A Persistent Entropy Automaton for the Dow Jones Stock Market

Marco Piangerelli(✉), Luca Tesei, and Emanuela Merelli

School of Sciences and Technology, University of Camerino,
Via Madonna delle Carceri 9, 62032 Camerino, MC, Italy
{marco.piangerelli,luca.tesei,emanuela.merelli}@unicam.it

Abstract. Complex systems are ubiquitous. Their components, agents, live in an environment perceiving its changes and reacting with appropriate actions; they also interact with each other causing changes in the environment itself. Modelling an environment that shows this feedback loop with agents is still a big issue because the model must take into account the emerging behaviour of the whole system. In this paper, following the S[B] paradigm, we exploit topological data analysis and the information power of persistent entropy for deriving a persistent entropy automaton to model a global emerging behaviour of the Dow Jones stock market index. We devise early warning states of the automaton that signal a possible evolution of the system towards a financial crisis.

Keywords: Complex systems · S[B] paradigm · Emerging behavior · Topological data analysis · Stock market

1 Introduction

A complex system is any system consisting of a great number of heterogeneous entities interacting with each other within an environment to shape an emerging behavior. Such emerging behaviour depends on a non-trivial space of correlations that derive from the interplay of agents *entangled in loops of non-linear interactions*. In the metaphor of the flock of starlings, any environmental change perceived by the starlings during their flight is visible in the formation of the flock shape due to their reaction. This implies that there is an underlying feedback loop between the agents and the global system.

Mastering the complexity of these systems has always been a challenge in almost every branch of science. In computer science, process-, actor- and agent-based models and languages have been developed for describing the behaviour of complex software systems [1,2,6,11]. All these approaches require an *a priori* knowledge of the basic rules governing the dynamic of the system in order to define the behaviour of the components and of the environment. Unfortunately, for natural or social phenomena, it is quite impossible to have enough knowledge about the real interaction rules. However, global information about the system

Published by Springer Nature Switzerland AG 2019
H. Hojjat and M. Massink (Eds.): FSEN 2019, LNCS 11761, pp. 37–42, 2019.
https://doi.org/10.1007/978-3-030-31517-7_3

is hidden inside phenomenological data produced by the individual components. Thus, we need suitable methods to extract a specific model of interest.

Topological Data Analysis (TDA) is a relatively new field of study in which topology driven methods are used to analyse big collections of data [4,5,8]. The $S[B]$ paradigm is a general framework of modelling in which a complex system is described as a pair of entangled systems: S, the global environment, and B, a set of interactive agents [13,14]. Persistent Entropy (PE) is a Shannon-like entropic measure able to describe the global dynamics of a complex system [16]. PE has been used for studying complex phenomena in different fields [15–17]. As shown in [16], by analysing the trajectories of PE and its derived quantities, an automaton, called Persistent Entropy Automaton (PEA), which models the global dynamics of the system under study, can be manually devised.

The global financial system is one of the most important, human-made, complex systems. This system is composed of multiple interacting autonomous components or "selfish" agents, who - very often - act for their own benefit, and of complex interactions among those components. Each component behaves according to his/its own strategies, under the influence of the environment and interacting with other heterogeneous components. Classical tools for analysing and modelling such systems operate under a range of rather unrealistic assumptions. For example, interactions are normally abstracted with equations: this implies that the system reaches the equilibrium through non-linear optimisation methods rather than emerging from the agents interactions [12].

In this paper we use TDA to construct a data space from the components of the Dow Jones stock market index. The considered data set is the time series of the daily log-returns of Dow Jones' components from 1987 to 2017. From the data space, we calculate the PE and we devise a PEA whose locations model global states of the stock market. We show that early warnings about the emergence of the already occurred financial crisis can be identified by the PEA.

2 Methods

TDA employs concepts and principles of the field of computational topology to reveal higher dimensional patterns hidden in big data sets [5,8]. Computational topology studies invariants of shapes among which the so called Betti numbers, or barcodes, that characterise the existence of n-dimensional holes in the topological data space. TDA builds a discrete topological space, a simplicial complex, following a filtration procedure. In this work we use the Vietoris-Rips complex filtration that works on point clouds [8]. At each step of the filtration the persistent homology is computed yielding a collection of barcodes that indicate the life span of the topological invariants. PE is then computed from the barcodes [15].

The 24 time series of the considered Dow Jones components were mapped into a point cloud using a sliding window of 50 days and scrolling one point at a time with superposition of 49 points. Each window then produced 50 points in $\mathbb{R}^{24}$. This technique has been demonstrate suitable for studying the time-varying properties of systems similar to the one we are studying [9,15].

The mechanism driving critical transitions in complex systems is called *tipping point*, which is an abrupt qualitative change in the behaviour of a dynamical system when one or more control parameters change. In approaching a tipping point, a complex system shows a phenomenon called Critical Slowing Down (CSD), which can be considered an Early Warnings Signal (EWS) for the critical transition [3,18]. Since PE describes a system globally, it contains a summary of the knowledge about the system. Moreover, it can be considered a time series itself and can be calculated for all the dimensions. Thus, we study the *total* PE time series ($\mathrm{PE_{tot}}$) - calculated as the sum of PEs for all the dimensions - with an analytical approach. The goal is to identify EWSs about a crisis by detecting the occurrence of tipping points. The obtained $\mathrm{PE_{tot}}$ is shown in Fig. 1.

To delimit the CSD areas we used an adaptation of the W_2 index, i.e. a combination of statistical indices (coefficient of variation, 1-lag autocorrelation, and kurtosis), described in [7]. W_2 is computed from $\mathrm{PE_{tot}}$ with the R package "earlywarnings" using another sliding window of size 450. Thus, W_2 is another time series and it is plotted along with its running average and 2σ confidence bands in Fig. 1. Potential areas of CSD are identified by finding points where $W_2 > \overline{W_2}+2\sigma$ [7]. These areas are shown in Fig. 1 with black bands and represent the EWSs in our system.

Identified CSD areas can be used to define PEA states. A PEA *monitors* the PE and derived functions w. r. t. equilibrium conditions that define its states [16]. It remains in a state s as long as the associated equilibrium condition $ec(s)$ is satisfied. When it is violated, the PEA exits s and starts a non-instantaneous transition, which can be seen as an adaptation phase. This adaptation may end into an adjacent PEA state s' as soon as $ec(s')$ is satisfied or may not terminate. This definition is based on the fact that PEA states are devised from the observed trajectories of PE and derived functions. Indeed, it may happen that the monitored functions exhibit evolutions that were not identified as equilibrium conditions. This is expected for natural complex systems for which "complete" models can not be established. The main difference between a PEA and a hybrid automaton, which is a top-down defined model not considering unknown evolutions [10], is essentially in this different perspective.

3 Persistent Entropy Automaton of Dow Jones

In the $S[B]$ paradigm the structural level S is a model of the global dynamics of the system and the behavioural level B is a model of the local interactions among the entities of the system [13]. In this study, the behavioural level B is represented by real human agents that produced the data that we use. The structural level S is defined as a PEA in the following.

We consider two financial crises, dot-Com and Lehman Brothers Crash, both represented with coloured bands in Fig. 1. We devise the PEA by monitoring the functions $\mathrm{PE_{tot}}(t)$, $W_2(t)$, their running average $\overline{\mathrm{PE_{tot}}}(t)$, $\overline{W_2}(t)$ together with their derivatives indicated with a dot over the symbol. Each discrete instant t corresponds to one day observation. The derived PEA is called PEA_{W_2}, is

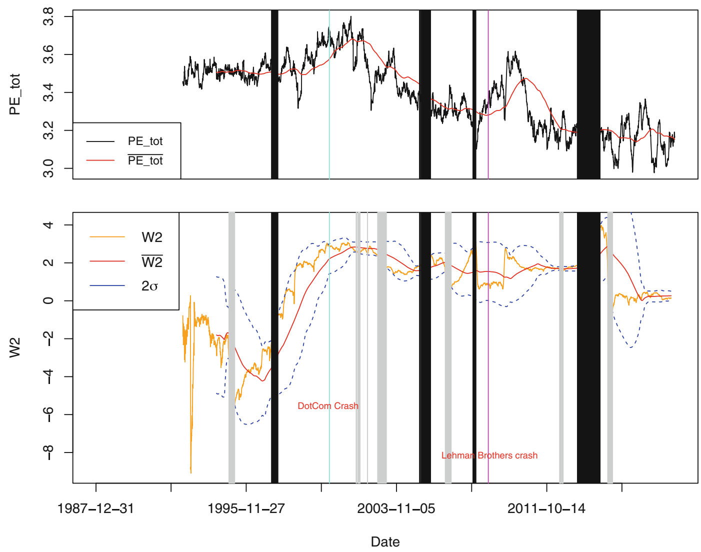

Fig. 1. Plot of PE_{tot} and its running average (above). Plot of W_2 and its running average with confidence bands (below). The grey and black vertical bands correspond to the relative states of the automaton in Fig. 2. The thin coloured vertical bands correspond to financial crises. (Color figure online)

depicted in Fig. 2 and its states are described in the following. A *stable* state is characterised by the equilibrium condition $|W_2 - \overline{W_2}| < 2\sigma$, that is W_2 does not exceed the confidence bands. The state called Stable is the initial one and holds this condition. As soon as the functions violate the stable condition, PEA_{W_2} exits state Stable and starts an adaptation. The only state in which the adaptation can end is the one called Grey, a state indicating that there was in the past at least one violation of the stable state condition. The equilibrium condition of Grey is $\dot{\overline{W_2}} \approx 0 \wedge \dot{\overline{\text{PE}_{\text{tot}}}} < 0 \wedge W_2 < \overline{W_2} - 2\sigma$, which means that the running average of W_2 has minimal oscillations, the running average of the total PE is decreasing and W_2 exits the confidence band -2σ. Visually, the periods in which PEA_{W_2} stays in this state are represented by the grey bands in Fig. 1. State White corresponds visually to the period after a grey band, it has the stable equilibrium condition and records the fact that the system entered at least once state Grey. After White, another grey band can occur (in this case the PEA goes back to Grey) or a black band occurs. A black band corresponds to state Black. This is the early warning state because from Black the system can only evolve to state Tipping that represents a tipping point. In Tipping a crisis, represented by the dashed transition towards state Stable, can occur or the system can return to state Grey. The dashed transition can be interpreted as the occurrence of a phase transition of the system.

PEA_{W_2} is then able to give a warning, in state Black, that a crisis may occur, without giving a prediction. However, if the current state is different from Black, the model says that a crisis cannot occur immediately: there must be at least one (or more) adaptations before the tipping point state is reached.

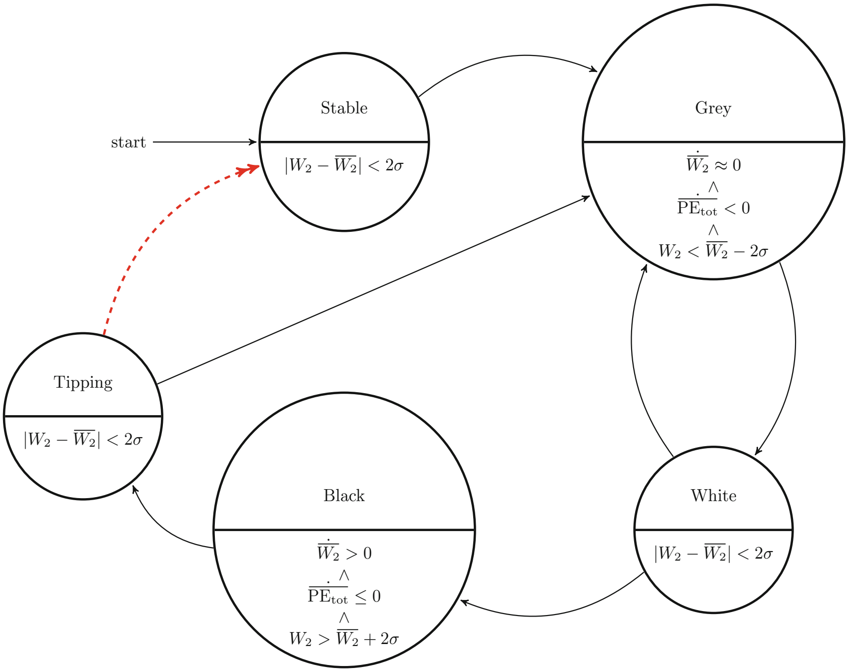

Fig. 2. PEA_{W_2}, describing the transitions among the global states of the system.

4 Conclusions

In this work we have modelled the global dynamics of a complex system by manually devising a PEA. TDA has been used for analysing the phenomenological data of the system and PE has been calculated from a topological space derived from a data set. The application domain is the Dow Jones stock market. The derived automaton models the global behaviour of the market and is able to recognise a tipping point state in which a financial crisis may occur and previous states in which there is some degree of warning but there is not an immediate alarm because some other adaptations are required to reach the tipping point state. The transition that goes from the tipping point state to the stable state can be interpreted as a phase transition of the system.

Despite the encouraging result we are aware that the proposed analysis presents some limitations: one is about the peculiarity of the data set that does not allow one to set up a statistical validation of the results because of the unicity of the phenomenon under study, for which other instances do not exist.

Another regards the computation of the indices for deriving the W_2, for which it is necessary to try different lengths of the sliding windows.

References

1. Arbab, F.: Reo: a channel-based coordination model for component composition. Math. Struct. Comput. Sci. **14**(3), 329–366 (2004)
2. Baier, C., Sirjani, M., Arbab, F., Rutten, J.: Modeling component connectors in Reo by constraint automata. Sci. Comput. Program. **61**(2), 75–113 (2006)
3. Battiston, S., et al.: Complexity theory and financial regulation. Science **351**(6275), 818–819 (2016)
4. Binchi, J., Merelli, E., Rucco, M., Petri, G., Vaccarino, F.: jHoles: a tool for understanding biological complex networks via clique weight rank persistent homology. Electron. Not. Theor. Comput. Sci. **306**, 5–18 (2014)
5. Carlsson, G.: Topology and data. Bull. AMS **46**(2), 255–308 (2009)
6. De Nicola, R., Ferrari, G.L., Pugliese, R.: KLAIM: a kernel language for agents interaction and mobility. IEEE Trans. Softw. Eng. **24**, 315–330 (1998)
7. Drake, J.M., Griffen, B.D.: Early warning signals of extinction in deteriorating environments. Nature **467**(7314), 456 EP - (2010)
8. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. AMS, Providence (2010)
9. Gidea, M., Katz, Y.: Topological data analysis of financial time series: landscapes of crashes. Phys. A **491**, 820–834 (2018)
10. Henzinger, T.: The theory of hybrid automata. In: Inan, M., Kurshan, R. (eds.) Verification of Digital and Hybrid Systems, NATO ASI Series (Series F: Computer and Systems Sciences), vol. 170, pp. 265–292. Springer, Heidelberg (2000). https://doi.org/10.1007/978-3-642-59615-5_13
11. Jennings, N.R.: An agent-based approach for building complex software systems. Commun. ACM **44**(4), 35–41 (2001)
12. Landini, S., Gallegati, M., Rosser, J.B.: Consistency and incompleteness in general equilibrium theory. J. Evol. Econ. (2018). https://doi.org/10.1007/s00191-018-0580-6
13. Merelli, E., Paoletti, N., Tesei, L.: Adaptability checking in complex systems. Sci. Comput. Program. **115–116**, 23–46 (2016)
14. Merelli, E., Pettini, M., Rasetti, M.: Topology driven modeling: the IS metaphor. Nat. Comput. **14**(3), 421–430 (2015)
15. Merelli, E., Piangerelli, M., Rucco, M., Toller, D.: A topological approach for multivariate time series characterization: the epileptic brain. In: EAI Endorsed Transactions on Self-Adaptive Systems (2016). https://doi.org/10.4108/eai.3-12-2015.2262525
16. Merelli, E., Rucco, M., Sloot, P., Tesei, L.: Topological characterization of complex systems: using persistent entropy. Entropy **17**(10), 6872–6892 (2015)
17. Piangerelli, M., Rucco, M., Tesei, L., Merelli, E.: Topological classifier for detecting the emergence of epileptic seizures. BMC Res. Not. **11**, 392 (2018)
18. Scheffer, M.: Complex systems: foreseeing tipping points. Nature **467**, 411 EP (2010)