

# SCIENTIFIC REPORTS



OPEN

## Exact results of the limited penetrable horizontal visibility graph associated to random time series and its application

Minggang Wang<sup>1,2,3</sup>, André L. M. Vilela<sup>3,4</sup>, Ruijin Du<sup>3,5</sup>, Longfeng Zhao<sup>3</sup>, Gaogao Dong<sup>3,5</sup>, Lixin Tian<sup>1,5</sup> & H. Eugene Stanley<sup>3</sup>

The limited penetrable horizontal visibility algorithm is an analysis tool that maps time series into complex networks and is a further development of the horizontal visibility algorithm. This paper presents exact results on the topological properties of the limited penetrable horizontal visibility graph associated with independent and identically distributed (*i.i.d.*) random series. We show that the *i.i.d.* random series maps on a limited penetrable horizontal visibility graph with exponential degree distribution, independent of the probability distribution from which the series was generated. We deduce the exact expressions of mean degree and clustering coefficient, demonstrate the long distance visibility property of the graph and perform numerical simulations to test the accuracy of our theoretical results. We then use the algorithm in several deterministic chaotic series, such as the logistic map, Hénon map, Lorenz system, energy price chaotic system and the real crude oil price. Our results show that the limited penetrable horizontal visibility algorithm is efficient to discriminate chaos from uncorrelated randomness and is able to measure the global evolution characteristics of the real time series.

Several methodologies for understanding the complicated behavior of nonlinear time series have been recently developed, including chaos analysis<sup>1,2</sup>, fractal analysis<sup>3,4</sup>, and complexity measurement<sup>5,6</sup>. With the development of complex network theories<sup>7–10</sup>, a new multidisciplinary methodology for characterizing nonlinear time series using complex network science has emerged and rapidly expanded<sup>11–26</sup>. The main tool of these methods is to use an algorithm to transform a nonlinear time series into a corresponding complex network and then use the topological structure of complex networks to analyze the properties of the nonlinear time series.

Currently there are several ways of converting univariate time series into complex networks. The first one is Pseudo-periodic time series transitions<sup>11</sup> that analyze pseudo-periodic time series. The second one is the visibility graph (VG) method, which was first proposed by Lacasa *et al.*<sup>12,13</sup> and simplified by Luque *et al.*<sup>14,15</sup>. Bezsudnov *et al.*<sup>16</sup> proposed a parameter visibility method and Gao *et al.*<sup>17</sup> also proposed a limited penetrable visibility method (LPVG) and multiscale limited penetrable horizontal visibility graph (MLPHVG). The third one is the phase space reconstruction method<sup>18,19</sup>. It begins with a phase space reconstruction of time series analysis, maps fixed-length time series segments into nodes of a network, and then uses the correlation coefficients (or distances) between these nodes to determine whether they are connected or not. The fourth one is recurrence networks method<sup>20,21</sup>. This method uses the concept of recurrences in phase space and the recurrence matrix of a time series is interpreted as the adjacency matrix of an associated complex network. The links of this network connects different points in time if the considered states are closely neighboured in phase space. And the last one is the coarse graining method<sup>22–25</sup> by which fluctuations of time series are transformed into signal sequences. A fixed-length signal sequence is treated as a network node that connects nodes of time series in chronological order, and a weighted

<sup>1</sup>School of Mathematical Science, Nanjing Normal University, Nanjing, 210042, Jiangsu, China. <sup>2</sup>Department of Mathematics, Nanjing Normal University Taizhou College, Taizhou, 225300, Jiangsu, China. <sup>3</sup>Center for Polymer Studies and Department of Physics, Boston University, Boston, MA, 02215, USA. <sup>4</sup>Universidade de Pernambuco, 50720-001, Recife, PE, Brazil. <sup>5</sup>Energy Development and Environmental Protection Strategy Research Center, Jiangsu University, Zhenjiang, 212013, Jiangsu, China. Correspondence and requests for materials should be addressed to L.T. (email: [tianlx@ujs.edu.cn](mailto:tianlx@ujs.edu.cn))

and directed network is then constructed. In recent years, researchers have also used complex network theories to study multivariate time series<sup>26–29</sup>. These methods all effectively maintain most of the properties of different types of time series, and they have been successfully used in many different fields<sup>30–37</sup>.

Although there have been abundant empirical results obtained using time series complex network algorithms<sup>11–29</sup>, rigorous theoretical results are still lacking. Only a small amount of literature has presented exact results on the properties of the horizontal visibility graphs (HVG) associated with random series<sup>12–15</sup>. Thus far no rigorous theory other than the above algorithms have been developed. Our goal here is to focus on one type of general horizontal visibility algorithm, the limited penetrable horizontal visibility graph (LPHVG). The parameter  $\rho$  of the limited penetrable horizontal visibility graph is called limited penetrable distance, when we set the limited penetrable distance  $\rho = 0$ , the limited penetrable horizontal visibility graph (LPHVG) degenerates into the horizontal visibility graph (HVG)<sup>14</sup>, thus LPHVG is the extended form of HVG. We derive exact results on the properties of the limited penetrable horizontal visibility graphs associated with independent and identically distributed (*i.i.d.*) random series. We prove that an *i.i.d.* random series can be mapped on a limited penetrable horizontal visibility graph with exponential degree distribution, which is an extension of the result presented by Luque *et al.*<sup>14</sup>. We deduce the exact mean degree and the clustering coefficient, and we prove that the limited penetrable horizontal visibility graph associated with any *i.i.d.* random series has small-world characteristics. To verify our theoretical solution, we acquire simulation results by using several deterministic chaotic series and a real crude oil price series that confirms the accuracy and usability of our results.

## Results

Here we show several exact results of LPHVG associated with *i.i.d.* random time series and apply them to the deterministic chaotic series of a logistic map, a Hénon map, a Lorenz system and energy price chaotic system and a crude oil price series.

**Degree distribution.** Let  $X(t)$  be a real valued bi-infinite time series of *i.i.d.* random variables with a probability density  $f(x)$  in which  $x \in [a, b]$ , and consider its associated LPHVG with the limited penetrable distance  $\rho = 1$ . Then

$$P(k) \sim \exp[-(k - 4)\ln(5/4)] \quad \text{with } k = 4, 5, \dots, \forall f(x), \quad (1)$$

where  $P(k)$  is the degree distribution and  $k$  is the degree of a node. To prove this conclusion we first calculate the probability that an arbitrary datum with value  $x_0$  has a limited penetrability at most a one-time visibility of  $k$  other data. We thus list all sets of possible configurations for data  $x_0$  with  $k = 4, 5$  and  $6$  (Figs S1, S2 and S3 in Supplementary Information), and we calculate the probability  $P(k = 4)$ ,  $P(k = 5)$  and  $P(k = 6)$  (Eqs (S4), (S9) and (S10), respectively). We then deduce the rules of when a given configuration contributes to  $P(k)$  (rules i–iv in Supplementary Information) and obtain a general expression for  $P(k)$  [see Eq. (S12)]. The detailed proof of this result is shown in *Supplementary Information Theorem S1*. This is an exact result for a limited penetrable horizontal visibility graph with the limited penetrable distance  $\rho = 1$ . We conclude that for every probability distribution  $f(x)$ , the degree distribution  $P(k)$  of the associated LPHVG has the same exponential form. In addition, from this result we can obtain the more general result (*Theorem S2 in Supplementary Information*) in which  $X(t)$  is a real bi-infinite time series of *i.i.d.* random variables with a probability distribution  $f(x)$  in which  $x \in [a, b]$ , and can examine its associated LPHVG with the limited penetrable distance  $\rho$ . Then

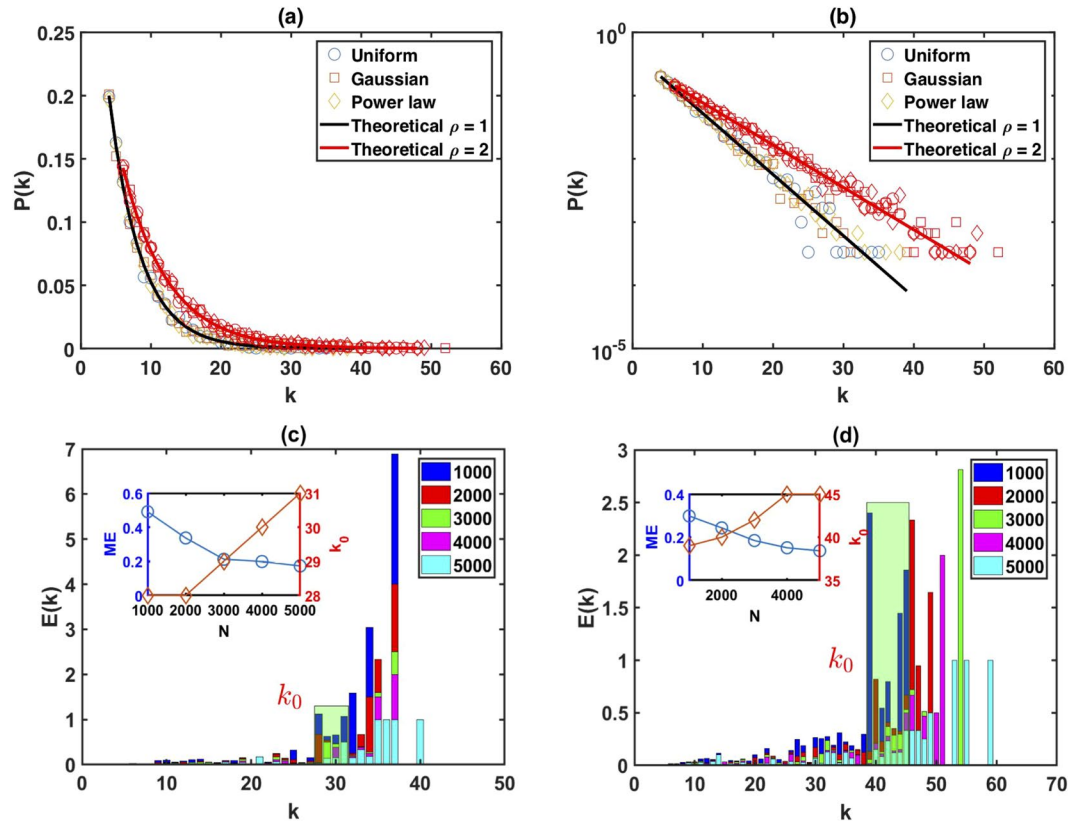
$$P(k) \sim \exp\{-(k - 2\rho - 2)\ln[(2\rho + 3)/(2\rho + 2)]\}, \quad \rho = 0, 1, 2, \dots \quad \text{with} \\ k = 2\rho + 2, 2\rho + 3, \dots, \forall f(x). \quad (2)$$

Note that when  $\rho = 0$ , then  $P(k) \sim \exp[-(k - 2)\ln(3/2)]$ <sup>14</sup>. For this case the LPHVG becomes the HVG (*Methods Section*). When  $\rho = 1$ , the result is Eq. (1). Therefore Eq. (2) is an extension of the previous result<sup>14</sup> that indicates that the degree distribution  $P(k)$  of LPHVG associated with *i.i.d.* random time series has a unified exponential form.

To further check the accuracy of our analytical results, we perform several numerical simulations. We generate a random series of 3000 data points from uniform, Gaussian, and power-law distributions and their associated limited penetrable horizontal visibility graphs. Figure 1(a) and (b) show plots of the degree distributions of the resulting graphs with a penetrable distance  $\rho = 1$  and  $\rho = 2$ . Here circles indicate a series extracted from a uniform distribution, and squares and diamonds indicate series extracted from Gaussian and power-law distributions, respectively. The solid lines indicate the theoretical results of Eq. (2). We find that the theoretical results agree with the numerics. Note that one prerequisite for our theoretical results is that the length of the time series must be infinitely long, i.e. the series size  $N \rightarrow \infty$ , so we can assert that the tail degree distribution of LPHVG associated to *i.i.d.* random series deviated from the theoretical result is only due to the effect of the finite size. In order to check the effect of the finite size, we define the relative error  $E(k)$  and the mean relative error  $ME$  to measure accurate between the numerical result under the finite size and the theoretical result. We also use a cutoff value  $k_0$  to denote the onset of finite size effects. Thus, we write

$$E(k) = \frac{|P_{num}(k) - P_{the}(k)|}{P_{the}(k)} \quad \text{and} \quad ME = \sum_k E(k), \quad (3)$$

where,  $P_{num}(k)$  and  $P_{the}(k)$  denote the degree distribution of the numerical and theoretical results, respectively. We generate the random series from uniform distribution with different series sizes  $N$  and 10 realizations for each size. Figure 1(c) shows the test results of the resulting graphs with penetrable distance  $\rho = 1$  and Fig. 1(d) with  $\rho = 2$ . The subplots in Fig. 1(c) and (d) show the relations between the mean relative error  $ME$  and the series size



**Figure 1.** Numerical simulations of the degree distribution. (a) Degree distribution of the resulting graphs with penetrable distance  $\rho = 1$  and  $\rho = 2$ , (b) semi-log plot of the degree distribution of the resulting graphs with penetrable distance  $\rho = 1$  and  $\rho = 2$ , (c) test results of the resulting graphs with penetrable distance  $\rho = 1$ , (d) the test results of the resulting graphs with penetrable distance  $\rho = 2$ , both averaged over 10 realizations.

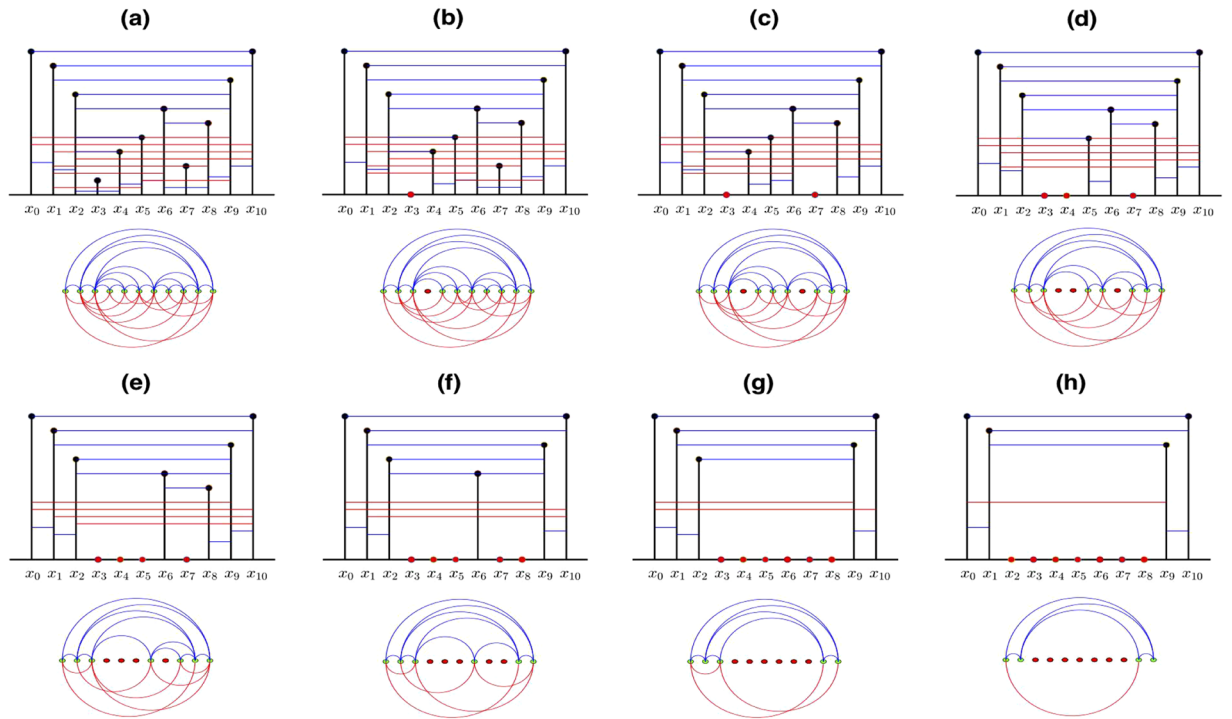
$N$ , and the relations between the cutoff value  $k_0$  and the series size  $N$ . We find that when the finite size  $N$  increases, the cutoff value  $k_0$  also increases but the mean relative error  $ME$  decreases, which is in agreement with our previous assertion.

**Mean degree.** Using Eq. (2) we calculate the mean degree  $\langle k \rangle$  of the LPHVG associated with an uncorrelated random series,

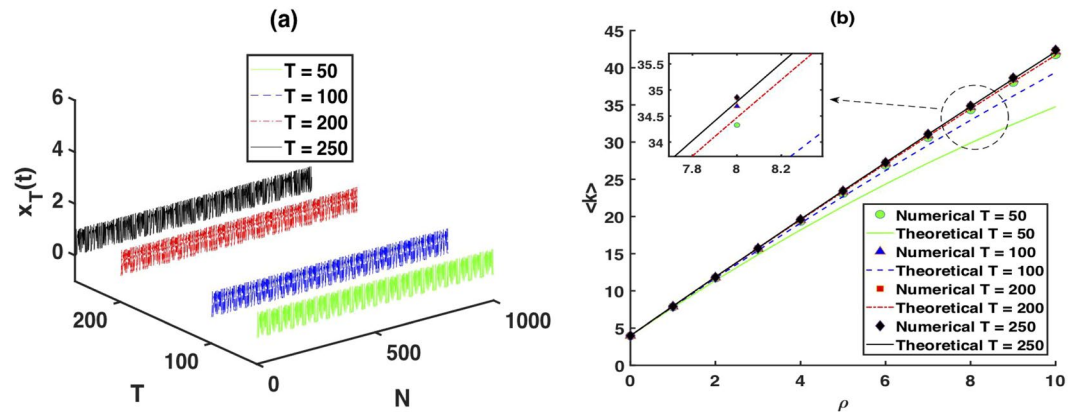
$$\langle k \rangle = \sum_k kP(k) = \sum_{k=2(\rho+1)}^{\infty} \frac{k}{2\rho+3} \left( \frac{2\rho+2}{2\rho+3} \right)^{k-2(\rho+1)} = 4(\rho+1). \tag{4}$$

Next, we deduce a more general expression of mean degree  $\langle k(T) \rangle$ , which depends on the period of a time series  $T$ . We consider an infinite periodic series of period  $T$  denoted by  $X_t = \{ \dots, x_0, x_1, x_2, \dots, x_T, x_1, x_2, \dots \}$ , where  $x_0 = x_T$  and no repeated values in a period. Let  $\rho \ll T$  for the subseries  $\tilde{X}_T = \{ x_0, x_1, x_2, \dots, x_T \}$ . Without loss of generality, we assume  $x_0 = x_T$  corresponds to the largest value of the subseries, and  $x_1, \dots, x_\rho, x_{T-\rho}, \dots, x_{T-1}$  corresponds to the  $(2\rho+1)$ nd largest value of the subseries. We then can construct the LPHVG associated with the subseries  $\tilde{X}_T$ . If the LPHVG has  $E$  links and  $x_i$  is smallest datum of  $\tilde{X}_T$ , the degree of  $x_i$  is  $2(\rho+1)$  during the construction of LPHVG. Figure S1 illustrates the result for  $\rho = 1$ . We delete node  $x_i$  and its  $2(\rho+1)$  links from the LPHVG and the resulting graph has  $E - 2(\rho+1)$  links and  $T$  nodes. We iterate this process  $T - (2\rho+1)$  times and in Fig. 2 we present a graphical illustration for  $\rho = 1$  and  $T = 10$ . The total number of deleted links is now  $E_d = 2(\rho+1)[T - (2\rho+1)]$  and the resulting graph has  $2(\rho+1)$  nodes, i.e.,  $x_0, x_1, \dots, x_\rho, x_{T-\rho}, \dots, x_{T-1}, x_T$  as shown in Fig. 2(h) for  $\rho = 1$  and  $T = 10$ . Because these  $2(\rho+1)$  nodes are connected by  $E_r = \binom{2(\rho+1)}{2}$  links, the mean degree of a limited penetrable horizontal visibility graph associated with  $X_t$  is

$$\begin{aligned} \langle k(T) \rangle &= 2 \left( \frac{E_d + E_r}{T} \right) = \frac{2[2(\rho+1)(T - (2\rho+1)) + (\rho+1)(2\rho+1)]}{T} \\ &= 4(\rho+1) \left( 1 - \frac{2\rho+1}{2T} \right), \quad \rho \ll T. \end{aligned} \tag{5}$$



**Figure 2.** Graphical illustration of the constructive proof of  $\langle k(T) \rangle$  considering a LPHVG with  $\rho=1$  extracted from a periodic series of period  $T=10$ .

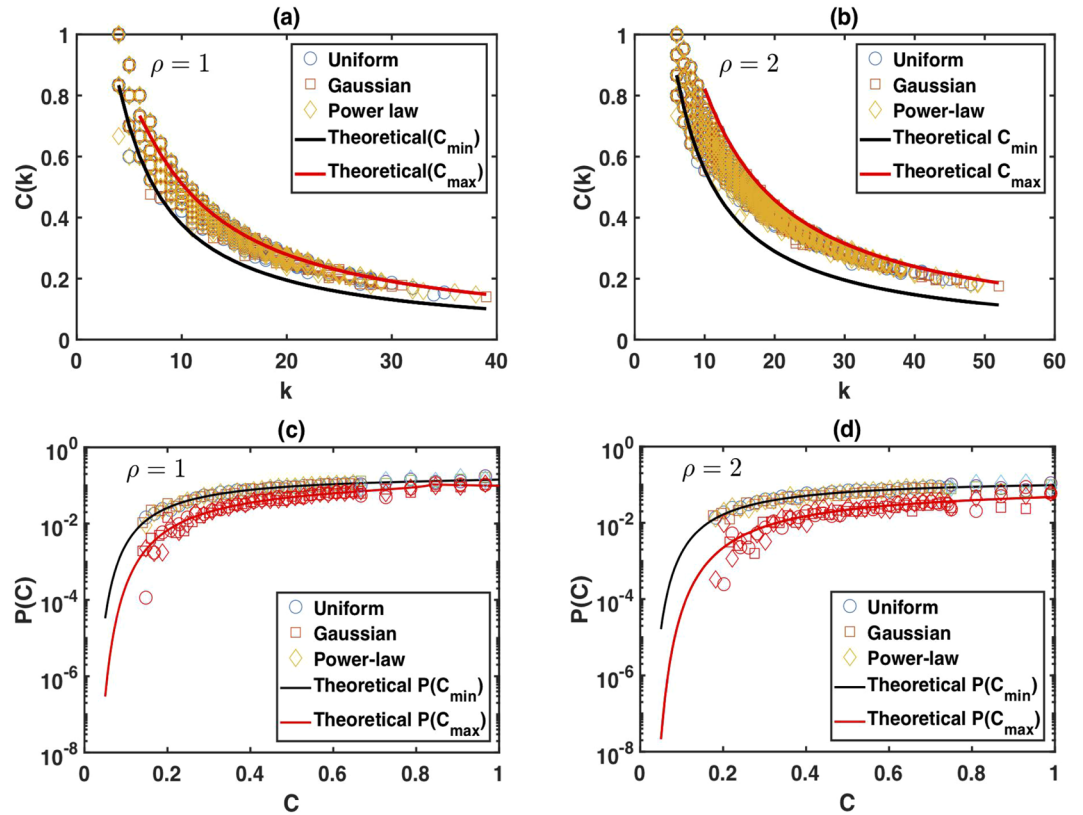


**Figure 3.** Numerical simulations of mean degree. (a) A simplified period-50, period-100, period-200 and period-250 time series of 1000 data. (b) The mean degree of the resulting LPHVGs with different  $\rho$ , where the time series with periodic-50, periodic-100, periodic-200 and periodic-250 are represented by circles, triangles, squares and diamonds, respectively. The solid lines correspond to the theoretical result for each time series.

Note that Eq. (5) holds for every periodic or aperiodic series in which  $T \rightarrow \infty$ . This is independent of the deterministic process that generates the series, since the only constraint in its derivation is that no data repetitions in a period is allowed. Note that one consequence of Eq. (5) is that every time series has an associated LPHVG with the maximum mean degree (achieved for aperiodic series)  $\langle k(\infty) \rangle = 4(\rho + 1)$ , which agrees with Eq. (4).

To check the accuracy of our analytical result, we generate simple period-50, period-100, period-200, and period-250 time series with 1000 data points, as illustrated in Fig. 3(a). We construct the limited penetrable horizontal visibility graphs with the penetrable distance  $\rho=0, 1, 2, \dots, 10$  associated with this periodic time series. Figure 3(b) is a plot of the mean degree of the resulting LPHVGs with different  $\rho$  values, where we found a good agreement with the numeric results for  $\rho \ll T$ .

**Local clustering coefficient.** In LPHVG, nodes with the same degree usually have different clustering coefficients since the degree of the node contribute with different configurations and structures (see proof of Theorem S1).



**Figure 4.** Clustering coefficient of LPHVG versus degree of the nodes for (a)  $\rho = 1$  and (b)  $\rho = 2$ . The theoretical prediction of  $C_{\min}(k)$  is shown in a solid black line and a red line denotes the prediction for  $C_{\max}(k)$ . In (c) and (d) we plot the clustering coefficient distribution and the theoretical prediction lines for  $\rho = 1$  and  $\rho = 2$ , respectively.

By calculating these coefficients for different configurations (Theorem S3), we find that they are irregular, but the minimum and the maximum clustering coefficient for nodes in LPHVG are regular. Therefore, based on results of degree distribution [Eq. (2)], we can deduce the minimum and maximum local clustering coefficient  $C_{\min}(k)$  and  $C_{\max}(k)$  of the LPHVG associated to *i.i.d.* random series. Thus, we write

$$C_{\min}(k) = \frac{2}{k} + \frac{2\rho(k-2)}{k(k-1)}, \quad \rho = 0, 1, 2 \quad \text{and} \quad k \geq 2(\rho + 1). \tag{6}$$

$$C_{\max}(k) = \frac{2}{k} + \frac{4\rho(k-3)}{k(k-1)}, \quad \rho = 0, 1, 2 \quad \text{and} \quad k \geq 2(2\rho + 1). \tag{7}$$

Using Eqs (2, 6) and (7) we also obtain the local clustering coefficient distribution  $P(C_{\min})$  and  $P(C_{\max})$ , i.e.,

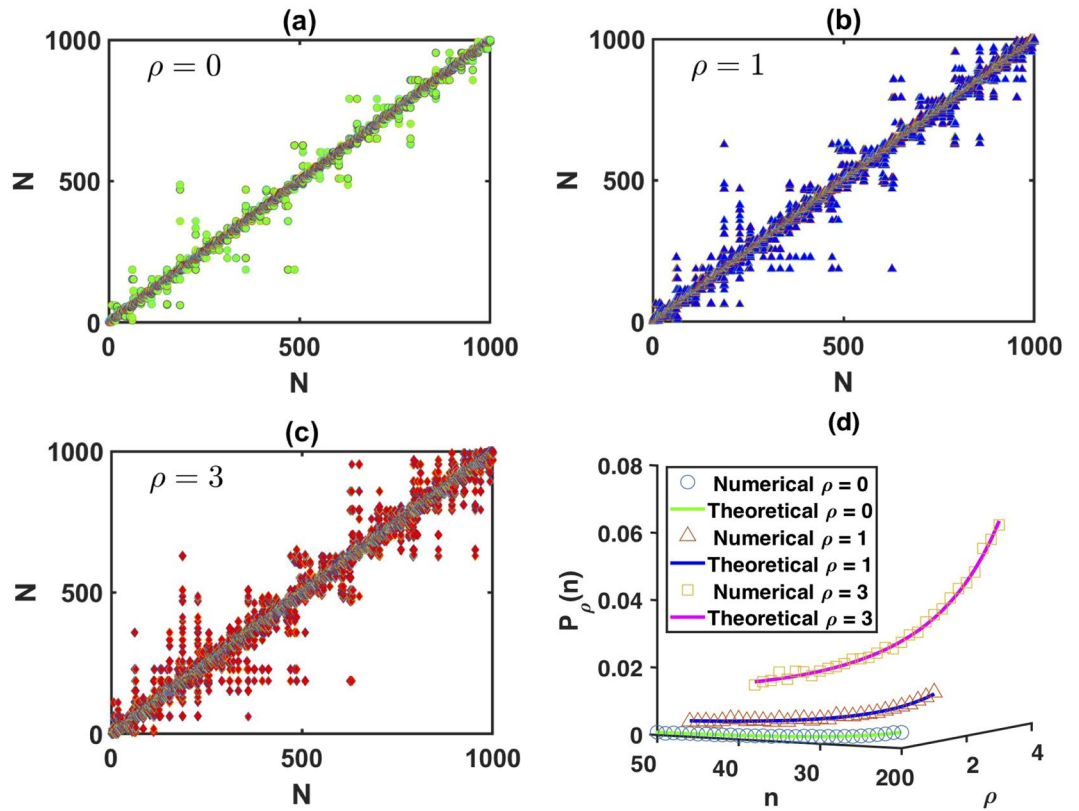
$$P(C_{\min}) = \frac{1}{2\rho + 3} \exp \left\{ \left[ \frac{\varphi + \sqrt{\varphi^2 - 8C_{\min}(2\rho + 1)}}{2C_{\min}} - 2(\rho + 1) \right] \ln \left( \frac{2\rho + 2}{2\rho + 3} \right) \right\},$$

$$\varphi = C_{\min} + 2\rho + 2, \tag{8}$$

$$P(C_{\max}) = \frac{1}{2\rho + 3} \exp \left\{ \left[ \frac{\phi + \sqrt{\phi^2 - 8C_{\max}(6\rho + 1)}}{2C_{\max}} - 2(\rho + 1) \right] \ln \left( \frac{2\rho + 2}{2\rho + 3} \right) \right\},$$

$$\phi = C_{\max} + 4\rho + 2. \tag{9}$$

For a proof of this result see Theorem S3 in the Supplementary Information. Figure 4 shows the clustering coefficient  $C(k)$  and its distribution  $P(C)$  for limited penetrable horizontal visibility graphs associated with different random series of 3000 data points. The solid black and red lines in Fig. 4(a) and (b) are the theoretical prediction of  $C_{\min}(k)$  and  $C_{\max}(k)$ , respectively [see Eqs (6) and (7)]. In Fig. 4(c) and (d) we show a similar analysis for  $P(C_{\min})$  and  $P(C_{\max})$  [see Eqs (8) and (9)].



**Figure 5.** Adjacency matrix of LPHVGs associated to a random series for (a)  $\rho=0$ , (b)  $\rho=1$  and (c)  $\rho=3$ . (d) Plot of the dependence of  $P_\rho(n)$  with  $\rho$  and  $n$  (the solid line corresponds to the theoretical result [Eq. (10)]. Circles, triangles and diamonds correspond to the result of numerical simulations for  $\rho=0$ ,  $\rho=1$  and  $\rho=3$ , respectively.

Figure 4 shows that the theoretical predictions of  $C_{\min}(k)$ ,  $C_{\max}(k)$ ,  $P(C_{\min})$ , and  $P(C_{\max})$  agree with the numerics.

**Long distance visibility.** Consider a limited penetrable horizontal visibility graph associated with a bi-finite sequence of *i.i.d.* random variables, which are extracted from a continuous probability density  $f(x)$ . The probability  $P_\rho(n)$  that two data points are connected, when they are separated by  $n$  intermediate data points, is given by

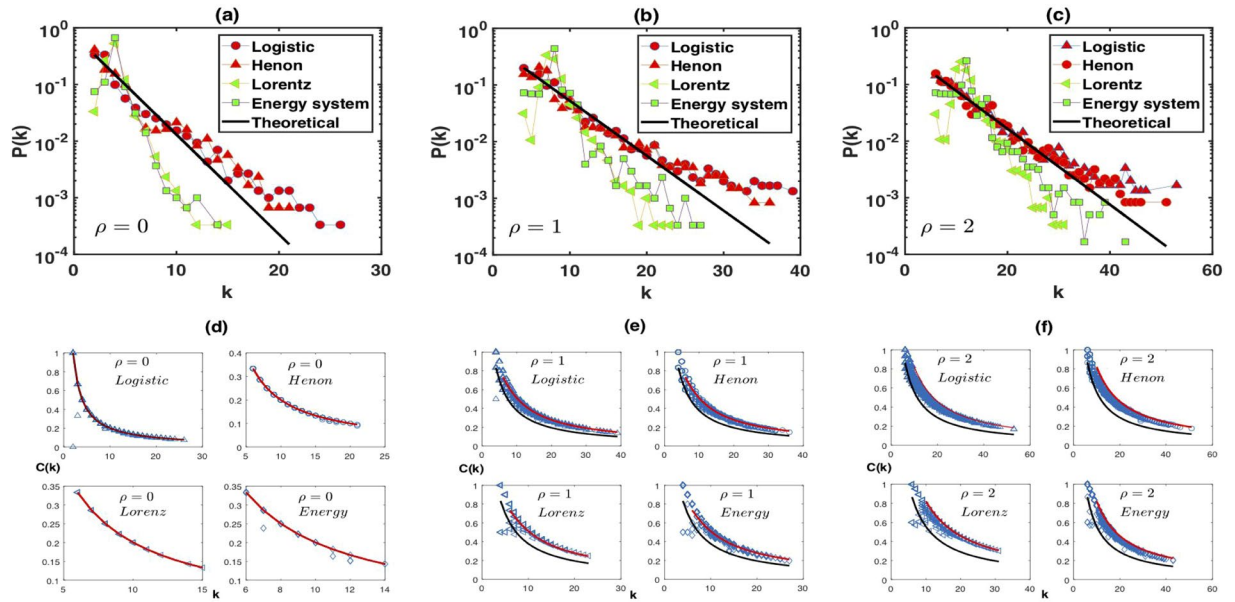
$$P_\rho(n) = \frac{2\rho(\rho + 1) + 2}{n(n + 1)}. \quad (10)$$

Note that  $P_\rho(n)$  is again independent of the probability distribution of the random variable  $x$ . For a detailed proof of this result see Theorem S4 in the Supplementary Information. Figure 5(a,b) and (c) show the adjacency matrix  $A$  of the limited penetrable horizontal visibility graph associated with a random series for different limited penetrable distances. When  $A(i, j) = 1$ , we plot  $\rho=0$  (circles),  $\rho=1$  (triangles), and  $\rho=3$  (diamonds) at each element of the matrix.

As we observed, Fig. 5(a,b) and (c) shows a typical homogeneous structure in which the adjacency matrix is filled around the main diagonal. In addition, the matrix indicates a superposed sparse structure caused by the limited penetrable visibility probability  $P_\rho(n)$ , introducing shortcuts into the limited penetrable horizontal visibility graph. These shortcuts indicate that the limited penetrable horizontal visibility graph present small-world features. Figure 5(b) shows that the theoretical result in Eq. (10) agrees with the numerics and these results are exact, with regard to the topological properties of the LPHVG associated with *i.i.d.* random series.

**Application to deterministic chaotic time series.** These results can be used to discriminate between random and chaotic signals. Methods to analyze random processes and identify its behavior as chaotic or stochastic has received extensive study in recent decades<sup>38–41</sup>. Most previous algorithms have been phenomenological and computationally complicated. Thus, new methods that can reliably distinguish stochastic from chaotic time series are needed. Recently Lacasa *et al.*<sup>13,14</sup> used the horizontal visibility algorithm to characterize and distinguish between chaotic and stochastic processes. Here we use our new theory to distinguish chaotic series from random series and compare with the horizontal visibility algorithm<sup>13,14</sup>.

We address four deterministic time series. First, generated by the Logistic map<sup>42</sup>, we have



**Figure 6.** Numerical results for different chaotic series with semilog plots for the degree distributions of LPHVGs associated to series generated by Logistic map, Hénon map, Lorenz chaotic system and Energy price-supply-economic growth system. In (a) we have  $\rho = 0$ , (b)  $\rho = 1$  and (c)  $\rho = 2$ . In bottom panels we show the relation between degree and clustering coefficient for (d)  $\rho = 0$ , (e)  $\rho = 1$  and (f)  $\rho = 2$ . The solid black and red lines in these latter panels are the theoretical prediction of  $C_{\min}(k)$  and  $C_{\max}(k)$ , respectively.

$$x_{t+1} = \mu x_t(1 - x_t) \quad \text{with} \quad \mu = 4.$$

Second, the Hénon map<sup>43</sup>

$$x_{t+1} = 1 + y_t - ax_t^2, \quad y_{t+1} = bx_t \quad \text{with} \quad a = 1.4 \quad \text{and} \quad b = 0.3.$$

The Lorenz chaos system<sup>44</sup>,

$$\dot{x} = a(y - x), \quad \dot{y} = cx - y - xz, \quad \dot{z} = xy - bz \quad \text{with} \quad a = 10 \quad \text{and} \quad b = 8/3, \quad \text{and} \quad c = 28;$$

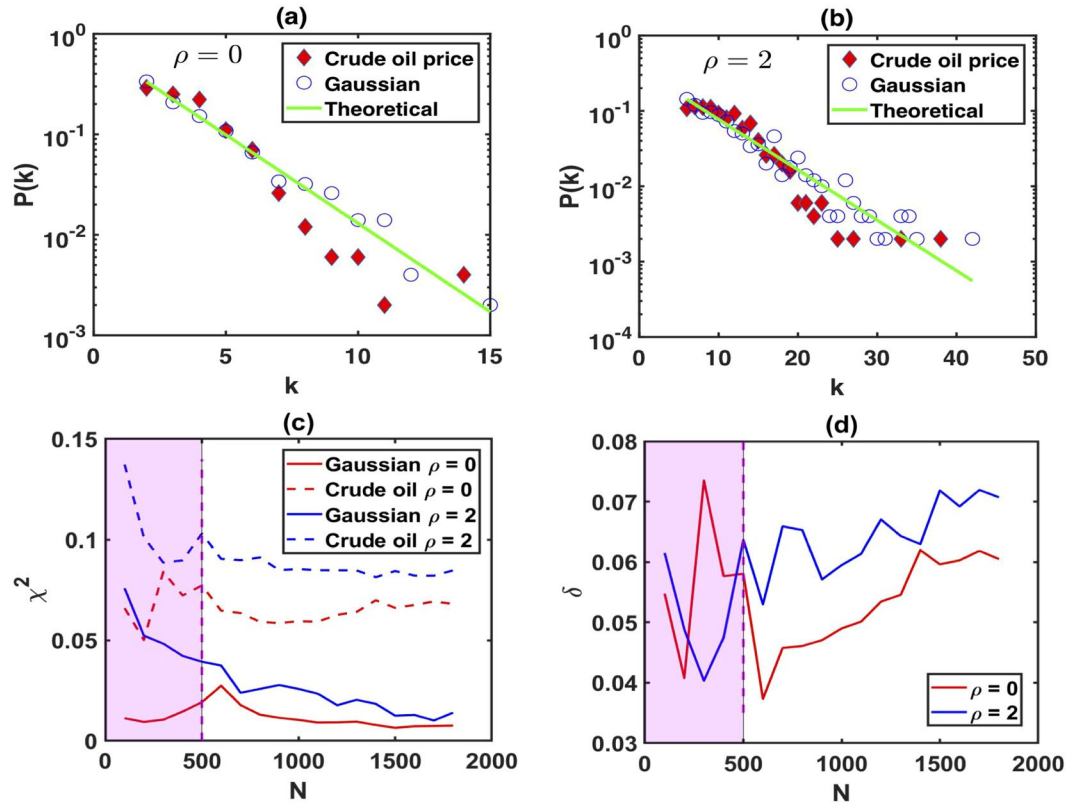
and last, the energy price-supply-economic growth system<sup>34</sup>,

$$\dot{x} = a_1x + a_2(C - y) + a_3(z - K_1), \quad \dot{y} = -b_1y + b_2x - b_3z(1 - z/K_2), \quad \dot{z} = c_1z(1 - z/L) + C_2yz,$$

where  $a_1 = 0.3$ ,  $C = 27$ ,  $a_2 = 0.5563$ ,  $a_3 = 0.15$ ,  $b_1 = 0.4$ ,  $b_2 = 0.6073$ ,  $b_3 = 0.3$ ,  $K_1 = 15$ ,  $K_2 = 15$ ,  $c_1 = 0.3$ ,  $c_2 = 0.006$  and  $L = 19$ .

Figure 6 shows the limited penetrable horizontal visibility graphs of 3000 data points extracted from two different chaotic maps and two different chaotic systems. Here, we show our results for the degree distribution calculated numerically for (a)  $\rho = 0$ , (b)  $\rho = 1$  and (c)  $\rho = 2$ . In the bottom panel, we plot the relation between clustering coefficient and degrees of the nodes. In every case  $P(k)$  deviates from Eq. (2) and  $C(k)$  deviates from Eqs (6) and (7). We also find that the degree distributions of the LPHVGs associated with these chaotic maps and systems can be approximated using the exponential function  $P(k) \sim \exp(-\hat{\lambda})k$ , but  $\hat{\lambda} \neq \lambda = \ln[(2\rho + 3)/(2\rho + 2)]$  in each case. We conjecture that there is a functional relation between the random and chaos dimensions<sup>13,14</sup>. Thus the parameter  $\lambda$  is the frontier between random series and chaotic series and can be used to distinguish randomness from chaos. The solid black and red lines in the bottom panel of Fig. 6 are the theoretical prediction of  $C_{\min}(k)$  and  $C_{\max}(k)$ , respectively [see Eqs (6) and (7)]. In particular, both lines collapse in Fig. 6(d). This result shows that the limited penetrable horizontal visibility graphs for (e)  $\rho = 1$  and (f)  $\rho = 2$  are better discriminators from the perspective of  $C(k)$  than the horizontal visibility graph of  $\rho = 0$  shown in Fig. 6(d).

**Application to real crude oil future price series.** As a further example, we use the data from the U.S. Energy Information Administration on the crude oil future contract 1 (Dollars per Barrel) from April 4, 1983 to March 28, 1985, and we found that they exhibit chaotic and long-range correlations<sup>45,46</sup>. We select 500 sample data points and demonstrate that we can use our method to distinguish chaotic series from random series when the data sample is small (although for theoretical results we need infinite data). Figure 7 shows the results of LPHVG with  $\rho = 0$ , i.e. HVG [see Fig. 7(a)] and LPHVG with  $\rho = 2$  [see Fig. 7(b)] of 500 data points extracted from Gaussian random series and crude oil futures. Here circles indicate a series extracted from a Gaussian distribution and diamonds indicate series extracted from crude oil futures. The solid green line is the theoretical value of  $P(k)$  given by Eq. (2). We can find that the degree distributions associated with Gaussian random series and crude



**Figure 7.** Semilog plot of the degree distributions for (a) HVG associated to crude oil price and Gaussian random series and (b) for the LPHVG with  $\rho = 1$ . (c) Plot of the numerical values of  $\chi^2$  of LPHVGs with  $\rho = 0$  and  $\rho = 2$  associated to crude oil price and Gaussian random series. (d) Plot for numerical values of  $\delta$  of LPHVGs with  $\rho = 0$   $\rho = 2$  for different data sizes.

oil price series both deviate from Eq. (2). From the above analysis, the deviations between the tails of the degree distributions associated with Gaussian random series are caused solely by finite size effects. However, how the degree distributions associated with crude oil price deviate from Gaussian random series is not clear. To quantify the deviations of these two cases, we compute the quantity  $\chi^2$

$$\chi^2 = \sum_k \frac{[P_{\text{num}}(k) - P_{\text{theo}}(k)]^2}{P_{\text{theo}}(k)}, \tag{11}$$

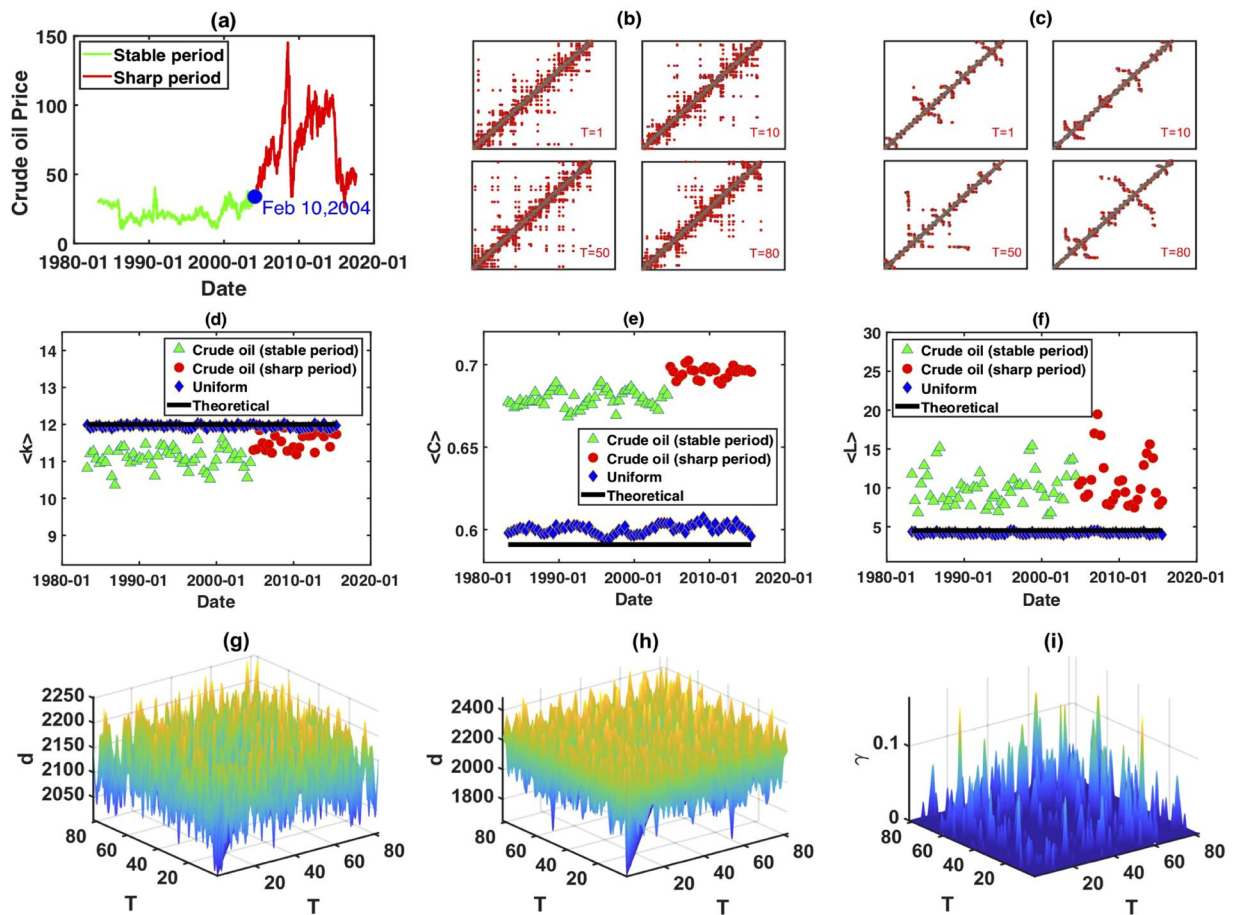
where  $P_{\text{num}}(k)$  is the degree distribution of the numerical result and  $P_{\text{theo}}(k)$  the theoretical result from Eq. (2). Note that  $\chi^2$  changes with the sample size  $N$ . Figure 7(c) shows a plot of the numerical values of  $\chi^2$  of LPHVGs with  $\rho = 0$  and  $\rho = 2$  associated with Gaussian random series and crude oil price series for different data sizes. From Fig. 7(c) we can clearly see that the degree distributions associated with crude oil price deviate from Gaussian random series, which means the crude oil future price sequence is not random but chaotic. Furthermore, we compute the absolute distance  $\delta$

$$\delta = |\chi_{\text{Gauss}}^2 - \chi_{\text{oil}}^2|, \tag{12}$$

where  $\chi_{\text{Gauss}}^2$  and  $\chi_{\text{oil}}^2$  are  $\chi^2$  value for Gaussian random series and the value of crude oil price series, respectively. The larger  $\delta$  means a better distinction between the series. Figure 7(d) shows a plot of the numerical values of  $\delta$  of LPHVGs with  $\rho = 0$  and  $\rho = 2$  of different data size. Comparing the results of HVG ( $\rho = 0$ ) and LPHVG ( $\rho = 2$ ), we find that when  $N \geq 500$  LPHVG works better than HVG to distinguish the real time series from uncorrelated randomness. Therefore, in the next section, we select 500 sample data in each sliding window, i.e.  $L = 500$ , to construct LPHVG for measuring the global evolution characteristics of crude oil price series.

Next we use LPVHG with  $\rho = 2$  to describe the global evolution of crude oil future prices (for the calculation process, see Methods). Our sample data is from the crude oil future contract 1 (in dollars per barrel) from April 4, 1983 to August 15, 2017 [Fig. 8(a)]. As we know, the fluctuations in crude oil future prices differ for several time periods<sup>24,47</sup>. The whole crude oil price time series can be divided into five different fluctuation periods, based on complex network perspective: stable fluctuation period (April 4, 1983 to February 10, 2004), sharp rise period (February 11, 2004 to June 30, 2007), sharp decline period (July 1, 2007 to November 20, 2008), sharp rise period (November 20, 2008 to June 19, 2013) and sharp decline period (June 20, 2013 to August 15, 2017). Here, instead of discussing the fluctuation characteristics of each period, we want to show that the fluctuation behavior





**Figure 8.** (a) Crude oil price series, where the green line represents the stable period and the red line represents the sharp period. (b) evolution of the adjacency matrix of LPHVGs associated to the random series extracted from a uniform distribution. (c) Evolution of the adjacency matrix of LPHVGs associated to crude oil price series. (d) Mean degree of the LPHVGs associated to random series, crude oil price series and the predicted theoretical value. (e) Mean clustering coefficient of the LPHVGs associated to random series, crude oil price series and the theoretical value. (f) Mean path length of the LPHVGs associated to random series, crude oil price series and the theoretical value. (g) Distance distribution of random series. (h) Distance distribution of crude oil price series. (i) Correlation index distribution of crude oil price series.

in different periods can be revealed by the topological structure of LPHVG. Thus, for simplicity, we combine the sharp rise and sharp decline periods into one sharp fluctuation period and we only consider two periods: a stable period from April 4, 1983 to 10 February 2004 and a period of sharp fluctuations from 11 February 2004 to 15 August 2017. Using our calculation method (see the Method section), we establish 82 windows where each window has 100 weeks in the time series (i.e.,  $L = 500$ ), the first of which is from 4 April 1983 to 28 March 1985. Because each window moves 20 weeks to generate the next window (i.e.,  $l = 100$ ), two adjacent windows have overlaps of 80 weeks. This enables information from one window to move to the next in succession. Each window contributes with 500 nodes to build the local limited penetrable visibility graph network. Figure 8(b) and (c) show the evolution of the adjacency matrix of LPHVGs associated with a random series extracted from a uniform distribution and from crude oil price, respectively. Note that adjacent matrices in the random time series and the crude oil price time series significantly differ, but their respective adjacent matrices in different time windows are similar. Figure 8(d–f) show the evolution of the mean degree, mean clustering coefficient and mean path length, respectively. We find that the mean degree, mean clustering coefficient and mean path length of the LPHVG associated with the random series agree with the theoretical values, but these three quantities of LPHVG associated with the crude oil price series do not. The levels of mean degree of the LPHVG associated with the crude oil price series are smaller than the theoretical values, but the mean clustering coefficient and mean path length are larger. They also show different trends in different fluctuation periods. All the quantities plotted present larger values in the sharp fluctuation period than in the stable fluctuation period. Figure 8(g) and (h) show the distance distribution for a random and for crude oil price series (see Eqs (14–15) in Methods). Comparing the results of Fig. 8(g) and (h), we find that the range of distance of crude oil price series is larger than random series, which means the correlation of crude oil price fluctuation presents high complexity. Based on the result of Fig. 8(g) and Eq. (16), we can determine the threshold value  $\theta = 2001$ , then using Eq. (17), we can obtain the correlation

index distribution of crude oil price series. Note that the *i.i.d.* random time series has neither short-range nor long-range correlations, but the crude oil price time series has both. Thus using LPHVG enables us to describe the global time series evolution.

## Discussion

We have introduced a more generalized case of the horizontal visibility algorithm<sup>13,14</sup>, named limited penetrable horizontal visibility algorithm, in which the limited penetrable distance is  $\rho$ . We obtain exact results on several properties of the limited penetrable horizontal visibility graph associated with a general uncorrelated random series, and its reliability has been confirmed by numerical simulations. In particular, the degree distribution of the graph has the exponential form and the expression for the mean degree holds for every periodic or aperiodic series, which does not depend on the deterministic process that generates them. We also found that the clustering coefficient  $C$  has a maximum and a minimum value, which depend on the degree  $k$  and distance  $\rho$ . We concluded that the probability  $P_\rho(n) = \frac{2\rho(\rho+1)+2}{n(n+1)}$  introduces shortcuts to the limited penetrable horizontal visibility graph that exhibit a small-world phenomenon. These results are independent of the distribution from which the series was generated and we observe that all uncorrelated random series have the same limited penetrable horizontal visibility graph and, in particular, the same degree distribution, mean degree, clustering coefficient distribution, and small-world characteristics.

This algorithm can thus be used as a simple test for discriminating uncorrelated randomness from chaos. We show that the method can distinguish between random series that follow the theoretical predictions, and chaotic series that deviate from them. In addition, we employ the method to measure the global evolution characteristics of time series by using LPHVG with  $\rho = 2$ , and the results confirmed its validity. Note that we set  $L = 500$ ,  $l = 100$  as an example to construct the continuous LPHVGs in this paper, moreover, we can set different scales of sliding window  $L$  and step length  $l$  depending on the needs of the analysis. If the goal is to study the global evolution characteristics based on short periods, scales  $L$  and  $l$  can be set to smaller values. On the contrary, the scales  $L$  and  $l$  can be set to larger values. Increasing the scale  $L$  and  $l$ , the mean degree and mean path length in the corresponding complex networks increase, as shown in Fig. 9(a) and (c) and the mean clustering coefficient decreases [Fig. 9(b)]. The parameters  $L$  and  $l$  determine the number of sliding windows, and for crude oil price time series, when  $L = 500$ ,  $l = 100$  the number of sliding windows is 82. This number is only 28 if  $L = 500$ ,  $l = 300$ . We observe that an increase in scale  $L$  and  $l$  will hide the characteristics of diversity of the real time series. On the other hand, for  $L = 50$ ,  $l = 10$  the number of sliding windows is more than 800, and they contain much noise. Thus, if the values of scale  $L$  and  $l$  are too large or too small, it is meaningless for studying the global evolution characteristics of the real time series. When we determine the appropriate parameters  $L$  and  $l$ , the corresponding threshold  $\theta$  can be calculated by using Eqs (14–16) and we can observe the similarly distribution characteristics of correlation index of crude oil price time series, as illustrated in Fig. 9(d–i).

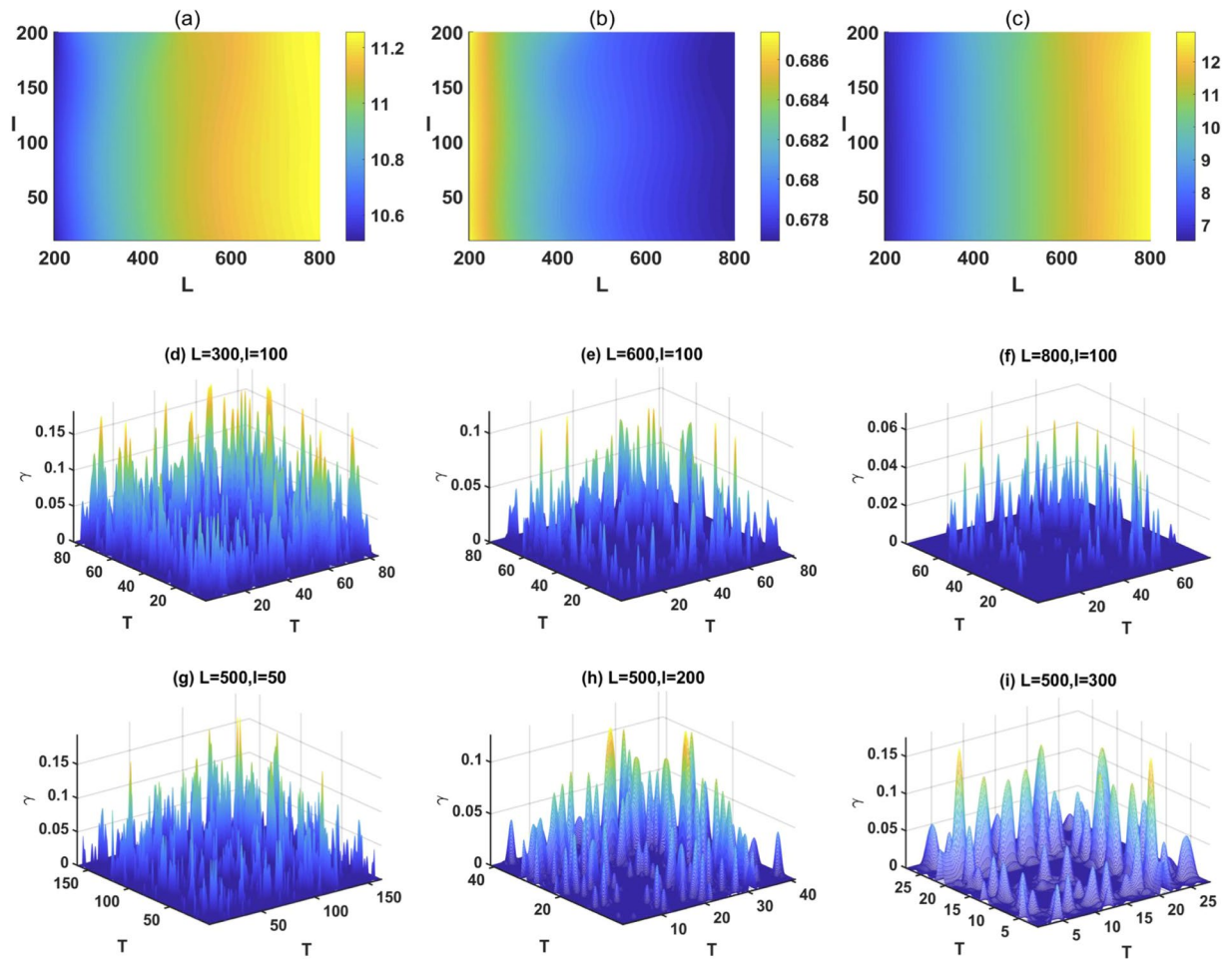
Our exact results presented here are extension of previous work<sup>14</sup> and the method can serve as a preliminary test for locating deterministic fingerprints in time series. To better distinguish chaos from uncorrelated randomness, the limited penetrable parameter  $\rho$  can be adjusted. If we determine that  $P(k)$  has an exponential tail that deviates from the theoretical values or that  $C(k)$  also deviates from the theoretical results [Eqs (2, 6) and (7)], we apply embedding methods to analyze the series. There are also some open problems unsolved in this work, and the main question is how to deduce the theoretical results of LPHVG associated with more complicated time series, such as logistic maps, fractional Brownian motions and Lorenz equations. Another question is how to use this algorithm to measure chaos and reveal the global evolution characteristics for other time series. Therefore, topics for further research could include whether this algorithm is also able to quantify chaos, the relation between such standard chaos indicators as Lyapunov exponents and the correlation dimension. Also, how to tune the limited penetrable parameter  $\rho$  and how to use the limited penetrable horizontal visibility graph to handle two-dimensional manifolds. The topological properties of the visibility graphs (VG), of the limited penetrable visibility graphs (LPVG) and expanded applications of LPHVG might be considered in future works.

## Methods

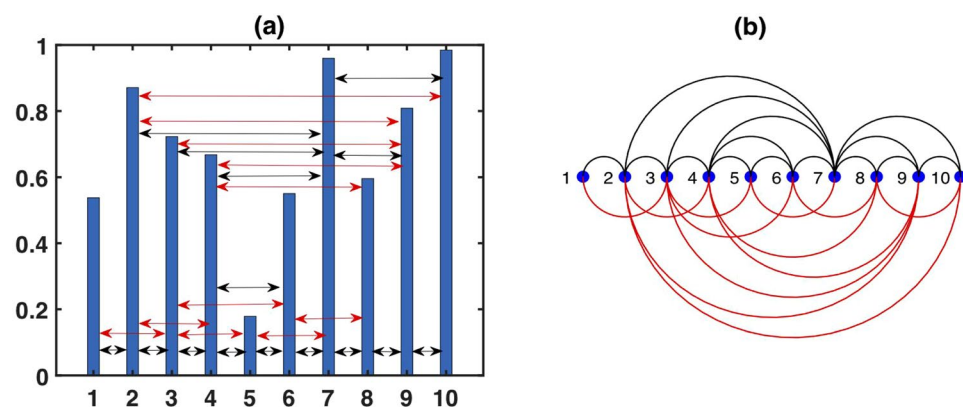
**Limited Penetrable Horizontal Visibility Graph [LPHVG].** The limited penetrable visibility graph (LPVG)<sup>37</sup> and the multiscale limited penetrable horizontal visibility graph (MLPHVG)<sup>17</sup> are a recent extension of the VG<sup>12</sup> and HVG<sup>13,14</sup> used to analyze nonlinear time series. The limited penetrable horizontal visibility graph (LPHVG) is geometrically more simple and an analytically solvable version of LPVG<sup>37</sup> and MLPHVG<sup>17</sup>. To define it we let  $X_i = \{x_i\}$ ,  $i = 1, 2, \dots, N$  be a time series of  $N$  real numbers. If we set the limited penetrable distance to  $\rho$ , LPHVG maps the time series into a graph with  $N$  nodes and an adjacency matrix  $A$ . Nodes  $i$  and  $j$  are connected through an undirected edge ( $A_{ij} = A_{ji} = 1$ ) when  $x_i$  and  $x_j$  have limited penetrable horizontal visibility (see Fig. 10), i.e., if at most  $\rho$ , the intermediate data  $x_q$  is

$$x_q \geq \inf\{x_i, x_j\}, \forall q \in (i, j), \quad (13)$$

this mapping is a limited penetrable horizontal visibility graph (LPHVG). When we set the limited penetrable distance  $\rho = 0$ , LPHVG degenerates into HVG<sup>14</sup>. When  $\rho \neq 0$ , there are more connections between any two nodes in LPHVG than in HVG. Figure 10(b) shows the new established connections (red lines) when we infer the LPHVG on the basis of HVG with a limited penetrable distance  $\rho = 1$ . Note that the limited penetrable horizontal visibility graph of a given time series has all the properties of its horizontal visibility graph, e.g., it is connected and invariant under all affine transformations of the series data<sup>12,14</sup>.



**Figure 9.** Sensitivity analysis on  $L$  and  $l$  for crude oil price series. In (a) we show the mean degree, (b) the mean clustering coefficient, (c) the mean path length. In (d–i) we show the correlation index distributions for different values of scale  $L$  and  $l$ .



**Figure 10.** Example of (a) a time series with 10 data values and (b) its corresponding LPHVG with the limited penetrable distance  $\rho = 1$ , where every node corresponds to a time series data. The limited penetrable horizontal visibility lines between data points define the links connecting nodes in the graph.

**Measurement of the Global Evolution Characteristics of Time Series using LPHVG.** To characterize the evolution of the time series  $X_t = \{x_i, i = 1, 2, \dots, N\}$  using LPHVG, we divide the entire scale of the time series into equal small-scale segments (or windows) and assume that the length of the sliding window is  $L$ . We define  $l$  as the step length between sliding time windows. To ensure that small-scale segments of the time

series are continuous, we require that  $l < L$ . This allows us to obtain  $T = [(N - L)/l + 1]$  for small-scale time windows, where [...] is the rounding function. For every small-scale time window  $t$ , we transform time series into the LPHVG of time  $t$  using the limited penetrable horizontal visibility algorithm. The topological structure of LPHVG changes with time  $t$ , therefore, we write  $LPHVG(t)$ . In order to describe this process, we use the Euclidean distance to measure the relationship between LPHVGs. We define the Euclidean distance between  $LPHVG(t_m)$  and  $LPHVG(t_n)$  to be

$$d[LPHVG(t_m), LPHVG(t_n)] = \sqrt{\sum_{i=1}^L \sum_{j=1}^L (a_{ij}^{(t_m)} - a_{ij}^{(t_n)})^2}, \quad \text{where } a_{ij}^{(t_m)} \in A^{t_m} \text{ and } a_{ij}^{(t_n)} \in A^{t_n}. \quad (14)$$

We then determine the distance matrix

$$D_{T \times T} = \{d_{t_m, t_n}\}_{t_m=1,2,\dots,T, t_n=1,2,\dots,T}, \quad (15)$$

and assign a threshold value to  $\theta$

$$\theta = \min\{d_{t_m, t_n}^{\text{rand}}\}_{t_m \neq t_n}, \quad d_{t_m, t_n}^{\text{rand}} \in D_{T \times T}^{\text{rand}}. \quad (16)$$

Here  $D_{T \times T}^{\text{rand}}$  is the distance matrix associated with the independent and identically distributed random time series. From Eq. (16), we can see that  $\theta$  can be obtained from the distance matrix of LPHVGs associated with the *i.i.d.* time series, which is the critical value to measure the correlation between the data in different time periods. Using the threshold  $\theta$ , we define the correlation index  $\gamma$ ,

$$\gamma_{t_m, t_n} = \begin{cases} 0, & d_{t_m, t_n} \geq \theta, \\ 1 - d_{t_m, t_n}/\theta, & d_{t_m, t_n} < \theta. \end{cases} \quad (17)$$

Here  $\gamma_{t_m, t_n}$  is the correlation degree of LPHVG at time  $t_m$  and time  $t_n$ , and  $\gamma_{t_m, t_n}$  can be visualized using a recursive graph constructed using the formula

$$\mathfrak{R}(t_m, t_n) = \Theta(\theta - d[LPHVG(t_m), LPHVG(t_n)]), \quad \Theta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (18)$$

where  $\Theta(x)$  is the Heaviside function. We use the formula to plot the dependence between  $LPHVG(t_m)$  and  $LPHVG(t_n)$  in two-dimensional coordinates, in which both the abscissa and the ordinate are the time  $t$ . In the recursive graph when the Euclidean distance between  $LPHVG(t_m)$  and  $LPHVG(t_n)$  is sufficiently close, i.e., when  $\mathfrak{R}(t_m, t_n) = 1$ , we plot a dot at  $(t_m, t_n)$  and  $(t_n, t_m)$ . Note that at  $(t_m, t_m)$  and  $(t_n, t_n)$ , i.e., the dots remain in the main diagonal (Fig. 8), and we can use this to characterize the global dynamic changes in correlation.

**Data Availability.** All data generated or analysed during this study are included in this published article.

## References

- Daw, C. S. *et al.* Self-Organization and Chaos in a Fluidized Bed. *Phys. Rev. Lett.* **75**, 2308–2311 (1995).
- Mutua, S., Gu, C. & Yang, H. Visibility graphlet approach to chaotic time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **26**(5), 053107 (2016).
- Peng, C. K. *et al.* Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
- Manshour, P. Complex network approach to fractional time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**(10), 103105 (2015).
- Podobnik, B. & Stanley, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett.* **100**, 084102 (2008).
- Marwan, N. *et al.* Recurrence plots for the analysis of complex systems. *Physics reports* **438**(5–6), 237–329 (2007).
- Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *science* **286**(5439), 509–512 (1999).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *nature* **393**(6684), 440–442 (1998).
- Newman, M. E. J. & Watts, D. J. Renormalization group analysis of the small-world network model. *Physics Letters A* **263**(4), 341–346 (1999).
- Erdős, P. & Rényi, A. On the existence of a factor of degree one of a connected random graph. *Acta Mathematica Hungarica* **17**(3–4), 359–368 (1996).
- Zhang, J. & Small, M. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical review letters* **96**(23), 238701 (2006).
- Lacasa, L. *et al.* From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences* **105**(13), 4972–4975 (2008).
- Lacasa, L. On the degree distribution of horizontal visibility graphs associated with Markov processes and dynamical systems: diagrammatic and variational approaches. *Nonlinearity* **27**(9), 2063 (2014).
- Luque, B. *et al.* Horizontal visibility graphs: Exact results for random time series. *Physical Review E* **80**(4), 046103 (2009).
- Lacasa, L. & Toral, R. Description of stochastic and chaotic series using visibility graphs. *Physical Review E* **82**(3), 036120 (2010).
- Bezsudnov, I. V. & Snarskii, A. A. From the time series to the complex networks: The parametric natural visibility graph. *Physica A: Statistical Mechanics and its Applications* **414**, 53–60 (2014).
- Gao, Z. K. *et al.* Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series. *Scientific Reports* **6**, 35622 (2016).
- Xu, X., Zhang, J. & Small, M. Superfamily phenomena and motifs of networks induced from time series. *Proceedings of the National Academy of Sciences* **105**(50), 19601–19605 (2008).
- Wang, M. & Tian, L. From time series to complex networks: The phase space coarse graining. *Physica A: Statistical Mechanics and its Applications* **461**, 456–468 (2016).
- Marwan, N. *et al.* Complex network approach for recurrence analysis of time series. *Physics Letters A* **373**(46), 4246–4254 (2009).

21. Donner, R. V. *et al.* Recurrence networks—a novel paradigm for nonlinear time series analysis. *New Journal of Physics* **12**(3), 033025 (2010).
22. Gao, X. *et al.* Characteristics of the transmission of autoregressive sub-patterns in financial time series. *Scientific reports*, **4** (2014).
23. Gao, X. *et al.* Transmission of linear regression patterns between time series: From relationship in time series to complex networks. *Physical Review E* **90**(1), 012818 (2014).
24. Wang, M. *et al.* Fluctuation behavior analysis of international crude oil and gasoline price based on complex network perspective. *Applied Energy* **175**, 109–127 (2016).
25. Chen, H. *et al.* Analysis of the Dynamic Evolutionary Behavior of American Heating Oil Spot and Futures Price Fluctuation Networks. *Sustainability* **9**(4), 574 (2017).
26. Gao, Z. K. *et al.* Multiscale complex network for analyzing experimental multivariate time series. *EPL (Europhysics Letters)* **109**(3), 30005 (2015).
27. Gao, Z. K. *et al.* Multilayer Network from Multivariate Time Series for Characterizing Nonlinear Flow Behavior. *International Journal of Bifurcation and Chaos* **27**(04), 1750059 (2017).
28. Lacasa, L., Nicosia, V., Latora, V. Network structure of multivariate time series. *Scientific reports*, **5** (2015).
29. Wang, M., Tian, L. & Du, R. Research on the interaction patterns among the global crude oil import dependency countries: A complex network approach. *Applied Energy* **180**, 779–791 (2016).
30. Marwan, N. How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos* **21**(04), 1003–1017 (2011).
31. Marwan, N. & Meinke, A. Extended recurrence plot analysis and its application to ERP data. *International Journal of Bifurcation and Chaos* **14**(02), 761–771 (2004).
32. Du, R. *et al.* A complex network perspective on interrelations and evolution features of international oil trade, 2002–2013. *Applied Energy* **196**, 142–151 (2017).
33. Wang, M. *et al.* Systemic risk and spatiotemporal dynamics of the consumer market of China. *Physica A: Statistical Mechanics and its Applications* **473**, 188–204 (2017).
34. Wang, M. & Tian, L. Regulating effect of the energy market: Theoretical and empirical analysis based on a novel energy prices–energy supply–economic growth dynamic system. *Applied Energy* **155**, 526–546 (2015).
35. An, H. *et al.* Research on patterns in the fluctuation of the co-movement between crude oil futures and spot prices: A complex network approach. *Applied Energy* **136**, 1067–1075 (2014).
36. Zhou, C. *et al.* Characterizing time series of near-miss accidents in metro construction via complex network theory. *Safety science* **98**, 145–158 (2017).
37. Gao, Z. K. *et al.* Limited penetrable visibility graph from two-phase flow for investigating flow pattern dynamics (2013).
38. Grassberger, P. & Procaccia, I. Characterization of strange attractors. *Physical review letters* **50**(5), 346 (1983).
39. Sugihara, G. *et al.* Distinguishing error from chaos in ecological time series. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **330**(1257), 235–251 (1990).
40. Kaplan, D. T. & Glass, L. Direct test for determinism in a time series. *Physical review letters* **68**(4), 427 (1992).
41. Rosso, O. A. *et al.* Distinguishing noise from chaos. *Physical review letters* **99**(15), 154102 (2007).
42. Kocarev, L. & Jakimoski, G. Logistic map as a block encryption algorithm. *Physics Letters A* **289**(4), 199–206 (2001).
43. Gallas, J. A. C. Structure of the parameter space of the Henon map. *Physical Review Letters* **70**(18), 2714 (1993).
44. Grigorenko, I. & Grigorenko, E. Chaotic dynamics of the fractional Lorenz system. *Physical review letters* **91**(3), 034101 (2003).
45. Panas, E. & Ninni, V. Are oil markets chaotic? A non-linear dynamic analysis. *Energy economics* **22**(5), 549–568 (2000).
46. Tabak, B. M. & Cajueiro, D. O. Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence in prices and volatility. *Energy Economics* **29**(1), 28–36 (2007).
47. Gao, X. *et al.* Detecting method for crude oil price fluctuation mechanism under different periodic time series. *Applied energy* **192**, 201–212 (2017).

## Acknowledgements

The Research was supported by the following foundations: The National Natural Science Foundation of China (71503132, 71690242, 91546118, 11731014, 71403105, 61403171), Qing Lan Project of Jiangsu Province (2017), University Natural Science Foundation of Jiangsu Province (14KJA110001), Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, CNPq, CAPES, FACEPE and UPE.

## Author Contributions

Lixin Tian, Minggang Wang and H. Eugene Stanley designed the structure of the research; Minggang Wang performed the research and prove the Theorem; Minggang Wang, André L.M.Vilela and Ruijin Du designed the model; Minggang Wang and Longfeng Zhao analyzed the data; Minggang Wang, André L.M.Vilela and Gaogao Dong wrote the paper. All co-authors of this paper reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-23388-1>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018