Ranking Influential and Influenced Shares Based on the Transfer Entropy Network

José de Paula Neves Neto[1], Daniel Ratton Figueiredo[1]

[1] Programa de Engenharia de Sistemas e Computação (PESC)
Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE)
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil

`{neves,daniel}@cos.ufrj.br`

Abstract. *Influence is a concept found in nature and society and is related to the interdependency among a set of objects. In the context of a stock market, the variation in price of shares can influence the variation in price of other shares, leading to influential and influenced shares. In this work we leverage the notion of transfer entropy to build a network of shares and pairwise directed influence that is used to rank the most influential and influenced shares. Classical network centrality metrics such as PageRank and HITS are leveraged to rank the nodes. We apply our methodology to the shares in the greater stock market in Brazil, we rank nodes to find source and destination of influence in that market, while also comparing the different rankings and their correlation with traded volume.*

1. Introduction

Influence is a fundamental concept in nature and society as all objects both exert and suffer some kind of influence from one another. Human intuition says that climate influences crop and unemployment rate influences wages, for example. Indeed, there are many situations in which the action of an object clearly affects the behavior of another. This kind of relationship between objects is what we refer to as *influence*.

Influence can also be seen as *information*. In particular, when one object influences another, one can say that the first object has passed information to the second. Thus, influence can be interpreted as *information transfer*. An approach to quantify this kind of influence has been recently proposed, known as *transfer entropy*, and is being widely applied in different contexts [Bossomaier et al. 2016].

Shares traded in a stock market intuitively influence one another, as the price variation in a share may trigger (or be correlated with) a subsequent price variation in another share. Our focus is the *influence* between shares in a stock market exchange. By considering a pairwise directed measure of influence, we construct a directed weighted network and leverage its structure to determine the most *influential* and the most *influenced* shares in the market.

To achieve this goal, we consider the time series to represent the price change (not the price value) of shares and measure the information transfer between every two ordered pair of shares. We construct a directed weighted network where nodes are shares and edges weights represent transfer entropy values. Thus, we can analyze this network using classic centrality measures to assess the importance of nodes (shares).

Note that the edge weight $w_{i,j}$ denotes how much share i influences share j. Intuitively the sum of incoming edges to node i characterizes how influenced is this share, while the sum of outgoing edges of node i characterizes its influence power. Ranking shares based on its influence power or its subjection to influence yields a metric of *importance* for shares.

The ranking of shares can be done using algorithms like PageRank or HITS, originally developed for ranking web pages [Brin and Page 1998, Page et al. 1999, Kleinberg 1999], which leverage the weights and structure of the network. Moreover, nodes can also be ranked by their incoming and outgoing total (sum) edge weights, here called *node weight*.

We apply this methodology to shares traded in the BM&FBOVESPA, the most important stock market in Brazil. Using public available data (time series of share prices over a 21 year period), we measure pairwise transfer entropy and construct the network. We rank nodes according to three different metrics yielding for each metric a ranking for influential and influenced shares. We assess the coherency among the top ranked shares across the different rankings, to better characterize influenced and influential shares. We also investigate the relationship between influence (according to the ranking of different metrics) and traded volume of shares, and results indicate there is very little correlation.

The novelty of this paper is to consider the relation between movements of different shares instead of the relation between the price of shares. Intuitively, an investor or a stockbroker observes the movement (and its bias) to make their subsequent decision about buying or selling a stock, based on the belief that it goes up or down in the future. We have not found other papers using actually the same approach. But, in this way, we have another measure to assess importance of stock market shares.

The reminder of this paper is organized as follows: In Section 2 we revise some fundamentals and other related works. The proposed methodology is presented in Section 3. Section 4 presents the empirical data that was collected and analyzed. Section 5 exposes the results of the different rankings and correlation analysis. Finally, Section 6 summarizes the work.

2. Related Works

2.1. Measuring Dependency

There are many ways to measure dependency and information and possibly the most widely accepted notion is that of *entropy*, introduced by Shannon [Shannon 1948], to measure information as follows. Given a random variable I, with probability distribution $p(i)$, the Shannon entropy H_I is defined as

$$H_I = - \sum_{i \in D(I)} p(i) \log_2 p(i) \tag{1}$$

where $D(I)$ is the image of random variable I. The larger the value of H_I the more information the random variable carries. In particular, the maximum value of H_I is achieved for a uniform random variable.

When dealing with two objects a useful measure is the *mutual information*

[Shannon 1948]. The mutual information M_{IJ} of two random variables I and J, with distribution $p(i)$ and $p(j)$ respectively, and joint-probability $p(i,j)$ is defined as

$$M_{IJ} = \sum_{\substack{i \in D(I) \\ j \in D(J)}} p(i,j) \log_2 \frac{p(i,j)}{p(i)p(j)} \tag{2}$$

Mutual information does not suit our goals because it does not consider the direction of the relationship nor the causality (which comes after) between the two objects. To account for these notions we use instead the concept of transfer entropy, defined below.

2.2. Transfer entropy

Transfer entropy was introduced by Schreiber [Schreiber 2000] and is based on the concept of *entropy* and *mutual information*. The novelty of Schreiber's proposal is the introduction of a time delay between the observed variables, a concept useful in several applications [Bossomaier et al. 2016]. Thus, it considers an object (event) that occurs "after" the others.

The transfer entropy $T_{J \to I}$ from random variable J to random variable I is given by

$$T_{J \to I} = \sum_{\substack{i_+, i \in D(I) \\ j \in D(J)}} p(i_+, i, j) \log_2 \frac{p(i_+ \mid i, j)}{p(i_+ \mid i)} \tag{3}$$

where i_+ is the value taken by random variable I in its next occurrence. Note that transfer entropy measures the ratio in the conditional probabilities of i_+ given i when j is also observed. Thus, if observing j allows to increase the probability of a particular i_+, reducing its entropy, we have transfered information from J to I.

2.3. Computing Transfer Entropy

Consider two time series I and J, both with length N. In order to apply Equation 3 to empirical data the (joint and conditional) probabilities of $p(i_+, i, j)$, $p(i_+ \mid i, j)$ and $p(i_+ \mid i)$, where $i_+, i \in I$ and $j \in J$, must be properly estimated. This is done using the time series to compute the relative frequency of the observed values.

To ease calculation, those probabilities are transformed into their versions without conditionals, yielding only the (joint) empirical probabilities $p(i)$, $p(i_+, i)$, $p(i, j)$ and $p(i_+, i, j)$. Thus, the number of occurrences for each instance (single value, pair of values and triple values) is counted across the two time series to determine the corresponding relative frequencies, which are used as their probabilities.

Finally, the transfer entropy from J to I is determined by the summation described below, running over all values contained in I and J series.

$$T_{J \to I} = \sum_{\substack{i_{n+1}, i_n \in D(I) \\ j_n \in D(J)}} p(i_{n+1}, i_n, j_n) \log_2 \frac{p(i_{n+1}, i_n, j_n) \cdot p(i_n)}{p(i_{n+1}, i_n) \cdot p(i_n, j_n)} \tag{4}$$

where $n \in 1, 2, \ldots, N-1$ and i_k, j_k are the k-th values into I and J series, respectively.

As discussed ahead in Sub-section 3.2, values for i_{n+1}, i_n and j_n are discrete and can assume only four different symbols, instead of a larger number of values or even a continuous scale, because this would enlarge the space of possible combinations of values, and would hamper estimation of respective probabilities and, consequently, the transfer entropy computation would be more noisy.

2.4. Applications of transfer entropy

Market networks can be defined as the network of equities created by enterprises and their relationships within the financial market. Market networks take advantage of network analysis tools and metrics which are leveraged to analyze the market, understand relationship between enterprises, support forecasting and portfolio strategies [Baker 1990, Kauê Dal'Maso Peron et al. 2012].

There are various works on market network analysis [Tabak et al. 2010, Leahy et al. 2014], market network measurements [Lee and Djauhari 2012], market network correlations [Shirokikh et al. 2013, Lyócsa et al. 2012, Namaki et al. 2011, Boginski et al. 2005], stock market network forecasting [Atsalakis and Valavanis 2009] and financial modeling [Petelin et al. 2011]. For example, Tabak et al. [Tabak et al. 2010] considers a market network where nodes are shares in a stock market and undirected edge weights denote a distance based on cross-correlation between the prices time series. In particular, correlation values of $+1$, 0 and -1 were mapped into distances of 0, 1.4142 and 2, respectively, which are then used to compute the Minimal Spanning Tree (MST) of the network.

Approaches based on transfer entropy have recently increased and are being addressed in different scenarios, including financial markets [Bossomaier et al. 2016, Schreiber 2000, Yang et al. 2017]. In particular, there are recent studies on information transfer between stock markets (of different countries) using transfer entropy methods [Kwon and Yang 2008, Marschinski and Kantz 2002, Yang et al. 2017, Korbel et al. 2017, He and Shang 2017], including the use of variations like *Effective Transfer Entropy*, *Phase Transfer Entropy*, *Rényi's Transfer Entropy* and *Effective Phase Transfer Entropy*. However, our focus is information transfer between shares of the same market.

2.5. Node Ranking in Networks

There are many available models and metrics for ranking nodes in directed weighted networks. Directed edges give us to distinguish two types of nodes in the network: *authorities* and *hubs*. Authorities are nodes which act as sinks, receiving many links (or valuable ones), while hubs are nodes with many outgoing edges (or valuable ones).

In order to rank network nodes two techniques that can be applied are PageRank [Page et al. 1999] and HITS [Kleinberg 1999, Chakrabarti et al. 1998]. PageRank ranks web pages by their importance, considering the importance of the pages that point to it. HITS has a similar recursive structure, but considering that nodes are both hubs and authorities, and uses this categorization in defining importance. In particular, an important authority is a node that is pointed to by important hubs. Similarly, a hub is important if it points to important authorities.

t	0	1	2	3	4	5	6	7	8	9	10
OP	8.69	0.28	9.18	5.66	2.77		1.51		8.85	8.88	4.98
CP	0.28	9.18	5.66	2.77	7.18		8.85		8.88	4.98	3.12
M	down	up	down	down	up	null	up	null	draw	down	down

t = time; OP = opening price; CP = closing price; M = movement symbol

Figure 1. Mapping stock prices series into stock biases series.

Since the graph that represents the market network is complete (all edges are present), such algorithms are only meaningful in their extended version that uses edge weights: *weighted PageRank* [Xing and Ghorbani 2004] and *weighted Hyperlink-induced topic search* [Li et al. 2002, Zhang et al. 2007].

3. Methodology

3.1. Time series considered

The basic time series element considered in many analysis is not the stock price movement, but its value. However, our interest has been on how a stock movement can influence a movement of another stock. In other words, does the price change of a certain stock today influence the price change of another stock tomorrow?

Several studies make comparisons between stock prices which are synchronized in time, that is, prices of two stocks A and B are compared in the same time interval $t_i, \ldots, t_f$, where t_i and t_f are initial and final time instants, respectively. Such studies search for a kind of binding between two stocks, trying to understand how and why they move together over time, possibly being influenced by a third party.

On the other hand, our study focuses on stock price movements. They are preferable to individual price values because we focus on understanding influence among stocks in a market. Trading decisions (to sell or to buy) a stock occur after observation and analysis (foresee) of a previous stock movement. There is clearly a time shift between the observed fact (one share movement) and the action (other share trading). Thus, the model we propose follows along this intuition.

In order to understand how past movements influence future movements we will consider a time shift between two time series. We convert prices into symbols that will represent our basic series elements, namely movements of a bond price.

3.2. Symbols

We will adopt a set of symbols to investigate influence in stock price movements. Using the daily *opening* and *closing* prices for shares we map a price series (float type) into a symbol series (discrete type) as shown in Figure 1. If a stock price goes up from market opening time to its close time, we have an *up* symbol for that stock in that day; if it goes down, we have a *down* symbol; if it remains *almost the same*, we have a *draw* symbol; and there are cases where we have no opening neither/or closing price for the share, because it is not traded in that day, and in this case we have a *null* symboBut, in this way, we have another measure to assess importance of stock market sharesl.

As a remark, when we say "*almost* the same", we mean prices whose opening and closing values are similar, with a val difference smaller than $\varepsilon > 0$. This avoids false positives for irrelevant *up* or *down* movements. Actually ε is an arbitrary value, but it is reasonably coherent to use a factor of $\varepsilon = 4.5 \cdot 10^{-5} = 0.0045\%$ because it represents a daily percentage that corresponds to a 1% (one percent) monthly variation divided by 22 (twenty two) work days of a month.

Therefore each stock has a time series in which elements are one of the four symbols and its length is the number of work days in time period.

3.3. Pairwise Transfer Entropy

The time series of movements will be used to compute the transfer entropy for each pair of stocks. The idea is to measure the transfer of information from one share to the other one, in a pairwise comparison. Notice that this is a directed measure, so we can evaluate the transfer of information in both directions. The transfer entropy will be calculated using a lag of one position in the series (*e.g.*, one day).

3.4. The Directed Network and Ranking

We will consider all ordered pairs of shares, and thus, our model generates a complete directed graph where nodes are shares and directed links have weights that correspond to the transfer entropy of the ordered pair.

In order to rank nodes and capture their influence in the network, we use PageRank and HITS. Given the directionality of edges, *hubs* are stocks that *influence* many others hubs, whereas *authorities* are stocks that are *influenced* by many others.

PageRank generates a measure, through a recursive method, where a node is more important the more important are the nodes that point to it, which is the similar idea of an authority in HITS.

Another possible metric to rank nodes is to give every node a number that is equal to the sum of all its incoming weighted edges. Nodes with higher numbers can be seen as authorities as well. Analogously, we can give every node a number equal to the sum of all its outgoing weighted edges, where nodes with high numbers can be seen as hubs.

4. Collected Data

BM&FBOVESPA, the main Brazilian stock exchange, is the target of this work. There are about 700 public companies which negotiate their bonds in BM&FBOVESPA and such information, including bond prices, is publicly available [bmf 2017].

Companies can have a large number of different bonds negotiated in the market and are also responsible for various actions concerning them, such as creation, discontinuation, changing names, joining, or splitting bonds. However, such activities impose limitations in the data analysis since records are not always publicly available, being hard to track names changes, or bonds that have joined or split. This poses some difficulty in tracking a particular bond over long time periods.

4.1. Dealt shares and records

BM&FBOVESPA offers an electronic data set containing daily transactions for each share since August 1980. Our initial evaluation revealed over 40,000 different bonds (each with

a code) negotiated in the market within that period, where each stock within BM&FBO-VESPA is identified by a unique code. However, many of them have a short life time (one month, for example) or a very low *liquidity* (not traded frequently). In terms of an analysis over a large time period, shares with one of these two characteristics are not interesting.

Therefore, in the present work, two main filters were applied to the original data. These filters identify stocks that are more relevant to the present study of influence in the market, discarding stocks that are of less importance. The following procedure reduced the amount of codes from over 40,000 to 210:

1. Only those share codes that were negotiated in the last month of the period (December 2016) are considered;
2. Only those share codes that were dealt at least in 50% of the work days in the period are considered.

5. Results

In this section, results obtained by applying the proposed methodology to data from BM&FBOVESPA are presented and discussed. We first discuss general properties of nodes and edges of the market graph, and then ranking of stocks based on three metrics.

5.1. Edges: transfer entropy between stocks

We start by presenting results on the edge weights among all the shares (a total of 43,890 edges), Recall that directed edge weights represent transfer entropy values between the respective shares, according to the direction of information transfer.

Figure 2(a) shows the complementary cumulative distribution function of edge weights. The additional colored grid lines indicate 50% (red), 10% (green) and 5% (yellow) of total points of the distributions. Note that a small amount of edges have a much larger value. The 20 largest values are represented by the green dots and their identification is shown in the box. Note that 50% of the values have a transfer entropy less than 4.27×10^{-3}, whereas 95% have a value less than 3.98×10^{-2}, with an average of 9.71×10^{-3}. Negative values are possible in the calculations and indicates that basically no information is transfered.

Note that some shares appear multiple times as the source of information transfer among the largest edges, while other shares appear multiple times as the target, suggesting that there are bonds with greater presence in each side of information transfer. For example, **ITUB4T** (2), **CSAN3** (2), **CSAN3F** (3) and **EMBR3** (2) appear more than once as source of information transfer among the top 20 (in parentheses one can see number of appearances) of *most influential* shares. Yet when considering not the shares but the organization, aggregating their bonds, **ITUB** (2), **CYRE** (2), **CSAN** (5), **ENBR** (3) and **LREN** (2) occur several times as influential source of information transfer.

Figure 2(b) shows all transfer entropy values at once, in matrix form, with values in logarithmic scale. For example, the four shares mentioned above are in lines: **ITUB4T** (118), **CSAN3** (53), **CSAN3F** (54) and **EMBR3** (74). In general, a "lighter" line suggests that the corresponding share is not an important source of information transfer, whereas a "darker' line indicate that the share transfers much information to other shares. Analogously, a "lighter" column suggests that the share does not receive much information from other shares, and a "darker" column suggests a share that receive information from many other shares, being therefore non susceptible to the action of others.
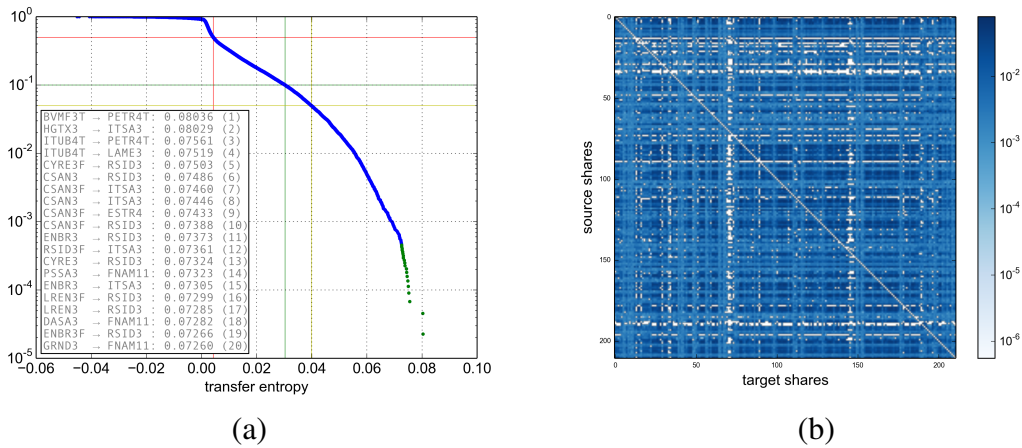
Figure 2. Weight values for all shares: (a) complementary cumulative distribution function, in semi-log scale; (b) color map, in log scale.
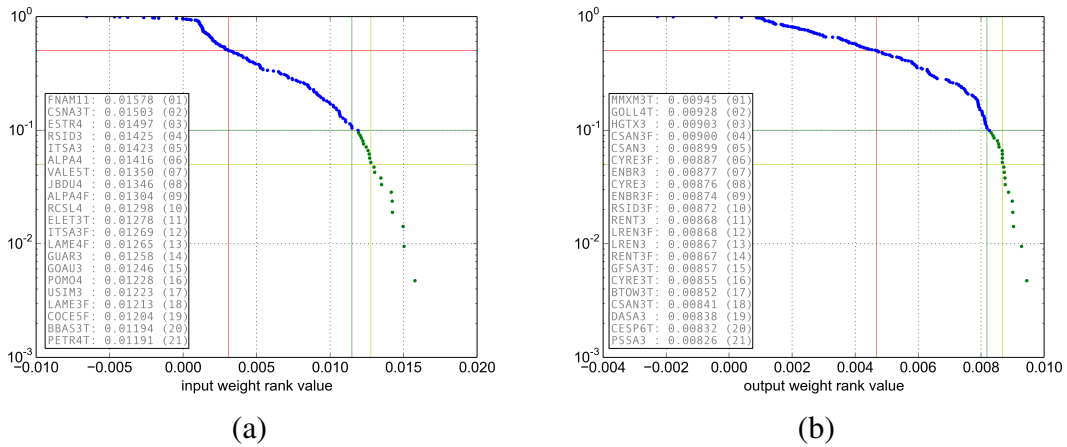

Figure 3. Complementary cumulative distribution function of node weights in semi-log scale: (a) incoming weight (influenced by others); (b) outgoing weight (influence others).

5.2. Incoming and outgoing weight

The outgoing weight of a node is given by the sum of all its outgoing edge weights. Analogously, the incoming weight of a node is the sum of all its incoming edges.

The outgoing and incoming weights of nodes allows to understand their importance in transferring (both sending and receiving) information in the market network. Figure 3 shows the complementary cumulative distribution for incoming and outgoing weights over all nodes. Thus, with respect to each share is now ranked both in terms of being *influenced* or *influential* of information transfer.

Shares with higher incoming weight are those that are more susceptible to changes of other shares, whereas shares with lower incoming weight are relatively robust to other shares movements. Both kinds of shares are interesting, depending on the objectives and market expectations. Shares with higher outgoing weight are the most influential ones, whereas shares with lower outgoing weight are relatively invisible to the other shares. In both cases the tail of the distribution shows that there are very few shares with weights
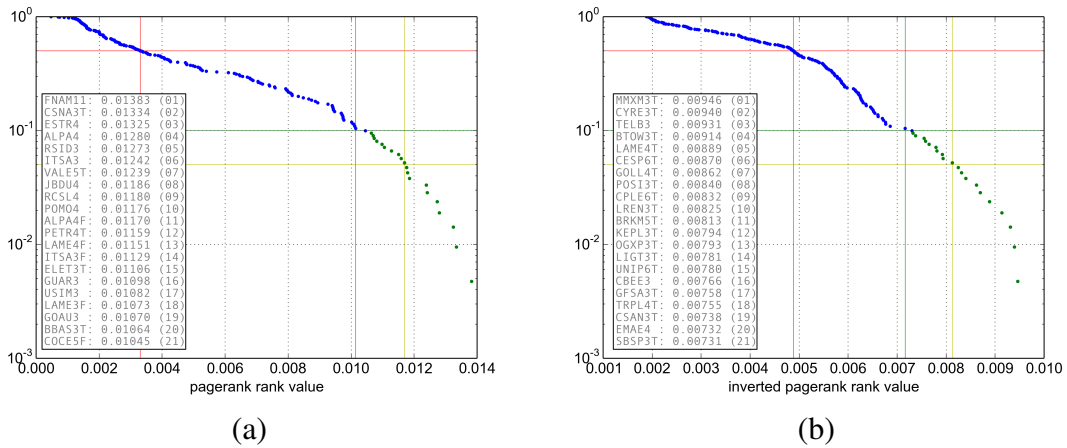
Figure 4. Complementary cumulative distribution function for PageRank values of shares in semi-log scale: (a) original network (influenced by other); (b) reverse edge network (influence others).

much larger than the average. This indicates that only a few shares are really influential and really influenced. Interestingly, there is no overlap between the top 20 largest incoming and outgoing weight.

5.3. Using PageRank algorithm

The weighted version of PageRank algorithm was also used to rank the shares, taking the complete network as input. Note that the weighted PageRank algorithm computes how important a share is with respect to being the target of information transfer or, in other words, how much the market influence the share. Larger PageRank values indicate shares that are more influenced.

In order to capture how influential are the shares, we reverse the direction of the edges in the network. With this transformation, a node with many heavy incoming edges is a great source of information transfer. We run the weighted version of PageRank in the network to identify the most influential nodes.

Figure 4(a) shows the cumulative distribution of PageRank values for the market graph, whereas Figure 4(b) shows the PageRank value distribution for the reverse edge graph. We again observe that a few nodes have Pagerank values much larger than the average, being much influenced by others (Figure 4(a)), or much influential (Figure 4(b)). These results can be compared to those in Figure 3 which shows incoming and outgoing node weight.

5.4. Using HITS algorithm

We also use HITS algorithm to rank shares in the network [Chakrabarti et al. 1998]. HITS algorithm assigns two numbers to every node in the network, namely a hub factor and an authority factor. A *hub* is related to the *source*, and an *authority* to the *target* of information transfer. Figure 5 shows the distribution for hub and authority values. As PageRank, HITS shows that there are some shares whose values are very far from the average.

Figure 5.

```
(a)
FNAM11: 0.01606 (01)
ESTR4 : 0.01513 (02)
CSNA3T: 0.01493 (03)
ITSA3 : 0.01488 (04)
RSID3 : 0.01457 (05)
ALPA4 : 0.01404 (06)
ITSA3F: 0.01353 (07)
ALPA4F: 0.01345 (08)
JBDU4 : 0.01340 (09)
VALE5T: 0.01329 (10)
RCSL4 : 0.01296 (11)
ELET3T: 0.01291 (12)
GUAR3 : 0.01283 (13)
LAME3F: 0.01282 (14)
GOAU3 : 0.01277 (15)
LAME4F: 0.01272 (16)
COCE5F: 0.01247 (17)
USIM3 : 0.01225 (18)
ITSA4T: 0.01218 (19)
BBAS3T: 0.01192 (20)
ETER3F: 0.01187 (21)
```
authority rank value

```
(b)
CSAN3F: 0.01014 (01)
CSAN3 : 0.01010 (02)
CYRE3F: 0.00996 (03)
CYRE3 : 0.00990 (04)
ENBR3 : 0.00989 (05)
ENBR3F: 0.00989 (06)
HGTX3 : 0.00986 (07)
LREN3 : 0.00981 (08)
RENT3F: 0.00981 (09)
RENT3 : 0.00979 (10)
LREN3F: 0.00979 (11)
RSID3F: 0.00974 (12)
PSSA3 : 0.00939 (13)
GRND3 : 0.00928 (14)
PSSA3F: 0.00928 (15)
CPFE3F: 0.00922 (16)
DASA3 : 0.00921 (17)
CPFE3 : 0.00919 (18)
SUZB5F: 0.00910 (19)
PIBB11: 0.00908 (20)
GOLL4T: 0.00908 (21)
```
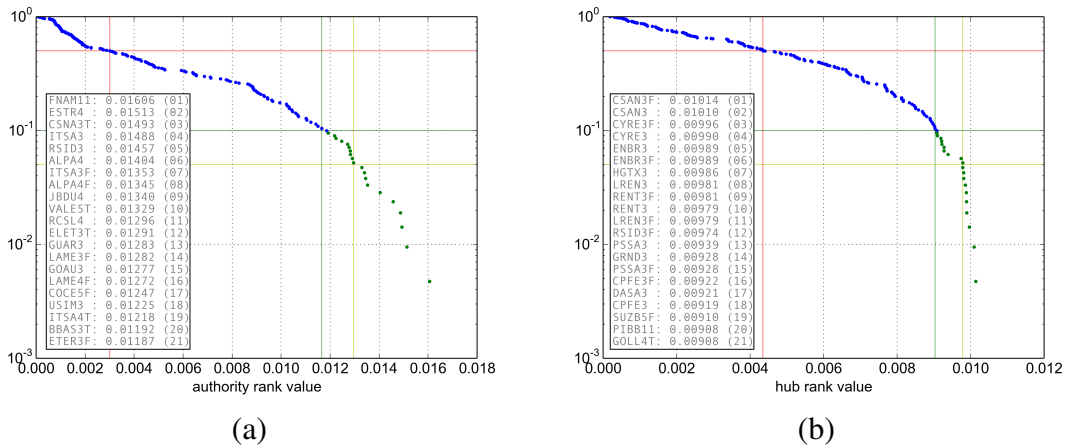hub rank value

Figure 5. Complementary cumulative distribution function for HITS values of shares in semi-log scale: (a) authorities (are influenced by other); (b) hubs (influence other).

Type of value	weight *vs.* PageRank	PageRank *vs.* HITS	HITS *vs.* weight	weight *vs.* HITS *vs.* PageRank
incoming	0.9048 (19)	0.8182 (18)	0.9048 (19)	0.8182 (18)
outgoing	0.2121 (7)	0 (0)	0.4815 (13)	0 (0)

Table 1. Jaccard coefficient for the top-20 shares among different ranking metrics, for incoming (influenced) and outgoing (influential) rankings.

5.5. Comparing between ranking methods

Note a coherence in the tail of the incoming weight distribution (Figure 3) and PageRank (Figure 4). In top 20 shares of these two rankings, 19 shares are common to both rankings. Both rankings have the exact same first three positions, **NAM11**, **CSNA3T** and **ESTR4**, in order, whereas others shares slightly change their position. HITS shows also consistent results with node weight and PageRank rankings.

However, the outgoing weight and reverse edge PageRank rankings do not show the same level of coherence. Among the 20 top shares in each ranking, only 7 are common to both, less than 50%. Analogously, in HITS, a share considered a top hub is not in the top of the ranking of the other two metrics. However, there is some consistency between hubs and outgoing weights.

The Jaccard coefficient was used to compare the top 20 incoming and outgoing values between the different rankings, summed up in Table 1. The first three columns show the comparison between pairs of rankings, while forth column compare all rankings together. In parentheses we have the number of coincident shares among the rankings.

The Spearman correlation coefficient (ρ) between rankings can also endorse the similarity between the rankings. Note some coherence for the different rankings are shown in Table 2. The Spearman coefficient was calculated using the entire rankings of all shares in the network.

The top 20 most influenced shares are the same, no matter what ranking metric is used. On the other hand, the most influential shares depend more on the ranking metric,

Metric	weight *vs.* PageRank	PageRank *vs.* HITS	HITS *vs.* weight
incoming	0.974	0.980	0.995
outgoing	0.779	0.687	0.984

Table 2. Spearman's ρ between rankings.

Metric	weight	PageRank	HITS
Spearman's ρ	0.033	-0.110	0.045

Table 3. Spearman's ρ between sensibility to influence (ranking) and influence power (ranking) of a share considering the same metric.

despite some agreement between HITS and node weights.

Finally, in general the influence power of a share does not agree with its sensibility to influence. In general, these things do not go together as we can see in Table 3, endorsing our intuition that says that the most influenced shares are not the most influential.

5.6. Computing financial volume

An important consideration is the relationship between the financial volume of the shares and their influence with respect to information transfer. Indeed, some shares have very large volumes (compared to the average) which could be related to their influence in the market.

Figure 6 shows the distribution of the total traded volume of the shares in the market for the observed period. It is notorious the volume of the most traded shares, for example, the volume of the top two shares, **PETR4** and **VALE5**, is one order of magnitude larger than 90% of the smaller shares! The top 6 shares by traded volume are shown in Table 4, along their position in every ranking metric (recall that the network has 210 shares).

Table 4 shows that financial volume and influence are not correlated. A highly traded share is not one that the most influences or is influenced by others. Indeed, the Spearman's coefficient between financial volume ranking and each one of the ranking metrics is shown in Table 5. Indeed, the very small values for the coefficient indicates that there is no correlation between the traded volume of a share and its influence in terms of information transfer.

Ranking method	PETR4	VALE5	BBDC4	PETR3	VALE3	BBAS3
incoming weight	163	155	138	146	69	131
outgoing weight	197	192	201	196	176	194
normal PageRank	175	163	141	153	68	140
inverted PageRank	189	200	201	206	177	197
HITS authority	163	156	135	149	69	133
HITS hub	200	188	199	195	175	193

Table 4. Position of certain high financial volume bonds in the lists of applied rank methods, considering 210 bonds.
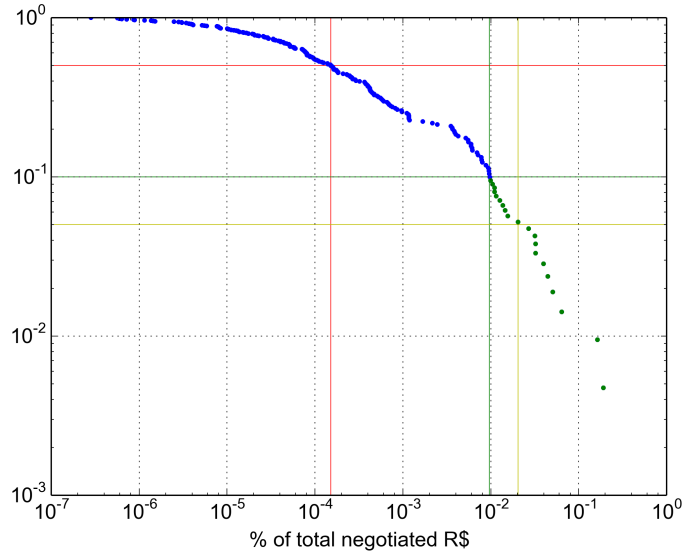
Figure 6. Complementary cumulative distribution function for market shares financial values (volume) in log scale.

Metric	volume *vs.* weight	volume *vs.* PageRank	volume *vs.* HITS
incoming	-0.108	-0.141	-0.124
outgoing	-0.101	-0.143	-0.108

Table 5. Spearman's ρ between volume ranking and influence rankings.

6. Conclusions

Transfer entropy has recently emerged as an ubiquitous metric to measure the influence between two dynamics, capturing the notion of information transfer. This work has leveraged this metric to measure influence among shares traded in the stock market exchange, BM&FBOVESPA. In particular, we have considered a 21-years daily record of stock price movement to characterize the influence (or information transfer) between shares in the market. Beyond computing the pairwise transfer entropy among the shares, we proposed a network-based approach to identify both shares that are influential and shares that are influenced across the market.

We construct a market graph, where shares correspond to nodes and transfer entropy values represent weighted directed edges between shares. Classic network centrality metrics such as PageRank and HITS were used to rank nodes, both in terms of incoming and outgoing edges, revealing influential and influenced shares in the market. We show that a small fraction of the shares have ranking metric values that are significantly larger than average, indicating their prevalence as either influential or influenced stocks. This observation is consistent across rankings.

We also find very good consistency between the three rankings concerning the most influenced shares in the market (top 20), while the agreement concerning the most influential is less pronounced among the three rankings. Moreover, our analysis indicates that the most traded shares (financial volume) in the market are not the most influential nor the most influenced, indicating a lack of correlation between traded volume and

information transfer.

The market network was constructed using the entire time series of price movements, across a period of 21 years. As future work, we plan to characterize influence over time by considering shorter time series and understanding the persistence of top influential and influenced shares.

References

(2017). Bolsa de valores, mercadorias e futuros do estado de são paulo.

Atsalakis, G. S. and Valavanis, K. P. (2009). Surveying stock market forecasting techniques - soft computing methods. *Expert Systems with Applications*, 36(3):5932 – 5941.

Baker, W. E. (1990). Market networks and corporate behavior. *American journal of sociology*, 96(3):589–625.

Boginski, V., Butenko, S., and Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48(2):431 – 443.

Bossomaier, T., Barnett, L., Harré, M., and Lizier, J. T. (2016). *An introduction to transfer entropy: information flow in complex systems*.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107 – 117.

Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., Gibson, D., and Kleinberg, J. (1998). Automatic resource compilation by analyzing hyperlink structure and associated text. 30(1):65–74.

He, J. and Shang, P. (2017). Comparison of transfer entropy methods for financial time series. *Physica A: Statistical Mechanics and its Applications*, 482:772–785.

Kauê Dal'Maso Peron, T., da Fontoura Costa, L., and Rodrigues, F. A. (2012). The structure and resilience of financial market networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1):013117.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632.

Korbel, J., Jiang, X., and Zheng, B. (2017). Transfer entropy between communities in complex networks. *ArXiv e-prints*.

Kwon, O. and Yang, J.-S. (2008). Information flow between stock indices. *Europhysics Letters*, 82(6):68003–p1–68003–p4.

Leahy, S. P., Levy Carciente, S., Stanley, H. E., and Kenett, D. Y. (2014). Structure and dynamics of the brazilian stock market: A correlation analysis.

Lee, G. S. and Djauhari, M. A. (2012). An overall centrality measure: The case of us stock market. *International Journal of Electrical & Computer Sciences*, 12(6).

Li, L., Shang, Y., and Zhang, W. (2002). Improvement of hits-based algorithms on web documents. In *Proc. Int. Conf. on World Wide Web*, pages 527–535.

Lyócsa, Š., Vÿrost, T., and Baumöhl, E. (2012). Stock market networks: the dynamic conditional correlation approach. *Physica A: Statistical Mechanics and its Applications*, 391(16):4147–4158.

Marschinski, R. and Kantz, H. (2002). Analysing the information flow between financial times series. *The European Physical Journal B*, 30:275–281.

Namaki, A., Shirazi, A., Raei, R., and Jafari, G. (2011). Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications*, 390(21):3835–3841.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: bringing order to the web. Technical Report 1999-66.

Petelin, D., Šindelàf, J., Přikryl, J., and Kocijan, J. (2011). Financial modeling using gaussian process models. *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, 2:672–677.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85:461–464.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shirokikh, O., Pastukhov, G., Boginski, V., and Butenko, S. (2013). Computational study of the us stock market evolution: a rank correlation-based network model. *Computational Management Science*, 10(2-3):81–103.

Tabak, B. M., Serra, T. R., and Cajueiro, D. O. (2010). Topological properties of stock market networks: The case of brazil. *Physica A: Statistical Mechanics and its Applications*, 389(16):3240–3249.

Xing, W. and Ghorbani, A. (2004). Weighted pagerank algorithm. In *Communication Networks and Services Research, 2004. Proceedings. Second Annual Conference on*, pages 305–314. IEEE.

Yang, P., Shang, P., and Lin, A. (2017). Financial time series analysis based on effective phase transfer entropy. *Physica A: Statistical Mechanics and its Applications*, 468:398–408.

Zhang, X., Yu, H., Zhang, C., and Liu, X. (2007). An improved weighted hits algorithm based on similarity and popularity. In *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)*, pages 477–480.