Indicator of serious flight delays with the approach of time - delay stability

Yan - Jun Wang,[1] Yun - Feng Zhu,[2] Chen - Ping Zhu,[2] Fan Wu,[1] Hui - Jie Yang,[3] Yong - Jie Yan,[4] and Chin-Kun Hu[5,6]

[1] *College of Civil Aviation, Nanjing University Aeronautics and Astronautics, Nanjing, 211106, China*
[2] *College of Science, Nanjing University Aeronautics and Astronautics, Nanjing, 211106, China*
[3] *College of Management, Shanghai University for Science and Technology, Shanghai, 200053, China*
[4] *State Key Laboratory of Air Traffic Management System and Technology, Nanjing, 210000, China*
[5] *Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[6] *Department of Physics, National Dong Hwa University, Hualien 97401, Taiwan*

(Dated: October 28, 2018)

Passenger flight delays, causing much disorder of air traffics, economic losses of airlines, and downgrading the travel quality of millions of people, are ubiquitous phenomena in airports all over the world. Investigation based on real data from the view point of statistical physics is rarely seen. In the present work, big data of such delay records over 20 years accumulated by Bureau of Transportation Statistics in the United States are downloaded and purified by us. We account the departure and arrival records of such flights between certain pair of airports as time series, and rectify them by defining dimensionless velocity of the flights. Furthermore, we find the varying cross-correlations among such time series with the approach of time delay stability, and describe the correlations with temporal networks for correlation states. Deterministic correspondences between the average degrees of temporal networks and delay ratios of passenger flights are verified in different sampling groups of flights with the longest records. The mean degrees of correlation networks usually emerge a peak prior to that of high delay ratios, which serves an indicator for the precaution to serious flight delays.

PACS numbers: 87.23.Kg, 89.40.Dd,89.90.+n

INTRODUCTION

Passenger flight delays happen every day in airports all over the world, which brings disorder in air traffics, billions dollars of economic losses of airlines, and unhappy travels of millions of people. To alleviate such phenomena, people need to know some fundamental laws of flight delays which are still a void of research. Statistical physics is well known tool being used to the analysis of various complex systems. However, it has rarely been applied in the investigation of flight delays, although a few works [1–4] have transplanted it into the modeling of air transportation. In this paper, we report some new results from the cross-correlation analysis of many time series from big data of domestic passenger flights in the USA for 20 years, which is done through the approach of time delay stability (TDS)[5]. A new indicator predicting high ratio of flight delays is found by the observation of the average degree of cross-correlation[5–21] network with TDS, which is also verified by traditional tool of Hurst index[22].

Domestic passenger flights are definitely correlated since some of them take off or land on common airports, use the same aircrafts at different flights, share the same air routes or the same sectors, or being affected by the same weather systems. Apart from the competition between different airlines, some counter-intuitive cross-correlations are also found to exist. Our task here is to take a reliable new approach to directly reveal the cross-correlations between flights from the time series obtained from historical records.

Algorithmic theories of cross-correlations between different time series have recently been much developed. V. Plerou et al. [6–8] used methods of random matrix to analyze the cross-matrix C of stock price changes of the largest 1000 US companies for the years 1994-1995, and found that the statistics of most eigenvalues in the spectrum of C agree with the predictions of random matrix theory, although with some deviations for a few of the largest eigenvalues. B. Podobnik and H. E. Stanley [9] proposed a new method – de-trend cross-correlation analysis (DCCA) to the analysis of two non-stationary time series, and illustrated the method by selected examples from physics, physiology and finance. W. X. Zhou [10] then proposed multi-fractal de-trend cross-correlation analysis to investigate the multi-fractal behaviors in the power-law cross-correlation between two time series or high dimensional quantities recorded simultaneously. Z. Q. Jiang and W. X. Zhou [11] further developed multi-fractal de-trend moving average cross-correlation analysis, and compared different efficacy with various methods in extensive numerical simulations. In this direction, X. Y. Qian [12] and her collaborators then proposed de-trend partial cross-correlation analysis (DPXA) of two non-stationary time series. They analyzed multi-fractal (MF) binomial measures masked with strong white noises and found that MF-DPXA succeeds while MF-DCCA failed to quantify the hidden multi-fractal nature. These theoretical tools have been widely applied in the analyses of various complex systems [13, 14]. For example, cross-correlation in financial dynamics [15, 16], price-volume relationships in agricultural commodity futures [17], multi-scale MF-DCCA of financial time series [18, 19], discrete scale-invariance in cross-correlations between time series [20], and directed networks of influence [21]. These works represent the main stream of cross-correlation analyses between time series.

The time delay stability method (TDS) proposed by Bashan et al.[5] presents a new idea of mining cross-correlations in a pair of time series. Just as what is well known, integrated dynamical systems are coupled by feedback and/or feed-forward loops with a broad range of time delays. In the presence of stable/strong interactions between two systems, transient modulations in the output signal of one system lead to corresponding changes that occur with a stable time lag in the output signal of another system, thus a maximum value of correlation intense appears for most suitable lag of one time series sliding along the other one. With certain thresholds to check correlation intense between time series, hence the links between records, they succeed in setting up correlation state networks for different moments, and find drastic changes of the average link numbers of such temporal networks deterministically, corresponding to the transitions between physiological states, e.g., deep or slight sleeps and awake.

Inspired by successful applications of TDS in the analysis of physiological systems [5], we use it to mine the possible cross-correlations between historical records of domestic passenger flights in USA, and set up a series of correlation state networks for all the time steps. By direct observation of average links of them, obvious peaks prior to large ratios of flight delays were revealed directly. Meanwhile, these kinds of correspondence are supported by similar Hurst indices of both time series in a traditional reference frame.

MODEL

Aiming at reducing the ratio of flight delays, we need to find cross-correlations among time series of all historical operation records of passenger flights. Primary big data are downloaded from the web site of Bureau of Transportation Statistics (BTS)[23] of the United States. Each item of such records contains date, flight code, aircraft code, departure/arrival moments and airports on the schedule, departure/arrival time practically operated. We concentrate ourselves on actual velocities in comparison with what are designed on the schedule, hence to concern actual delay correlation behaviors between flights in the whole historic records. To this end, we define the dimensionless velocity (DV) by dividing time interval Δt_s on the schedule by the corresponding time interval Δt_a it takes by the actual flight. In this way, geometric distance from departure airport to the arrival airport is reduced. We use v_d as the metric to measure the speed of a flight in the following formula

$$v_d = \frac{\Delta t_s}{\Delta t_a} = \frac{t_{sa} - t_{sd}}{t_{aa} - t_{ad}} \tag{1}$$

where t_{sa} and t_{sd} represent arrival and departure time on the schedule, while t_{aa} and t_{ad} represent arrival and departure time the flight actually operated by the same flight. Obviously, it delays when $v_d < 1$, and it gains time when $v_d > 1$. The smaller the v_d, the more serious it delays. Depending on different habits, passengers usually sense late arrival or departure of an aircraft as delay phenomena, instead of the criterion $v_d < 1$ used here. For more convenient mining correlations of en route processes of different flights, we chose v_d. Actually, in the present stage, we have dissolved delay problems into different topics. We focus on DA processes in this work. With DA we mean departure form airport 1 and arrival at airport 2. Actually we have another work dealing with departure delays due to the late arrival of last flight (see arxiv.1701.05556 [24], for reference).

We define the flights on the route with the same pair of original and destination airports and operated by the same airline as the same flight. Thus it accumulates a few tens of thousands of departure records in 20 years. Therefore, the same number of dimensionless velocities are accumulated for each flight as a time series in the same route. Indeed, flights do not always take off simultaneously, we should count their cross-correlations within a certain length of time interval, which is done by equally dividing the whole 20-year time over the total number of flights in the same route. Moreover, we go further to calculate the cross-correlations between different flights with the smallest time-difference of taking off moments as possible.

To determine if any two time series of dimensionless velocities are correlated at moment k, in the following scheme we adopt the TDS approach by Bashan et.al. essentially. We denote the constructed dimensionless velocity series with, $S = \{S_{\mathrm{m,n}}, m = 1, 2, \cdots, M; n = 1, 2, \cdots, N\}$, where M is the total number of air routes and N the length of flight records. For each discrete time moment k, we have a time window with specified width T, let it slide with the step Δ

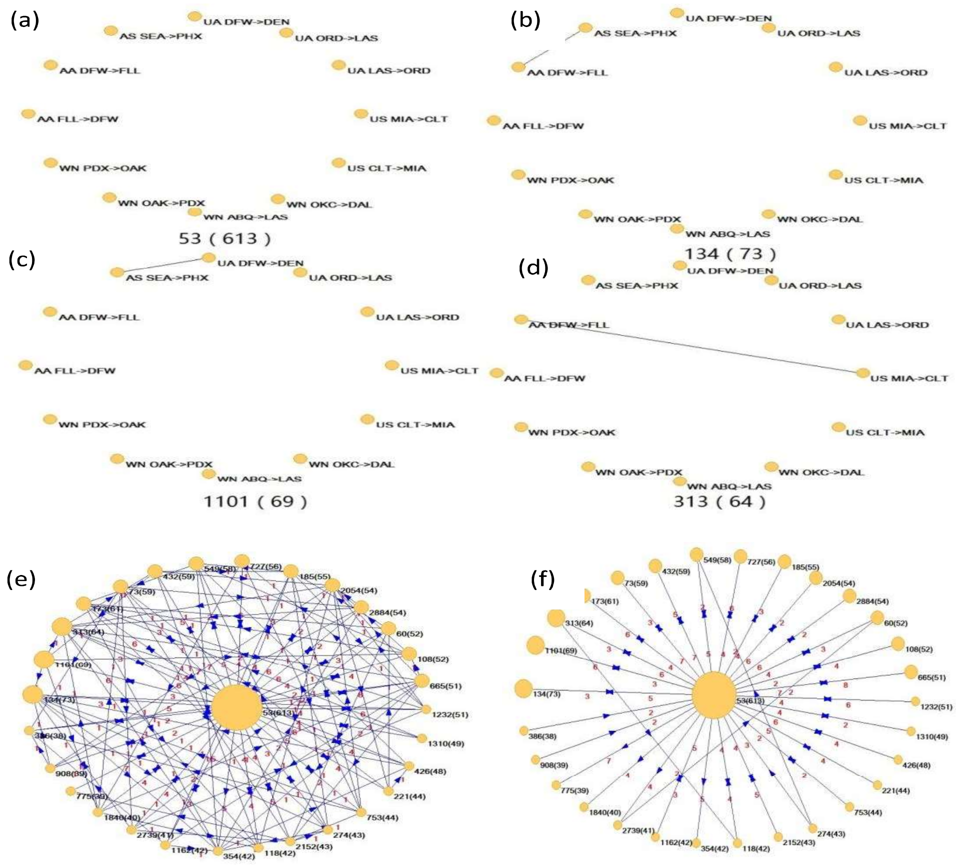


Figure 1: The temporal networks of correlation states with the top-4 times of appearance, and the evolution of top-30 dense correlation states. (a)53(613);(b)134(73);(c)1101(313);(d)313(64).(a)-(d):The left numbers represent the moments they appears the first time, the right one in brackets represent the times they appear. (e)Temporal evolution of correlation states with the top-30 times of appearance, with the sizes of the nodes proportional to the times it appears. The number beside a node(correlation state) represents the moment it appears the first time, and the number in the bracket represents the time it appears. The arrow of a link indicates the transition direction of the correlation states, with the number beside it indicating the transition times. (f)The simplified temporal evolution from (e) with all links appearing only once deleted.

along S to produce a series of overlapping segments as time elapses, $S(k) \equiv \left\{S_{\mathrm{m},(\mathrm{k}-1)\Delta+\mathrm{i},\mathrm{m}=1,2,\cdots,\mathrm{M};\mathrm{i}=1,2,\cdots,\mathrm{T}}\right\}, k = 1, 2, \cdots, \left[\frac{N-T}{\Delta}\right]$.(see Fig.2(a) The pattern of cross-correlations between all the pairs of rows in the segment $S(k)$ is used cooperatively to represent the system's state at the k-th moment. In this way we can monitor the evolutionary behavior of the system. The key idea of TDS is that if there exists a causal relation from one to another flight, signals from the source to the destination may vary in amplitude and persistence etc., but the required time for a signal being transferred (time delay) is determined by intrinsic properties of the influential mechanism and keeps subsequently unchanged, namely, the time delay is stable. This fact will guarantee a high confidence of influence.

For illustration, we consider two rows in $S(k)$ that corresponding to the air routes numbered a and b, namely, $\left\{S_{a,(k-1)\Delta+i}, i = 1, 2, \cdots, T\right\}$ and $\left\{S_{b,(k-1)\Delta+i}, i = 1, 2, \cdots, T\right\}$.(see Fig.2(b)). The length-L sub-segments starting at the s'th elements read, $x \equiv \left[S_{a,(k-1)\Delta+s+t}, t = 0, 1, \cdots, L-1\right\}, y \equiv \left[S_{b,(k-1)\Delta+s+t}, t = 0, 1, \cdots, L-1\right\}$, respectively. Let y slide along x. At each step one calculates the cross-correlation coefficient between the matched parts. For simplicity, x is usually extended periodically(see Fig.2(c)). The number of slide steps at which the absolute value of Pearson coefficient C reaches maxima is defined to be the time delay for the influence between a and b(see Fig.2(d)). Scanning the starting element s in the interval of $[1, T-L+1]$, we obtain a total of $T-L+1$ time delays. Then we

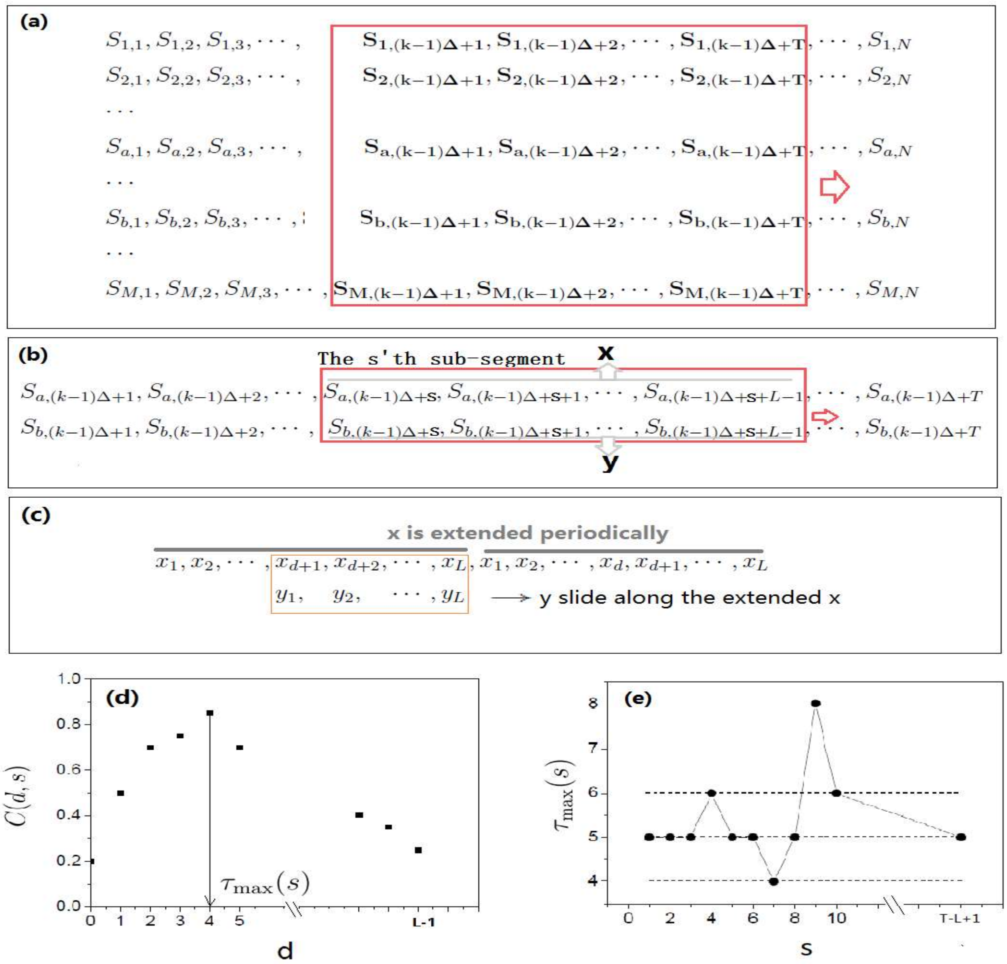


Figure 2: Procedure to determine Time-Delay Stability at moment k. (a) A time window (red frame) with size T slides along the time series of dimensionless velocity with a step of Δ to extract a total of $\left[\frac{N-T}{\Delta}\right]$ successive overlapping segments. The k-th segment in the red frame corresponds to the k-th moment. (b) Let a sub-window with size L (red frame) slide along the series of both a and b in the kth segment, which produces a total of $T-L+1$ overlapping sub-segments in parallel, denoted with x and y, respectively. (c) Let y slide along the periodically extended x, and calculate the Pearson coefficient of the matched parts. (d) A total of L Pearson coefficients with delays numbered $0,1,2,...,L-1$ are obtained. The delay corresponding to the maximum of absolutes of cross-correlation coefficient is defined to be Time Delay of influence from y to x. (e) For the kth segment, a total of $T-L+1$ time delays are obtained. Count the number of the time delays which are very close to (± 1) or identical to that for the first sub-segment, if the ratio of number over $T-L+1$ is no less than a specified ratio γ, the influence from y to x (the flight b to the flight a at the moment k has Time Delay Stability.

count the number of the time delays identical or very close (± 1) to the first one. If the ratio of such number over all $T-l+1$ sub-segments, is larger than a specified threshold γ ,the time delay is regarded to be stable. Therefore, we confirm a reliable influence between the two flights in air routes a and b at moment k (see Fig.2(e)), and connect two flights (filled circles in Fig.1(a)–(d),Fig.6 and Fig.8)denoted by airlines and origin-destination airports with links to form temporal networks of correlation states.

During the 20 years, flight numbers deviates quite a lot at different air routes, and these flights last in various time intervals. Therefore, the lengths deviate from each other from route to route. However, we need to take the lengths of time series as consistence as possible. In the present work, we calculate the cross-correlations between time series under the following conditions: to select the time series as long as possible; to get a lowest limit of the length with the shortest time series in the same group. To be more specific, we have the longest 12 time series with their lengths from 40500 to 40700, and cut off others to make up the time series with the equal length 40500 in this sampling group. For the sampling groups with the lengths of time series in the ranges $(36700, 40700)$ and (35700, 40700), we have 84

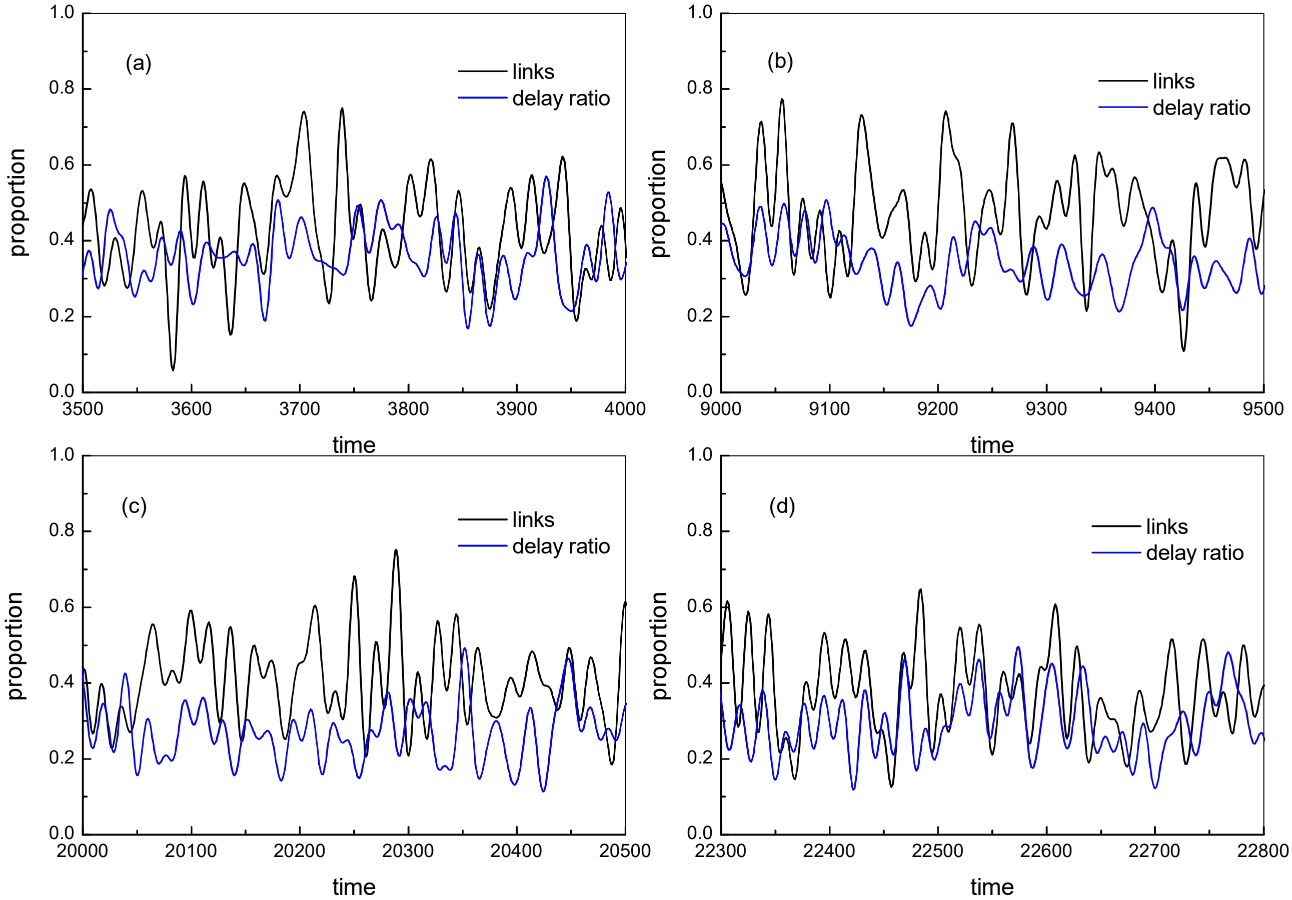


Figure 3: Typical examples comparing the ratios between the link numbers in the temporal networks of correlation states and that of the densest one with the temporal variation of delay ratio of the flights in the sampling group with 12 flights. (a)From the moment 3500 to 4000.(b) From the moment 9000 to 9500.(c) From the moment 2000 to 20500.(d)From the moment 22300 to 22800.

and 102 flights, we select out 4 records of the same flight uniformly every day, so that we get the minimum standard error of the time series of the intervals between the records. Here the time intervals of any two take-off records of the same flight vary in the range $(4, 8)$ hours, and take the average value as 6 hours. According to the scheme of TDS[5], $M = N - T + 1$, we have $N = 40500, 36700$ and 35700, hence $M = 40492, 36692$ and 35692 for the group with $12, 84$, and 102 flights with the same parameter set in the TDS analysis. For the sampling groups with different numbers of flights N, we all have $T = 9$, $L = 5$, $\Delta = 1$, $C_0 = 0.6$, and $\gamma = 0.8$. In Fig.1(a)-(d), any link between two flights at moment k in its temporal network of cross-correlated states satisfies the conditions that the maximum correlation value C is higher than or equal to the threshold $C_0 = 0.6$ and the time delays are stable in the range $\tau_{max} \pm 1$. Based on TDS, the length N and segment length T, we have totally 40492 networks of M cross-correlation states for the sampling group with 12 flights. Some of such networks of correlation states reappear many times, which can be called attractors. In Fig.1, we label any one with the moment code k indicating its first appearance, and the number in the bracket beside it to label the times it appears. Altogether, 32425 independent correlation networks appear among which 2762 are attractors, taking the fraction of 8.5 percent.

RESULTS

Four networks of correlation states with the highest appearing frequencies are shown in Fig.1(a), (b), (c) and (d). Any node in it stands for a time series of a flight of certain airline, while a link connecting any pair of nodes represents the correlation determined by TDS. The first appearing moments of 4 attractors are 53, 134, 1101 and 313, respectively. And the times of appearance are $613, 73, 69$ and 64, respectively. The most frequently appeared attractor number 53 is actually a void network, and only one link exists in the remaining 3 networks. Other high-frequently appeared

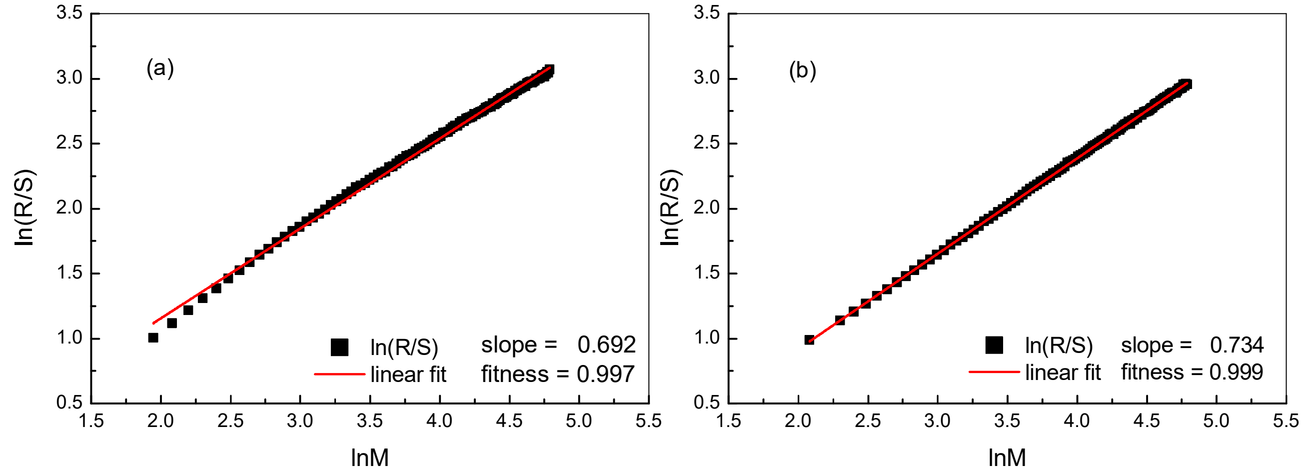


Figure 4: (a)Hurst index of the time series of mean degrees in the temporal networks of correlation states, for the sampling group with top-12 flights. (b)Hurst index of the time series of delay ratio of the flights in the sampling group with top-12 flights.

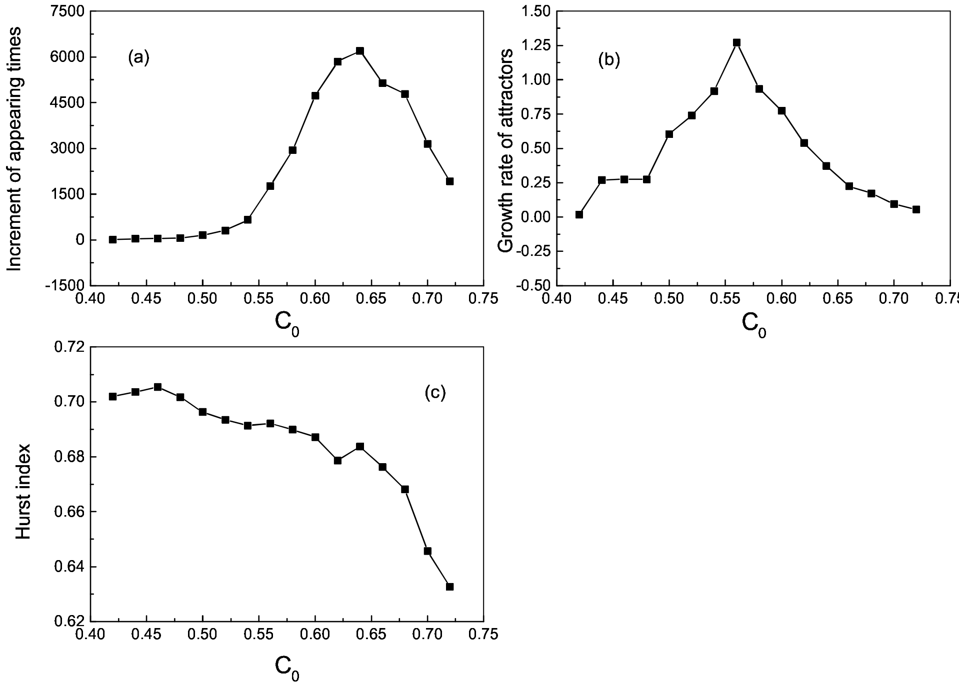


Figure 5: (a)Increment of appearing times of attractors of the correlation states versus the thresholds C_0.(b)Growth rate of attractors versus the thresholds C_0.(c)Hurst index of the time series of link numbers versus the thresholds C_0. All for the sampling group with top-12 flights.

networks are all sparsely connected networks. On contrary, large amount temporal networks which appears only once have denser links. In general, the weight-averaged link number is around 5.39. We need to construct an evolution network to describe the temporal variation of the networks of correlation states for all the moments. Every node in such a network represents a correlation state, while a direct link represents the evolution from this one to that one being pointed at. Self-loop is allowed to describe the reappearance of the same correlation state. It is difficult to depict the entire evolution process for all ever appeared correlation states since the number of independent correlation states are above two thousands. In Fig.1(e) we display the evolution network of the top 30 frequently appeared correlation states, i.e, 30 nodes, with their sizes representing the times of appearance in the evolution. The number by a node represents the moment it appears the first moment, while the number inside the bracket represents the times of its appearance. Direct links describe the sequence of the appearance of the correlated states, with the number beside

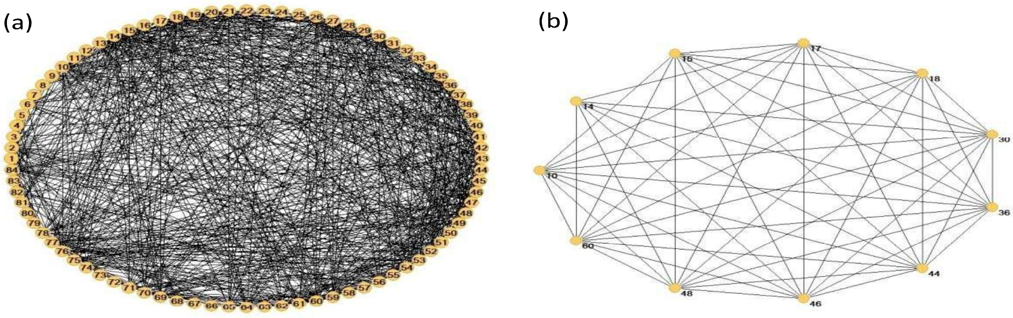


Figure 6: (a)The densest temporal network of correlation states, No. 9807.(b)The simplified version of (a) with the nodes having the degrees smaller than 30 deleted.

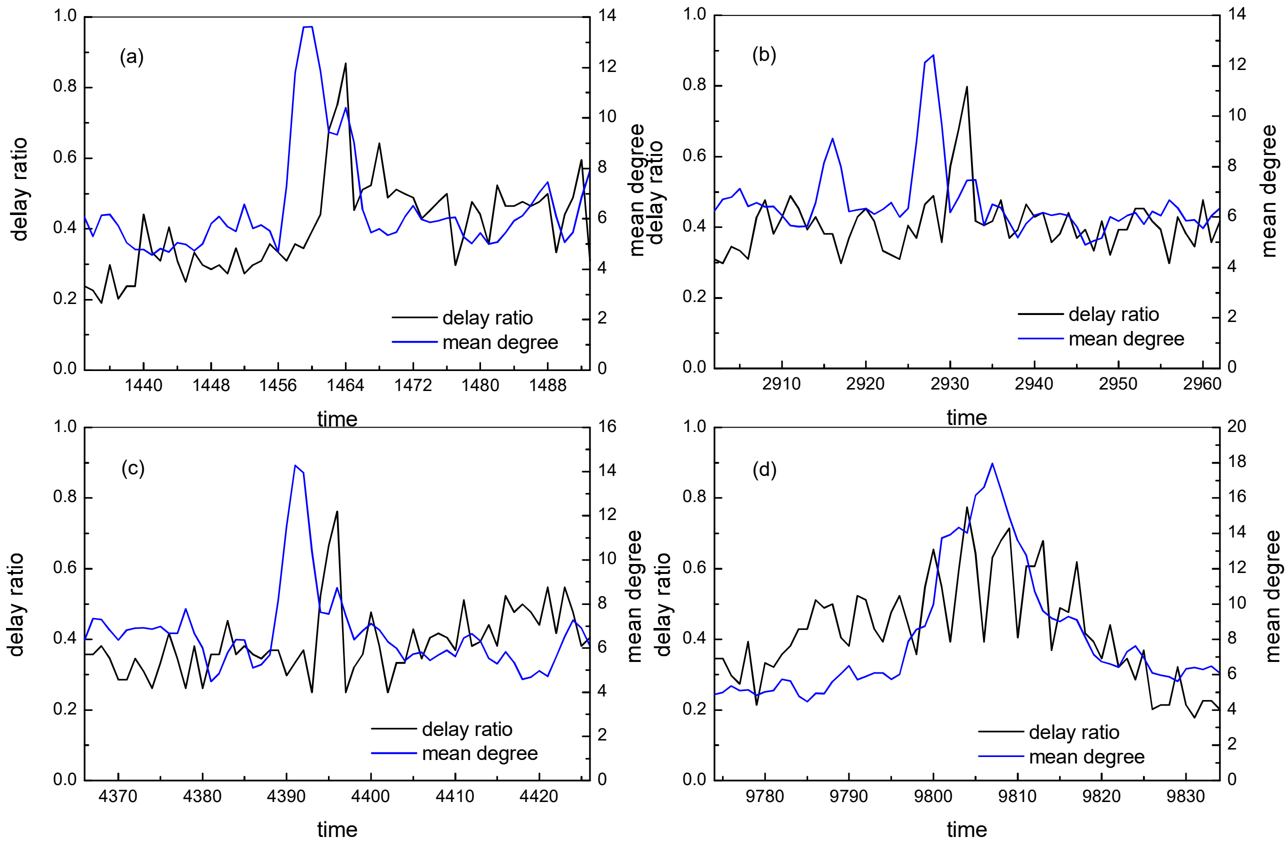


Figure 7: A single peak of the mean degree of temporal networks of correlation states appears prior to a single peak of serious (above half) flight delays for the sampling group with top - 84 flights. While two kinds of peaks appear almost simultaneously for multi-peaks of the delay ratios. The time series of two quantities in four typical segments are shown for the comparison.(a) From the moment 1432 to 1493; (b)From the moment 2902 to 2962; (c)From the moment 4365 to 4426;(d)From the moment 9774 to 9834. Note the heights of delay ratio peaks are all above 60 percent.

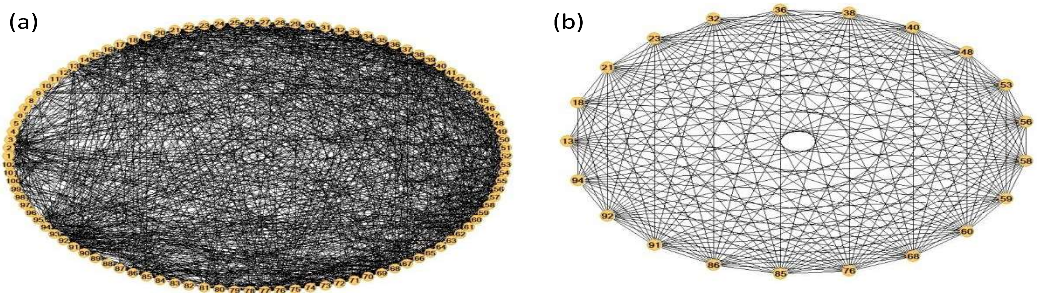


Figure 8: (a)The densest temporal networks of correlation states, No.9807, for the sampling group with 102 flights.(b)The simplified version of (a) with the nodes having the degrees smaller than 35 deleted.

them representing the times of the appearances in such state-evolution.

To show the key processes of evolution of correlation states, we erase all the directed links appears just once. A simplified result is shown in Fig.1(c). One can see that the evolution is always directed for all links appearing only once. Moreover, some repeatedly appeared links are also directed. With the threshold $C_0 = 0.6$, correlation state $No.53$ possesses the largest degree.

We need to pursuit the air-transportation meanings from our simulations since the primary data are records of operated flights. An intuitive idea inspired by the investigation of Bashan et al.[5] is to compare the temporal behaviors of link-numbers of the networks of correlation states with the variation of delay ratio in the sampling group (12, 84 or 102) investigated. First we do normalization of link-numbers by dividing link-numbers at all moments with the averaged number appeared in the evolution process. Then we smoothen the mean degree of the time series of temporal correlation networks and flight delay ratio respectively with the technique of fast Fourier Transformation (FFT) since both of them do not vary continuously in the exact sense. Fig.3 shows four typical examples of contrary panels between smoothen link number variations and delay ratio variations. At many moments, e.g., at moment $k = 3700, 3750, 3850, 9050, 9420, 20040, 20450, 22460, 22550,$ and 22600, the peaks as well as valleys of both smoothen time series fit well with each other. From the limited sampling of the top longest time series, positive correlation between the link number of correlation states and delay ratio of flights is dominant (shown in panels (a), (b) and (d)) to their negative correlations as shown in panel (c). Of course, as shown in panel (d) in both Fig.7 and Fig.9, the wider peaks of the link - numbers of correlation networks appear almost simultaneously with multi-peaks of flight delay ratios, which still show certain correspondence between the both, although the former no longer appear prior to the later. For precise meaning, we define the delay ratios above 50 percent as serious delay as shown with the black peaks in both Fig.7 and Fig.9.

In comparison, we calculate Hurst Indices with rescaled range analysis for the series of link numbers in correlation states, and for the series of ratios of delayed flights. The slopes of the fitting lines in both panels of Fig.4 are their Hurst indices, respectively. It is 0.692 for the series of link numbers, and 0.734 for the series of delaying ratios of the flights. Both are larger than 0.5 and close to each other in their values, which means they are the persistent series, and their abilities to keep the increasing tendency or decreasing tendency are quite similar.

The change of link numbers in correlation states and that of ratios of delayed flights behave in similar ways, which could be understood from the propagation of flight delays. A delay of one flight can hardly propagate to other flights when the links in the correlation states are quite sparse. Consequently, the delay ratio of the whole civil aviation system is quite low. In contrast, a delay of one flight can propagate to other flights easily when the links in the correlation state are dense. Consequently, the delay ratio becomes high. Therefore, the link numbers in the correlation states can be regarded as the indicators of flight delay ratios.

As we increase the threshold C_0, the networks of correlation states become sparser and sparser, possible new networks available become less and less. The state space becomes smaller and smaller as the evolution of the states, which induces growing number of attractors till all networks become a void one, *i.e.*, the state space collapses into a single point. Moreover, the change of link numbers can predict the change of delay ratio in some degree. From the view point of Hurst index, to take the threshold C_0 between 0.5 and 0.6 is suitable, since the variation in this range looks stable than in other ranges, which is shown in Fig.5. As the side evidence, the peaks of attractor growth rate

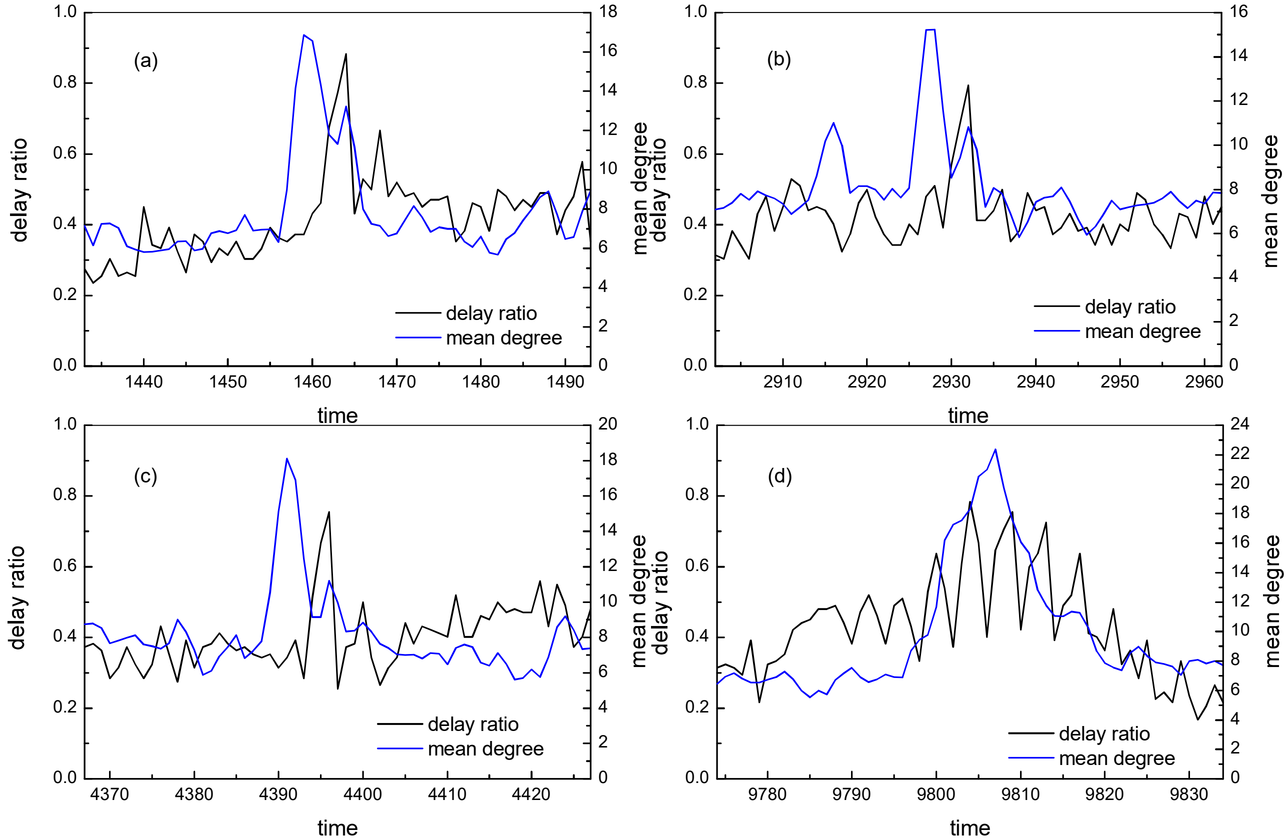


Figure 9: A single peak of the mean degrees of temporal networks of correlation states appears prior to a single peak of serious(above half) flight delays for the sampling group with top - 102 flights. While two kinds of peaks appear almost simultaneously for multi-peaks of the delay ratios. The time series of two quantities in four typical segments are shown for the comparison.(a) From the moment 1433 to 1493;(b)From 2902 to 2962; (c)From 4367 to 4427;(d)From 8774 to 9834. Note the heights of the delay ratio peaks are all above 60 percent.

(with threshold interval 0.02 appears) around the thresholds in (0.5, 0.6). Therefore, the selected threshold as 0.6 is suitable in our calculation.

For the sampling group with top-84 flights, we take the same set of parameters as in the group with top-12 flights. In contrast, we can not see any attractor in this group. As a typical example, we show here the temporal network of correlation states at the moment $k = 9807$ since it is the densest network during the whole evolution. Again nodes represent the flights, each link between a pair of flights is determined correlation by TDS and the threshold $C_0 = 0.6$. The temporal networks constructed this way for every moment are quite complex, so we just show the densest one, *i.e.*, No. 9807 in Fig.6(a) and its simplified version with the nodes having the degrees smaller than 30 deleted. This version is shown in Fig.6(b), and much similar to a complete graph. We compare the behaviors of the time series of link numbers of the temporal networks of correlation states and that of delay ratios. It is difficult to display the details of the whole evolution result of these time series. So we display 4 typical segments in Fig.7 to show a deterministic correspondence for all the cases with delay ratios higher than 65 percent. Every time once a peak emerges for the average degree of the temporal network of correlation states(solid blue lines), immediately a peak of flight delay ratios(dashed black lines) definitely follows it. This correspondence emerges without the operation of FFT as in the case of the sampling group with 12 flights, which means the increased number of flights causes the statistical property to improve. The obvious correspondence of such phenomena can be understood in the same way as the mechanism for the group with top-12 flights. The more connections in such networks, the delays are easier to propagate to other flights.

For the sampling group with top-102 flights, the situation is even more complex than that of the group with top-84 flights. It is impossible to see any attractor in this case, and very difficult to clearly display the temporal networks of correlation states at every moment. We show the densest temporal network at moment $k = 9807$ (the same as in the group 84) in Fig. 8(a), and show the simplified version with the nodes only have the degrees larger than 35 in

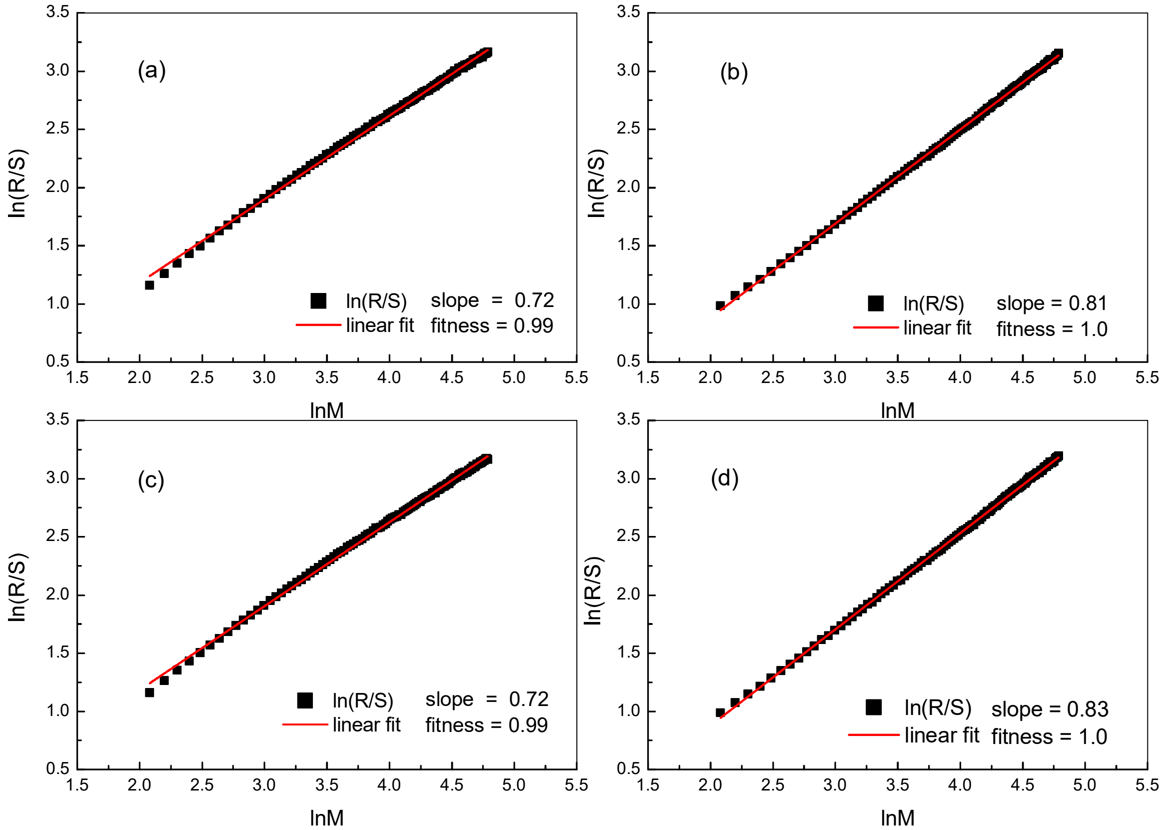


Figure 10: (a)Hurst index $H = 0.72$ for the time series of link numbers of correlation states of the sampling group of top-84 flights; (b)Hurst index $H = 0.81$ for the time series of delay ratios of the flights in the same group as (a). (c)Hurst index $H = 0.72$ for the time series of link numbers of correlation states of the sampling group of top-102 flights; (d)Hurst index $H = 0.83$ for the time series of delay ratios of the flights in the same group as (c).

Fig.8(b). Again we see a network much close to a complete graph for the simplified version. Once again, we display 4 typical segments to show a deterministic correspondence for all the cases with delay ratios higher than 65 percent. Once a peak emerges for the average degree of the temporal network of correlation states, immediately a peak of flight delay ratios definitely follows it without the use of FFT. The results in this group are shown in Fig.9(a) to (d), which provides further examples of correspondence of such two kinds of peaks. For more sureness we can say that the peaks of the networks of correlation states are prior to those of delay ratios. Therefore, we can reckon the peaks of the average link number (hence its ratio) of the correlation states as the precaution signals before serious flight delays(above 50 percent delay ratio shown in both Fig.7 and Fig.9) of domestic passenger flights.

The correspondence between the time series of the networks link - numbers of correlation states and that of delay ratios can also be supported by the similarity of the Hurst indices of both of them. Fig.10(a) and (b) display the case for the group with top-84 flights, with the Hurst indices 0.718 and 0.809 for the time series of link numbers and delay ratios, respectively. While these Hurst indices for the case of the group with top-102 flights are 0.721 and 0.827, respectively(see Fig.10(c)and (d)). The values for the present two groups are higher than those in the case with top-12 flights, which implies higher persistence of the time series than that with smaller number of flights. More importantly, Fig.10 gives more intense impression of statistical similarity between the temporal variations of two types of time series. Besides, it implies that the similarity between them would not change essentially with the increased number of passenger flights sampled.

CONCLUSION

In conclusion, we investigated the cross-correlations between the time series of dimensionless velocities of different domestic passenger flights of the USA with the TDS method. By sampling the longest records in three groups, we have verified the correspondence between the peaks of link number density of temporal networks of correlation states and the peaks of flight delay ratios, which erects a new indicator prior to serious flight delays. Moreover, the results are supported by the similar Hurst indices of both time series of them. In the fields of economy, ecology, systems biology, epidemics, electronic business and "science of science", time series tend to cross correlate to form the collective behaviors. Therefore, a lot of applications may be developed with TDS to set up temporal networks of correlation states. The present work provides an example of the analysis of big data of flight delays with the tool of statistic physics, and serves a new example to find cross-correlation from different real time series in complex systems. Better precaution signals or global measurement[24] on flight delays are hopefully found in future works.

ACKNOWLEDGEMENT

We acknowledge financial support from NSFC under Grant Nos.11775111, 61773203 and U1833126. Y.J. Wang is supported by State Key Laboratory of Air Traffic Management System and Technology under Grant No.SKLATM201707. CKH has been supported by Grant MOST 106-2112-M-001-027 and MOST 107-2112-M-259 -006".

[1] E. B. Peterson, K. Neels, N. Barczi and T. Graham, The economic cost of airline flight delay. J. Trans. Eco. and Pol. 47(2013) 107.
[2] L. I. Xiong, G. C. Liu, M. C. Yan and W. Zhang, The economic loss of airlines and passengers caused by flight delays. Syst. Eng. 25(2007) 20.
[3] M. Ball, C. Barnhart, M. Dresner, M. Hansen, K. Neels and A. Odoni, et al. Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the united states. The National Center of Excellence for Aviation Operations Research (NEXTOR) Nev. 2010.
[4] S. A. Morrison and C. Winston, Another look at airport congestion pricing. Ame. Eco. Rev. 97(2007) 1970.
[5] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin and P. C. Ivanov, Network physiology reveals relations between network topology and physiological function. Nat. Comm. 3(2012) 702.
[6] V. Plerou, P. Gopikrishnan,B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley, Random matrix approach to cross correlations in financial data. Phys. Rev. E 65(2002) 066126.
[7] V. Plerou, P. Gopikrishnan,B. Rosenow, L. A. N. Amaral and H. E. Stanley, Collective behavior of stock price movementsa random matrix theory approach. Physica A 299(2001) 175.
[8] B. Rosenow, V. Plerou, P. Gopikrishnan, L. A. N. Amaral and H. E. Stanley, Application of random matrix theory to study cross-correlations of stock prices, Inter. J. Theo. Appl. Finan. 3(2002) 399.
[9] D. Horvatic, H. E. Stanley and B. Podobnik, Detrended cross-correlation analysis for non-stationary time series with periodic trends. Europhys. Lett. 94(2011) 18007.
[10] W. X. Zhou, Multifractal detrended cross-correlation analysis for two nonstationary signals. Phys. Rev. E 77(2008) 066211.
[11] Z. Q. Jiang and W. X. Zhou, Multifractal detrending moving-average cross-correlation analysis. Phys. Rev. E 84(2011) 016106.
[12] X. Y. Qian, Y. M. Liu, Z. Q. Jiang, B. Podobnik, W. X. Zhou and H. E. Stanley, Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces. Phys. Rev. E 91(2015) 062816.
[13] N. Yuan, Z. Fu, H. Zhang, L. Piao, E. Xoplaki and J. Luterbacher, Detrended partial-cross-correlation analysis: a new method for analyzing correlations in complex system. Sci. Rep. 5(2015) 8143.
[14] N. Yuan, E. Xoplaki, C. Zhu and J. Luterbacher, Detrended partial-cross-correlation analysis: a new method for analyzing correlations on different time scales in complex system. Sci. Rep. 6(2016) 27707.
[15] J. Shen, and B. Zheng, Cross-correlation in financial dynamics. 86(2009) 48005.
[16] T. Conlon, H. J. Ruskin and M. Crane, Multiscaled cross-correlation dynamics in financial time-series. Adv. Comp. Sys. 12(2009) 439.
[17] L. Y. He and S. P. Chen, Nonlinear bivariate dependency of priceCvolume relationships in agricultural commodity futures markets: a perspective from multifractal detrended cross-correlation analysis. Physica A 390(2011) 297.
[18] W. Shi, P. Shang, J. Wang and A. Lin, Multiscale multifractal detrended cross-correlation analysis of financial time series. Physica A 403(2014) 35.
[19] S. Ghosh, P. Manimaran and P. K. Panigrahi, Characterizing multi-scale self-similar behavior and non-statistical properties of fluctuations in financial time series. Physica A 390(2010) 4304.
[20] Q. Xiao, X. Pan, M. Stephen, Y. Yang, X. Li and H. Yang, Discrete scale-invariance in cross-correlations between time series. Physica A 421(2015) 161.
[21] N. Masuda, Y. Kawamura and H. Kori, Analysis of relative influence of nodes in directed networks. Phys. Rev. E 80(2009) 046114.
[22] H. E. Hurst, Long-term storage capacity of reservoirs. Trans. Ame. Soc. Civ. Eng. 116(1951) 770.
[23] http://www.bts.gov

[24] Yan-Jun Wang, Ya-Kun Cao, Chen-Ping Zhu, Fan Wu, Ming-Hua Hu, Baruch Barzel and H. Eugene Stanley, phys.soc-ph. arXiv:1701.05556(2017).

Figure1

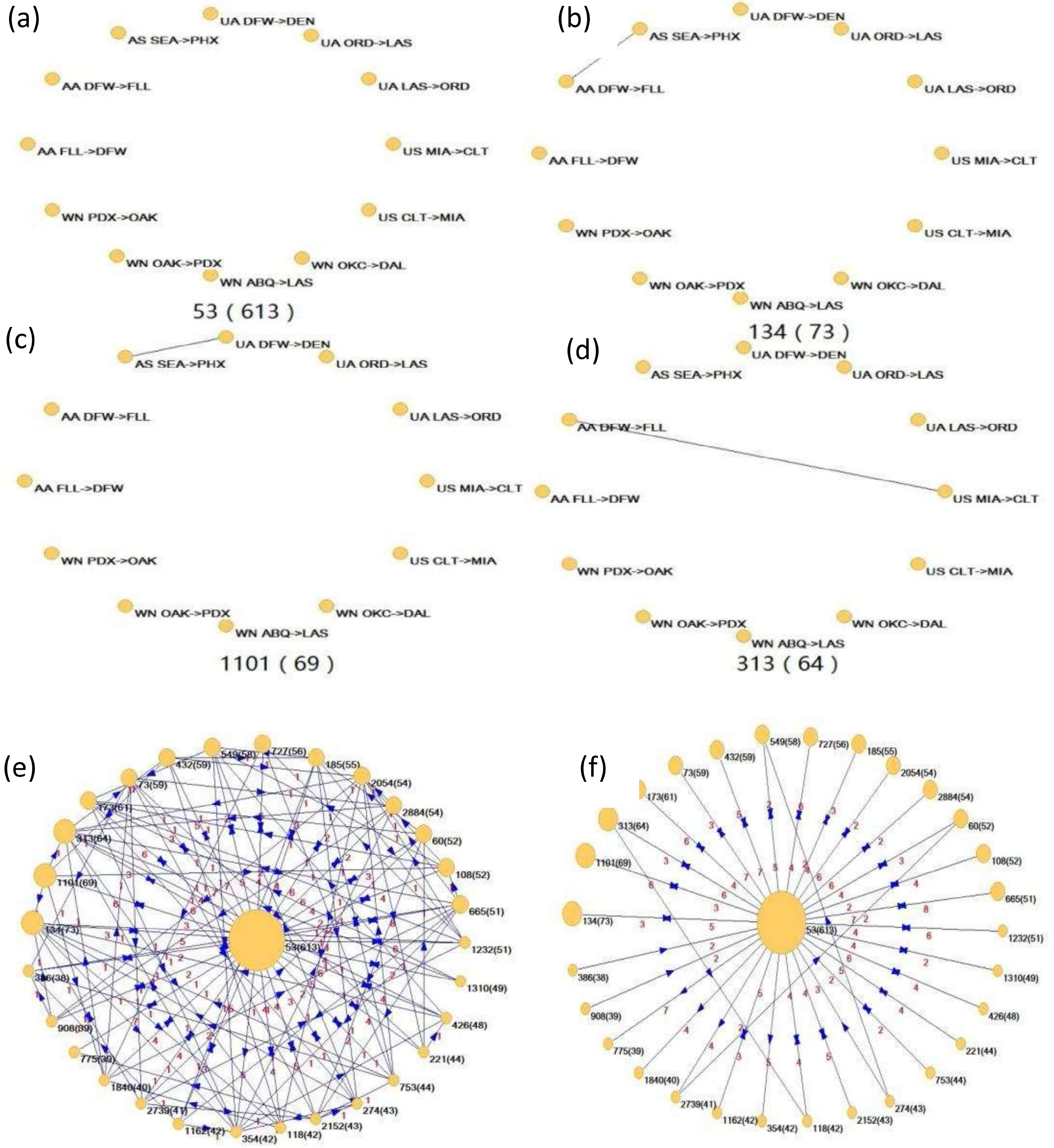
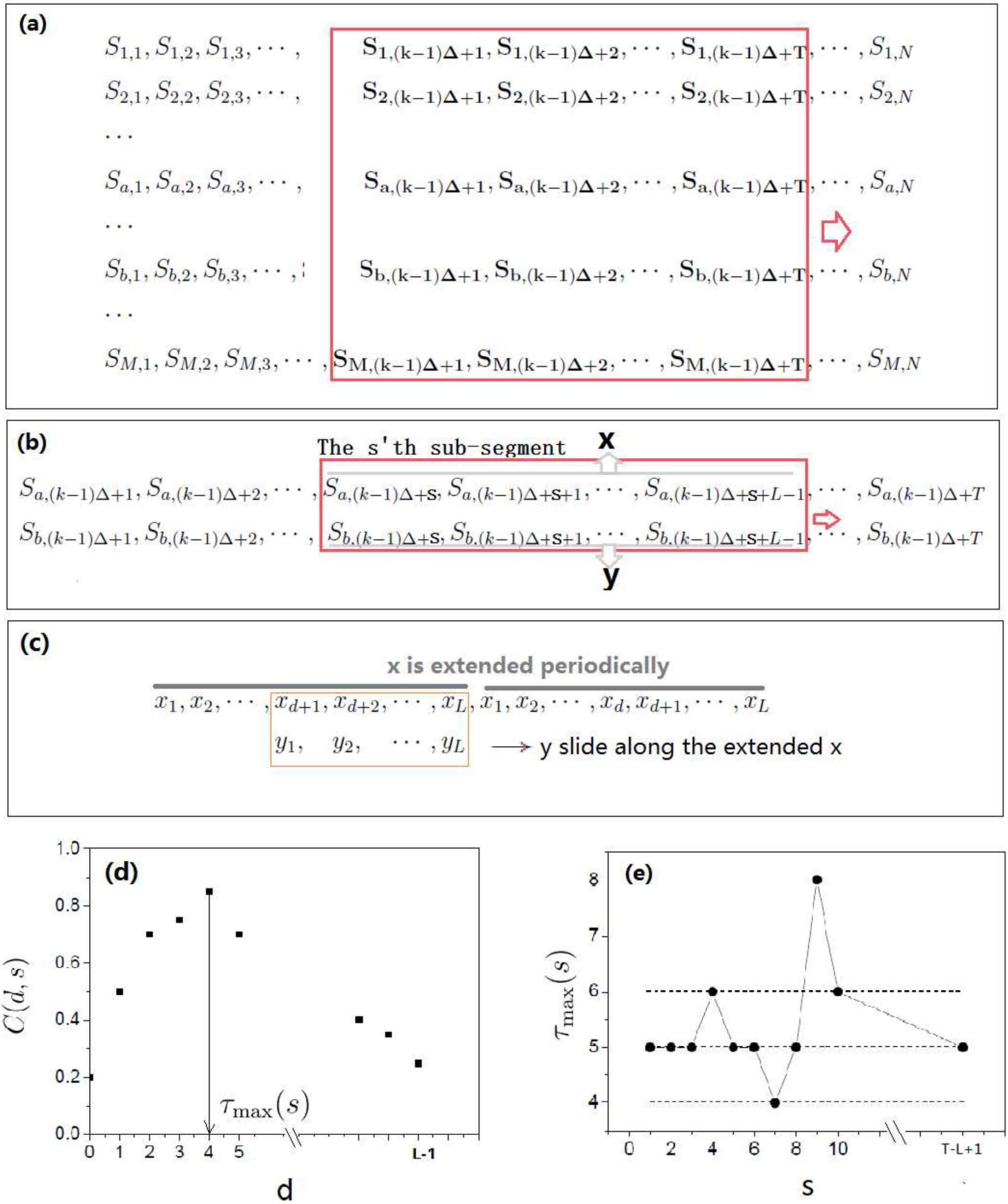
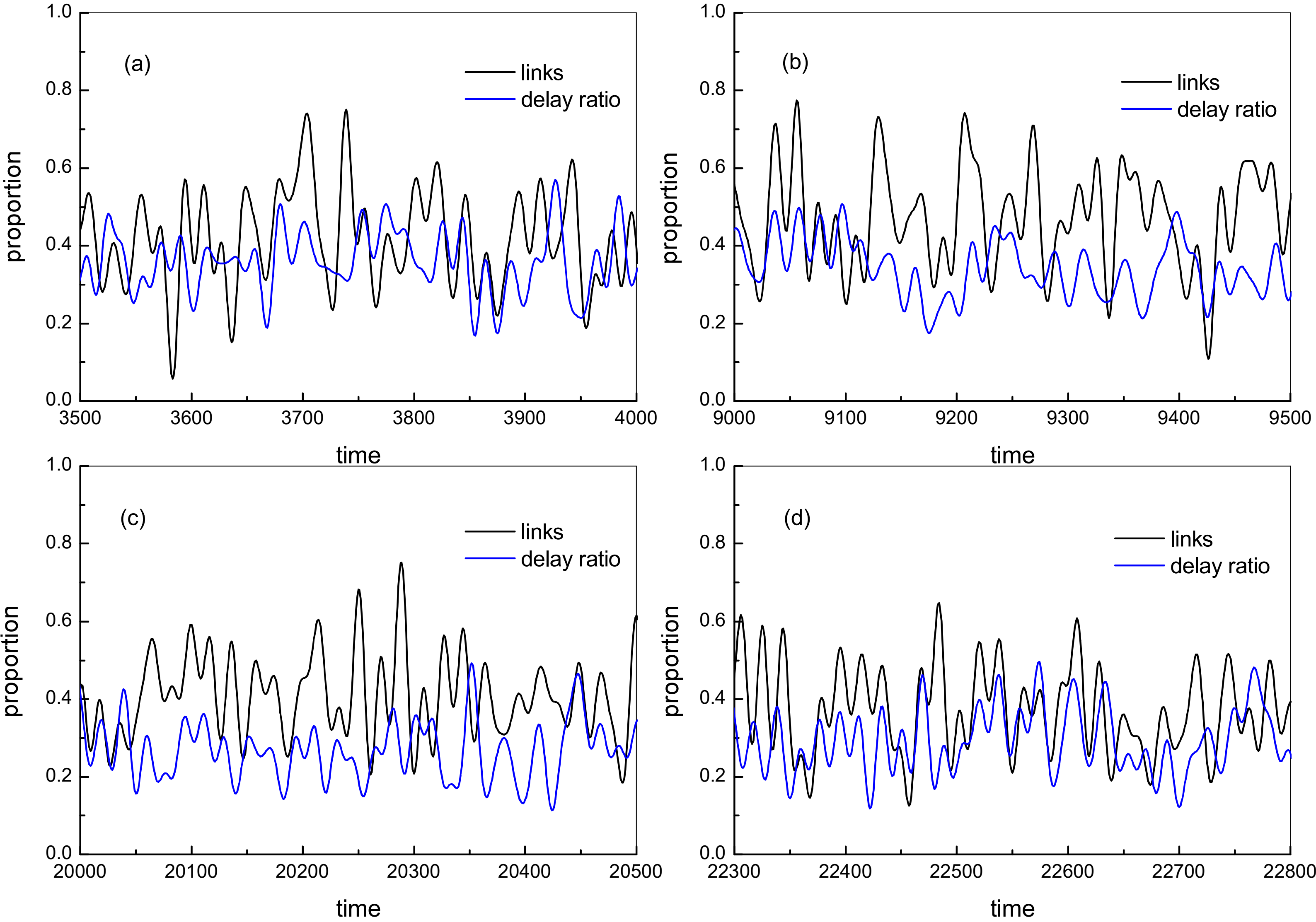

(a)
UA DFW->DEN
AS SEA->PHX
UA ORD->LAS
AA DFW->FLL
UA LAS->ORD
AA FLL->DFW
US MIA->CLT
WN PDX->OAK
US CLT->MIA
WN OAK->PDX
WN OKC->DAL
WN ABQ->LAS
53 (613)
(b)
134 (73)
(c)
1101 (69)
(d)
313 (64)
(e)
(f)

Figure2



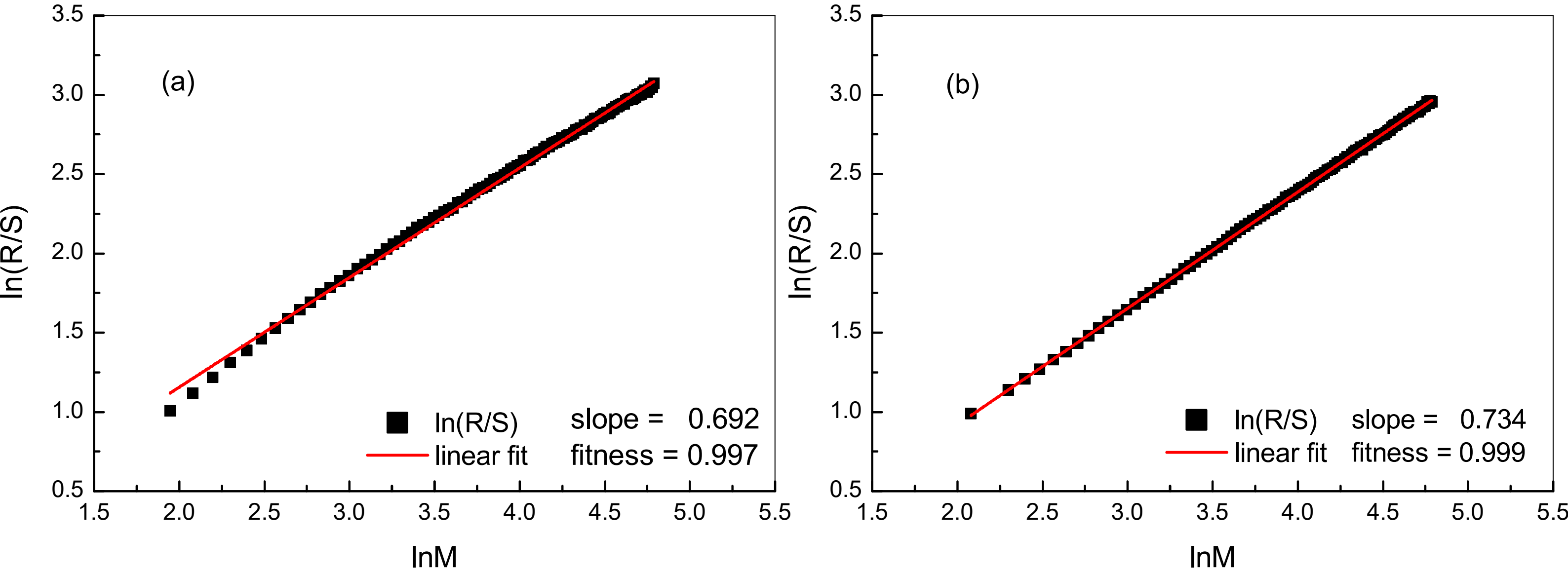
(a)
$S_{1,1}, S_{1,2}, S_{1,3}, \cdots, S_{1,(k-1)\Delta+1}, S_{1,(k-1)\Delta+2}, \cdots, S_{1,(k-1)\Delta+T}, \cdots, S_{1,N}$
$S_{2,1}, S_{2,2}, S_{2,3}, \cdots, S_{2,(k-1)\Delta+1}, S_{2,(k-1)\Delta+2}, \cdots, S_{2,(k-1)\Delta+T}, \cdots, S_{2,N}$
$S_{a,1}, S_{a,2}, S_{a,3}, \cdots, S_{a,(k-1)\Delta+1}, S_{a,(k-1)\Delta+2}, \cdots, S_{a,(k-1)\Delta+T}, \cdots, S_{a,N}$
$S_{b,1}, S_{b,2}, S_{b,3}, \cdots, S_{b,(k-1)\Delta+1}, S_{b,(k-1)\Delta+2}, \cdots, S_{b,(k-1)\Delta+T}, \cdots, S_{b,N}$
$S_{M,1}, S_{M,2}, S_{M,3}, \cdots, S_{M,(k-1)\Delta+1}, S_{M,(k-1)\Delta+2}, \cdots, S_{M,(k-1)\Delta+T}, \cdots, S_{M,N}$
(b)
The s'th sub-segment
x
$S_{a,(k-1)\Delta+1}, S_{a,(k-1)\Delta+2}, \cdots, S_{a,(k-1)\Delta+s}, S_{a,(k-1)\Delta+s+1}, \cdots, S_{a,(k-1)\Delta+s+L-1}, \cdots, S_{a,(k-1)\Delta+T}$
$S_{b,(k-1)\Delta+1}, S_{b,(k-1)\Delta+2}, \cdots, S_{b,(k-1)\Delta+s}, S_{b,(k-1)\Delta+s+1}, \cdots, S_{b,(k-1)\Delta+s+L-1}, \cdots, S_{b,(k-1)\Delta+T}$
y
(c)
x is extended periodically
$x_1, x_2, \cdots, x_{d+1}, x_{d+2}, \cdots, x_L, x_1, x_2, \cdots, x_d, x_{d+1}, \cdots, x_L$
$y_1, y_2, \cdots, y_L$
y slide along the extended x
(d)
$C(d,s)$
$\tau_{\max}(s)$
d
L-1
(e)
$\tau_{\max}(s)$
s
T-L+1

Figure3



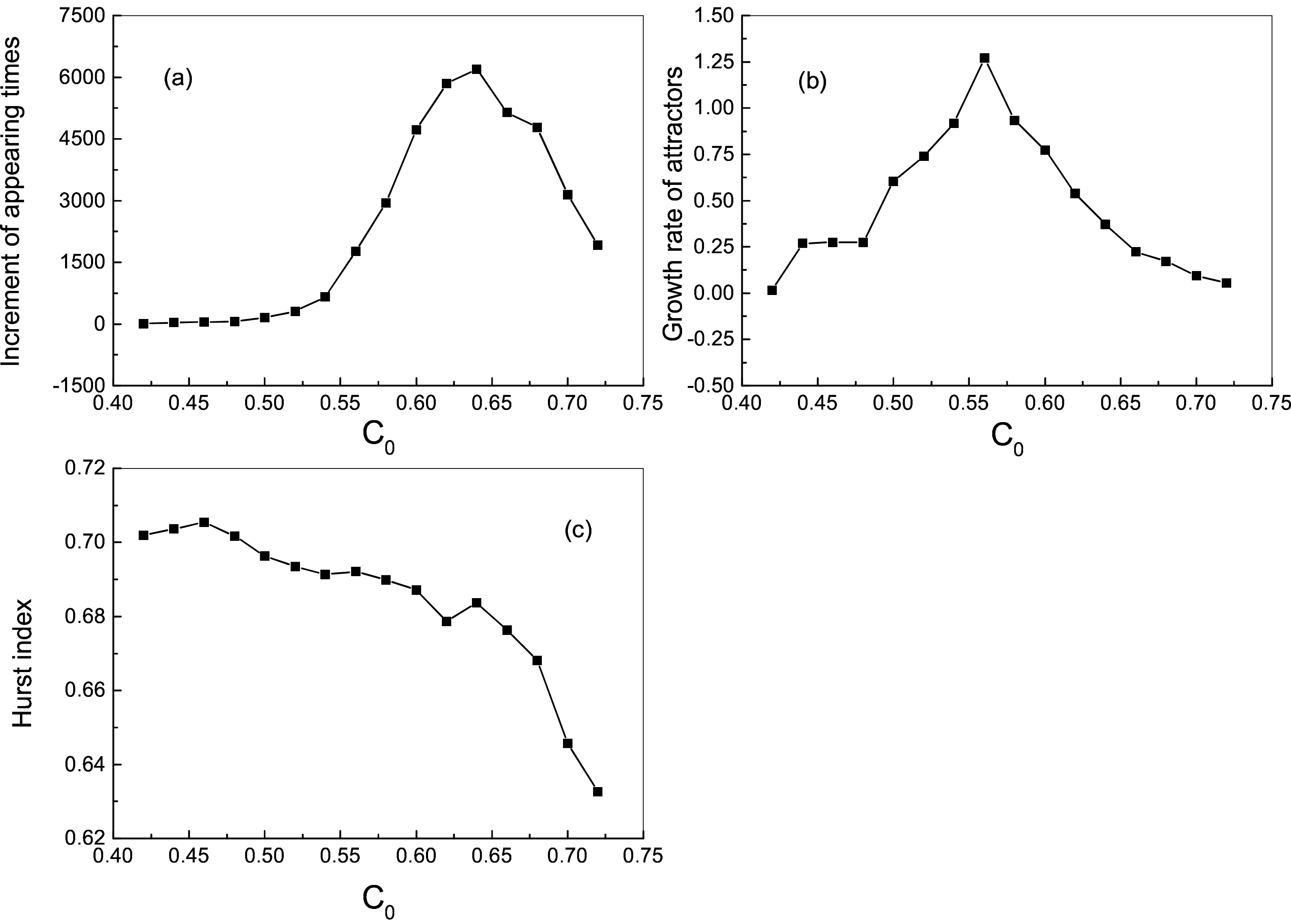
(a)
links
delay ratio
proportion
time
(b)
links
delay ratio
proportion
time
(c)
links
delay ratio
proportion
time
(d)
links
delay ratio
proportion
time

Figure4



3.5
3.0
2.5
2.0
1.5
1.0
0.5
ln(R/S)
(a)
ln(R/S)
linear fit
slope = 0.692
fitness = 0.997
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
lnM
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ln(R/S)
(b)
ln(R/S)
linear fit
slope = 0.734
fitness = 0.999
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
lnM

Figure5



(a)
Increment of appearing times
7500
6000
4500
3000
1500
0
-1500
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
C_0
(b)
Growth rate of attractors
1.50
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
C_0
(c)
Hurst index
0.72
0.70
0.68
0.66
0.64
0.62
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
C_0

Figure6

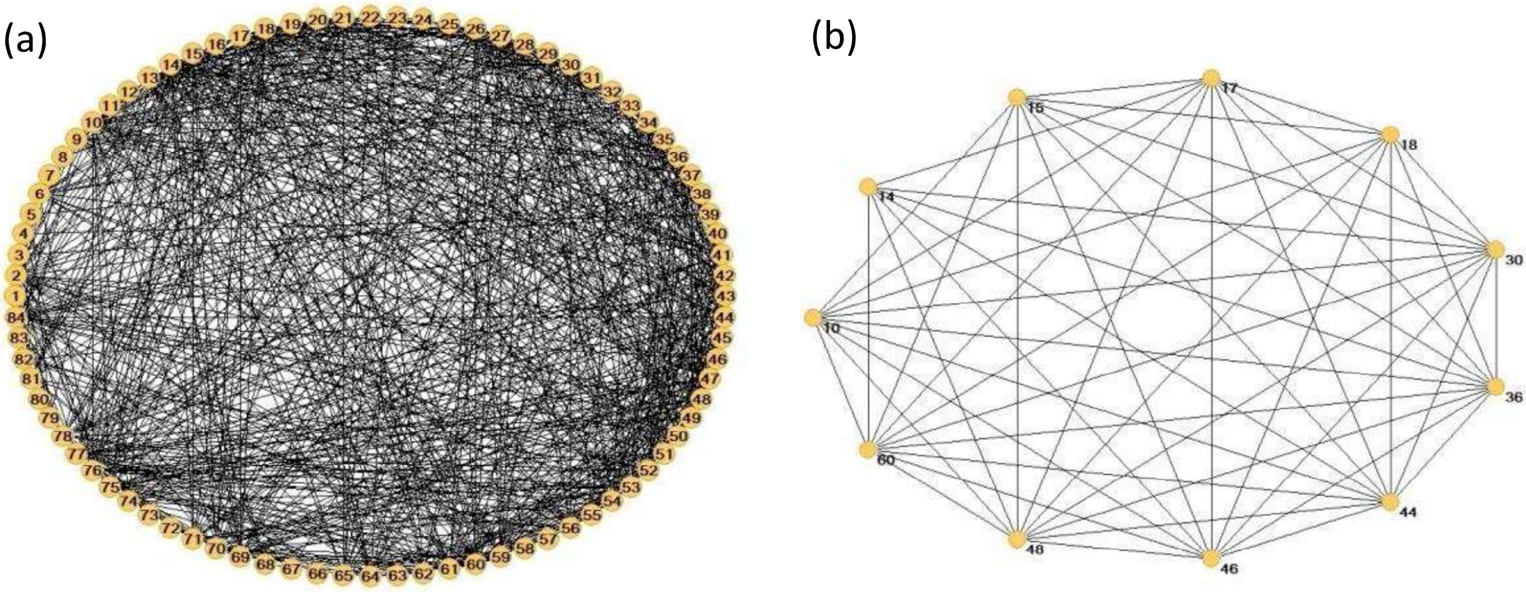


Figure7

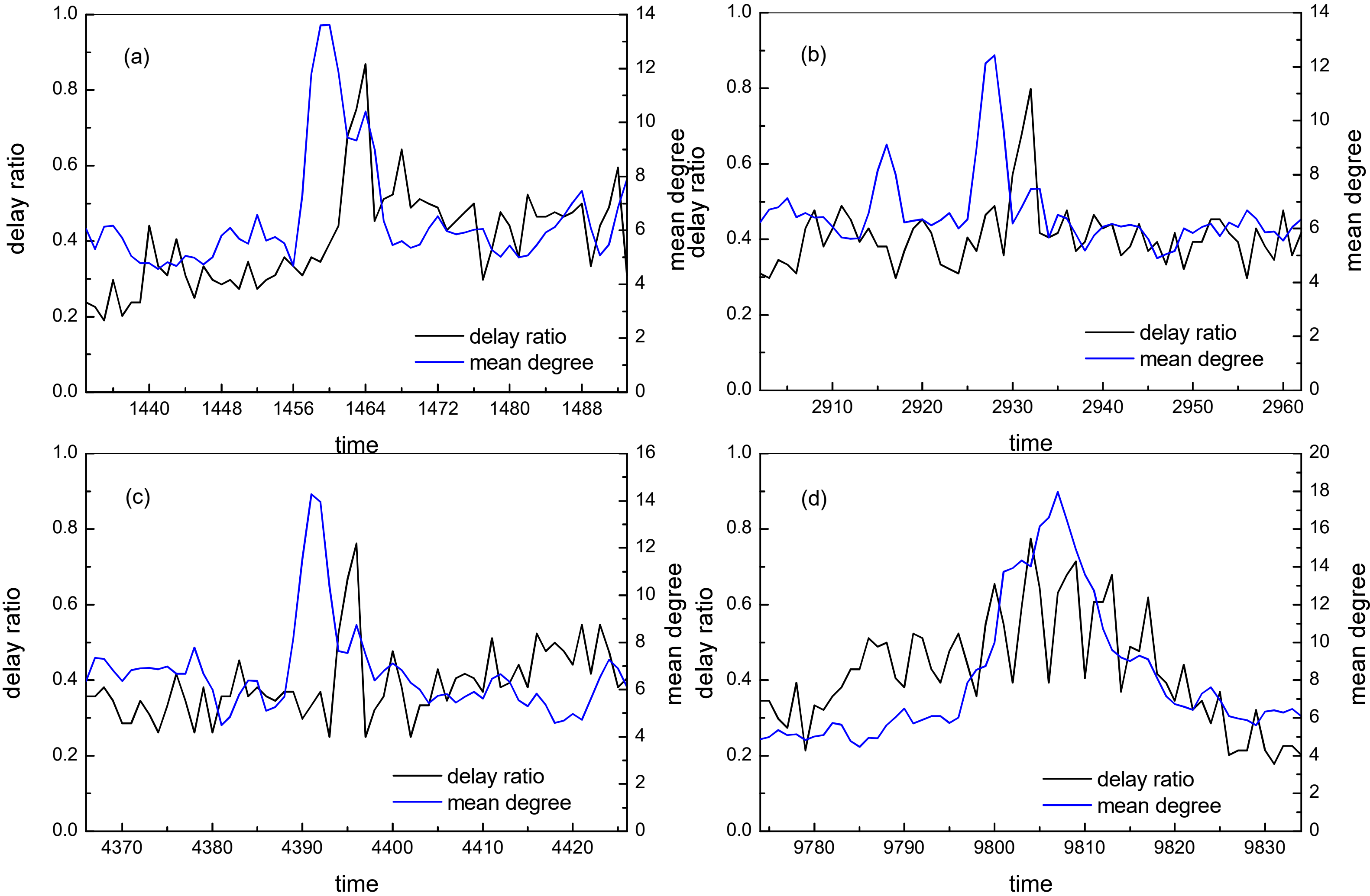


Figure8

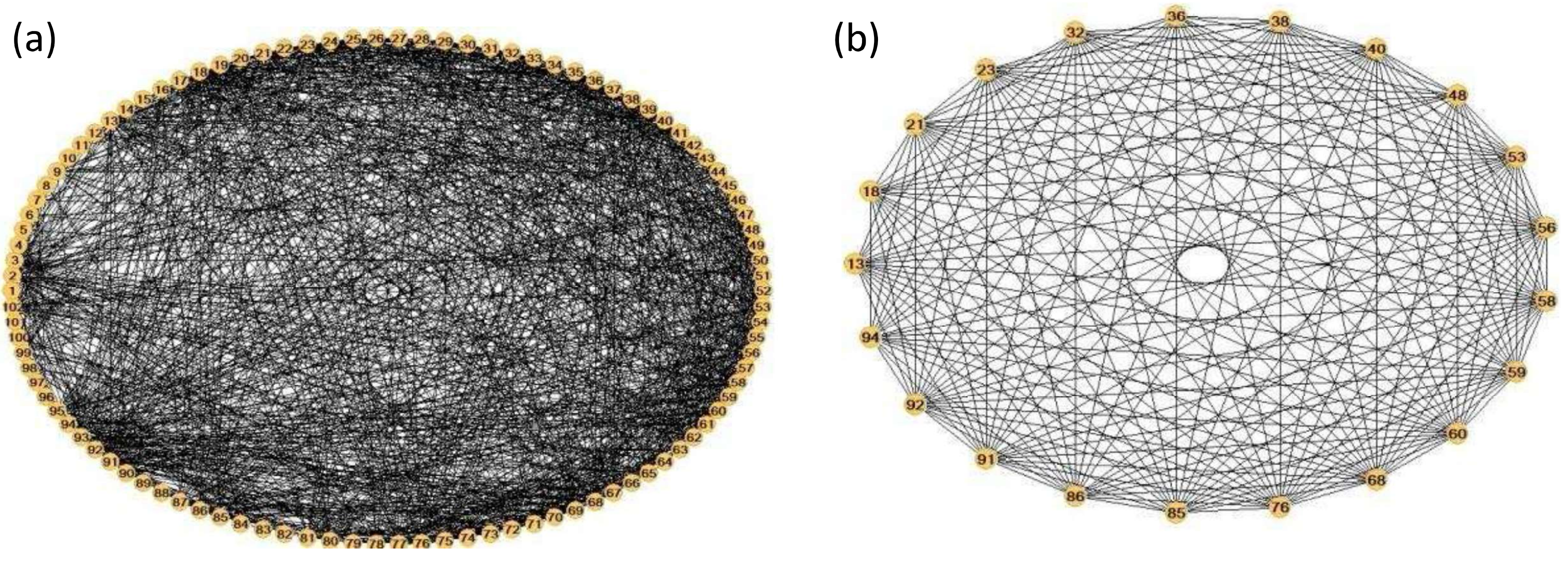


Figure9

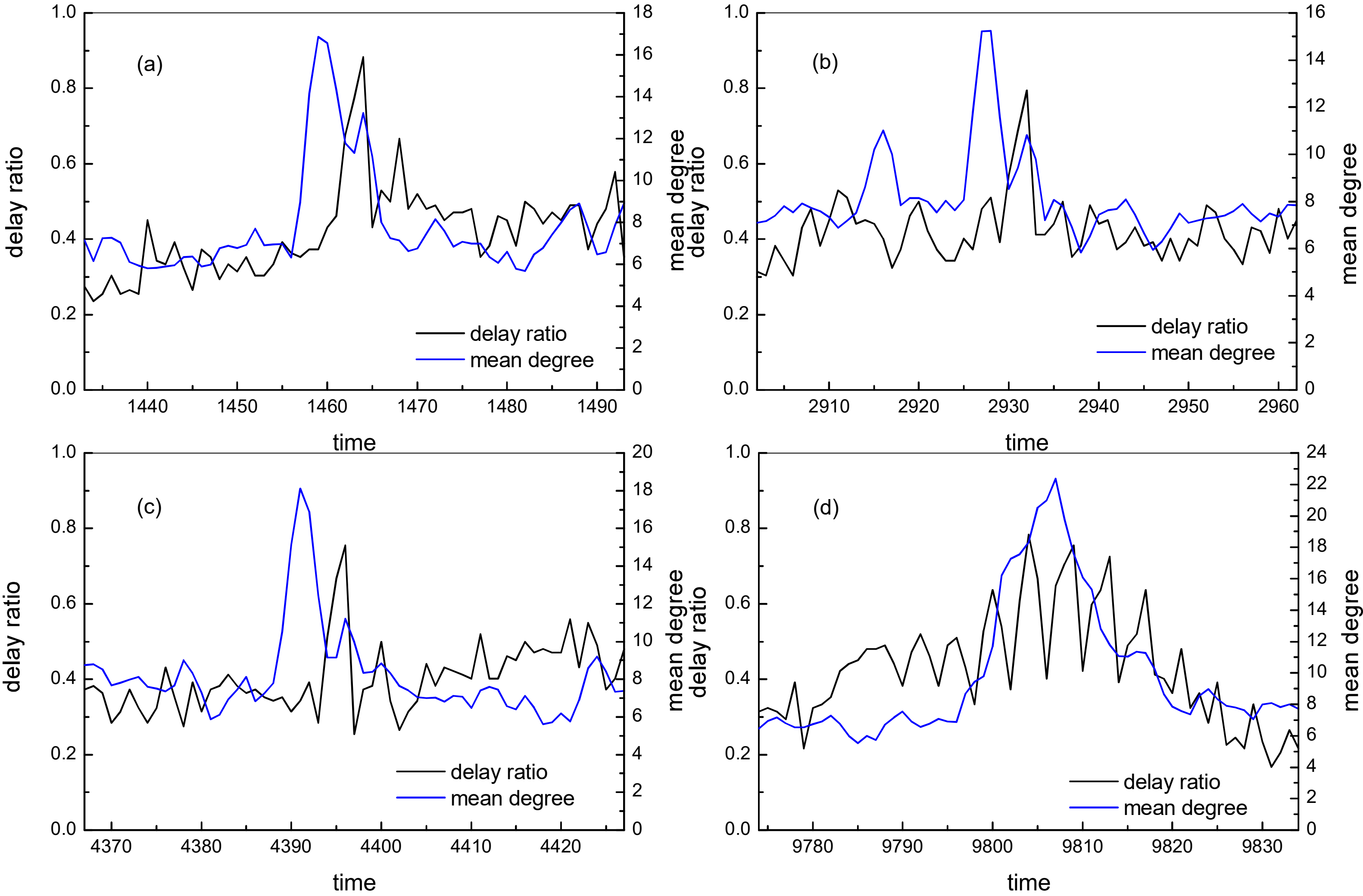
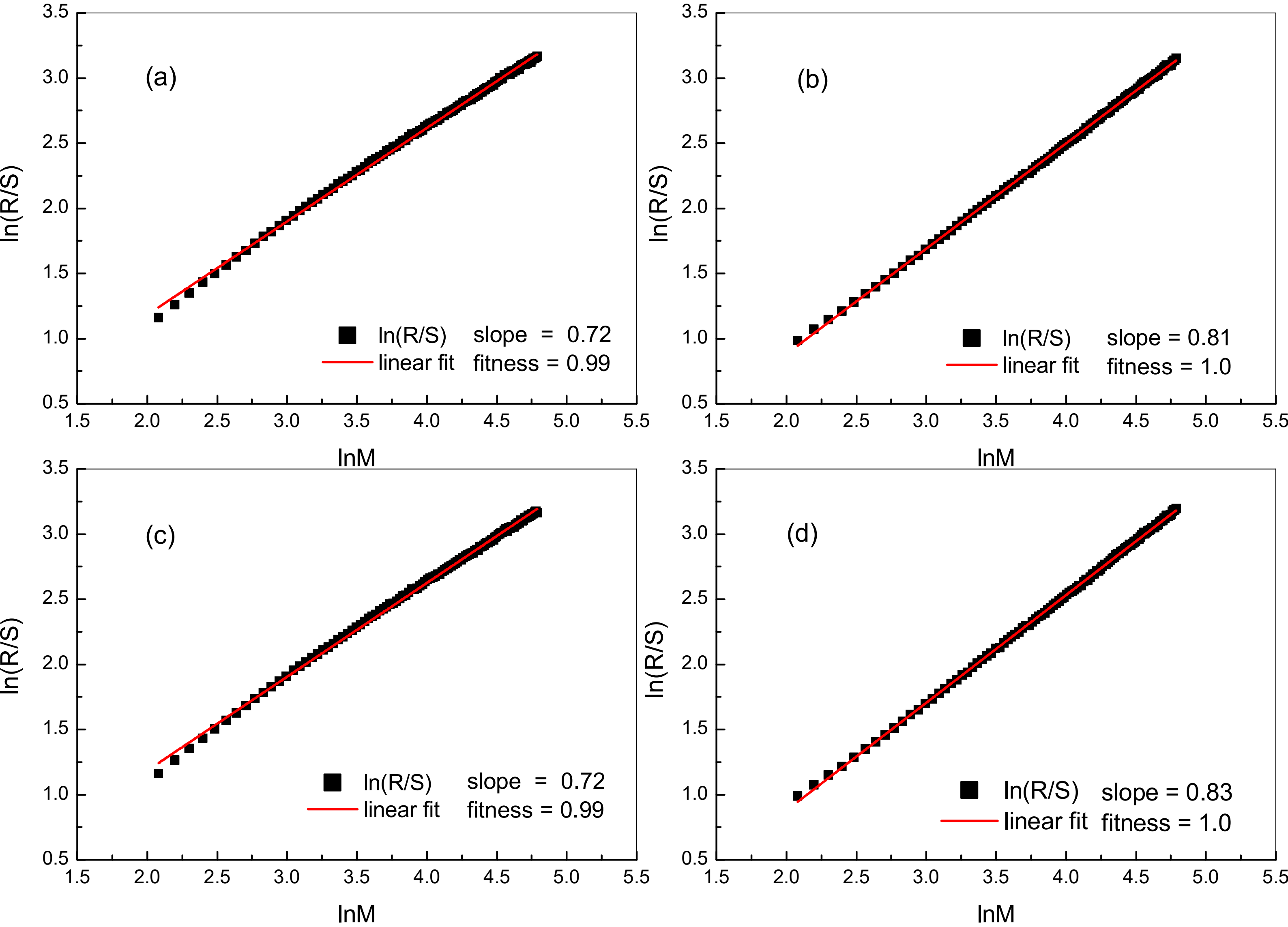

(a)
(b)
(c)
(d)
delay ratio
mean degree
time
1.0
0.8
0.6
0.4
0.2
0.0
1440
1450
1460
1470
1480
1490
2910
2920
2930
2940
2950
2960
4370
4380
4390
4400
4410
4420
9780
9790
9800
9810
9820
9830

Figure10



(a)
ln(R/S) slope = 0.72
linear fit fitness = 0.99
ln(R/S)
lnM
(b)
ln(R/S) slope = 0.81
linear fit fitness = 1.0
ln(R/S)
lnM
(c)
ln(R/S) slope = 0.72
linear fit fitness = 0.99
ln(R/S)
lnM
(d)
ln(R/S) slope = 0.83
linear fit fitness = 1.0
ln(R/S)
lnM

LaTeX Souce Files

Click here to download LaTeX Souce Files: indicator0503J.tex