

Collective behavior in financial market

Thomas Kauê Dal'Maso Peron

*Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São Carlense 400,
Caixa Postal 369, CEP 13560-970, São Carlos, São Paulo, Brazil*

Francisco A. Rodrigues*

*Departamento de Matemática Aplicada e Estatística,
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo,
Caixa Postal 668, 13560-970 São Carlos, São Paulo, Brazil*

Financial market is an example of complex system, which is characterized by a highly intricate organization and the emergence of collective behavior. In this paper, we quantify this emergent dynamics in the financial market by using concepts of network synchronization. We consider networks constructed by the correlation matrix of asset returns and study the time evolution of the phase coherence among stock prices. It is verified that during financial crisis a synchronous state emerges in the system, defining the market's direction. Furthermore, the paper proposes a statistical regression model able to identify the topological features that mostly influence such an emergence. The coefficients of the proposed model indicate that the average shortest path length is the measurement most related to network synchronization. Therefore, during economic crisis, the stock prices present a similar evolution, which tends to shorten the distances between stocks, indicating a collective dynamics.

PACS numbers: 89.75.Fb, 05.45.Xt, 02.10.Ox

INTRODUCTION

Stock market is an example of complex systems, which are characterized by self-organization and emergent behavior [1]. The market regulates the relative security prices of companies worldwide without an external or central control. More specifically, individual actions made by independent investors cause an emergent behavior (the market's direction). Agents, or investors, have information about only a limited number of companies within their portfolio and must follow the regulatory rules of the market and analyze the transactions either independently or in large groupings. Since this complex system is composed of elements that interact nonlinearly generating an intricate organization, it can be naturally represented by networks [2, 3]. In this case, nodes correspond to stocks and the connection between two stocks can be defined according to their price dynamics [4–7]. This representation allows studying financial markets by taking into account methods, tools and concepts of complex networks theory [8, 9].

The early studies involving the representation of financial markets in terms of networks considered the correlation matrix of asset returns [4], in order to find a hierarchical arrangement of stocks through studying the clustering of companies [4]. In this way, the initial interest was focused mainly on the topological analysis of financial networks, in which a minimum spanning tree was generated, so as to select the most important connections and group companies according to their asset returns. Subsequent works investigated dynamical aspects of financial networks. For instance, Onnela et. al. [10]

studied the resilience of the minimal spanning tree and the consequences of economical events on its structure. They verified that the evolving analysis could capture economic instabilities, such as the *Black Monday*, which occurred on October 19th, 1987.

In the current work, we are also interested in examining dynamic aspects of financial market networks. More specifically, it is analyzed the emergence of collective behavior by considering concepts of network synchronization [11]. The emergence of collective behavior occurs when stock prices exhibit a similar tendency, defining the market's direction. We have verified that such collective dynamics occurs during financial instabilities. For instance, networks generated during financial crisis (e.g. the Black Monday and the the global economic crisis of 2008) present a higher degree of synchronization than networks generated in other periods.

Besides the dynamical analysis, we examine how the network organization changes during financial crisis, which cause the collective behavior. A regression model [12] is adopted to verify the network properties that influence the degree of synchronization. It revealed that the increase in the average shortest path length decreased the level of synchronization. Thus, during economic instabilities, since stock prices tend to evolve similarly, as verified by the increase in the network synchronization, the network average shortest path is reduced. The same effect is observed in the average clustering coefficient. Thus, the emergence of the collective behavior is a consequence of network reorganization during financial crashes.

The following sections discuss the results mentioned above in more detail, describing the construction of fi-

financial market networks, the concepts of synchronization and how we can relate the structure and dynamics of these networks by considering regression analysis.

COLLECTIVE DYNAMICS

Financial market networks are constructed by considering correlation matrices [4]. More specifically, we take into account the correlation coefficient between the time series of two stocks i and j , which is given by

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \quad (1)$$

where Y_i is the return of a stock i given by $Y_i = \ln P_i(t) - \ln P_i(t-1)$ and $P_i(t)$ is the closure at day t . The distance between two stocks i and j can be calculated by [4]

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

This distance is close to zero if two stocks present similar price evolution, i.e. correlated time series. On the other hand, two stocks whose prices have contrary tendency have distance close to two. These distances between N stocks form a symmetric $N \times N$ distance matrix \mathbf{D} , which is the weighted matrix that represents the topology of the financial market.

We constructed financial networks by taking into account a database formed by the daily prices of 3,799 stocks traded at New York Stock Exchange. These stock prices are available at the *Yahoo!* financial website [23]. We have selected $N = 348$ stocks from this set, which are the stocks that have historical data from January 1986 to February 2011. From these data, an amount of 6,008 closure prices per stock is generated.

In order to study the collective behavior of the financial market, it is necessary to consider the time evolution of the respective networks. By setting a time window of length $\Delta t = 28$ days and moving this window along time, we can obtain a sequence of networks — each one describing the market organization inside each window. This window is moved by an amount of $\delta t = 1$ day and a new network is obtained after each displacement. More specifically, the first network is constructed from the time series starting at day $t_1^1 = 1$ and ending at day $t_2^1 = 28$, the second network from the time series starting at day $t_1^2 = 2$ and ending at day $t_2^2 = 29$, and so on. This process is repeated until the end of the original time series has been reached. Since the whole database corresponds to a time series composed of 6,007 returns per stock, a total of 5,979 networks was achieved. Differently from previous works (e.g. [4, 10]), instead of using the spanning trees, the fully connected weighted networks were taken into account to represent the financial networks. This choice allowed considering the whole information about network

topology, since the spanning tree methodology uses only the most important connections, ignoring the majority of the links.

The collective dynamics of complex systems can be studied by taking into account the concepts of network synchronization [13]. This dynamic process causes the emergence of collective phenomena when there occurs the onset of synchronization [11]. A model to describe the synchronization of a system was proposed by Kuramoto [14]. In complex networks, each oscillator i obeys an equation of motion given by

$$\dot{\theta}_i = \omega_i + \lambda \sum_{j=1}^N a_{ij} \sin(\theta_j - \theta_i), \quad i = 1, \dots, N, \quad (3)$$

where λ is the coupling strength, ω_i is the natural frequency of oscillator i (generally distributed according to some function $g(w)$), and a_{ij} are the elements of the adjacency matrix A , which represent the topology of the complex systems. More specifically, elements $a_{ij} = 1$ if two nodes i and j are connected, and $a_{ij} = 0$, otherwise. Coupling strengths higher than a determined threshold λ_c produce the onset of synchronization. As the coupling strength is increased, more and more oscillators present individual phase around the average phase of the whole system and the network settles in the complete synchronized state.

In order to study the synchronization dynamics in financial market networks, we adapted the Kuramoto model [15] by considering the distance between stocks, d_{ij} , which are the elements of the distance matrix D (see Equation 2).

$$\dot{\theta}_i = \omega_i + \lambda \sum_{j=1}^N e^{-\alpha d_{ij}} \sin(\theta_j - \theta_i), \quad i = 1, \dots, N. \quad (4)$$

We considered a value of $\lambda = 0.1$. Parameter α is a constant and we adopted $\alpha = 5$. The collective dynamics of the whole system can be measured by the macroscopic complex order parameter,

$$r(t) = \left| \frac{1}{N} \sum_{j=1}^N e^{i\theta_j(t)} \right|, \quad (5)$$

where $0 \leq r(t) \leq 1$ measures the phase coherence of populations. When $r(t) \approx 1$, all nodes oscillate with similar phases

Figure 1 presents the evolution of the macroscopic complex order parameter for the generated financial market networks for $t \approx 10^4$. During most of the time, the synchronization level does not suffer a high variation and it is kept around $r = 0.05$. On the other hand, three prominent peaks can be observed in specific times. The first, at the beginning of the time series, corresponds to the networks constructed between 09/16/1987 and

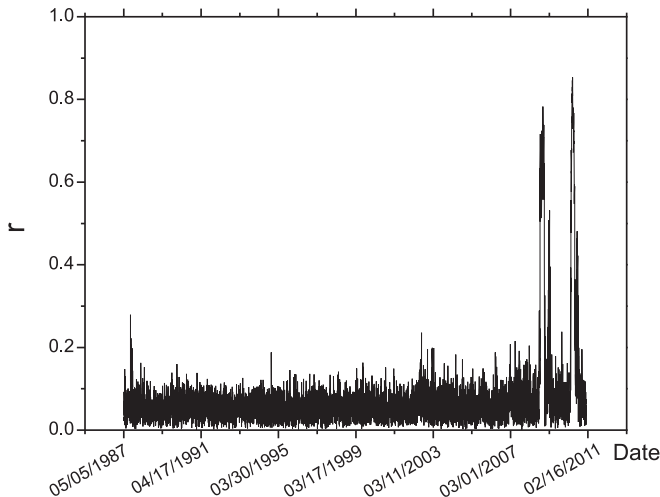


FIG. 1: Time evolution of the network synchronization.

10/26/1987. A financial instability known as *Black Monday* occurred inside this interval, *i.e.* on October 19th, 1987. The other two high peaks correspond to networks generated during the financial global crisis, which started in 2008. The last peak started decreasing only in 2010, when the global crisis was attenuated. Therefore, the time series indicate that during economic instabilities, financial networks tend to become more synchronizable, *i.e.* the phase coherence increases, indicating the emergence of a collective behavior of stock prices, since most of them tend to have a similar evolution. During financial crashes, stock prices move to the same direction presenting a behavior similar to those observed in other type of complex systems, such as cellular movements in tissue formation, flock of birds, and social behavior in human societies. On the other hand, in “normal” periods, stock prices tend to evolve in a more independent fashion and the behavior of each stock is not synchronized as verified in financial crisis.

STRUCTURE AND DYNAMICS

The synchronization of financial market networks is determined by the network organization. Thus, we investigate the topological properties that influence such synchronization and therefore understanding how financial market networks evolve in different situations. This investigation is one of the main problems in complex networks research [2]. For instance, Watts and Strogatz [13] suggested that the decrease in the average shortest path length in small-world networks facilitated a more efficient coupling and, therefore, enhanced synchronization. They examined this tendency by increasing the probability of rewiring, which creates more shortcuts between pairs of vertices. Other works have investigated the influence of the distribution of connectivity [16], clustering coefficient

(*e.g.* [17]) and degree correlations [18]. In all these papers, networks are modified in order to keep some properties constant while others are varied. However, when the parameters characterizing the original network are modified, other network properties also change, raising difficulties to draw conclusions about the relationship between one single statistical property of network and its synchronization level [11]. To overcome this difficulty, we propose here the use of statistical regression analysis [12]. This methodology helps understand how the value of the dependent variable changes when any independent variable is varied — while the other independent variables are held fixed. Thus, regression allows verifying the relationship between synchronization and network properties in a proper fashion.

Although a regression model does not imply in a cause-effect relationship between variables, it allows to obtain an association between variables. In the case of complex networks theory, it is known that the topology have a direct effect on dynamical process [2]. So, the cause-effect relationship is determined theoretically and the regression allows to quantify such a relationship.

Three network measurements are considered to determine those that influence the network synchronization. The weight of the connection between two stocks is given by their distance (see Equation 2). The strength of a node, s , is defined as the sum of the weights of its corresponding edges. From this measurement, it is possible to quantify the network strength heterogeneity by taking into account the Shannon entropy of the strength distribution, $P(s)$, *i.e.*

$$H = - \sum_i P(s_i) \log_2 P(s_i). \quad (6)$$

The clustering coefficient for weighted networks is given by [19]

$$cc_i^w = \frac{1}{k_i(k_i - 1)} \sum_{j,k} (\hat{w}_{ij} \hat{w}_{ik} \hat{w}_{jk})^{1/3}, \quad (7)$$

where the weights are normalized by the maximum weight in the network, $\hat{w}_{ij} = w_{ij} / \max_{pq} (w_{pq})$, $p, q = 1, 2, \dots, N$. The respective global measurement is the average cluster coefficient

$$C = \frac{1}{N} \sum_i cc_i^w. \quad (8)$$

The weighted average shortest path length permits quantifying the average topological distance between the stocks,

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} \tau_{ij}. \quad (9)$$

where τ_{ij} is the length of the shortest distance between stocks i and j . Note that a more similar time evolution

of a pair of stock prices implies shorter distance between them.

Considering these measurements, we have proposed a regression model to determine the relationship between the structure and dynamics of financial markets. This model is given by

$$r = \beta_0 + \beta_1 H + \beta_2 C + \beta_3 \ell + \varepsilon. \quad (10)$$

where ε is assumed to be normally distributed with mean zero and standard deviation σ . This variable captures all other factors which influence the dependent variable r other than the network measurements considered. Thus, in the proposed regression model, the dependent variable r represents the order parameter $r(t)$ obtained for a large value of $t \approx 10^4$. On the other hand, the independent variables are the network topological measurements, i.e. the Shannon entropy of the strength distribution (H), the average clustering coefficient (C), and the average shortest path length (ℓ). The contribution of each network measurement i for the synchronization is given by coefficient β_i .

The regression analysis adopted here considers the least-squares estimator. In this case, some assumptions with respect to the database must be done. First, the networks considered must be independent. Thus, we take into account only networks constructed by disjoint time series, i.e. time series without time intersections. More specifically, networks constructed from windows starting at day 1, day 29, day 57, and so on. The second assumption is related to the fact that the independent variable must present, at least approximately, a normal distribution. Figure 2 shows the normal probability plot, which is a graphic technique for normality testing. Despite some outliers, the points lie close to a straight line, which allows concluding that the distribution of r is consistent with a sample from a normal distribution. Interesting to note that the outliers correspond to the networks obtained during financial crisis, i.e. they correspond to the highest values of r (see Figure 1). Thus, during financial instabilities, a financial network has an anomalous organization, differing strongly from normal periods. The presence of these outliers makes the least squares estimation inefficient and can also be biased. Therefore, we adopted the robust regression [20, 21] to obtain the coefficients of the regression model, since this method allows treating data with outliers. Robust regression works by assigning a weight to each data point. At the first step, equal weights are assigned to each point and ordinary least squares are used for estimation. In the next step, weights are recalculated so that points farther from predictions in the previous iteration are given lower weight. The coefficients of the model are recalculated by using ordinary least squares. The process is repeated until the values of the coefficient estimates have converged within a specified tolerance.

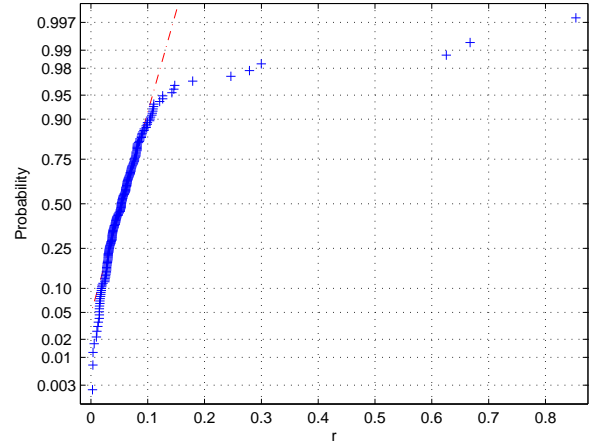


FIG. 2: Normal probability plot for r .

By performing the robust regression with ordinary least squares, the following linear model was obtained,

$$r = 1.74 - 0.090H - 0.103C - 0.826\ell + \varepsilon. \quad (11)$$

Coefficients $\beta_1 = -0.09$ suggests that the value of the Shannon entropy of the strength distribution does not influence significantly the network synchronization. Thus, the heterogeneity in the strength distribution causes a small decrease in the synchronization level. The clustering coefficient and the average shortest path length also contribute negatively to the onset synchronization. A large clustering coefficient implies many transitive connections and, consequently, more redundant paths in the network, which tend to prejudice the emergent behavior. The major influence on the degree of synchronization is due to the average shortest path length, $\beta_3 = -0.826$. This network property has a negative influence on the network synchronization, since larger distances delay the appearance of the fully synchronized state. During financial crisis, the average shortest path diminishes, leading to an increase in the network synchronization level. This event indicates that the stock prices present a collective behavior, which implies an increase in the correlation between pairs of nodes. During economic crisis, the stock prices tend to exhibit a similar evolution. This result was expected, since the prices tend to suffer a strong decrease during crashes, defining the market's direction. Thus, the increasing in the degree of synchronization is mainly due to the decrease in the average shortest path length during financial crisis.

CONCLUSIONS

Financial markets are examples of complex systems which can be represented as networks. The emergence

of a collective behavior has been observed in many complex systems, such as flocking of birds and swarm intelligence in ant colonies [22]. This phenomenon is a consequence of the agents which interact according to local rules. Since financial markets are examples of complex systems, a collective behavior is expected to be also observed. The current work analyzes financial complex networks whose nodes represent stocks and the connections between two nodes are established according to a measurement related to the correlation between the temporal price evolutions of the respective stocks. We take into account the Kuramoto model, which describes the synchronization of a system, to study the emergence of collective dynamics. Our results suggest that during financial instabilities, stock prices tend to move to the same direction, becoming more synchronizable, which indicates the emergence of a collective behavior. This tendency was observed during the Black Monday on October 19th, 1987, and during the last global crisis, which started in 2008 and was attenuated in the early 2010. In these periods, financial networks presented the highest degrees of synchronization.

In order to determine the topological factors that influence such a synchronization and understand the reorganization of the financial market networks during financial instabilities, a statistical regression model was used. Since the data considered here present some outliers, which correspond to the networks generated during financial crashes, we have taken into account the robust regression approach. The regression model suggests that the clustering coefficient and the average shortest paths length tend to prejudice the network synchronization. Thus, during economic crisis, the stock prices present similar evolution, which results in a higher correlation between time series. Such a correlation tends shorten distances in the networks, enhancing synchronization.

In summary, the results presented in the current work help to understand how the collective dynamics emerge in financial markets. In addition, the regression analysis permits determining how structural properties influence this network dynamics.

Acknowledgments

Francisco A. Rodrigues would like to acknowledge CNPq (305940/2010-4) and FAPESP (2010/19440-2) for the financial support given to this research. Thomas K. D. M. Peron would like to acknowledge Fapesp for spon-

sorship.

* Electronic address: francisco@icmc.usp.br

- [1] L. Amaral and J. Ottino, The European Physical Journal B-Condensed Matter **38**, 147 (2004).
- [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Physics reports **424**, 175 (2006).
- [3] L. Costa, F. Rodrigues, G. Travieso, and P. Boas, Advances in Physics **56**, 167 (2007).
- [4] R. Mantegna, The European Physical Journal B-Condensed Matter and Complex Systems **11**, 193 (1999).
- [5] N. Vandewalle, F. Brisbois, and X. Tordoir, Quantitative Finance **1**, 372 (2001).
- [6] J. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, Physica Scripta **2003**, 48 (2003).
- [7] J. Onnela, A. Chakraborti, K. Kaski, and J. Kertész, Physica A: Statistical Mechanics and its Applications **324**, 247 (2003).
- [8] M. Newman, *Networks: an introduction* (Oxford Univ Pr, 2010), ISBN 0199206651.
- [9] L. Costa, O. Oliveira Jr, G. Travieso, F. Rodrigues, P. Boas, L. Antiqueira, M. Viana, and L. Da Rocha, Advances in Physics **60**, 329 (2011).
- [10] J. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, Physical Review E **68**, 056110 (2003).
- [11] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Physics Reports **469**, 93 (2008).
- [12] D. Montgomery, E. Peck, G. Vining, and J. Vining, *Introduction to linear regression analysis* (Wiley New York, 2001).
- [13] D. Watts and S. Strogatz, Nature **393**, 440 (1998).
- [14] J. Acebrón, L. Bonilla, C. Vicente, F. Ritort, and R. Spigler, Reviews of Modern Physics **77**, 137 (2005).
- [15] N. Nakagawa and Y. Kuramoto, Physica D: Nonlinear Phenomena **75**, 74 (1994).
- [16] T. Nishikawa, A. Motter, Y. Lai, and F. Hoppensteadt, Physical review letters **91**, 14101 (2003).
- [17] P. McGraw and M. Menzinger, Physical Review E **72**, 015101 (2005).
- [18] A. Motter, C. Zhou, and J. Kurths, EPL (Europhysics Letters) **69**, 334 (2005).
- [19] J. Onnela, J. Saram, "aki, J. Kertész, and K. Kaski, Physical Review E **71**, 065103 (2005).
- [20] P. Holland and R. Welsch, Communications in Statistics-Theory and Methods **6**, 813 (1977).
- [21] J. Street, R. Carroll, and D. Ruppert, American Statistician pp. 152–154 (1988).
- [22] M. Mitchell, *Complexity: A guided tour* (Oxford University Press, USA, 2009).
- [23] <http://finance.yahoo.com>