A relationist and descriptive approach to stationary time series

Aurélien Hazan, Vincent Vigneron

A relationist and descriptive approach to stationary time series

Aurélien Hazan[1] and Vincent Vigneron[1,2]

[1] TADIB, CNRS FRE 2873, 91020 Evry-Courcouronnes, France,
{aurelien.hazan,vincent.vigneron}@ibisc.univ-evry.fr
[2] CES SAMOS-MATISSE CNRS URM 8095, 90 rue de Tolbiac, 75634 Paris cedex 13, France
vigneron@univ-paris1.fr

Summary. This article addresses the issue of building discrete topological spaces from continuous data measured on a complex system and then the statistical characterization of the obtained space. As an illustration, the sensitivity of graphs properties to thresholding is analysed. A possible way to cope with that flaw is the multilevel point of view. We extend this approach to *n*-ary relations using *simplicial complexes*; statistical independence is shown to be an appropriate framework for characterizing the obtained space.

Key words: knowledge representation, random graph, binary relations, correlation, independence, simplicial complex, persistent homology

1 Introduction

In the complex systems litterature, a whole field or research is dedicated to expressing the organizational principles that shape large-scale networks [1] and their evolution [**?**].
Let's take an example. In biology, three types of macromolecular networks rule the inner organization of cells: metabolic, protein-protein, and genetic regulation networks. It has been proposed that metabolic networks should be encoded in a graph theoretic way, which allows random graph theory to characterize them [3]: the elementary relation that units two metabolites of the network is the existence of a reaction catalyzed by given enzymes. Similar principles rule intracellular processes in many organisms, in a scale-free manner that entails for example a remarkable resistance to errors. However, the great heterogeneity of reaction strengths [2] questions the rationale of using unweighted graphs to represent the network activity.
In neurophysiology, brain networks can be examined from several points of view; structural or anatomical studies on one hand and functional and effective ones on the other [**?**]. The first area deals with the physical connection at different possible scales, whether at the level of individual neurons or of brain areas. The former involves large-scale network while the latter lays stress on small-scale networks. We don't pick example in the field of structural connectivity, but rather from functional and effective connectivity cases that both examine the activity either of neurons of brain areas. The functional case favors statistical interdependence irrespective of causality while the effective connectivity case is preoccupied by causal explanation of activities of neural areas. The often quoted articles [7] illustrate the first approach. See also [**?**, **?**] in the artificial networks context.
The two examples above underline one limit of graph-like representations: the topology of the graph may strongly depend on the definition of the binary relation that conditions the existence of an edge between two nodes. One solution would be to assign weights to edges [**?**, 3] that take the stand to generalize invariants defined for unweighted graphs to weighted ones. In this article we explore an alternative proposition: first we define a threshold-dependent relation to ground the existence of edges, and to superpose several graphs for different threshold values, then we characterize the global structure.

In the following, section 2 reviews relationship representation in from the experimental context, then we state definitions of relationship between several variables in statistical terms. Section 3 is devoted to examining a detailed example of a binary relation that involves correlation and thresholding while section 4 gives a multilevel point of view on graph that allows a characterization that includes threshold shifts. Section 5 puts forward tools from computational topology by extending works on binary relationship to n-ary relationship.

2 Organizing similarities in time series

The system under study consists of a n units (we suppose n is big) whose activities are interdependent in an unknown manner. We assume this system can be *observed* by means of a finite set of scalar variables, whose values are indexed by discrete time instants. We hold those time-varying activities to be random, and stationary. The outputs are time-dependent stationary and continuous signal. The purpose of this article is to discuss the *organization*, rather than the explanation of such data. In particular, we look for a representation of the overall interaction between theses units. The meaning of the relationship between units will be rooted in statistical inference, since little is known about the processes, except their stationarity.
A first approach would be to define a function of n variables, whose behaviour would be examined and would reveal interdependence. But as n grows this method tends to become untractable. By limiting the output domain onto $\{0, 1\}$, one may state a satisfactory answer if we lay stress on the organization of a set of relationships (see sections 3–5).
Let u_i and u_j be two units whose activities are measured. To make things clear, "units u_i and u_j are related" is often understood as "*correlated enough*". However this term says nothing about two rv being "sufficiently correlated". The *correlation coefficient* is a meaningful measure of dependence

$$\mathrm{corr}(x, y) = \rho \triangleq E[XY] = E[g(X, Y)]\Big|_{g(x,y)=xy} \tag{1}$$

so that $0 \leq \rho^2 \leq 1$, the latter upper inequality holding if X and Y are in strict linear functional relationship. Indeed, ρ is a coefficient of linear dependence, and it does not capture more complex forms of interdependence. It remains an open question: which function of ρ should be used as a measure of *interdependence* ? ρ^2 is more directly interpretable than ρ itself. On the other hand, if $\rho = 0$ does not imply *independence*, it is difficult to interpret ρ as a measure of interdependence.

Example 1 (Naive correlations). Consider the random variables (rv) $X \sim \mathcal{N}(0, 1)$ *and* $Y = X^2$. *The covariance of the two rv* X *and* Y

$$c_{XY} = E[XY] - \bar{x}\bar{y} = \int_{-\infty}^{\infty} = \frac{x^3}{\sqrt{2\pi}} e^{-x^2/2} dx = 0. \tag{2}$$

where the integral vanishes because the integrand is odd-symetrical about $x = 0$. $E[.]$ *and* $\overline{x}$ *respectively stand for the expectation and the mean of* X. □

In the following we define $R(\cdot, \cdot)$ as an indicator function for grounding statistically the statement "U_i and U_j are correlated enough"

$$R_\epsilon(U_i, U_j) \iff \mathbf{1}_{\mathrm{corr}(U_i, U_j) > \epsilon}. \tag{3}$$

But, this general statement says nothing about *how* U_i and U_j are related. A wide variety of other measures of correlation, with respect to tests for independence, is available – *e.g. intraclass* correlation, *tetrachoric* correlation, *biserial* correlation, etc. Daniels [6] defined a class of correlation coefficients based on the expression

$$r_D = \frac{\sum_{i,j} a_{ij} b_{ij}}{\sqrt{\sum_{i,j} a_{ij}^2 \sum_{i,j} b_{ij}^2}}, \tag{4}$$

where a_{ij} and b_{ij} depend on the n-uple (x_i, x_j) and (y_i, y_j), respectively. Though correlation constitutes a fundamental tool it has important limitations: (*i*) the linearity of the functional link between rv, (*ii*) it deals only with two variables. Since our goal, expressed at the beginning of this section, is to account for the interdependences between the outputs of a great number of functions of m arguments ($m \leq n$), we need to generalize this definition to m-ary relations. We often met difficulties when a variable is correlated with a set of variables. If we find that holding another variable fixed reduces the correlation between two other variables, we infer that their interdependence arises in part – *i.e.* conditionnally – through this other variable. This function is known as *partial correlation*. Conversely, if the partial correlation is larger than the original one, we infer that the other variable was masking the correlation. Remember that we cannot assume a *causal* connection. We shall revert to Scharf [11, pp. 292] for demonstrations of the basic results.

Example 2 (Partial correlations). :*Suppose we have n observations on 3 (< n) variates*

$$x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, \ldots, x_{n1}, x_{n2}, x_{n3}$$

that are multinormally distributed (such an approach is not necessary but simplifies the development) and standardized. The conditional distribution of $\mathbf{x}_1 = (x_1, x_2)^T$ *given* x_3 *is multinormal so that*

$$corr(x_1, x_2|x_3) = \rho_{12|3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}. \tag{5}$$

The extension of (5) for the conditional distribution of (x_1, x_2, x_3) *given* $\mathbf{x}_K$*, where* $\mathbf{x}_K$ *denotes any subset of* $(x_4, \ldots, x_p)$ *gives*

$$corr(x_1, x_2|x_3, \mathbf{x}_K) = \frac{\rho_{12|K} - \rho_{13|K}\rho_{23|K}}{\sqrt{(1-\rho_{13|K}^2)(1-\rho_{23|K}^2)}}. \tag{6}$$

□

But most certainly, (6) says nothing simple about corr(X, Y) when corr(X, Y, Z) is greater than ϵ, and this hampers the rest of our approach, for reasons that will appear in the following.
As an alternative in the n-ary case, independence between variables may provide a solution since independence between two continuous rv holds when $P(X < x, Y < y) = P(X < x)P(Y < y)$, which can be generalized to n-variables.
Stating that two units are related is half the job: given an intricate set of (measured) relations that hold between units, what do we learn from the overall activity ? Knowledge Representation Theory usually builds a graph, once a relation is defined among n units, as examplified by Fig.1. Then, this graph can be characterized in many ways and the mutual interactional structure can be analyzed and explained. Furthermore graphs are limited to binary relations, and as we mentionned earlier the possibility of grounding n-ary relations in statistical inference, one can build hypergraphs out of continuous signals as will be shown in section 4.

3 Organization of binary relations

In this section we limit the discussion to the case of binary relations between different units. As mentionned before, stating relation (3) is equivalent to associating a graph to the set of measured activities of the units.
Now displaying an invariant of the interaction supposes giving a characterization of the set of relations as a whole; to do so we look for inspiration in classical tools from computer science, namely graph (spectral) theory and random matrix theory. The first step is naming the graphs we work with conveniently: we notice that the binary relation between two units is symmetric since $R_\epsilon(X, Y) = R_\epsilon(Y, X)$. Consequently vertices that represent the units in the graph can be either disconnected, or connected in an unoriented way. Though many extensions can be imagined (see 6), we will focus only on undirected graph.
Among many standard tools available from random graph theory (see [**?**]), we choose to depict

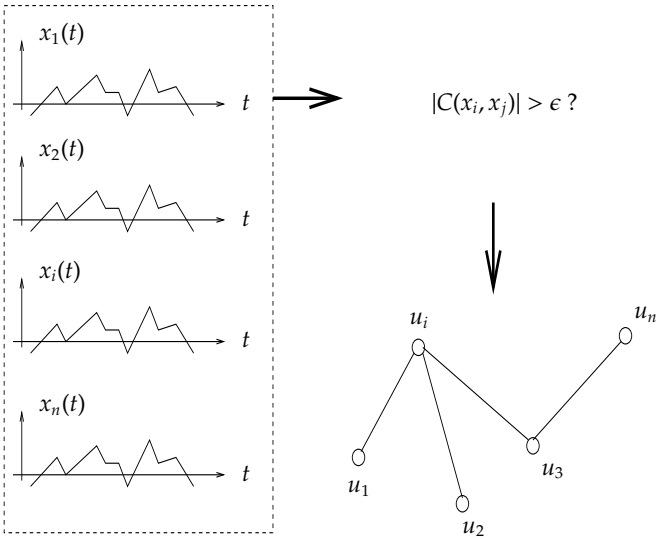


Fig. 1. From stationary time dependent signals to graph via correlation.

graphs in term of degree, that quantifies the typical number of connections of a given vertex. This quantity can be seen in a probabilistic context: for instance the estimated probability density function (pdf) of the degree.
The process of characterization of a graph using selected elements of random graph and random matrix theory is depicted by Fig.2; the following section now applies this scheme to real signals.

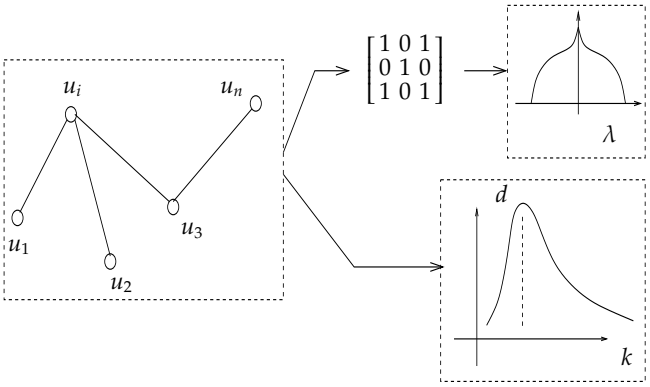


Fig. 2. Two characteristics computed from the graph: (*up*) the spectrum of the adjacency matrix of the graph, (*down*) the probability density function of the degree $k \to P(K = k)$.

In that perspective, we aim at comparing graphs associated to the activity of different sets of units. Rather than specifying complex interdependence patterns between time-dependent activities of several units, we found it convenient to consider a stochastic process defined as a family $(X_t)_{t\in I}$ of rv indexed by t taken from a continuous interval I, from which we extract

a finite set of rv $\{X_{t_1}, \ldots, X_{t_i}, \ldots X_{t_n}\}$, as shown by Fig.3. Thus, to each type of stochastic process is associated a set of vertices, whose relations can be computed from a finite number of realizations of the stochastic process over a finite time interval. For the sake of diversiy, we generate graphs from different sorts of stochastic processes (random walk, long-range dependent process) as well as deterministic time series generated by a Lorenz system in chaotic regime.

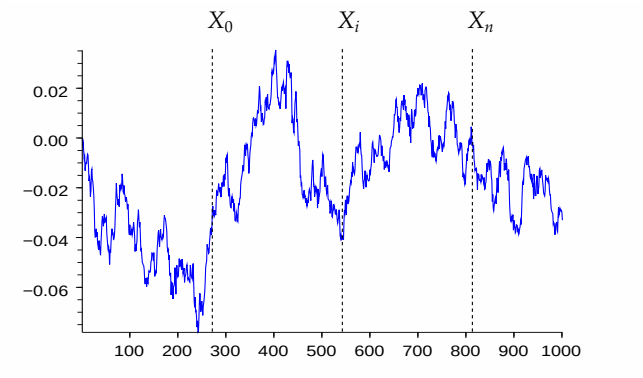
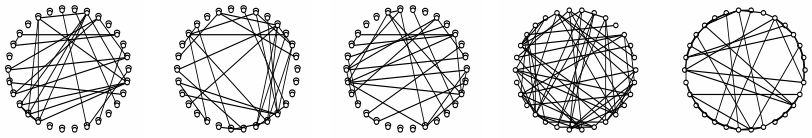


Fig. 3. Realizations of rv taken from one realization of a brownian stochastic process.

Questions are the following: is there a type of process corresponding to a given type of random graph such as Erdös-Rényi random graphs, small-world or scale-free graphs [**?**, **?**, **?**]. At first sight, graphs built from signals using R_ϵ look quite similar, as evidenced by Fig. 4.



(a) Brownian motion (b) Long-range dependent (c) Lorenz (d) Erdös-Rényi graph (e) Small World graph

Fig. 4. Graph topologies. *(a,b,c)* are built from signals using R_ϵ, for values of ϵ ensuring that the number of edges ($e = 60$) and the number of nodes ($n = 30$) are uniform. On the contrary *(d,e)* are classical random graphs, with a probability of rewiring of $1/4$ in the Small World case.

Hence we turn to a more quantitative comparison, as evoked earlier, with the *degree distribution*. Two possibilities were explored: the first one[3] is to choose a signal and to generate a set of graphs for that type of signal, then to compute the degree distribution for each graph realization, and lastly to build a global histogram for each signal in order to approximate the underlying distribution. For Erdös-Rényi and Small World graphs, the theoretical distribution

[3] The second method was to compare pairwise each realization of the degree distribution, with the help of a statistical test (e.g. Kolmogorov-Smirnov or Wilcoxon), and to count the number of positive tests, for all possible combinations (e.g. Lorenz generated graph compared to Small-World graph), but this method didn't prove usefull

is known (see [**?**], [**?**]), and following the law of large numbers, the histogram converges in probability to the theoretical distributions. Accordingly, Fig.5 displays empirical degree distributions generated from five types of graphs, three of which were built thanks to R_ϵ, the last two being Erdös-Rényi random and Small World graphs, as in Fig.4.

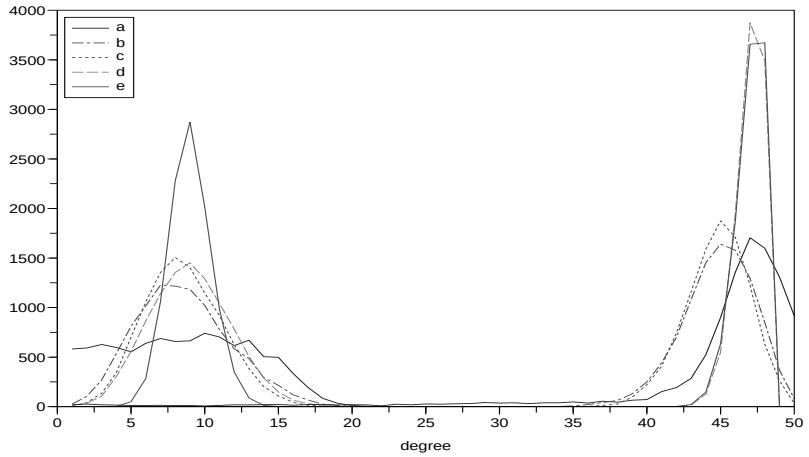


Fig. 5. Degree histograms (a) Brownian motion, (b) Lorenz, (c) Long-range dependent noise (d) Erdös-Rényi (e) Small World. In all cases, graphs are composed of 50 nodes, however the number of edges varies: on the left part, 250 edges are present, against 1200 on the right side.

Before commenting on these results, we must remark that in random graph theory if one needs to compare two graphs from their properties, it may be necessary to ensure that their number of nodes (and edges) is of the same order of magnitude[4].
Now, we noticed that depending on the threshold ϵ used to make a graph out of a process, the number of edges depends of the type of signal (this fact is easily explained thanks to the difference of autocorrelation functions). The degree distribution are made comparable so that disparity doesn't fit with the constraint just stated, that edges and number of nodes should be approximately equal. Consequently, R_ϵ-induced graphs can be compared from random graph theory, we must choose different values of ϵ depending on the type of signal before proceeding so that the number of edges is kept constant.
This limit being clearly exposed, we can now compare the degree distribution of different graphs. Fig. 5 shows two sets of curves, obtained for two distinct edge numbers, and we shall focus on the left part first. Three clusters can grossly be isolated: line (a), line (e), and lines (b, c, d). On the right part of the figure, we remark that two clusters can now be identified: curves (a, b, c) on one hand, and (d, e) on the other. This counterexample shows first that classical graphs such as Erdös-Rényi random graph or small-world graphs hardly approximate the

[4] "Consequently in random-graph theory the occupation probability is defined as a function of the system size: p represents the fraction of the edges that are present from the possible $N(N^2-1)/2$. Larger graphs with the same p will contain more edges, and consequently properties like the appearance of cycles could occur for smaller p in large graphs than in smaller ones. This means that for many properties Q in random graphs there is no unique, N-independent threshold, but we have to define a threshold function that depends on the system size", [**?**] p.55

properties of R_ϵ-induced graphs, and second that depending on the threshold used to build graphs, their properties -the degree distribution in that case- are not constant.

4 Multilevel organization

Comparing graph with the same threshold is inappropriate because both the number of edges and nodes matter. Considerable differences will appear for a same graph at distinct levels of normalization if these conditions were standardized. A natural idea developped in that section, is therefore to take into account the history of the graph when the normalization level moves. We admit that for Erdös-Rényi random graphs the degree distribution obeys a binomial[5] law $\mathcal{B}(n,p)$, that can be approximated by a normal law. Now, a normal law is completely described by its mean and standard deviation (μ,σ). Though approximating the distributions of various graphs encountered so far by a normal law would diserve more careful justification, we admit this hypothesis just to illustrate the idea of characterizing the graph simultaneously at different normalization levels.
It is easy to distinguish graphs by simply focusing on the degree distribution. Figure 6 illustrate this by plotting the parametrized curves $\mathcal{C} : e \to (\mu(e), \sigma^2(e))$ when the number of nodes N is kept constant, but the number of edges e grows linearly: curves corresponding to different graphs are easily distinguished, even if they rely on a simple characteristic such as the degree distribution.

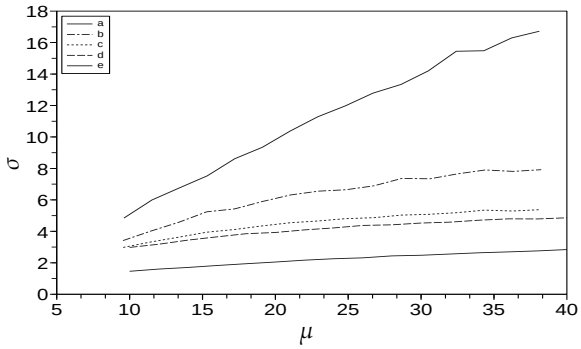


Fig. 6. Mean and standard deviation parametrized by the number of edges $e \to (\mu(e), \sigma(e))$ of the degree distributions. e varies from 500 to 2000, while the number of nodes N remains equal to 100. (a) Brownian motion, (b) Lorenz, (c) Long-range dependent noise (d) Erdös-Rényi (e) Small World.

This confirm the dependence to thresholding evidenced by section 3, and then prove the existence of an alternative position, based on embracing in a single representation several levels of detail. It appears clearly that not only the number of edges should be taken into account to build this multilevel representation, but the number of nodes as well, which raise the issue of hierarchical agglomeration of variables.
Here we do not deal with several phenomenologically distinct levels of description, and the objects we're concerned with remain the same even when the normalization conditions evolve. One can take advantage of the combinatorial nature of relations defined on a finite set of vertices to elaborate hierarchical multilevel approaches [**?**, 5].

[5] In the case of a Erdös-Rényi random graph, $n = N - 1$ where N is the number of nodes, while p is the probability for two nodes to be connected, cf [**?**] p.56

5 *n*-ary relations

In this section, we extends the framework presented so far to n-ary relations. The underlying idea is to identify a structure of relationships and a topological space. But this cannot be achieved directly: correlation is inappropriate to ground a n-ary relationship. Then, to meet computational requirements we need to take advantage of algebraic topology that allows algorithmic processing. To take into account the multilevel stand put forward in previous section, we introduce *filtrations*. Lastly we give experimental results.

5.1 Causality in *n*-ary relations

Let us enumerate some possibilities offered by statistics to express n-ary relation. (**??**) is limited by the arity of the correlation function: the predominance of second-order moments is a consequence of the prevalence of the Gaussian distribution in models if not in nature. Indeed:

A. the Gaussian distribution is *completely* described by its first two moments.
B. instead of describing an unknown distribution, it may seem more natural to first compare it to the normal law and to provide some distance from it.

One possibility would be to define a n-ary relation based on binary relations, *e.g.* $(X_1, \dots, X_n)$ are related if each couple (X_i, X_j) is related[6], but in many cases pairwise relations say nothing about relations between, say 3 variables.

Example 3 (Pairewise independent variables). Let X_1 and X_2 two independent rv with values in $\{0,1\}$, with a probability $\frac{1}{2}$, and the rv $X_3 = X_1X_2 + (1-X_1)(1-X_3)$. X_3 has also values in $\{0,1\}$ with a probability $\frac{1}{2}$ because $P(X_3 = 0) = P(X_1 = 0)P(X_2 = 1) + P(X_1 = 1)P(X_2 = 0) = \frac{1}{2}$ and $P(X_3 = 1) = 1 - P(X_3 = 0) = \frac{1}{2}$. X_1, X_2, X_3 are pairewise independent since:

$$P(X_1 = 0, X_3 = 0) = P(X_1 = 0)P(X_2 = 1) = \tfrac{1}{4} = P(X_1 = 0)P(X_3 = 0)$$
$$P(X_1 = 0, X_3 = 1) = P(X_1 = 0)P(X_2 = 0) = \tfrac{1}{4} = P(X_1 = 0)P(X_3 = 1)$$
$$P(X_1 = 1, X_3 = 0) = P(X_1 = 1)P(X_2 = 0) = \tfrac{1}{4} = P(X_1 = 1)P(X_3 = 0)$$
$$P(X_1 = 0, X_3 = 1) = P(X_1 = 1)P(X_2 = 1) = \tfrac{1}{4} = P(X_1 = 1)P(X_3 = 1)$$

Analog equalities can be found for X_2, X_3. However we have the relation

$$P(X_1 = 0, X_2 = 0, X_3 = 0) = 0 \neq P(X_1 = 0)P(X_2 = 0)P(X_3 = 0)$$

Hence X_1, X_2, X_3 are not independent. □

Example 4 (Case of three pairewise independent rv). Consider a random vector $X = (X_1, X_2, X_3)$ uniformly distributed onto the tetraedron whose vertices are the points $\{(0,0,0), (0,1,1), (1,0,1), (1,1,0)\}$ with the pdf

$$f(x) = \frac{1}{2}\mathbf{1}_{[0,1]}(x_1)\mathbf{1}_{[0,1]}(x_2)\mathbf{1}_{[0,1]}(x_3)[\delta(x_1+x_2-x_3) + \delta(x_2+x_3-x_1) + \delta(x_1+x_2+x_3-2)], \quad (7)$$

where $\mathbf{1}_{[0,1]}$ is the indicator function of the interval $[0,1]$. Any coordonnate of this vector is uniformly distributed in the interval $[0,1]$ and its projection onto the plan $x_1 + x_2 + x_3 = 0$ is uniformly distributed inside $[0,1]^2$. Hence, variables X_1, X_2 et X_3 are pairewise independent. However they are dependent because otehrwise the distribution of X would be uniform inside the cube $[0,1]^3$. □

[6] this is similar to the use of Rips complex instead of Čech complex in computational topology, see [**?**] for definitions

A fundamental result of Information Theory is that a Gaussian variable has the largest entropy among all random variables of equal variance [9], in other word the Gaussian distribution is the "most random" or the least structured of all distributions. This means that entropy could be used as a measure of *nongaussiannity*.
A statistical relationship, however strong and suggestive, can *never* establish a *causal* connection: ideas on causation must come from outside statistics. For instance, we may be interested in whether there is a relationship between an alarm and an earthquake: put this way it is a problem of interdependence. But if we are interested in detecting the alarm to convey information about the earthquake, we are considering the dependence of the latter upon the former. This is clearly an asymmetrical relation: earthquake 'causes' alarm to activate, but we are certain that alarm do not affect the earthquake, so we measure the dependence of alarm upon earthquake. Even if they were in perfect functional correspondence, we cannot reverse the "obvious" *causal connection*.
At this stage, we ought to define what we mean by *cause*. We shall content of the following definition: *x is a cause of y if and only if the value of y can be changed by manipulating only x*. The issue of causality cannot be overlooked and the result of a statistical investigation is in support of a *causal* relationship. In regresson analysis, it is reasonnable to admit that changes in the *dependent* (or *response*) variables are caused by the changes in the inputs. The notion of *conditional* independence has an important role to play in disentangling relationships between variables. Rubin [10] provides a framework for causal inference. Granger [8] describes a form of causality based on time ordering of the variables.

5.2 Organization of measured relationships

In this section, we devise tools capable to organize the interdependence relationships between a number of variables or in the dependence of one or more variables upon others. Suppose that we have agreed to select an type of operator $R_\epsilon(\dots)$ that accepts k arguments $(2 < k \leq n)$ – the vertices $\{v_1, \dots, v_n\}$ –. We look for order and regularity in subset of vertices that are related according to R_ϵ, for instance the following sets

$$\begin{gathered} R_\epsilon(v_1, v_2) \\ R_\epsilon(v_2, v_4, v_5) \\ R_\epsilon(v_2, v_i, v_{n-1}, v_n) \\ \vdots \end{gathered}$$

In the same way that a set of pairs of vertices $\{(v_i, v_j) | i \in I, j \in J\}$ defines a graph, the previous set of subsets defines a hypergraph, that extends graphs to larger dimensions. By analogy with ideas put forward in Knowledge Representation where topologial properties of relation-induced graphs are studied – *i.e.* number of connected components, graph connectivity[7] – we consider that set of subsets as a topological space and borrow relevant tools to examine it.
By "examining" we mean the search of an *invariant* that maps the same element to spaces that share the same topology. Invariants are often used via contrapositives: when two topological spaces have different invariants, their types differ. Nevertheless if the invariant is the same, it might have an insufficient discriminating power, and it is not guaranteed that the two spaces really are of the same topological type.

Simplicial homology

Simplicial homology theory provides us with such invariants, as will be examplified by section 5.3. First, let us set some landmarks about simplicial homology and related fields [**?**]. The main idea here is to compare different spaces, to decide whether or not they are equivalent from the point of view of topology, and finally to constitute *equivalence classes*. Of course the acceptions of "equivalence" are manifold:

[7] the problem of graph connectivity is determining the smallest subset of vertices (or edges) whose deletion would disconnect the graph.

a. *homeomorphy*: let $\mathbb{X}$ and $\mathbb{Y}$ be two topological spaces. If there exist a continuous and bijective map $f : \mathbb{X} \to \mathbb{Y}$ such that f^{-1} is continuous then $\mathbb{X}$ and $\mathbb{Y}$ are said to be homeomorphic, and have the same topological type.
b. *homotopy*: the formal statement being counter intuitive, we settle for the following: $\mathbb{X}$ and $\mathbb{Y}$ are homotopy-equivalent if they can be transformed into one another by bending, shrinking and expansion.
c. *homology*: instead of working directly on spaces thanks to a map defined between them, homology introduces intermediate algebraic structures that correspond to the topological spaces (e.g. group structures in that case), so that from those algebraic structures, invariants can be built and compared as mentioned earlier.
d. *simplicial homology*: this form of homology is defined in a combinatorial setting (i.e. when the set of points that form the space is countable), more precisely for a particular type of topological space -namely simplicial complexes- that add constraints to the hypergraph structure.

Zomorodian [**?**] compares these different notions borrowed from topology and algebraic topology, on the basis of their computational tractability and focusses on simplicial homology.

Simplicial complexes

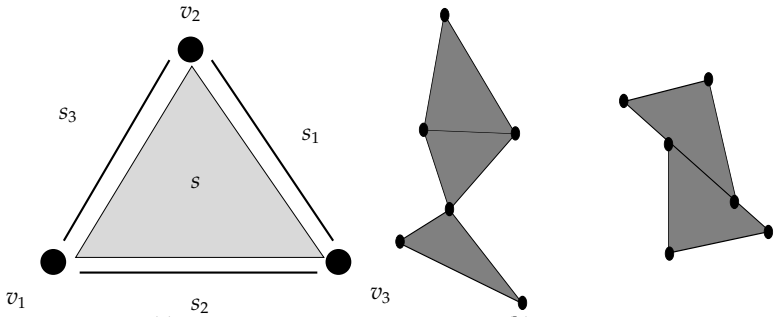
Now the price to pay for casting topological features of spaces in computational terms is to restrict the scope of possible topological spaces to simplicial complexes, that may be defined grossly as a countable set of vertices $V = \{v_i\}_{i \in I}$, and a set of simplices that intersect along their faces. There are important additional requirements, but instead of giving an axiomatic presentation (see [**?**]), we state definitions in a more intuitive way:

- every simplex is constituted of faces. For example, the 2-simplex $S = \{v_1, v_2, v_3\}$ has the following simplices the following faces, as illustrated by Fig. 7(a) in the case of a triangle:

$$\{\{v_1\}, \{v_2\}, \{v_3\}, \{v_1, v_2\}, \{v_2, v_3\}, \{v_1, v_3\}, \{v_1, v_2, v_3\}\}$$

- C_1: if a simplex s belongs to a simplicial complex K then all its faces belong to K.
- C_2: intersections in a simplicial complex must occur along shared faces, as shown by Fig. 7(b).

So far we've first justified the choice of simplicial homology, before stating the condition to be met by the topological space under study. In section 5.3 we give a characterization method derived from the homological framework just depicted, then in section 5.4 we draw the consequences, in statistical terms, of constraint C_1 discussed above.



(a) Faces (b) Intersections

Fig. 7. (a) 2-simplex $s = \{v_1, v_2, v_3\}$, and faces $\{\{v_1\}, \{v_2\}, \{v_3\}, \{v_1, v_2\}, \{v_2, v_3\}, \{v_1, v_3\}, \{v_1, v_2, v_3\}\}$. (b) simplicial complex with (left) allowed (right) forbidden intersections.

Filtrations

A filtration is a growing sequence of simplicial subcomplexes of a complex K, as shown by Fig. 8. One way to describe it is to imagine a map from a continuous scalar space such as $[0, 1]$ to K: for each parameter value t we get a subcomplex of K, and as t increases contiuously from 0 to 1 we first obtain an empty subcomplex to finally get the full complex K.

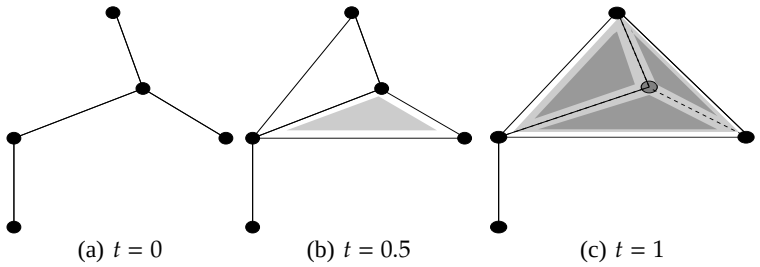


Fig. 8. Filtration at different parameter levels.

This structure is made necessary to take into account the multilevel structure that supposes to organizes the relations computed simultaneously at several threshold levels.
We review in section 5.3 some computational characteristics of simplicial complexes taken from the field of computational topology that first allow to compute invariants for simplicial complexes, then for filtrations.

5.3 Characterization of simplicial complexes

Section 5.2 precises the way to identify a set of n-ary relations with a topological space. Here, we aim at deriving computable characteristics of those spaces.
In experiments not reported in this article we first intended to characterize simplicial complexes in a quite naive way, computing the relative proportion of k-simplexes in the complex, for several threshold values; however this approach displayed little discriminative power. The second idea which turned out not to be pertinent was generalizing the idea of degree distribution for each simplex order: what is the probability for a vertex to be simultaneously part of exactly k_0 0-simplex, k_1 1-simplex, etc but that would involved an n-dimensional probability distributions (depending on the maximum order allowed by the n-ary relation); parametrized by the threshold level, so we take advantage of the framework, where the set of relations is assimilated to a particular type of topological space before being characterized using simplicial homology theory.
In section 5.2, the framework relies on associating a group structure to each simplicial complex. Ggiving the details of that structures is far beyond the scope of this article. Suppose we deal with a simplicial complex in dimension 3; a way to characterize it is to count the number of voids enclosed inside the complex, and the number of tunnels that go through the space. Fig. 9 illustrates this with two examples: an empty sphere and a torus. Intuitively, finding that these spaces are of different topological types seems obvious since one cannot be deformed continuously one into the other; the homological way to state this is to note that the sphere encloses a void space, as does the torus, however there is a "tunnel" going through the torus, not through the sphere.
In the topological litterature, the Betti numbers of order k β_k encode those invariant properties of spaces:

- β_0 can be interpreted[8] as the number of *connected components* in the simplicial complex.

[8] in dimension 3 for torsion-free spaces

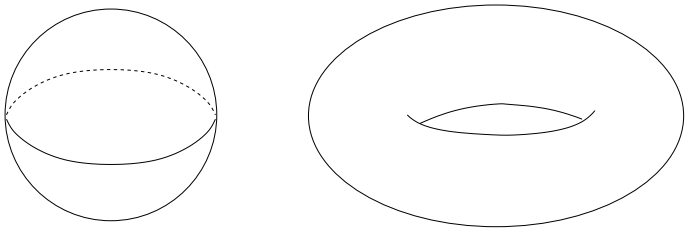


Fig. 9. Empty sphere and torus.

- β_1 is the *number of tunnels* enclosed by the space.
- β_2 is the *number of voids* enclosed by the space.

Now recall from section 4 that we'we adopted a multilevel stand, to cope with the parametrization of relations. Thus instead of organizing a set of relations at a given threshold level, we take into account simultaneously several levels and build a filtration, as mentioned in section 5.2. The last step is thus to adapt the characterization of a simplicial complex in the case of a filtration. This was achieved by Edelsbrunner *et al.* in [**?**], and lead to the notion of *persistent homology*, that captures long living Betti number when the continuous value that parametrizes the simplicial complex in the filtration is varied.
As this will be examplified in section 5.5, we now turn to ensuring compatibility between constraint C_1 imposed by the structure of simplicial complexes in 5.2, and the statistical grounding of relations as in 2.

5.4 From correlatedness to independence

As already suggested, independence is a much stronger property than uncorrelatedness. This can be stated by saying independence implies *nonlinear uncorrelatedness*. If x_1 and x_2 are independent rv, then any nonlinear transformations $g(x_1)$ and $h(x_2)$ are uncorrelated. Mathematically, statistical independence is defined in terms of probability densities [9]. For simplicity, X is independent of Y if knowing the value of Y does not give any information on the values of X. In words, the joint density $p_{X,Y}(X,Y)$ must factorize into the product of their marginal densities $p_X(X)$ and $p_Y(Y)$. Uncorrelated Gaussian rv are also independent, a property which is not shared by other distributions in general.
Mutual information is a measure of the information that a set of rv have on the other rv in the set. Using entropy, we can define the mutual information I between n rv $X_1, \dots X_n$, as follows

$$I(x_1, \dots, x_n) = \sum_{i=1}^{n} H(x_i) - H(\mathbf{x}), \tag{8}$$

where $\mathbf{x}$ is the vector containing all the x_i. Mutual information can be interpreted by using the interpretation of entropy as code length. The terms $H(x_i)$ give the lengths of code for the x_i when these are coded separatedly, and $H(\mathbf{x})$ gives the code length when $\mathbf{x}$ is coded as a random vector, *i.e.* all the components are coded in the same code. Mutual information thus shows what code length reduction is obtained by coding the whole vector instead of the separate components. In general, better codes can be obtained by coding the whole vector. However if the x_i are independent, they give no information on each other.
Alternatively, mutual information can be interpreted as a distance between two probability densities, because, as the Kullback-Leibler divergence, it is always non negative and zero iif the two distributions are equal. Thus one might measure the independence of the X_i as the mutual information between the real density $p_{\mathbf{X}}(\mathbf{x})$ and the factorized density $p_{X_1}(x_1) \dots p_{X_n}(x_n)$.
Moreover, any extracted subsequence of variables forms an independent set of variables.
If the space of multidimensional densities comes equipped with a metric structure then from (3), the n-ary relation based on 'approximate' independence can be stated as follows:

$$R_\epsilon(X_1, \dots, X_n) \iff I(x_1, \dots, x_n) < \epsilon. \tag{9}$$

since (9) respects the following conditions:

(*i*) the variable arity: $R_\epsilon(X_1, \ldots, X_k)$ is well-defined for $k \leq n$.
(*ii*) for all subsequence $(X_i, \ldots, X_j) \subseteq (X_1, \ldots, X_k)$, then $R_\epsilon(X_1, \ldots, X_k) \Rightarrow R_\epsilon(X_i, \ldots, X_j)$.
(*iii*) computational tractability.

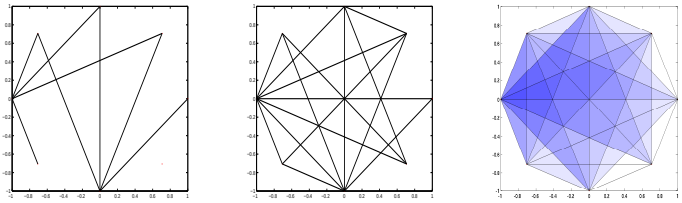
5.5 Experimental results

The mutual information is a function of densities. This makes the problem much more complicated because the estimation of densities is, in general, a nonparametric problem. Nonparametric means that it cannot be reduced to the estimation of a finite parameter set. Nonparametric estimation of densities is known to be a difficult problem. One way to solve the problem of density estimation is to approximate the densities of the components by a family of densities that are specified by a finite number of parameters[9]. For instance, we consider the following log-densities:

$$\log p^+(x) = \alpha_1 - 2\log\cosh(x) \tag{10}$$

$$\log p^-(x) = \alpha_2 - [\frac{x^2}{2} - \log\cosh(x)] \tag{11}$$

where α_1, α_2 are positive parameters that are fixed to make p^- and p^+ probability densities. p^- is *subgaussian* whereas p^+ is *supergaussian*.
Densities could be estimated using basic density estimation methods such as kernel estimators: such a simple approach would be very error prone, however, because the estimator would depend on the correct choice of the kernel parameters, greedy of samples, computationally rather complicated for a large number of dimensions [**?**].
The validity of the approach can be found in the detail of the experiments discussed below, which were carried on using this method[10]. At this stage, it is therefore licit to add to the simplex whose vertices correspond to the rv $X_1, \ldots, X_i$ the simplicial complex $\mathcal{C}$. It would then be a simple matter of iteration to achieve the incremental building of the simplicial complex.



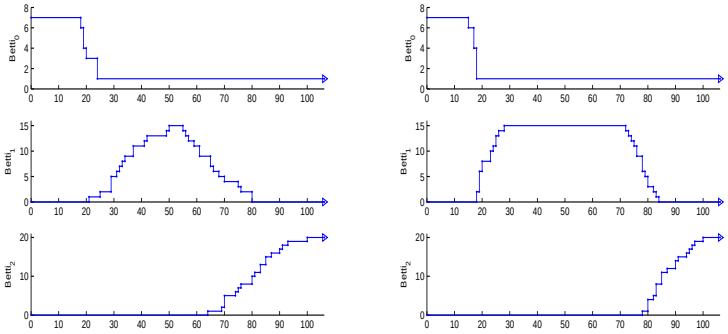
(a) $\epsilon = 0.3$ (b) $\epsilon = 0.5$ (c) $\epsilon = 0.9$

Fig. 10. Simplicial complexes extracted from the filtration for different threshold values, limited to 2-simplices.

Instead of iterating the building process depicted above for each distance threshold, we store directly the distances for each possible combination of $(X_1, \ldots, X_i) \subseteq (X_1, \ldots, X_n)$. These distances play the role of birth dates necessary to specify the filtration structure. Consequently, we get the simplicial complex corresponding to a given threshold value by just extracting it from the filtration, as illustrated by Figure 10 for a series of arbitrary threshold values, at a

[9] Classical approximations by cumulants, –*e.g.* Edgeworth expansion when Gaussian distribution is assumed – is computationally very difficult.
[10] Source code made available by [**?**].

limited order.
Finally we compute the persistent homology of a filtration of simplicial complexes for complexes corresponding to different signals. Firstly we consider just one complex per type of signal, built from n-realizations of a process as shown by Fig.11, composed of three plots, each corresponding to the Betti number of k-simplices where k varies from 0 to 2. From one signal to the other, the general shape of the curves are similar – at the same order k –, even though the range of abscissa differ: at level 0, two steady states are separated by a quick decrease. Conversely at level 2, two steady states are separated by a quick increase. In between, at level 1 for both signal types, we observe an increase followed by a decrease when the distance linearly increases. Are theses signatures enough to distinguish two types of signals ? Even if in the selected particular cases the persistent Betti numbers of order 0 and 2 are quite similar from one signal to the other , Betti numbers of order 1 seem to exhibit a specific shape. However the exposed result hold only for one realization of the complex, and say little about the statistical properties of persistent Betti numbers.



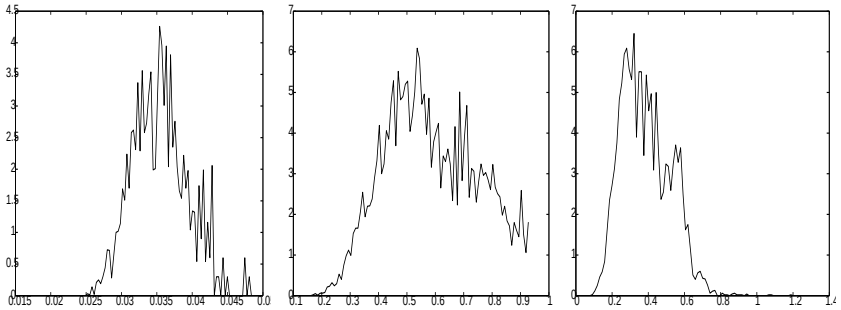
(a) Brownian motion (b) Long-Range Dependent

Fig. 11. Persistent homology upo to order 2.

To assess the statistical significance of these results, we compute the empirical mean of persistent Betti numbers at order $k = 1$, which seems the most discriminant from previous results. Fig.12 compares the averages over $N = 50$ realizations of filtrations, for three signal types. Clearly shape differences are blurred by averaging, however the positions of the curves on the absissa permits alone an easy discrimination.

6 Conclusions

The aim pursued in this article was to take into account weighted relations to characterize a graph in a way that redefine classical invariants. We have defined a comparison method that relies on statistical independence for n-ary relations. In the first case experimental assessment involves stochastic processes of different types where distinct instants on time are selected to mimick dependent units. We show that the network characterization depends strongly on the parametrization. The n-ary case where the suggested representation, inspired from computational topology is encoded as a filtration. Statistical independence can ground n-ary relations, and provide an associated characterization in the case of n-ary relations thanks to simplicial homology. Lastly we assess experimentally the relevance of this framework.
One technical improvement resides in the type of statistical relation that roots the definition of edges. Depending for instance on *time dependence* of measured activity, one could consider looking for various models such as time-regressive methods, be they linear or not. Another



(a) Long Range Dependent (b) Brownian (c) Lorenz

Fig. 12. Averaged persistent homology at order $k = 1$ over $N = 50$ realizations. Abscissa scales where kept distinct to allow shape comparison.

work direction could be the use of causal inference methods that first associates a discrete alphabet to continuous values before identifying causal states that allow to identify the phenomenon with a Markov chain [**?**].
One topic of interest in complex systems studies is to represent phenomena that occur at distinct scales, each level interfering with the others. In cell biology, intracellular biological networks of distinct types are often studied separately. Yet some studies tend to link several levels through hierarchical network building (see [**?**] for biological networks, [**?**] in a more abstract setting). We plan next to examine the effect of uncertainty at a microscopic level on higher scales when hierarchical structures are assembled. To do so we will first compare several methods to take into account the weights of edges when changing levels, then we will borrow tools from computational physics to evaluate uncertainty propagation and compare hierarchy building strategies when weights are ignored or taken into account. Lastly we propose to confront such approaches and methods from multiscale Physics such as mean-field or renormalization theory.

References

1. Albert, Réka, Barabási, and Albert-László. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, Jan 2002.
2. Almaas, Kovacs, Vicsek, Oltvai, and Barabási. Global organization of metabolic fluxes in the bacterium escherichia coli. *Nature*, 427:839–843, Feb. 2004.
3. A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *PNAS*, 101(11):3747–3752, 2004.
4. J.P. Cardoso. Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-89*, 1989.
5. J.L. Casti. *Reality Rules II, Picturing the world in mathematics, the frontier*. Wiley Interscience, 1997.
6. H.E. Daniels. The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33(129), 1944.
7. S. Dodel, J.M. Herrmann, and T. Geisel. Functional connectivity by cross-correlation clustering. *Neurocomputing*, 44-46, 2002.
8. C.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(424), 1969.
9. A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley series on Adaptive and Learning Systems. John Wiley & Sons, 2001.
10. D.B. Rubin. Formal models of statistical inference for causal effects. *J. Statist. Planning Inf.*, 25(279), 1990.
11. L.L. Scharf. *Statistical signal processing. Detection, estimation and times series analysis.* Addison Wesley, july 1991.