



Mapping individual behavior in financial markets: synchronization and anticipation

Mario Gutiérrez-Roig^{1,4*} , Javier Borge-Holthoefer², Alex Arenas³ and Josep Perelló^{4,5}

*Correspondence:

Mario.Gutierrez-Roig@wbs.ac.uk

¹Data Science Lab, Warwick Business School, University of Warwick, Coventry, UK

⁴Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona, Spain
Full list of author information is available at the end of the article

Abstract

In this paper we develop a methodology, based on Mutual Information and Transfer of Entropy, that allows to identify, quantify and map on a network the synchronization and anticipation relationships between financial traders. We apply this methodology to a dataset containing 410,612 real buy and sell operations, made by 566 non-professional investors from a private investment firm on 8 different assets from the Spanish IBEX market during a period of time from 2000 to 2008. These networks present a peculiar topology significantly different from the random networks. We seek alternative features based on human behavior that might explain part of those 12,158 synchronization links and 1031 anticipation links. Thus, we detect that daily synchronization with price (present in 64.90% of investors) and the one-day delay with respect to price (present in 4.38% of investors) play a significant role in the network structure. We find that individuals reaction to daily price changes explains around 20% of the links in the Synchronization Network, and has significant effects on the Anticipation Network. Finally, we show how using these networks we substantially improve the prediction accuracy when Random Forest models are used to nowcast and predict the activity of individual investors.

Keywords: Financial markets; Behavioral economics; Transfer of entropy; Mutual information; Networks

1 Introduction

Human collective behavior has been increasingly studied due to an unprecedented amount of data available from the digital world [1]. A new research topic has been thus opened to an extensive use of multidisciplinary strategies, that are aimed to dive into the empirics by using a wide variety of styles and techniques. Approaches in the literature to find dynamical patterns in data or even address fundamental research questions are today rich and diverse. Still, one of the most intriguing aspects that needs further understanding is the non-trivial relationship between individual actions and the aggregated bulk of actions of large collectivities [2].

Rather evident contexts where it is possible to study the phenomena are social networks. It is possible to observe coordination effects, amplifying for instance the impact of a street protest in a microblogging platform such as Twitter [3]. The links through which information flows can bring out macroscopic emergent patterns. However, other situations differ from this perspective, and then allow us to neatly focus on how the macroscopic signal leads to individual actions simply because there is no direct communication among the

individuals. This can also be considered the case of our dataset containing clients' activity from a trading firm whose orders have no significant impact in asset price evolution.

Within the study of collective behavior in financial markets, there are several lines of research [4, 5]: from computational agent-based models aiming to better understand phenomena such as herding behavior [6–9] to pure empirical analysis on investor's activity [10] or eventually through data-driven models [11]. Some of these studies focus on the bursty trading activity data [12, 13], and the impact of external information flows on price and then provide new indicators to measure the degree to which a particular news item attracts attention from investors [14]. Price shifts due to trading activity and order book imbalances are being studied observing universal patterns that link macroscopic price formation and individual market and limit orders placed in the order book [15, 16]. Tick-by-tick trading activity indeed describes a multifractal behavior explained by a highly heterogeneous nature of executed tasks, mostly due to the large diversity of investor's profiles [12, 13]. The marked peaks of trading activity and the clusters with very intense activity emerging between calm periods are also observed to be linked with the bursty evolution of market volatility [17, 18] which is a very relevant indicator in traders decisions mechanisms. The investor's behavior is also behind the interpretation of the non-trivial market phenomena, such as the leverage effect where daily price drops increase volatility of the following few days [19, 20].

The non-stationary nature of the financial series, together with the fact that investors are heterogeneous, meaning for instance that they operate at different volume scales and time-horizons, asks for a careful analysis and the application of the most adequate techniques. It is precisely under this context where non-parametric statistics deploys all its powerful methods. Thus, our analysis is mostly grounded on Mutual Information [21] and Symbolic Transfer of Entropy [22] (STE), which allows to quantitatively study individual behavioral aspects, like synchronization and information flows, key elements to identify higher properties like structural hubs, coordinated communities, critical transitions or sudden collapses [23]. STE analysis is in fact a rather new tool in the context of financial markets, which has mostly being used to analyze cross-market effects [24, 25] and to identify dynamic causal linkages as a way to complement other techniques such as network analysis [26, 27], which might have important consequences in optimizing portfolio composition. In this sense, Mutual Information and mostly STE respectively represent an alternative approach to statistically validated synchronous networks [28] and its much more recent evolution under the form of statistically validated lead-lag networks [29, 30]. These two methods have already been explored recently in the context of financial market at a nanolevel with trader-resolved data [29–33].

Unfortunately, data records at individual level are not easily available for research purposes, what limits scientists in the exploration of this crucial aspect of financial markets dynamics. For this reason, this work makes the database accessible in order to support research activity in a field that still lacks extensive exploration. One of the first researchers to look in this class of data was Terrance Odean who, in 1998, after studying the performance of 10,000 accounts where individual activity is available, proved that investors mostly sell the winning stocks, while keeping the losers [34]. A subsequent study by the same author [35] was also analyzing return patterns and investor's purchases finding that overall trading for a particular group of investors is excessive. In the 2000's Grinblatt and Keloharju took advantage of transparent Finnish stock market, where traders' IDs are recorded in

every transaction, and used a database of this market to analyze the performance of different types of traders, categorized as pro-momentum or contrarians in a first study [36]; and sensation seekers or overconfident traders subsequently [37]. Other efforts [11] were made with clients database from one of the greatest on-line Swiss broker which found empirical relationships between turnovers (contrarian strategies) account values and number of assets in which a trader is investing.

Tumminello et al. [31] made a first attempt in 2012 to identify clusters of investors in the Finnish market with statistically validated synchronous networks [28] and this effort has also served to go deeper in trading profiles identification [38]. More recently same methods have been applied to the clusters of investors with similar trading profiles in a robust and reliable way understand their long-term ecology based on what Musciotto et al. call adaptive market hypothesis [33] or even to study systemic risk [32]. Other recent works explore the possibility to find trading similarities of Swedish investors with similar portfolios [39], while Lillo et al. [40] have also investigated how news (an exogenous signal) affect the trading behavior of different categories of investors or even how different. Pairwise synchronization between traders' activity is been used to detect communities and define groups of traders. Recently, Challet et al. [29] infer lead-lag networks to predict the sign of the order flow and the volume weighted average price of broker clients over the next hour. And even more recently Cordi et al. [30] use the same methods to give reason of asset price time reversal asymmetry.

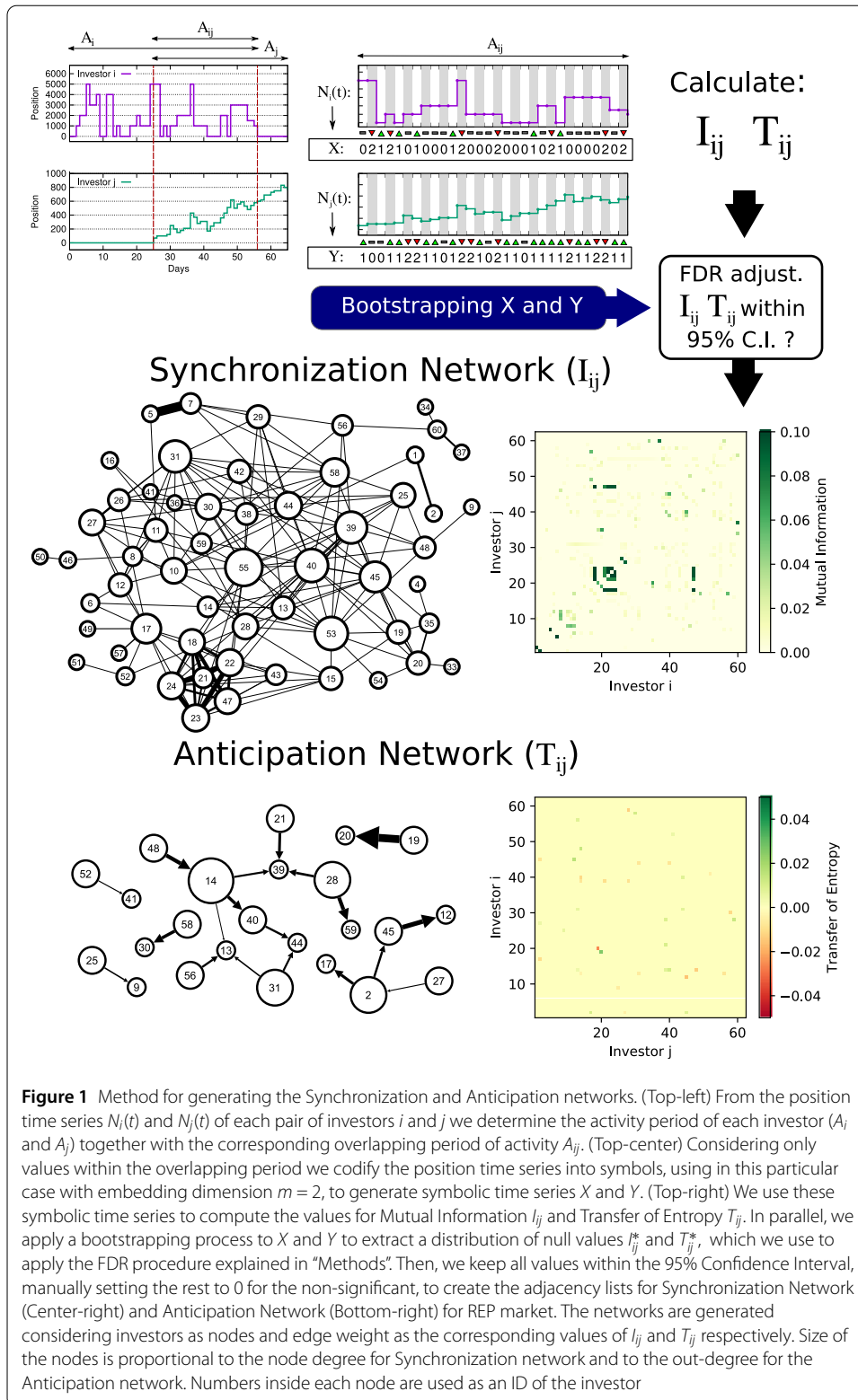
2 Methods

In this section we present and thoroughly describe the process to build the Synchronization and Anticipation Networks from the raw data of investors' performance. All the process is summarized in Fig. 1.

2.1 Data

Most of the markets do not allow traders to directly access the market and their orders are placed through trading firms. Our raw database reports trading activity of 29,930 clients, who are small size non-professional traders that invest their own savings. All their decisions are genuinely human and not taken by any kind of algorithmic trading robot, although it is possible that some of them are influenced by some external factors. Investors traded over 120 different assets in the BME Spanish stock exchange from 01/01/2000 until 12/31/2008 (1969 trading days). This period does not present a general global trend, although the initial range (2000–2002) has been qualified by experts as a *bearish* period (that is: Spanish market prices had an overall negative trend during that period); while the subsequent one (2003–2008) has been considered *bullish* (that is: Spanish market prices had an overall positive trend during that period). In total, the dataset contains 3,303,695 transactions, where each record includes the ID of the client, the ID of the investment firm's associated manager, the date of the transaction, the price of the asset, and the number of shares being sold or bought. Our database contrasts with previous studies from Finnish market where records compiled trading activity of all actors in the trading floor (including households and financial corporations) [33, 38, 40]. Our database keeps some similarity with the one being used in Refs. [11] and [31].

Since our interest is to map a collectivity of investors based on their individual performance, our database needs to fulfill two general criteria: (i) bearing enough data from each



investor, so that behavior at both individual and aggregate levels can be studied. Therefore, a first filter consists on limiting the analysis to those investors for whom their daily position balance (number of shares bought minus number shares sold) is different from

zero for at least 20 days. And (ii), for those investors passing the first restriction, we consider only transactions over the 8 most traded assets, that contain at least 30 investors after filtering. Those assets are very heterogeneous regarding their number of transactions, but also in terms of the business sector of the companies. In decreasing order, these are: Telefonica (TEF), from communications sector, with 415 most active investors accounting for 131,518 transactions; Santander (SAN), from finance sector, with 219 most active investors accounting for 71,463 transactions; BBVA (BBVA), from finance sector, with 113 most active investors accounting for 53,388 transactions; Endesa (ELE), from utilities sector, with 86 most active investors accounting for 45,468 transactions; Ezentis (EZE), from industrial sector, with 88 most active investors accounting for 31,207 transactions; Zeltia (ZEL), from health care sector, with 71 active investors accounting for 19,021 transactions; Repsol (REP), from energy sector, with 62 active investors accounting for 36,354 transactions; Gas Natural (GAS), from utilities sector, with 30 active investors accounting for 22,193 transactions.

In summary, and considering that a single investor can trade with different assets, we have analyzed the performance of up to 566 different individuals accounting for 410,612 transactions. While these quantities are large enough to study a community of investors, it is nonetheless impossible that this collectivity has any impact on market price of any of specified assets, specially considering the daily total volume traded for any of them in the Spanish stock market. Consequently, price signal should never be considered as something endogenous or generated by these communities of investors.

The filtered dataset is publicly accessible as described in “Availability of data and materials” section.

2.2 Performance time series and activity periods

Comparing the behavior and performance between two investors only makes sense when they hold or are trading with the same asset. Therefore, in our analysis we are going to treat each of the asset managements as a different and separated scenario. Thus, for each asset we define $N_i(t) > 0$ as the total number of shares that investor i is holding at the end of day t . Equivalently, $\Delta N_i(t) \equiv N_i(t) - N_i(t - 1)$ is the daily cumulative change in her position, size of assets bought minus size of assets sold, by that particular investor during the day t . Thus, if $\Delta N_i(t) > 0$, her trading volume is dominated by buying orders; selling orders are predominant if $\Delta N_i(t) < 0$. Also, note that $\Delta N_i(t) = 0$ does not imply necessarily that investor i hasn't traded in the day t : it might well be that she has behaved like an intra-day trader holding the same number of shares at the beginning and at the end of the day. Alternatively, since our data resolution is at daily level, the information is more relevant when $\Delta N_i(t) \neq 0$ because it implies not only that individual i has been active but also that her decisions incorporate a specific market daily orientation.

Once $N_i(t) > 0$ is determined for every investor, her activity period A_i can also be defined as the time period from its first until last recorded transaction. When comparing two different investors i and j , respectively with activity periods A_i and A_j , we constrain the analysis to the overlapping period between both investors A_{ij} . To avoid any confusion in further discussions, we here want to point out that the activity joint period A_{ij} is not an element of a matrix. In the case that the overlapping period is smaller than 20 days ($A_{ij} > 20$), the pair of investors is considered as non-contemporary and measures for synchronization and anticipation are not computed for this specific pair of investors. In order

to ensure a sample size big enough, we also disregard the pair of investors if within the overlapping time period any of the two investors does not show activity, i.e. the position does not change, for at least 20 days.

2.3 Symbolization, mutual information and transfer of entropy

Considering the nature of the time series $N_i(t) > 0$, we cannot assume that these are linear nor stationary. Moreover, the strong disparity in their nature invites us not to use any kind of analysis grounded on linear assumptions [41], and choose instead more sophisticated tools [42, 43] which can handle implicit non-linear dynamics. In this context, symbolization seems appropriate when it comes to compare agents' behavior, regardless of their capital or typical transactions size. We thus adopt the framework of Bandt and Pompe [44] to symbolize the investor's position $N_i(t) > 0$ in order to compute Symbolic Mutual Information (SMI) and Symbolic Transfer of Entropy (STE) [22] between investors later on. We also introduce a new important feature in the symbolization process due to the nature of our time series: here we consider an additional symbol representing unchanging values in $N_i(t)$, that is when $\Delta N_i(t) = 0$. In their work, Bandt and Pompe neglected unchanging values in the series, because their fluctuations were generated by a continuous distribution. That is, the probability to observe a chain of constant values was negligible. However, in the present case the situation is quite the opposite, where $N_i(t) = N_i(t + 1)$ is a common situation (see Figure S1).

In order to preserve the original nomenclature, we redefine the original daily time series for two investors i and j as

$$\begin{aligned} N_i(t) &\rightarrow X = \{x_0, x_1, \dots, x_t, \dots, x_n\}, \\ N_j(t) &\rightarrow Y = \{y_0, y_1, \dots, y_t, \dots, y_n\}, \end{aligned} \quad (1)$$

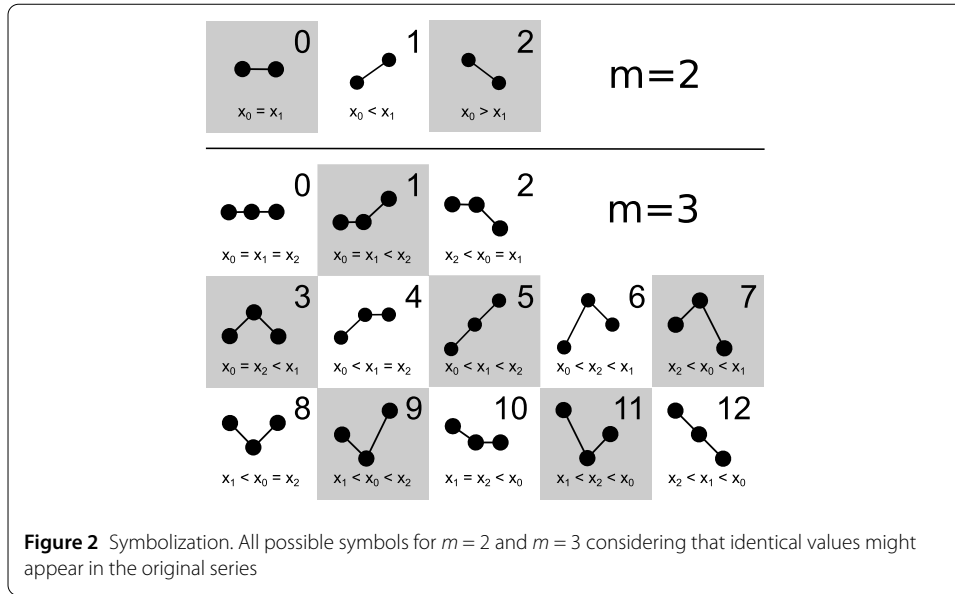
being t within the overlapping activity period A_{ij} , and sub-indices 0 and n representing the first day and last day of this period, respectively. From here, we can transform these numerical time series into a series of symbols that depend on sub-pieces of consecutive numerical values. The length of these pieces is given by the embedding dimension m , which in turn defines the number of possible symbols (see Fig. 2). We can thus read

$$\begin{aligned} X &\rightarrow \hat{X} = \{\hat{x}_0, \hat{x}_1, \dots, \hat{x}_t, \dots, \hat{x}_n\}, \\ Y &\rightarrow \hat{Y} = \{\hat{y}_0, \hat{y}_1, \dots, \hat{y}_t, \dots, \hat{y}_n\}, \end{aligned} \quad (2)$$

where hat represents the fact that series are now codified in symbols, instead of the original numbers. Now, according to the definition of Shannon [21], we compute Symbolic Mutual Information (SMI) as

$$I(\hat{X}, \hat{Y}) = \sum_{\hat{y}} \sum_{\hat{x}} p(\hat{x}_t, \hat{y}_t) \log_m \frac{p(\hat{x}_t, \hat{y}_t)}{p(\hat{x}_t)p(\hat{y}_t)}, \quad (3)$$

where the sum is over all symbols, $p(\hat{x}_t, \hat{y}_t)$ is the joint probability that two specific symbols appear together and $p(\hat{x}_t)$ and $p(\hat{y}_t)$ are the marginal probabilities. If both series \hat{X} and \hat{Y} are independent, then $I(\hat{X}, \hat{Y}) = 0$ which means that both investors i and j performances are unrelated. Instead, if i and j are completely synchronized, $\hat{x}_t = \hat{y}_t$ ($\forall t$), $I(\hat{X}, \hat{Y})$ will take



the maximum value which depends on the number of symbols and the embedding dimension m .

Similarly, Symbolic Transfer of Entropy (STE) [22] between investors i and j can also be computed. Thus, STE from \hat{Y} (investor j) to \hat{X} (investor i) reads

$$T(\hat{Y} \rightarrow \hat{X}) = \sum_{\hat{y}} \sum_{\hat{x}} p(\hat{x}_{t+1}, \hat{x}_t, \hat{y}_t) \log_m \frac{p(\hat{x}_{t+1} | \hat{x}_t, \hat{y}_t)}{p(\hat{x}_{k+1} | \hat{x}_t)}. \tag{4}$$

The sum is again over all symbols, and now both joint probability $p(\hat{x}_{k+1}, \hat{x}_k, \hat{y}_k)$ and conditional probabilities $p(\hat{x}_{k+1} | \hat{x}_k)$ and $p(\hat{x}_{k+1} | \hat{x}_k, \hat{y}_k)$, include a third element that considers certain time delay by shifting events one-day ahead. Thus, in order to assess the direction of the entropy transfer flow we need to calculate

$$T(\hat{X}, \hat{Y}) = T(\hat{Y} \rightarrow \hat{X}) - T(\hat{X} \rightarrow \hat{Y}), \tag{5}$$

where a positive value means that \hat{Y} (investor j) is anticipating with respect to \hat{X} (investor i), and the opposite for negative values. It is important to remark that here we are using the concept of Transfer of Entropy as a tool to reveal information flows and predictive power between variables. Since we do not have any access to the complete context and circumstances of all investors, we cannot therefore use it to establish any causal relationship between them [45, 46].

Finally, we need to determine the embedding dimension m , i.e. the number of consecutive daily records considered to generate all possible symbols. In this work we initially tested both $m = 2$ and $m = 3$, generating time series with 3 and 13 symbols respectively. However, given the daily nature of our time series, the results for $m = 3$ were very noisy and the networks barely had any significant link. Notwithstanding, $m > 2$ could still be useful when applied to a longer time series or with a higher frequency because could lead to a more refined study. Hence, we report here results for $m = 2$, what leads to encode the time series for the position using three different kind of symbols: positive change in position $N_i(t) > 0$ (\uparrow), negative change in position $N_i(t) < 0$ down (\downarrow) or null change in position

$N_i(t) = 0$ or price ($-$). Note that for the specific case of $m = 2$ we generate networks very similar than co-occurrence networks [28, 31] or lead-lag networks [29]. However, there is an important difference between lead-lag and the anticipation networks based on Transfer of Entropy. The former ones build a co-occurrence network over a pair of time series where one is lagged with respect to the other, whereas the Transfer of Entropy considers not only the lag with respect the second time series but also the lag of the first one (see the conditional probabilities in Eq. (4)). This allows to measure the neat flow of information between two time series.

2.4 Bootstrapping and network construction

Once I_{ij} and T_{ij} have been calculated for each pair of investors we carry out a bootstrapping process of 10,000 iterations in order to establish the significance level of each link. In each of those iterations we shuffle the symbolized series of the investors position in the overlapping period A_{ij} , \hat{X} and \hat{Y} , and subsequently compute the corresponding value for the Mutual Information and Transfer of Entropy, I_{ij}^* and T_{ij}^* respectively. We determine the significance of the original values for I_{ij} and T_{ij} based on such distribution. Since this implies multiple hypothesis testing, we must control the false positive rate and adjust the p -values accordingly. Here we use the FDR controlling procedure called Benjamini–Hochberg (from here codenamed as “FDR”) and FWER controlling procedure called Bonferroni correction (from here codenamed as “Bonferroni”). These two procedures are very standard and also used in similar cases through statistically validated networks in [31] and [29].

Bonferroni correction modifies the original significance threshold $\alpha = 0.05$, setting it to α/m , where m is the number of independent test performed, which in our case is the number of pairs of investors with an overlapping time-window A_{ij} that fulfills the conditions defined above. We then sort the distribution of the shuffled values for I_{ij}^* and T_{ij}^* and set the significance thresholds given by “Bonferroni”, one-sided for the case of Mutual Information and two-sided for Transfer of Entropy. If the original value is outside those intervals we keep it otherwise we manually set it to 0.

Whereas Bonferroni correction can be very conservative, other criteria such as Benjamini–Hochberg, can still control the false positive rate but in a less strict way that allows for more true positives. This method is based on sorting the p -values for the Mutual Information and Transfer of Entropy, and then consider significant all the p -values smaller than the largest p -value fulfilling

$$p^{(k)} \leq \frac{k}{m} \alpha, \quad (6)$$

where $p^{(k)}$ is the k th p -value, m the number of investors pairs, and $\alpha = 0.05$ the original significance threshold. This method requires first to compute the actual p -values. Such task is not trivial since the proportion of the symbols in the underlying series might modify the mean and variance of the distribution of the null values, as we demonstrate in the Figure S2 of Additional file 1. The strategy we follow here consist on computing the mean and standard deviation of the shuffled values. We then determine the p -value of the original I_{ij} from a gamma distribution [47] and the p -value of the original T_{ij} from a normal distribution. In all cases, for each pair of investors we parametrize those functions with the mean and standard deviation of the distribution for I_{ij}^* and T_{ij}^* respectively.

Finally, adjacency matrices are built from I_{ij} and T_{ij} quantities to generate Synchronization and Anticipation networks as Fig. 1 shows. In the first case we obtain a weighted undirected network whose nodes represent investors and edges how synchronized they are. In the second, we obtain a weighted directed network whose nodes represent investors, and arrows indicate who anticipates whom.

3 Results

A general overview of network properties can be observed in Tables 1 and 2, while the adjacency matrices of all networks are shown in Figures S3–S11 in Additional file 1. There are some important differences between Synchronization and Anticipation networks at the structural level, apart from the fundamental fact that the former is undirected whereas the latter is directed. The first difference is related to the number of edges, and therefore to the average degree. The synchronization networks are much denser, which means that finding a pair of synchronized agents is much more common than finding an investor that anticipates another. As for the degree distributions, almost all of them significantly deviate from a Poisson distribution, associated to random networks, creating more high connected groups and hubs than someone would expect in the random case. Consequently,

Table 1 Synchronization Network Features. Number of nodes, edges and average degree is shown in the first three columns for all networks. Fourth column refers to a Null Hypothesis testing based on Kolmogorov–Smirnov (KS) statistic for rejecting the hypothesis that underlying distribution for node out-degrees is a Poisson distribution, only p -value is shown. Fifth and sixth column show the Clustering Coefficient and Degree Assortativity of the undirected graph. The networks in this table have been built using “FDR” as described in methods section. Similar results can be found when using Bonferroni (see table S1)

Asset	Number of nodes	Number of edges	Average degree	Poisson KS test p -value	Clustering coefficient	Assortativity coefficient
TEF	400	8774	43.87	$<10^{-5}$	0.38	0.14
SAN	210	2086	19.87	$<10^{-5}$	0.32	0.10
BBVA	92	397	8.63	0.00036	0.37	0.00
ELE	77	314	8.16	$<10^{-5}$	0.36	0.20
EZE	77	180	4.68	0.00201	0.21	0.00
ZEL	60	173	5.77	0.08251	0.31	-0.01
REP	58	179	6.17	0.00220	0.32	0.13
GAS	25	55	4.40	0.63592	0.27	-0.01

Table 2 Anticipation network features. Number of nodes, edges and average degree are shown in the first three columns for all networks. Fourth column refers to a Null Hypothesis testing based on Kolmogorov–Smirnov (KS) statistic for rejecting the hypothesis that underlying distribution for node out-degrees is a Poisson distribution, only p -value is shown. Fifth column shows the Degree Assortativity of the directed graph. The networks in this table have been built using “FDR” as described in methods section. Similar results can be found when using Bonferroni (see table S2)

Asset	Number of nodes	Number of edges	Average out-degree	Poisson KS test p -value	Assortativity coefficient
TEF	326	777	2.38	$<10^{-5}$	0.01
SAN	130	140	1.08	$<10^{-5}$	0.05
BBVA	44	30	0.68	$<10^{-5}$	-0.13
ELE	38	30	0.79	$<10^{-5}$	-0.26
EZE	27	18	0.67	$<10^{-5}$	0.04
ZEL	24	16	0.67	$<10^{-5}$	0.10
REP	24	19	0.79	0.00005	0.33
GAS	2	1	0.50	0.30964	0.00

synchronization networks also reveal a clustering coefficient systematically greater than what would correspond to a random graph, i.e. the average degree divided by number of nodes [48] (ranging from 0.06 to 0.18 in this case). Such structural feature might arise because if investor i is synchronized with investor j , and j is synchronized with k , it is likely that i and k are going to be synchronized as well. As for the assortativity coefficient, synchronization networks present 4 cases of relatively high assortative networks, whereas in the anticipation networks we observe that 2 of them are very disassortative and 1 very assortative, while the rest present values equivalent to random networks.

In the next sub-sections we explore some of the possible explanations for the creation of links in those networks.

3.1 Measuring individual reaction to price

One of the immediate candidates to be a behavior driver is the reaction of each of the investors to price. In order to measure it we apply the same methodology explained above, but instead of two investors position time series we consider for each investor her position and the price. Both Mutual Information between investors' position and price, I_{ip} , and Transfer of Entropy between investors' position and price, T_{ip} , can be computed from symbolized series, as well as the bootstrap procedure applied for significance tests (see Fig. 3). As before, we consider the price time series only within the activity period of each investor A_i and apply the FDR adjust in order to correctly set the significance threshold.

The first element to notice from distributions of aggregated values in Fig. 3 and detailed by markets in Table 3 is the amount of significant values. Such results reinforce the idea that prices at t and $t - 1$ are definitely good candidate drivers for investors' behavior, as some other studies with different approaches have pointed out [49]. Indeed, our investor population sample is found to be very sensitive to either today or yesterday's price change. From 1074 investors considered, a majority of them 697 (64.90%) shows significant values for the synchronization with price, while some of them 47 (4.38%) significantly react to what the market did in the previous day. Such result is consistent across all studied assets. Notice that, by keeping most active investors in order to guarantee enough statistics to compute SMI and STE, we can also be filtering out less active investors and keeping the most active ones and therefore more likely to react to immediate changes. As for the STE between price and investors position, we observe in Table 3 a systematic deviation towards negative values, which is consistent with the fact that some of the investors can be driven by price, but hardly the performance of any investor can anticipate the price. The single case in TEF market can be considered a false positive because when we apply more strict methods, such as Bonferroni, there are no investors anticipating the price at all.

These measures of individual reaction to price are features that might drive investor's behavior and could be important to explain part of the edges in Synchronization and Anticipation networks between investors. While we cannot test causal relationships (limited historical data), we can at least provide a measure of how much could be explained based on individual behavior features.

The next two sections address this issue by quantifying how much of this connections in the networks could be generated by certain degree of coincidence in the way investors react to price.

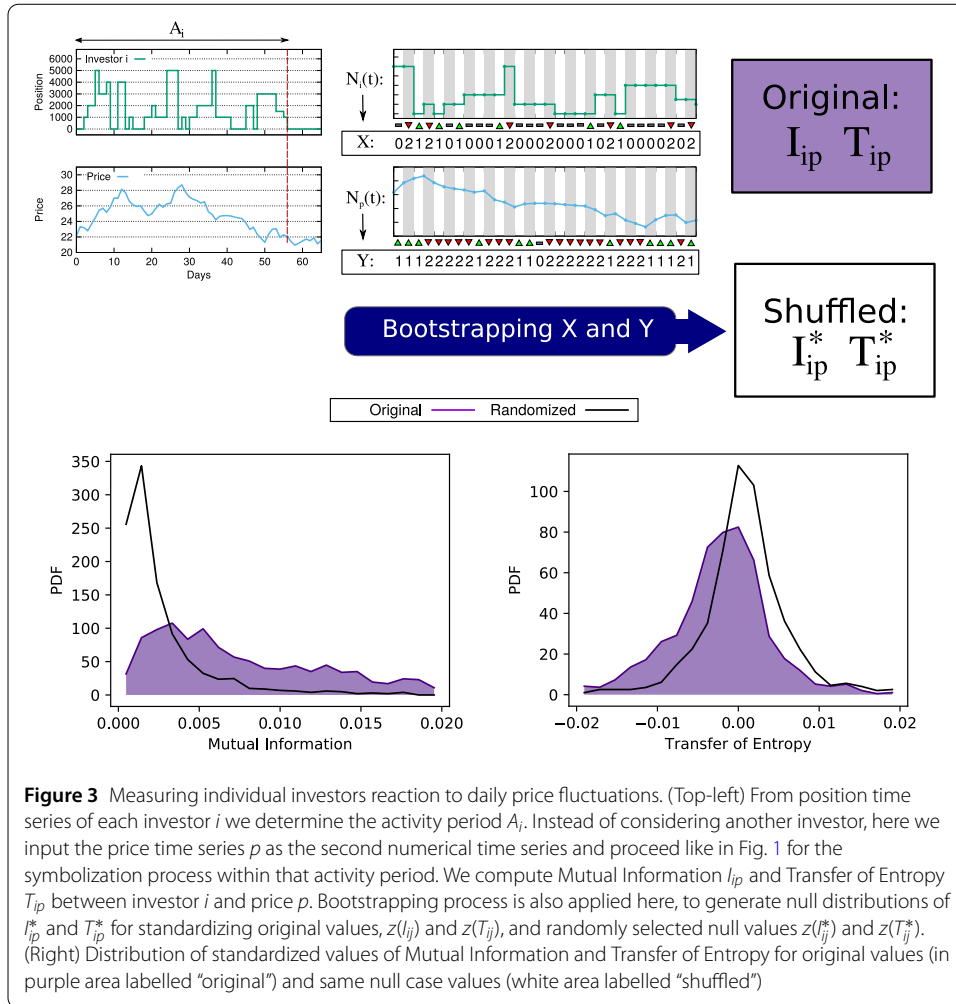


Figure 3 Measuring individual investors reaction to daily price fluctuations. (Top-left) From position time series of each investor i we determine the activity period A_i . Instead of considering another investor, here we input the price time series p as the second numerical time series and proceed like in Fig. 1 for the symbolization process within that activity period. We compute Mutual Information I_{ip} and Transfer of Entropy T_{ip} between investor i and price p . Bootstrapping process is also applied here, to generate null distributions of I_{ip}^* and T_{ip}^* for standardizing original values, $z(I_{ij})$ and $z(T_{ij})$, and randomly selected null values $z(I_{ij}^*)$ and $z(T_{ij}^*)$. (Right) Distribution of standardized values of Mutual Information and Transfer of Entropy for original values (in purple area labelled “original”) and same null case values (white area labelled “shuffled”)

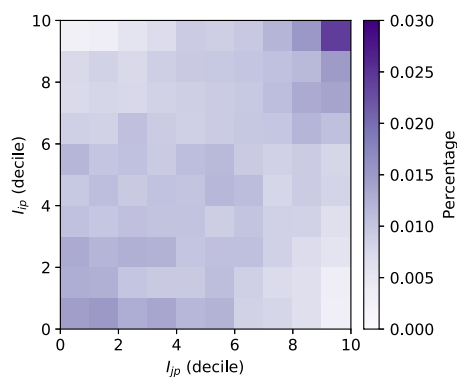
3.2 Individual reaction to price effect on synchronization network

If two investors react in the same way to the current price, chances are that they will be connected in the synchronization network. In that case, reaction to price would work as a hidden variable that explains the significant value for synchronization measured by I_{ij} , instead of a direct interaction. We can quantify such scenario by using non-parametric statistics, so without having to assume any kind of distribution for the reaction to price. Thus, we compute I_{ip} values for all investors and divide the distribution in deciles, so that each bin contains exactly 10% of the investors according to I_{ip} . We then create a 10×10 matrix with all possible investor–investor combinations in terms of deciles. Now, in the cell we calculate the fraction of the edges that go from one investor i (origin) with her corresponding I_{ip} and assigned to a certain decile, to another investor j (destination) with her corresponding I_{jp} and assigned to a specific decile. By construction, if the synchronization with price had no effect on the structure of the network, each cell should contain around 1% of all edges regardless of I_{ip} or I_{jp} of the nodes. Instead, Fig. 4 shows an uneven pattern with the most populated cells along the diagonal, with the top-right corner being the area where the effect is stronger. This basically means that investors who are synchronized between each other tend to have similar values for the synchronization with the price. This effect is even stronger when the synchronization with price is high. Thus, the top-left cor-

Table 3 Number of individuals with significant Symbolic Mutual Information (SMI) and Symbolic Transfer of Entropy (STE). The SMI I_{ip} and STE T_{ip} are calculated over the complete investor's activity period A_i in eight assets of the BME Spanish stock exchange. Significant values for SMI and STE are computed using the FDR approach described in the section "Methods". Additional file 1 contains the same table but using the Bonferroni approach showing very similar results. First column gives the number of individuals with >95% confidence interval of sharing pattern with price evolution (among the total number individuals, in parenthesis). Three last columns present the equivalent analysis for STE. The number of individuals being in the <2.5% confidence interval are those showing a significant $T_{ip} < 0$ (they are one-day delayed with respect to price change) while the number of in the >97.5% Confidence Interval are those with a significant $T_{ip} > 0$ (they are anticipating price change). The column labelled as "NS" accounts for the number of individuals that do not present any significant STE between performance and price

Asset	Investors	T_{ip}			
		I_{ip} >95%	<2.5%	NS	>97.5%
TEF	146	259	33	371	1
SAN	57	162	8	211	0
BBVA	26	87	0	113	0
ELE	24	62	1	85	0
EZE	48	40	0	88	0
ZEL	44	27	0	71	0
REP	24	38	1	61	0
GAS	8	22	4	26	0
ALL	377	697	47	1026	1

Figure 4 Synchronization with price as driver of synchronization between investors. Axis x and y represent deciles so that generate a 10×10 matrix. Each cell coordinates are given by the deciles of I_{ip} and I_{jp} respectively, and the value showing refers to the frequency of events with statistically significant Mutual Information I_{ij} , i.e. all not null edges of the Synchronization network



ner cell concentrates the highest number of cases with a total 2.42% of cases and thus deviating 1.58σ 's from the uncorrelated randomized null case. The effect is even accentuated when using the Bonferroni method to build the networks (see Figure S12), deviating 3.23% (1.32σ). Finally, the cumulative deviation across all cells is of 19.35%. Therefore we claim that the effect of the investors synchronized with price explains around one fifth of the links between investors in the synchronization network.

3.3 Individual reaction to price effect on anticipation network

In a similar way, we measure the effects of individual reaction to price in the Anticipation network. Thus, if we consider the case where an investor i anticipates investor j on a daily basis ($T_{ij} > 0$), there are two main reasons (among the ones that can we measure here) that consider the reaction to price as a possible origin of such anticipation. First scenario is where investor i is synchronized with the price whereas j is delayed. This one-day offset might generate significant values for anticipation of i with respect to j . The second scenario is where investor i anticipates price while j is synchronized to price. Although

Table 4 Influence of individual reaction to price over the Anticipation network. The table is divided in four groups across the 8 different markets plus the aggregation of all them. For all significant links from i to j ($T_{ij} > 0$) for each network (market), we computed the probability of i being more synchronized with price than j (first group), i being less synchronized than j (second column), i having a higher value for the STE with respect to price than j (third group) and i having a higher value for the STE with respect to price than j (fourth group). FDR approach described in “Methods” section is used here to discriminate the significant links. Additional file 1 contains an equivalent table using Bonferroni instead of FDR method displaying similar results. Numbers between brackets account for the frequency. Asterisks refer to different confidence interval levels, * for 90%, ** for 95%, and *** for 99%

Asset	$\rho(I_{ip} > I_{jp} T_{ij} > 0)$	$\rho(I_{ip} < I_{jp} T_{ij} > 0)$	$\rho(T_{ip} > T_{jp} T_{ij} > 0)$	$\rho(T_{ip} < T_{jp} T_{ij} > 0)$
TEF	0.54*** (420)	0.46*** (357)	0.50 (390)	0.50 (386)
SAN	0.57* (80)	0.43* (60)	0.54 (76)	0.46 (64)
BBVA	0.60 (18)	0.40 (12)	0.43 (13)	0.57 (17)
ELE	0.70** (21)	0.30** (9)	0.43 (13)	0.57 (17)
EZE	0.61 (11)	0.39 (7)	0.56 (10)	0.44 (8)
ZEL	0.69 (11)	0.31 (5)	0.44 (7)	0.56 (9)
REP	0.63 (12)	0.37 (7)	0.63 (12)	0.37 (7)
GAS	0.00 (0)	1.00 (1)	1.00 (1)	0.00 (0)
ALL	0.56*** (573)	0.44*** (458)	0.51 (522)	0.49 (508)

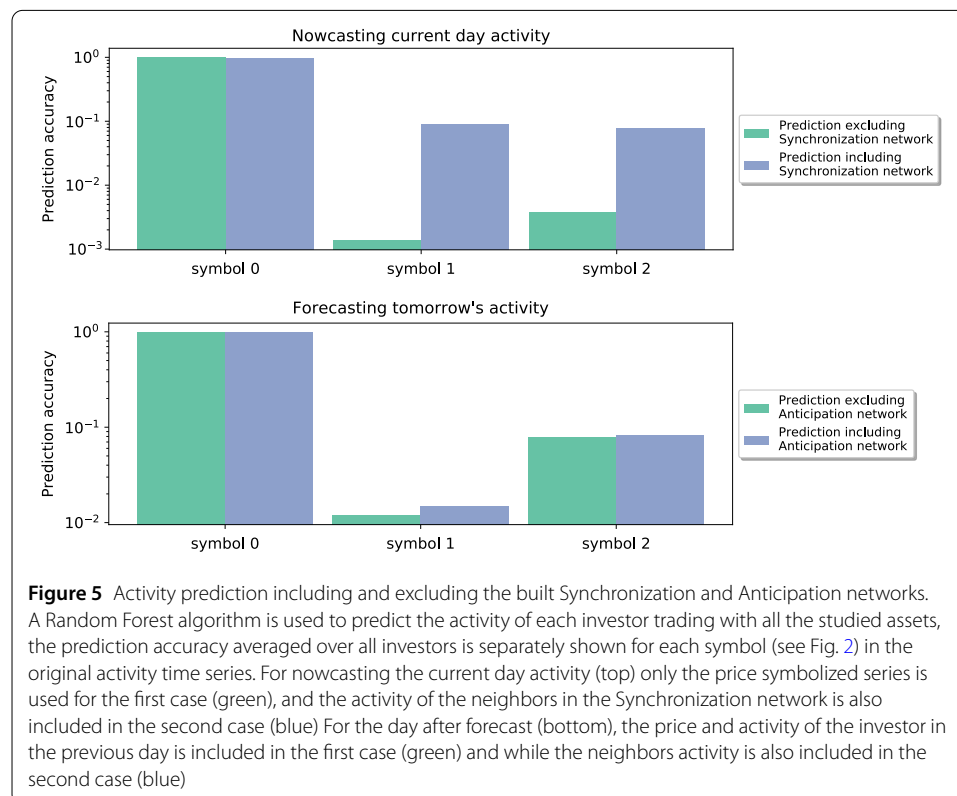
more restrictive (given the low number of users anticipating to price), it could enable the generation of one-day anticipation of i with respect to j . Both scenarios can be tested by computing the probability of the event $I_{ip} > I_{jp}$ (scenario 1) and $T_{ip} > T_{jp}$ (scenario 2) for each link in the anticipation network ($T_{ij} \neq 0$) within the period A_{ij} . Note that in the null case, where there was no influence of those variables, the probability should be 0.5. Table 4 shows that effects are limited but still significant, especially in the first scenario. When pooling together all the edges of all networks, results show a significant deviation (at 99% C.I.) from the null case of nearly 6% for the first scenario. This result is systematically consistent across all assets considered. As for scenario 2, we do not obtain significant results.

3.4 Synchronization and anticipation networks improve investors’ activity prediction

Predicting the activity of investors can be very challenging, specially considering the high degree of heterogeneity in the activity levels. Even keeping the most active investors, as detailed in the methods section, still the fraction of symbols for $m = 2$ that represent no significant activity is around 96% (Figure S1). Similar levels of sparseness can be found in other real datasets [50, 51] widely used to test all kinds of machine learning classifiers and recommender systems. Indeed, sparseness is a big challenge for machine learning algorithms that heavily rely not only on high amount of records but also on the density (non-zero instances) of the dataset [52]. In our particular case, the symbolization of the time series transforms the problem from a regression type of machine learning problem to a multi-class machine learning problem. Thus, Random Forests (RF) become one of the natural choices for predicting such kind of data. RF perform a multi-class prediction with a low risk of overfitting [53] while have been tested on sparse datasets like in language modeling [54]. The aim of this exercise is not to successfully develop a very accurate method to predict investor’s behavior but to demonstrate how using the synchronization and anticipation networks developed above substantially improves the prediction accuracy.

For each investor and asset we train two different versions of a RF in two different scenarios. In the first scenario we test the improvement of nowcasting accuracy when the information about the Synchronization network is used. Thus, two versions of RF model are trained, first one only considers the information about the price whereas the second also takes into account the activity of the connected investors in the aforementioned network. In contrast, in the second scenario we test how the accuracy increases when the Anticipation network is used to forecast the behavior of the day after. For this purpose we consider again two cases. In the first one the RF is fed with the current price and position in order to predict investor's tomorrow behavior. In a similar procedure than before, in the second version of the RF we also add the current position of the investors connected in the Anticipation network as an information source.

Predictions for the version of the RF that does not include the network information are mostly flat, i.e. almost all "symbol 0" along the whole time series, generating high prediction accuracy values due to the sparseness of activity events. Despite having shown above that when the investor acts she is strongly driven by price on average, the overwhelming cases of symbols that encode no activity make the flat prediction very successful. The accuracy of the first RF version surpasses the 90% threshold. However, the real challenge is to generate prediction series that not only successfully predict no-activity events, but also when the real investor has acted. It is under these conditions that the RF version that uses the network information clearly outperforms the version that does not. Figure 5 shows that the probability to successfully predict the events when the investor presents some activity, i.e. symbols 1 and 2 in the original investor activity series, is systematically higher both for nowcasting or predicting the day after.



Random Forest models also provide the weight from 0 to 1 of the importance of each feature, i.e. each vector used to predict the activity, used during the training process. In the second RF version we can compare the contributions of neighbors activity with the rest of the features like price or activity of the investor in the day before (for the forecasting scenario). Thus, we see that the importance of the network adds up to 0.85 for nowcasting predictions and 0.51 for forecasting predictions.

4 Discussion

In complex systems research and real-world networks [55], and in economics in particular, emergent behavior is one of the most striking phenomenon. The price of an asset itself is the result of the interaction of multiple individual agents when buying or selling shares of financial assets [34, 35]. The study of features and common properties of the individual behavior is important, because often they are behind macroscopic phenomena like bubbles, crashes and other price dynamics [56]. Agents receive multiple stimuli before making a decision. There is a myriad of possible reasons behind a decision of buying or selling. However, thanks to the development of electronic markets, we can identify certain elements that statistically drive individual behavior. Several studies have focused on exogenous factors [40]. Here, we complete this vision by studying how endogenous factors may affect the individual behavior, at least in the case of non-expert (non-professional) investors. Particularly, we demonstrate that price drives the individual behavior for the majority of non-expert investors who work within a one-day time window. This confirms the results of Gutiérrez-Roig et al. [49] in a previous study, where imitation was found as an intuitive strategy to cope with the uncertainties of the market.

The method used in the analysis of non-expert investors, Symbolic Transfer of Entropy, stands as an appropriate tool for the treatment of non-linear time series –ubiquitous in the field of econophysics, and social systems in general [57]. It is also important to highlight the adaptation of such method for the symbolization of market position series. In the original description of the symbolization technique [44] identical numerical values, which are a very often event in investor position series, were not considered. Our method suggested here improves this weakness for such kind of time series and could be useful for further studies using Symbolic Transfer of Entropy applied on investor position series or similar.

The use of this method allows to map and link investors according to their behavior in a alternative manner than some recent studies [29, 33, 38]. Thus, far from observing random networks, we are able to detect groups of synchronized investors as well as 'leaders' that anticipate with respect to the others looking at the synchronization and anticipation networks. As mentioned above, the price as a driver and the reaction of the investors to it have a strong influence in the creation of these links between investors. But since the relationship between investors is measured considering their empirical performance, this method could also be useful to measure other kind of important behavioral effects in financial markets and economy, such as herding behavior [40, 49, 58, 59]. Accessing those maps of investors communities connected by similar behavior complements previous studies [11, 31, 33, 38] and it is also of interest for investment firms and financial institutions: to begin with, for better sampling –identifying key actors in the network, and studying their behavior in depth, could eventually enable their use as a proxy to estimate the behavior of the entire community. Secondly, they could also improve the prediction

of the reaction of their clients, and therefore anticipate and respond efficiently to their impact. In fact, we demonstrate a substantial improvement in accuracy prediction when using the information of behavioral networks.

Further studies along this line could consider the heterogeneity in the investment horizons that investors actually have [33], by extending the symbols to shorter and longer periods or testing predictions at different time horizons. However, data resolution and activity patterns of investor population in our dataset is restricted to daily time windows. Thus, future work would include the validation of our results for shorter and longer trading time windows, for longer time series, and for a larger collection of investors. Such study would demonstrate our hypothesis that synchronized investors are in fact anticipated/delayed even in a shorter time-scale. Finally, another possible improvement to better understand how price modulates the interaction between investors could consist on considering triplets of variables (two investors and price) when calculating Transfer of Entropy, rather than only doing it for pairwise investors behavior. This technique has already presented promising results in neuroscience when applied to time-series of cortical data [60, 61].

Additional material

Additional file 1: Supplementary information (PDF 2.7 MB)

Acknowledgements

This work was supported by MINECO (Spain) FIS2013-47532-C3-2-P (MG-R and JP), FIS2016-78904-C3-2-P (JP); by Generalitat de Catalunya (Spain) through Complexity Lab Barcelona (contracts no. 2014 SGR 608, MG-R and JP, and 2017 SGR 1064, JP). We finally want to specially acknowledge anonymous referees for their comments, which have helped to highly improve the results of our research and the final version of the manuscript.

Abbreviations

SMI, Symbolic Mutual Information; STE, Symbolic Transfer of Entropy.

Availability of data and materials

The dataset used in this paper is Zenodo repository with the DOI reference 10.5281/zenodo.2573031. The python codes used to symbolize the time series data and to compute the Symbolic Mutual Information and Symbolic Transfer of Entropy are stored in the following GitHub repository: <https://github.com/mariogutierrezroig/smite>.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MG-R, JP conceived and designed the study. MG-R and JB-H analyzed the data. MG-R, JP, JB-H, AA discussed the analysis results, MG-R, JP, JB-H, AA wrote the manuscript. All four authors reviewed and approved the paper.

Author details

¹Data Science Lab, Warwick Business School, University of Warwick, Coventry, UK. ²Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Spain. ³Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain. ⁴Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona, Spain. ⁵Institute of Complex Systems UBICS, Universitat de Barcelona, Barcelona, Spain.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 August 2018 Accepted: 12 March 2019 Published online: 27 March 2019

References

1. King G (2011) Ensuring the data-rich future of the social sciences. *Science* 331(6018):719–721
2. Schelling TC (2006) *Micromotives and macrobehavior*. Norton, New York
3. González-Bailón S, Borge-Holthoefer J, Moreno Y (2013) Broadcasters and hidden influentials in online protest diffusion. *Am Behav Sci* 57(7):943–965

4. Bouchaud JP, Bonart J, Donier J, Gould M (2018) *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, Cambridge
5. Bouchaud JP (2013) Crises and collective socio-economic phenomena: simple models and challenges. *J Stat Phys* 151(3–4):567–606
6. Iori G (2002) A microsimulation of traders activity in the stock market: the role of heterogeneity, agents' interactions and trade frictions. *J Econ Behav Organ* 49(2):269–285
7. Chiarella C, Iori G, Perelló J (2009) The impact of heterogeneous trading rules on the limit order book and order flows. *J Econ Dyn Control* 33(3):525–537
8. Tedeschi G, Iori G, Gallegati M (2012) Herding effects in order driven markets: the rise and fall of gurus. *J Econ Behav Organ* 81(1):82–96
9. Farmer JD, Foley D (2009) The economy needs agent-based modelling. *Nature* 460(7256):685–686
10. Mike S, Farmer JD (2008) An empirical behavioral model of liquidity and volatility. *J Econ Dyn Control* 32(1):200–234
11. de Lachapelle DM, Challet D (2010) Turnover, account value and diversification of real traders: evidence of collective portfolio optimizing behavior. *New J Phys* 12(7):075039
12. Perelló J, Masoliver J, Kasprzak A, Kutner R (2008) Model for interevent times with long tails and multifractality in human communications: an application to financial trading. *Phys Rev E* 78(3):036108
13. Barabasi A-L (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211
14. Mizuno T, Ohnishi T, Watanabe T (2017) Novel and topical business news and their impact on stock market activity. *EPJ Data Sci* 6(1):26
15. Patzelt F, Bouchaud J-P (2018) Universal scaling and nonlinearity of aggregate price impact in financial markets. *Phys Rev E* 97(1):012304
16. Bouchaud J-P, Gefen Y, Potters M, Wyart M (2004) Fluctuations and response in financial markets: the subtle nature of random price changes. *Quant Finance* 4(2):176–190
17. Eisler Z, Perelló J, Masoliver J (2007) Volatility: a hidden Markov process in financial time series. *Phys Rev E* 76(5):056105
18. Gillemot L, Farmer JD, Lillo F (2006) There's more to volatility than volume. *Quant Finance* 6(5):371–384
19. Perelló J, Masoliver J (2003) Random diffusion and leverage effect in financial markets. *Phys Rev E* 67(3):037102
20. Thurner S, Farmer JD, Geanakoplos J (2012) Leverage causes fat tails and clustered volatility. *Quant Finance* 12(5):695–707
21. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
22. Staniek M, Lehnertz K (2008) Symbolic transfer entropy. *Phys Rev Lett* 100(15):158101
23. Ni K-Y, Lu T-C (2014) Information dynamic spectrum characterizes system instability toward critical transitions. *EPJ Data Sci* 3(1):28
24. Chen X, Tian Y, Zhao R (2017) Study of the cross-market effects of brexit based on the improved symbolic transfer entropy garch model. An empirical analysis of stock-bond correlation. *PLoS ONE* 12(8):0183194
25. Zhang N, Lin A, Shang P (2017) Multiscale symbolic phase transfer entropy in financial time series classification. *Fluct Noise Lett* 16(2):1750019
26. Bekiros S, Nguyen D, Junior L, Uddin GS (2017) Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *Eur J Oper Res* 256:945–961
27. Rocchi J, Tsui EYL, Saad D (2017) Emerging interdependence between stock values during financial crashes. *PLoS ONE* 12(5):0176764
28. Tumminello M, Micciché S, Lillo F, Piilo J, Mantegna RN (2011) Statistically validated networks in bipartite complex systems. *PLoS ONE* 6(3):e17994
29. Challet D, Chicheportiche R, Lallouache M, Kassibrakis S (2018) Statistically validated lead-lag networks and inventory prediction in the foreign exchange market. *Adv Complex Syst* 21(08):1850019
30. Cordi M, Challet D, Kassibrakis S (2019) The market nanostructure origin of asset price time reversal asymmetry. Preprint. [arXiv:1901.00834](https://arxiv.org/abs/1901.00834)
31. Tumminello M, Lillo F, Piilo J, Mantegna RN (2012) Identification of clusters of investors from their real trading activity in a financial market. *New J Phys* 14(1):013041
32. Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D (2016) Statistically validated network of portfolio overlaps and systemic risk. *Sci Rep* 6:39467
33. Musciotto F, Marotta L, Piilo J, Mantegna RN (2018) Long-term ecology of investors in a financial market. *Palgrave Commun* 4(1):92
34. Odean T (1998) Are investors reluctant to realize their losses? *J Finance* 53(5):1775–1798
35. Odean T (1999) Do investors trade too much? *Am Econ Rev* 89(5):1279–1298
36. Grinblatt M, Keloharju M (2000) The investment behavior and performance of various investor types: a study of Finland's unique data set. *J Financ Econ* 55(1):43–67
37. Grinblatt M, Keloharju M (2009) Sensation seeking, overconfidence, and trading activity. *J Finance* 64(2):549–578
38. Musciotto F, Marotta L, Micciché S, Piilo J, Mantegna RN (2016) Patterns of trading profiles at the nordic stock exchange. A correlation-based approach. *Chaos Solitons Fractals* 88:267–278
39. Bohlin L, Rosvall M (2014) Stock portfolio structure of individual investors infers future trading behavior. *PLoS ONE* 9(7):103006
40. Lillo F, Micciché S, Tumminello M, Piilo J, Mantegna RN (2015) How news affects the trading behaviour of different categories of investors in a financial market. *Quant Finance* 15(2):213–229
41. Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3):424–438
42. Ver Steeg G, Galstyan A (2012) Information transfer in social media. In: *Proceedings of the 21st International Conference on World Wide Web*, pp 509–518
43. Lungarella M, Ishiguro K, Kuniyoshi Y, Otsu N (2007) Methods for quantifying the causal structure of bivariate time series. *Int J Bifurc Chaos Appl Sci Eng* 17(03):903–921
44. Bandt C, Pompe B (2002) Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 88(17):174102

45. Lizier JT, Prokopenko M (2010) Differentiating information transfer and causal effect. *Eur Phys J B* 73(4):605–615
46. Barrett AB, Barnett L (2013) Granger causality is designed to measure effect, not mechanism. *Front neuroinform* 7:6
47. Hutter M (2002) Distribution of mutual information. In: *Advances in neural information processing systems*, pp 399–406
48. Newman MEJ (2010) *Networks: an introduction*. Oxford university press, Oxford
49. Gutiérrez-Roig M, Segura C, Duch J, Perelló J (2016) Market imitation and win-stay lose-shift strategies emerge as unintended patterns in market direction guesses. *PLoS ONE* 11(8):0159078
50. Bennett J, Lanning S (2007) The netflix prize. In: *Proceedings of KDD cup and workshop*, p 35
51. Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the Flickr social network. In: *Proceedings of the 18th International Conference on World Wide Web*, pp 721–730
52. Li X, Ling CX, Wang H (2016) The convergence behavior of naive Bayes on large sparse datasets. *ACM Trans Knowl Discov Data* 11(1):10
53. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
54. Xu P, Jelinek F (2007) Random forests and the data sparseness problem in language modeling. *Comput Speech Lang* 21(1):105–152
55. Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G (2019) The statistical physics of real-world networks. *Nature Rev Phys* 1(1):58–71
56. Bouchaud J-P, Bonart J, Donier J, Gould M (2018) *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, Cambridge
57. Borge-Holthoefer J, Perra N, Gonçalves B, González-Bailón S, Arenas A, Moreno Y, Vespignani A (2016) The dynamics of information-driven coordination phenomena: a transfer entropy analysis. *Sci Adv* 2(4):1501158
58. Bouchaud J-P (2018) Agent-based models for market impact and volatility. In: *Handbook of computational economics*, vol 4. Springer, Berlin, pp 393–436
59. Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–292
60. Faes L, Marinazzo D, Stramaglia S (2017) Multiscale information decomposition: exact computation for multivariate Gaussian processes. *Entropy* 19(8):408
61. Erramuzpe A, Ortega GJ, Pastor J, de Sola RG, Marinazzo D, Stramaglia S, Cortes JM (2015) Identification of redundant and synergetic circuits in triplets of electrophysiological data. *J Neural Eng* 12(6):066007

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
