Discriminating Chaotic Time Series with Visibility Graph Eigenvalues

Vincenzo Fioriti
Alberto Tofani
Antonio Di Pietro
Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), CR Casaccia Labs
S. Maria in Galeria
301,00130 Rome, Italy

Time series can be transformed into graphs called horizontal visibility graphs (HVGs) in order to gain useful insights. Here, the maximum eigenvalue of the adjacency matrix associated to the HVG derived from several time series is calculated. The maximum eigenvalue methodology is able to discriminate between chaos and randomness and is suitable for short time series, hence for experimental results. An application to the United States gross domestic product data is given.

1. Introduction

Discriminating between chaos and randomness is an inescapable, difficult problem of time series analysis. Different fields of science and engineering such as biology, medicine, economics, meteorology, and mechanics are deeply interested in developing suitable tools. Derived from computational geometry, the visibility algorithm [1, 2] is a tool intended to study the time series from a relatively new point of view. It has been developed by L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño and presented in a number of seminal papers [1, 3]. Basically, it transforms a time series into a graph, thus allowing the usage of graph theory techniques. Recently, it has been applied to finance, fluid dynamics, and analyzing the atmosphere [4–6]. The mapping from the time series to the graph is simple: two arbitrary samples $[t_a, y_a]$ and $[t_b, y_b]$ in the time series have visibility and become two nodes in the associated graph, if any other data $[t_c, y_c]$ such that $t_a < t_c < t_b$, fulfills:

$$y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a}. \tag{1}$$

By modifying the visibility rule of equation (1), a so-called horizontal visibility graph (HVG) is obtained as shown in Figure 1. The new rule assigns each sample to a node in the HVG below. Two nodes

https://doi.org/10.25088/ComplexSystems.21.3.193

i and j in the graph are connected if a horizontal line can be drawn connecting x_i and x_j as in Figure 1. Hence, i and j are two adjacent nodes if:

$$x_i, x_j > x_n \text{ for all } n, \text{ such that } i < n < j.$$

The HVG is a subgraph of the original visibility graph. Moreover, it is connected, invariant under affine transformation, and is reversible if links are weighted.

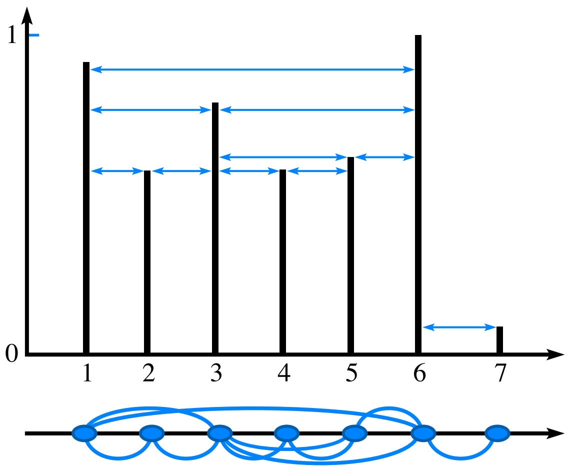

Figure 1. The time series and the associated graph. Lines connecting samples of the time series indicate the visibility. In the visibility graph, each node indicates a sample that is connected to another node if visibility between the two exists. The time series have been normalized.

2. Degree Distribution

An interesting result has been obtained in [1] regarding random time series. For every probability distribution, the degree distribution $P(d)$ of the associated HVG has the same form:

$$P(d) = \frac{1}{3}\left(\frac{2}{3}\right)^{d-2} \tag{2}$$

where d is the degree of the node, that is, the total number of links originating from a node. The result of equation (2) is exact when the number of nodes (samples in the HVG framework) $N \to \infty$. To discriminate chaotic (even high-dimensional) time series from true random time series, it suffices [1] to verify if equation (2) is respected by means of a simple visual inspection of the $P(d)$ versus d plot. Actually, noise interferes with the visual analysis when $d > 20$ and standard

https://doi.org/10.25088/ComplexSystems.21.3.193

statistic tests may confirm the analysis. It should be stressed that the procedure does not quantify chaos; nevertheless, being based on an exact result, it is very reliable. In the following it will be shown that the maximum eigenvalue L_{max} of the adjacency matrix associated to the HVG describes equally random and chaotic time series. The largest eigenvalue of the network adjacency matrix is a key parameter relevant for a variety of dynamical network processes such as synchronization, percolation, and disease spreading. In this paper, the usual point of view is changed by means of the HVG framework.

3. Horizontal Visibility Graph Spectral Analysis

It was proved in [7] that, for all values of the edge probability $p(n)$, the largest eigenvalue of a random graph $G(n, p)$ satisfies almost surely:

$$L_{max} = (1 + o(1)) \cdot \max\left[\sqrt{d_{max}}, np\right] \tag{3}$$

where d_{max} is the maximum degree of G. The $O(1)$ term tends to zero as $\max\left[\sqrt{d_{max}}, np\right]$ tends to infinity. Moreover, in [8] it is shown that the largest eigenvalue of the adjacency matrix of a random graph with a given expected degree sequence is determined by $\sqrt{d_{max}}$ and $\langle d^2 \rangle$. The maximum eigenvalue of the adjacency matrix is almost surely:

$$L_{max} = (1 + o(1)) \cdot \max\left[\langle d^2 \rangle, \sqrt{d_{max}}\right] \tag{4}$$

provided some minor conditions are satisfied.

A further simplification can be found in [9]. If the $O(1)$ term tends to zero, the maximum eigenvalue for a graph (and for the HVG) may be estimated as:

$$L_{max} \approx \sqrt{d_{max}} \qquad \text{when } N \to \infty. \tag{5}$$

This is also valid for nonrandom and directed graphs, and only requires knowledge of the adjacency matrix. Thus, equation (5) is used much in the same way equation (2) is used in [1], finding an approximation for the eigenvalues extracted from the time series by the horizontal visibility algorithm. However, it must be noted carefully that the approximation also depends on the degree distribution, and therefore cannot be considered as a general rule. To study the pattern for finite values of N in order to distinguish random and chaotic time series and verify equation (5), a plot of L_{max} versus N is shown in Figure 2.

In Figure 2, the upper curves are from a Gaussian and a uniform distribution, and the lower three are from the Lorentz, Rossler, and Glass time series. As $N \to \infty$, they are clearly separated from the curves below (periodic, quasi-periodic, and chaotic series) while L_{max} converges quickly.

https://doi.org/10.25088/ComplexSystems.21.3.193

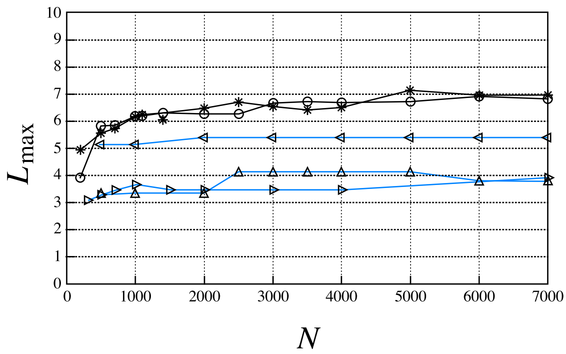

Figure 2. Plot of the maximum eigenvalue versus node number N. Gaussian: star, circle (uniform); Lorentz: left triangle; Rossler: up triangle; Glass: right triangle. Random and chaotic distributions are clearly separated and distinguishable in the lower part of the figure.

Figure 3 shows the actual trend of L_{max} versus N and the trend as calculated according to equation (5). The approximation is very good, although for the uniform time series (circle) and its approximation (square) the finite size effect is quite strong for low N, while in the cases of Rossler and Glass the error is log(2). It should be observed that we used only 7000 data points at most (the authors of [1] used 10^6). This is an important point because it is well known how difficult it is to collect experimental data for a long time, for example, in climate analysis and medicine. Moreover, the convergence to constant values seems to allow a "measure" of the complexity of the time series that deserves further investigation.

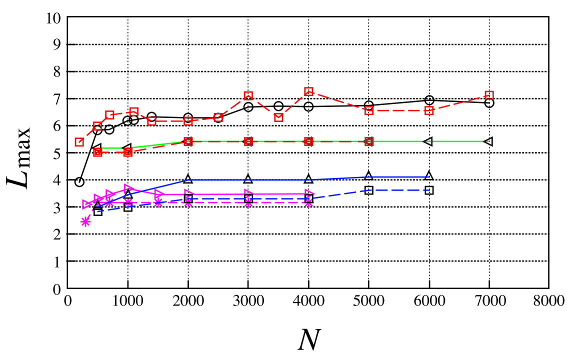

Figure 3. Actual and calculated value for L_{max}. Uniform: circle, dotted square (calculated); Lorentz: left triangle, dotted square (calculated); Rossler: up triangle, dotted square (calculated); Glass: right triangle, dotted square (calculated). L_{max} converges to 7 for the random distribution, 5.3 for Lorentz, 3.8 – log(2) for Rossler, and 3.5 – log(2) for Glass.

https://doi.org/10.25088/ComplexSystems.21.3.193

4. The Gross Domestic Product Time Series Transformed into an Undirected Graph

To appreciate some capabilities of the HVG method, a simple application to the United States gross domestic product (GDP) percentage quarterly increments during the years 1945–2011 is given in Figure 4 (further details can be found in [10]). It is remarkable how the topological structure of the graph resembles the GDP time series variability, as the GDP is divided into a first part with larger oscillations and a second part that is more stable and predictable. The turning point is clearly visible in the graph as the bridge between the two subgraphs corresponding to the yellow dotted line in Figure 5. The largest eigenvalue of the GDP graph is $L_{\max} = 5.3296$ while the data length of the time series is 258. From Figure 6, it can be seen that the point identified by the eigenvalue and the data length is compatible with a Gaussian time series. As a matter of fact, there is an ambiguity between the Gaussian and the Lorentz distribution, but it should be remembered that the time series is only 258 data points long. Actually, the GDP time series has a distribution very close to the Gaussian (see Figure 4).

Many parameters from graph theory are available to establish dynamic and static characteristics: the eigenspectra, degree, closeness, betweenness, other recent centrality measures, and so on. All these tools can be used to infer behaviors of the system not clearly seen in the time series. In the present case the old problem afflicting graph analysis, that is, the extremely large number of nodes/links, is avoided, because in the given time domain the problem is the lack of experimental data samples. Therefore, the nodes/links number of the associated graph will always be small, meaning that graph theory tools are immediately available.

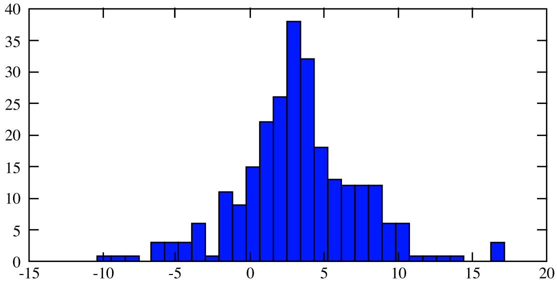

Figure 4. Distribution of the GDP quarterly percentage increments time series. The distribution is close to a Gaussian, in agreement with the point $L_{\max} = 5.3296$, $N = 258$ of Figure 2 that belongs to the Gaussian time series.

https://doi.org/10.25088/ComplexSystems.21.3.193

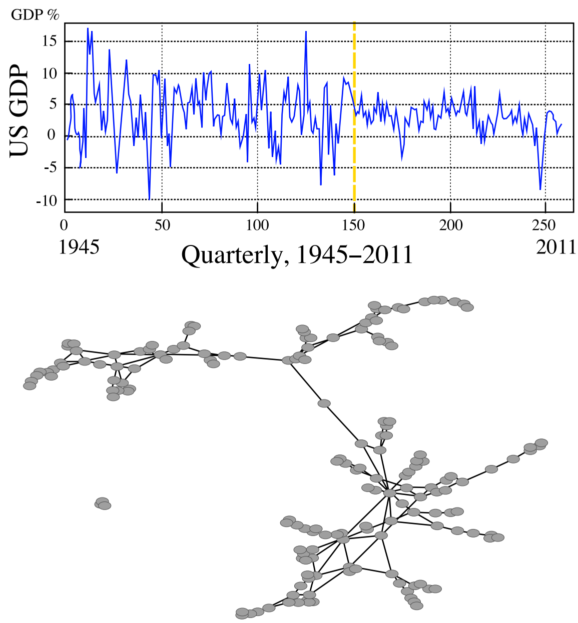

Figure 5. GDP percentage increments during the years 1945–2011. The time series of 258 samples is transformed into a graph, clearly split into two subgraphs resembling the first part and the second part of the time series divided by the yellow dotted line. The lower subgraph corresponds to the first part of the time series.

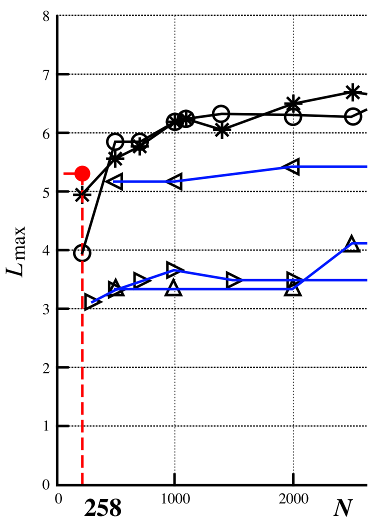

Figure 6. Enlargement of the point of coordinates (258, 5.3296) in Figure 2. The position is compatible with a Gaussian or a Lorentz time series.

https://doi.org/10.25088/ComplexSystems.21.3.193

5. Conclusions

A relatively new tool, the horizontal visibility graph (HVG), has been presented by Luque et al. to discriminate random from chaotic time series, mapping the samples into nodes of a graph and comparing the degree distributions associated to the series. It is possible to obtain the same result by the largest eigenvalue of the adjacency matrix performing a spectral analysis of HVGs. The discrimination is clear also for short time series. The largest eigenvalue methodology has been applied to the United States gross domestic product (GDP) data.

Acknowledgments

We thank M. Ruscitti, S. Ruzzante, A. Arbore, and N. Sigismondi for useful discussions.

References

[1] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal Visibility Graphs: Exact Results for Random Time Series," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, **80**(4), 2009 p. 046103. doi:10.1103/PhysRevE.80.046103.

[2] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, 2nd ed., Berlin, New York: Springer-Verlag, 2000.

[3] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From Time Series to Complex Networks: The Visibility Graph," *Proceedings of the National Academy of Sciences of the United States of America*, **105**(13), 2008 pp. 4972–4975. doi:10.1073/pnas.0709247105.

[4] C. Liu, W-X. Zhou, and W-K. Yuan, "Statistical Properties of Visibility Graph of Energy Dissipation Rates in Three-Dimensional Fully Developed Turbulence." arxiv.org/abs/0905.1831, 2009.

[5] J. B. Elsner, T. H. Jagger, and E. A. Fogarty, "Visibility Network of United States Hurricanes," *Geophysical Research Letters*, **36**, 2009 L16702. doi:10.1029/2009GL039129.

[6] Y. Yang, J. Wang, H. Yang, and J. Mang, "Visibility Graph Approach to Exchange Rate Series," *Physica A: Statistical Mechanics and Its Applications*, **388**(20), 2009 pp. 4431–4437. doi:10.1016/j.physa.2009.07.016.

[7] M. Krivelevich and B. Sudakov, "The Largest Eigenvalue of Sparse Random Graphs," *Combinatorics, Probability and Computing*, **12**(1), 2003 pp. 61–72. doi:10.1017/S0963548302005424.

[8] F. Chung, L. Lu, and V. Vu, "Spectra of Random Graphs with Given Expected Degrees," *Proceedings of the National Academy of Sciences of the United States of America*, **100**(11), 2003 pp. 6313–6318. doi:10.1073/pnas.0937490100.

[9] J. G. Restrepo, E. Ott, and B. R. Hunt, "Approximating the Largest Eigenvalue of Network Adjacency Matrices," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, **76**(5), 2007 p. 056119. doi:10.1103/PhysRevE.76.056119.

[10] V. Fioriti and M. Chinnici, "The Synchronization of Kuramoto Oscillator Networks: Forecasting Financial Index Critical Points," talk given at Interop-VLAB 5th Workshop, CNR-IASI, Rome, September 28, 2012 (available from vincenzo.fioriti@enea.it).

https://doi.org/10.25088/ComplexSystems.21.3.193