scientific reports

OPEN

An electrostatics method for converting a time-series into a weighted complex network

Dimitrios Tsiotas[1,2,3✉], Lykourgos Magafas[3] & Panos Argyrakis[3,4]

This paper proposes a new method for converting a time-series into a weighted graph (complex network), which builds on electrostatics in physics. The proposed method conceptualizes a time-series as a series of stationary, electrically charged particles, on which Coulomb-like forces can be computed. This allows generating electrostatic-like graphs associated with time-series that, additionally to the existing transformations, can be also weighted and sometimes disconnected. Within this context, this paper examines the structural similarity between five different types of time-series and their associated graphs that are generated by the proposed algorithm and the visibility graph, which is currently the most popular algorithm in the literature. The analysis compares the source (original) time-series with the node-series generated by network measures (that are arranged into the node-ordering of the source time-series), in terms of a linear trend, chaotic behaviour, stationarity, periodicity, and cyclical structure. It is shown that the proposed electrostatic graph algorithm generates graphs with node-measures that are more representative of the structure of the source time-series than the visibility graph. This makes the proposed algorithm more natural rather than algebraic, in comparison with existing physics-defined methods. The overall approach also suggests a methodological framework for evaluating the structural relevance between the source time-series and their associated graphs produced by any possible transformation.

The multidisciplinary nature of networks[1–3] has introduced new directions in the time-series research that led to the emergence of the complex network analysis of time-series. This newly established research field showed a remarkable development, at a multidisciplinary level[4], when scholars conceptualized[5–7] that transforming a time-series into a graph can produce insights that are not visible by current time-series approaches. In general, studying the topology of a graph instead of the structure of a time-series promotes time-series analysis because it enlarges the embedding of the available information, from a first-order tensor (i.e. the time-series vector) into a second-order tensor (i.e. the graph connectivity matrix)[8]. Within this context, Zhang and Small[7] were the first who constructed graphs from pseudo-periodic time-series, and Yang and Yang[6] applied thresholds to the correlation matrix to convert it into a connectivity matrix. Xu et al.[9] proposed a transformation for creating graphs from time-series based on different dynamic systems. Lacasa et al.[5] built on the intuition of considering a time-series as a landscape and introduced a connectivity criterion based on visibility from optics physics. Gao and Zin[10] proposed methods (i.e. flow pattern complex network, dynamic complex network, and fluid–structure complex network) to construct complex networks from experimental flow signals, and Donner et al.[11] introduced a recurrence method converting graphs from time-series based on the phase-space of a dynamical system. Amongst the existing methods, the natural visibility graph (NVG) or, in synonym, the visibility graph algorithm (VGA) of Lacasa et al.[5] seems to prevail in the literature either in terms of citations, or in the number of applications[12–14], or the number of derivative methods, such as the horizontal visibility graph of Luque et al.[15], and the visibility expansion algorithm of Tsiotas and Charakopoulos[8,16]. The popularity of VGA can be either due to its intuitive conceptualization from optics physics, which makes comprehension and interpretation of results easier, or to its topological consistency to convert periodic time-series to regular graphs, random time-series to random graphs, and fractal time-series to scale-free graphs. However, this method builds on a binary connectivity criterion, which leads to the development of binary connections and thus to unweighted graphs[5]. Therefore the VGA is by definition restricted in generating visibility graphs that are disassociated from the numerical scale of the source (original) time-series.

[1]Department of Regional and Economic Development, Agricultural University of Athens, Amfissa, Greece. [2]Adjunct Academic Staff, School of Social Sciences, Hellenic Open University, 10677 Athens, Greece. [3]Laboratory of Complex Systems, Department of Physics, International Hellenic University, Kavala, Greece. [4]Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece. ✉email: tsiotas@aua.gr
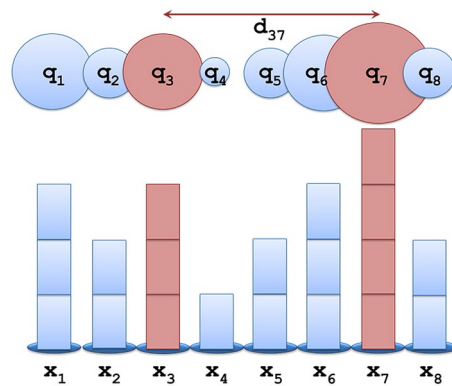
Figure 1. Example of the electrostatic graph (ESG) algorithm conceptualization. The volume of electrical charge (q_i, $i = 1,..,8$) in each node is shown proportionally to the node size (x_i, $i = 1,..,8$) of the time-series.

Aiming at serving the demand of promoting a weighted conceptualization in the complex network analysis of time-series, this paper introduces a method for converting a time-series into a weighted graph by using an electrostatics transformation algorithm based on Coulomb's law. The proposed method is driven by a dual motivation: the first builds on the example of the VGA[5], which implies that physics-defined transformations can be more intuitive and easily comprehensive than the algebraic (or computational) ones. The second one is based on the universality of Coulomb's inverse square law[17], which grounded the development of essential research in electromagnetism but also inspired multidisciplinary research in economics[18], urban and spatial planning and transport engineering[19,20], biology[21], geophysics[22], computational[23] and communication sciences[24], etc. Within this multidisciplinary context, the proposed method conceptualizes a time-series as a sequence of stationary and electrically charged particles (nodes) and generates an electrostatic graph based on pair-wise calculations of Coulomb's law across the time-series nodes. The Coulomb-like forces are assigned as weights in the connectivity matrix of the electrostatic graph and can be seen as a measure of relevance between two nodes, in terms of the sign, scale, and spatial proximity. This approach allows quantifying the interaction between the time-series nodes and thus to conceptualize the dynamics of a time-series through the effect of electrostatic forces applied between the nodes.

The remainder of the paper is organized as follows: Sect. 2 (methods) describes the proposed ESG algorithm and its modeling context, it introduces the node-series of network measures concept in the ESG, and it briefly describes the methods used for testing the performance of the proposed algorithm. Section 3 (Results) shows the results of the multilevel analysis testing the performance of the proposed algorithm in comparison with a well-established method of converting a time-series to a graph. Finally, in Sect. 4, conclusions are given.

Methods

The proposed ESG algorithm. Let us consider a time-series $X = \{x_1, x_2, \ldots, x_n\}$ with $n \in \mathbb{N}$ number of nodes $i \in X$, where each one has a real numeric value $X(i) = x_i \in \mathbb{R}$. If we assume that every node i in the time-series can be seen as a static particle of electrical charge $q(i) \equiv q_i = x_i$, we can define an (either attractive or repulsive) electrostatic force F_{ij} applied between any pair of nodes i,j (Fig. 1), according to the inverse-square Coulomb's law expressed by the relation[17]:

$$F_{ij} = k_e \cdot \frac{q_i \cdot q_j}{\left(d_{ij}\right)^2} = k_e \cdot \frac{x_i \cdot x_j}{\left(d_{ij}\right)^2}, \tag{1}$$

where q_i and q_j are the electrostatic charges of nodes i and j, d_{ij} is the intermediate discrete distance between nodes i,j that expresses steps of separation and is defined by the difference $(i-j)$, and k_e is the Coulomb's constant.

This assumption allows considering a time-series X as a series of stationary and electrically charged particles (i.e. time-series nodes), on which we can compute a square matrix with the Coulomb-like forces $F(X) = \{F_{ij} \mid i,j = 1, \ldots, n\}$, according to the relation:

$$(1) \overset{q_i = X(i) = x_i}{\underset{k_e \equiv 1}{\Leftrightarrow}} F(X) = \{F(X(i), X(j)) \equiv F_{ij} = \frac{x_i \cdot x_j}{(i-j)^2} \mid i,j = 1, \ldots, n\}, \tag{2}$$

where $d_{ij} = (i-j)$ and k_e is the Coulomb's constant[17], which can be considered as a scale factor and in this paper is set to $k_e = 1$.

The square structure of the $F(X)$ matrix (with the Coulomb-like forces) can be seen as an electrostatic graph-model ESG, where each element $F_{ij} \in \mathbb{R}$ expresses the (attractive or repulsive) electrostatic force applied between any pair of nodes i,j. When it is important to note that the ESG is associated with the time-series X, we can symbolize the electrostatic graph as ESG(X). In terms of graph theory[25], $F(X)$ is the weighted connectivity matrix of an undirected graph $G_{ESG}(V,E)$, where V is the node-set and E is the edge-set. The weights (w_{ij}) in the ESG's weighted connectivity matrix are equal to the Coulomb-like forces ($w_{ij} = F_{ij}$) and can be seen as a measure of similarity between two nodes, in terms of the sign, scale, and spatial proximity. In particular, positive weights

$(w_{ij} > 0)$ indicate that nodes i,j have homogeneous arithmetic signs, where negative cases ($w_{ij} < 0$) imply that they have heterogeneous signs. Also, high w_{ij} scores may imply either that nodes i,j are close in the time-series line, in terms of spatial proximity, or that they have relatively high arithmetic values or both. Within the context of the electrostatic conceptualization, the attraction expressed by a negative Coulomb-like force ($w_{ij} = -F_{ij}$) can be seen as a tendency of the nodes to balance their heterogeneity and converge toward the horizontal axis, whereas the repulsion expressed by a positive force ($w_{ij} = +F_{ij}$) can be seen as a tendency of the nodes to escape from their homonymous electrostatic balance and thus to evolve (either increasingly or decreasingly) through time.

By definition, Coulomb's law determines a field of infinite distance, where the electrostatic forces are noticed at infinity, although they are negligible. This property makes the ESG by default a fully connected (complete) graph K_n, namely a graph where all nodes are linked to any other. Provided that a complete graph K_n has a trivial topology, in terms of complexity (since the average degree is always $\langle k \rangle = n-1$ and most of other metrics, such as average path length, network diameter, graph density, and clustering coefficient are equal to one), we filter the set E of the ESG connections, aiming to generate more complex topologies of electrostatic graphs. In particular, we consider a threshold F_c, defined within the interval $F_c \in \left(\min\{F_{ij}\}, \max\{F_{ij}\} \right)$, so that the weighted connectivity matrix W_{ESG} include those values that are equal or above F_c, as it is expressed by the relation:

$$W_{ESG} = \{F_{ij} \neq 0 \in F(X) : F_{ij} \geq F_c\} \subseteq F(X), \tag{3}$$

where $F(X)$ is the Coulomb-like matrix defined in relation (1). This filtering allows considering numerous electrostatic graphs ESG(X), which are expressed as a function $W_{ESG} = f(F_c)$ of the threshold-variable F_c. To introduce a reference value to the threshold-variable F_c, we define a typical value f_z by the relation:

$$F_c(f_z) = f_z = \frac{1}{n-1} \cdot \sum_n x_n = \frac{n}{n-1} \cdot \langle x \rangle = n \cdot \text{sgn}(\langle x \rangle) \cdot \frac{\sqrt{|\langle x \rangle|} \cdot \sqrt{|\langle x \rangle|}}{\left(\sqrt{n-1}\right)^2}, \tag{4}$$

where n is the number of time-series nodes, $\langle \cdot \rangle$ is the average operator, and $\text{sgn}(\cdot)$ is the sign (or signum) function[26]. In numeric terms, the f_z filtering describes that non-zero elements of the weighted ESG's connectivity matrix are those with values higher than the adjusted mean-value $\frac{n}{n-1} \cdot \langle x \rangle$ of the time-series X. In physical (electrostatic) terms, f_z describes an electrostatic force that is n-times greater than this applied to a pair of particles with electrical charges $q_i, q_j = \sqrt{|\langle x \rangle|}$(i.e. equal to the square-root of the absolute mean-value of the time-series X), which are $d_{ij} = \sqrt{n-1}$ (i.e. equal to the square-root of the time-series length) steps of separation distant.

Within this context, the proposed ESG algorithm is implemented in four steps, as it is shown in Fig. 2. First, we compute the matrix $F(X)$ of Coulomb-like forces, according to the relations (1) and (2). Secondly, we apply to $F(X)$ the connectivity filter and compute the weighted connectivity matrix W_{ESG}, according to the relations (3) and (4). Thirdly, we manage the disconnected data (i.e. mainly the diagonal element yielding infinite computations due to zero distances included in the denominator) of $F(X)$, by substituting "inf" (infinite) values by zeros. Fourthly, we create the graph-layout of the ESG(X) based on the weighted connectivity matrix W_{ESG}.

According to the first four steps of the algorithm, we can generate the electrostatic graph ESG(X), which is associated with a time-series X and is an undirected and weighted graph with a non-trivial topology. In this graph model, we can further compute several network measures and metrics and thus reveal the topological properties of the ESG. Therefore, at the fifth and final step of the algorithm, we compute node-series of network measures of the ESG(X), and afterward, we compare their structural relevance with this of the source time-series X. The procedure is described in more detail in the following paragraphs.

Node-series of network measures.

The electrostatic graph ESG(X) is a graph-model $G_{ESG}(V,E)$ where each network node $v_i \in V$ is the same with a time-series node $i \in X$, namely $v_i \equiv i \in V, X$. Therefore, for every node-measure Y (e.g. node degree, local clustering coefficient, closeness, betweenness, and eigenvector centrality, etc.) of the ESG, we can arrange the scores $Y(v_i) = y_i$ into the time-series $X = \{x_1, x_2, \ldots, x_n\}$ ordering, and thus to configure node-series $X(Y) = \{y_1, y_2, \ldots, y_n\}$ of the ESG network measures that are associated with the source time-series. This allows comparing the source time-series X with these of the ESG node-series $X(Ys)$ and detecting possible structural similarities that can be seen as a measure of relevance between the time-series and the ESG. The available network (node) measures that participate in the construction of the node-series are shown in Table 1.

In terms of notation, for a (source) time-series $X = \{x_1, x_2, \ldots, x_n\}$, where $n \in \mathbb{N}$ and $x_i \in \mathbb{R}$, we can write its associated node-series for the network measure Y as $X(Y) = \{Y(x_1), Y(x_2), \ldots, Y(x_n)\} = \{y_1, y_2, \ldots, y_n\}$. We can read that $X(Y)$ is "the node-series of the network measure Y, which is computed for the ESG that is associated with the time-series X" or, in brief, that $X(Y)$ is "the node-series of (the measure) Y for the ESG". Within this context, we can compute the node-series for the measures of degree $X(Y = k) = \{k_1, k_2, \ldots, k_n\}$, strength $X(s) = \{s_1, s_2, \ldots, s_n\}$, clustering coefficient $X(C) = \{C_1, C_2, \ldots, C_n\}$, betweenness centrality $X(CB) = \{CB_1, CB_2, \ldots, CB_n\}$, closeness centrality $X(CC) = \{CC_1, CC_2, \ldots, CC_n\}$, and eigenvector centrality $X(CE) = \{CE_1, CE_2, \ldots, CE_n\}$, according to the mathematical formulas shown in Table 1. Provided that we can generate a node-series for any graph $G(X)$ that is associated with a time-series X, we can include a subscript index in the notation $X_G(Y)$ when necessary (e.g. $X_{ESG}(k)$) to denote the type of graph that the time-series is associated with.

The effect of the connectivity threshold on the ESG topology.

The connectivity threshold F_c that is applied to the Coulomb-like matrix, according to relation (3), is determinative for the configuration of the ESG topology. To illustrate this, let us consider the series $X_{1:100} = \{1, 2, \ldots, 100\}$ of the first hundred natural numbers. By applying to this series sequentially the connectivity thresholds $F_c = 0$, $F_c = 1$, $F_c = 5$, $F_c = 10$, $F_c = 25$, $F_c = 50$, $F_c = f_z$, $F_c = 75$, and $F_c = 100$, we get various ESGs, as it is shown in Fig. 3.
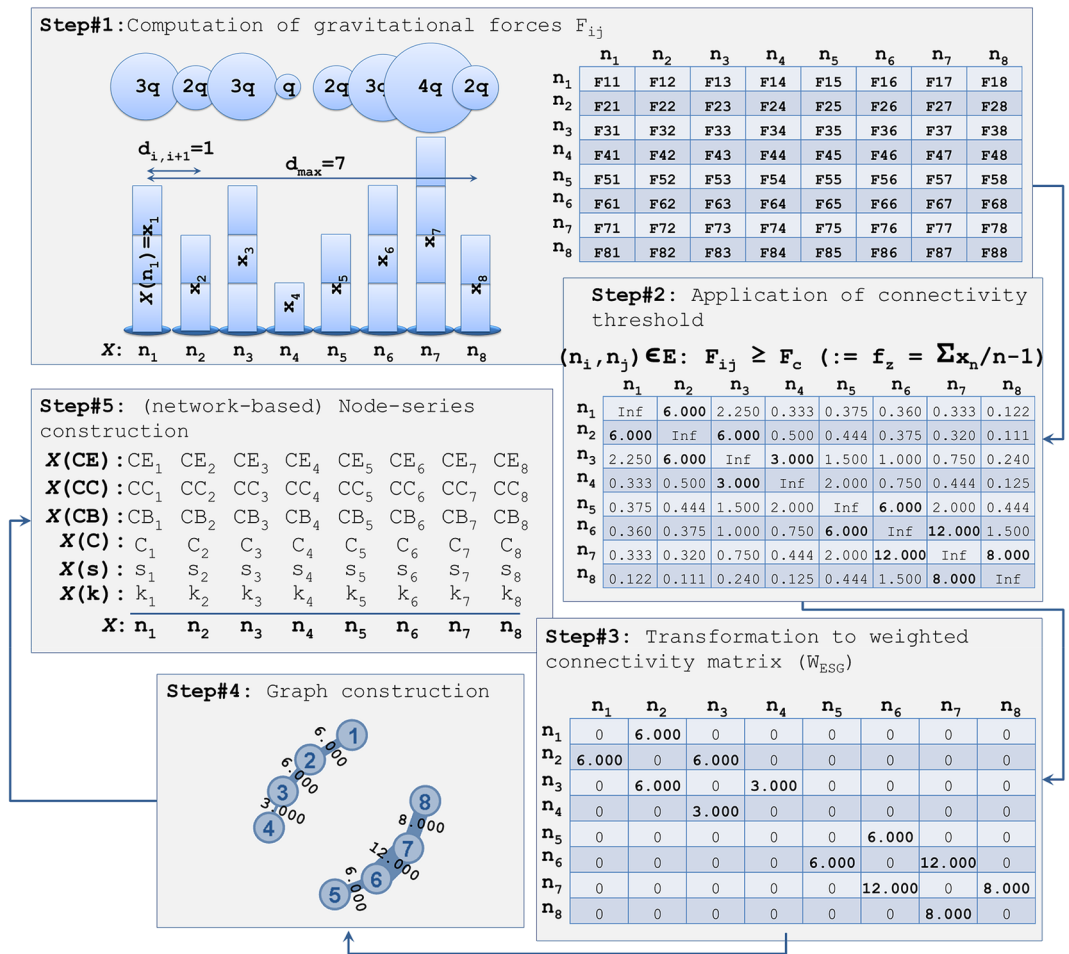
Figure 2. The methodological framework of the study. Steps #1–#4 describe the ESG algorithm generating an electrostatic graph from a time-series $X = \{x_i \mid i = 1, \ldots, 8\}$. Step#5 describes the process of generating secondary time-series from the network measures of ESG(X).

Measure	Description	Mathematical Expression
Node Degree (k)	The number of edges being adjacent to a node i	$k_i = k(i) = \sum_{j \in V} \delta_{ij}$, where $\delta_{ij} = \begin{cases} 1, & \text{if } e_{ij} \in E \\ 0, & \text{otherwise} \end{cases}$
Node strength (s)	The sum of edge weights being adjacent to a given node i	$s_i = s(i) = \sum_{j \in V(G)} \delta_{ij} \cdot w_{ij}$, where $w_{ij} = w(e_{ij})$
Local Clustering Coefficient (C)	The number of a node's connected neighbors E(i), divided by the number of the total triplets $k_i(k_i-1)$ shaped by the node i	$C(i) = \frac{E(i)}{k_i \cdot (k_i - 1)}$
Closeness Centrality (CC)	Total binary distance d(i,j) computed on the shortest paths originating from a given node i and having destination all the other nodes j in the network. This measure expresses the node's reachability in terms of steps of separation	$CC(i) = \frac{1}{n-1} \cdot \sum_{j=1, i \neq j}^{n} d_{ij} = \overline{d}_i$
Betweenness Centrality (CB)	Fraction of all shortest paths $\sigma(i)$ including a given node i, to the number σ of all the shortest paths in the network	$CB(i) = \sigma(i) / \sigma$
Eigenvector Centrality (CB)	Spectral measure expressing the influence of node i in the network. In the formula N(i) expresses the neighborhood of node i, a_{ij} an element of the adjacency, x_j the j-th component of the adjacency's eigenvector with eigenvalue equal to λ	$CE(i) = \frac{1}{\lambda} \cdot \sum_{j \in N(i)} a_{ij} \cdot x_j$

Table 1. The node measures that are considered in the analysis [1,27-29]

As it can be observed, the ESGs shown in Fig. 3 appear quite different in terms of graph density and node arrangement in the adjacency matrix. In particular, as the F_c becomes greater, the connectivity strip toward the main diagonal in the adjacency matrix becomes narrower, expressing each time a separate connectivity pattern. To examine whether and how the network topology is affected by changes in F_c, we compute a set of network
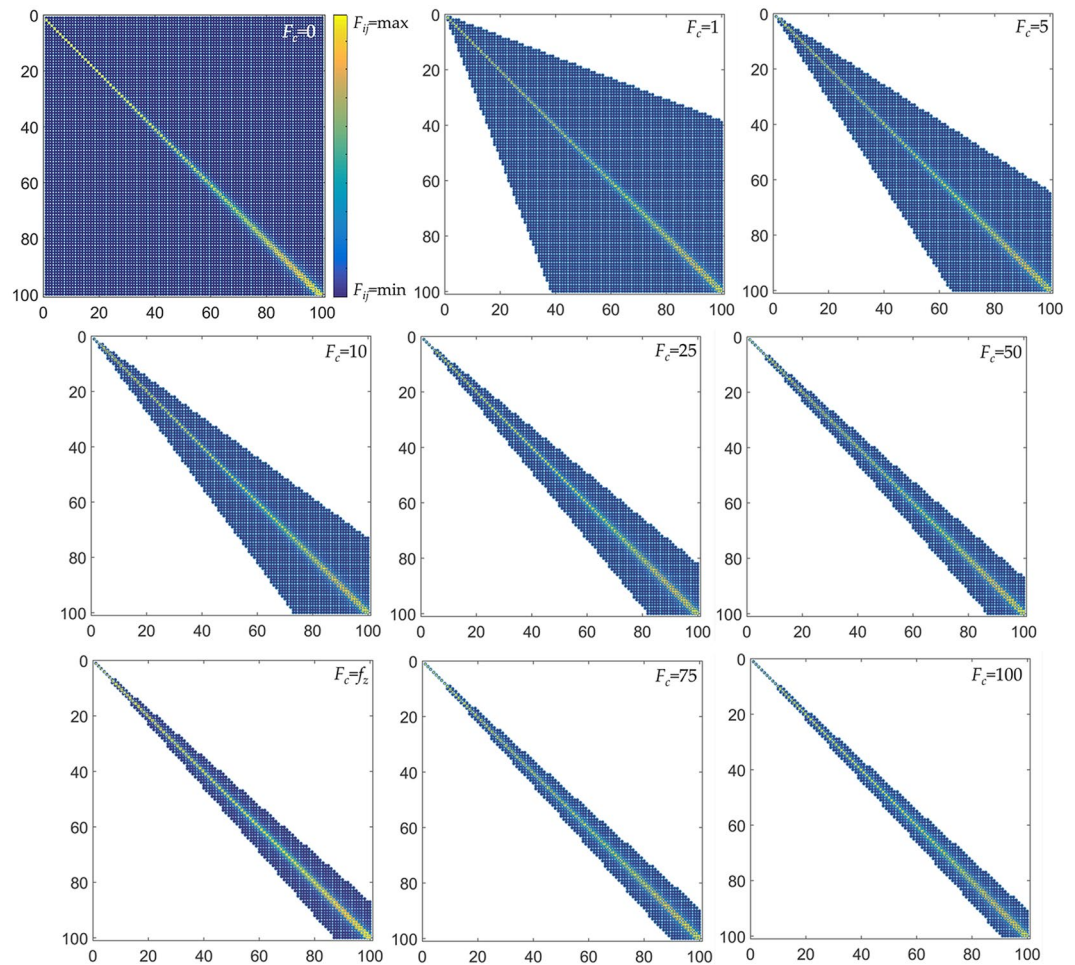
Figure 3. Sparsity (spy) plots of the ESGs that are associated with the series $X_{1:100} = \{1, 2, \ldots, 100\}$ and are computed for the connectivity thresholds $F_c = 0$, $F_c = 1$, $F_c = 5$, $F_c = 10$, $F_c = 25$, $F_c = 50$, $F_c = f_z$, $F_c = 75$, and $F_c = 100$.

measures and metrics (average degree $\langle k \rangle$, clustering coefficient C, graph density ρ, modularity Q, average path length $\langle l \rangle$, network diameter $d(G)$, and the number of components) for a series of ESGs that are generated by applying connectivity thresholds ranging within the interval $F_c \in [0, n^2 = \max\{ X_{1:100}\}^2 = 10^4]$. This approach assumes that the network topology is collectively approximated by the set of available network measures, where each measure represents a certain topological aspect. The results of the analysis are shown in Fig. 4, where each network measure is expressed as a function of the connectivity threshold F_c.

Also, is evident that all network measures considerably fluctuate as the connectivity threshold F_c changes. The cases of average degree $\langle k \rangle$(Fig. 4a), clustering coefficient C (Fig. 4b), and graph density ρ (Fig. 4c) follow a declining pattern to the changes of F_c, the cases of average path length $\langle l \rangle$(Fig. 4e) and graph density $d(G)$(Fig. 4f) follow a bell-shaped pattern of negative skew (asymmetry), whereas the number of components (Fig. 4g) follows an increasing pattern. For the case of modularity Q (Fig. 4d), the performance of this measure appears considerably invariant along the biggest part of the F_c's interval. As far as the typical value f_z is concerned, we can observe that this value cannot be related to border (i.e. min or max) distribution values, but it can be quite indicative of the average performance of the topological aspects of the ESGs. This indication can support the goodness of the choice of defining the typical f_z value within a physical (Coulomb-like) context, as it is shown in relation (4).

Overall, this analysis shows that the choice of the connectivity threshold F_c can be determinative to the topological features and generally the topology of the resulting ESG. This observation is evident even by the examination of a simple linear series of natural ascending numbers, which can only be considered as an indicative approach for the ESG construction. However, even this simple consideration sufficed to highlight the dependence between the connectivity threshold and the ESG's network topology and thus to introduce a methodological path for optimally defining the F_c value. The examination of the optimum or most representative threshold is a matter of specialized optimization analysis that introduces avenues of further research and falls outside the scope of this paper. However, the physically defined approaches, as this of the Coulomb-like definition of the F_c shown in relation (4), or others utilizing methods from other disciplines can become insightful toward this optimization direction and are suggested for further research promoting multidisciplinary conceptualization. For instance, further research on this topic can apply to different types of time-series and more thorough optimization analysis in the choice of F_c. For the scope of this paper, the choice of the typical value f_z for the connectivity
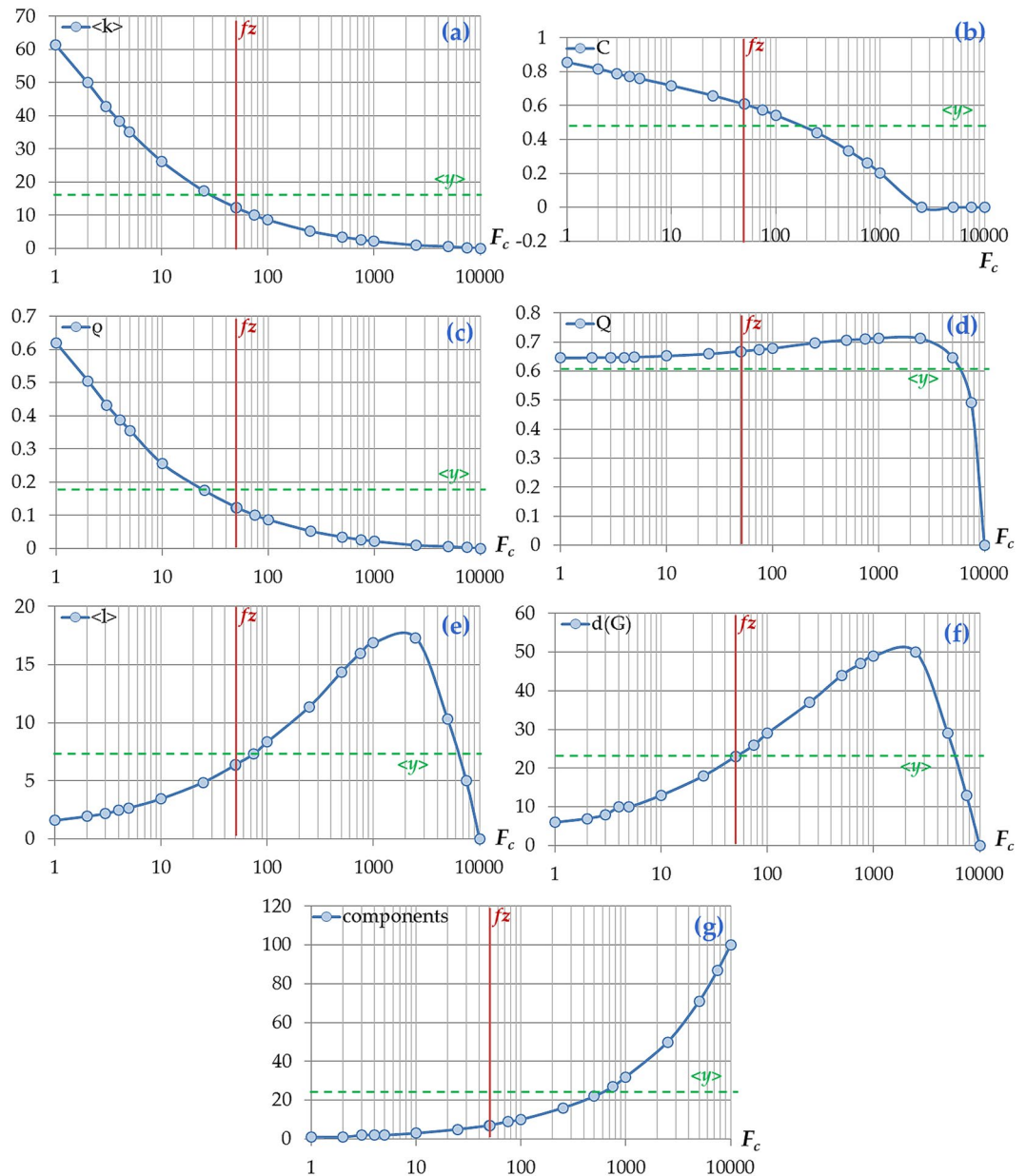
Figure 4. . Line diagrams showing how the network measures of (**a**) average node degree $\langle k \rangle$, (**b**) clustering coefficient C, (**c**) graph density ρ, (**d**) modularity Q, (**e**) average path length $\langle l \rangle$, (**f**) network diameter $d(G)$, and (**g**) number of components change as a function of the connectivity threshold F_c, for a series $X_{1:100} = \{1, 2,\ldots, 100\}$, where F_c rangeswithin the interval $F_c \in [0, \max\{X_{1:100}\}^2]$. The bold vertical line represents the typical value f_z defined at relation (4), whereas the dashed horizontal line a point estimate of the average y values.

threshold is considered satisfactory to provide a reference value that is representative of the average topological features of the ESGs.

Testing the performance of the ESG algorithm. The analysis examines five different types of time-series, as it is shown in Fig. 5. The first one (Fig. 5a) was extracted from AirPassengers[30] and is a time-series with a linear trend (abbreviated: $X_a \equiv$ AIR), including the monthly totals of US airline passengers for the period 1949 to 1960 (144 cases). The second one (Fig. 5b) was extracted from LorentzTS[31] and is a typical Lorentz chaotic time-series ($X_b \equiv$ CHAOS) generated from the Lorenz differential equations, on standard values sigma = 10.0, r = 28.0, and b = 8/3. This time-series has a length of 1900 cases. The third one (Fig. 5c) was extracted from DEOK.hourly[32] and is a part (the first 5000 cases) of a broader stationary time-series (of 57,739 cases) including estimated energy consumption, in Megawatts (MW), for the Duke Energy Ohio/Kentucky ($X_c \equiv$ DEOK). Next,
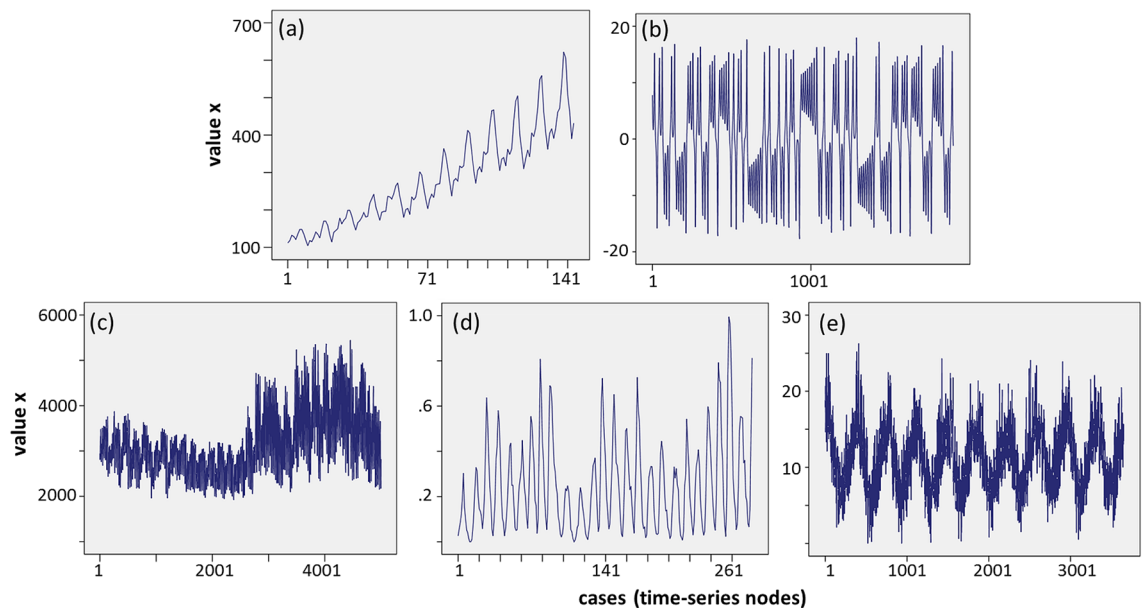
Figure 5. The source (reference) time-series considered in the analysis represent distinctive different patterns, where (**a**) is an air-passengers time-series with linear trend (X_a: 144 cases, including the monthly totals of a US airline passengers for the period 1949 to 1960), (**b**) is the typical Lorentz chaotic time-series (X_b: 1900 cases, created from the Lorenz equations, on standard values sigma = 10.0, r = 28.0, and b = 8/3), (**c**) is a part (X_c: 5000 cases) of a broader stationary time-series including estimated energy consumption, in Megawatts (MW), for the Duke Energy Ohio/Kentucky, (**d**) is a periodical time-series (X_d: 280 cases, including wolfer sunspot numbers for the period 1770 to 1771), and (**e**) is a cyclic time-series (X_c: 3650 cases, including daily minimum temperatures in Melbourne, Australia, for the period 1981–1990).

the fourth one (Fig. 5d) was extracted from Wolfer-sunspot-numbers[33] and is a periodic time-series including Wolfer sunspot numbers (X_d≡SUNSPOTS), for the period 1770 to 1771 (280 cases). The fifth one (Fig. 5e) was extracted from Daily-minimum-temperatures-in-me[34] and is a cyclical time-series including daily minimum temperatures in Melbourne, Australia (X_e≡TEMP), for the period 1981–1990 (3650 cases). Links to the time-series databases are available in the reference list.

To examine the effectiveness of the proposed algorithm, we firstly compare the structure of the source time-series X with its node-series $X_{ESG}(Ys)$ of the ESG node measures (Ys). Such comparisons are driven by the rationale that the ESG is a transformation (conversion) of a time-series to a complex network and therefore possible similarities that can be detected in the structural properties (e.g. data variability, linear trend, chaotic, stationary, periodic, and cyclical structure) between the original time-series and its associated ESG node-series can be seen as aspects of homeomorphism describing this transformation. In general, this approach is expected to illustrate the level at which the topology of the associated electrostatic graph ESG(X) sufficiently incorporates structural information of the source time-series X. Secondly, we compare the structure of the $X_{ESG}(Ys)$ node-series with this of their concordant node-series $X_{VGA}(Ys)$ of the node measures (Ys) computed in the visibility graphs defined by Lacasa et al.[5]. The comparisons between the source time-series and its associated node-series (either of ESG or VGA conversion) build on a multilevel analysis consisting of five tests; the first one detects similarities in data-variability (i.e. whether the original time-series and the node-series have the same fluctuation patterns) based on the Pearson's bivariate coefficient of correlation[35,36], the second one in linear-trend by using the Linear Regression (LSLR) fitting[36], the third one in chaotic-structure based on the correlation dimension and embedding dimension diagram[37], the fourth one in stationary-structure based on the augmented Dickey-Fuller test (ADF) for a unit root [38], and the fifth one in periodic-structure based on autocorrelation function[38]. Each test is briefly described in the following paragraphs.

The visibility graph algorithm. The natural visibility algorithm (NVG) was proposed by Lacasa et al.[5] and builds on the intuition of considering a time-series as a path of successive mountains of different height, where each represents the value of the time-series at a certain time. In this time-series landscape, an "observer" standing on the top of a mountain can see (either forward or backward) as far as possible, provided that no other top obstructs its visibility field (Fig. 6).

In mathematical terms, each time-series node (t_i, $x(t_i)$) corresponds to a graph node $i≡(t_i, x(t_i))∈V$, and thus two nodes $i,j∈V$ are connected $(i,j)∈E$ in the visibility graph when the following inequality (NVG connectivity criterion) is satisfied:

$$X(t_k) < X(t_i) + (X(t_j) - X(t_i))\frac{t_k - t_i}{t_j - t_i}, \forall k \in (i,j),$$ (6)
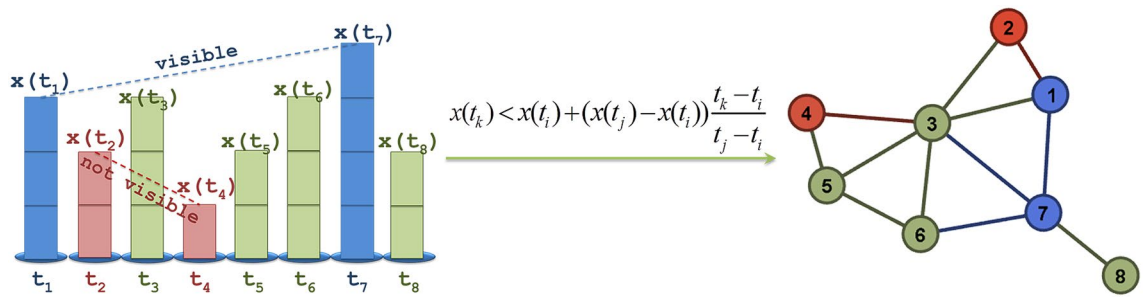
Figure 6. (left) Example a pair of visible (shown in blue colour) and another of not visible (shown in red colour) time-series nodes (generally shown in green colour) defined according to the natural visibility algorithm (NVG), (right) the visibility graph generated from the time-series shown at the left side.

where $X(t_i)$ and $X(t_j)$ are the numerical values of the time-series nodes $(t_i, x(t_i)) \equiv i$ and $(t_j, x(t_j)) \equiv j$ and t_i, t_j express their time points. In geometric terms, a visibility line can be drawn between two time-series nodes $i,j \in V$, if no other intermediating node $(t_k, x(t_k)) \equiv k$ obstructs their visibility. That is, two time-series nodes are connected in the visibility graph whether no other intermediary node is higher so that to intersect the visibility line defined by this pair of nodes (Fig. 6). Therefore, two time-series nodes can enjoy a connection in the associated visibility graph if they are visible through a visibility line. The visibility algorithm conceptualizes the time-series as a landscape and generates a visibility graph associated with this landscape. The associated (to the time-series) visibility graph is a complex network where complex network analysis can be further applied[8,16].

Correlation analysis. At the first step of the analysis, we detect linear correlations between the source time-series X and the available (ESG and VGA) node-series. This approach examines whether the original time-series X and the node-series $\{X_i(k), X_i(s), X_i(C), X_i(CB), X_i(CC),$ and $X_i(CE) \mid i = \text{ESG,VGA}\}$ have the same fluctuation patterns and thus they can be considered as relevant in terms of data variability. In this analysis, the Pearson's bivariate coefficient of correlation[35,36] is used, which ranges at the interval $r_{X,Y} \in [-1,1]$ and detects linear (either positive or negative) correlations when $|r_{XY}| \to 1$.

Test of the linear trend. To detect a linear trend, we apply linear fittings to the source time-series X and to its associated node-series $\{X_i(k), X_i(s), X_i(C), X_i(CB), X_i(CC),$ and $X_i(CE) \mid i = \text{ESG,VGA}\}$. According to this approach[36], a linear curve $\hat{y} = b \cdot f(x) + c$ is fitted to the available data that bests describes their variability. The curve fitting algorithm estimates the parameters b, c minimizing the square differences $y_i - \hat{y}_i$[36], according to the relation:

$$\min \left\{ e = \sum_{i=1}^{n} [y_i - \hat{y}_i]^2 = \sum_{i=1}^{n} \left[y_i - \left(\sum b_i f_i(x) + c \right) \right]^2 \right\}, \tag{7}$$

where y_i express the observed and $\hat{y}_i$ the estimated values. The optimization method that is used is the Least-Squares Linear Regression (LSLR) method[36], which assumes that the differences $e = \sum_{i=1}^{n} [y_i - \hat{y}_i]^2$ follow the normal distribution $e \sim N(0, \sigma_e^2)$. The goodness of the model fit is measured by the coefficient of determination R^2, which is defined by the expression[35,36]:

$$R^2 = \left(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \right) \Big/ \left(\sum_{i=1}^{n} (y_i - \bar{y})^2 \right), \tag{8}$$

where $\bar{y}$ is the average of the observations and n is the number of cases (i.e. the series length). The coefficient of determination expresses the amount of variability of the response variable that is expressed by the linear model and ranges within the interval $[0,1]$, indicating perfect linear determination when $R^2 = 1$[35,36]. Within this context, amongst the ESG and VGA node-series, those being closer to the source time-series X in determination and model configuration (i.e. values in b and c estimators) are considered as more relevant to X in terms of linear trend.

Detection of chaotic structure. To detect chaotic structure in a time-series, we examine the patterns of the correlation (v) versus the embedding dimension (m) scatter plots (v,m). According to the Chaos theory[37], the correlation dimension (v) is a measure of the dimensionality of the space occupied by a set of random points and thus is used to determine the dimension of the fractal objects, which is often called fractal dimension. For a time-series $X = \{x_i \mid i = 1, …, n\}$, the correlation integral $C(\varepsilon)$ is calculated by the expression[39,40]:

$$C(\varepsilon) = \lim_{n \to \infty} \frac{N(\varepsilon)}{n^2} \sim \varepsilon^v, \tag{9}$$

where $N(\varepsilon)$ is the total number of pairs of time-series points (x_i, x_j) with a distance smaller than ε, namely $d(x_i, x_j) = d_{ij} < \varepsilon$. As the number of points tends to infinity $(n \to \infty)$, and therefore as their corresponding distances

tend to zero ($d_{ij} \to 0$), the correlation integral tends to the quantity $C(\varepsilon) \sim \varepsilon^\nu$, where ν is the so-called correlation dimension. Intuitively, the correlation dimension expresses the ways to which the points can be close to each other along different dimensions and is expected to rise faster when the space of embedding is of a higher dimension. Therefore, the correlation (ν) versus the embedding dimension (m) diagram (ν,m) can provide insights into how the time-series points are close to each other, as the dimensionality of the space of embedding increases[39,40]. Within this context, amongst the ESG and VGA node-series, those with the (ν,m) diagram being closer to the source time-series X are considered as more relevant to the original time-series, in terms of chaotic structure.

Detection of stationarity. To detect stationarity in the available series we apply the augmented Dickey-Fuller test (ADF) for a unit root [38]. The ADF algorithm examines the null hypothesis (H_o) that a unit-root is present in the model's time-series data, which is expressed by the relation:

$$y_t = c + \delta t + \phi \cdot y_{t-1} + \beta_1 \cdot \Delta y_{t-1} + \cdots + \beta_p \cdot \Delta y_{t-p} + \varepsilon_t, \tag{10}$$

where Δ is the differencing operator ($\Delta y_t = y_t - y_{t-1}$), p is the number of lagged difference terms (specified by the user), c is a drift term, δ is a deterministic trend coefficient, ϕ is an autoregressive coefficient, β_i are the regression coefficients of the lag differences, and ε_t is a mean zero innovation process. According to Eq. (10), the unit-root hypothesis testing is expressed as follows[38]:

$$H_o : \phi = 1 \text{ vs.} H_1 : \phi < 1, \tag{11}$$

and the (lag adjusted) test statistic DF is defined by the expression[38]:

$$DF = \frac{N(\widehat{\phi} - 1)}{(1 - \widehat{\beta}_1 - \ldots - \widehat{\beta}_p)}, \tag{12}$$

where the uppercase symbol '^' expresses an estimator. Within this context, amongst the node-series of ESG and VGA, first those satisfying the null hypothesis and then those that have more similar DF statistics with the source time-series X are considered as more relevant to the original time-series, in terms of stationarity.

Detection of periodicity and cyclical structure. To detect periodicity in the available time-series we use the autocorrelation function (ACF), which is defined as:

$$\rho(s,t) = \frac{\gamma_x(s,t)}{\sqrt{\gamma_x(s,s)\gamma_x(t,t)}}, \tag{13}$$

where (s,t) are time points and $\gamma_x(s,t)$ is the autocovariance function of the variable x[38]. In general, the ACF measures the linear predictability of the series at time t by using only the value x_s (at time s) with a time-lag $dt = t - s$. The ACF lies within the interval $-1 \le \rho(s,t) \le 1$, where positive coefficient values imply a positive linear trend and negative values a negative one. Based on the ACF, we construct a set of ACF-variables, where the first refers to the source time-series X and the others to the node-series $X_i(k)$, $X_i(s)$, $X_i(C)$, $X_i(CB)$, $X_i(CC)$, and $X_i(CE)$, where index i can be either $i = $ ESG or VGA. Each variable includes 30 elements corresponding to ACFs of lag $dt = 1,2,\ldots,30$, respectively, namely:

$$ACF(X) = \{\rho(t, t+1), \rho(t, t+2), \ldots, \rho(t, t+30)\}. \tag{14}$$

By constructing these ACF-variables, we compute the Pearson's bivariate coefficient of correlation[35,36] to detect linear correlations between the $ACF(X)$ variable of the source time-series X and the other node-series variables. Within this context, amongst the available ESG and VGA node-series variables, those being higher correlated with the source time-series X are considered as more relevant to the original time-series in terms of periodicity and cyclical (i.e. periodic with a constant oscillation height) structure.

Results

Spy plots and graph layouts. The spy plots and graph layouts of the ESG(X) and VGA(X) graphs associated with the time-series X are shown in Fig.A1-A5 (in the Appendix). The spy plots are matrix-plots displaying with dots the non-zero elements of the adjacency matrix and they can thus represent the graph topology within the matrix-space[3,41]. On the other hand, network visualization is implemented by using the "Force-Atlas" layout, which is available in the open-source software of Bastian et al.[42]. This layout is generated by a force-directed algorithm, which applies repulsion strengths between network hubs while it arranges hubs' connections into surrounding clusters. Graph models that are represented in this layout have therefore their hubs centered and mutually distant (i.e. intermediate distance between hubs is the highest possible), whereas lower-degree nodes are placed as closely as possible to their hubs[3].

As it can be observed in Fig.A1 (Appendix), the ESG(X_a) spy plot has a connectivity pattern configuring a tie (along the main diagonal) of increasing width (Fig.A1.a,c, Appendix), which appears indicative of the increasing trend of the source time-series ($X_a = $ AIR). An aspect of such trend is also evident in the chain-like ESG(X_a) graph layout (Fig.A1.e, Appendix), where a cluster of hubs appears on the right side that resembles the tie configuration shaped in the spy plot. Also, the saw-like pattern of the source time-series appears smoother in the pattern of the 2d ESG(X_a) spy plot (Fig.A1.a, Appendix), whereas is more evident in the diagonal arrangement of the 3d ESG(X_a) spy plot (Fig.A1.c, Appendix). On the other hand, the VGA(X_a) spy plot configures a periodic pattern

(Fig.A1.b,d, Appendix), where no linear trends are visible. This can be also observed in the VGA(X_a) graph layout (Fig.A1.f, Appendix), which shapes an almost symmetric hub-and-spoke pattern.

In Fig.A2, the ESG(X_b) spy plot configures a fractal-like tiling (Fig.A2.a, Appendix) illustrating a chaotic structure. Although such structure in the ESG(X_b) graph layout (Fig.A2.f, Appendix) is not that clear, we can observe two major components composing the electrostatic graph of X_b (Lorentz time-series). This is a result of the positive and negative values in the structure of the source time-series (X_b), illustrating the ability of the electrostatic graph (ESG) algorithm to generate disconnected graphs[1,41,27]. Although connectivity is generally a desirable property in complex networks, the ability of the ESG algorithm to generate disconnected graphs can be insightful for removing past or unnecessary information (noise) of the time-series, therefore proposing avenues for further research. On the other hand, the VGA(X_b) graph layout (Fig.A2.f, Appendix) better illustrates a chaotic structure than its spy plot (Fig.A2.b,d Appendix) does, which is more illustrative to a periodic than to chaotic structure.

Next, in Fig.A3 (Appendix) the ESG(X_c) spy plot (Fig.A3.a,c, Appendix) configures a tie (along the main diagonal), with an almost constant width, which complies with the stationary structure of the source time-series (X_c = DEOK). Some evidence of stationarity can be also observed in the concentrated (solid-like) pattern of the ESG(X_c) graph layout (Fig.A3.e, Appendix). On the other hand, neither the VGA(X_c) spy plot (Fig.A3.b,d) nor graph layout (Fig.A3.f, Appendix) are illustrative of a stationary structure describing the original time-series (X_c).

In Fig.A4 (Appendix), the ESG(X_d) spy plot also configures a tie (along the main diagonal) with repeated knot-concentrations (Fig.A4.a,c, Appendix), which complies with the periodic structure of the source time-series (X_d = SUNSPOTS). Some insightful indications of such periodicity can be also observed in the clustered (torus-like) pattern that is shown in the ESG(X_d) graph layout (Fig.A4.e, Appendix). On the other hand, the VGA(X_c) spy plot (Fig.A4.b,d, Appendix) has an interesting periodic pattern, which is slightly mixed by the square areas of the other connections. However, the VGA(X_d) graph layout (Fig.A4.f, Appendix) does not appear illustrative of the periodic structure describing the source time-series (X_d).

Finally, the ESG(X_e) spy plot configures a tie (along the main diagonal) with repeated slightly thicker segments (Fig.A5.a,c, Appendix), which can relate to the cyclical structure describing the source time-series (X_e = TEMP). However, such cyclical structure is almost hidden in the chain pattern of graph components that have an odd arrangement in the ESG(X_d) graph layout (Fig.A5.e, Appendix). Periodicity can become clearer whether the layout will be further stretched to succeed symmetric arrangement similar to this of Fig.A4.e (Appendix). On the other hand, the VGA(X_c) spy plot shapes a clearer periodic pattern (Fig.A5.b,d, Appendix), which (although difficult) can be observed in the graph layout (Fig.A5.f, Appendix). Overall, the proposed ESG algorithm appears at least as capable as the VGA is in generating graphs of topologies representative of their source time-series. This observation will be also quantitatively tested in the following sections.

Correlation analysis. To compare patterns in data variability between the source and the ESG and VGA node-series (see Fig. A6-A10, Appendix), we apply a Pearson's bivariate correlation analysis, the results of which are shown in Table 2. Amongst the available correlation coefficients, we compare concordant pairs ($r(X,X_{ESG}(z)$, $r(X,X_{VGA}(z)|z=k$, C, CB, CC, and CE) between ESG and VGA node-series and we denote pairwise maxima ($\max\{(r(X,X_{ESG}(z), r(X,X_{VGA}(z))\}$ in bold font. Cases with the $X_{ESG}(s)$ node-series are paired with those of corresponding degree $X_{VGA}(k)$, due to the similarity of the measures of node degree (k) and node strength (s), for the binary and weighted networks. Within this context, according to Table 2, in the case of the X_a time-series, the variability of the ESG node-series is overall closer to this of the source time-series (X_a) than the variability of the VGA node-series overall is, because the ESG node-series count 5 out of 6 maxima, whereas the VGA node-series count just one. This observation implies that the ESG transformation generates graphs that better preserve fluctuations with a linear trend of the original time-series than the VGA does. On the contrary, in the case of the chaotic time-series (X_b), the VGA node-series count 5 out of 6 maxima (a double count is given for the k,s pair), whereas the ESG node-series count just one. This observation implies that the VGA transformation better preserves chaotic fluctuations of the original time-series than the ESG does.

In the case of the X_c (DEOK), the ESG node-series count 4 out of 6 maxima, whereas the VGA node-series count 2 out of 6, which implies that the ESG transformation better preserves stationary fluctuations of the original time-series than the ESG does. In the case of the X_d (SUNSPOTS), the ESG node series count 5 out of 6 maxima, whereas the VGA node-series count just one, which implies that the ESG transformation better preserves periodical fluctuations of the original time-series than the VGA does. In the case of the X_e (TEMP) both the ESG and the VGA node-series count 3 out of 6 maxima, showing a balanced performance. As far as the measure of strength (s) (see Fig. A11, Appendix) is concerned, the analysis shows that, for all types of time-series except the chaotic one (X_b, CHAOS), the ESG node-series have higher performance than the VGAs. Overall, this pairwise consideration illustrates that the variability of ESG node-series is closer to the source time-series (X_i) than of the VGAs, since the first count 18 out of 30 maximum cases, whereas the latter count 12 out of 30 maxima.

Test of the linear trend. The test of the linear trend was applied to ESG and VGA node-series associated with the X_a (AIR) time-series, which is a time-series with a known linear trend. The results of the analysis are shown in Table 3, where first it can be observed that the source (X_a: AIR) time-series is well described by a linear regression model ($R^2 = 0.8536$). However, none of the VGA node-series can sufficiently retain this linear structure, as is evident by the low coefficients of determination ranging from $R^2 = 0.0002$ to $R^2 = 0.0132$.

On the contrary, the ESG node-series of degree $X_{ESG}(k)$, strength $X_{ESG}(s)$, and eigenvector centrality $X_{ESG}(CE)$ have a considerable linear structure, as is denoted by their respective coefficients of determination $R^2 = 0.6916$, $R^2 = 0.8012$, and $R^2 = 0.7579$. It should be noted that among these cases, the strength node-series $X_{ESG}(s)$ have the

www.nature.com/scientificreports/

VariableX_i (source time-series)	Measure	Variable$X_{ij}(z^{(a)})$ (node-series)										
		j = VGA					j = ESG					
		z = k	z = C	z = CB	z = CC	z = CE	z = k	z = s[b]	z = C	z = CB	z = CC	z = CE
i = a (AIR)	$r(X_i,X_{ij}(z))$	0.331**	-0.158	**0.250****	-0.232**	0.254**	**0.805****	<u>**0.981****</u>	**0.358****	-0.144	**-0.359****	**0.837****
	sig.[c]	0.000	0.059	**0.002**	0.005	0.002	**0.000**	**0.000**	**0.000**	0.085	**0.000**	**0.000**
	n[d]						144					
i = b (CHAOS)	$r(X_i,X_{ij}(z))$	<u>**0.516****</u>	0.019	**0.188****	**-0.185****	**0.201****	-0.008	0.002	**-0.022**	0.020	0.103**	-0.132**
	sig	**0.000**	0.401	**0.000**	**0.000**	**0.000**	0.712	0.919	**0.345**	0.375	0.000	0.000
	n						1900					
i = c (DEOK)	$r(X_i,X_{ij}(z))$	0.354**	**-0.570****	0.148**	**-0.140****	0.157**	**0.890****	<u>**0.989****</u>	-0.549**	**0.179****	0.088**	**0.725****
	sig	0.000	**0.000**	0.000	**0.000**	0.000	**0.000**	**0.000**	0.000	**0.000**	0.000	**0.000**
	n						5000					
i = d (SUNSPOTS)	$r(X_i,X_{ij}(z))$	0.496**	**-0.666****	0.478**	-0.567**	0.309**	**0.768****	<u>**0.773****</u>	.[b]	**0.593****	**0.712****	**0.733****
	sig	0.000	**0.000**	0.000	0.000	0.000	**0.000**	**0.000**		**0.000**	**0.000**	**0.000**
	n						280					
i = e (TEMP)	$r(X_i,X_{ij}(z))$	0.437**	**-0.439****	**0.211****	-0.520**	0.164**	**0.908****	<u>**0.944****</u>	0.405**	0.123**	-0.133**	**0.752****
	sig	0.000	**0.000**	**0.000**	0.000	0.000	**0.000**	**0.000**	0.000	0.000	0.000	**0.000**
	n						3650					

Table 2. Results of the Pearson's bivariate correlation analysis. [a]k = node degree, C = clustering coefficient, CC = closeness centrality, CB = betweenness centrality, CE = eigenvector centrality. c. In pairwise consideration, $X_{ESG}(s)$ is paired with the $X_{VGA}(k)$. c. 2-tailed significance. d. Number of cases. **Correlation is significant at the 0.01 level (2-tailed). Cases shown in **bold** indicate maximum coefficients (in absolute terms) between concordant ESG versus VGA pairs, $\max\{r(X,X_{ESG}(z)), r(X,X_{VGA}(z))\}$. <u>Underlined</u> cases indicate max coefficients (in absolute terms) within each row (for each time-series type).

highest determination. Overall, this analysis illustrates that the ESG algorithm appears more capable than the VGA in generating graphs that can preserve aspects of the linear trend of the source time-series.

Detection of chaotic structure. In this part of the analysis, the correlation versus the embedding dimension diagrams (v,m) of the VGA and the ESGs node-series are compared for preserving the chaotic structure of the source time-series X_b (CHAOS), which is already known as a chaotic time-series constructed on the Lorenz equations. The results are shown in Fig. A7 (Appendix), where all (v,m) diagrams of the ESG node-series (except this of eigenvector centrality $X_{b,ESG}(CE)$) illustrate the chaotic structure, but of different characteristics than the source chaotic time-series X_b. However, the (v,m) diagrams of strength $X_{b,ESG}(s)$ and the original time-series X_b almost coincide, a fact that implies a relevant chaotic structure between these time-series. On the other hand, the degree $X_{b,VGA}(k)$, and possibly the eigenvector centrality $X_{b,VGA}(CE)$ VGA node-series illustrate a chaotic structure of high dimensionality, which are also of different characteristics than the original chaotic time-series X_b. Overall, the chaos analysis shows that the ESG is a more capable transformation in incorporating the chaotic structure of the source time-series in the network topology. Particularly, the measure of strength shows the most relevant chaotic structure that almost coincides with this of source time-series.

Detection of stationarity. The test of stationarity was applied to the X_c (DEOK) time-series, which is a part of an already known stationary time-series. The results of the analysis are shown in Table 4, where, first, it can be observed that is 7.03% likely for X_c to have a unit-root and thus to be a non-stationary time-series. This result implies that the null-hypothesis (stating a null unit-root) cannot be rejected, and thus that the source (X_c) time-series cannot be considered as a stationary one. As it can be observed, the results for all VGA node-series imply that all cases are statistically safe to be considered as stationary series, which opposes the indication of the original time-series.

On the other hand, the ESG results imply that 4 out of 5 ESG node-series cannot be considered as stationary ones and thus resembling the structure of the original time-series. An interesting observation here is that the p-values of the VGA node-series are (although insufficient indications to retain the null hypothesis) closer than those of the ESG node-series, in terms of distance. These results imply that the non-stationary effects, which are immanent in the source time-series, probably appear more intensely in the structure of the ESG node-series than of the VGA ones.

Detection of periodicity and cyclical structure. This part of analysis builds on bivariate correlations, which are applied to autocorrelation variables $ACF(X)$ that are defined in relation (14) with lag 1,2,…,30, where $X = X_d$ (SUNSPOTS time-series), X_e (TEMP time-series), k (degree node-series), C (clustering coefficient node-series), CB (betweenness centrality node-series), CC (closeness centrality node-series), and CE (eigenvector centrality node-series). The results of the analysis are shown in Table 5, where the correlation coefficients r_{XY} and their significances are provided, with $X \in \{ACF(X_d), ACF(X_e)\}$ and $Y \in \{ACF_i(k), ACF_i(s), ACF_i(C), ACF_i(CB), ACF_i(CC), ACF_i(CE) \mid$ where $i = VGA, ESG\}$.

Time-series/ Measure		Linear regression	
		Mathematical expression	Determination
x_a (AIR)		$y = 2.6572x + 87.653$	$R^2 = 0.8536$
VGAs	$k(x_a)$	$y = 0.0118x + 7.0629$	$R^2 = 0.0096$
	$C(x_a)$	$y = 0.0006x + 0.7174$	$R^2 = 0.0132$
	$CB(x_a)$	$y = 0.174x + 141.63$	$R^2 = 0.0003$
	$CC(x_a)$	$y = -0.0002x + 3.1682$	$R^2 = 0.0002$
	$CE(x_a)$	$y = 0.0001x + 0.2013$	$R^2 = 0.0009$
ESGs	$k(x_a)$	$y = 0.2149x + 13.656$	$R^2 = 0.6916$
	$s(x_a)$	$y = 4813.4x - 62,583$	$\underline{R^2 = 0.8012}$
	$C(x_a)$	$y = 0.0009x + 0.6969$	$R^2 = 0.1985$
	$CB(x_a)$	$y = -1.5134x + 315.61$	$R^2 = 0.0663$
	$CC(x_a)$	$y = -0.0106x + 4.6484$	$R^2 = 0.2034$
	CE	$y = 0.0069x - 0.1143$	$R^2 = 0.7579$

Table 3. Linear regression fittings for the X_a (Air) time-series. Cases shown in **bold** indicate high (> 0.65) coefficients of determination. <u>Underlined</u> cases highlight the max coefficient, in total.

		ADF test			
		$h^{(a)}$	p-Value	DFt (stat)	cValue
x_c (DEOK)		**0**	**0.0703**	**-1.7876**	**-1.9416**
VGAs	k	1	**0.0010$^{(b)}$**	-16.8555	
	C	1	**0.0010**	-8.8160	
	CB	1	**0.0010**	-63.9394	**-1.9416**
	CC	1	**0.0176**	**-2.3683**	
	CE	1	**0.0010**	-28.4443	
ESGs	k	0	0.2176	**-1.1860**	
	s	0	0.3876	**-0.7177**	
	C	0	0.3458	**-0.8359**	
	CB	1	0.0010	**-20.2674**	**-1.9416**
	CC	0	0.1739	**-1.3179**	
	CE	0	0.2251	**-1.1656**	

Table 4. ADF test for stationarity of the X_c (DEOK) time-series. ^{a}$h = 0$ indicates failure to reject the unit-root null (indication: non-stationary series). $h = 1$ indicates rejection of the unit-root null in favor of the alternative model (indication: stationary series). b. Cases shown in **bold** are the closest between concordant measures to the source time-series scores.

For the case of X_d (SUNSPOTS) time-series, we can observe that 4 out of 6 VGA node-series (k, $k \equiv s$, C, CE) and 3 out of 6 ESG node-series (k, s, CC) are significantly correlated with the original time series X_d. Amongst these significant results, the VGA node-series have 2 maxima of concordant pairs, whereas the ESG node-series have also 2 maxima. Moreover, the node-series of strength ($X_{d,\mathrm{ESG}}(s)$) has the highest correlation coefficient amongst all available node-series for the SUNSPOTS (X_d) typology, illustrating a better performance of the ESG algorithm to preserve periodicity, probably due to its capability in generating weighted electrostatic networks. For the case of X_e (TEMP) time-series, 1 VGA node-series (closeness centrality) is significantly correlated with the source time-series, where all ESG node-series are significantly correlated with the original time-series. In terms of pairwise comparisons, the VGA node-series count 1 (out of 6) maximum case, whereas the ESG node-series count 5 out of 6 maxima. However, although is high, the strength does suggest the highest of the maxima of the TEMP (X_e) time-series concordant pairs. Overall, this analysis shows that the ESG node-series appear more capable than the VGA ones in preserving periodic and cyclical characteristics of the source time-series.

Conclusions

This paper proposed a new algorithm, the Electrostatic Graph Algorithm (ESG), for converting a time-series into a graph (complex network). The ESG builds on the conceptualization of considering a time-series as a series of stationary and electrically charged particles, on which Coulomb-like forces can be computed. The proposed algorithm provides an added value to complex network analysis of time-series due to its ability to produce weighted graphs, which is currently not applicable. This additional property was quantitatively examined in this paper and was found to produce graphs that are more representative of the structure of the source (original) time-series, implying that the proposed algorithm suggests a transformation that is more natural rather than

Y-variable		X-variable			
		ACF(X_d) [SUNSPOTS]		ACF(X_e) [TEMP]	
		r_{XY}	Sig	r_{XY}	Sig
VGA	ACF(k)	**0.8397**	0	-0.0803	0.613
	ACF(C)	0.7953	0	-0.1311	0.408
	ACF(CB)	-0.082	0.6057	0.2807	0.0717
	ACF(CC)	0.4225	0.0053	<u>1</u>	0
	ACF(CE)	**0.823**	0	0.3306	0.0325
ESG	ACF(k)	0.6513	0	**0.9999**	0
	ACF(s)	<u>**0.9379**</u>	0	**0.9196**	0
	ACF(C)	NaN**	NaN	**0.9999**	0
	ACF(CB)	0.2398	0.1261	**0.9998**	0
	ACF(CC)	**0.7105**	0	0.9999	0
	ACF(CE)	0.5601	0.0001	**0.9999**	0

Table 5. Correlations of ACFs$^{(*)}$ between the source and the node-series (Sunspots and Temp). *ACFs are computed on lags from 1:30 (relation 13). **Not a number (unable to compute due to zero entries). Cases shown in **bold** indicate max coefficients (in absolute terms) according to pairwise (ESG *vs.* VGA) comparisons (only statistically significant cases are shown). <u>Underlined</u> cases indicate max coefficients (in absolute terms) within each column (for each time-series type).

algebraic, in comparison with the existing methods. In particular, the analysis showed that the ESG node-series can better preserve the linear trend and stationary structural properties of the source time-series in comparison with the VGA node-series and that they appear slightly better in preserving periodical and cyclical structural properties of the original time-series than the VGA node-series can. On the other hand, the VGA node-series appeared slightly better in preserving the chaotic structural properties of the original time-series in comparison with the ESG node-series, which complies with the claim of the VGA authors regarding the added value of their algorithm. However, in almost all the parts of the analysis, the ESG node-series of the measure of strength outperformed their concordant VGA node-series. This result highlighted the added value of the proposed algorithm in generating weighted graphs, in which the measure of node strength can only be computed. Therefore the ESG algorithm attributes to the generated graphs information that is more representative of the source time-series, due to the weights included in the graph structure. Another property of the proposed ESG algorithm to generate disconnected graphs was indirectly examined by the detection of chaotic and periodic structures, where the ESG algorithm sufficed to provide disconnected graphs, whereas the VGA did not. This analysis showed that insufficient connectivity does not restrict the ESG node-series to preserve the structural characteristics of the source time-series, since the generated electrostatic graphs were representative of the structure of the original time-series. The authors believe that the property of insufficient connectivity introduces avenues of further research in the field of noise reduction in the time-series analysis. Other avenues of further research can emerge towards the direction of either choosing the optimum or most representative connectivity threshold to produce the ESGs or examining the applicability of the proposed algorithm to solve problems where standard methods fail to analyze efficiently the time-series, such as the time evolution of stock price, within the framework of Black Scholes model, and others. The overall approach also suggests a methodological framework for evaluating the structural relevance between the source time-series and their associated graphs produced by any possible transformation.

References

1. Barabasi, A.-L. Network science. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **371**(1987), 20120375 (2013).
2. Brandes, U., Robins, G., McCranie, A. & Wasserman, S. What is network science?. *Netw. Sci.* **1**, 1–15 (2013).
3. Tsiotas, D. Detecting different topologies immanent in scale-free networks with the same degree distribution. *Proc. Natl. Acad. Sci. U. S. A. (PNAS)* **116**(14), 6701–6706 (2019).
4. Gao, Z.-K., Small, M. & Kurths, J. Complex network analysis of time-series. *Europhys. Lett.* **116**, 50001 (2016).
5. Lacasa, L., Luque, B., Ballesteros, F., Luque, J. & Nuno, J. C. From time-series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci.* **105**(13), 4972–4975 (2008).
6. Yang, Y. & Yang, H. Complex node-series analysis. *Phys. A* **387**(5), 1381–1386 (2008).
7. Zhang, J. & Small, M. Complex network from pseudoperiodic time-series: Topology versus dynamics. *Phys. Rev. Lett.* **96**(23), 238701 (2006).
8. Tsiotas, D. & Charakopoulos, A. VisExpA: Visibility expansion algorithm in the topology of complex networks. *SoftwareX* **11**, 100379 (2020).
9. Xu, X., Zhang, J. & Small, M. Superfamily phenomena and motifs of networks induced from time-series. *Proc. Natl. Acad. Sci.* **105**(50), 19601–19605 (2008).
10. Gao, Z.-K. & Zin, N. Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks. *Phys. Rev. E* **79**(6), 066303 (2009).

11. Donner, R. V., Zou, Y., Donges, J. F., Marwan, N. & Kurths, J. Recurrence networks—a novel paradigm for nonlinear time-series analysis. *New J. Phys.* **12**(3), 033025 (2010).
12. Iacobello, G., Scarsoglio, S. & Ridolfi, L. Visibility graph analysis of wall turbulence time-series. *Phys. Lett. A* https://doi.org/10.1016/j.physleta.2017.10.027 (2017).
13. Jiang, W., Wei, B., Zhan, J., Xie, C. & Zhou, D. A visibility graph power averaging aggregation operator: A methodology based on network analysis. *Comput. Ind. Eng.* **101**, 260–268 (2016).
14. Liu, C., Zhou, W. X. & Yuan, W. K. Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. *Phys. A* **389**(13), 2675–2681 (2010).
15. Luque, B., Lacasa, L., Ballesteros, F. & Luque, J. Horizontal visibility graphs: Exact results for random time-series. *Phys. Rev. E* **80**(4), 046103 (2009).
16. Tsiotas, D. & Charakopoulos, A. Visibility in the topology of complex networks: Introducing a new approach. *Physica A* **505**, 280–292 (2018).
17. Serway, R. *Physics for Scientists & Engineers with Modern Physics* 6th edn. (Thomson Books, California, 2004).
18. Ghasemi, M., Ghavidel, S., Aghaei, J., Akbari, E. & Li, L. CFA optimizer: A new and powerful algorithm inspired by Franklin's and Coulomb's laws theory for solving the economic load dispatch problems. *Int. Trans. Electr. Energy Syst.* **28**(5), e2536 (2018).
19. Sousa, W. & de Oliveira, R. Coulomb's law discretization method: A new methodology of spatial discretization for the radial point interpolation method. *IEEE Antennas Propag. Mag.* **57**(2), 277–293 (2015).
20. Lai, Y., Lv, Z., Li, K. C. & Liao, M. Urban traffic Coulomb's law: A new approach for taxi route recommendation. *IEEE Trans. Intell. Transp. Syst.* **20**(8), 3024–3037 (2018).
21. Van Hooydonk, G. Gauge symmetry, chirality and parity effects in four-particle systems: Coulomb's law as a universal function for diatomic molecules. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **56**(12), 2273–2331 (2000).
22. Cagnoli, B., & Manga, M. Granular mass flows and Coulomb's friction in shear cell experiments: Implications for geophysical flows. *J.Geophys. Res. Earth Surf.* **109**(F4), 1–12 (2004).
23. Rahkar-Farshi, T. & Behjat-Jamal, S. A multimodal firefly optimization algorithm based on Coulomb's Law. *Int. J. Adv. Comput. Sci. Appl.* **7**(5), 134–141 (2016).
24. Zhang, H., Wei, D., Hu, Y., Lan, X. & Deng, Y. Modeling the self-similarity in complex networks based on Coulomb's law. *Commun. Nonlinear Sci. Numer. Simul.* **35**, 97–104 (2016).
25. West, D. B. Introduction to Graph Theory, 2nd Ed. (Prentice Hall, Upper Saddle River, 2001).
26. Yun, B. I & Petkovic, M. S. Iterative methods based on the signum function approach for solving nonlinear equations. *Numer. Algorithms* **52**(4), 649–662 (2009).
27. Boccaletti, S. et al. The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).
28. Koschutzki, D., Lehmann, K., Peeters, L. & Richter, S. Centrality indices. In *Network Analysis* (eds Brandes, U. & Erlebach, T.) 16–61 (Springer, Berlin, 2005).
29. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
30. AirPassengers. Monthly totals of a US airline passengers from 1949 to 1960 (2020). https:// https://www.kaggle.com/chirag19/air-passengers [accessed: 29/6/20]
31. Lorentz, T.S. Lorentz typical chaotic time-series created by Runge-Kutta integration of the Lorenz equations, on standard values sigma=10.0, r=28.0, and b=8/3. time-series database (2020). http://www.physics.emory.edu/faculty/weeks/research/tseries1.html [accessed: 29/6/20]
32. DEOK.hourly. Duke Energy Ohio/Kentucky (DEOK) estimated energy consumption in Megawatts (MW) (2020). https://www.kaggle.com/robikscube/hourly-energy-consumption?select=DEOK_hourly.csv [accessed: 29/6/20]
33. Wolfer-sunspot-numbers, (2020) "Wolfer sunspot numbers, 1770 to 1771", time-series, time-series database available at the URL: https://www.kaggle.com/dougcresswell/time-series-practice-datasets?select=wolfer-sunspot-numbers-1770-to-1.csv [accessed: 29/6/20].
34. Daily-minimum-temperatures-in-me. Daily minimum temperatures in Melbourne, Australia, 1981–1990", time-series (2020). https://www.coursehero.com/file/26192773/daily-minimum-temperatures-in-mecsvxlsx/ [accessed: 29/6/20]
35. Norusis, M. *SPSS 16.0 Statistical Procedures Companion* (Prentice Hall Publications, New Jersey, USA, 2008).
36. Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. *Probability & Statistics for Engineers & Scientists* 9th edn. (Prentice Hall Publications, New York, 2012).
37. Alligood, K. T., Sauer, T. D. & Yorke, J. A. *Chaos* 105–147 (Springer, New York, 1996).
38. Shumway, R. H. & Stoffer, D. S. *Time Series Analysis and Its Applications, With R Examples* 4th edn. (Springer, Switzerland, 2017).
39. Hanias, M., Tsakonas, S., Magafas, L., Thalassinos, E. I. & Zachilas, L. Deterministic chaos and forecasting in Amazon's share prices. *Equilib. Q. J. Econ. Econ. Policy* **15**(2), 253–273 (2020).
40. Magafas, L., Hanias, M., Tablatou, A. & Konstantaki, P. Non-linear properties of the VIX index. *Int. J. Prod. Manag. Assess. Technol.* **5**(2), 16–24 (2017).
41. Tsiotas, D. Detecting differences in the topology of scale-free networks grown under time-dynamic topological fitness. *Sci. Rep.* **10**(1), 10630 (2020).
42. Bastian, M., Heymann, S., Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the Third International ICWSM Conference*, pp. 361–362 (2009).

Author contributions

D.T. designed research; D.T. and L.M. performed research; D.T., L.M., and P.A. contributed new reagents/analytic tools; D.T., L.M., and P.A. analyzed data; D.T., L.M., and P.A. wrote the manuscript; All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89552-2.

Correspondence and requests for materials should be addressed to D.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.