Analysis of the determinism of time-series extracted from social and biological systems

Fortuna Luigi, Frasca Mattia, Gambuzza Lucia
Valentina, Sarra Fiore Angelo
DIEES – Faculty of Engineering
University of Catania
Catania, Italy

Ramzy S. Ali, Mofeed T. Rashid
Electrical Engineering Department
University of Basrah
Basrah, Iraq

Abstract—**Self-organizing systems arise in many different fields. In this work we analyze data from social and biological systems. A central question is to demonstrate the presence of the determinism in time-series extracted from such systems that appear apparently not correlated but that are two good benchmarks for the study of complexity in real systems. We will apply the Kaplan test and we will define an order parameter for the biological data to characterize the complexity of the system.**

I. INTRODUCTION

A particular process may be studied with the analysis of a sequence of data points (Y_1, Y_2, ..., Y_t) measured at successive times spaced at uniform time intervals (1, 2, ..., t): its time series. In fact, with this analysis, it is possible to find existing internal structure (such as autocorrelation, trend and seasonal variation) that can not be discovered using other methods of analysis.

There are many interesting things in time series analysis: the observed Y_t may be the result of a summation of "noise" values (ε_t) and of a smooth signal (η_t): $Y_t = \varepsilon_t + \eta_t$ and we want to find the value of η_t; we may want to develop a mathematical model, and to estimate its parameters, which explains the observed pattern of Y_1, Y_2, ..., Y_t; on the basis of observations Y_1, Y_2, ..., Y_t, we may want to predict the value of Y_{t+L} (L ≥ 1), with indications of the uncertainty in the prediction; we may wish to intervene with the process to modify the future values [1].

Standard analyses of time series involve:
- smoothing/trend assessment;
- assessment of/accounting for seasonality;
- assessment of/exploiting "serial correlation".

A very interesting problem in the study of a periodic and apparently irregular time series is to determine if the time series arises from a stochastic process or has deterministic component, generated from chaotic dynamics with finite degrees of freedom. In deterministic case it is possible to predict the behavior of the process with low error, while in the stochastic case this is not possible and prediction errors have to be taken into account. Another interesting case is the mixed-one where the deterministic and stochastic components coexist and the behavior is the sum of these two simple components.

The presence or not of a deterministic component in a time series determines which approaches are appropriate for investigating the time series and its generating system [2].

Several methods of nonlinear dynamical analysis have been developed in years to detect determinism in time series. To assert the existence of a deterministic component in a process, the most useful test is the so-called Kaplan test [3][4]. It is a powerful method to characterize a time series, based on a statistics indicating the degree to which trajectories passing through a small region of an embedding space are parallel. It provides for the calculus of a parameter, Ln, that is the mean value of the vectors of the trajectories for all the boxes that have n elements. The evolution of the value of Ln can be considered significant to characterize the nature of a system: a value of Ln close to 1 indicates the presence of determinism while a value close to 0 indicates that the system is fully stochastic.

Another parameter for the analysis of the time series investigated derives from the specific problem analyzed. In his paper [5] Vicsek proposed a model to explain the emergence of collective motion in a system of interacting self-propelled particles. In this paper a parameter, v_a, that is the normalized modulus of the mean velocity of particles and may be considered an index that measures the tendency of the system to an ordered behavior, is defined. If the motions of particles is random then the value of v_a tends to 0 while it tends to 1 if there is an order, in fact, in this case all the velocity are oriented on the same directions.

In this paper we analyze time series obtained from a social system: Facebook and from a biological systems: the motions of Artemia Salina (brine shrimps), with the aim to investigate the presence of the deterministic component.

II. FACEBOOK TIME SERIES

Facebook, originally developed for students of U.S. colleges, is the most popular on-line social network today that put in relation people of the whole world while initially it worked only for a list of supported academic institutes.

A user may share pictures, video, music, add friends and talk with them with private messages or with messages that are available for all in the so called "wall", add applications for gaming, update personal profiles to notify friends about themselves, join groups and networks organized by city, workplace, school, and region. Facebook is available in more than 65 translations, and at present there are more than 300 millions of active users in the world with an average number of friends for user of about 130 persons.

For this work we have developed an application able to monitor the number of on-line friends of a Facebook user with about 300 friends. An important thing is that the number of friends is constant during our analysis to avoid erroneous measurements. We have used this application for a week with sampling frequency of 5 minutes, obtaining the trend shown in Figure 1. It is possible to observe that the maximum number of on-line friends is 27 while the minimum is 2.
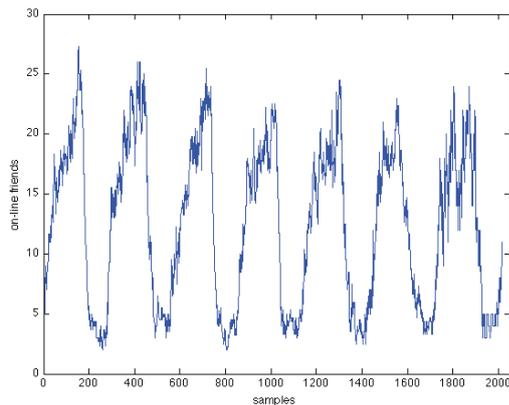

Figure 1. Trend of the number of on-line friends of a Facebook user for a week, sampled with frequency f=5 min.

To study correctly this trend we have chosen to filter the time series to eliminate the sinusoidal component that represents the day and night cycle. To obtain this we have analyzed the power spectrum of the time series to know at which frequency apply a filter. So we applied a second order Butterworth high pass filter with cutoff frequency of 0.2Hz obtaining the signal shown in Figure 2.
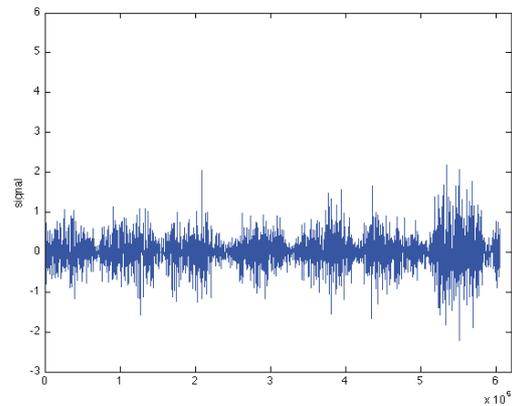

Figure 2. Signal obtained applying the Butterworth high-pass filter.

We have applied the Kaplan test to detect determinism in this time series. The results of this analysis are shown in Figure 3. where the trend of Ln parameter for the Kaplan test of the filtered time series is shown.
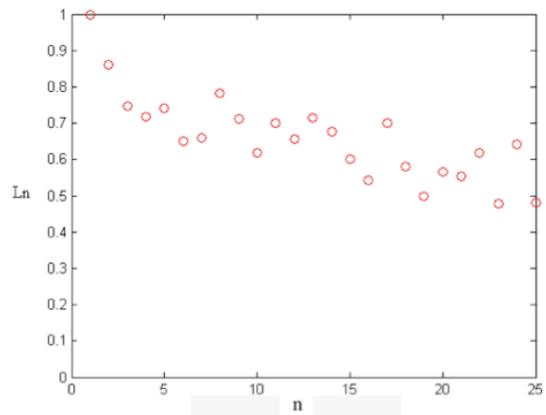

Figure 3. Trend of the Ln parameter for the Kaplan test of the filtered time series.

For comparison Figure 4. shows the trend of the Ln parameter for the Kaplan test for a surrogate signal with the same mean and variance of the filtered signal.
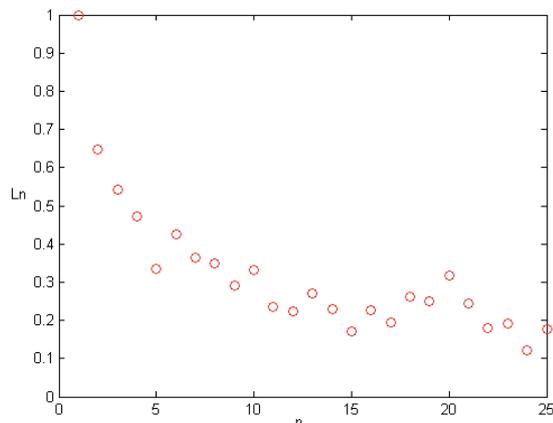
Figure 4.   Trend of the Ln parameter for the Kaplan test of a random signal with with the same mean and variance of the filtered signal

A deterministic component in the time series obtained starting from on-line Facebook users is evident: in fact the trend of Ln parameter in the first case is totally different than in the second one.

III.   ARTEMIA TIME SERIES

Artemia, known as brine shrimp, is a small aquatic crustacean that lives in salt marshes, where water is salty and there are no predators; in this way it may grow and reproduce itself without problems, reaching up to a maximum of four months with a maximum size of about 2 cm. Artemia produces eggs that may give rise to nauplies (larvae) or, if the conditions are not favorable, it can produce cysts that can last for decades and then give rise to nauplies when favorable conditions return. Due to this extraordinary qualitities of resistance, it lives on Earth from 100 million years and is, in fact, contemporary of the dinosaurs. A photo of a group of nauplies and eggs of Artemia is shown in Figure 5. [6][7].
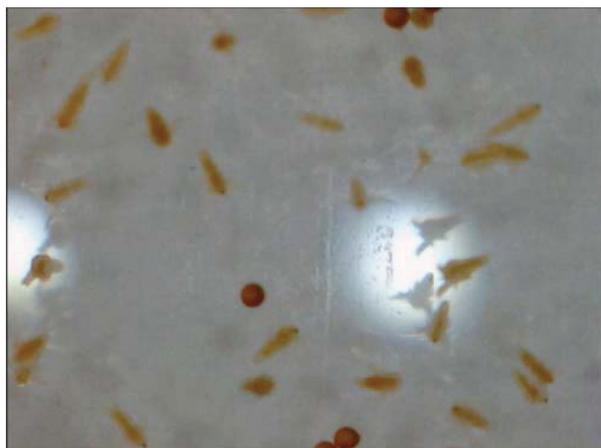

Figure 5.   Microscope photo of a group of eggs and nauplies

In this work we have analyzed the motion of Artemia in two different conditions: with diffused light and with two LEDs in opposite sides of the box that contains the Artemias.

To obtain these datasets we have divided the problem into three steps:

- Realization of an experimental setup and data acquisition;

- Implementation of a video tracking algorithm;

- Analysis of obtained time series of directions and speed in the two different conditions with sampling time of 30 frames per second.

For the first step we have used the following materials:

- Nauplies six days after hatching in a layer of water thick 4 mm;
- Video camera SAMSUNG HMX-H100P;
- Extra Bright LED 5mm.

In the second step we have developed an algorithm for the analysis of the videos based on these operations:

- Conversion of images from the standard RGB codification of colors, stored as three digital numbers (Red, Green and Blue) to gray scale;

- Binarization of images;

- Noise filtering with morphological operation, i.e. dilation and erosion;

- Calculation of centroids of identified moving artemia;

- XOR to point out the motions;

- Calculation of trajectories.

For the third step of our analysis we have used the Kaplan test and we have defined a parameter derived from the Vicsek model to show the presence of determinism [3]-[5].

In this case we have two sets of three recordings on the motions of 17 Artemias, identified with progressive numbers. The time series are similar in all Artemias analyzed for each of the considered conditions.

Figure 6. shows the trend of the velocity for an Artemia in frames, with sampling time of 30 frames per second.
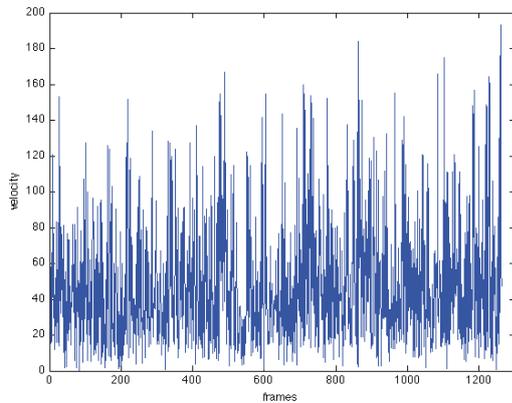
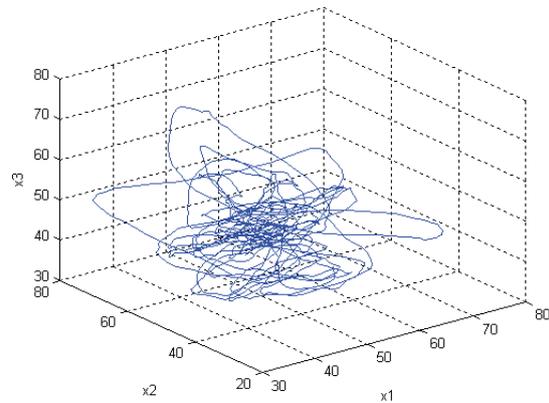Figure 6.   Trend of the speed of an artemia

Figure 8.   Representation in the phase space of the velocity time series in diffusive light conditions

In our study we have used the TISEAN package [8] for the representation of the variation of the velocity for second in the phase space. In Figure 7. and Figure 8. the autocorrelation and the representation in the phase space of the averaged velocity evaluated every 30 samples, where it is possible to see a random behavior, are shown.
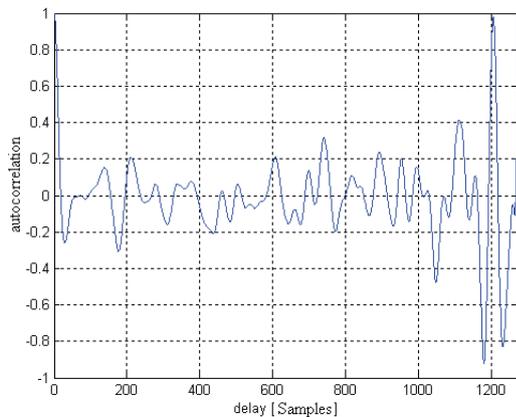
To confirm what it is possible to deduce from the representation in the phase space, the absence of a deterministic component, we have used the Kaplan test. In Figure 9. and in Figure 10. the intersection of the actractor with a 16x16x16 grid and the trend of Ln of the speed in diffusive light condition are shown respectively. From these two figures it is possible to say that velocity of Artemia in diffusive light condition has no deterministic component and a random trend.

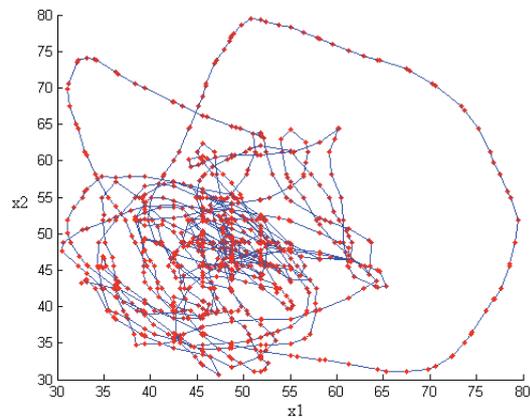Figure 7.   Autocorrelation of the averaged velocity

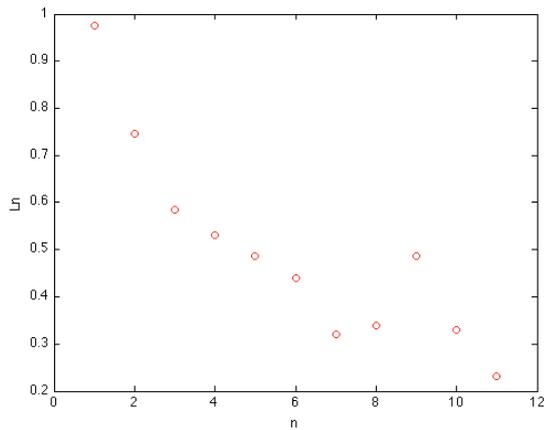Figure 9.   Intersection with a 16x16x16 grid

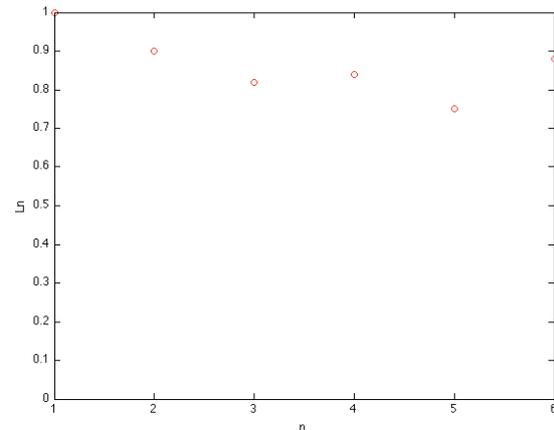Figure 10. Ln of the velocity in diffusive light conditions

Figure 12. Ln of the velocity in LED light conditions

The same kind of analysis done in diffusive light conditions is made in light LED conditions. In this case there are two LEDs in oppositive positions. In a first moment a LED is on and the other one is off, then this LED is turned off and is turned on the other one. In this moment we start to capture the video for our analysis. By a visual analysis all the Artemia go neatly to the side with the turned on LED.

The considered dataset is short so we have not filtered the time series, to avoid modifying the real trend of the signal. In Figure 11. and in Figure 12. the phase space and the trend of Ln for the velocity in LED light conditions are shown.

With respect to the case of diffusive light conditions it is possible to conclude that in this second case there is a deterministic component, in fact, the Ln is high.

We have calculated an order parameter derived from the Vicsek model [5] on a dataset that takes into account the same number of Artemia and the same number of frames in the two cases of diffusive light and LED light.
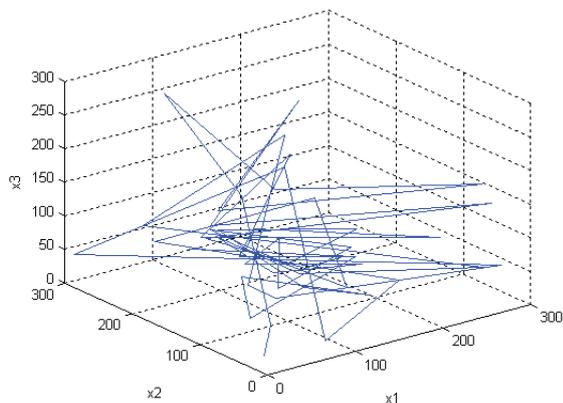

Figure 13. Order parameter (v_a) in diffusive light conditions

Figure 11. Representation in the phase space of the velocity time series in LED light conditions

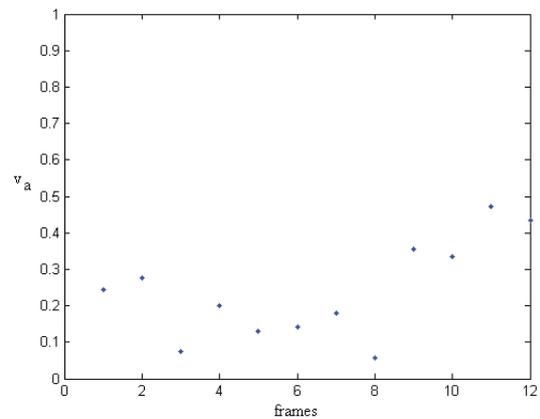The results show that v_a tends to 0, as shown in Figure 13. in the diffusive light conditions while tends to 1 in LED light conditions, as shown in Figure 14. emphasizing that in this case all the Artemias move on a direction.
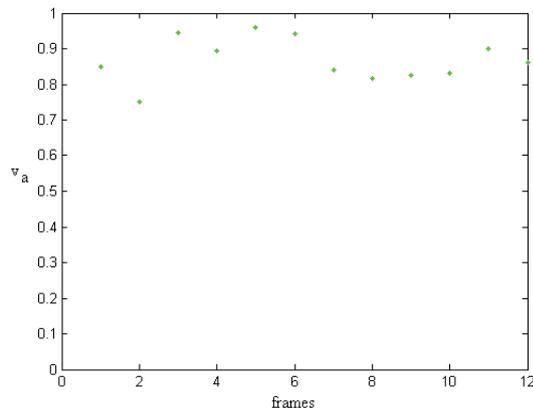
Figure 14. Order parameter (v$_a$) in LED light conditions

In TABLE I. values of mean and standard deviation in both light conditions are shown.

TABLE I.   MEAN AND STANDARD DEVIATION IN THE TWO CASES

	DIFFUSIVE LIGHT CONDITIONS	**LED LIGHT CONDITIONS**
Mean	0.3061	0.8678
Standard deviation	0.1656	0.0622

IV.   CONCLUSIONS

In this work determinism in time series from two different complex systems has been studied. Often a simple analysis of the signal trend in time is not enough to discover a possible deterministic component, while its representation in phase space may reveal this component.

We have collected the number of on-line friends using a Facebook account, filtered the obtained signal using a Butterworth high pass, to remove the periodic component, and then we have analyzed it with the Kaplan test that showed the presence of a deterministic component.

In the case of time series obtained from biological systems, we have analyzed the motion of some brine shrimps in two different light conditions. This was done by using the Kaplan test and with the definition of an order parameter for moving particles. In diffusive light conditions there is not a deterministic component while in LED light condition there is a mechanism of self-organization that leads the Artemias to take parallel directions with a deterministic component in time series.

In this work we have analyzed two real systems: Facebook, social system, and Artemias, biological data that, in a certain sense, are two benchmarks for the study of complexity in real systems. In fact Facebook may be analyzed and monitored and Artemias are one of the reproducible experimental example.

REFERENCES

[1] J. C. Sprott, "Chaos and Time-Series Analysis", Oxford University Press, 2003

[2] J. Jeong, J. C. Gore, and B. S. Peterson, "A method for determinism in short time series, and its application to stationary EEG", IEEE Transactions on biomedical engineering, vol. 49, No. 11, pp. 1374-1379, November 2002

[3] D. T. Kaplan, L. Glass, "Coarse grained embeddings of time series:random walks, Gaussian random processes,and deterministic chaos", Physica D: Nonlinear Phenomena, vol. 64, issue 4, pp. 431-454, April 1993.

[4] D. T. Kaplan, L. Glass, "Direct test for determinism in a time series", Physical review letters, vol. 68, No. 4, pp. 427-430, January 1992.

[5] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles", Physical review letters, vol. 75, issue 6 , pp. 1226-1229, August 1995.

[6] A. Asem, "Historical record on (brine shrimps) Artemia more than one thousand years ago from Urmia Lake", Journal of Biological Research, vol. 9, pp. 113-114, 2008.

[7] T. A. WILLIAMS, "The Nauplius Larva of Crustaceans: Functional Diversity and the Phylotypic Stage", Amer. Zool., vol. 34, pp. 562-569, 1994.

[8] R. Hegger, H. Kantz., "Practical implementation of nonlinear timeseries methods: the tisean package", Chaos, vol. 9, No. 2 , 413-435, 1999