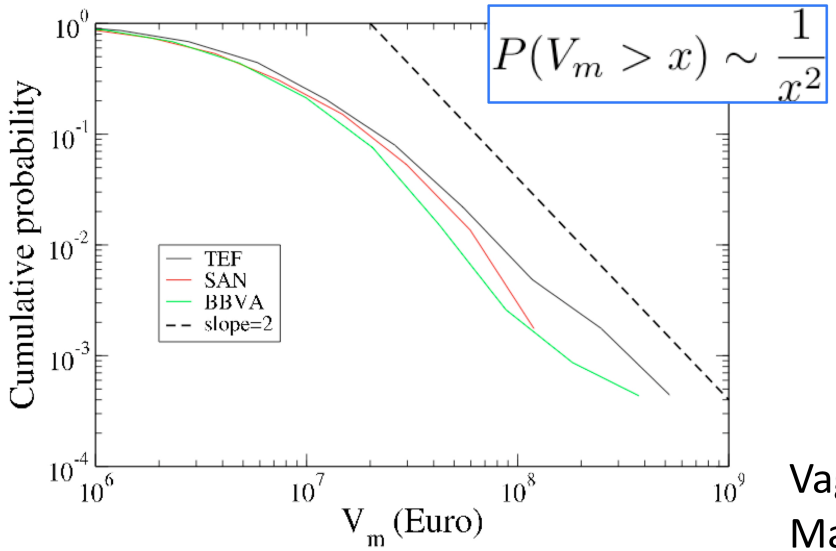
Investment horizon



Number of transactions



Volume of the order



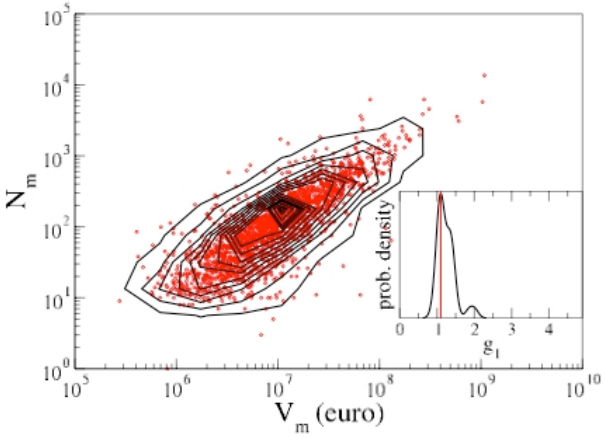
Circles and squares are data taken from Chan and Lakonishok at NYSE (1995) and Gallagher and Looi at Australian Stock Exchange (2006)

Vaglica, Lillo, Moro, Mantegna 2008

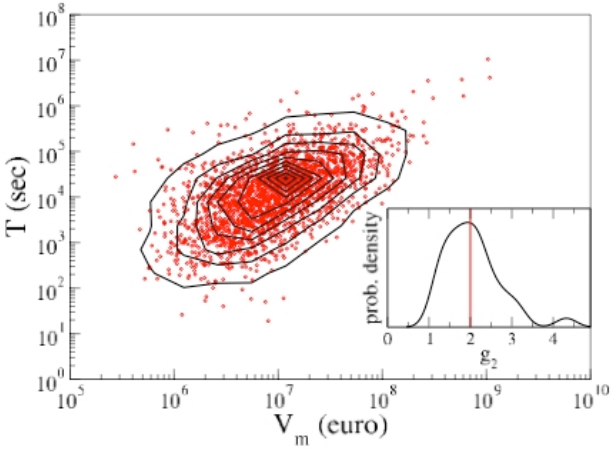
OCS

Allometric relations

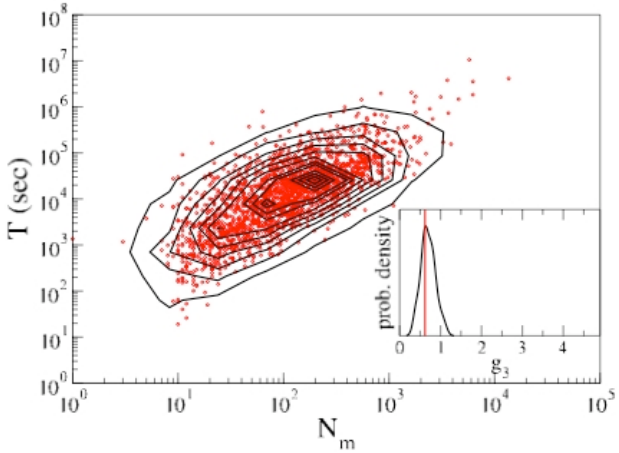
$N_m \sim V_m^{1.1}$



$T \sim V_m^{1.9}$



$N_m \sim T^{0.66}$



	BBVA (2104)	SAN (2086)	TEF (2062)
ζ_{V_m}	2.3 (1.9;2.7)	2.0 (1.7;2.3)	1.9 (1.6;2.2)
ζ_{N_m}	2.0 (1.7;2.3)	1.7 (1.4;2.0)	1.7 (1.4;2.0)
ζ_T	1.5 (1.3;1.7)	1.5 (1.3;1.7)	1.2 (1.0;1.4)
g_1	1.08 (1.05;1.12)	1.06 (1.01;1.10)	1.07 (1.04;1.11)
g_2	1.81 (1.69;1.93)	1.81 (1.68;1.94)	2.00 (1.88;2.14)
g_3	0.68 (0.65;0.71)	0.68 (0.65;0.70)	0.62 (0.59;0.64)
T	75 (15/20)	63 (17/27)	77 (24/31)
N_m	90 (18/20)	100 (27/27)	100 (31/31)
V_m	90 (18/20)	100 (27/27)	94 (29/31)
λ_1	90 (18/20)	85 (23/27)	87 (27/31)
λ_2	15 (3/20)	18 (5/27)	22 (7/31)

OCS

Which is the major source of heterogeneity?

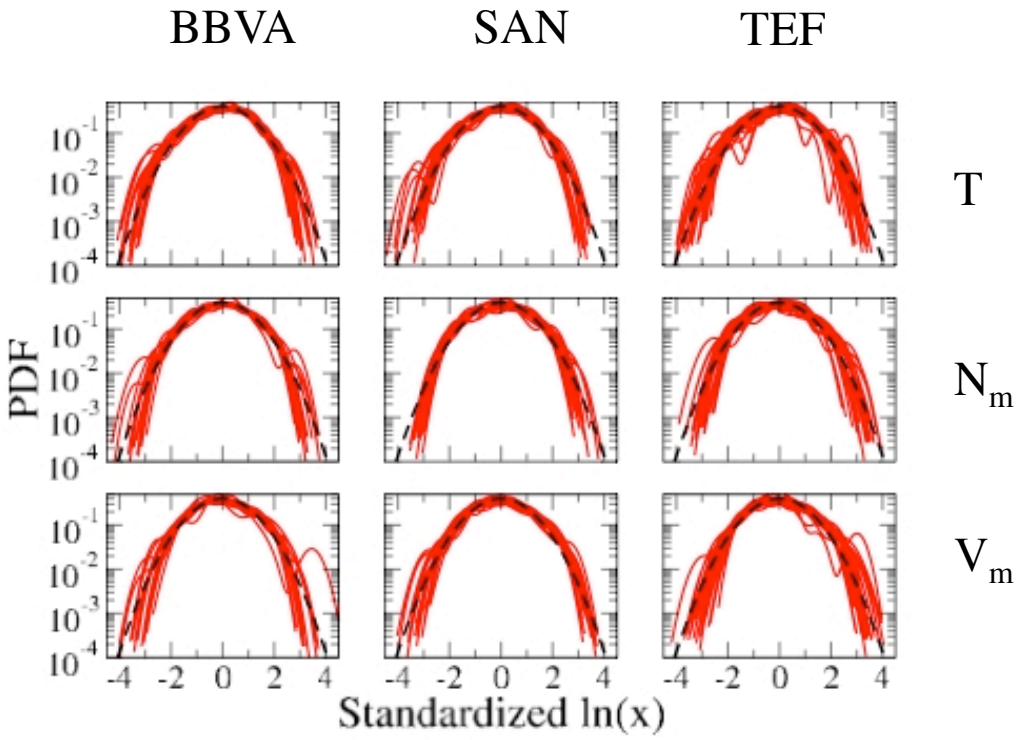


FIG. 5. (Color online) Probability density function of the standardized logarithm of the variables T, N_m, and V_m of the firms for which the Jarque-Bera test of log-normality cannot be rejected. Spe-

Most of the firms show a log-normal profile

The power-law profile in the complete set is due to the heterogeneity of the system

OCS

The NCSD database.

From January 1st, 1995 the Nordic Center Securities Data (NCSD) (today Euroclear) collects a database recording the daily ownership of financial portfolios and trading records of different classes of investors domiciled in Finland (companies, institutional governmental investors, foreign investors, no profit organizations, financial institutions and households).

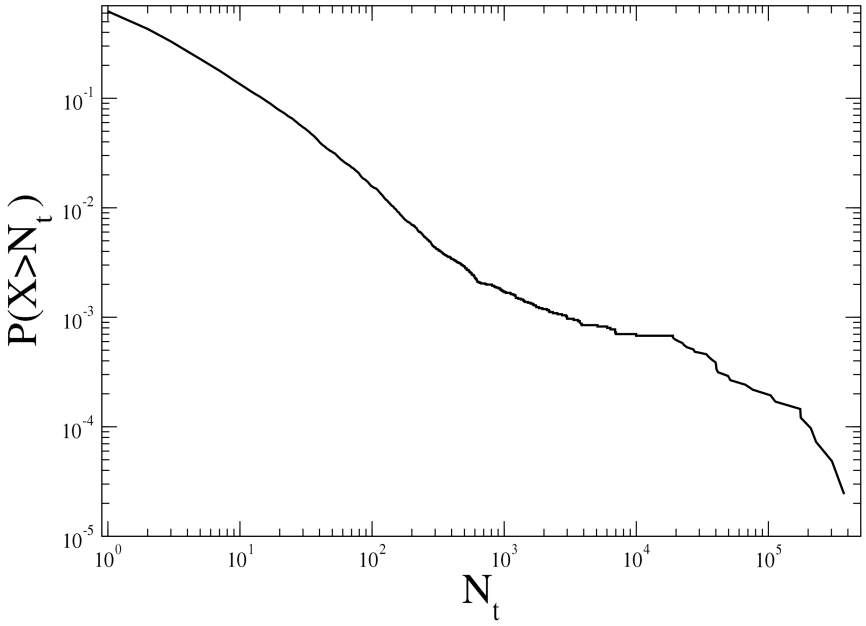
Identity of the investors is coded for privacy reasons.

This database has been extensively investigated at the aggregated level of classes of investors in the financial literature by Grinblatt and Keloharju (2000, 2001,2009).

OCS

The set of investors is quite heterogeneous in terms of volume and frequency of trading. Therefore a comprehensive analysis done in terms of correlation of variables monitoring their trading activity like, for example, inventory variation is not feasible.

Ex: Cumulative probability density $P(X > N_t)$ of the number of transactions N_t performed by 41250 Finnish legal entities trading the Nokia stock (typically but not exclusively at the Nordic Stock Exchange) during 2003.



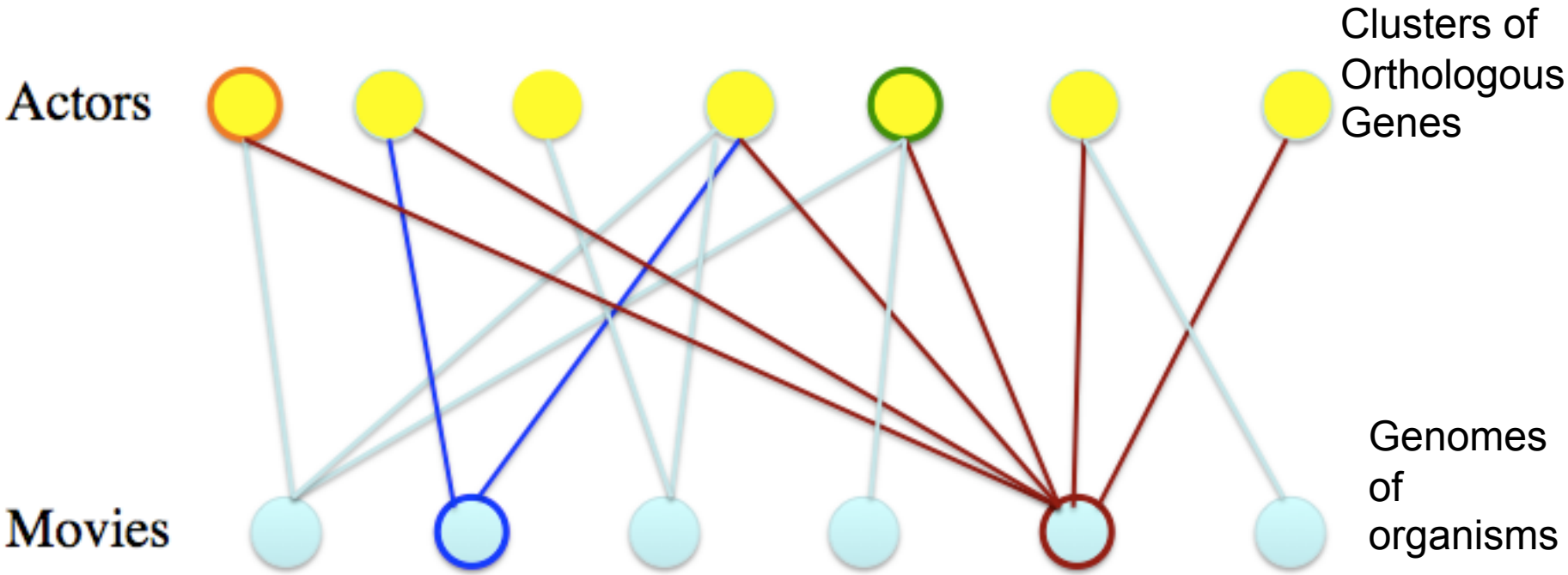
OCS

Our investigation is performed by using categorical variables. This choice allows us to perform a comprehensive analysis of a very large set of investors in spite of the huge level of heterogeneity present in it.

OCS

Networks in bipartite complex systems

Investors

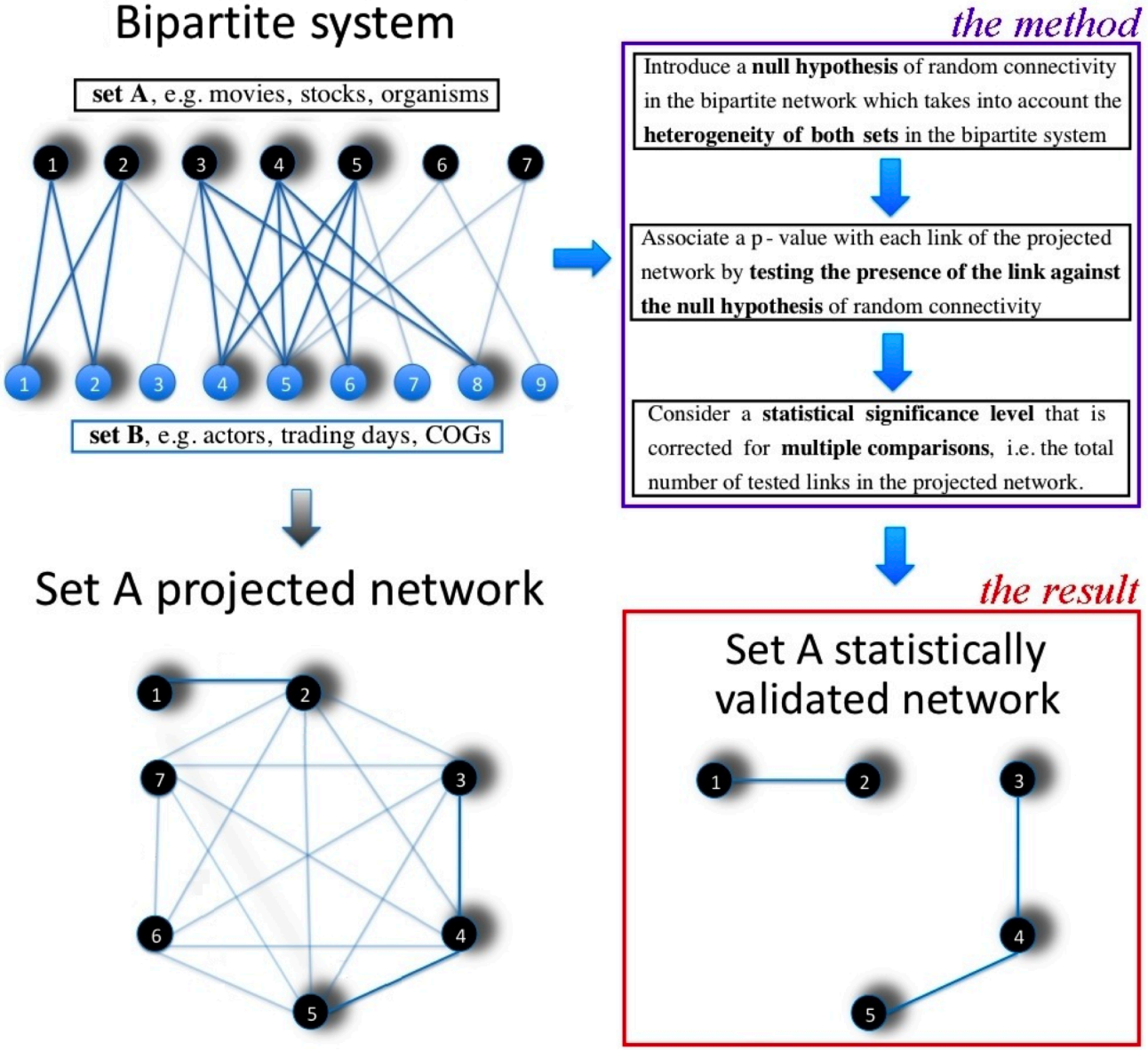


Investment decision on a specific day

Heterogeneity is observed on both kind of elements of the bipartite network

OCS

Statistically validated networks



Tumminello M, Miccichè S, Lillo F, Piilo J, Mantegna RN (2011) Statistically Validated Networks in Bipartite Complex Systems. PLoS ONE 6(3): e17994. doi:10.1371/journal.pone.0017994

OCS

Investment decisions at the Nordic Stock Exchange (1998-2003)

OCS

We study the buy/sell profile of investors trading the NOKIA stock during the time period
19 Oct 1998 – 29 Dec 2003

Approximately 170,000 different investors were recorded by NCSD trading NOKIA shares during the considered period. This number reduces to approximately 14,000 when we require that they are performing at least 20 different market transactions in the considered period.

OCS

Definition of categorical variables

V_B is the volume of an asset bought by a given investor in a given time horizon (in our case one day).

V_S is the volume of an asset sold by a given investor in a given time horizon (in our case one day).

We define 3 states:

Buy state: The investor is primarily buying.

Sell state: The investor is primarily selling.

BuySell state: The investor is buying and selling but closes her position with a very limited relative inventory variation.

The threshold procedure we use to obtain categorical variables is the following:

$$\frac{V_B - V_S}{V_B + V_S} > +\vartheta$$

Buy state

$$\frac{V_B - V_S}{V_B + V_S} < -\vartheta$$

Sell state

$$-\vartheta < \frac{V_B - V_S}{V_B + V_S} < +\vartheta \quad \text{with } V_B > 0 \text{ and } V_S > 0$$

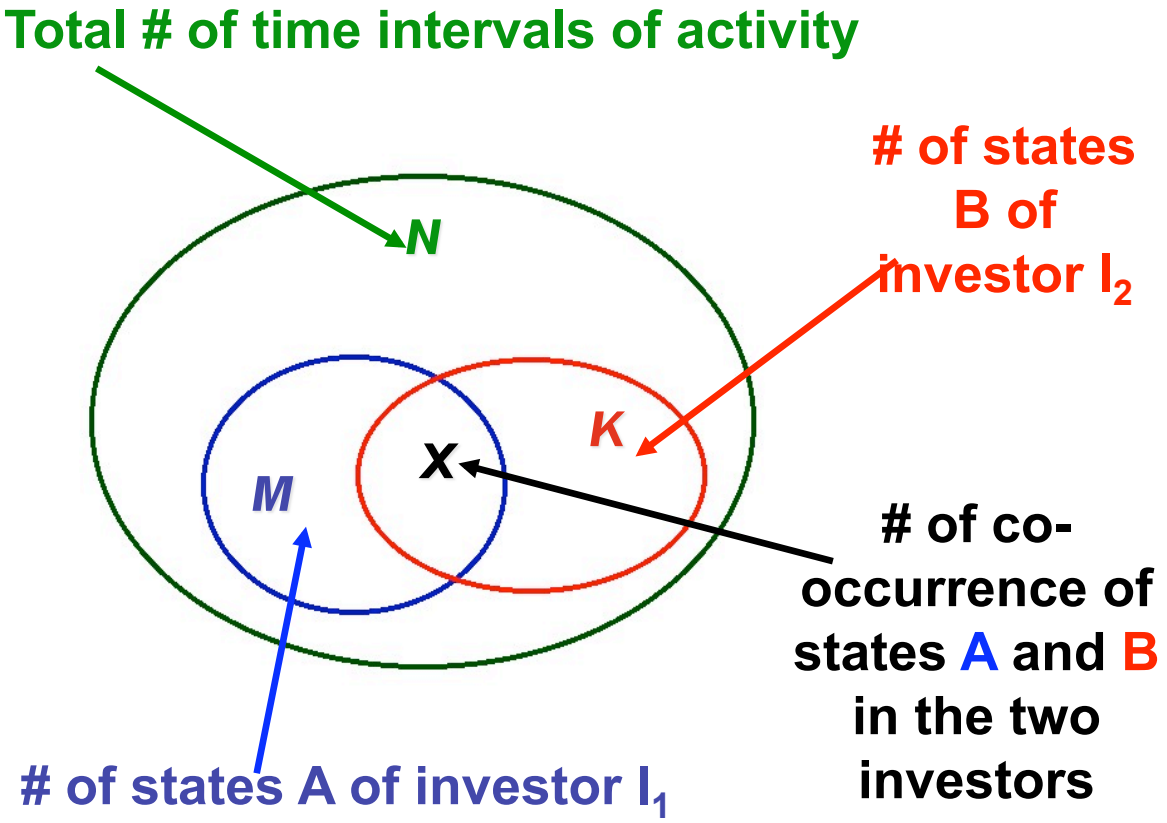
BuySell state

with $\vartheta = 0.25$ and $\vartheta = 0.01$

OCS

A statistical validation of co-occurrence

Suppose there are N time intervals in the investigated set. Suppose we are interested to compare the occurrence of two given states A and B of two investors I_1 and I_2. Investor I_1 is M times in the state A whereas investor I_2 is K times in the state B. Let us call X the co-occurrence of states A and B of the two investors.



The question characterizing the null hypothesis is:

what is the probability that the number X occurs by chance?

OCS

A solution of this problem exists

The probability of having exactly X co-occurrence of states **A** and **B** between investors I_1 and I_2 is given by the Hypergeometric distribution

Hypergeometric distribution:

$$P(X \mid N,M,K) = \frac{\binom{M}{X}\binom{N-M}{K-X}}{\binom{N}{K}}$$

Expected number of co-occorrence:

$$\langle X \rangle = \sum x \; P(x \mid N,M,K)$$

p-value associated to a detection of co-occurrence ≥ X:

$$p = 1 - \sum_{i=0}^{X-1} \frac{\binom{M}{i}\binom{N-M}{K-i}}{\binom{N}{K}}$$

OCS

Network construction and multiple hypothesis testing correction

We link two vertices (e.g. investors, movies, genomes, etc) if the associated *p*-value is below a given statistical threshold.

We are performing a multiple hypothesis test and therefore we need a multiple hypothesis test correction.

OCS

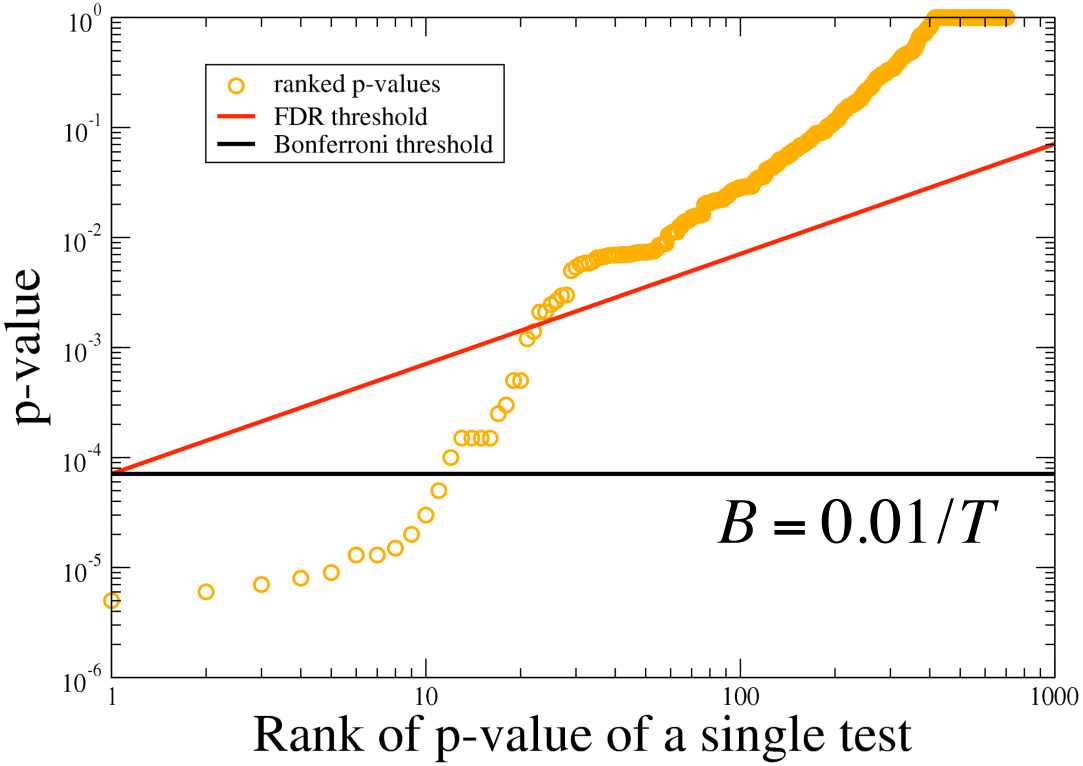
The most restrictive multiple hypothesis testing correction is the so-called: **Bonferroni correction.**

It is defined as follow: by requiring a 0.01 individual statistical threshold the threshold B for the multiple test procedure is set to B=0.01/T where T is the total number of tested hypotheses.

We address the statistically validated network obtained with the Bonferroni correction as the **Bonferroni network**

OCS

There is a less restrictive multiple hypothesis test correction widely used in statistics. It is called **False Discovery Rate correction**



In the False Discovery Rate (FDR) correction the threshold F is set to F= n B where n is the smallest integer such that there are n p-values smaller than n B in the system.

We address the statistically validated network obtained with the FDR correction as the **FDR network**

OCS

Multi-link statistically validated network

Each investor can participate to 9 different types of co-occurrences of the 3 states **Buy**, **Sell** and BuySell with any other investor.

Two investors participating to the Bonferroni or the FDR network will be characterized by a link describing one, or more than one, of the 9 possible validated co-occurrences.

OCS

Multi-link statistically validated network

Investor 1 \ Investor 2	Buy	Sell	BuySell
Buy	1	0	0
Sell	0	1	0
BuySell	0	0	0

With this matrix representation we mean that the FDR link indicates co-occurrence of the **Buy I_1 – Buy I_2** and **Sell I_1 – Sell I_2** activities

The number of different co-occurrence combinations is $2^9=512$

OCS

During 1998-2003 14,735 Nokia investors did more than 20 transactions

The **Bonferroni** network comprises 3,118 investors which are connected by 36,664 multi-links

The **FDR** network comprises 10,435 investors which are connected by 330,404 multi-links

TABLE I: Most populated co-occurrence combinations in Bonferroni and FDR network

Label	Co-occurrence combination	Bonferroni (36664)	FDR (330404)	Color label
C1	(i_b, j_b)	7,716 (21.0)	120,655 (36.5)	magenta
C2	(i_s, j_s)	6,254 (17.1)	91,219 (27.6)	green
C3	(i_{bs}, j_{bs})	1,732 (4.72)	19,227 (5.82)	apricot
C4	(i_b, j_b) (i_s, j_s)	20,243 (55.2)	66,692 (20.2)	black
C5	(i_b, j_{bs})	312 (0.85)	13,494 (4.08)	blue
C6	(i_s, j_{bs})	157 (0.43)	9,592 (2.90)	orange
C7	(i_s, j_b)	12 (0.033) *	2,662 (0.81)	tan
C8	(i_b, j_b) (i_s, j_s) (i_{bs}, j_{bs})	137 (0.37) *	2,304 (0.70)	brown
C9	(i_b, j_{bs}) (i_s, j_{bs})	43 (0.12) *	1,414 (0.43)	purple

More than 99% of multi-links belong to just 9 different co-occurrence combinations

OCS

Top multi-links of the FDR network of investors

When θ=0.01 we detect a network of **10435 investors** connected by **330404 multi-links**. The most common kinds of multi-links are

	B	S	BS
B	**1**	0	0
S	0	0	0
BS	0	0	0

1 120655

magenta

	B	S	BS
B	0	0	0
S	0	**1**	0
BS	0	0	0

2 91219

light green

	B	S	BS
B	**1**	0	0
S	0	**1**	0
BS	0	0	0

3 66692

black

	B	S	BS
B	0	0	0
S	0	0	0
BS	0	0	**1**

4 19227

apricot

	B	S	BS
B	0	0	**1**
S	0	0	0
BS	0	0	0

5 13494

blue

	B	S	BS
B	0	0	0
S	0	0	**1**
BS	0	0	0

6 9592

orange

	B	S	BS
B	0	0	0
S	**1**	0	0
BS	0	0	0

7 2662

tan

	B	S	BS
B	**1**	0	0
S	0	**1**	0
BS	0	0	**1**

8 2304

brown

	B	S	BS
B	0	0	0
S	0	0	0
BS	**1**	**1**	0

9 1414

light purple

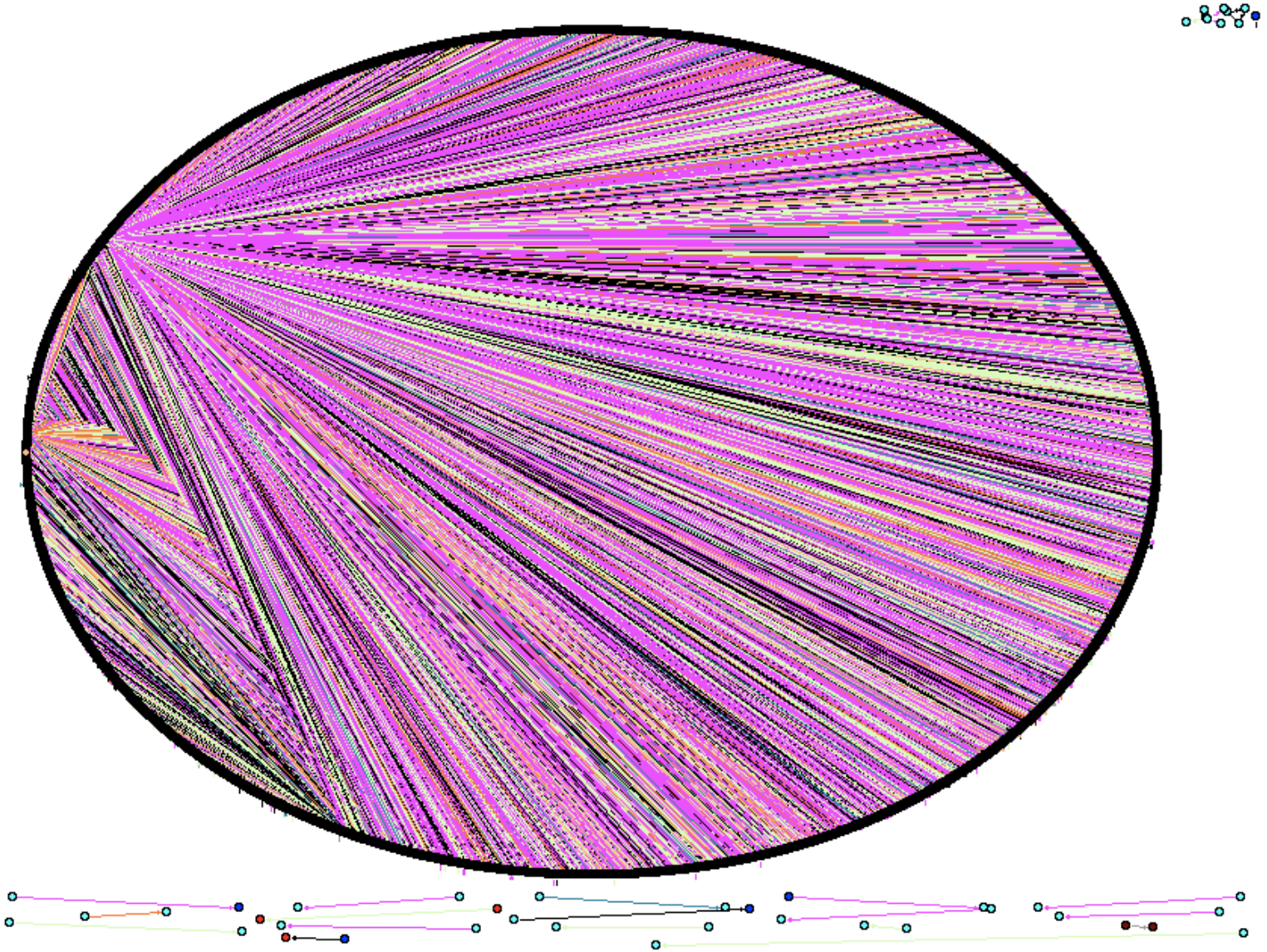
	B	S	BS
B	**1**	0	0
S	0	0	0
BS	0	**1**	0

10 460

gray

OCS

The FDR network has a giant component covering almost entirely the set of investors (10392/10435)



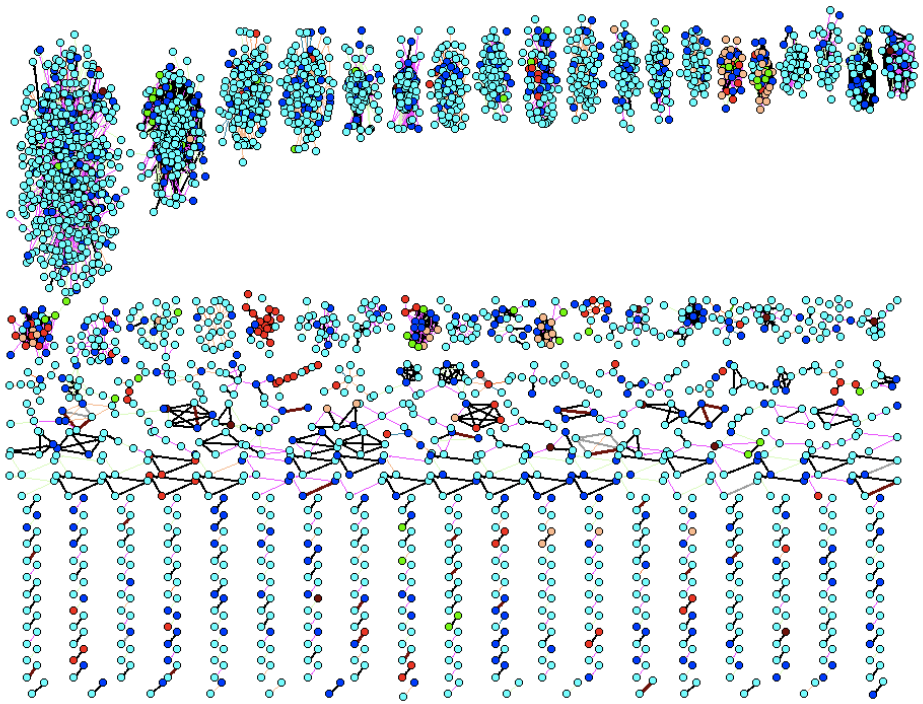
OCS

We perform unsupervised community (cluster) detection in the statistically validated networks

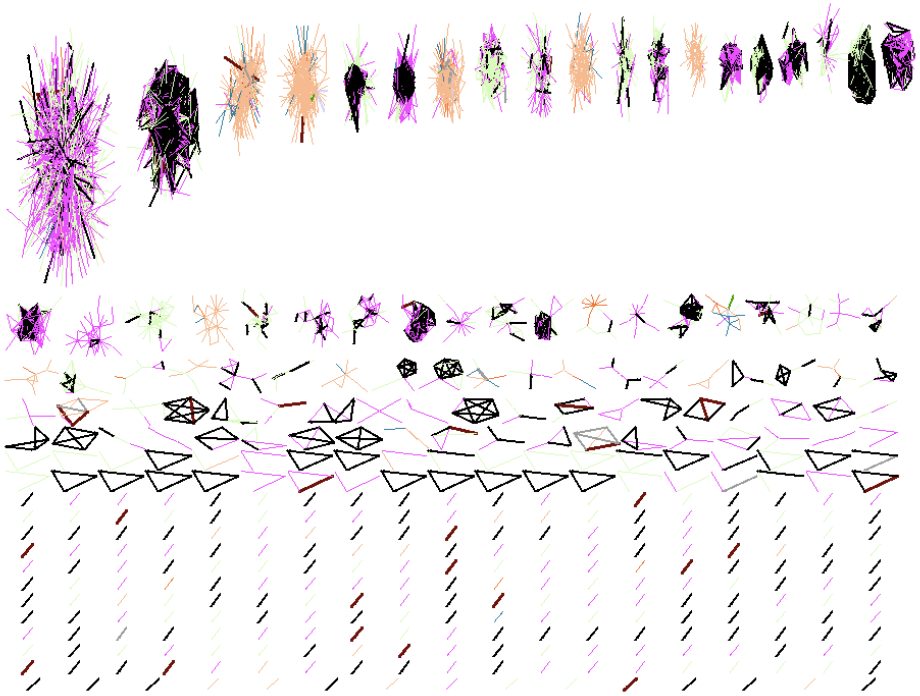
Example: The Infomap partition of the Bonferroni network of investors

OCS

Clusters of the Bonferroni network



focus on vertices



focus on links

- Companies
- Gov. inst.
- Financial inst.
- Non profit org.
- Foreign inst.
- Households

B.B

S.S

B.B S.S

BS.BS

B.BS

OCS

Over-expression validation of vertex or link attributes

For a given set of elements (investors, links, etc) we count how many of them are present in our reference set. We count the same information also inside each subset of interest. For the sake of simplicity, let us focus on investor classes but similar conclusion applies for different attributes. For each subset **a** and for each investor class **k** we have the number $\mathbf{N_{a,k}}$ of investors of class **k** present in the subset **a**, the number $\mathbf{N_a}$ is the number of investors of subset **a**, $\mathbf{N_k}$ is the number of investors of class **k** in the subset and the number $\mathbf{N_n}$ is the number of investors in the reference set. The probability that **X** investors of subset **a** belongs to class **k** under a random null hypothesis is again given by the hypergeometric distribution $\mathbf{H(X|N_n,N_a,N_k)}$ and a *p*-value can therefore be associated to the observation of $\mathbf{N_{a,k}}$ occurrence.

OCS

Over-expression validation of vertex or link attributes

Again this is a multiple hypothesis test procedure and a Bonferroni threshold is set as $\mathbf{0.01/N_{att}}$ for each test of each partition, where $\mathbf{N_{att}}$ is the number of different attributes that are tested. In the example the number of different investor classes.

Michele Tumminello , Salvatore Micciché , Fabrizio Lillo , Jan Varho , Jyrki Piilo and Rosario N Mantegna, Community characterization of heterogeneous complex systems J. Stat. Mech. (2011) P01019 doi: 10.1088/1742-5468/2011/01/P01019

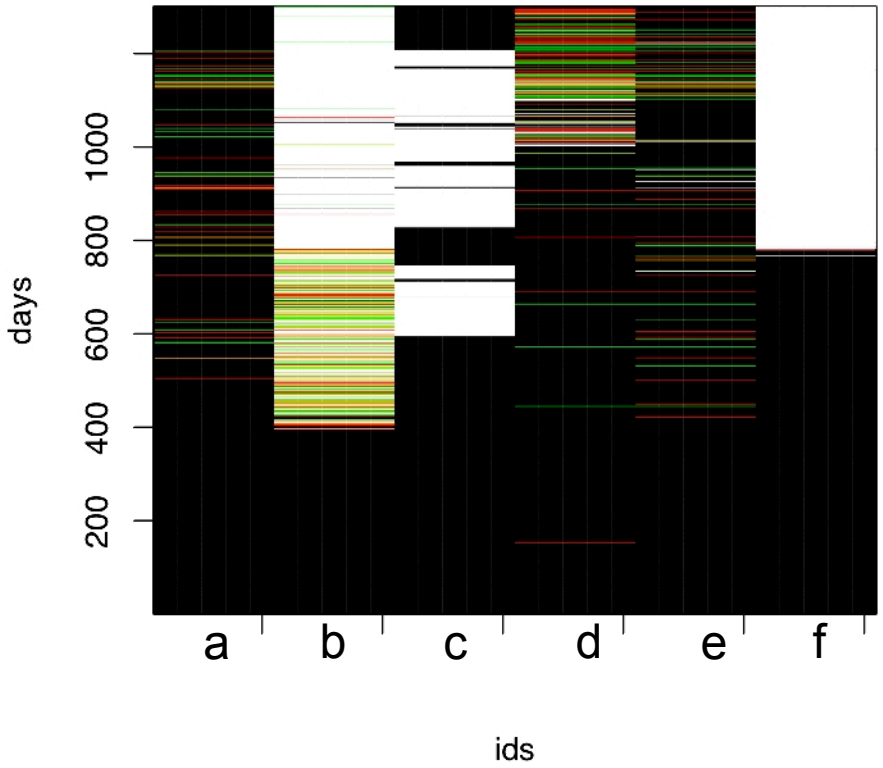
OCS

A visualization procedure:

Buy state

Sell state

BuySell state

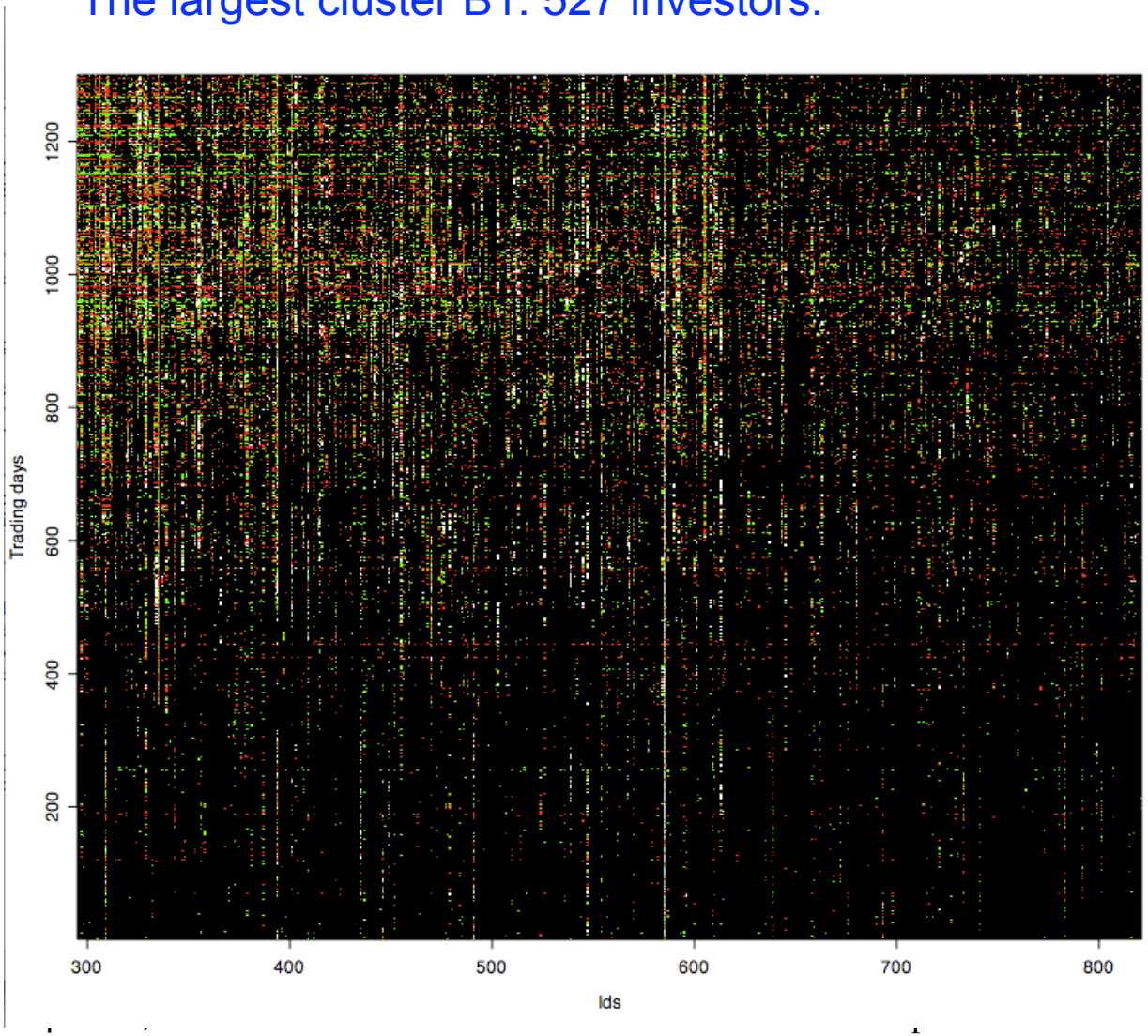


Label	Id	Class	Color
a)	41669	Gov. Investors	Apricot
b)	543896	Companies	Blue
c)	1165458	Foreign inv	Brown
d)	406444	Households	Cyan
e)	45768	No profit	Green
f)	101294890	Fin. Institutions	Red

OCS

Different clusters have different trading profiles and some of them have an over-expression of specific attributes of vertices and links

The largest cluster B1: 527 investors.



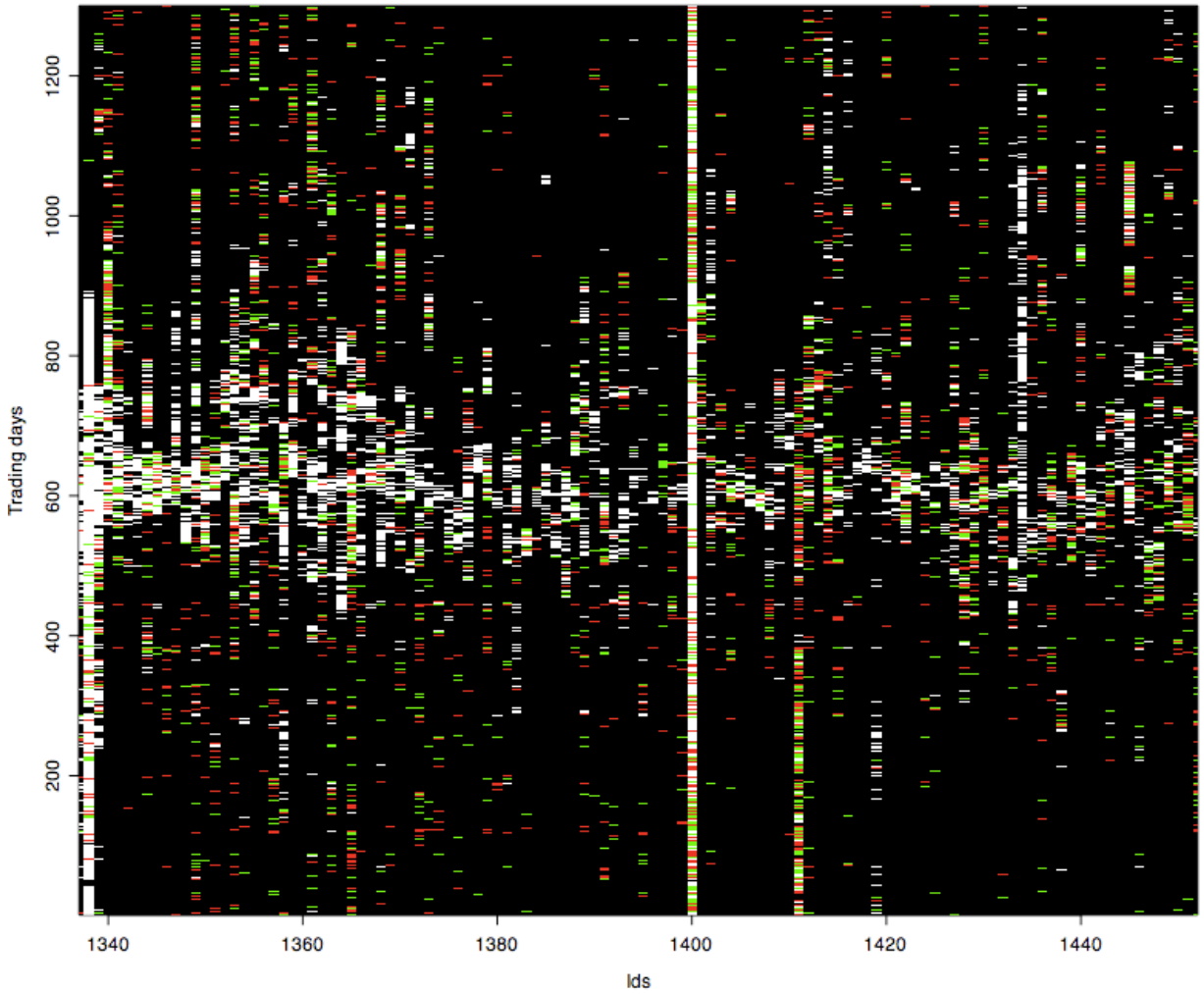
Over-expression of:

- Households investors;
- C1 (B.B) and C2 (S.S) links;
- Age cohort1941-1950
- Male gender

The trading activity is pretty frequent and almost continuous over a period of time spanning a period of time longer than a year

OCS

B4: 116 investors.

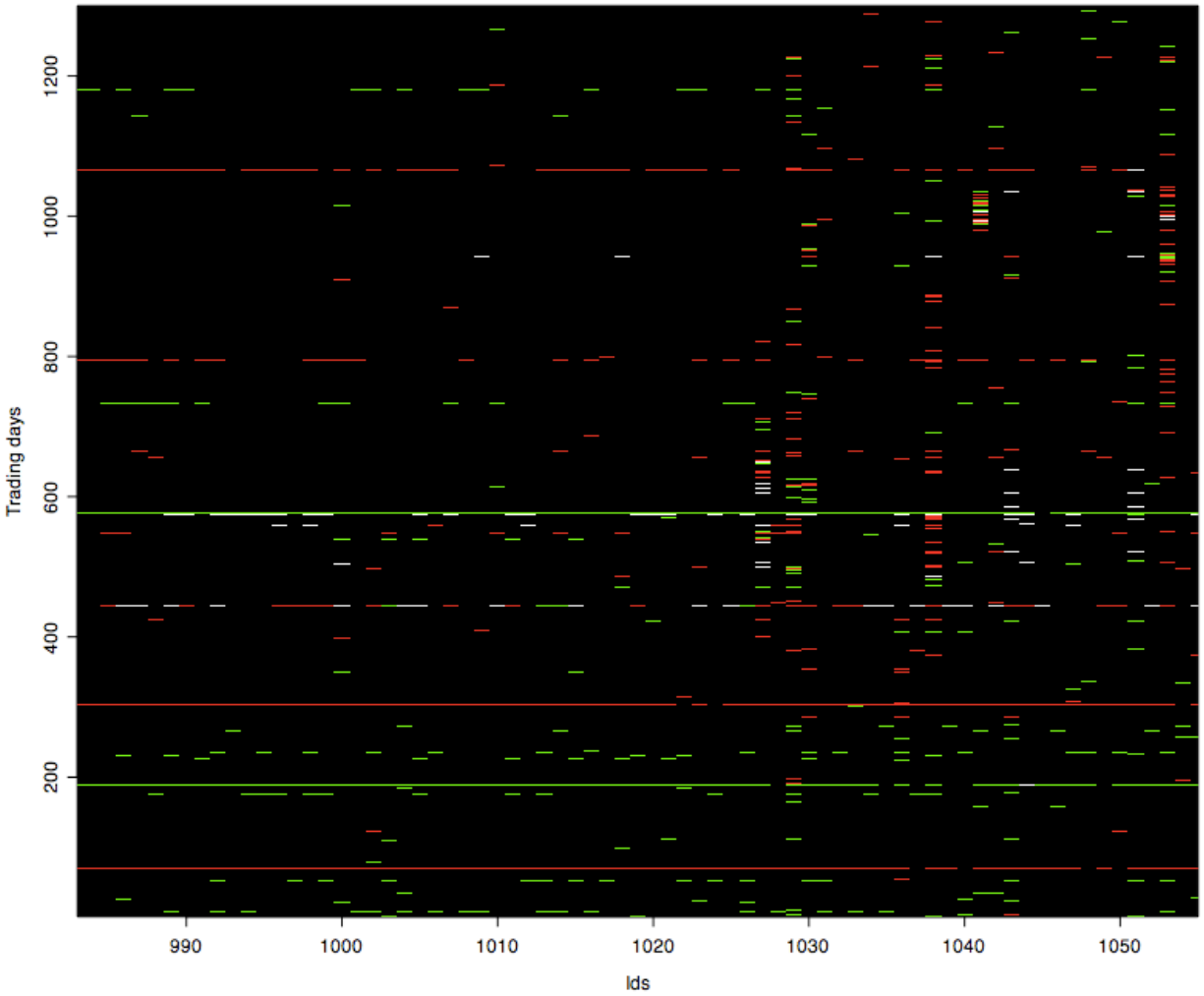


Over-expression of:
- C3 (BS.BS);

The trading activity is sometime pretty localized in a specific period of time.

OCS

B8: 73 investors.

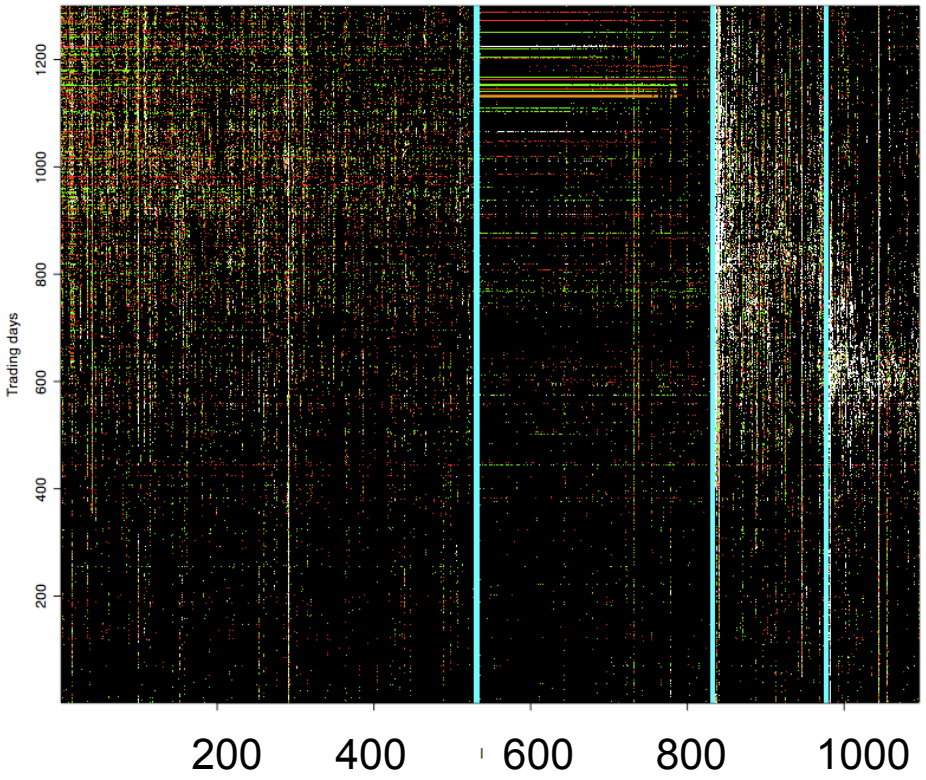


Over-expression of:
- C2 (S.S);

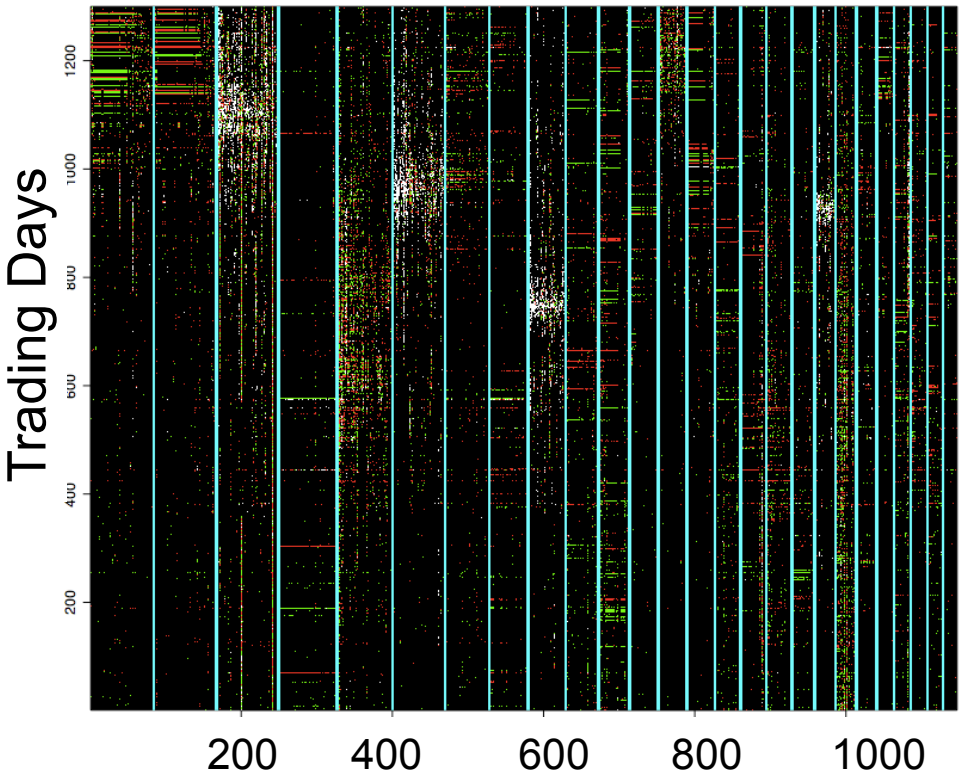
In other cases the frequency of the trading activity is much lower but synchronicity in the trading decisions is seen on a time period spanning several years.

OCS

Trading activity profile of the 30 largest clusters of the Bonferroni network



4 largest clusters



next 26 clusters

H = Households
C= Companies
G= Gov. Inst.
FI= Financial Inst.
NP= Non Profit Inst.
C1= B.B
C2= S.S
C3= BS.BS
C4= B.B S.S
C5= B.BS
C6= S.BS
C7= B.S
C8= B.B S.S BS.BS
C9= B.BS S.BS
0=Legal entity
1=1902-1940
2=1941-1950
3=1951-1960
4=1961-1970
5=1971-2000
0=Legal entity
1=Male
2=Female

Characterization of link and vertex attributes of clusters of the Bonferroni network

TABLE III: Summary statistics of the 30 most populated clusters of the Bonferroni network detected with Infomap. For each cluster we statistically validate the over-expression or under-expression of investors belonging to a specific class: companies (C), institutional governmental organizations (G), foreign organizations (FO), non-profit organizations (NP), financial institutions (FI) and households (H). We also statistically validate the over-expression or under-expression of multi-links belonging to a specific co-occurrence combination. The list of most frequent co-occurrence combinations are given in Table I.

Cluster	Investors	Over-expr. investor class	Under-expr. Investor class	Over-expr. co-occur. comb.	Under-expr. co-occur. comb.	Age class	Gender	Juridical class.	Postcode area	Investor code
B1	527	H	C G NP FI	C1 C2	C3 C4	2	1	11		500
B2	294		FI	C4	C1 C2 C3 C5 C6 C9		2			520
B3	138			C3 C5 C6 C9	C1 C2 C4		1			500
B4	116			C3	C1 C2 C4					
B5	82			C4	C1 C2 C3					
B6	79			C1 C4 C5	C2 C3 C8					
B7	78			C3 C5 C6 C9	C1 C2 C4					
B8	73			C2	C1 C3 C4					
B9	70			C1 C2	C4					
B10	65			C3 C5	C1 C2 C4	4				
B11	55			C2	C1 C3 C4				5	
B12	47			C1 C2	C3 C4				5	
B13	46			C3	C1 C2 C4					
B14	39	G NP	H	C1	C2 C3 C4	0	0	34 51 52	1	351 430
B15	37	G NP	H	C2	C1 C3	0	0	34		351 430
B16	34				C3				1	520
B17	34			C1 C2	C4					
B18	33			C4	C1 C3					
B19	30	FI	H	C1	C3 C4	0	0	32 41 90		221
B20	30			C1	C2 C3 C4					
B21	30			C1	C2 C4					
B22	26			C2	C4				8	
B23	24			C3	C4					
B24	23	FI	H	C2	C4	0	2	51	1	260
B25	23			C1	C4					
B26	19			C1	C4					
B27	18	G NP	H	C1	C2	0	0	17 51 63		320 430
B28	18			C1	C4					
B29	17	G	H			0	0	34		351
B30	17			C2	C4					

H = Households
C= Companies
G= Gov. Inst.
FI= Financial Inst.
NP= Non Profit Inst.
C1= B.B
C2= S.S
C3= BS.BS
C4= B.B S.S
C5= B.BS
C6= S.BS
C7= B.S
C8= B.B S.S BS.BS
C9= B.BS S.BS
0=Legal entity
1=1902-1940
2=1941-1950
3=1951-1960
4=1961-1970
5=1971-2000
0=Legal entity
1=Male
2=Female

Characterization of link and vertex attributes of clusters of the FDR network

TABLE IV: Summary statistics of the 30 most populated clusters of the FDR network detected with Infomap. For each cluster we statistically validate the over-expression or under-expression of investors belonging to a specific class: companies (C), institutional governmental organizations (G), foreign organizations (FO), non-profit organizations (NP), financial institutions (FI) and households (H). We also statistically validate the over-expression or under-expression of multi-links belonging to a specific co-occurrence combination. The list of most frequent co-occurrence combinations are given in Table I.

Cluster	Investors	Over-expr. investor class	Under-expr. Investor class	Over-expr. co-occur. comb.	Under-expr. co-occur. comb.	Age class	Gender	Juridical class.	Postcode area	Investor code
F1	3000	H	G NP FI	C1 C2 C5 C6 C9	C4 C3 C8	2	1	11	5	500
F2	1851	H	C G	C1	C2 C3 C4 C5 C6 C8 C9	1	2	11 12		520
F3	931		G	C3 C5 C6 C9	C1 C2 C4 C8	3	1			500
F4	639			C1 C4 C9	C2 C3 C5 C6 C8					500
F5	438	C NP	H	C4 C8	C1 C2 C3 C5 C6 C9	0	0 2	31	1	121 430
F6	312	FI		C2 C5 C6	C4 C8					
F7	223			C3 C5 C6	C1 C2 C4					
F8	205	C G FI NP	H	C4	C1 C2 C3 C5 C6 C8 C9	0	0	31 34 41 51 52 63 71 90	1	121 221 260 320 351 430
F9	140			C3 C5 C6 C9	C1 C2 C4					
F10	129			C2 C4	C1 C3 C5 C6 C9					
F11	127			C3 C5 C6 C9	C1 C2 C4					
F12	85			C2	C1 C3 C4 C5 C6				8	512
F13	68			C4	C1 C3 C5 C6				5	
F14	54			C3 C5 C6	C1 C2 C4		0			
F15	40			C4	C2 C3 C5				1	520
F16	39			C4	C1 C2 C3 C5 C6				1	
F17	39			C4	C2 C3 C5 C6					
F18	37			C1						
F19	29			C4	C2					
F20	26			C2	C1					
F21	26			C6	C3					
F22	24			C6						
F23	22			C4 C8	C1					
F24	20			C8	C2					
F25	19			C4	C1					
F26	19			C2	C1				4	
F27	17									
F28	16									
F29	16					5				
F30	16									

OCS

Inclusiveness relationship among FDR and Bonferroni clusters

TABLE II: Inclusiveness relationships of the 30 most populated Bonferroni clusters. The relationship is indicated when more than 75% of the elements are present into a single FDR cluster. An asterisk indicates that more than 90% of the elements are present in the corresponding FDR cluster.

FDR Cluster	Bonferroni clusters
F1	B1 (*) B10(*) B11 B23 (*)
F2	B21 (*)
F3	B4 B13 (*)
F4	B5 B6 (*) B17 (*)
F5	B2 (*) B26 (*)
F8	B14 (*) B15 (*) B19 (*) B27 B29 (*)
F10	B8 (*)
F12	B22 (*)
F13	B12 (*)
F15	B25 (*)
F16	B16
F17	B20

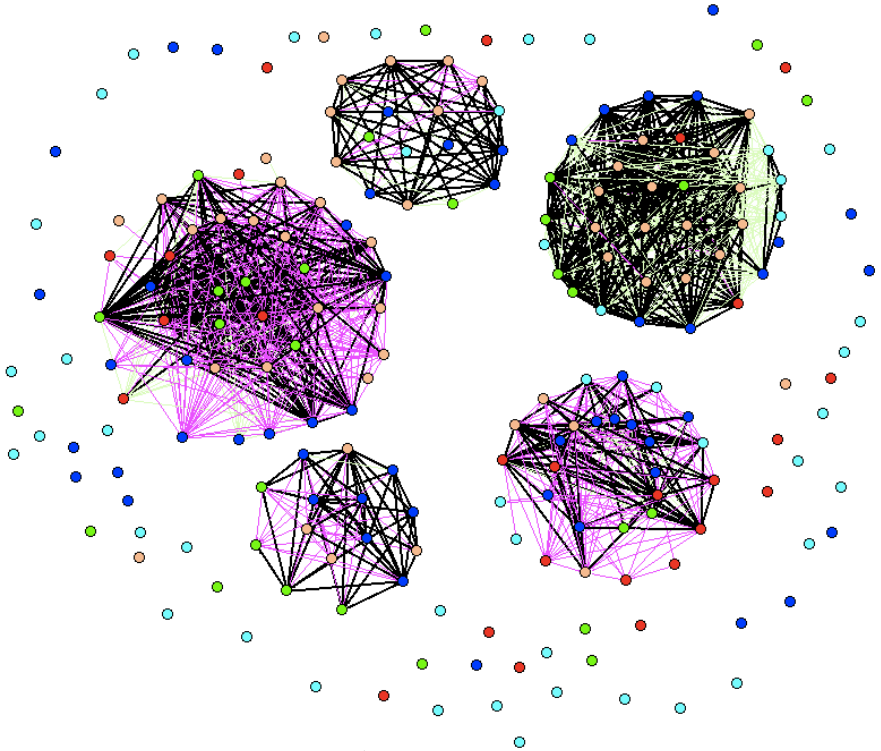
OCS

The percent of Bonferroni multi-links which are changing nature when detected into the FDR network is close to 37%.

However, 30% of them concerns the co-occurrence combinations C1 (16%) and C2 (14%).

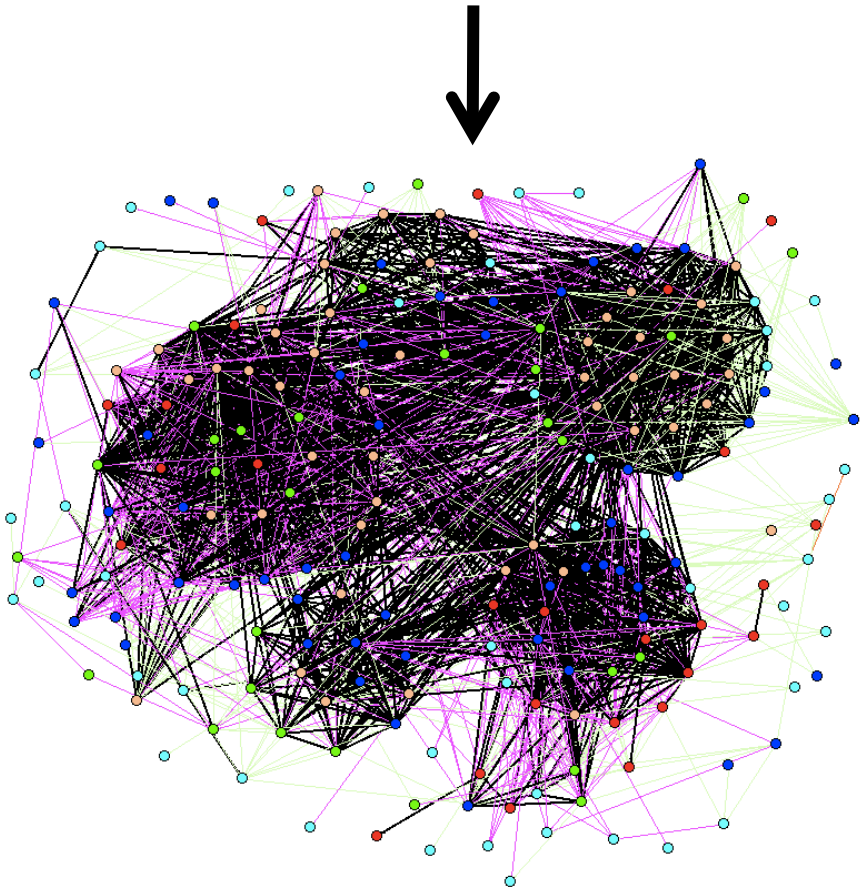
TABLE I: Most populated co-occurrence combinations in Bonferroni and FDR network

Label	Co-occurrence combination	Bonferroni (36664)	FDR (330404)	Color label
C1	(i_b,j_b)	7,716 (21.0)	120,655 (36.5)	magenta
C2	(i_s,j_s)	6,254 (17.1)	91,219 (27.6)	green
C3	(i_{bs},j_{bs})	1,732 (4.72)	19,227 (5.82)	apricot
C4	(i_b,j_b) (i_s,j_s)	20,243 (55.2)	66,692 (20.2)	black
C5	(i_b,j_{bs})	312 (0.85)	13,494 (4.08)	blue
C6	(i_s,j_{bs})	157 (0.43)	9,592 (2.90)	orange
C7	(i_s,j_b)	12 (0.033) *	2,662 (0.81)	tan
C8	(i_b,j_b) (i_s,j_s) (i_{bs},j_{bs})	137 (0.37) *	2,304 (0.70)	brown
C9	(i_b,j_{bs}) (i_s,j_{bs})	43 (0.12) *	1,414 (0.43)	purple

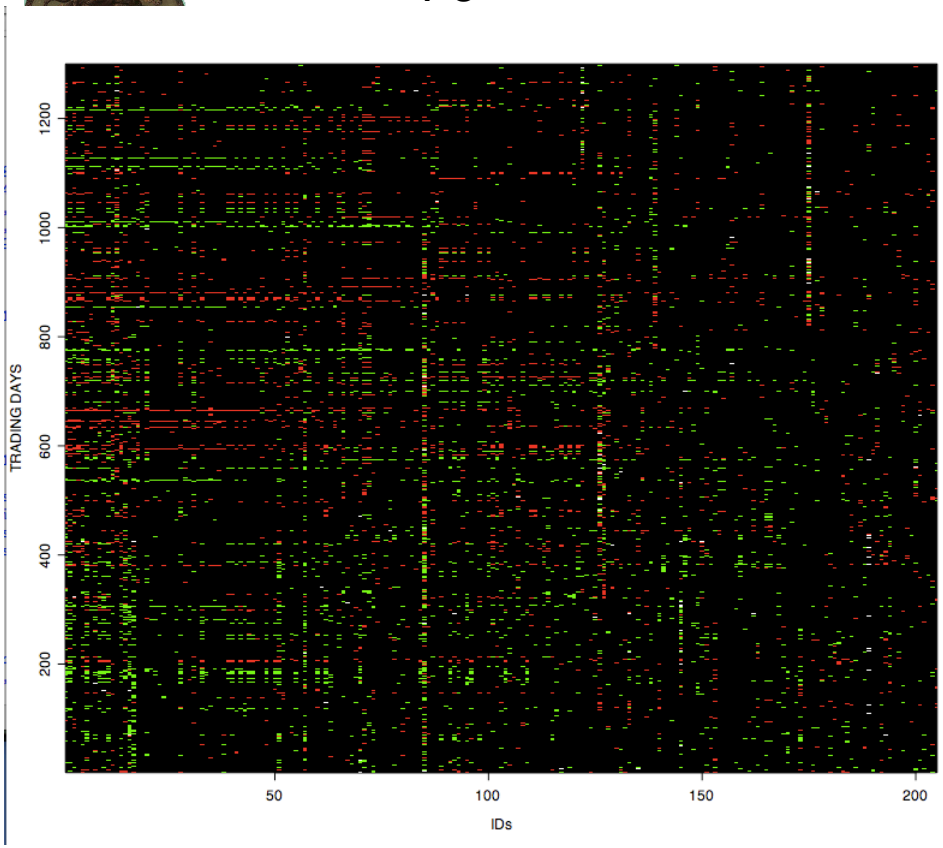


B14, B15, B19, B27 and B29 clusters of the Bonferroni network. Links are the link present in the Bonferroni network

F8 cluster of the FDR network. Links are the link present in the FDR network

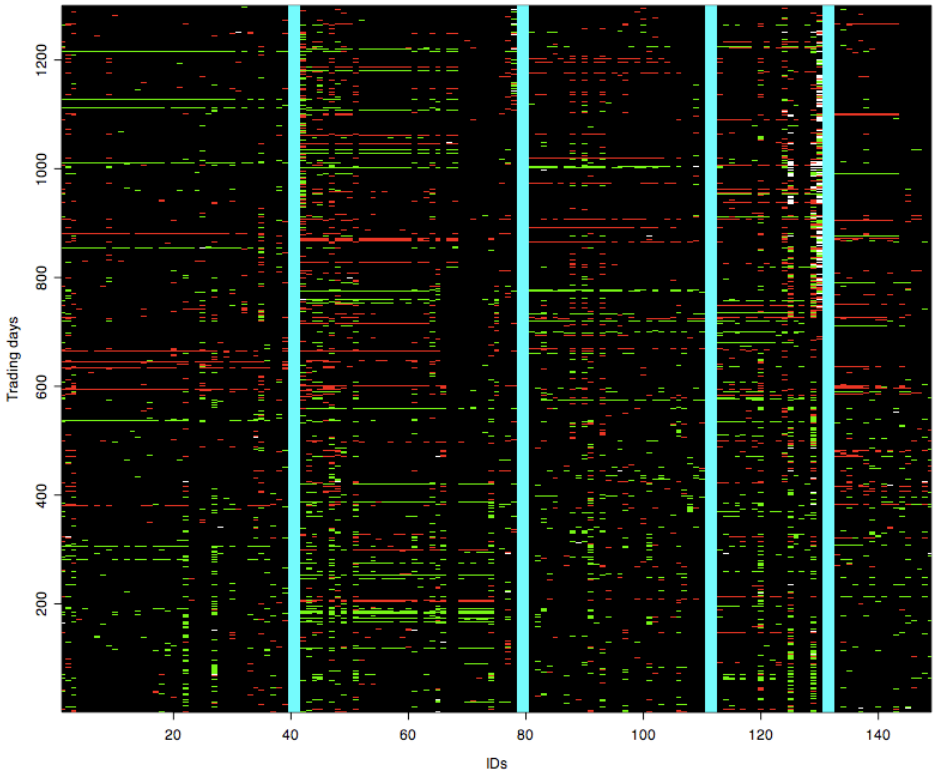


F8



Trading activity of the 205 investors of the F8 FDR cluster

B14 B15 B19 B27 B29



Trading activity of the 141 investors of the B14, B15, B19, B27 and B29 Bonferroni clusters. 135 elements are also in the F8 FDR cluster

OCS

Conclusions

- **In financial markets, investors' trading profile can be investigated down to the level of individual investors.**

- **An essential heterogeneity of investors is observed in the time dynamics of market activity.**

- **Resulting "strategies" are often persistent over very long period of times.**

- **Clusters of investors can be detected in heterogeneous systems.**