Persistent homology of fractional Gaussian noise

H. Masoomy [1], B. Askari,[1] M. N. Najafi,[2,*] and S. M. S. Movahed [1,†]

[1]*Department of Physics, Shahid Beheshti University, 1983969411, Tehran, Iran*
[2]*Department of Physics, University of Mohaghegh Ardabili, P.O. Box 179, Ardabil, Iran*

In this paper, we employ the persistent homology (PH) technique to examine the topological properties of fractional Gaussian noise (fGn). We develop the weighted natural visibility graph algorithm, and the associated simplicial complexes through the filtration process are quantified by PH. The evolution of the homology group dimension represented by Betti numbers demonstrates a strong dependency on the Hurst exponent (H). The coefficients of the birth and death curves of the k-dimensional topological holes (k-holes) at a given threshold depend on H which is almost not affected by finite sample size. We show that the distribution function of a lifetime for k-holes decays exponentially and the corresponding slope is an increasing function versus H and, more interestingly, the sample size effect completely disappears in this quantity. The persistence entropy logarithmically grows with the size of the visibility graph of a system with almost H-dependent prefactors. On the contrary, the local statistical features are not able to determine the corresponding Hurst exponent of fGn data, while the moments of eigenvalue distribution (M_n) for $n \geqslant 1$ reveal a dependency on H, containing the sample size effect. Finally, the PH shows the correlated behavior of electroencephalography for both healthy and schizophrenic samples.

I. INTRODUCTION

A powerful approach to study different types of data sets ranging from point cloud data and scalar fields to a complex network (graph), particularly high-dimensional data, is called topological data analysis (TDA) [1–6]. TDA as an application of algebraic topology [7–9] and a branch of computational topology [10]; it analyzes the shape of high-dimensional complex data in terms of global features (number of connected components, loops, voids, etc.) of topological space underlying the data set. In the persistent homology (PH) technique, as the main part of TDA, the topological approximation of phase space of any type of data set which is called the simplicial complex is assigned to the underlying data, and then topological invariants are computed.

The PH aims to capture topological evolution of data set by varying scale (parameter), and extracts topological invariants of the data set in each scale summarizing them in different representations, e.g., persistence barcode (PB) [11,12], persistence diagram (PD) [13], persistence landscape (PL) [14], persistence image (PI) [15], persistence surface (PS), and β-curve, which reveal topological information of the data set. Being robust to noise, PH can clarify the essential features of the systems with high internal degrees of freedom and is capable of classifying data sets [16,17]. The PH technique has attracted much attention due to its vast applications in analyzing complex networks [18–20]. Also it has been used in various systems (see, e.g., [21–29] and references therein).

A mathematical model containing correlation tuned by one parameter (the Hurst exponent H [30]) is the fractional Brownian motion (fBm) and its increment is known as fractional Gaussian noise (fGn) [31]. It is also used to model self-similar phenomena of various types ranging from meteorology, engineering, econophysics, and astronomy to biology (see, e.g., [32,33] and references therein). To quantify the properties of a given self-similar data set whose power spectrum behaves as a power law in the frequency (wavelength) domain, many methods have been proposed. A well-studied method is multifractal detrended fluctuation analysis (MFDFA) [34–36]. Taking into account the higher-order detrended covariance, the multifractal detrended cross-correlation analysis (MFDXA) has been introduced [37]. Despite many advantages brought by the mentioned methods, the impact of more complicated trends and finite-size effects have not been diminished completely in many previous approaches [38–41]. On the other hand, the finite size of time series affects the accurate estimation of the Hurst exponent by some of previous methods [42].

Although different methods can be found in the literature to determine the scaling exponent of fBm or fGn series, very little attention has been paid to deal with the topological properties and the probable capability of TDA for estimation of the Hurst exponent. In this regard, knowing a given time series belongs to an fGn class, some relevant questions can be raised: (i) Does the topological aspect of time series depend on the Hurst exponent? (ii) What are the effects of sample size, trends, and irregularity of fGn signal on the PH of topological motifs generated from the data set? (iii) How is the multifractality expressed by PH? Motivated by the mentioned questions, in this paper, we concentrate on the topological properties (homology group) of fGn.

─────────────
*nattagh.najafi@uma.ac.ir
†m.s.movahed@ipm.ir

A way to construct the higher-dimensional manifold from a typical time series in order to evaluate the evolution of k-dimensional holes through the filtration process is assigning a weighted graph to the underlying time series. There are various methods to assign a network to a typical time series, e.g., the proximity network, cycle network, visibility graph, correlation network, recurrence network, and transition network (see, e.g., [43] and references therein). The idea of the visibility graph (VG) is a complex network constructed by considering the *visibility* algorithm, proposed by Lacasa *et al.* as a novel way to analyze time series in terms of complex network language [44] proposed as an alternative method to estimate H for fBm and fGn series [45].

In this paper, we rely on statistical and topological properties (homology group) of natural VGs (NVGs) associated with such signals. First, we propose a weighted version of natural VGs (WNVGs) by defining a well-defined weight function to quantify the quality of visibility between data points of the signal for extracting nonsensitive (robust) features in the presence of possible noises in the signal. Then, we apply the filtration process on the weighted clique simplicial complex corresponding to WNVG by considering the weight (visibility value) of links as a threshold parameter and higher-order connections (k-cliques; $k > 2$) as building blocks of such higher-order structure. Accordingly, we figure out the evolution of robust topological features explaining the global structure of WNVG. Considering other techniques for making networks from time series and applying PH could provide interesting results, but is time consuming from computational points of view; the possibility of distinguishing the nonstationary and stationary series, the sensitivity of corresponding results to the value of the Hurst exponent, and the robustness of results with sample size are some of our criteria for taking VG in its weighted version.

Our research presented in this paper has the following advantages.

(1) Inspired by network science and a self-similar process characterized by a scaling exponent called the "Hurst exponent," we implement the weighted natural visibility graph (WNVG) to make a network from fractional Gaussian noise (fGn). In this approach, the topological motifs survive; consequently, their evolution concerning self-similar exponents can be examined. Our results approve that such evolution is a robust feature for determining the Hurst exponent.

(2) We will demonstrate that the statistical analysis of WNVGs such as probability distribution functions of eigenvectors, betweenness, and closeness centralities are almost H-independent, and therefore, they cannot recognize the type of correlation of fGn series.

(3) We rely on TDA and implement the PH, which is a robust method to examine the evolution of global properties during the filtration process. Almost all relevant results are sensitive enough to the Hurst exponent of the underlying data set and it is also possible to check whether the data are stationary or not. More precisely, irrespective of either stationarity or nonstationarity of underlying times series, to capture the homology generators, the constructed network should be weighted. It is worth noting that the nonstationary behavior may be produced due to the various types of trends superimposed on the intrinsic fluctuation; therefore, in the presence

of any prior information regarding trends, the preprocessing procedure including at least one of the following algorithms must be employed to produce clean series, e.g., the singular value decomposition algorithm [46], the adaptive detrending algorithm [47], and empirical mode decomposition [48].

(4) Finally, we emphasize that the behavior of the local (statistical) observables depends weakly on H, whereas the coefficients of global (topological) observables are almost strongly H-dependent and even for a part of measures, the size effect is almost diminished. We notify that TDA provides a new type of measure to quantify the Hurst exponent, and therefore, the scaling exponents of the correlation function, power spectrum, and fractal dimension can be specified with reliable approaches.

The rest of paper is organized as follows: In the next section, our methodology is introduced briefly. The construction of VGs for a time series and the key idea of topological network analysis are clarified in Sec. II. The numerical results of synthetic fGn time series, which are covered in two subsections, local statistical properties and topological properties, are presented in Sec. III. The implementation of realistic data is described in Sec. IV. We give some ideas for applying the proposed method to various problems in Sec. V, and close the paper with a conclusion. More details on synthetic data generation, statistical network analysis, algebraic topology, and persistent homology are explained in appendices.

II. METHODOLOGY

In this research, we aim to analyze the complex network of the visibility graph constructed from a fractional Gaussian noise (fGn) (see Appendix A for more details of generating a synthetic fGn series), with an emphasis on the topological aspects. Besides this, we also compute some conventional statistical properties of the mentioned signal.

A. Visibility graph method

Among growing applicability of complex networks in many fields and interdisciplinary branches in science [49], a technique has been suggested which converts a time series to a network, the so-called visibility graph method [44]. Generally, suppose $\{x\} : \{x(t_i), i = 1, \dots, N\}$ represents a real-valued time series. One can construct a network, the so-called visibility graph, denoted by $G = (V, E, w)$, $V \equiv \{v_i\}_{i=1}^{N}$ is the node (vertex) set, E is link (edge) set, and w is a map from E to the real numbers. The VG is defined by using the bijection as follows,

$$f : V \equiv \{v_i\}_{i=1}^{N} \leftrightarrow T \equiv (t_i)_{i=1}^{N}, \quad f(v_i) = t_i, \quad (1)$$

and the connections are constructed according to the *visibility condition* between the nodes, i.e., the nodes v_i and v_j are connected if the node v_j is visible from the node v_i and vice versa, and therefore the resulting graph is undirected (for more details on the properties of VGs, see [44]). In general, there are two ways to construct a network (graph) from a time series: the horizontal visibility graph (HVG) [50–52] and the natural visibility graph (NVG) [53–55]; the former is more sparse than the latter case and in this work we focus on the NVG. In Fig. 1, we show how an HVG
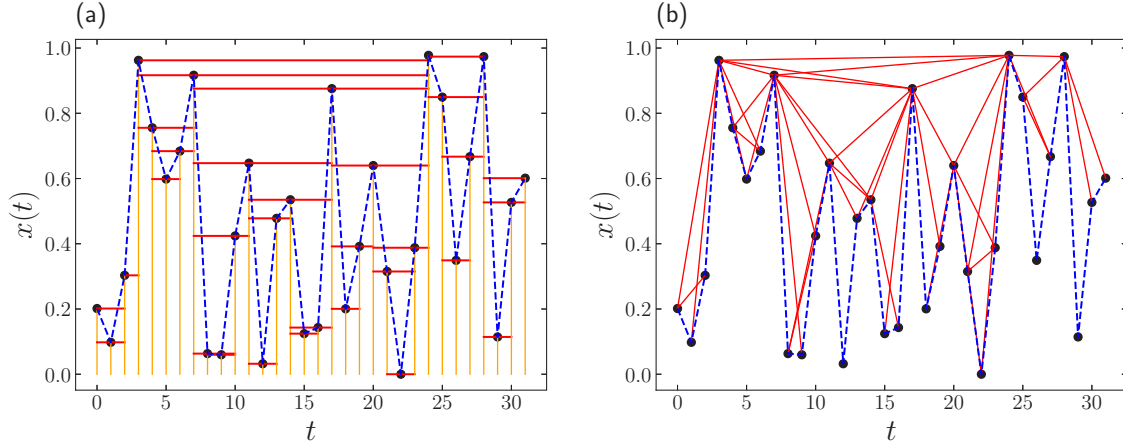
FIG. 1. The schematic representation of making a network for a typical data set using VG algorithm. (a) The HVG and (b) NVG of a synthetic fGn with $H = 0.5$ (white noise). To make more sense, we took the sample size equal to $N = 32$.

[panel (a)] and an NVG [panel (b)] for a synthetic fGn series can be constructed. In a binary setup, the corresponding visibility graph chooses the range of the weights from a binary set, $w^{(B)} : E \rightarrow \mathbb{Z}_2 \equiv \{0, 1\}$; e.g., for a binary NVG (BNVG) the weight function can be written according to following relation,

$$
w_{ij}^{(BN)} \equiv
\begin{cases}
1, & |f(v_i) - f(v_j)| = 1, \\
\prod\limits_{k=i+1}^{j-1} \Theta(s_{ij} - s_{ik}), & |f(v_i) - f(v_j)| > 1,
\end{cases}
\tag{2}
$$

where Θ is the step function, and $s_{ij} \equiv \frac{x(f(v_j)) - x(f(v_i))}{f(v_j) - f(v_i)}$. The argument of the Θ function being positive guarantees that the node v_j is visible from the node v_i and vice versa. Since the edge in a BNVG has the weight 0 or 1, it is unsuitable for continuous filtering. To take into account the quality of visibility between nodes, we suggest the weighted version of the natural visibility graph (WNVG), by considering the weight function as follows:

$$
w_{ij}^{(WN)} \equiv
\begin{cases}
|s_{ij}|, & |f(v_i) - f(v_j)| = 1, \\
\{\prod\limits_{k=i+1}^{j-1} \Theta(s_{ij} - s_{ik})[s_{ij} - s_{ik}]\}^{1/(j-i-1)}, & |f(v_i) - f(v_j)| > 1.
\end{cases}
\tag{3}
$$

There are two factors inside the product. In the second branch of Eq. (3), one is the step function just like the binary graph, and the other is the weight which is proportional to "how visible is the site j from i and vice versa"; i.e., the more distinguishable the data points are in the original time series, the higher the corresponding weight is in the constructed network. The term $\frac{1}{j-i-1}$ is necessary to make the weights reasonable numbers for comparison reasons. In the absence of this exponent, the more the distance between the nodes is, the higher the corresponding weights are. For both statistical and topological analysis, we use this weight function which admits continuous filtering.

B. Topological network analysis

A conventional approach to quantify the properties of a typical network is determining degree, the number of nodes straightly connected with the underlying node by nonzero weight, and centrality measures such as betweenness centrality, the number of shortest paths between any pair of nodes going through a distinct node, closeness centrality, eigenvector centrality, etc. (see Appendix B for more details). Besides the mentioned approach, other formalisms have been proposed in which not only the pairwise connections (links) are examined but also a systematic pipeline is considered to incorporate higher-order connections (k-cliques), including the simplicial complex and hypergraph [56–58].

To analyze such higher-order structures, one can generalize the well-known statistical quantities based on higher-order connections. On the other hand, for global analysis of these structures newly born topological tools are proposed. By the term "global," we mean the essential features of the object which are not affected by geometric transformations. The well-known topological quantities describing global properties of the space (or any object mapped to a topological space) are Betti numbers. The kth Betti number of an object (β_k) is a

topological invariant. The dimension of the k-homology group of the topological space corresponding to the object counts the number of k-dimensional topological holes (k-holes) of the object. Intuitively, β_0 indicates the number of connected components, β_1 is the number of topological (unshrinkable) loops, and β_2 counts the number of topological (unshrinkable) voids (see Appendix C for more details). According to the mentioned properties of the topological measures and weight-based analysis of weighted complex networks, the applied algebraic topology tool, known as the persistent homology (PH) technique, is useful to study global structures of the weighted complex network in higher dimensions [10,59,60].

In the PH-based analysis of a weighted complex network, the weighted network is mapped to a weighted clique simplicial complex as a higher-order structure containing any $(k + 1)$-clique of the network as the k-simplex. Therefore, PH creates a nested sequence of the clique simplicial complex, so-called filtration, by considering the weight as the parameter, such that any link (1-simplex) presents in the complex at a distinct weight if the weight of the link is less than the distinct weight. Accordingly, the k-holes of the complex appear and disappear when the weight (threshold) varies. To summarize the evolution of these k-dimensional topological features of the system, PH assigns an ordered tuple, the so-called persistence pair (PP), to this global descriptor to visualize the topological variation of the system in the birth-death space. This kind of visualization for the evolution of k-holes is called the persistence diagram (PD) for the k-homology group. PD for the k-homology group includes hidden topological information of the k-holes in the weighted complex. For example, one can compute the population of birth, death, and lifetime of k-holes using the distribution of PPs in PD for the k-homology group. Persistence entropy (PE) for the k-homology group is another quantity defined by the Shannon entropy of topological persistence (lifetime) of PPs of the k-homology group (see Appendix D for more details).

III. RESULTS

To evaluate the statistical and topological features for the constructed VG from a series, at first, we focus on the synthetic fractional Gaussian noise (fGn) as a self-similar series. There are some exact and approximate algorithms for generating a self-similar time series. The Hosking [61], Cholesky [62], and Davies-Harte methods [63] are the examples of exact approach to generate fBm and fGn series. Throughout this paper, we implement the Davies-Harte method to generate fGn series with different self-similar exponents and sizes. To reduce any bias in the nominal Hurst exponent as much as possible, all relevant results are determined by doing an ensemble average over 10^4 realizations generated by the Davies-Harte method for each H. To ensure the correctness of the Hurst exponent value associated with each simulated time series, we have computed again the value of H by our code based on the detrended fluctuation analysis (DFA) method [36] and compare the expected and computed Hurst exponents. Our results show that both Hurst exponents are in agreement with each other.

Now we turn to the statistical and topological properties of the VGs constructed from the fGn time series and investigate their behavior concerning the Hurst exponent. The networks of sizes $N = 2^7, 2^8, 2^9, 2^{10}, 2^{11}$, and 2^{12} (for which a desktop with 128 GB memory is capable of performing matrix operations) are considered. The Python toolbox NetworkX [64] is employed for the matrix operations on the graphs. In the topological analysis, we especially focus on the Betti-0 (represented by the β_0 defined as the number of connected components of the network) and Betti-1 (represented by β_1 defined as the number of loops) features, which are extracted by using the "Dionysus" Python package [65]. The persistence statistics, containing the lifetime (the interval between birth and death) of the topological features, and its Shannon entropy are also analyzed.

Each exponent has been estimated by Bayesian statistics accordingly; the $\{\mathcal{D}\}$ and $\{\Upsilon\}$ reveal the data and model free parameters, respectively. The posterior function is defined by

$$\mathcal{P}(\Upsilon|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\Upsilon)\mathcal{P}(\Upsilon)}{\int \mathcal{L}(\mathcal{D}|\Upsilon)\mathcal{P}(\Upsilon)d\Upsilon}, \qquad (4)$$

where $\mathcal{L}$ is the likelihood and $\mathcal{P}(\Upsilon)$ is the prior probability function containing all information concerning model parameters. Here we adopt the top-hat function for the prior function whose window's size depends on the expected range of the corresponding exponent. Taking into account the central limit theorem, the functional form of likelihood becomes multivariate Gaussian, i.e., $\mathcal{L}(\mathcal{D}|\Upsilon) \sim \exp(-\chi^2/2)$. The χ^2 for determining the best-fit value for the scaling exponent reads as

$$\chi^2(\Upsilon) \equiv \Delta^\dagger \cdot C^{-1} \cdot \Delta, \qquad (5)$$

where Δ is a column vector whose elements are determined by difference between computed value and theoretical form for each measure, and C is the corresponding covariance matrix. Finally, the best fit value of the considered exponent is computed by maximizing the likelihood probability distribution, and the associated error bar is given by

$$68.3\% = \int_{-\sigma_\Upsilon}^{+\sigma_\Upsilon} \mathcal{L}(\mathcal{D}|\Upsilon)d\Upsilon. \qquad (6)$$

Subsequently, we report the best value of the scaling exponent at a 1σ confidence interval as $\Upsilon_{-\sigma_\Upsilon}^{+\sigma_\Upsilon}$.

A. Local statistical properties

By local properties, we mean the properties which are node-dependent and are not necessarily globally defined. It has been confirmed that the distribution function of the node degree of VGs of the fBms and fGns is power law $p(k) \propto k^{-\gamma}$ with the exponent $\gamma(H) = 3 - 2H$ and $\gamma(H) = 5 - 2H$, respectively [44,45]. In this subsection, we perform our computation for the WNVG, introduced in this paper.

The probability distribution function of eigenvector (c^E) and betweenness (c^B) centralities are indicated in panels (a) and (b) of Fig. 2 in log-log scale, in terms of H for WNVGs. The scaling behavior of mentioned distributions has been checked by the Kolmogorov-Smirnov test (KS test) and the
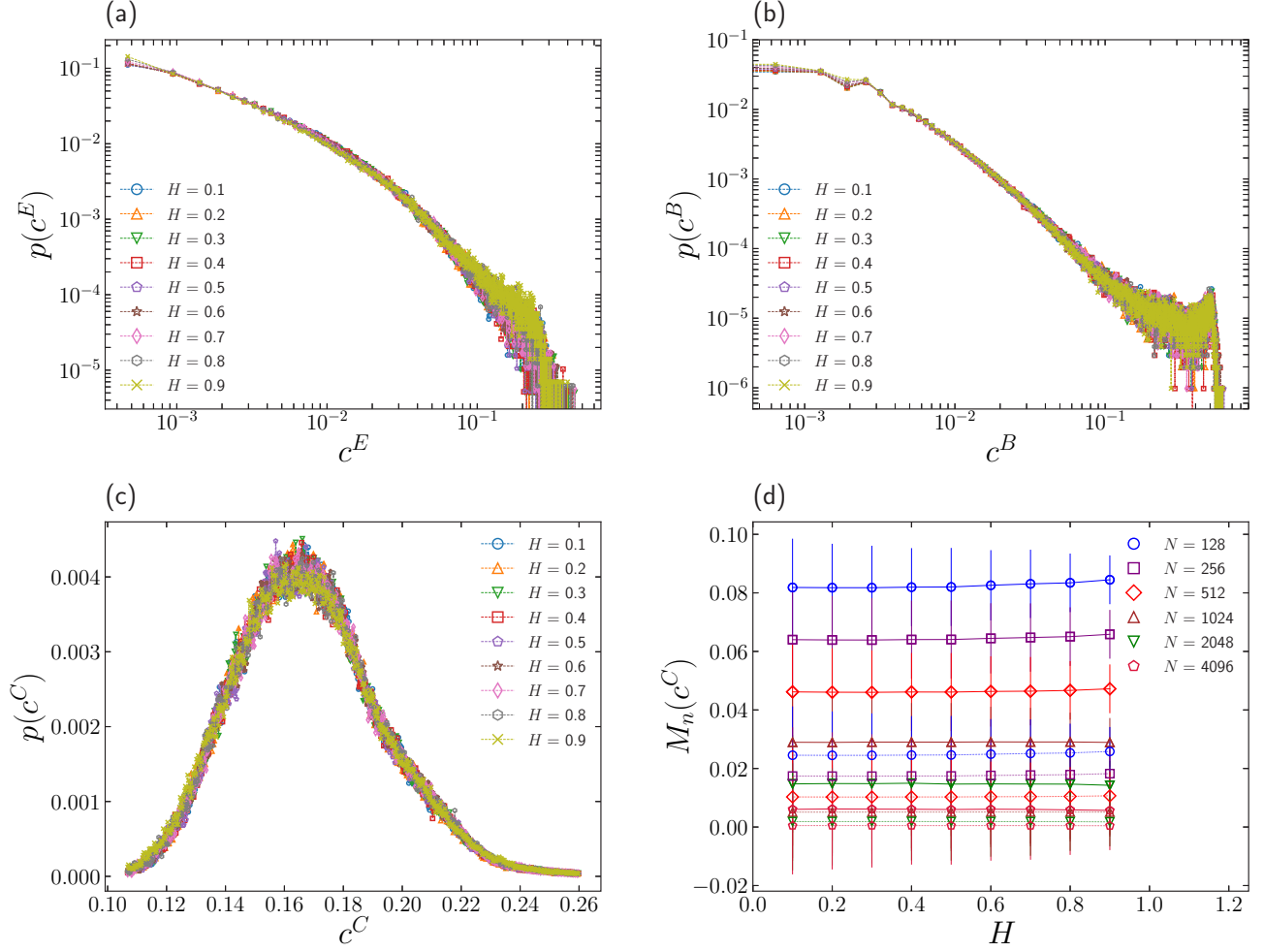
FIG. 2. Probability distribution function of local features of WNVGs constructed from fGns for various Hurst exponents. Panel (a) is the $p(c^E)$ while panel (b) shows the probability distribution of betweenness centrality for $N = 2^{10}$ in log-log scale. Panels (c) and (d) respectively illustrate the $p(c^C)$ and corresponding moments, M_n, for $n = 2$ (solid line) and $n = 3$ (dashed line) versus H. Different symbols are taken for various sizes.

results confirm that the power-law behavior of the computed $p(c^E)$ and $p(c^B)$ is highly H- and size-dependent [see panels (a) and (b) of Fig. 2]. The probability distribution of closeness centrality and corresponding moments M_n for $n = 2$ and $n = 3$ are illustrated in panels (c) and (d) of Fig. 2, respectively. This part depicts that not only the Hurst dependency is not confirmed but also the overall shape of probability distribution functions of statistical measures depends on the size of the underlying fGn series. Subsequently, such distributions cannot discriminate between different fGn series.

The full spectrum of the eigenvalues [Eq. (B2)] is illustrated in panel (a) of Fig. 3. We see that the impact of H is changing the range of the spectrum, and by increasing the Hurst exponent, the range of the spectrum for WNVGs becomes tight. This phenomenon can be understood by recalling that correlations (obtained by increasing H) smooth the underlying time series, causing the corresponding network to have more links at low weight. For more smoothed time series, the typical slopes for the associated WNVG become down, leading to lower weights according to Eq. (3), and equivalently making a shorter range for the distribution of λ's. Panel (b) of Fig. 3 indicates different moments for $p(\lambda)$. The solid

and dashed lines correspond to M_2 and M_3, respectively. The nth moment of this distribution behaves as an H-dependent quantity. The higher the order of moments, the stronger the dependency on sample size is.

B. Topological properties

For any given value of Hurst exponent, the associated BNVG of fGn is a *topological tree*; therefore, the BN-VGs are topologically equivalent to each other. In other words, various BNVGs are homeomorphic for different H according to the homeomorphism theorem. The loopless-ness of BNVGs indicates that these networks do not contain some classes of network motifs corresponding to topological loops, so-called topological motifs [52]. To make more sense, suppose 4 successive data points as $\{x_j\}_{j=i}^{i+3}$, ($j = 1, \ldots, N-4$) from a given time series $\{x_i\}_{i=1}^N$ to make a minimal topological loop. From a network perspective, the necessary (not sufficient) condition to construct a topological loop is that $w_{i,i+1} = w_{i,i+3} = w_{i+1,i+2} = w_{i+2,i+3} = 1$ and the sufficient condition is $w_{i,i+2} = w_{i+1,i+3} = 0$. The weights here are determined by Eq. (2). Geometrically, the
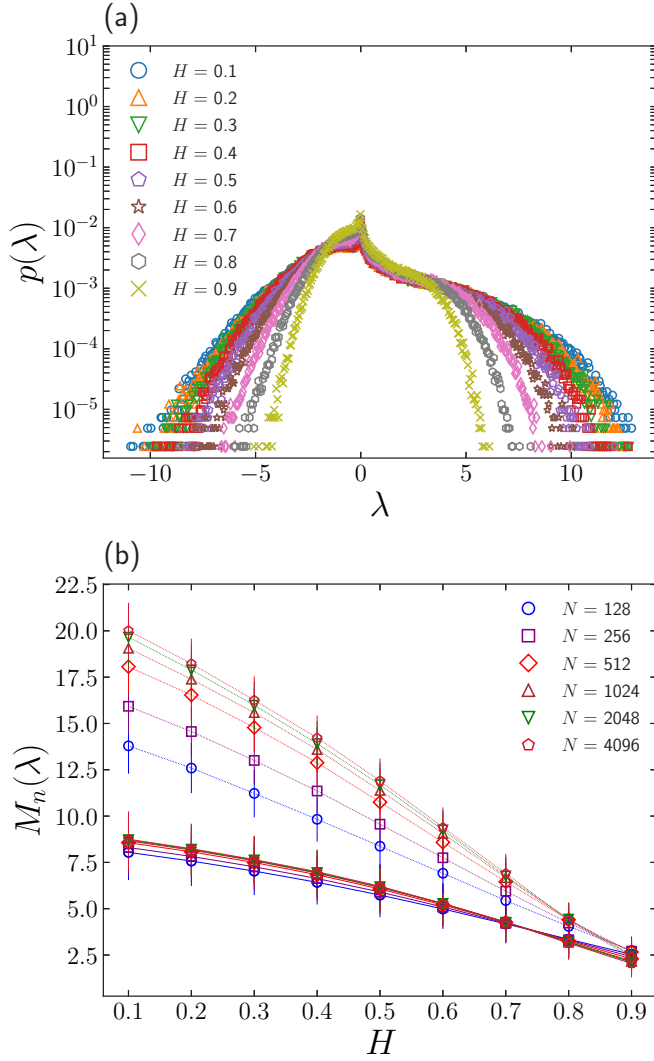
FIG. 3. (a) The probability distribution function of eigenvalue for $N = 2^{10}$ when the vertical axis is plotted in log scale. (b) The corresponding moments, M_n, for $n = 2$ (solid line) and $n = 3$ (dashed line) versus H. Different symbols are taken for various sizes.

necessary condition is equivalent to the case where data points have upward concavity. The visibility condition does not satisfy the sufficient condition, such that either $w_{i,i+2} = 1$ (creating two 2-simplices $[v_i, v_{i+1}, v_{i+2}]$ and $[v_i, v_{i+2}, v_{i+3}]$) or $w_{i+1,i+3} = 1$ (making two 2-simplices $[v_i, v_{i+1}, v_{i+3}]$ and $[v_i, v_{i+2}, v_{i+3}]$) or even $w_{i,i+2} = w_{i+1,i+3} = 1$ (constructing a 3-simplex $[v_i, v_{i+1}, v_{i+2}, v_{i+3}]$) killing the topological loop and leads the binary network to be topologically tree. Therefore, the BNVG does not contain a certain class of network motifs equivalent to topological holes, whereas in the WNVG case these local patterns can be captured as topological loops. Also mentioned loops are extracted through the filtration of the corresponding weighted simplicial complex. According to the PH language, the sufficient condition can be treated as $w_{i,i+1}, w_{i+1,i+2}, w_{i+2,i+3}, w_{i,i+3} < w_{i,i+2}, w_{i+1,i+3}$; therefore, the associated topological motifs now remain. Subsequently, we only focus on the topological aspects of WNVGs. Using this terminology, we can analyze local behaviors of a generic

fGn using the statistics of corresponding patterns (topological motifs) in WNVGs which are presented as PPs in PD.

Now, we are going to evaluate topological measures for WNVG constructed from a generic time series. Since in this paper, we are interested in examining the capability of the PH technique in determining the Hurst exponent of a typical self-similar process, therefore we take fGn which is an increment of fBm. Classifying the stationary and nonstationary series is in principle possible when the weighted network is constructed from time series in our approach. Intuitively, by increasing the value of the Hurst exponent, the underlying series becomes more smooth and the statistics of weights grow for lower values of weight. For an fBm data set which is the profile set of fGn [36,66], the maximum value of distribution function of weights occurs at lower weight values. Subsequently, the PH of WNVG which is represented by topological curves can recognize the stationary and nonstationary series. The constructed networks according to the visibility graph possess the footprint of the Hurst exponent empirically demonstrated in [45]. The higher value of H causes making the denser network. Also, the number of k-simplices at lower value of threshold increase in the weighted version of the visibility graph; on the other hand, such simplices influence the evolution of k-holes in associated simplicial complexes. Therefore, the coefficients of topological curves depend on Hurst exponent. Hereafter, we consider fGn as our case study. Panel (a) of Fig. 4 indicates the β_0 as a function of filtration parameter (threshold) for clique complexes of WNVG produced for various time series with different values of Hurst exponent. To reduce the effect of sample size, we divide the Betti numbers by the sample size and call these *normalized Betti numbers*. By increasing the value of the Hurst exponent, the normalized β_0 versus threshold decreases. Namely, the number of connected components in the network decreases with increasing H (the graphs become steeper). Interestingly, the slope of normalized β_0 as a function of w is different for different H; consequently, the WNVG of fGn sets reaches to the connected regime (path-connected), $\beta_0 = 1$ [in panel (a) of Fig. 4, we take $N = 2^{10}$], by different rates and at different thresholds, w_0. Panel (b) of Fig. 4 represents the value of w_0 as a function of Hurst exponent for different network sizes. As depicted in the mentioned panel, the value of w_0 behaves as a sample-size-dependent quantity and by increasing H, such dependency becomes negligible.

We define the $\beta_k^{(\text{birth})}(w)$ and $\beta_k^{(\text{death})}(w)$ as the number of k-dimensional holes that are born and die at a given threshold, w, respectively. It turns out that $\beta_0^{(\text{birth})} = 0$ for $w > 0$, since at $w = 0$ the underlying network has N connected components; therefore, all connected components are born at $w = 0$. Panel (c) of Fig. 4 illustrates the $\beta_0^{(\text{death})}/N$ as a function of w^2. As shown in this plot, one of the proper fitting functions to describe the normalized $\beta_0^{(\text{death})}$ is given by $\beta_0^{(\text{death})}(w) \sim \exp[-\alpha_0^{(\text{death})} w^2]$ for $2 \lesssim w^2 \lesssim w_{\max}^2$, where the value of $w_{\max}$ depends on the H value. The $\alpha_0^{(\text{death})}$ depends on the Hurst exponent as increasing function and it behaves as an almost independent function on size of series [panel (d) of Fig. 4].

Panel (a) of Fig. 5 is devoted to β_1/N for various Hurst exponents. As we expect, for a trivial threshold, $w = 0$, we have N connected components ($\beta_0 = N$) and therefore the
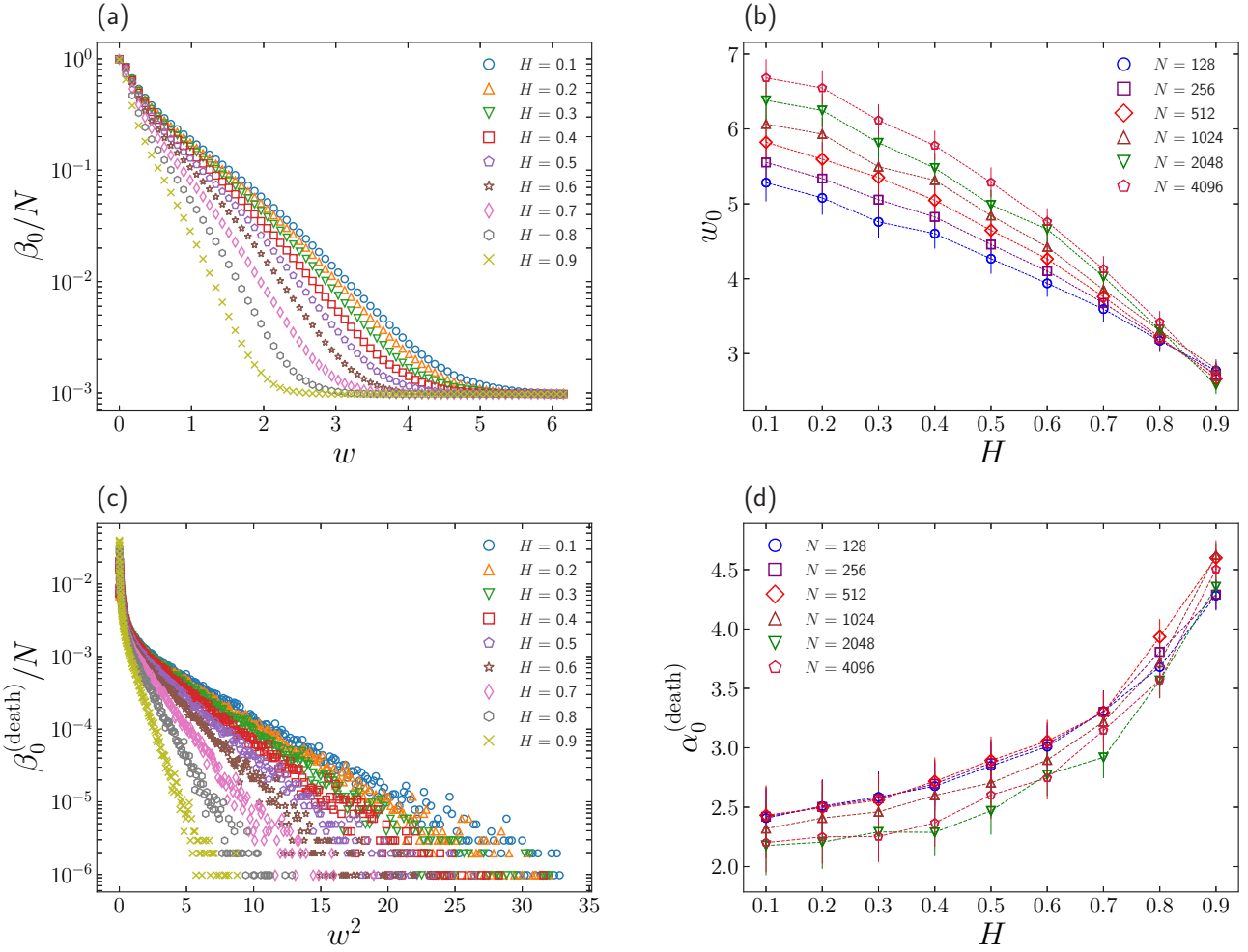
FIG. 4. (a) The normalized β_0 curve in log scale for clique complexes of WNVGs associated with fGns of various Hurst exponents versus threshold. (b) The w_0 as a function of H. (c) The normalized $\beta_0^{(\text{death})}$ in log scale as a function of w^2 (see the text) for various Hurst exponents as a function of threshold. (d) The value of $\alpha_0^{(\text{death})}$ as a function of Hurst exponent.

number of loops is identically zero ($\beta_1 = 0$). By increasing the threshold, the higher value of the Hurst exponent leads to a more rapidly increasing rate of β_1. On the other hand, for a high enough value of the threshold, again the underlying data set behaves like a topological tree without any topological loops. Therefore, the normalized β_1 goes asymptotically to zero, and such descending is more rapid for higher H. We also determine the lowest nontrivial threshold for which there is no loop in the underlying network denoted by w_1, and depict this threshold versus H for different samples size in panel (b) of Fig. 5. The w_1 is also a size-dependent quantity. Comparing the β_0/N and β_1/N demonstrate that by increasing threshold value, the WNVGs of the fGn series reach the loopless regime ($\beta_1 = 0$) before appearing in the connected regime (for which $\beta_0 = 1$), irrespective of the Hurst exponent, i.e., $w_0(H) > w_1(H)$. The quantities $\beta_1^{(\text{birth})} \propto \exp[-\alpha_1^{(\text{birth})} w]$, $\beta_1^{(\text{death})} \propto \exp[-\alpha_1^{(\text{death})} w]$ versus thresholds and corresponding coefficients are illustrated in panels (c)–(f) of Fig. 5, respectively. The $\alpha_1^{(\text{birth})}$ and $\alpha_1^{(\text{death})}$ are almost size-independent and they grow by increasing H. The local upward concavity behavior of time series satisfying the loop condition and associated topological motifs (1-holes) appears

and disappears earlier as the Hurst exponent of the time series increases.

Another interesting property to assess is the probability distribution of lifetime for 1-holes which is the difference between the death and birth thresholds of a typical measure in a 1-homology class. Panel (a) of Fig. 6 shows the probability distribution of topological 1-dimensional hole lifetime, ℓ_1, for various synthetic data sets with different values for the Hurst exponent when the vertical axis is plotted in log scale. Our results confirm that $p(\ell_1) \propto \exp[-\alpha_1^{(\text{lifetime})} \ell_1]$. The H dependency of $\alpha_1^{(\text{lifetime})}$ for various systems sizes is depicted in panel (b) of Fig. 6. This result confirms that $\alpha_1^{(\text{lifetime})}$ can be considered as a robust measure for determining the Hurst exponent of the fGn signal which is not affected by sample size even compared to $\alpha_0^{(\text{death})}$, $\alpha_1^{(\text{birth})}$, and $\alpha_1^{(\text{death})}$. The increasing behavior of the lifetime coefficient $\alpha_1^{(\text{lifetime})}$ versus the Hurst exponent also indicates that the anticorrelated signals ($H < 0.5$) contain more persistent topological motifs (1-holes). This behavior also demonstrates a considerable difference between the amount of visibility between the inner and the outer nodes which are corresponding to
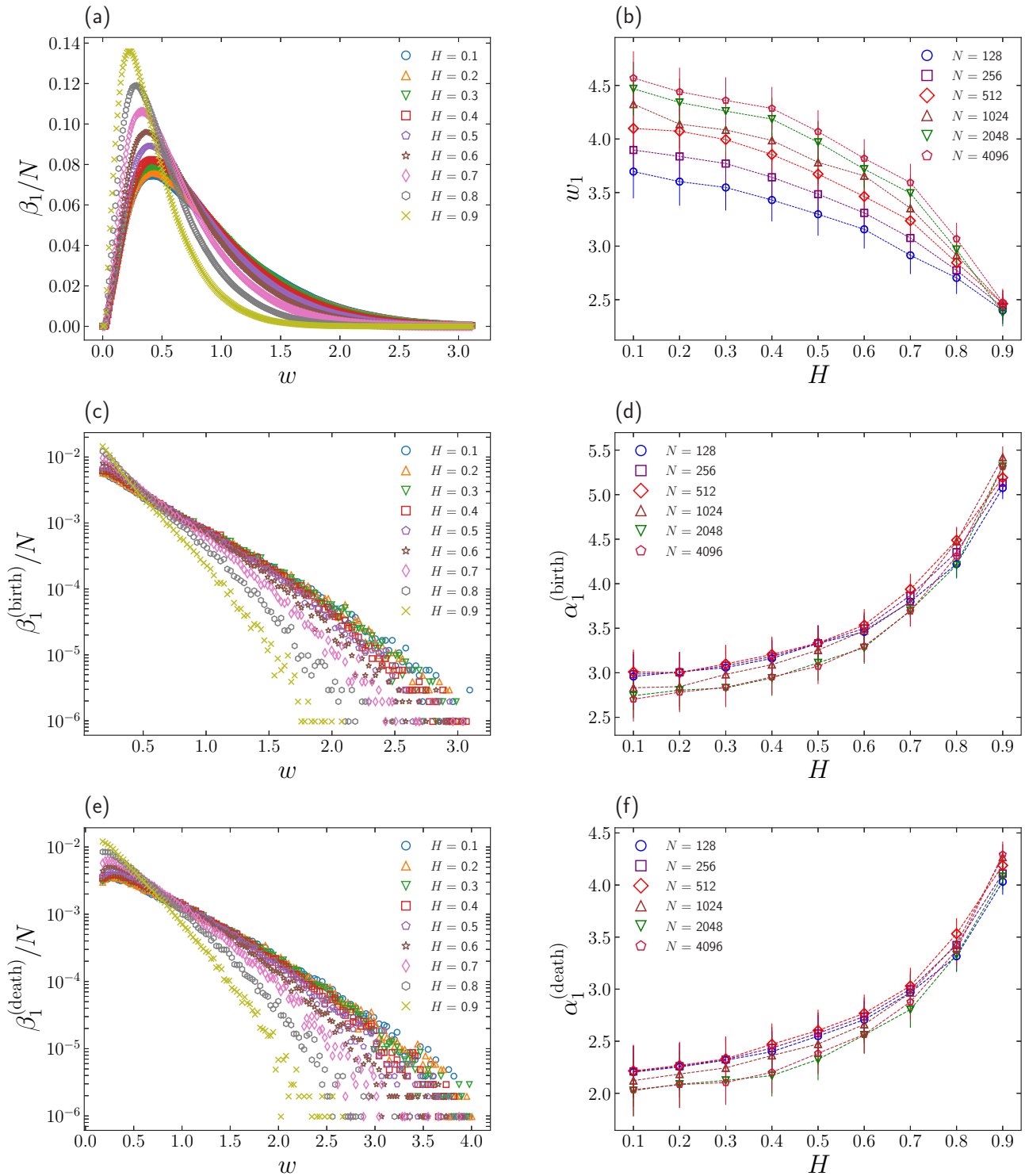
FIG. 5. (a) The normalized β_1 curve for clique complexes of WNVGs associated with fGns of various Hurst exponents versus threshold. (b) The w_1 as a function of H for different lengths of data set. (c) The distribution of the persistence diagram in log scale versus the birth axis as a function of threshold. (d) The corresponding coefficient of $\beta_1^{(\text{birth})}$ in terms of threshold known as $\alpha_1^{(\text{birth})}$ as a function of Hurst exponent. Panels (e) and (f) are the same as panels (c) and (d), respectively, just for dying 1-hole statistics. In this plot for computing β_1, we took $N = 2^{10}$.

the local upward concavity behaving data points in these signals.

Figure 7 indicates the persistence diagram (PD) and persistence barcode (PB) for three types of fGn signals for (a) $H = 0.2$ (anticorrelated), (b) $H = 0.5$ (uncorrelated), and (c)

$H = 0.8$ (correlated). The open circle and triangle symbols correspond, respectively, to the 0- and 1-homology groups in a persistence diagram for WNVGs of fGn. Each symbol depicts a pair $(w_{\text{birth}}, w_{\text{death}})$ of a k-dimensional hole. As expected, all 0-holes are born in $w_{\text{birth}} = 0$, and also always $w_{\text{death}} > w_{\text{birth}}$.
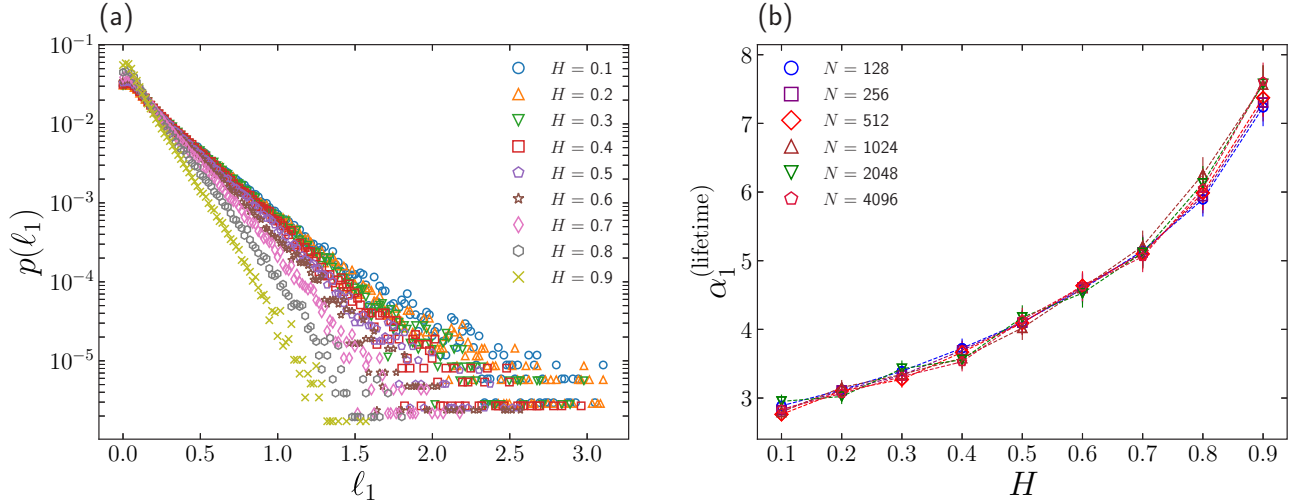
FIG. 6. (a) The probability distribution of lifetime for 1-dimensional holes. Different symbols correspond to various H. This diagram has been obtained by doing an ensemble average. We set the vertical axis in log scale. (b) The $\alpha_1^{(\text{lifetime})}$ coefficient for different sizes represented by different symbols.

In the barcode representation (inset plot), the horizontal lines (blue lines for $k = 0$ and orange lines for $k = 1$) start and end on the threshold values on which k-dimensional holes are born and die, respectively.

The associated persistence entropies (PEs) defined by Eq. (D4) are obtained using the persistence diagram. Panels (a) and (b) of Fig. 8 depict the PE_0 and PE_1, as a function of sample size, respectively. We compute persistence entropy for all available Hurst exponent values represented by different symbols. Our results demonstrate that $PE_k = \mathcal{A}_k(H) \log_{10} N$ for $k = 0, 1$. The behavior of prefactor $\mathcal{A}_k$ versus H is represented in panel (c) of Fig. 8. The $\mathcal{A}_0$ is almost an increasing function versus Hurst exponent, while the $\mathcal{A}_1$ becomes constant.

IV. IMPLEMENTATION ON REALISTIC DATA

To illustrate the applicability of PH to characterize the self-similar nature of real data, in this section, we use the physiological data set studied in [67]. We implement the DFA

algorithm to verify the scaling behavior of their correlation functions. Then, we apply our approach to the channel F8 of the standard 10–20 electroencephalography (EEG) montage recorded from 14 patients (7 males: 27.9 ± 3.3 years, 7 females: 28.3 ± 4.1 years) with paranoid schizophrenia and 14 healthy controls (7 males: 26.8 ± 2.9, 7 females: 28.7 ± 3.4 years). The EEG data were recorded in all subjects during an eyes-closed resting-state condition for 15 minutes with the sampling frequency of 250 Hz. We construct the WNVG from the EEG signals and the evolution of topological motifs through the filtration process is examined as depicted in Fig. 9. To infer the statistical significance of the results, we carry out the KS test and we obtain the exponential behavior of results indicated in Fig. 9. Then, according to the likelihood approach, we estimate the topological coefficients and corresponding error bars. The value of coefficients $\alpha_1^{(\text{birth})}$, $\alpha_1^{(\text{death})}$, $\alpha_1^{(\text{lifetime})}$ and associated Hurst exponents for both healthy and schizophrenic cases are reported in Table I. Comparing the topological coefficients with our results for estimation of the Hurst exponent (Figs. 5 and 6) enables us to compute
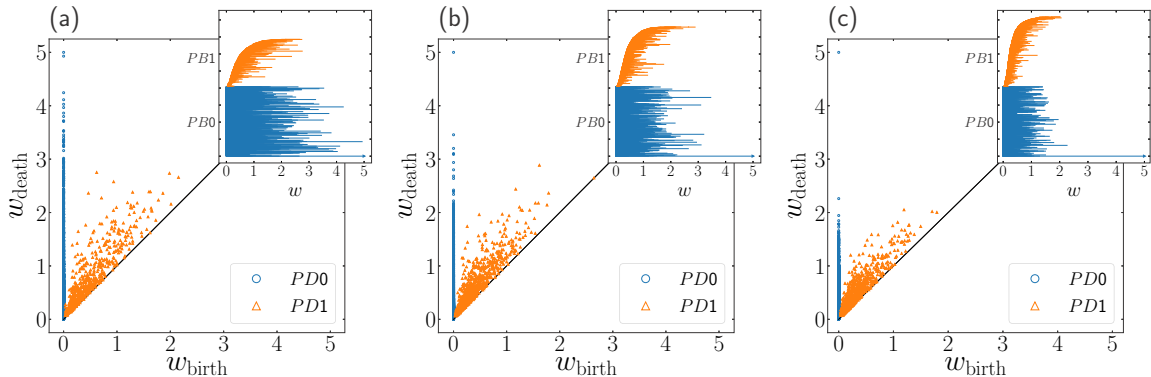

FIG. 7. The persistence diagram and persistence barcode (inset) of weighted clique complex of WNVG corresponding to fGn with (a) $H = 0.2$ (anticorrelated noise), (b) $H = 0.5$ (uncorrelated noise), and (c) $H = 0.8$ (correlated noise). The blue circle and orange triangle symbols are considered for the 0-dimensional and 1-dimensional homology generators, respectively.
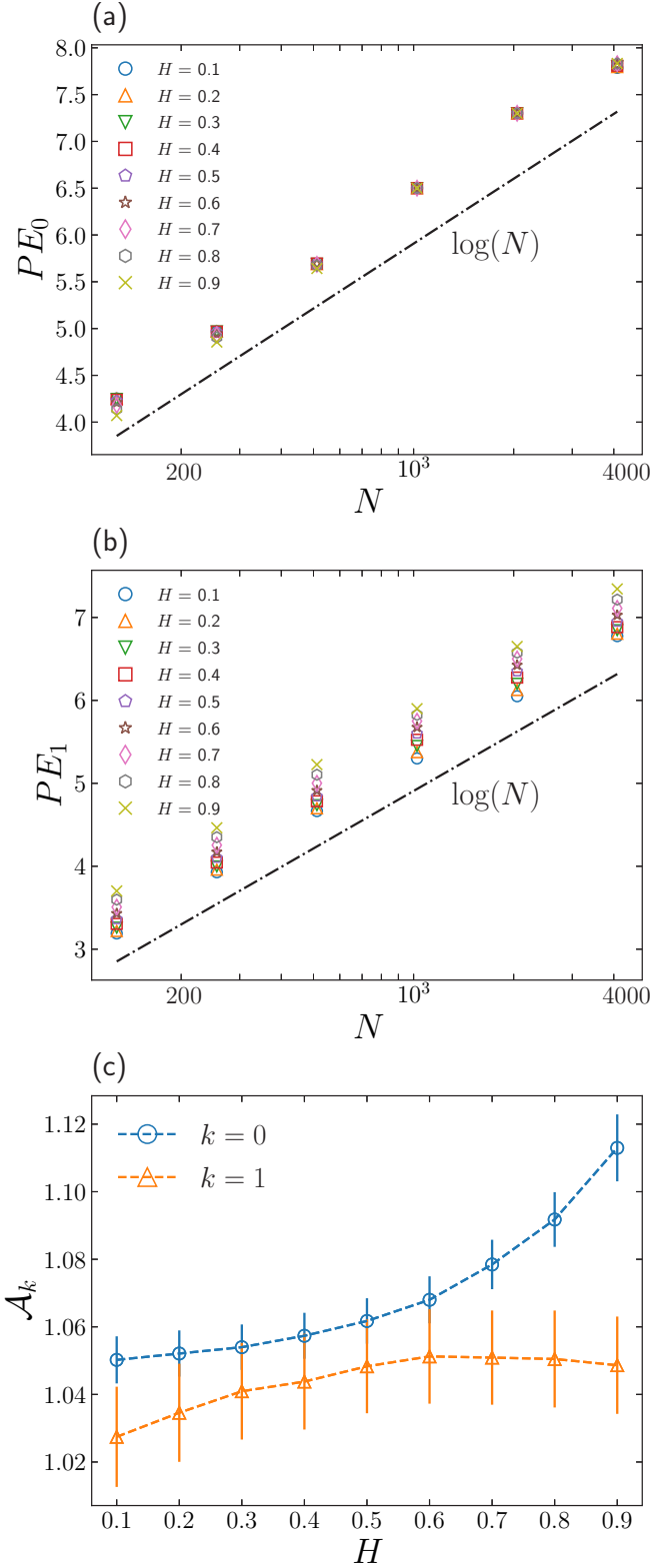
FIG. 8. (a) The persistence entropy for the 0-homology group, PE_0. (b) The PE_1 for 1 hole as a function of sample size in the log scale. Different symbols correspond to various values of the Hurst exponent. (c) The prefactor of persistence entropy as a function of H computed by the ensemble average.

TABLE I. The value of topological coefficients and corresponding Hurst exponents for healthy and schizophrenic cases at 1σ level of confidence.

Measure	Healthy	Schizophrenic
$\alpha_1^{(\text{birth})}$	4.85 ± 0.04	3.52 ± 0.05
$H_{(\text{birth})}$	0.84 ± 0.02	0.66 ± 0.03
$\alpha_1^{(\text{death})}$	3.96 ± 0.04	2.83 ± 0.04
$H_{(\text{death})}$	0.85 ± 0.02	0.68 ± 0.03
$\alpha_1^{(\text{lifetime})}$	7.03 ± 0.03	5.00 ± 0.03
$H_{(\text{lifetime})}$	0.89 ± 0.02	0.73 ± 0.03

the associated Hurst exponent. The value of the Hurst exponent for each coefficient is reported in Table I. Averaging on three kinds of Hurst exponents for EEG signals leads to $H = 0.86 \pm 0.02$ and $H = 0.69 \pm 0.03$ at the 1σ confidence interval for healthy and schizophrenic cases, respectively. Our results also reveal a correlated behavior in both data sets. It is worth noting that the PH can classify different types of data.

V. SUMMARY AND CONCLUDING REMARKS

Applying the various tools developed in network science on a network constructed from a typical time series reveals nontrivial information. Particularly, some of the features of the network for a generic fractional Gaussian noise (fGn), characterized by Hurst exponent H, exhibit a promising relation to the corresponding Hurst exponent. In this paper, we developed a method to generate a WNVG from the fGn (an increment of fBm) series which is a more general network compared to a binary network class. According to the filtration process, the evolution of topological holes in the associated WNVG can be captured. To this end, we used the PH technique, to examine the topological motifs.

The statistical properties of the WNVG were analyzed using the standard network measures. Our main results are as follows:

(1) The probability distribution functions of eigenvector, betweenness, and closeness centralities computed for WN-VGs constructed from fGn series [panel (b) of Fig. 1] do not depend on H and even the overall shape of mentioned distributions is highly size-dependent. The main message is that such kinds of statistical measures are not proper measures for determining the Hurst exponent of the fGn series (Fig. 2).

(2) Increasing the amount of correlation results in obtaining more dense networks and shifts the weights of the links to the lower values and, consequently, the width of the distribution function of the eigenvalues to be tighter. The nth moment of this distribution behaves as an H-dependent quantity. The higher the order of moments, the stronger the dependency on sample size (Fig. 3).

In the second part, we considered the topological properties of synthetic fGn. Our main results are summarized below:

(3) The corresponding BNVG of a typical fGn is a *topological tree* which means that all topological motifs have identically vanished, while taking into account a weight function to construct the so-called weighted NVG (WNVG) leads
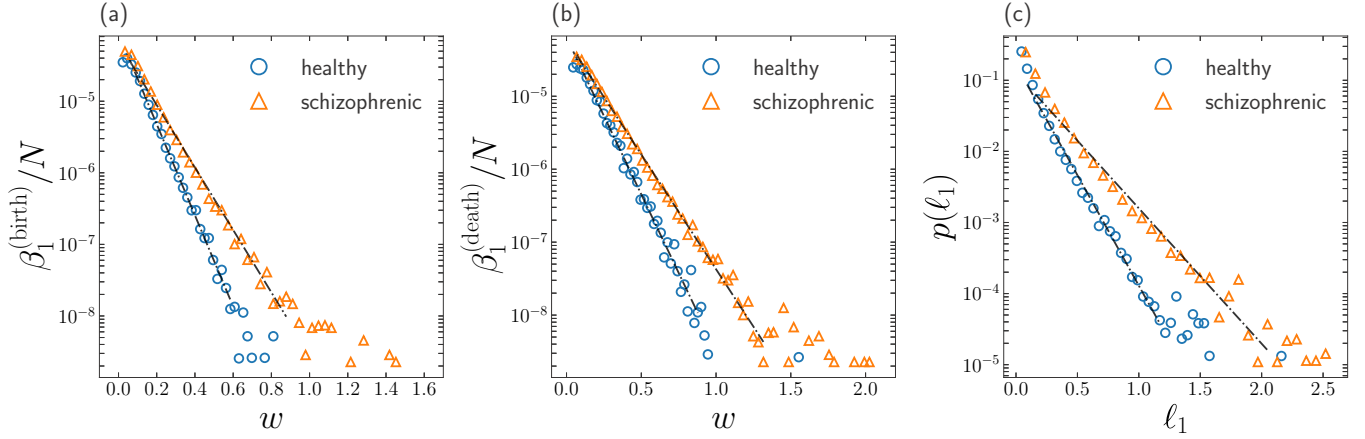
FIG. 9. (a) The distribution of persistence diagram versus birth axis ($\beta_1^{(\text{birth})}$) in log scale for clique complexes of WNVGs associated with EEG signals for healthy (blue circle symbols) and schizophrenic cases (orange triangle symbols) as a function of threshold. (b) The $\beta_1^{(\text{death})}$ in log scale as a function of threshold. (c) The lifetime distribution in log scale for 1-homology generator statistics.

to the survival of topological motifs and therefore their evolution with respect to self-similar exponents can be examined. Such evolution reveals a robust feature for determining the Hurst exponent reliably.

(4) The Betti-0 is a representative of the number of connected components of the weighted clique simplicial complex associated with the WNVG with respect to the threshold (weight). The decreasing rate of Betti-0 increases by increasing H. The threshold value for which the underlying WNVG reaches the connected regime (path-connected) indicates an H dependency [panels (a) and (b) of Fig. 4].

(5) Our result also confirms that the coefficient corresponding to the exponentially decaying function of the number of connected components dying (merging), $\alpha_0^{(\text{death})}$, as a function of threshold, depends on the Hurst exponent and interestingly it is almost size-independent [panels (c) and (d) of Fig. 4].

(6) The statistics of 1-holes (topological loops) show that the vanishing threshold for the number of topological loops, w_1, is H-dependent and it contains the sample size effect. The coefficients corresponding to the exponentially decaying function of the number of topological loops appearing ($\alpha_1^{(\text{birth})}$) and disappearing ($\alpha_1^{(\text{death})}$) as a function of threshold reveal the proper criteria for measuring the Hurst exponent (Fig. 5). These coefficients also indicate that the topological motifs (upward concavity behavior data points of time series satisfying loop condition) evolve (appear and disappear) in low weights for correlated signals.

(7) The probability distribution of the lifetime of 1-holes (ℓ_1) confirms that the corresponding coefficient ($\alpha_1^{(\text{lifetime})}$) is an increasing function versus H emphasizing that the sample size effect is completely diminished in this quantity. Consequently, for a self-similar time series in the absence of trends, this coefficient can be a reliable measure for estimating the Hurst exponent even for a small sample size irrespective of the value of the H exponent (Fig. 6). This coefficient also indicates that the fGn signals living in negatively correlated regimes incorporate more robust topological motifs (1-holes) which shows a significant difference between the quality of visibility of inner and outer nodes corresponding

to the local upward concavity behaving data points in these signals.

(8) The persistence pairs (PPs) in the persistence diagram (PD) and the persistence barcode (PB) for the weighted clique complex of WNVGs which are indicators of persistent homology have been computed and with increasing the value of the Hurst exponent shrink to the origin of the coordinate (Fig. 7). We also computed persistence entropies (PEs) for 0-homology and 1-homology groups. Both quantities depend on sample size as expected and the corresponding slopes in semilog scale were almost H-dependent (Fig. 8). Implementing the PH method on real data and computing the topological parameters, we could determine the corresponding Hurst exponent. Interestingly, PH admitted its capability for the classification of various data sets.

(9) Employing PH on the constructed weighted network from the realistic series confirmed the correlated behavior of electroencephalography for both healthy and schizophrenic samples.

Finally, we emphasize that the behavior of the local (statistical) observables depends weakly on H, whereas the exponents of global (topological) observables are almost strongly H-dependent and even for the $\alpha_1^{(\text{lifetime})}$, the size effect is completely diminished. A take-home message is that the TDA provides a new type of measure to quantify the Hurst exponent and, therefore, the scaling exponents of the correlation function, power spectrum, and fractal dimension can be specified with reliable approaches.

In this paper, we have not verified whether the persistent homology technique is capable of recognizing that the underlying time series is a self-similar set or not. Indeed, this purpose is beyond the scope of this paper. To address the capability of our method to discriminate between stationary and nonstationary times series, more simulations for different types of nonstationary cases should be performed; this is left for our future study. Also, it would be interesting to examine the effect of trends and irregularity which may occur in a wide range of events in nature in the context of the TDA and more precisely via the PH approach, and we leave this too for future research. The above analysis can be done on dif-

ferent phenomena ranging from cosmology, astrophysics, and economics to biology [32,68,69] with different approaches to make graphs from time series.

APPENDIX A: SYNTHETIC FRACTIONAL GAUSSIAN NOISE

To model the stochastic fractal processes, Mandelbrot and Van Ness introduced the fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) [31]. The theory of fBm is a mathematical generalization of the classical random walk and Brownian motion [31,70]. A 1-dimensional fBm is represented by $B \equiv \{B(t) : t \geqslant 0\}$, with power-law variance, for which $\text{Var}[B(at)] \triangleq \text{Var}[a^H B(t)] = a^{2H}\text{Var}[B(t)]$, where $H \in (0, 1)$ is called the Hurst exponent. For this random force, the Markov property and the stationarity are violated (note that when we have domain Markov property, stationarity, and continuity for a time series, then it should be proportional to a 1-dimensional Brownian motion). A model for generating fBm (denoted by B_H to emphasize its Hurst exponent H) is a generalization of the Brownian motion which is nonstationary and non-Markovian [71], and is given by the Holmgren-Riemann-Liouville fractional integral,

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t - s)^{H - \frac{1}{2}} dB(s), \quad \text{(A1)}$$

where Γ is the Gamma function, $dB(s) \equiv B(s + ds) - B(s)$ is the increment of 1-dimensional Brownian motion, and it has the following covariance:

$$\langle B_H(t)B_H(s) \rangle = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}), \quad \text{(A2)}$$

where $\sigma^2 \equiv \langle B(0) \rangle$ and also $\langle B_H \rangle = 0$. The increments, $x_H(t) \equiv \delta_t(B_H(t + \delta t) - B_H(t))$, are known as fractional Gaussian noise (fGn). The power spectrum of fBm and fGn behaves as $S(f) \sim f^{-\xi}$, where $\xi(H) = 2H + 1$ and $\xi(H) = 2H - 1$ for fBm and fGn, respectively. For $H > \frac{1}{2}$ ($H < \frac{1}{2}$) the corresponding fGn is positively (negatively) correlated. According to the results provided by detrended fluctuation analysis (DFA), the relation between the scaling exponent derived by the DFA method and associated Hurst exponent of fBm is $H + 1$, while constructing the fGn series by making the increment of fBm confirmed that the scaling exponent of fluctuation functions computed by DFA is directly related to the Hurst exponent [36,66].

APPENDIX B: STATISTICAL NETWORK ANALYSIS

Suppose that a network (graph) is represented by $G = (V, E, w)$. Here, $V \equiv \{v_i\}_{i=1}^N$ is a node (vertex) set, $E = V \times V \equiv \{e_{ij} = (v_i, v_j)|v_i, v_j \in V\}_{i,j=1}^N$ is a link (edge) set, and

$w : E \to \mathbb{R}$ is a weight function (threshold). Subsequently, the degree of ith node (v_i) is the number of nodes straightly connected with the underlying node by nonzero weight, and it is denoted by $k_i \equiv \sum_j (1 - \delta_{0,w_{ij}})$. The degree centrality (c_i^D) is defined by $\frac{k_i}{N-1}$, which is apparently related directly to how important the underlying node is, since it is the number of agents that have a connection with it. An important function concerning this quantity is the degree distribution showing the probability distribution function for degree of all nodes in the network:

$$p(k) = \frac{1}{N} \sum_{i=1}^N \delta_{k,k_i}. \quad \text{(B1)}$$

The eigenvector centrality (c_i^E) is also defined via the eigenvalue equation

$$\lambda c_i^E = \sum_{j=1}^N w_{ij} c_j^E, \quad \text{(B2)}$$

where λ is the eigenvalue. The maximum value of the λ spectrum, i.e., $\lambda_{\max}$ plays the dominant role in the network properties, the corresponding eigenvectors of which are denoted by $c_{i,\max}^E$ revealing the importance of the nodes. Let us denote the shortest distance between nodes v_i and v_j by d_{ij} which is assumed to be N when there is no path connecting them (disconnected graphs). Then the closeness centrality is defined by

$$c_i^C = \frac{N - 1}{\sum_{j=1}^N d_{ij}}, \quad \text{(B3)}$$

and the betweenness centrality is as follows:

$$c_i^B = \frac{2}{(N - 1)(N - 2)} \sum_{j=1, j \neq i}^N \sum_{k=1, k \neq i, j}^N \frac{n_{jk}(i)}{n_{jk}}, \quad \text{(B4)}$$

where n_{jk} is the number of geodesics from v_j to v_k, and $n_{jk}(i)$ is the number of geodesics from v_j to v_k which pass through node v_i.

APPENDIX C: ALGEBRAIC TOPOLOGY

Topology generally refers to the global features in contrast to the geometrical invariants of underlying objects or sets. We have two spaces represented by X and Y, and they have the same local (geometric) features if any relevant features are invariant under *congruence*, while the mentioned spaces are topologically equivalent if the associated features are invariant under *homeomorphisms*. In other words, they are homeomorphic. Homology theory plays a crucial role in the mathematical description of the relevant building blocks of a typical topological space and reveals the connectedness of underlying space [10,59,60]. Based on such properties, for the sake of clarity, we will give a brief review on the building blocks of algebraic topology which are useful to set up homology groups.

Simplex. A k-simplex (σ_k) is a convex hull of any geometrically independent subset, accordingly $\sigma_k \equiv [v_0, v_1, \ldots, v_k] \subseteq \mathbb{R}^D$. By this definition, a 0-simplex is a point, a 1-simplex is a
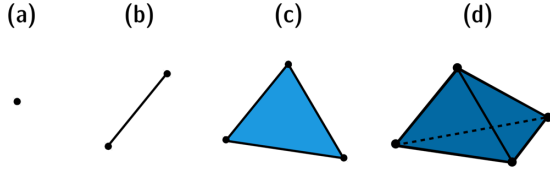
(a) (b) (c) (d)

FIG. 10. Low-dimensional simplices: (a) 0-simplex (point), (b) 1-simplex (line segment), (c) 2-simplex (filled triangle), and (d) 3-simplex (filled tetrahedron).

FIG. 11. Betti numbers of some topological spaces: point, line, circle, sphere, torus, 2-torus.

segment of a line, a 2-simplex is a filled triangle, a 3-simplex is a filled tetrahedron, and so on (Fig. 10).

Face. An l-simplex which is denoted by σ_l is a subset of the k-simplex ($\sigma_l \subseteq \sigma_k$) and it is called the l-face of the k-simplex.

Simplicial complex. A simplicial complex (ψ) is a collection of simplices such that any l-face of any k-simplex of a typical complex ($0 < l < k$) is a member of the complex. In addition, the nonempty intersection of any two simplices, σ_k, and σ_m, from the complex is an l-face of both simplices. The dimension of a complex is the maximum dimension of all simplices of the complex. According to the definition of the complex, one can define its k-ordered subcollection of complexes ψ as follows:

$$\Sigma_k(\psi) \equiv \{\sigma \in \psi \mid \dim(\sigma) = k\}. \tag{C1}$$

Chain. For a given simplicial complex, a k-chain (k-dimensional chain) is a linear combination of k-simplices of ψ, defined by

$$c_k \equiv \sum_{i=1}^{|\Sigma_k(\psi)|} a_i\, \sigma_k^{(i)}, \quad \sigma_k^{(i)} \in \Sigma_k(\psi), \tag{C2}$$

where $|\Sigma_k(\psi)|$ corresponds to the cardinality of the k-ordered subcollection of complexes and the coefficients, a_i, belong to a field, which is usually considered as $\mathbb{F} = \mathbb{Z}_2 \equiv \{0, 1\}$. The collection of all possible k-chains in the simplicial complex is called the k-chain group as

$$C_k(\psi) \equiv \left\{ c_k \,\middle|\, c_k = \sum_{i=1}^{|\Sigma_k(\psi)|} a_i\sigma_k^{(i)}; \sigma_k^{(i)} \in \Sigma_k(\psi), a_i \in \mathbb{F} \right\}. \tag{C3}$$

Boundary operator. For the simplices in any dimension, the boundary operator ∂_k is an operator mapping σ_k to its boundary according to

$$\partial_k(\sigma_k) \equiv \sum_{j=0}^{k} (-1)^j\, [x_0, x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k] \subseteq \sigma_k. \tag{C4}$$

Boundary. A k-chain which is the boundary of a $(k+1)$-chain is called the k-boundary, denoted by b_k. The k-boundary group is the collection of all k-boundaries in complex ψ,

$$B_k(\psi) \equiv \{c_k \in C_k(\psi) | \exists c_{k+1} \in C_{k+1}(\psi); \partial_{k+1}(c_{k+1}) = c_k\}$$
$$\equiv \{b_k^{(i)}\}_i^{|B_k(\psi)|} \subseteq C_k(\psi). \tag{C5}$$

Cycle. A k-chain that has no boundary is called a k-cycle denoted by z_k as

$$\partial_k(z_k) = \varnothing. \tag{C6}$$

The k-cycle group is defined as the collection of all k-cycles in complex ψ,

$$Z_k(\psi) \equiv \{c_k \in C_k(\psi) \mid \partial_k(c_k) = \varnothing\}$$
$$\equiv \{z_k^{(i)}\}_i^{|Z_k(\psi)|} \subseteq C_k(\psi). \tag{C7}$$

Since "boundaries have no boundary," therefore, we have

$$\partial_k(b_k) = \partial_k(\partial_{k+1}(c_{k+1})) = \varnothing; \tag{C8}$$

hence

$$B_k(\psi) \subseteq Z_k(\psi) \subseteq C_k(\psi). \tag{C9}$$

Homology group. The k-homology group is defined by the quotient group of the k-cycles group by the k-boundary group,

$$H_k(\psi) \equiv Z_k(\psi)/B_k(\psi). \tag{C10}$$

The kth *Betti number* of a simplicial complex, denoted by $\beta_k(\psi)$, is a topological invariant which counts the number of k-homology classes corresponding to the number of k-dimensional holes of complex ψ (Fig. 11),

$$\beta_k(\psi) \equiv \dim[H_k(\psi)]. \tag{C11}$$

The clique (flag) simplicial complex of an unweighted (binary) network, $G = (V, E, w \in \{0, 1\})$, is a simplicial complex, denoted by $\psi^{(G)}$, such that any k-simplex of each dimension in the complex corresponds to a $(k+1)$-clique in the network and vice versa (Fig. 12). A binary network is a *topological tree* if and only if it is topologically holeless in all dimensions. Namely, the associated clique simplicial complex
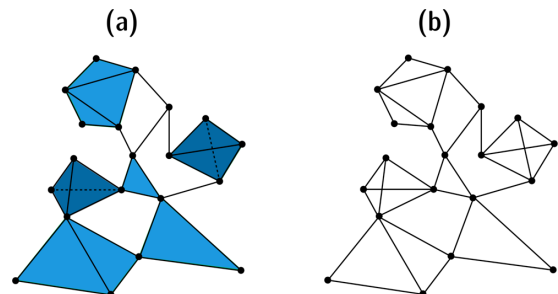

(a) (b)

FIG. 12. (a) Clique complex of (b) a typical network.

has the following property:

$$\beta_k(\psi^{(G)}) = \min(\beta_k) = \begin{cases} 1, & k = 0, \\ 0, & k > 0. \end{cases} \quad \text{(C12)}$$

APPENDIX D: PERSISTENT HOMOLOGY

In the context of topological data analysis (TDA), we are interested in statistical analysis of the structure in data. Generally, in TDA-based analysis, data of any type (point cloud data, scalar field, time series, network, etc.) when mapped to a weighted simplicial complex are worked out topologically in terms of the parameters present inherently in the original data, e.g., the weight of links in a weighted network, or the pairwise distance between data points in point cloud data. More precisely, TDA maps parameter-dependent data, $X(w)$, (where w is a typical parameter, like the threshold value for any weighted network) to a weighted simplicial complex, $\psi^X(w)$. Such approach produces the chain group, the cycle group, the boundary group, the homology group, and particularly, the kth Betti number [7,10].

Persistent homology (PH) as a powerful tool of TDA examines the creation (birth) and destruction (death) of topological invariants associated with homology classes during a mathematical process called filtration [10]. Filtration, ϕ, is a nested sequence of weighted complex $\psi(w)$ in which any complex with a distinct weight is a subcollection of any complex with higher weight,

$$\phi(\psi(w)) \equiv (\psi(w) \mid \forall w' < w'' : \psi(w') \subseteq \psi(w''))_{w_{\min}}^{w_{\max}}. \quad \text{(D1)}$$

More precisely, the PH technique enumerates the kth Betti number of any subcomplex in ϕ and assigns an ordered tuple $w^{(h_k)} \equiv (w_{\text{birth}}^{(h_k)}, w_{\text{death}}^{(h_k)})$ to the existing k-dimensional topological hole. Here $w_{\text{birth}}^{(h_k)}$ and $w_{\text{death}}^{(h_k)}$ are the thresholds for which h_k appears (birth) and disappears (death), respectively. Since $w_{\text{birth}}^{(h_k)} < w_{\text{death}}^{(h_k)}$, we can define the positive-value quantity $\ell^{(h_k)} \equiv w_{\text{death}}^{(h_k)} - w_{\text{birth}}^{(h_k)}$ as persistency (lifetime) of a k-dimensional hole. The persistence barcode (PB) and equivalently the persistence diagram (PD) are the famous representations of PH. As an illustration, the k-dimensional persistence diagram of weighted complex $\psi(w)$ is a multiset $PD_k(\phi(\psi(w))) \equiv (\mathcal{M}, \mathcal{N})$, where

$$\mathcal{M} \equiv \{w^{(h_k)} \mid h_k \in H_k(\psi(w)), \; \psi(w) \in \phi\}, \quad \text{(D2)}$$

and $\mathcal{N} : \mathcal{M} \to \mathbb{N}$ is the count function. Inspired by Shannon entropy for a typical state probability, one can define the persistence entropy (PE) of the kth PD (PB). To this end, we construct the probability for lifetime of homology classes as

$$p(\ell^{(h_k)}) \equiv \frac{\ell^{(h_k)}}{\mathcal{L}}, \quad \mathcal{L} \equiv \sum_{w^{(h_k)} \in \mathcal{M}(PD_k)} \ell^{(h_k)}. \quad \text{(D3)}$$

Therefore, the PE for the k-dimensional topological hole is defined by [72,73]

$$\text{PE}_k = - \sum_{w^{(h_k)} \in \mathcal{M}(PD_k)} p(\ell^{(h_k)}) \log_{10} p(\ell^{(h_k)}). \quad \text{(D4)}$$

[1] H. Edelsbrunner, J. Harer, and A. Zomorodian, in *Proceedings of the Seventeenth Annual, ACM, Symposium on Computational Geometry* (ACM, Medford, MA, USA, 2001), pp. 70–79.

[2] A. Zomorodian and G. Carlsson, Discrete Comput. Geom. **33**, 249 (2005).

[3] A. J. Zomorodian, *Topology for Computing* (Cambridge University Press, Cambridge, 2005), Vol. 16.

[4] G. Carlsson, Bull. Am. Math. Soc. **46**, 255 (2009).

[5] A. Zomorodian, Adv. Appl. Comput. Topol. **70**, 1 (2012).

[6] L. Wasserman, Annu. Rev. Stat. Appl. **5**, 501 (2018).

[7] M. Nakahara, *Geometry, Topology and Physics* (Taylor and Francis, 2018).

[8] J. R. Munkres, *Elements of Algebraic Topology* (CRC Press, 2018).

[9] A. Hatcher, *Algebraic Topology* (Cambridge, Cambridge University Press, 2005).

[10] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, 2010).

[11] R. Ghrist, Bull. Am. Math. Soc. **45**, 61 (2008).

[12] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, Int. J. Shape Model. **11**, 149 (2005).

[13] A. Patel, J. Appl. Comput. Topol. **1**, 397 (2018).

[14] P. Bubenik and P. Dłotko, J. Symb. Comput. **78**, 91 (2017).

[15] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, J. Mach. Learn. Research **18**, 47 (2017).

[16] F. Chazal and B. Michel, arXiv:1710.04019.

[17] E. Munch, J. Learn. Anal. **4**, 47 (2017).

[18] D. H. Serrano, J. Hernández-Serrano, and D. S. Gómez, Chaos, Solitons Fractals **137**, 109839 (2020).

[19] M. Saggar, O. Sporns, J. Gonzalez-Castillo, P. A. Bandettini, G. Carlsson, G. Glover, and A. L. Reiss, Nat. Commun. **9**, 1399 (2018).

[20] D. Horak, S. Maletić, and M. Rajković, J. Stat. Mech.: Theory Exp. (2009) P03034.

[21] C. M. Topaz, L. Ziegelmeier, and T. Halverson, PLoS ONE **10**, e0126383 (2015).

[22] A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett, Network Neurosci. **3**, 656 (2019).

[23] P. Pranav, H. Edelsbrunner, R. Van de Weygaert, G. Vegter, M. Kerber, B. J. Jones, and M. Wintraecken, Mon. Not. R. Astron. Soc. **465**, 4281 (2017).

[24] J. Jaquette and B. Schweinhart, Commun. Nonlinear Sci. Numer. Simul. **84**, 105163 (2020).

[25] L. Speidel, H. A. Harrington, S. J. Chapman, and M. A. Porter, Phys. Rev. E **98**, 012318 (2018).

[26] V. Salnikov, D. Cassese, and R. Lambiotte, Eur. J. Phys. **40**, 014001 (2018).

[27] O. Bobrowski and P. Skraba, Phys. Rev. E **101**, 032304 (2020).

[28] Q. H. Tran, M. Chen, and Y. Hasegawa, Phys. Rev. E **103**, 052127 (2021).

[29] B. Olsthoorn, J. Hellsvik, and A. V. Balatsky, Phys. Rev. Research **2**, 043308 (2020).

[30] H. E. Hurst, Trans. Am. Soc. Civ. Eng. **116**, 770 (1951).

[31] B. B. Mandelbrot and J. W. Van Ness, SIAM Rev. **10**, 422 (1968).

[32] I. Eghdami, H. Panahi, and S. Movahed, Astrophys. J. **864**, 162 (2018).

[33] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, and D. Sornette, Rep. Prog. Phys. **82**, 125901 (2019).

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

[35] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, Chaos: Interdisc. J. Nonlinear Sci. **5**, 82 (1995).

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

[37] W.-X. Zhou *et al.*, Phys. Rev. E **77**, 066211 (2008).

[38] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Phys. Rev. E **64**, 011114 (2001).

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

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

[41] R. Nagarajan and R. G. Kavasseri, Chaos, Solitons Fractals **26**, 777 (2005).

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

[43] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, Phys. Rep. **787**, 1 (2019).

[44] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, Proc. Natl. Acad. Sci. USA **105**, 4972 (2008).

[45] L. Lacasa, B. Luque, J. Luque, and J. C. Nuno, Europhys. Lett. **86**, 30001 (2009).

[46] R. Nagarajan and R. G. Kavasseri, Physica A **354**, 182 (2005).

[47] J. Hu, J. Gao, and X. Wang, J. Stat. Mech.: Theory Exp. (2009) P02066.

[48] Z. Wu, N. E. Huang, S. R. Long, and C.-K. Peng, Proc. Natl. Acad. Sci. USA **104**, 14889 (2007).

[49] L. d. F. Costa, O. N. Oliveira Jr., G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha, Adv. Phys. **60**, 329 (2011).

[50] B. A. Gonçalves, L. Carpi, O. A. Rosso, and M. G. Ravetti, Physica A **464**, 93 (2016).

[51] Z.-K. Gao, Q. Cai, Y.-X. Yang, W.-D. Dang, and S.-S. Zhang, Sci. Rep. **6**, 35622 (2016).

[52] W.-J. Xie and W.-X. Zhou, Physica A **390**, 3592 (2011).

[53] M. Zheng, S. Domanskyi, C. Piermarocchi, and G. I. Mias, Sci. Rep. **11**, 5623 (2021).

[54] Y. Yang, J. Wang, H. Yang, and J. Mang, Physica A **388**, 4431 (2009).

[55] M. Ahmadlou, H. Adeli, and A. Adeli, J. Neural Transm. **117**, 1099 (2010).

[56] G. Bianconi and C. Rahmede, Phys. Rev. E **93**, 032315 (2016).

[57] K. Kovalenko, I. Sendiña-Nadal, N. Khalil, A. Dainiak, D. Musatov, A. M. Raigorodskii, K. Alfaro-Bittner, B. Barzel, and S. Boccaletti, Commun. Phys. **4**, 43 (2021).

[58] J.-G. Young, G. Petri, and T. P. Peixoto, Commun. Phys. **4**, 135 (2021).

[59] G. Rote and G. Vegter, in *Effective Computational Geometry for Curves and Surfaces* (Springer, 2006), pp. 277–312.

[60] A. Zomorodian, in *Algorithms and Theory of Computation Handbook*, Vol. 2 (CRC Press, 2009).

[61] J. R. Hosking, Water Resour. Res. **20**, 1898 (1984).

[62] A. B. Dieker and M. Mandjes, Prob. Eng. Infor. Sci. **17**, 417 (2003).

[63] R. B. Davies and D. Harte, Biometrika **74**, 95 (1987).

[64] See, https://networkx.org.

[65] See, https://mrzv.org/software/dionysus2.

[66] M. S. Taqqu, V. Teverovsky, and W. Willinger, Fractals **3**, 785 (1995).

[67] E. Olejarczyk and W. Jernajczyk, PLoS ONE **12**, e0188629 (2017).

[68] Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, K. Benabed *et al.*, Astron. Astrophys. **641**, A4 (2020).

[69] A. Vafaei Sadr, M. Farhang, S. Movahed, B. Bassett, and M. Kunz, Mon. Not. R. Astron. Soc. **478**, 1132 (2018).

[70] J.-P. Kahane, *Some Random Series of Functions* (Cambridge University Press, Cambridge, 1993), Vol. 5.

[71] I. S. Reed, P. Lee, and T.-K. Truong, IEEE Trans. Inf. Theory **41**, 1439 (1995).

[72] E. Merelli, M. Rucco, P. Sloot, and L. Tesei, Entropy **17**, 6872 (2015).

[73] N. Atienza, R. Gonzalez-Diaz, and M. Rucco, J. Intell. Inf. Syst. **52**, 637 (2019).