Complex network approach to characterize the statistical features of the sunspot series

New Journal of Physics
The open access journal at the forefront of physics
Deutsche Physikalische Gesellschaft DPG | IOP Institute of Physics

Complex network approach to characterize the statistical features of the sunspot series

Yong Zou[1,2,6], **Michael Small**[3], **Zonghua Liu**[1] **and Jürgen Kurths**[2,4,5]

[1] Department of Physics, East China Normal University, Shanghai, People's Republic of China
[2] Potsdam Institute for Climate Impact Research, Potsdam, Germany
[3] School of Mathematics and Statistics, University of Western Australia, Crawley, Australia
[4] Department of Physics, Humboldt University Berlin, Berlin, Germany
[5] Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK
E-mail: yzou@phy.ecnu.edu.cn

Received 21 October 2013
Accepted for publication 8 January 2014
Published 30 January 2014

New Journal of Physics **16** (2014) 013051

doi:10.1088/1367-2630/16/1/013051

Abstract

Complex network approaches have been recently developed as an alternative framework to study the statistical features of time-series data. We perform a visibility-graph analysis on both the daily and monthly sunspot series. Based on the data, we propose two ways to construct the network: one is from the original observable measurements and the other is from a negative-inverse-transformed series. The degree distribution of the derived networks for the strong maxima has clear non-Gaussian properties, while the degree distribution for minima is bimodal. The long-term variation of the cycles is reflected by hubs in the network that span relatively large time intervals. Based on standard network structural measures, we propose to characterize the long-term correlations by waiting times between two subsequent events. The persistence range of the solar cycles has been identified over 15–1000 days by a power-law regime with scaling exponent $\gamma = 2.04$ of the occurrence time of two subsequent strong minima. In contrast, a persistent trend is not present in the maximal numbers, although maxima do have significant deviations from an exponential form. Our results suggest some new insights for evaluating existing models.

[6] Author to whom any correspondence should be addressed.

1. Introduction

Solar-cycle prediction, i.e. forecasting the amplitude and/or the epoch of an upcoming maximum, is of great importance as solar activity has a fundamental impact on the weather conditions of the Earth, especially with increasing concern over the various climate change scenarios. However, predictions have been notoriously wayward in the past [1, 2]. There are basically two classes of methods for solar cycle predictions: empirical data-analysis-driven methods and methods based on dynamo models. Most successful methods in this regard can give reasonably accurate predictions only when a cycle is well advanced (e.g. 3 years after the minimum) or with guidance from its past [3, 4]. Hence, these methods show very limited power in forecasting a cycle which has not yet started. The theoretical reproduction of a sunspot series by most current models shows convincingly the 'illustrative nature' of the existing record [5]. However, they generally failed to predict the slow start of the present cycle 24 [6]. One reason cited for this is the emergence of prolonged periods of extremely low activity. The existence of these periods of low activity brings a big challenge for solar-cycle prediction and reconstruction by the two classes of methods described above, and hence prompted the development of special ways to evaluate the appearance of these minima [7]. Moreover, there is increasing interest in the minima since they are known to provide insight for predicting the next maximum [8].

Some earlier authors have both observed and made claims for the chaotic or fractal features of the observed cycles, but the true origin of such features has not yet been fully resolved. For instance, the Hurst exponent has been used as a measure of the long-term memory in time series [9, 10]—an index of long-range dependence that can be often estimated by a rescaled range analysis. The majority of Hurst exponents reported so far for the sunspot numbers are well above 0.5, indicating some level of predictability in the data. Nonetheless, it is not clear whether such characterization is due to an underlying chaotic mechanism or the presence of correlated changes due to the quasi-11-year cycle [2, 11]. By removing the sinusoidal trend of the time series by Fourier truncation in advance, the Hurst exponent is reduced to about 0.12, demonstrating a rather different estimation for the multifractal nature of the fluctuations [12]. This argument should be carefully interpreted since a properly designed adaptive detrending algorithm to remove the 11-year cycle in sunspots yields an estimation of the Hurst exponent to be 0.74 [13]. Anyway, it is the irregularity (including the wide variations in both amplitudes and cycle lengths) that makes the characterization of the possible multifractal features of the solar activity an interesting, challenging and, as yet, unsolved issue. In contrast to the 11-year cycle *per se*, we concentrate on the recently proposed hypothetical long-range memory mechanism on time scales shorter than the quasi-periodic 11-year cycle [14].

In this work, we provide a distinct perspective on the temporal sunspots evolution (in particular the strong maximal activities and quiescent minima) by means of the so-called visibility graph (VG) analysis (as explained in figure 1). Such graphs (mathematical graphs, in the sense of networks) have recently emerged as one alternative to describe various statistical properties of complex systems. In addition to applying the standard method, we generalize the technique further—making it more suitable for studying the observational records of the solar cycles.

2. Data and network construction

Both the international sunspot number (ISN) and the sunspot area (SSA) series [15] are used in this work, and we have obtained consistent conclusions in either case. The lengths of the data

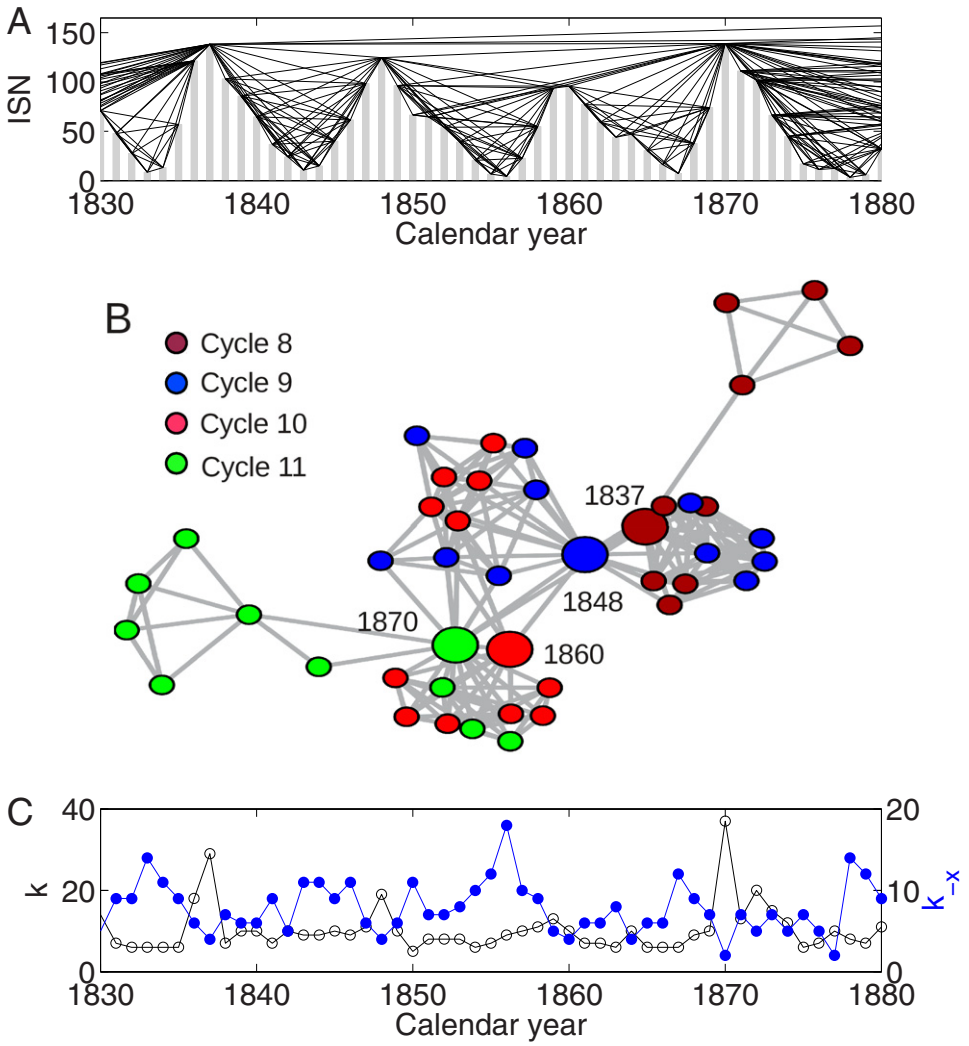

Figure 1. The procedure to construct a VG. Note that an enlargement of the particular time span of the annual data is used here only for visualization purposes, while the analysis performed throughout is based on both monthly and daily series. (A) Sunspot numbers in gray bar plot, where two time points fulfilling the visibility condition are connected by a line; (B) complex network representation, one network cluster usually includes time points of two subsequent solar cycles and (C) degree sequences of the original series $x(t_i)$ (open circles) and the negatively inverted series $-x(t_i)$ (filled circles), respectively.

sets are summarized in table 1. We perform a VG analysis using both monthly and daily sunspot series, which yields, respectively, month-to-month and day-to-day correlation patterns of the sunspot activities. Note that we depict the annual numbers *only* for graphical visualization and demonstration purposes (we use the annual numbers to demonstrate our method—the actual analysis is performed in daily and monthly data). We discuss the results with the ISN (in figures 2 and 3) in the main text and illustrate the results for the SSA (in figures 4 and 5) with notes in the captions. Moreover, we compare our findings based on observational records to the results obtained from data produced by simulations from computational models [16, 17].

Table 1. Temporal resolution and the length of the data sets. Values in parentheses are the number of points of each series. Note that the annual ISN is used *only* for graphical visualization purposes and to provide a reference time interval for models.

	ISN	SSA
Day	1 January 1849–31 October 2011 (59473)	1 May 1874–31 October 2011 (50222)
Month	January 1849–October 2011 (1944)	May 1874–October 2011 (1650)
Year	1700–2010 (311)	Not used

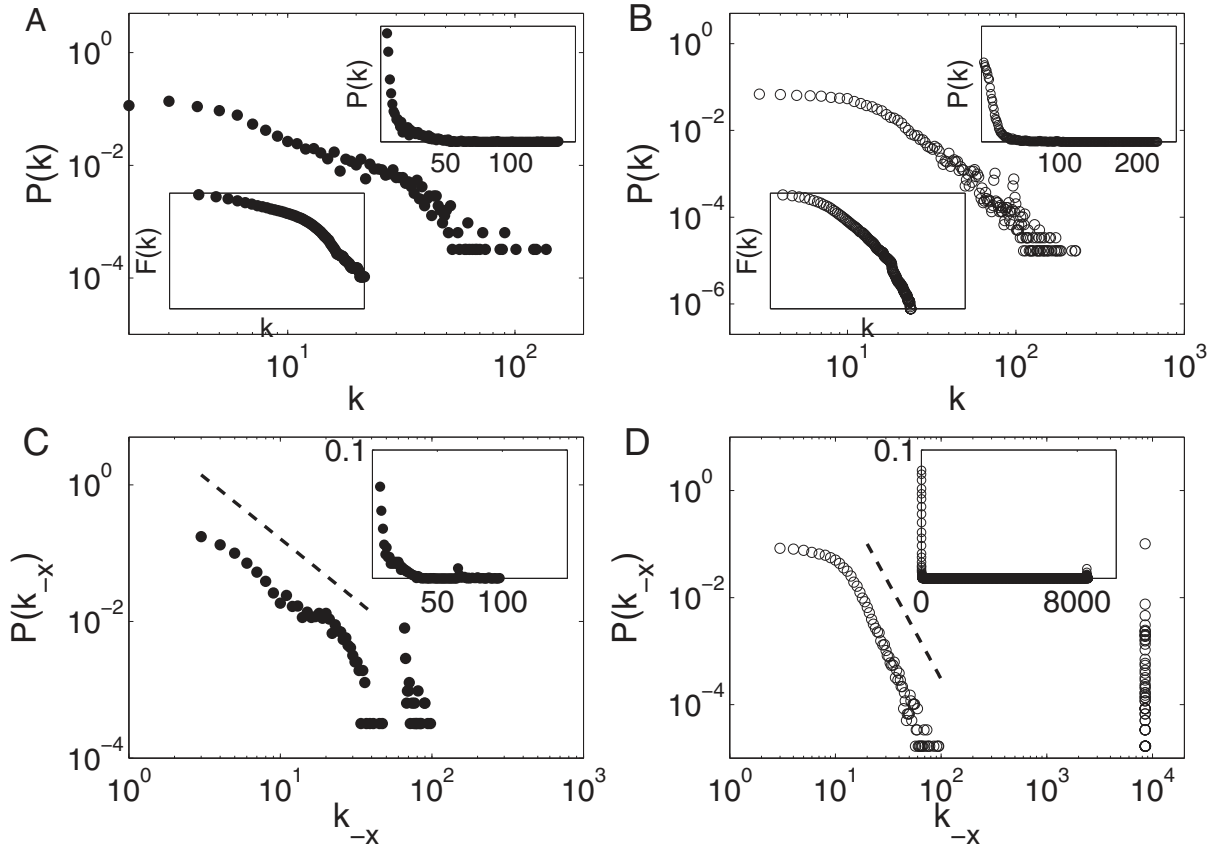

Figure 2. Degree distribution $p(k)$ of VGs from monthly (A), (C) and daily data (B), (D). (A), (B) are for $x(t_i)$ and (C), (D) for $-x(t_i)$. Upper insets of all plots are $p(k)$ in linear scale, while the lower insets of (A), (B) show cumulative distribution $F(k)$ in double logarithmic scale, where a straight line is expected if $F(k)$ followed a power law $\sim k^{-(\gamma-1)}$. In (A), (B) γ could be suspected to be in the range [2.32, 2.64], while a fit to the first part of $p(k_{-x})$ yields that the slope of the dashed line in (C) is 1.79, and that in (D) is 3.61, *but* all p-values are 0, rejecting the hypothetical power laws.

Recently a variety of methods has been proposed for studying time series from a complex networks viewpoint, providing us with many new and distinct statistical properties [18–22]. In this work, we restrict ourselves to the concept of the VG, where individual observations are

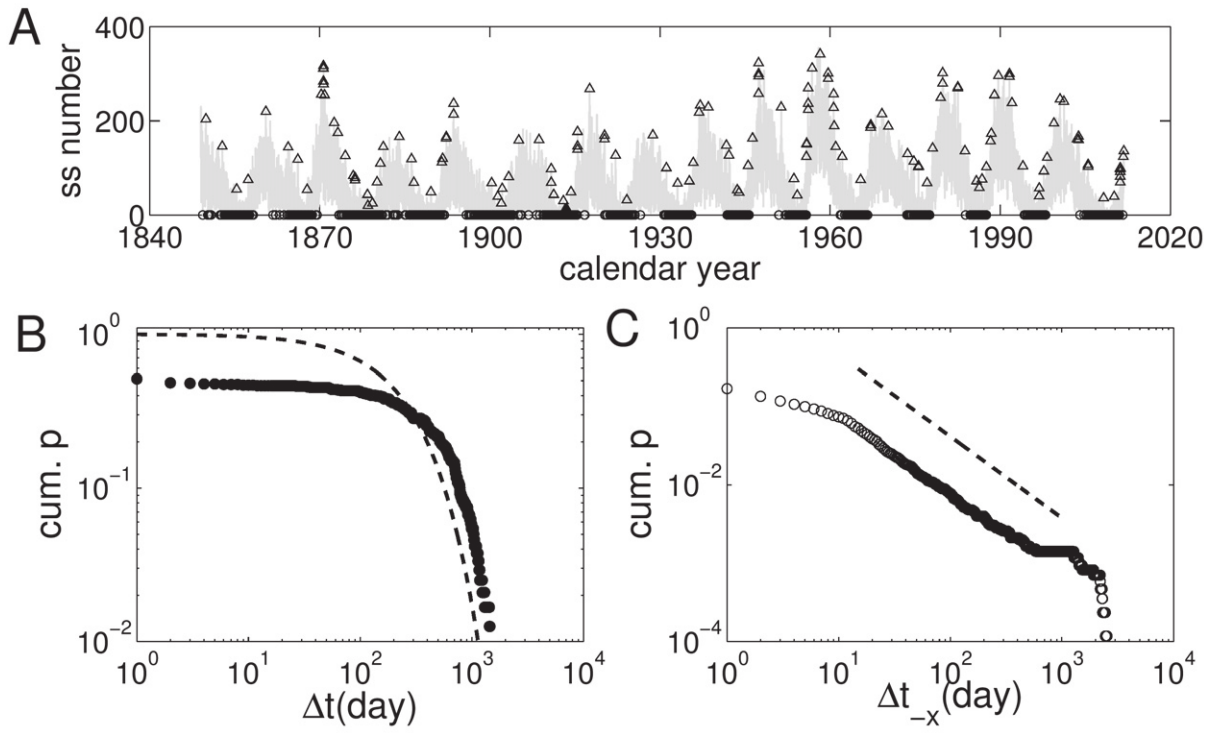

Figure 3. (A) Time points annotated by vertices with large degrees. Triangles (strong maxima) correspond to the VG constructed from the original series $x(t_i)$, while circles (strong minima) are for the VG from $-x(t_i)$. (B), (C) Cumulative probability distribution of the time intervals between subsequent strong maxima (B) and minima (C). A cumulative exponential distribution is plotted as a dashed line in (B), while the dashed line in (C) is a linear fit with slope equal to $\gamma - 1 = 1.04$. The corresponding p-value of (C) is 0.45, indicating that the power law is a plausible hypothesis for the waiting times of strong minima.

considered as vertices and edges are introduced whenever vertices are visible. More specifically, given a univariate time series $x(t_i)_{i=1,\ldots,N}$, we construct the 0–1 binary adjacency matrix $A_{N\times N}$ of the network. The algorithm for deciding non-zero entries of $A_{i,j}$ considers two time points t_i and t_j as being mutually connected vertices of the associated VG if the following criterion

$$\frac{x(t_i) - x(t_k)}{t_k - t_i} > \frac{x(t_i) - x(t_j)}{t_j - t_i} \tag{1}$$

is fulfilled for all time points t_k with $t_i < t_k < t_j$ [23]. Therefore, the edges of the network take into account the temporal information explicitly. By default, two consecutive observations are connected and the graph forms a completely connected component without disjoint subgraphs. Furthermore, the VG is not affected by choice of algorithmic parameters—most other methods of constructing complex networks from time series data are dependent on the choice of some parameters (e.g. the threshold ε of recurrence networks, see more details in [21]). While the inclusion of these parameters makes these alternative schemes more complicated, they do gain the power to reconstruct (with sufficient data) the underlying dynamical system. For the current discussion, and with the limited data we consider here, we prefer the simplicity of the VG.

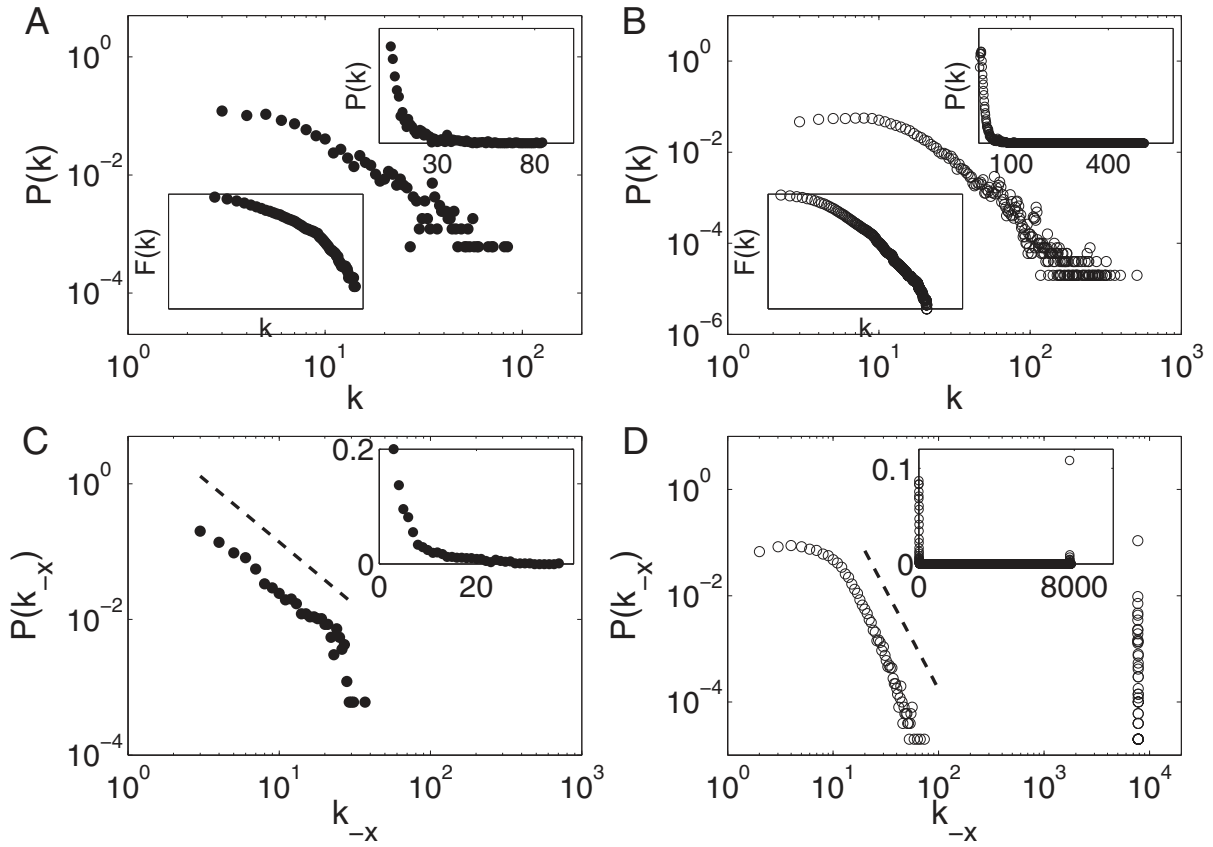

Figure 4. For the SSA data. Degree distribution $p(k)$ of VGs from monthly (A), (C) and daily data (B), (D). (A), (B) are for $x(t_i)$ and (C), (D) for $-x(t_i)$. Upper insets of all plots are $p(k)$ in linear scale, while lower insets of (A), (B) show cumulative distribution $F(k)$ in double logarithmic scale, where a straight line is expected if $F(k)$ followed a power law $\sim k^{-(\gamma-1)}$. In (C), (D), γ could be fitted by dashed lines (slopes: C, 1.86; D, 3.73 respectively), *but* all p-values are 0 rejecting the hypothetical power laws.

The VG approach is particularly interesting for certain stochastic processes where the statistical properties of the resulting network can be directly related to the fractal properties of the time series [24–28].

Figure 1 illustrates an example of how we construct a VG for the sunspot time series. It is well known that the solar cycle has an approximately 11-year period, which shows that most of the temporal points of the decreasing phase of one solar cycle are connected to those points of the increasing phase of the next cycle (figure 1(A)). Therefore, the network is clustered into communities, each of which mainly consists of the temporal information of two subsequent solar cycles (figure 1(B)). When the sunspot number reaches a stronger but more infrequent extreme maximum, we have inter-community connections, since they have better visibility contact with more neighbors than other time points—hence, forming hubs in the graph. The inter-community connections extend over several consecutive solar cycles encompassing the temporal cycle-to-cycle information.

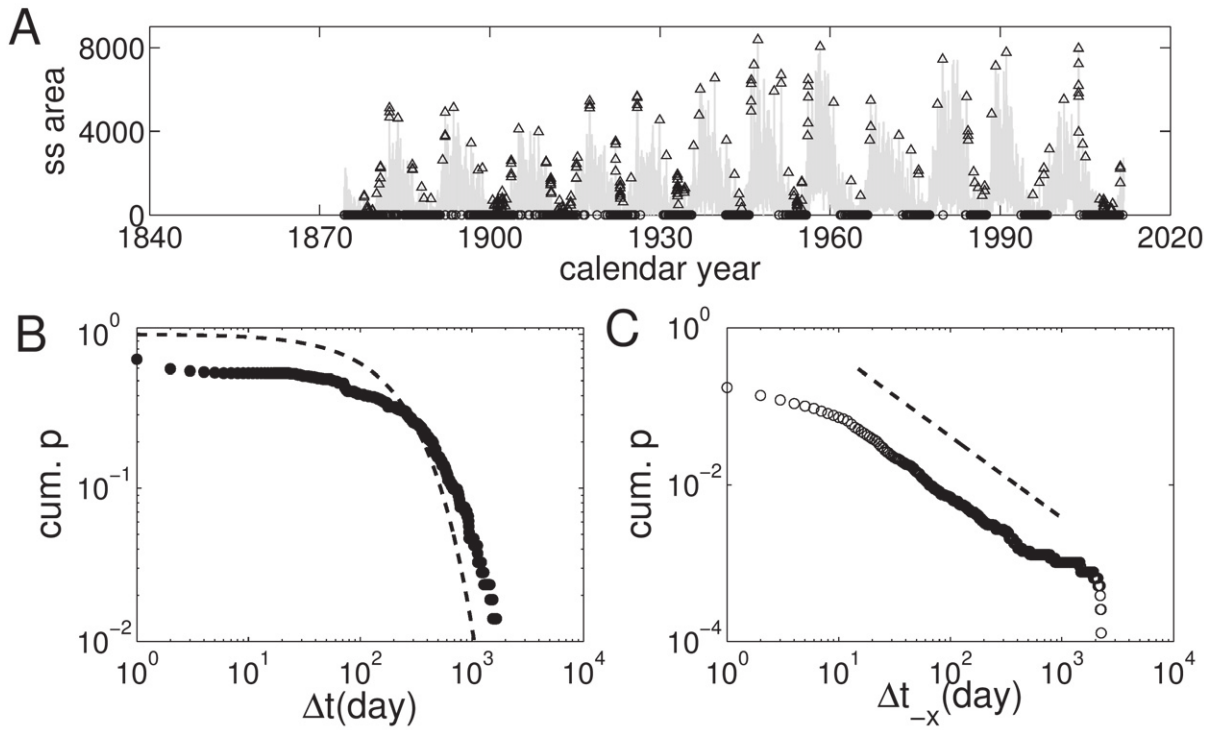

Figure 5. For the SSA data. (A) Time points annotated by vertices with large degrees. Triangles (strong maxima) correspond to the VG constructed from the original series $x(t_i)$, while circles (strong minima) are that of the VG from $-x(t_i)$. (B), (C) Cumulative probability distribution of the time intervals between subsequent strong maxima (B) and minima (C). A cumulative exponential distribution is plotted as a dashed line in (B), while the dashed line in (C) is a linear fit with slope equal to $\gamma - 1 = 1.04$. The corresponding p-value of (C) is 0.78, indicating that the power law is a plausible hypothesis for the waiting times of strong minima.

Depending on various notions of 'importance' of a vertex with respect to the entire network, various centrality measures have been proposed to quantify the structural characteristics of a network (cf [29]). Recent work on VGs has mainly concentrated on the properties of the degree and its probability distribution $p(k)$, where degree k measures the number of direct connections that a randomly chosen vertex i has, namely, $k_i = \sum_j A_{i,j}$. The degree sequence reflects the maximal visibility of the corresponding observation in comparison with its neighbors in the time series (figure 1(C)). Based on the variation of the degree sequence k_i, we consider 1837, 1848, 1860 and 1870 as hubs of the network, which can be used to identify the approximately 11-year cycle reasonably well (figure 1(B)).

Furthermore, in the case of sunspot time series, one is often required to investigate what contributions local minimum values make to the network—something that has been largely overlooked by the traditional VGs. One simple solution is to study the negatively inverted counterpart of the original time series, namely, $-x(t_i)$, which quantifies the properties of the local minima. We use k_{-x} and $p(k_{-x})$ to denote the case of $-x(t_i)$. Here, we remark that this simple inversion of the time series allows us to create an entirely different complex network—this is because the VG algorithm itself is extremely simple and does not attempt to reconstruct an underlying dynamical system. As shown in figure 1(C), k_{-x} captures the variation

of the local minima rather well. We will use this technique later to understand the long-term behavior of strong minima of the solar cycles.

The degree distribution $p(k)$ is defined to be the fraction of nodes in the network with degree k. Thus if there are N nodes in total in a network and n_k of them have degree k, we have $p(k) = n_k/N$. For many networks from various origins, $p(k)$ has been observed to exhibit power-law behavior: $p(k) \sim k^{-\gamma}$. In the case of VGs, $p(k)$ is related to the dynamical properties of the underlying processes [23]. More specifically, for a periodic signal, $p(k)$ consists of several distinct degrees indicating the regular structure of the series; for white noise, $p(k)$ has an exponential form; for fractal processes, the resulting VGs often have power law distributions $p(k) \sim k^{-\gamma}$ with the exponent γ being related to the Hurst exponent H of the underlying time series [24]. It is worth pointing out that when one seeks to estimate the exponent γ it is often better to employ the cumulative probability distribution $F(k) = \sum_{k>k_0} p(k)$ so as to have a more robust statistical fit.

Estimating the exponent γ of the hypothetical power law model for the degree sequence of VG can be done rather straightforwardly, but, the statistical uncertainties resulting from the observability of sunspots are a challenge for reliable interpretation. Meanwhile, fitting a power law to empirical data and assessing the accuracy of the exponent γ is a complicated issue. In general, there are many examples which claimed to have power laws but turned out not to be statistically justifiable. We apply the recipe of [30] to analyze the hypothetical power-law distributed data, namely, (i) estimating the scaling parameter γ by the maximum likelihood and (ii) generating a p-value that quantifies the plausibility of the hypothesis by the Kolmogorov–Smirnov statistic.

Aside from the aforementioned degree k_i and degree distribution $p(k)$, in appendix A, some alternative higher order network measures are suggested which may also be applied to uncover deeper dynamical properties of the time series from the VG.

3. Results

Figures 2(A) and (B) show the degree distributions $p(k)$ of the VGs derived from the ISN $x(t_i)$ with heavy-tails corresponding to hubs of the graph, which clearly deviates from Gaussian properties. In contrast, $p(k_{-x})$ of the negatively inverted sunspot series $-x(t_i)$ shows a completely different distribution, consisting of a bimodal property (figures 2(C) and (D), extra large degrees are at least two orders of magnitude larger than most of the vertices (figure 2(D)).

Since well-defined scaling regimes are absent in either $p(k)$ or $p(k_{-x})$ (nor do they appear in the cumulative distributions as shown in the insets, see caption of figure 2 for details of the statistical tests that we apply), we may reject the hypothetical power laws—in contrast to what has been reported in other contexts [23, 24].

In the context of studying grand maxima/minima over (multi-)millennial timescale using some particular indirect proxy time series, the main idea lies in an appropriately chosen threshold, excursions above which are defined as grand maxima, respectively, below which are defined as grand minima [31, 32]. In this work, we use a similar concept but here in terms of degrees of the corresponding VG. We define a strong maximum if its degree k is larger than 100 (figures 2(A) and (B)), a strong minimum if its degree k_{-x} is over 1000 (figures 2(C) and (D)). Note that, in general, our definition of strong maxima/minima coincides *neither* with the local maximal/minimal sunspot profiles, since degree k takes into account the longer term inter-cycle variations, *nor* with those defined for an individual cycle. Our definition avoids the choice

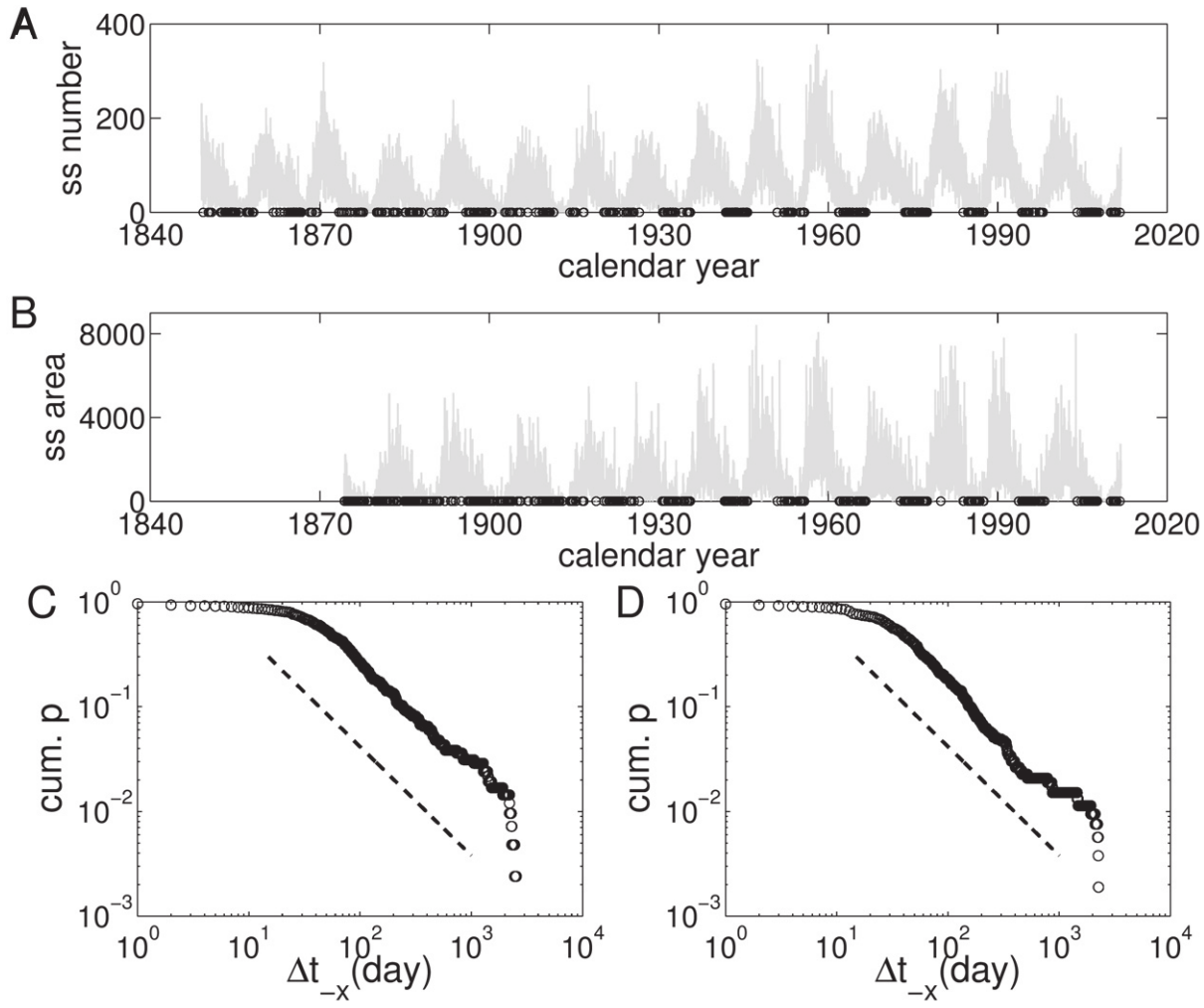

Figure 6. Time points annotated by vertices with strong minima, which are defined by network degrees ($k > 7000$) of VGs reconstructed from negatively inverted series. (A) ISN and (B) SSA. Cumulative probability distribution of the time intervals between subsequent strong minima, (C) ISN and (D) SSA. The dashed lines in (C), (D) are linear fits, which are obtained in the same way as shown in figures 3(C) and 5(C), respectively.

of maximum/minimum for one cycle, which suffers from moving-average effects (e.g. [1]). In contrast, our results below are robust with respect to the choice of the threshold degrees, especially in the case of the definition of strong minima, since large degrees are very well separated from others. The gray line in figure 3(A) shows the sunspot numbers overlaid by the maxima/minima identified by the large degrees. We find that the positions of strong maxima are largely homogeneously distributed over the time domain, while that of the strong minima are much more clustered in the time axis—although irregularly (figure 3(A)). We emphasize that the clustering behavior of the strong minima on the time axis as shown in figures 3(A) and 5(A) does *not* change if threshold degrees are varied in the interval of (200, 8000) to define strong minima, i.e. $k > 7000$, as shown in figures 6(A) and (B). Therefore, the bimodality as observed in figures 2(C) and (D) and 4(C) and (D) is not due to the finite size effects of time series.

The hidden regularity of the time positions of maxima/minima can be further characterized by the waiting time distribution [32]: the interval between two successive events is called

the waiting time. The statistical distribution of waiting-time intervals reflects the nature of a process that produces the studied events. For instance, an exponential distribution is an indicator of a random memoryless process, where the behavior of a system does not depend on its preceding states on both short or long time scales. Any significant deviation from an exponential law suggests that the underlying event occurrence process has a certain level of temporal dependence. One representative of the large class of non-exponential distributions is the power laws, which have been observed in many different contexts, ranging from the energy accumulation and release property of earthquakes to social contacting patterns of humans [33].

In the framework of VGs, the possible long temporal correlations are captured by edges that connect different communities (the increasing and decreasing phases of one solar cycle belong to two temporal consecutive clusters). As shown in figure 3(B), the distribution of the waiting time between two subsequent maxima sunspot deviates significantly from an exponential function, although the tail part could be an indicator of an exponential form. In contrast, we show in figure 3(C) that the waiting times between subsequent strong minima have a heavy-tail distribution where the exponent $\gamma \approx 2.04$ is estimated in the scaling regime. This suggests that the process of the strong minima has a positive long term correlation, which might be well developed over the time between 15 and 1000 days where a power-law fit is taken. Waiting time intervals outside this range are due to either noise effects on shorter scales or the finite length of observations on longer time scales. Again, the power law regimes identified by the waiting time distributions are robust for various threshold degree values in the interval (200, 8000) (for instance, the case of $k > 7000$ is shown in figures 6(C) and (D)).

4. Discussion

In contrast to the computations described above with observational data, we now demonstrate the inadequacy of two models of solar cycles. By applying the VG methods to model simulations we demonstrate that the observed data have, according to the complex network perspective, features absent in the models. We first choose a rather simple yet stochastic model that describes the temporal complexity of the problem, the Barnes model [16], consisting of an autoregressive moving average ARMA(2, 2) model with a nonlinear transformation

$$z_i = \alpha_1 z_{i-1} + \alpha_2 z_{i-2} + a_i - \beta_1 a_{i-1} - \beta_2 a_{i-2}, \tag{2}$$

$$s_i = z_i^2 + \gamma (z_i^2 - z_{i-1}^2)^2, \tag{3}$$

where $\alpha_1 = 1.906\,93$, $\alpha_2 = -0.987\,51$, $\beta_1 = 0.785\,12$, $\beta_2 = -0.406\,62$, $\gamma = 0.03$ and a_i are identically independent distributed Gaussian random variables with zero mean and standard deviation SD $= 0.4$. The second model is a stochastic relaxation Van der Pol oscillator, which is obtained from a spatial truncation of the dynamo equations [17]. The equations read

$$\dot{x} = y, \tag{4}$$

$$\dot{y} = -\omega^2 x - \mu y[3(\xi_0 + r\xi_s)x^2 - 1], \tag{5}$$

where $\omega = 0.2993$, $\mu = 0.2044$, $\xi = 0.0102$, ξ_s is Gaussian noise with zero mean and SD $= 1$, and r is adjustable but often chosen to be 0.02. The variable x is associated with the mean toroidal magnetic field, and therefore the sunspot number is considered to be proportional to x^2, which prompts us to construct VGs from x^2 ($-x^2$ respectively). Both models reproduce the

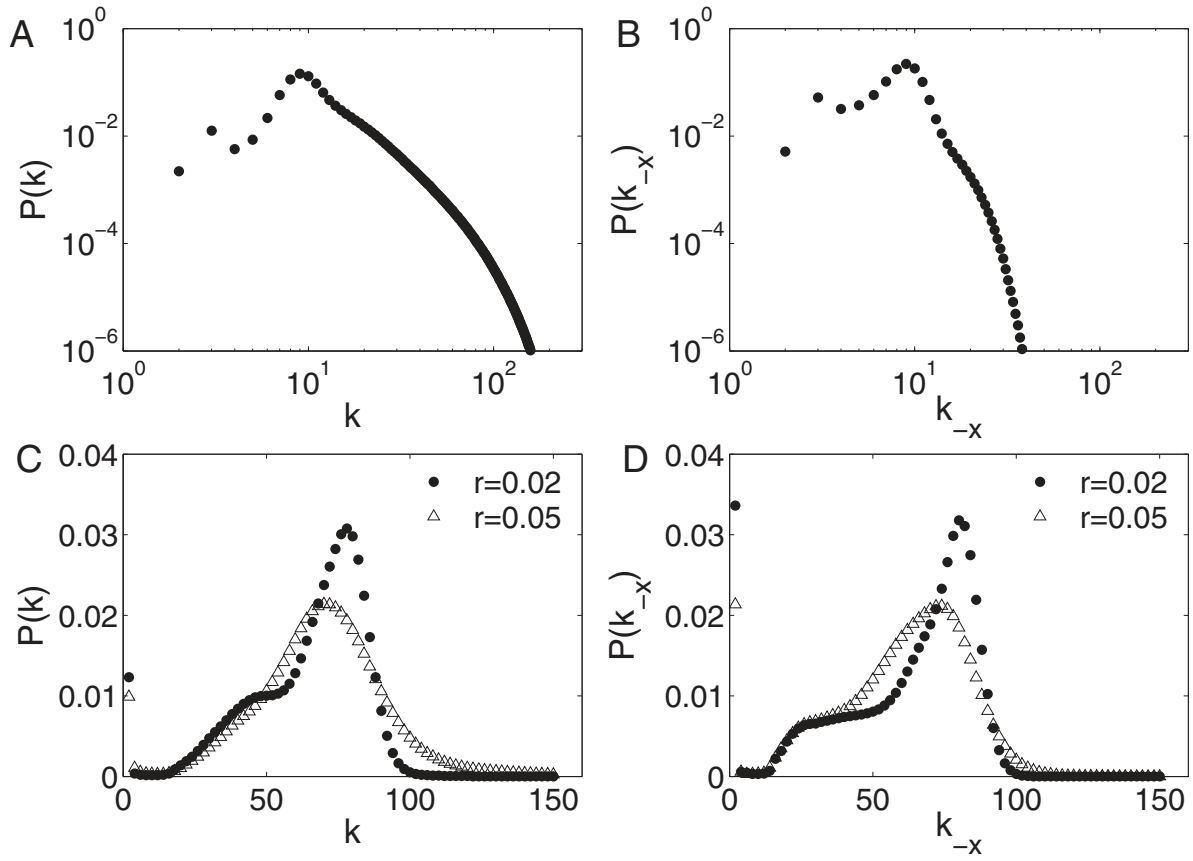

Figure 7. Degree distribution $p(k)$ of VGs constructed from: (A), (B) $x(t_i)$ and $-x(t_i)$ of Barnes' model and (C), (D) x^2 and $-x^2$ of Mininni's model. All $p(k)$ and $p(k_{-x})$ are estimated by an average over 10 000 independent realizations using a kernel smoother. In (C), (D) $r = 0.02$ (filled circles) and $r = 0.05$ (open triangles).

rapid growth in the increasing phase and slow decay in the decreasing phase of the activity cycles adequately. For the truncated model of the dynamo equations, a statistical significant correlation between instantaneous amplitude and frequency has been established, while the Barnes model shows virtually no correlation [34], which is generally termed the Waldmeier effect.

From both models, we generate 10 000 independent realizations, each of them has a 1 month temporal resolution and the same time span as we have for the observations (namely, over 300 years). We then construct VGs from both x_{t_i} and $-x_{t_i}$ from each realization in the same way as we processed for the original observation. As shown in figure 7, $p(k)$ can neither mimic the heavy tails of the distributions we observed in figures 2(A) and (B), nor can $p(k_{-x})$ capture the bimodality of the large degrees for strong minima as we have observed in the case of observational raw records in figures 2(C) and (D). This does not occur even if the parameter r of the second model is adjusted (figures 7(C) and (D)). One reason for the absence of any bimodality of $p(k)$ in the nonlinear oscillator model is the fact that the model was designed to reproduce smoothed sunspot number time series. The often-used 13-month running average method in the literature is known to suppress the maximum/minimum amplitudes of the series [1]. Therefore, the visibility condition for each time point is changed if the 13-month

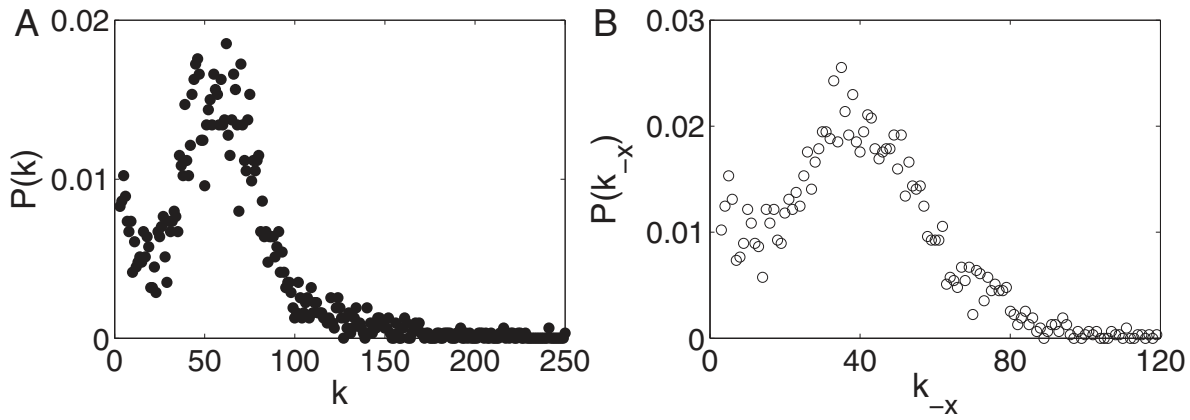

Figure 8. Degree distribution $p(k)$ of VGs reconstructed from smoothed monthly ISN data. (A) is for $x(t_i)$ and (B) for $x(t_i)$.

smoothing technique is applied to the original data. We show the degree distribution $p(k)$ of VGs reconstructed from smoothed ISN series in figure 8(B), where the bimodality is absent.

As shown in figure 7, strong maxima/minima are *not* well separated. Consequently we are prevented from identifying unique waiting time sequences. The corresponding analysis then depends significantly on the choice of threshold degrees.

Dynamo theory provides several hints that might explain the features observed in the long-term evolution of the solar activity. The two theoretical models tested in this work can reproduce the main qualitative features of the system reasonably well, however, within the context performed in this study, cannot yield a complete and conclusive rendition of the statistical properties of $p(k)$. The power-law regimes obtained from waiting time sequences suggest that the interaction patterns for two subsequent minima can be much more complicated than has been previously described as the instantaneous amplitudes–frequencies correlation using rather simple models [34, 35].

5. Conclusions

In this work, we apply a recently proposed network approach, namely the VG, to disclose the intricate dynamical behavior of both strong maximal and minimal sunspot numbers with observational records. More specifically, we show that:

(i) There is a power law regime of time scale of 15–1000 days in the occurrence of low-activity episodes, observed as clusters of inactive days. The identified persistence time scale of the strong minima agrees with the recently proposed hypothetical long range memory on time scales shorter than the 11-year cycle [14].

(ii) The occurrence of high activity episodes appears nearly completely random—as demonstrated by visual inspection of figures 3(A) and 5(A). In these figures the positions of strong maxima appear homogeneously distributed in the time domain. This suggests that the strong active regions appear more or less independently of each other. The distinctive long-term correlations of the strong maxima and strong minima are reflected by the structural asymmetries of the degree distributions $p(k)$ of the respective VGs.

(iii) There is no evidence for a long term inter-cycle memory. This is in agreement with the present paradigm based on alternative methods (see e.g. reviews by Petrovay [5] and Usoskin [36]), and provides an observational constraint for solar-activity models.

Therefore we propose that our results on the difference between maxima and minima could be used for evaluating models for solar activity because they reflect important properties that are not included in other measures reported in the literature.

From the methodological perspective, we are the first to propose a generalization for the construction of VGs from the negatively inverted time series. This has been seen, via our analysis of sunspot observations, to show complementary aspects of the original series. Note that the negatively inverse transformation is crucial for understanding when asymmetry is preserved in the time series. Therefore, from the methodological viewpoint, it is worth analyzing the dependence of the resulted VGs on an arbitrary monotonic nonlinear transformation. Furthermore, as presented in appendix B, a systematic investigation of the general conclusion as to whether large sampled points correspond to hubs of VGs will be a subject of future work—especially in the presence of cyclicity and asymmetry. On the other hand, one can study the fluctuation properties of the sunspot series by using any de-trending techniques as a preprocessing. Note that here we study the network properties of VGs from the original time series, instead of constructing a network from local fluctuations. In the case of studying the statistics of the fluctuations, it will be an interesting task to test the robustness of the network approach with respect to different de-trending algorithms to the sunspot series [12, 13].

Note that we construct VGs directly based on the raw sunspot series without any preprocessing. Many researchers prefer to base their studies on some kind of transformed series, since most common methods of data analysis in the literature rely on the assumption that the solar activity follows Gaussian statistics [2]. It is certain that the conclusions will then show some deviations depending on the parameters chosen for the preprocessing. The pronounced peaked and asymmetrical sunspot-cycle profiles prompt one to develop techniques such that the possible bias due to the unavoidable choice of parameters should be minimized. The complex network perspective offered by VG analysis has the clear advantage of being independent of any *a priori* parameter selection.

The procedure for network analysis outlined here can be directly applied to other solar activity indicators, for instance, the total solar irradiance and the solar-flare index. Some earlier analysis has revealed that, on average, the duration, rise and decay times of temporal flares show distinct asymmetric properties, which may vary either in phase or statistically show no correlation with the solar cycles [37, 38]. Note that our results reported here mainly concern the asymmetry between the maximal and minimal sunspots (from a VG perspective). We expect that our network approach could be used to study more asymmetric properties of the temporal sunspot series, e.g. to disclose the asymmetric distribution of sunspots between the northern and southern hemispheres [4, 39]. We compared our results to two rather empirical models and showed that the distinctive correlation patterns of maximal and minimal sunspots are currently absent from these two models. Using our analysis for more refined dynamo-based models (e.g. [40]) would be straightforward. Certainly, further work on this line of research will examine any differences given by the particular quantity and strengthen the understanding of the hypothetical long-range memory process of the solar activity from a much broader overview.

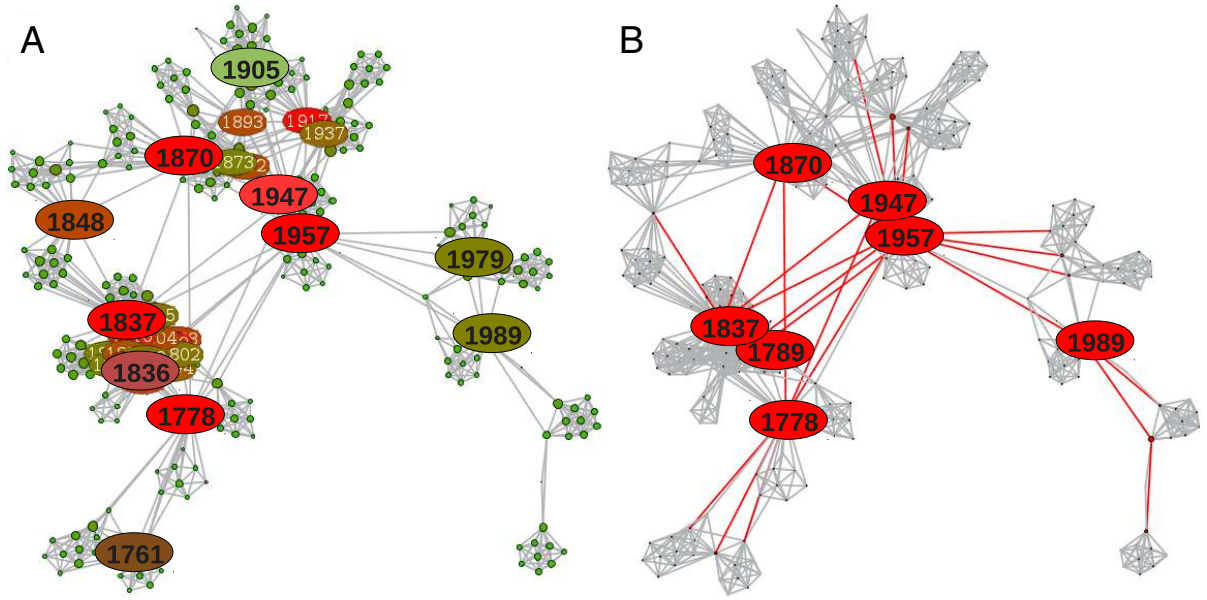

Figure A.1. Network representations of the VG constructed from the annual sunspot numbers of the entire series. Highlighted visible nodes are: (A) large degrees ($k_i > 15$) and (B) high betweenness centrality ($b_i > 0.2$).

Acknowledgments

This work was partially supported by the German BMBF (projects PROGRESS), the National Natural Science Foundation of China (grant numbers 11305062, 11135001, 11075056), Specialized Research Fund for the Doctoral Program (20130076120003) and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ043, SRF for ROCS, SEM. MS is supported by an Australian Research Council Future Fellowship (FT110100896).

Appendix A. Graph visualization of degree k_i and betweenness centrality b_i

Besides the fact that the maximal sunspot numbers are identified as hubs of VGs by the degree sequence k_i, convincing links between further network-theoretic measures and distinct dynamic properties can provide some additional interesting understanding for the time series [27]. In this study, we provide a graphical visualization on the relationship between degree k_i and node betweenness centrality b_i, which characterizes the node's ability to transport information from one place to another along the shortest path.

Using the annual ISN as a graphical illustration, here only the relationship between some relatively large degrees ($k_i > 15$) and betweenness centrality values ($b_i > 0.2$) is highlighted in figure A.1 for the entire series available. For instance, in the annual series, time points 1778, 1789, 1837, 1870, 1947, 1957 and 1989 are all identified simultaneously as large degrees and high betweenness, indicating strong positive correlations.

Certainly many other measures can be directly applied to the sunspot numbers, however, providing the appropriate (quantitative) interpretations of the results in terms of the particular

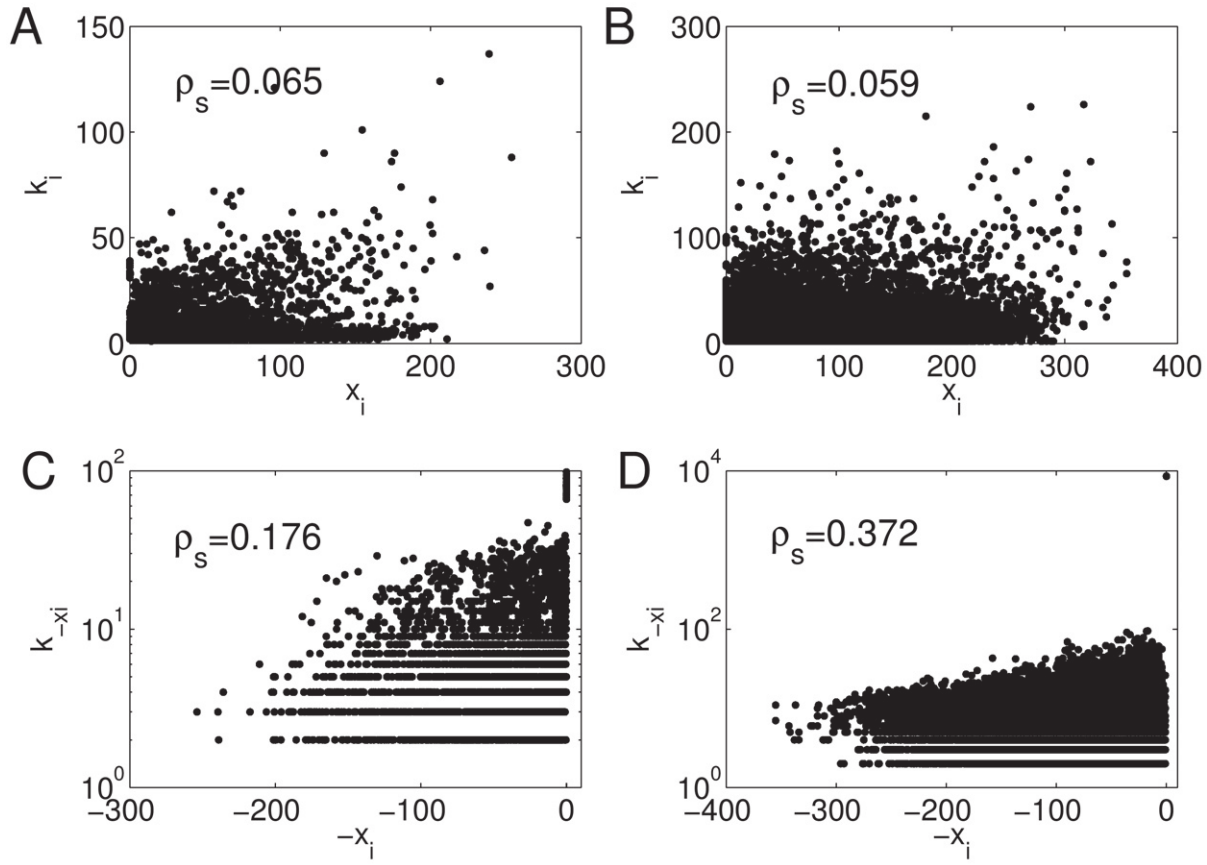

Figure B.1. Scatter plot of the degree sequence $k_i - x_i$ for VG constructed from ISN series based on (A) monthly and (B) daily data respectively. Spearman ρ is indicated. Panels (C), (D) are based on the VGs constructed from negatively inverted series $-x_i$.

underlying geophysical mechanisms remains a challenging task and is largely open for future work. Note that there is in general a strong interdependence between these different network structural quantities.

Appendix B. Correlation between k_i and x_i

A general rule of understanding the scale-free property of the degree distribution of complex networks is the effects of a very few hubs having a large number of connections. In the particular case of VGs, hubs are related to maxima of the time series, since they have better visibility contact with more vertices than other sampled points. However, this result cannot be generalized to all situations, for instance, it is easy to generate a time series such that its maxima are not always mapped to hubs in the VGs.

One simple way to better explore this correlation is to use scatter plots between the degree sequence and the sunspot time series. As we show in figure B.1, the Spearman correlation coefficients ρ are very small (still significantly larger than zero) in the case of VGs reconstructed from the original time series. This provides an important cautionary note on the interpretation of hubs of VGs by local maximal values of the sunspot numbers. In contrast, if the network

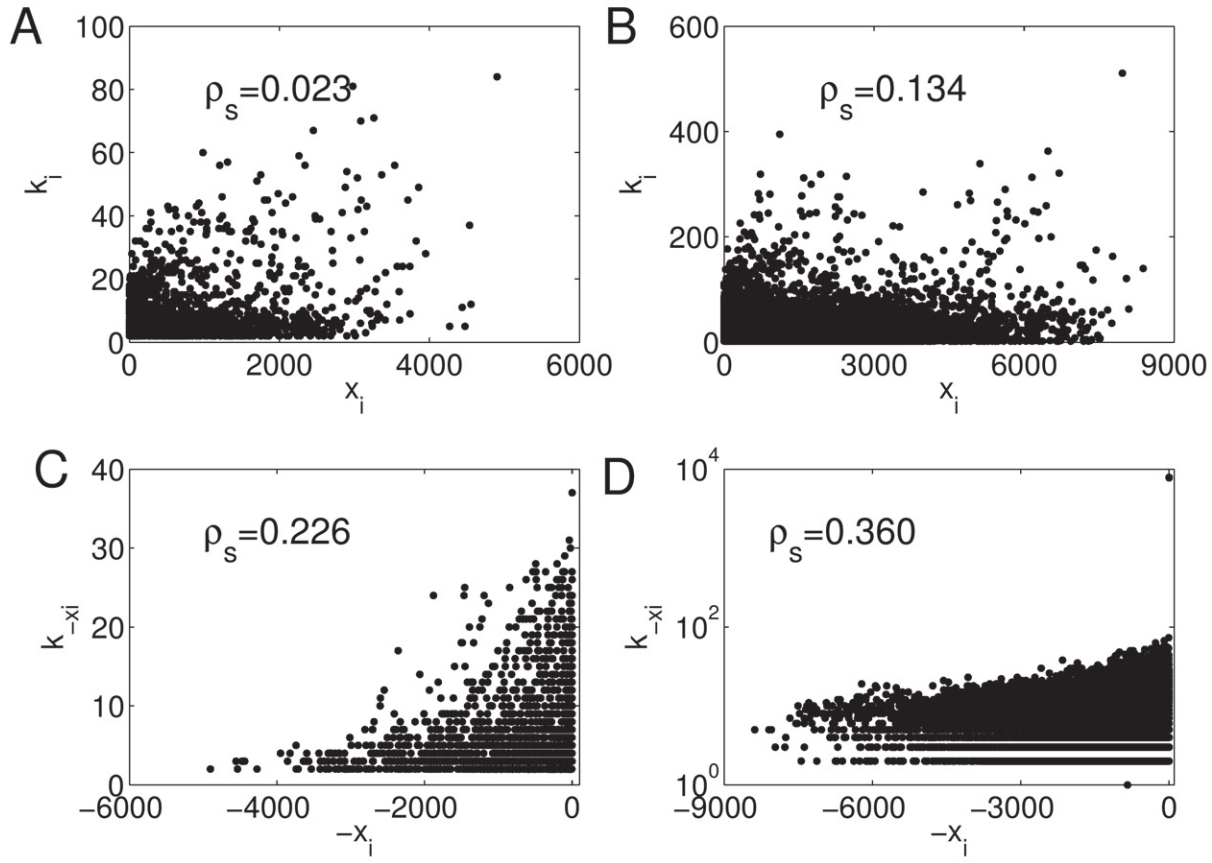

Figure B.2. Scatter plot of the degree sequence $k_i - x_i$ for VG constructed from SSA series based on (A) monthly and (B) daily data respectively. Spearman ρ is indicated. Panels (C), (D) are based on the VGs constructed from negatively inverted series $-x_i$.

is reconstructed from $-x$, hubs of VGs could be better interpreted by local minimal values of the sunspot data since the correlations become larger. These results hold for both the ISN and SSA series (figure B.2). One reason for the lack of strong correlation between the degree k_i and x_i is because of the (quasi-)cyclicity of the particular time series, which has a similar effect as the Conway series [23]. It is this concave behavior over the time axis (although quasi-periodic from cycle to cycle) that prevents the local maxima from having highly connected vertices. It remains unclear how local maxima of a time series are mapped to hubs of VGs—we defer this topic for future work especially in the presence of cyclicity. This situation becomes even more challenging if some sort of asymmetric property is preserved in the data, as we have found for the sunspot series.

References

[1] Hathaway D H 2010 The solar cycle *Living Rev. Sol. Phys.* **7** 1
[2] Pesnell W D 2012 Solar cycle predictions (invited review) *Sol. Phys.* **281** 507–32
[3] Kurths J and Ruzmaikin A A 1990 On forecasting the sunspot numbers *Sol. Phys.* **126** 407–10
[4] Hathaway D H, Wilson R M and Reichmann E J 1994 The shape of the sunspot cycle *Sol. Phys.* **151** 177–90

[5] Petrovay K 2010 Solar cycle prediction *Living Rev. Sol. Phys.* **7** 6
[6] Solanki S K and Krivova N A 2011 Analyzing solar cycles *Science* **334** 916–7
[7] Brajša R, Wöhl H, Hanslmeier A, Verbanac G, Ruždjak D, Cliver E, Svalgaard L and Roth M 2009 On solar cycle predictions and reconstructions *Astron. Astrophys.* **496** 855–61
[8] Ramesh K B and Bhagya Lakshmi N 2012 The amplitude of sunspot minimum as a favorable precursor for the prediction of the amplitude of the next solar maximum and the limit of the Waldmeier effect *Sol. Phys.* **276** 395–406
[9] Mandelbrot B B and Wallis J R 1969 Some long-run properties of geophysical records *Water Resour. Res.* **5** 321–40
[10] Ruzmaikin A, Feynman J and Robinson P 1994 Long-term persistence of solar activity *Sol. Phys.* **149** 395–403
[11] Oliver R and Ballester J L 1998 Is there memory in solar activity? *Phys. Rev.* E **58** 5650–4
[12] Movahed M S, Jafari G R, Ghasemi F, Rahvar S and Rahimi Tabar M R 2006 Multifractal detrended fluctuation analysis of sunspot time series *J. Stat. Mech.* **2006** P02003
[13] Hu J, Gao J and Wang X 2009 Multifractal analysis of sunspot time series: the effects of the 11-year cycle and Fourier truncation *J. Stat. Mech.* **2009** P02066
[14] Rypdal M and Rypdal K 2012 Is there long-range memory in solar activity on timescales shorter than the sunspot period? *J. Geophys. Res.* **117** A04103
[15] SIDC-team 2011 The international sunspot number & sunspot area data *Monthly Report on the International Sunspot Number* Royal Observatory Greenwich (www.sidc.be/sunspot-data/), (http://solarscience.msfc.nasa.gov/greenwch.shtml/)
[16] Barnes J A, Tryon P V and Sargent H H III 1980 Sunspot cycle simulation using random noise *The Ancient Sun: Fossil Record in the Earth, Moon and Meteorites* ed R O Pepin, J A Eddy and R B Merrill (Oxford: Pergamon) pp 159–63
[17] Mininni P D, Gómez D O and Mindlin G B 2000 Stochastic relaxation oscillator model for the solar cycle *Phys. Rev. Lett.* **85** 5476–9
[18] Zhang J and Small M 2006 Complex network from pseudoperiodic time series: topology versus dynamics *Phys. Rev. Lett.* **96** 238701
[19] Xu X, Zhang J and Small M 2008 Superfamily phenomena and motifs of networks induced from time series *Proc. Natl Acad. Sci. USA* **105** 19601–5
[20] Marwan N, Donges J F, Zou Y, Donner R V and Kurths J 2009 Complex network approach for recurrence analysis of time series *Phys. Lett.* A **373** 4246–54
[21] Donner R V, Zou Y, Donges J F, Marwan N and Kurths J 2010 Recurrence networks—a novel paradigm for nonlinear time series analysis *New J. Phys.* **12** 033025
[22] Donner R V, Small M, Donges J F, Marwan N, Zou Y, Xiang R and Kurths J 2011 Recurrence-based time series analysis by means of complex network methods *Int. J. Bifurcation Chaos* **21** 1019–46
[23] Lacasa L, Luque B, Ballesteros F, Luque J and Nuño J C 2008 From time series to complex networks: the visibility graph *Proc. Natl Acad. Sci. USA* **105** 4972–5
[24] Lacasa L, Luque B, Luque J and Nuno J C 2009 The visibility graph: a new method for estimating the hurst exponent of fractional Brownian motion *Europhys. Lett.* **86** 30001
[25] Elsner J B, Jagger T H and Fogarty E A 2009 Visibility network of United States hurricanes *Geophys. Res. Lett.* **36** L16702
[26] Nuñez A M, Lacasa L, Gomez J P and Luque B 2012 Visibility algorithms: a short review *New Frontiers in Graph Theory* ed Y Zhang (Rijeka, Croatia: InTech) pp 119–52
[27] Donner R V and Donges J F 2012 Visibility graph analysis of geophysical time series: potentials and possible pitfalls *Acta Geophys.* **60** 589–623
[28] Donges J, Donner R and Kurths J 2013 Testing time series irreversibility using complex network methods *Europhys. Lett.* **102** 10004
[29] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256

[30] Clauset A, Shalizi C R and Newman M E J 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703
[31] Voss H, Kurths J and Schwarz U 1996 Reconstruction of grand minima of solar activity from Δ^{14}C data: linear and nonlinear signal analysis *J. Geophys. Res.* **101** 15637–43
[32] Usoskin I G, Solanki S K and Kovaltsov G A 2007 Grand minima and maxima of solar activity: new observational constraints *Astron. Astrophys.* **471** 301–9
[33] Wu Y, Zhou C S, Xiao J H, Kurths J and Schellnhuber H J 2010 Evidence for a bimodal distribution in human communication *Proc. Natl Acad. Sci. USA* **107** 18803–8
[34] Mininni P D, Gomez D O and Mindlin G B 2002 Instantaneous phase and amplitude correlation in the solar cycle *Sol. Phys.* **208** 167–79
[35] Paluš M and Novotná D 1999 Sunspot cycle: a driven nonlinear oscillator? *Phys. Rev. Lett.* **83** 3406–9
[36] Usoskin I G 2013 A history of solar activity over millennia *Living Rev. Sol. Phys.* **10** 1
[37] Temmer M, Veronig A, Hanslmeier A, Otruba W and Messerotti M 2001 Statistical analysis of solar Hα flares *Astron. Astrophys.* **375** 1049–61
[38] Melnik V N 2003 Velocities of beam-plasma structures at their propagation in the solar corona *Sol. Phys.* **212** 111–9
[39] Donner R and Thiel M 2007 Scale-resolved phase coherence analysis of hemispheric sunspot activity: a new look at the north–south asymmetry *Astron. Astrophys.* **475** L33–6
[40] Charbonneau P 2010 Dynamo models of the solar cycle *Living Rev. Sol. Phys.* **7** 3