Dual-induced multifractality in online viewing activity

Yu-Hao Qin,[1,2] Zhi-Dan Zhao,[1,3] Shi-Min Cai,[1,3,4,a)] Liang Gao,[2,5,b)] and H. Eugene Stanley[4]

[1]*Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China*
[2]*Institute of Systems Science, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, People's Republic of China*
[3]*Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China*
[4]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[5]*Yinchuan Municipal Bureau of Big Data Management and Service, Yinchuan 750011, People's Republic of China*

Although recent studies have found that the long-term correlations relating to the fat-tailed distribution of inter-event times exist in human activity and that these correlations indicate the presence of fractality, the property of fractality and its origin have not been analyzed. We use both detrended fluctuation analysis and multifractal detrended fluctuation analysis to analyze the time series in online viewing activity separating from Movielens and Netflix. We find long-term correlations at both the individual and communal levels and that the extent of correlation at the individual level is determined by the activity level. These long-term correlations also indicate that there is fractality in the pattern of online viewing. We first find a multifractality that results from the combined effect of the fat-tailed distribution of inter-event times (i.e., the times between successive viewing actions of individuals) and the long-term correlations in online viewing activity and verify this finding using three synthesized series. Therefore, it can be concluded that the multifractality in online viewing activity is caused by both the fat-tailed distribution of inter-event times and the long-term correlations and that this enlarges the generic property of human activity to include not just physical space but also cyberspace. *Published by AIP Publishing.* https://doi.org/10.1063/1.5003100

To better understand the long-term correlations and multifractality in human activity, we analyze the time series in online viewing activity at both the individual and communal levels via famous detrended fluctuation analysis (DFA) and multifractal detrended fluctuation analysis (MFDFA) methods. We find that the long-term correlations at both the individual and communal levels are generic to human activity, and at the individual level, the extent of correlation is determined by the activity level. These long-term correlations suggest the fractal pattern in online viewing activity. We further find a multifractality that results from the combined effect of the fat-tailed distribution of inter-event times (i.e., the times between successive viewing actions of individuals) and the long-term correlations, which is verified by using synthesized series and surrogate methods. These empirical results enlarge this generic property of human activity to include not just physical space but also cyberspace.

I. INTRODUCTION

It is difficult to characterize and understand complex systems because splitting a complex system into simpler subsystems changes its dynamical properties.[1] Thus, researchers focus on macroscopic properties, e.g., analyzing a time series in which the behavioral evolution of a complex system is characterized by output records restricted by time scale. Output records from real-world complex systems, e.g., stock price fluctuations,[2] heart rate variations,[3–7] and inter-spike intervals,[8–10] usually follow a non-Gaussian probability density function (PDF) and involve fractal dynamics.

Human activity is itself a complex system. Time series analysis, e.g., detrended fluctuation analysis (DFA),[11–13] has recently discovered several macroscopic properties in human activity. For example, a periodic pattern has been found in such human activities as Internet surfing,[14] online game log-ins,[15] task submissions to a Linux server,[16] and e-commerce purchases.[17] Long-term correlations found in many physical, biological, economic, and ecological systems[3,18–26] have also been found in human interactive patterns, and these long-term correlations become strong as the human activity level increases.[27]

Rybski *et al.*[28,29] investigated human communication in a social network, found a relationship between long-term correlations and inter-event clustering, and provided a model to explain these correlations. They show that at the individual level, the long-term correlations in a time series of events are determined by the power-law distribution of the inter-event times (the "Levy correlations" in Ref. 29) but that at the communal level, they are a generic property of the system caused by interdependencies among community activities. Zhao *et al.*[30] analyzed the time series of inter-event times in online viewing activity, found that the long-term correlations (i.e., memory) are restricted by the activity level, and showed an abnormal scaling behavior associated with

a)Electronic mail: shimin.cai81@gmail.com
b)Electronic mail: lianggao@bjtu.edu.cn

long-term anticorrelations caused by the bimodal distribution of inter-event times. Kivelä and Porter[31] systematically estimated the inter-event time distribution from finite observation periods in communication networks.

Although these long-term correlations imply the existence of fractality in these time series of human activity, two unanswered questions remain: (i) what category of fractality applies in a time series of human activity and (ii) what is the origin of the fractality? Using the Internet technology, we examine the time series in the human activity of two movie viewing websites, Movielens and Netflix, analyze the long-term correlations, and find fractality. At the individual level, we apply DFA to the time series of records composed by users with the same activity level and to the corresponding shuffled time series in which each user preserves the inter-event times. Long-term correlations become strong as the activity level increases. Because the distributions of inter-event times at different activity levels do not follow a power law, there is a trivial difference between the Hurst exponents[32,33] of the original and the shuffled time series. The empirical result differs somewhat from that found in human communication activity.[29] At the communal level, we similarly analyze the time series of records aggregated from all user activities and find a stronger long-term correlation with Hurst exponents, approximately 0.9 and 1.0 for Movielens and Netflix, respectively.

To more accurately categorize the fractality and understand its origin, we use multifractal detrended fluctuation analysis (MFDFA) and probe the singularity spectrum. We find a dependence between the generalized Hurst exponent and q-order statistical moments which indicates multifractality in the time series of records at the communal level. Although multifractality remains after the time series are shuffled, changes occur in the value of the generalized Hurst exponent. One result is obviously suggested by the singularity spectrum. Reference 29 hypothesizes that multifractality relates to the broad PDF of inter-event times and the long-term correlations.[34] Because this hypothesis is verified by our empirical results and our synthesized series, we conclude that multifractality exists in human viewing activity and that the combined effect of the fat-tailed distribution of inter-event times and long-term correlations causes such multifractality.

II. DATA

The experimental datasets, released by Movielens and Netflix, record views and ratings to movies at a given time. Each user's account is anonymous with a hash tag done by Movielens and Netflix. The total users are approximately 7000 for Movielens and 17 774 for Netflix, respectively. The users' activity levels are hierarchically distributed both in Movielens and Netflix, as shown in Fig. 1. We filter datasets according to user's activity level, $M \geq 55$ (see the definition in Sec. III). Herein, we set $M \geq 55$ to guarantee that there are abundant records for converting into time series in a long duration. We finally obtain 26 884 users (38.4% of the total users) and 10 000 054 records for a duration of 4703 days (nearly 13 years) for Movielens and 17 703 users (99.6% of the total users) and 100 477 917 records for a duration of
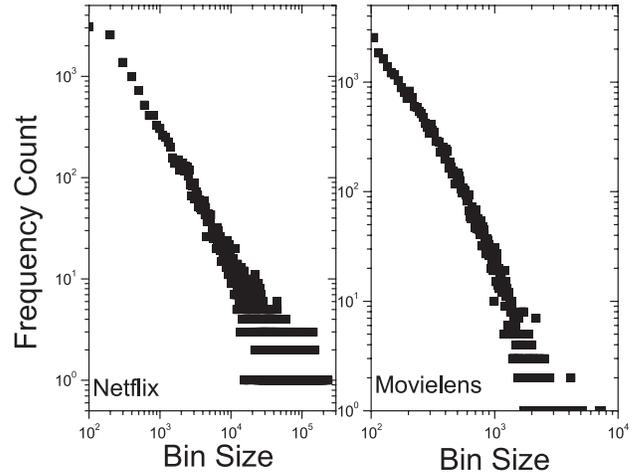

FIG. 1. The PDFs of user activity levels for Netflix and Movielens at a log-log scale. They both show a fat-tailed distribution, suggesting the hierarchical user activities.

2243 days (nearly 6 years) for Netflix. Although the Movielens records begin at its creation date and are thus noisy, the size of both filtered user datasets exceeds 10^5.

To convert these records into time series, we introduce two variables: the records per day of a single user $x(t)$ and the records per day of the entire community $x_{\text{tot}}(t)$. Note that the number of records quantifies the number of movies viewed by a single user $x(t)$, which is constrained by the number of hours in a day. We use these time series in our subsequent analysis. Figures 2(a) and 2(b) show the viewing actions of two typical users when they comment movies in Movielens and Netflix, respectively. Figures 2(c) and 2(d) show the corresponding time series at the individual level. Figures 2(e) and 2(f) show the time series at the communal level. In Fig. 2, we can also observe the clusters of activity records that suggest the burstiness occurring in these online viewing activities.
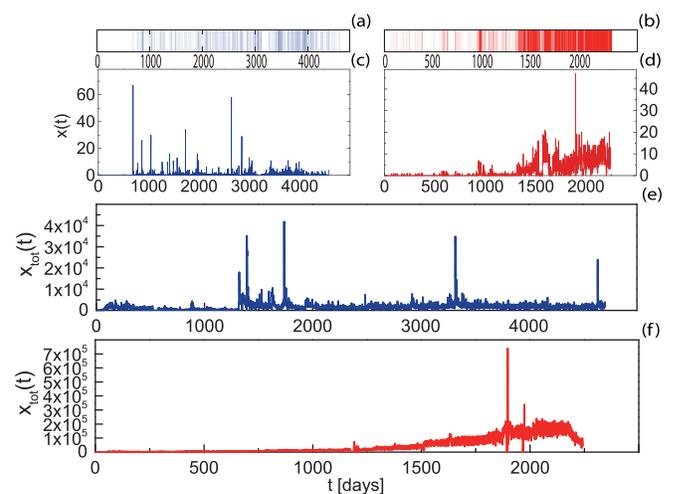

FIG. 2. A visual illustration of activity records of typical users and corresponding time series at both individual and communal levels. (a) and (b) The viewing actions of two typical users, user23172 ($M = 1308$) and user12228 ($M = 5606$), when they comment movies in the websites. (c) and (d) Corresponding time series for these two typical users. (e) and (f) Time series for the whole community. The dark blue and red lines denote Movielens and Netflix, respectively.

III. METHOD

A. Detrended fluctuation analysis

The DFA is a proven method for measuring long-term correlations of time series[11–13] and is less sensitive to the additional detrending process,[35,36] nonstationarities,[37] noise,[38] and missing data.[39,40] To keep our description self-contained, we briefly introduce the steps of this method as follows:

(i) Calculate the profile $Y(t')$ of time series $x(t)$

$$Y(t') = \sum_{t=1}^{t'} x(t) - \langle x(t) \rangle, \quad t' = 1, \dots, N. \quad (1)$$

(ii) Divide $Y(t')$ into N_s non-overlapping segments of length s in increasing order. Because N is often not equal to the product of s and N_s [i.e., $(N_s = \lfloor \frac{N}{s} \rfloor)$], suggesting that the last segment of $Y(t')$ is overlooked, we divide $Y(t')$ from the opposite direction in order to incorporate the entire $Y(t')$. Thus, there are $2N_s$ different segments. We sample the value of s from the logarithmic space, $s = \frac{N}{2^{int(\log_2^N - 2)}}, \dots, \frac{N}{2^3}, \frac{N}{2^2}$, which maintains the smoothness of the curve between $F(s)$ and s.

(iii) Given s, the profile $Y(t')$ in each segment is detrended separately. The least-squares fit is used to determine the χ^2-functions for each segment, such as for $v = 1, 2, \dots, N_s$,

$$F^2(v, s) = \frac{1}{s} \sum_{j=1}^{s} \left[Y((v-1)s + j) - \omega_v^n(j) \right]^2, \quad (2)$$

and for $v = N_s + 1, \dots, 2N_s$,

$$F^2(v, s) = \frac{1}{s} \sum_{j=1}^{s} \left[Y((N - v - N_s)s + j) - \omega_v^n(j) \right]^2, \quad (3)$$

where w_v^n is the n-order polynomial fitting of segment v.

(iv) Calculate the fluctuation function

$$F(s) = \left[\frac{1}{2N_s} \sum_{v=1}^{2N_s} F^2(v, s) \right]^{\frac{1}{2}} \sim s^H, \quad (4)$$

where H is the Hurst exponent. The value of H measures the long-term correlation of time series. It indicates a long-term anticorrelation for $0 < H < 0.5$, no correlation for $H = 0.5$, and a long-term correlation for $H > 0.5$.

B. Multifractal detrended fluctuation analysis

DFA gives us the long-term correlations of time series, which indicates its fractality. To further analyze this fractality and its origin, we modify DFA and introduce MFDFA,[5–7,34,41–43] where Eq. (4) is modified to become

$$F(s) = \left[\frac{1}{2N_s} \sum_{v=1}^{2N_s} \left[F^2(v, s) \right]^{q/2} \right]^{\frac{1}{q}} \sim s^{H(q)}. \quad (5)$$

Here, $H(q)$ is the generalized Hurst exponent. When a time series is monofractal, $H(q)$ is independent of q, and

when it is multifractal, $H(q)$ is dependent on q. This multifractality is caused by such key factors as long-term correlations and PDF of inter-event times. To determine the origin of the multifractality, we randomly shuffle the time series to reduce long-term correlations but preserve the PDF and once again apply MFDFA. If the PDF is the only source of the multifractality, it will be reserved in the shuffled time series. If the long-term correlations are the only source, it will disappear. If both the long-term correlations and PDF affect the time series, the multifractality will remain, but the value of $H(q)$ will change.

The singularity spectrum $f(\alpha)$ provides a clearer way of characterizing a multifractal time series. The horizontal span of $f(\alpha)$ quantifies multifractality. A narrow $f(\alpha)$ indicates a monofractal time series, and a wide $f(\alpha)$ indicates a multifractal time series. To determine its analytical relationship with α, we introduce Renyi exponent $\tau(q)$[44,45] using the equation

$$\tau(q) = qH(q) + 1. \quad (6)$$

Applying the Legendre transformation[46] gives us the relationship between $f(\alpha)$ and α

$$\alpha = \tau'(q) \text{ and } f(\alpha) = q\alpha - \tau(q), \quad (7)$$

or equivalently [using Eq. (6)],

$$\alpha = H(q) + qH'(q) \quad \text{and} \quad f(\alpha) = q[\alpha - H(q)] + 1. \quad (8)$$

IV. RESULTS

A. Long-term correlation in individual activity

To sort users by the activity level, we define M_i to be the total records of a single user i ($M_i = \sum_{i=1}^{N} x_i(t)$, where N is the length of the series) and convert M_i into a logarithmic scale, $L_i = \lfloor (\ln M_i) \rfloor$. Here, the range of L is from 4 to 8 in Movielens and from 4 to 12 in Netflix.

According to logarithmic activity levels, we first present the distributions of inter-event times in Fig. 3, where the left and right panels indicate Movielens and Netflix, respectively. As shown in Fig. 3, both of them show fat tails, which suggests the burstiness occurring in online viewing activity.[47]
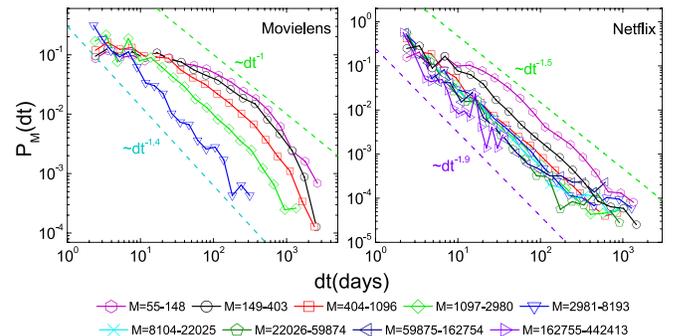

FIG. 3. Inter-event time distribution at different activity levels. The left and right panels indicate Movielens and Netflix, respectively. The dashed lines are the guide for power-law distributions. The fat tails are found in inter-event time distributions for both Movielens and Netflix, which suggests the burstiness of online viewing activity.
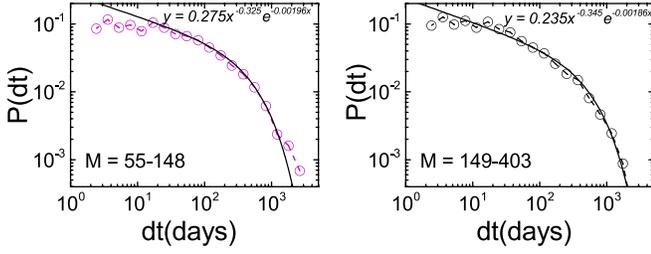
FIG. 4. Inter-event time distributions of users in Movielens whose activity levels are $L = 4$ and $L = 5$, respectively. Compared with exact power-law distributions, they can be fitted by exponential cut-off power-law distributions via the least squares estimating method.

More concretely, for these users with lower activity levels (e.g., $L < 6$), their inter-event times are not exactly power law distributed. For example, in Fig. 4, the inter-event times of users with activity levels $L = 4$ and $L = 5$ in Movielens apparently follow exponential cut-off power-law distributions via the least squares estimating method, while for these users with larger activity levels (e.g., $L > 7$), their distributions of inter-event times are the approximate power law. Thus, the power-law distribution is not the only type to characterize the fat tail of inter-event clustering (i.e., burstiness), and this differs somewhat from the empirical data in human communication[29] and stock trading.[48,49]

Because the burstiness of online reviewing activity (or fat tailed inter-event time distribution) potentially suggests that the time series of records have long-term correlations,[29] we use DFA to calculate the Hurst exponent in each time series of a single user. Note that the least squares estimating method is applied for the fitting trend, and the F-statistic test confirms the significance of fitting results (see more in Appendix B). These Hurst exponents are then scaled according to the user activity level and averaged. Figure 5 shows that the average Hurst exponents as a function of activity levels are greater than 0.5, and they are not strictly restricted

by the order of DFA. Thus, it can be claimed that the long-term correlations exist in these time series of records at the individual level. It is also worthy to be noted that there is an approximately positive relationship between the Hurst exponent and the activity level both for Movielens and Netflix, similar to that in the traded stock market and communication activity.[29,50,51] In addition, there is also a trivially different extent of the long-term correlation between Movielens and Netflix. We assume that this difference is to some extent caused by diverse individual activity patterns in Movielens and Netflix. The commercial website, Netflix, more easily urges users to form the cluster of consecutively occurred viewing actions and enhance long-term correlations.

We have found that both the long-term correlations and fat-tailed inter-event time distribution exist in online viewing activity from Netfilx and Movielens. To further analyze the relationship between the long-term correlation and inter-event time distribution, we shuffle the time series of records but preserve the distribution of inter-event times for each user. The procedure is shown as follows: (i) extract inter-event times of each user; (ii) shuffle the extracted data; and (iii) keep the first time stamp constant and rebuild the time series of records using the shuffled data.

We reuse the DFA to obtain the Hurst exponents of the new time series. Figure 5 shows that they differ only trivially from those of original data, indicating that at the individual level, the long-term correlations of time series of records are associated with the fat-tailed distribution of the inter-event times. Because the inter-event times are not strictly power-law distributed, we cannot infer the long-term correlations from a Levy correlation.[29]

B. Long-term correlation in community activity

The inter-event time distributions with respect to activity levels have a fat tail, but we still must determine whether this property is maintained throughout the community (i.e., the entire system). Figure 6 shows the inter-event time distributions of Movielens and Netflix at the communal level. It can be seen that it is fitted by an exponential cut-off power law distribution for Movielens and the approximate power law one for Netflix, which suggests that the fat tail is generic to the system. Furthermore, we aggregate the records from all users in the community, and investigate the resulting time series to quantify the long-term correlations. Figure 7 shows that although the fluctuation functions are somewhat affected by the oscillations associated with periodic patterns of
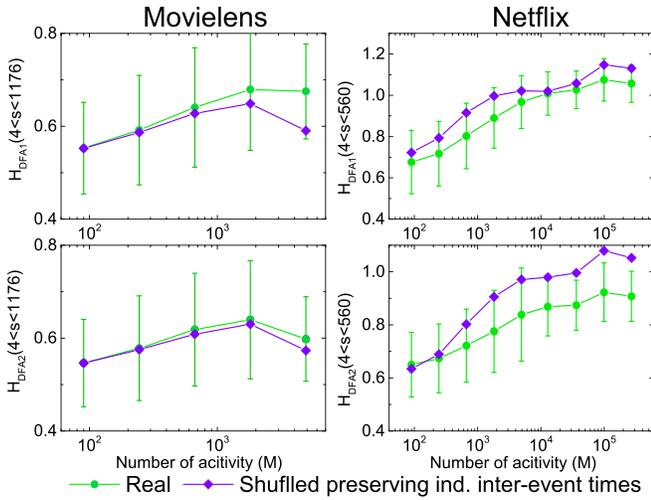

FIG. 5. Average Hurst exponents as a function of the activity level for Movielens and Netflix. The results obtained from the original time series of records and shuffled ones are plotted with green circles and blue squares, respectively. With the increase in activity levels, the long-term correlations become stronger. Moreover, the trivial difference between them reveals the long-term correlations having a potential relation with fat-tailed inter-event time distribution. The error bar is the deviation of Hurst exponents from users at the same activity level.
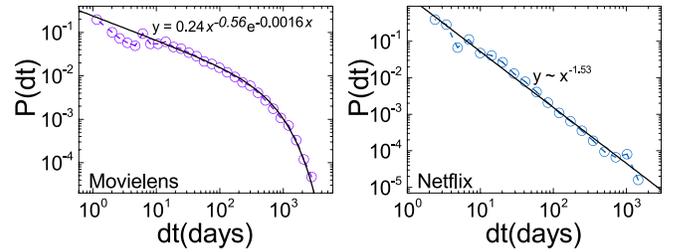

FIG. 6. Inter-event time distribution at the communal level. The power-law with an exponential cut-off relation behaves in Movielens, while the power-law relation behaves in Netflix. This result shows that the burstiness is generic to the system.
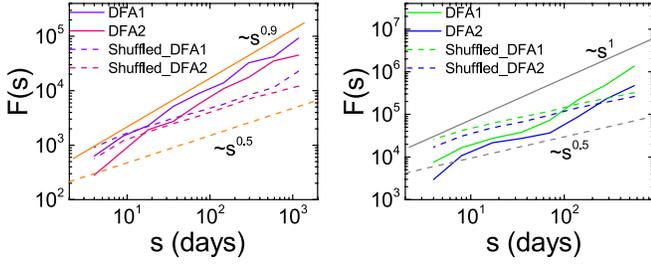
FIG. 7. The results of 1-order and 2-order DFA for Movielens and Netflix at the communal level. The Hurst exponents of the original time series of records obtained via the least squares estimating method are 0.9 and 1 in Movielens and Netflix, respectively. When they are randomly shuffled, the Hurst exponents approximately reduce to 0.5. This result demonstrates that strongly long-term correlations exist in both Movielens and Netflix. Note that solid and dashed lines indicate the original time series of records and shuffled one, respectively.
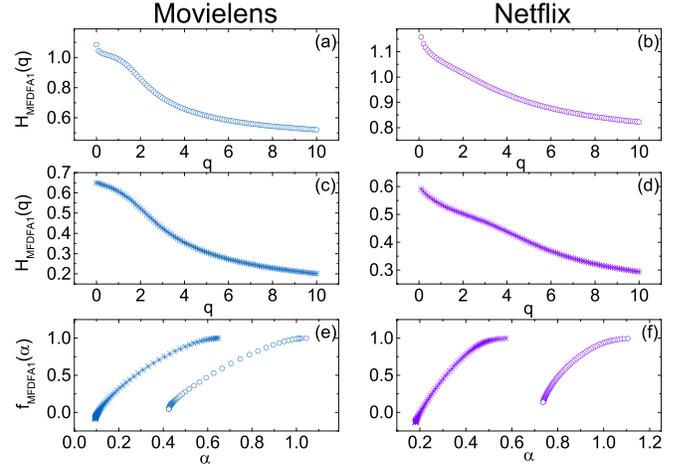

FIG. 8. Relationship between $H(q)$ and q deriving from 1-order MFDFA (a)–(d) and the corresponding singularity spectrum (e) and (f), where (a) and (b) are obtained from the original time series, while (c) and (d) are obtained from the shuffled ones. Although the multifractality remains, there are significant changes that happened to the values of H and α. This result reveals the existence of multifractality for Netflix and Movielens and its formation due to the combined effect of the long-term correlations and the broad PDF of inter-event times.

activity, the Hurst exponents we obtain from 1-order and 2-order DFA are robust and approximately 0.9 and 1 for Movielens and Netflix, respectively. They exhibit strong long-term correlations and are also associated with a time series spectrum with $\frac{1}{f}$ scaling, suggesting that there is self-organized criticality in the system. We shuffle these time series of the entire community to preserve the distribution of inter-event times and find that the Hurst exponents reduce to 0.5, which suggests that the long-term systemic correlation is due to the interdependence (which Ref. 29 calls "true correlation").

C. Multifractality in community activity

The long-term correlations in online viewing activity of separate individuals and of the entire community indicate the presence of fractality, but little research has analyzed the type of fractality involved or its origin. The results of DFA at the community level fluctuating at double logarithmic coordination indicate possible multifractality (see Fig. 7). Inspired by Ref. 34, we introduce MFDFA and analyze the datasets to determine whether the fractality is monofractal or multifractal.

Using 1-order MFDFA, we fix a certain value of q and fit $F_q(s)$ and s at double logarithmic coordination with the least squares estimating method to obtain the value of generalized Hurts exponent $H(q)$. Herein, we set q in an interval $(0, 10]$ with a step length 0.1. Figures 8(a) and 8(b) show $H(q)$ as a function of q via 1-order MFDFA for Movielens and Netflix, respectively. We found that both for Movielens and Netflix, $H(q)$ decreases as q increases, i.e., the dependence between $H(q)$ and q suggests multifractality in community activity.

To determine the origin of such multifractality, we randomly shuffle the time series of records at the communal level by preserving the inter-event time distribution and applying 1-order MFDFA once again on the shuffled one. Figures 8(c) and 8(d) show that although $H(q)$ for both Movielens and Netflix is clearly smaller than the original, the dependence between $H(q)$ and q remains and multifractality is still present.

Much more legible results describing the extent of multifractality in online viewing activity for Movielens and Netflix are characterized by the singularity spectrum $f(\alpha)$, as shown in Figs. 8(e) and 8(f). Note that the horizon span of $f(\alpha)$ both for the original and the shuffled time series is trivially different, which is suggested by the difference of the

asymptotical values of $H(q)$, $\Delta\alpha = 1.38$ (original) and $\Delta\alpha = 1.05$ (shuffled) for Movielens and $\Delta\alpha = 0.78$ (original) and $\Delta\alpha = 0.64$ (shuffled) for Netflix. The more large changes happen to the values of α. We derive these results using the relationship between $H(q)$ and q described in Refs. 52 and 53. It confirms the dependence between $H(q)$ and q and also suggests that the multifractality in community activity is not solely induced by the long-term correlations (see in Appendix C).

Our analysis leads us to assume that the multifractality in online viewing activity from Netflix and Movielens is induced by the combined effect of the long-term correlations and the broad PDF of inter-event times. To verify our hypothesis, we analyze the multifractality of three synthetic time series that are analogous to real time series. The first is a random series that obeys a power-law distribution $(y \sim x^{-2})$, the second is a monofractal series with strong long-term correlations $(H = 0.9)$, and the third is a combination of the first two (see Appendix A). We also obtain the corresponding shuffled time series and use MFDFA to derive the multifractality (see Fig. 9).

Figures 9(a) and 9(d) show that in the random series that obeys a power-law distribution, there is a remarkable dependence between $H(q)$ and q (there is a broad singularity spectrum), which indicates multifractality dominated by the power-law distribution. They also show that the absence of long-term correlations causes an overlap in results between the original and shuffled time series. Figures 9(b) and 9(e) show that $H(q)$ and q are independent (there is a narrow singularity spectrum), which indicates the monofractility in the monofractal series with strong long-term correlations. The long-term correlations cause $H(q)$ to change from 0.9 to 0.5. Figures 9(c) and 9(f) show that in the time series that combines the other two, the significant horizon span of the singularity spectrum and the change in $H(q)$ produce results that are similar to empirical findings. Our analysis strongly indicates that the multifractality in online viewing activity is
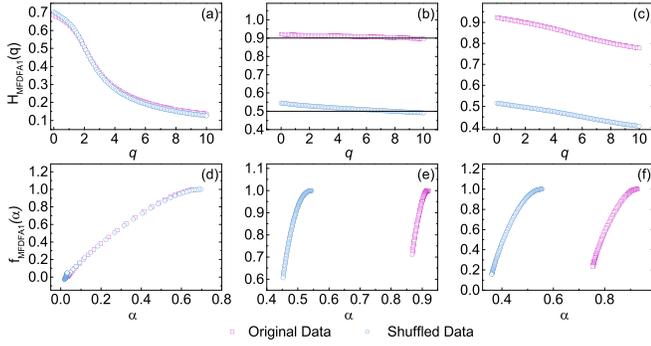
FIG. 9. Relationship between $H(q)$ and q obtained from 1-order MFDFA (a)–(c) and the corresponding singularity spectrum (d)–(f) of three types of synthetic time series, where the pink square and blue circle indicate the results of original and shuffled time series, respectively. Note that the first column shows a synthetic time series obeying a power law distribution $y \sim x^{-2}$, the second one describes a synthetic time series whose Hurst exponent is 0.9, and the third one represents a synthetic time series that combines the properties of the former two time series. Through carefully analyzing them, we can find that only the third time series gives similar results to empirical findings.

caused by the broad PDF of inter-event times and the presence of long-term correlations.

V. CONCLUSION

We have analyzed the datasets of online viewing activity from Movielens and Netflix at both the individual and communal levels. At the individual level, we find fat-tailed inter-event time distributions and the dependence on long-term correlations. Our analytical results are not exact, as in Ref. 29 because the inter-event time distributions are restricted to the activity levels and not strictly the power law. At the communal level, we find properties that are similar to those at the individual level, but here the long-term correlations are caused by the interdependence of community activity. Furthermore, the long-term correlations characterized by the Hurst exponent derived from DFA imply the presence of fractality in online viewing activity.

To determine the type of such fractality and its origin, we apply MFDFA and find multifractality at the communal level. We hypothesize that this is caused by the combined effect of the broad PDF of inter-event times and the long-term correlations. We verify this by analyzing three types of synthetic time series that have at least properties in common with a real time series. Thus, we can conclude that a dual-induced multifractality exists in online viewing activity, which enlarges this generic property commonly found in human activity from physical space to cyberspace. Nevertheless, it should not be ignored that an appropriate model is lacking for explaining the mechanism of reproducing such time series. According to Refs. 54–56, the time series of online viewing activity can be decomposed into magnitude and sign series, and by systematically analyzing them, we may obtain more dynamical properties to explain the mechanism of multifractality in online viewing activity. We hope that future work will solve these problems.

ACKNOWLEDGMENTS

APPENDIX A: CONSTRUCTION OF SYNTHETIC TIME SERIES

To construct the three synthetic time series, we first synthesize a random time series $x(t)$ that obeys a power-law distribution $p(x) = \beta x^{-(1+\beta)}$. We use the central limit theorem and generate it to be

$$x(t) = (r(t)/\beta)^{-\frac{1}{1+\beta}}, \qquad (A1)$$

where $r(t)$ is a time series sampled from a uniform distribution $U(0, 1)$. Here, we set $\beta = 1$, which causes $x(t)$ to obey a power-law distribution $p(x) = x^{-2}$.

We then apply the Flourier filtering method proposed in Ref. 57 to generate a monofractal time series with long-term correlations. The procedure is as follows:

(i)   Generate a 1-dimensional random time series U_i that follows a Gaussian distribution and derive its Fourier transform coefficients U_q.
(ii) Obtain S_q from the Fourier transformation of C_l, where $C_l = \langle \mu_i \mu_{i+l} \rangle = (1 + l^2)^{-\gamma/2}$.
(iii) Calculate $N_q = [S_q]^{1/2} U_q$.
(iv) Derive the time series N_r using the inverse Fourier transformation of N_q. In this way, we transform N_r using $N_r = N_r - min(N_r) + 1$.

We combine these two time series and synthesize the third time series

$$X(t) = (N_r(t)/\beta)^{-\frac{1}{1+\beta}}, \qquad (A2)$$

where N_r is the time series with long-term correlations.

APPENDIX B: F-TEST FOR LINEAR REGRESSION

We first show the fluctuation functions of DFA at a log-log scale from two users. Figure 10 shows the power-law relations of fluctuation functions across multiple scales for both Movielens and Netflix. We estimate the scale exponents of long-term correlations via a fitting linear relationship between $\log(F(s))$ and $\log(s)$. Then, we test the linear relationship to determine whether the linear relationship between two variables x and y is significant (i.e., $y = \beta_0 + \beta_1 x + \epsilon$). We do this by testing the following hypothesis and its alternative:

H_0: $\beta_1 = 0$, the relationship between these two variables is not significant.
H_1: $\beta_1 \neq 0$, the relationship between these two variables is significant.

In the F-test, we introduce F to measure the strength of the connection between x and y, which is defined as

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2}{\frac{1}{n-2} \sum (y_i - \hat{y}_i)^2} \sim F(1, n-2), \qquad (B1)$$

where $\hat{y}_i$ is derived from the equation $y = \beta_0 + \beta_1 x$, $\bar{y}$ is the mean value of y_i, and x_i, and y_i are the real data.
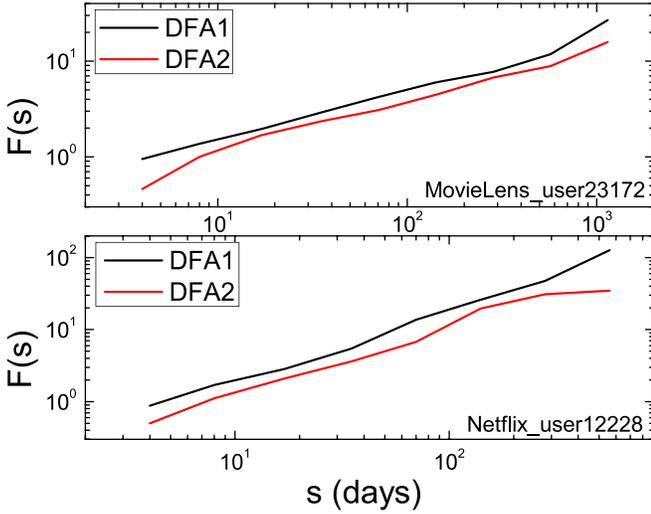
FIG. 10. The fluctuation functions of DFA at a log-log scale from two users. They clearly show the power-law relation across multiple scales for both Movielens and Netflix. For each user, the power-law relations obtained from DFA-1 and DFA-2 show approximately the same trend across multiple scales, which suggest that the scaling exponents of long-term correlations are not sensitive to the order of DFA.
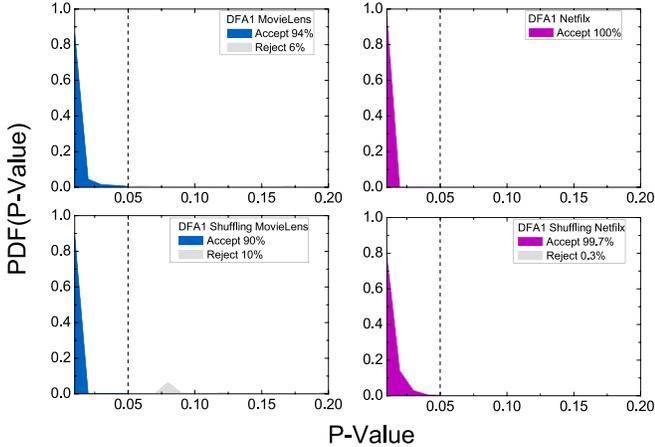

FIG. 11. The distribution of users' P-values for original data and corresponding shuffled one. It shows the strongly linear correlation between $\log(F(s))$ and $\log(s)$ for most users, which demonstrates the reliability of the results in Fig.4.

Using each F-value, we can derive the P-value for the distribution $F(1, n-2)$. At confidence level α (here, $\alpha = 0.05$), when $P < \alpha$, we accept H_1 and thus accept the linear relationship. When $P > \alpha$, we reject H_1 and thus reject the linear relationship. We compute the P value for each user to fit the relationship between $\log(F(s))$ and $\log(s)$ derived from the DFA. Figure 11 shows the P-value distribution that indicates a strong linear correlation between $\log(F(s))$ and $\log(s)$ for most users in both the original data and the shuffled data. This confirms the reliability of the results shown in Fig. 5.

APPENDIX C: SURROGATE METHODS FOR ANALYZING MULTIFRACTALITY

As mentioned above, there are two key factors that affect the multifractality in a time series of records: (i) long-term correlations in the fluctuations and (ii) a broad PDF of inter-event times. When we analyze the contributions of
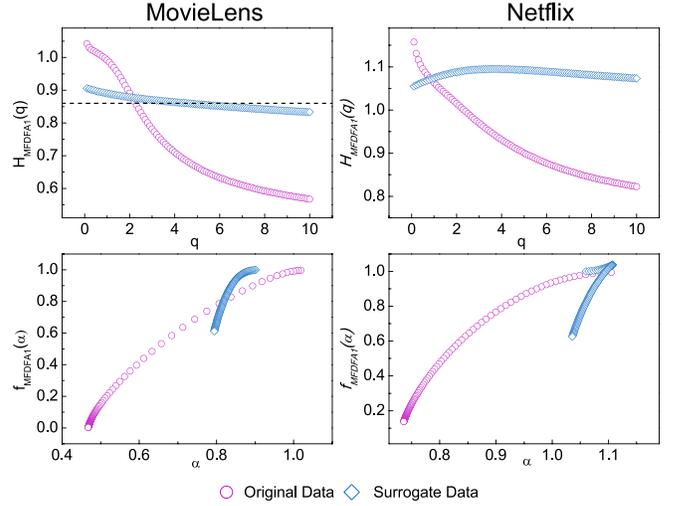

FIG. 12. Relationship between $H(q)$ and q deriving from 1-order MFDFA and the corresponding singularity spectrum. The pink circle indicates the original data, while the blue square represents the surrogate data produced by *only* Fourier phase randomization. When the broad PDF of inter-event times is disrupted, the long-term correlations still lie in the surrogate data, but the multifractality is dramatically weakened.

these two factors affecting multifractality separately, we generate many surrogate time series through shuffling and phase randomization.[58] The shuffling procedure preserves the PDF of the time series of records but destroys any long-term correlations. Thus, we randomly sort the entire time series at least 10 times. Fourier phase randomization maintains the long-term correlations but disrupts the broad PDF of inter-event times.[59,60]

Figures 12 and 13 show multifractality generated by shuffling and Fourier phase randomization. Figure 12 shows that when using only Fourier phase randomization to destroy the broad PDF of inter-event times, the long-term correlations remain in the surrogate data, but the multifractality is
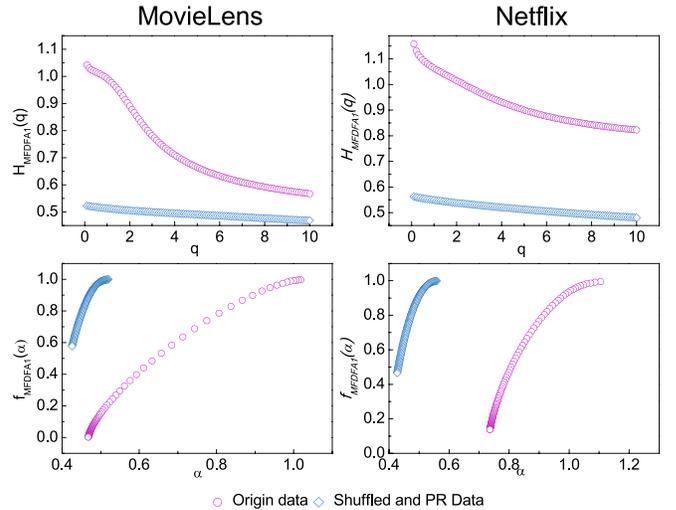

FIG. 13. Relationship between $H(q)$ and q deriving from 1-order MFDFA and the corresponding singularity spectrum. The pink circle indicates the original data, while the blue square represents the surrogate data produced by *both* shuffling and Fourier phase randomization. When the broad PDF of both inter-event times and long-term correlations are missed, $H(q)$ is close to 0.5 and has a trivial relationship with q. Moreover, the singularity spectrum also has a very narrow horizon span.

weakened. Figure 13 shows that when applying both shuffling and Fourier phase randomization to reduce long-term correlations and destroy the broad PDF of inter-event times, the $H(q)$ value is close to 0.5 and has a trivial relationship with q, and the horizon of the singularity spectrum is narrowed. These results verify the presence of dual-induced multifractality in online viewing activity.

[1] J. W. Kantelhardt, *Encyclopedia of Complexity and Systems Science: Fractal and Multifractal Time Series* (Springer, 2009), pp. 3754–3779.

[2] R. N. Mantegna and H. E. Stanley, Nature **376**, 46 (1995).

[3] C. K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger, Phys. Rev. Lett. **70**, 1343 (1993).

[4] L. S. Liebovitch, A. T. Todorov, M. Zochowski, D. Scheurle, L. Colgin, M. A. Wood, K. A. Ellenbogen, J. M. Herre, and R. C. Bernstein, Phys. Rev. E **59**, 3312 (1999).

[5] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature **399**, 461 (1999).

[6] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, H. E. Stanley, and Z. R. Struzik, Chaos **11**, 641 (2001).

[7] P. C. Ivanov, Q. D. Y. Ma, R. P. Bartsch, J. M. Hausdorff, L. A. N. Amaral, V. Schulte-Frohlinde, H. E. Stanley, and M. Yoneyama, Phys. Rev. E **79**, 041920 (2009).

[8] S. B. Lowen, L. S. Liebovitch, and J. A. White, Phys. Rev. E **59**, 5970 (1999).

[9] L. S. Liebovitch, D. Scheurle, and M. Zochowski, Method **24**, 359 (2001).

[10] C. Bédard, H. Kröger, and A. Destexhe, Phys. Rev. Lett. **97**, 118102 (2006).

[11] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J. H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).

[12] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).

[13] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin, and A. Bunde, Physica A **295**, 441 (2001).

[14] B. Gonçalves and J. J. Ramasco, Phys. Rev. E **78**, 026123 (2008).

[15] Z. Q. Jiang, F. Ren, G. F. Gu, Q. Z. Tan, and W. X. Zhou, Physica A **389**, 807 (2010).

[16] S. K. Baek, T. Y. Kim, and B. J. Kim, Physica A **387**, 3660 (2008).

[17] Y. W. Dong, S. M. Cai, and M. S. Shang, Acta Phys. Sin. **62**(2), 028901 (2013).

[18] C. K. Peng, S. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature **356**, 168 (1992).

[19] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. J. Schellnhuber, Phys. Rev. Lett. **81**, 729 (1998).

[20] H. A. Makse, S. Havlin, and H. E. Stanley, Nature **377**, 608 (1995).

[21] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, and H. E. Stanley, Phys. Rev. E **58**, 7054 (1998).

[22] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C. K. Peng, and H. E. Stanley, Phys. Rev. E **60**, 1390 (1999).

[23] S. M. Cai, P. L. Zhou, H. J. Yang, C. X. Yang, B. H. Wang, and T. Zhou, Chin. Phys. Lett. **23**, 754 (2006).

[24] S. M. Cai, Z. Q. Fu, T. Zhou, J. Gu, and P. L. Zhou, Europhys. Lett. **87**, 68001 (2009).

[25] H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley, and H. A. Makse, Proc. Natl. Acad. Sci. U. S. A. **105**, 18702 (2008).

[26] D. Rybski, H. D. Rozenfeld, and J. P. Kropp, Europhys. Lett. **90**, 28002 (2010).

[27] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Proc. Natl. Acad. Sci. U. S. A. **106**, 12640 (2009).

[28] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Eur. Phys. J. B **84**, 147 (2011).

[29] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Sci. Rep. **2**, 560 (2012).

[30] Z. D. Zhao, S. M. Cai, J. Huang, Y. Fu, and T. Zhou, Europhys. Lett. **100**, 48004 (2012).

[31] M. Kivelä and A. Porter, Phys. Rev. E **92**, 052813 (2015).

[32] H. E. Hurst, Trans. Amer. Soc. Civil. Eng. **116**, 770 (1951).

[33] H. E. Hurst, in *ICE Proceedings (Thomas Telford)* (1956), Vol. 5, p. 519.

[34] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Physica A **316**, 87 (2002).

[35] L. Xu, P. C. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, Phys. Rev. E **71**, 051101 (2005).

[36] Y. H. Shao, G. F. Gu, Z. Q. Jiang, W. X. Zhou, and D. Sornette, Sci. Rep. **2**, 835 (2012).

[37] Z. Chen, P. C. Ivanov, K. Hu, and H. E. Stanley, Phys. Rev. E **65**, 041107 (2002).

[38] Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. C. Ivanov, Phys. Rev. E **71**, 011104 (2005).

[39] Q. D. Y. Ma, R. P. Bartsch, P. Bernaola-Galva, M. Yoneyama, and P. C. Ivanov, Phys. Rev. E **81**, 031101 (2010).

[40] Y. Xu, Q. D. Y. Ma, D. T. Schmitt, P. Bernaola-Galva, and P. C. Ivanov, Physica A **390**, 4057 (2011).

[41] M. S. Movahed, G. R. Jafari, F. Ghasemi, S. Rahvar, and M. R. R. Tabar, J. Stat. Mech. **2006**, P02003 (2006).

[42] G. Lim, S. Kim, H. Lee, K. Kim, and D. I. Lee, Physica A **386**, 259 (2007).

[43] J. Ludescher, M. Bogachev, J. W. Kantelhardt, A. Schumann, and A. Bunde, Physica A **390**, 2480 (2011).

[44] P. Meakin, Adv. Colloid Interface Sci. **28**, 249 (1987).

[45] H. O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science* (Springer, 2004).

[46] W. M. Tulczyjew, *Annales De L'institut Henri Poincaré (a) Physique Théorique* (Gauthier-villars, 1977), Vol. 27, p. 101.

[47] A. L. Barabási, Nature **435**, 207 (2005).

[48] P. C. Ivanov, A. Yuen, B. Podonbnik, and Y. Lee, Phys. Rev. E **69**, 056107 (2004).

[49] P. C. Ivanov, A. Yuen, and P. Perakakis, PLoS ONE **9**, e92885 (2014).

[50] Z. Eisler and J. Kertész, Phys. Rev. E **73**, 046109 (2006).

[51] Z. Eisler, I. Bartos, and J. Kertész, Adv. Phys. **57**, 89 (2008).

[52] J. W. Kantelhardt, D. Rybski, S. A. Zschiegner, P. Braun, E. Koscielny-Bunde, V. Livina, S. Havlin, and A. Bunde, Physica A **330**, 240 (2003).

[53] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, J. Geophys. Res. **111**, D01106, https://doi.org/10.1029/2005JD005881 (2006).

[54] Y. Ashkenazy, S. Havlin, P. C. Ivanov, C. K. Peng, V. Schulter-Frohlinde, and H. E. Stanley, Physica A **323**, 19 (2003).

[55] M. Gómez-Extremera, P. Carpena, P. C. Ivanov, and P. Bernaola-, Galvan, Phys. Rev. E **93**, 042201 (2016).

[56] D. T. Schmitt, P. K. Stein, and P. C. Ivanov, IEEE Trans. Biomed. Eng. **56**, 1564 (2009).

[57] H. A. Makse, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. E **53**, 5445 (1996).

[58] P. Norouzzadeh and B. Rahmani, Physica A **367**, 328 (2006).

[59] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).

[60] D. Prichard and J. Theiler, Phys. Rev. Lett. **73**(7), 951 (1994).