Kernel-Granger causality and the analysis of dynamical networks

D. Marinazzo,[*] M. Pellicoro,[†] and S. Stramaglia[‡]

Dipartimento Interateneo di Fisica, Università di Bari, I-70126 Bari, Italy;
TIRES-Center of Innovative Technologies for Signal Detection and Processing, Università di Bari, Italy;
and I.N.F.N., Sezione di Bari, I-70126 Bari, Italy
(Received 20 March 2008; published 27 May 2008)

We propose a method of analysis of dynamical networks based on a recent measure of Granger causality between time series, based on kernel methods. The generalization of kernel-Granger causality to the multivariate case, here presented, shares the following features with the bivariate measures: (i) the nonlinearity of the regression model can be controlled by choosing the kernel function and (ii) the problem of false causalities, arising as the complexity of the model increases, is addressed by a selection strategy of the eigenvectors of a reduced Gram matrix whose range represents the additional features due to the second time series. Moreover, there is no *a priori* assumption that the network must be a directed acyclic graph. We apply the proposed approach to a network of chaotic maps and to a simulated genetic regulatory network: it is shown that the underlying topology of the network can be reconstructed from time series of node's dynamics, provided that a sufficient number of samples is available. Considering a linear dynamical network, built by preferential attachment scheme, we show that for limited data use of the bivariate Granger causality is a better choice than methods using $L1$ minimization. Finally we consider real expression data from HeLa cells, 94 genes and 48 time points. The analysis of static correlations between genes reveals two modules corresponding to well-known transcription factors; Granger analysis puts in evidence 19 causal relationships, all involving genes related to tumor development.

I. INTRODUCTION

Dynamical networks [1] model physical and biological behavior in many applications; examples range from networks of neurons [2], Josephson junctions arrays [3] to genetic networks [4], protein interaction nets [5], and metabolic networks [6]. Synchronization in dynamical networks is influenced by the topology of the network [7]. A great need exists for the development of effective methods of inferring network structure from time series data; a method for detecting the topology of dynamical networks, based on chaotic synchronization, has been proposed in [8]; a recent approach deals with the case of a low number of samples and proposed methods rooted on $L1$ minimization [9].

Granger causality has become the method of choice to determine whether and how two time series exert causal influences on each other [10]. This approach is based on prediction: if the prediction error of the first time series is reduced by including measurements from the second one in the linear regression model, then the second time series is said to have a causal influence on the first one. This frame has been used in many fields of science, including neural systems [11], reo-chaos [12], and cardiovascular variability [13]. The estimation of linear Granger causality from Fourier and wavelet transforms of time series data has been recently addressed [14].

Kernel algorithms work by embedding data into a Hilbert space, and searching for linear relations in that space [15]. The embedding is performed implicitly, by specifying the inner product between pairs of points [16]. We have recently exploited the properties of kernels to provide nonlinear measures of bivariate Granger causality [17]. We reformulated linear Granger causality and introduced a statistical procedure to handle overfitting [18] in the linear case. Our formulation was then generalized to the nonlinear case by means of the kernel trick [16], thus obtaining a method with the following two main features: (i) the nonlinearity of the regression model can be controlled by choosing the kernel function; (ii) the problem of false causalities, which arises as the complexity of the model increases, is addressed by a selection strategy of the eigenvectors of a reduced Gram matrix whose range represents the additional features due to the second time series.

In this paper we describe in detail the kernel-Granger approach and address use of Granger causality to estimate, from multivariate time series data, the topology and the drive-response relationships of a dynamical network. To this aim, we generalize our method in [17] to the case of multivariate data.

The paper is organized as follows. In the next section we describe the kernel-Granger causality method in the bivariate case, adding some details to the presentation in [17]. In Sec. III we generalize the method to process multivariate time series data, and show that the proposed method can discern whether an interaction is direct or mediated by a third time series. The case of dynamical networks is described in Sec. IV: we analyze a system of interacting chaotic maps and a model of gene regulatory networks. In Sec. V we consider

---
[*]daniele.marinazzo@ba.infn.it
[†]mario.pellicoro@ba.infn.it
[‡]sebastiano.stramaglia@ba.infn.it

systems with sparse connectivity and limited data, and compare the bivariate Granger approach with a multivariate method based on $L1$ minimization; then we analyze a real data set, gene expressions of HeLa cells. Section VI summarizes our conclusions.

II. BIVARIATE GRANGER CAUSALITY

In this section we review the kernel method for Granger causality proposed in [17]. Let us start with the linear case.

A. Linear Granger causality

Suppose to model the temporal dynamics of a stationary time series $\{\xi_n\}_{n=1,..,N+m}$ by an autoregressive model of order m,

$$\xi_n = \sum_{j=1}^{m} A_j \xi_{n-j} + E_n,$$

and by a bivariate autoregressive model which takes into account also a simultaneously recorded time series $\{\eta_n\}_{n=1,..,N+m}$,

$$\xi_n = \sum_{j=1}^{m} A_j' \xi_{n-j} + \sum_{j=1}^{m} B_j \eta_{n-j} + E_n'.$$

The coefficients of the models are calculated by standard least squares optimization; m is usually chosen according to Akaike criterion [19] or embedding techniques from the theory of nonlinear dynamical systems [20].

The concept of Granger causality is [10] η Granger causes ξ if the variance of residuals E' is significantly smaller than the variance of residuals E, as it happens when coefficients B_j are jointly significantly different from zero. This can be tested by performing an F-test or Levene's test for the equality of variances [21]. An index measuring the strength of the causal interaction is then defined as

$$\delta = 1 - \frac{\langle E'^2 \rangle}{\langle E^2 \rangle}, \tag{1}$$

where $\langle \cdot \rangle$ means averaging over n (note that $\langle E \rangle = \langle E' \rangle = 0$). Exchanging the roles of the two time series, one may equally test causality in the opposite direction, i.e., to check whether ξ Granger causes η.

We use the following shorthand notations:

$$X_i = (\xi_i, \ldots, \xi_{i+m-1})^\top,$$

$$Y_i = (\eta_i, \ldots, \eta_{i+m-1})^\top,$$

and $x_i = \xi_{i+m}$, for $i = 1, \ldots, N$. We treat these quantities as N realizations of the stochastic variables X, Y, and x. Let us denote $\mathbf{X}$ the $m \times N$ matrix having vectors X_i as columns, and $\mathbf{Z}$ the $2m \times N$ matrix having vectors $Z_i = (X_i^\top, Y_i^\top)^\top$ as columns. The values of x are organized in a vector $\mathbf{x} = (x_1, \ldots, x_N)^\top$. In full generality we assume that each component of X and Y has zero mean, and that vector $\mathbf{x}$ has zero mean and is normalized, i.e., $\mathbf{x}^\top \mathbf{x} = 1$.

Now, for each $i = 1, \ldots, N$, we define

$$\tilde{x}_i = \sum_{j=1}^{m} A_j \xi_{i+m-j},$$

$$\tilde{x}_i' = \sum_{j=1}^{m} A_j' \xi_{i+m-j} + \sum_{j=1}^{m} B_j \eta_{i+m-j}.$$

The vectors $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_N)^\top$ and $\tilde{\mathbf{x}}' = (\tilde{x}_1', \ldots, \tilde{x}_N')^\top$ are the estimated values by linear regression, in the two cases. It is easy to show that $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ have the following geometrical interpretation. Let $H \subseteq \mathfrak{R}^N$ be the range of the $N \times N$ matrix $\mathbf{K} = \mathbf{X}^\top \mathbf{X}$. Then $\tilde{\mathbf{x}}$ is the projection of $\mathbf{x}$ on H. In other words, calling $\mathbf{v}_1, \ldots, \mathbf{v_m}$ the (orthonormal) eigenvectors of $\mathbf{K}$ with nonvanishing eigenvalue and

$$P = \sum_{i=1}^{m} \mathbf{v_i} \mathbf{v_i}^\top$$

the projector on the space H, we have $\tilde{\mathbf{x}} = P\mathbf{x}$. Let us define $\mathbf{y} = \mathbf{x} - P\mathbf{x}$. Analogously $\tilde{\mathbf{x}}' = P'\mathbf{x}$, P' being the projector on the $2m$-dimensional space $H' \subseteq \mathfrak{R}^N$, equal to the range of the matrix $\mathbf{K}' = \mathbf{Z}^\top \mathbf{Z}$. Moreover, it is easy to show that

$$\delta = \frac{\tilde{\mathbf{x}}'^\top \tilde{\mathbf{x}}' - \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}}}{1 - \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}}}. \tag{2}$$

Now note that $H \subseteq H'$. Therefore we may decompose H' as follows: $H' = H \oplus H^\perp$, where $H^\perp$ is the space of all vectors of H' orthogonal to all vectors of H. $H^\perp$ corresponds to the additional features due to the inclusion of $\{\eta\}$ variables. Calling $P^\perp$ the projector on $H^\perp$, we can write

$$\delta = \frac{\|P^\perp \mathbf{y}\|^2}{1 - \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}}}. \tag{3}$$

Now we note that $H^\perp$ is the range of the matrix

$$\tilde{\mathbf{K}} = \mathbf{K}' - \mathbf{K}'P - P(\mathbf{K}' - \mathbf{K}'P) = \mathbf{K}' - P\mathbf{K}' - \mathbf{K}'P + P\mathbf{K}'P.$$

Indeed, for any $\mathbf{u} \in \mathfrak{R}^N$, we have $\tilde{\mathbf{K}}\mathbf{u} = \mathbf{v} - P\mathbf{v}$, where $\mathbf{v} = \mathbf{K}'(\mathbf{I} - P)\mathbf{u} \in H'$, and $\tilde{\mathbf{K}}\mathbf{u} \in H^\perp$. It follows that $H^\perp$ is spanned by the set of the eigenvectors, with nonvanishing eigenvalue, of $\tilde{\mathbf{K}}$. Calling $\mathbf{t}_1, \ldots, \mathbf{t_m}$ these eigenvectors, we have

$$\delta = \sum_{i=1}^{m} r_i^2, \tag{4}$$

where r_i is the Pearson's correlation coefficient of $\mathbf{y}$ and $\mathbf{t_i}$ (since the overall sign of $\mathbf{t_i}$ is arbitrary, we can assume that r_i is positive). Let π_i be the probability that r_i is due to chance, obtained by Student's t test. Since we are dealing with multiple comparison, we use the Bonferroni correction to select the eigenvectors $\mathbf{t}_{i'}$, correlated with $\mathbf{y}$, with expected fraction of false positive q (equal to 0.05). Therefore we calculate a causality index by summing, in Eq. (4), only over the $\{r_{i'}\}$ such that $\pi_{i'} < q/m$, thus obtaining a *filtered* linear Granger causality index,

$$\delta_F = \sum_{i'} r_{i'}^2. \tag{5}$$

It is assumed that δ_F measures the causality $\eta \to \xi$, without further statistical test. We conclude this section noting that, in this frame, static correlations are neglected.

B. Kernel-Granger causality

In this section we describe the generalization of linear Granger causality to the nonlinear case, using methods from the theory of reproducing kernel Hilbert spaces (RKHS) [16]. Given a kernel function K, with spectral representation $K(X,X') = \Sigma_a \lambda_a \phi_a(X) \phi_a(X')$ (see Mercer's theorem [15]), we consider H, the range of the $N \times N$ Gram matrix **K** with elements $K(X_i, X_j)$. In order to make the mean of all variables $\phi_a(X)$ zero, we replace $\mathbf{K} \to \mathbf{K} - P_0\mathbf{K} - \mathbf{K}P_0 + P_0\mathbf{K}P_0$, where P_0 is the projector onto the one-dimensional subspace spanned by the vector such that each component is equal to the unity [16]; in the following we assume that this operation has been performed on each Gram matrix. As in the linear case, we calculate $\widetilde{\mathbf{x}}$, the projection of **x** onto H. Due to the spectral representation of K, $\widetilde{\mathbf{x}}$ coincides with the linear regression of **x** in the feature space spanned by $\sqrt{\lambda_a} \phi_a$, the eigenfunctions of K; the regression is nonlinear in the original variables.

While using both X and Y to predict x, we evaluate the Gram matrix $\mathbf{K}'$ with elements $K'_{ij} = K(Z_i, Z_j)$. The regression values now form the vector $\widetilde{\mathbf{x}}'$ equal to the projection of **x** on H', the range of $\mathbf{K}'$. Before evaluating the filtered causality index, as in the linear case, we note that not all kernels may be used to evaluate Granger causality. Indeed if Y is statistically independent of X and x, then $\widetilde{\mathbf{x}}'$ and $\widetilde{\mathbf{x}}$ should coincide in the limit $N \to \infty$. This property, invariance of the risk minimizer when statistically independent variables are added to the set of input variables, is satisfied only by suitable kernels, as discussed in [22]. In the following we consider two possible choices, which fulfill the invariance requirement.

1. Inhomogeneous polynomial kernel

The inhomogeneous polynomial (IP) kernel of integer order p is

$$K_p(X,X') = (1 + X^\top X')^p.$$

In this case the eigenfunctions are made of all the monomials, in the input variables, up to the pth degree. The dimension of the space H is $m_1 = 1/B(p+1, m+1) - 1$, where B is the beta function, and $p=1$ corresponds to the linear regression. The dimension of space H' is $m_2 = 1/B(p+1, 2m+1) - 1$. As in the linear case, we note that $H \subseteq H'$ and decompose $H' = H \oplus H^\perp$. Subsequently we calculate $\widetilde{\mathbf{K}} = \mathbf{K}' - \mathbf{P}\mathbf{K}' - \mathbf{K}'\mathbf{P} + \mathbf{P}\mathbf{K}'\mathbf{P}$; the dimension of the range of $\widetilde{\mathbf{K}}$ is $m_3 = m_2 - m_1$. Along the same lines as those described in the linear case, we construct the kernel-Granger causality taking into account only the eigenvectors of $\widetilde{\mathbf{K}}$ which pass the Bonferroni test,

$$\delta_F^K = \sum_{i'} r_{i'}^2, \tag{6}$$

the sum being only over the eigenvectors of $\widetilde{\mathbf{K}}$ with probability $\pi_{i'} < q/m_3$.

2. Gaussian kernel

The Gaussian kernel reads

$$K_\sigma(X,X') = \exp\left(-\frac{(X-X')^\top(X-X')}{2\sigma^2} \right), \tag{7}$$

and depends on the width σ. σ controls the complexity of the model: the dimension of the range of the Gram matrix **K** decreases as σ increases. As in previous cases, we may consider H, the range of the Gram matrix **K**, and H', the range of $\mathbf{K}'$, but in this case the condition $H \subseteq H'$ would not necessarily hold; therefore some differences in the approach must be undertaken. We call L the m_1-dimensional span of the eigenvectors of **K** whose eigenvalue is not smaller than $\mu\lambda_{max}$, where λ_{max} is the largest eigenvalue of **K** and μ is a small number (we use 10^{-6}). We evaluate $\widetilde{\mathbf{x}} = P\mathbf{x}$, where P is the projector on L. After evaluating the Gram matrix $\mathbf{K}'$, the following matrix is considered:

$$\mathbf{K}^* = \sum_{i=1}^{m_2} \rho_i \mathbf{w}_i \mathbf{w}_i^\top, \tag{8}$$

where $\{\mathbf{w}\}$ are the eigenvectors of $\mathbf{K}'$, and the sum is over the eigenvalues $\{\rho_i\}$ not smaller than μ times the largest eigenvalue of $\mathbf{K}'$. Then we evaluate $\widetilde{\mathbf{K}} = \mathbf{K}^* - \mathbf{P}\mathbf{K}^* - \mathbf{K}^*\mathbf{P} + \mathbf{P}\mathbf{K}^*\mathbf{P}$, and denote $P^\perp$ the projector onto the m_3-dimensional range of $\widetilde{\mathbf{K}}$. Note that the condition $m_2 = m_1 + m_3$ may not be strictly satisfied in this case (however in our experiments we find that the violation of this relation is always very small, if any). The kernel Granger causality index for the Gaussian kernel is then constructed as in the previous case, see Eq. (6).

III. MULTIVARIATE KERNEL CAUSALITY

Let $\{\xi(a)_n\}_{n=1,..,N+m}$, $a = 1, \ldots, M$, be M simultaneously recorded time series. In order to put in evidence the drive-response pattern in this system, one may evaluate the bivariate Granger causality, described in the previous sections, between every pair of time series. It is recommended, however, to treat the data set as a whole, thus generalizing kernel Granger causality to the multivariate case, as follows. We denote

$$X(c)_i = (\xi(c)_i, \ldots, \xi(c)_{i+m-1})^\top,$$

for $c = 1, \ldots, M$ and $i = 1, \ldots, N$. In order to evaluate the causality $\{\xi(a)\} \to \{\xi(b)\}$, we define, for $i = 1, \ldots, N$,

$$Z_i = (X(1)_i^\top, \ldots, X(a)_i^\top, \ldots, X(M)_i^\top)^\top,$$

containing all the input variables, and

$$X_i = (X(1)_i^\top, \ldots, X(M)_i^\top)^\top,$$

containing all the input variables but those related to $\{\xi(a)\}$. Gram matrices **K** and $\mathbf{K}'$ are then evaluated: $K_{ij} = K(X_i, X_j)$ and $K'_{ij} = K(Z_i, Z_j)$. The target vector is now
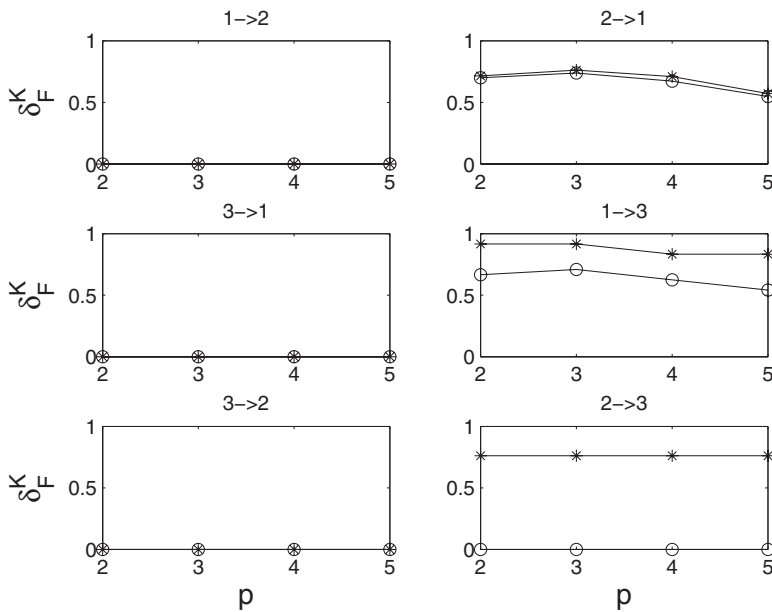
FIG. 1. The causal relationships between all pairs of maps, in the example of three logistic maps described in the text, using multivariate Granger causality (empty circles) and bivariate Granger causality (stars). Here $m=1$ and the IP kernel, with various values of p, is used.

$\mathbf{x}=(\xi(b)_{1+m},\ldots,\xi(b)_{N+m})^\top$. Along the same lines as in the bivariate case, for IP kernel or the Gaussian one, we then calculate the causality index as in Eq. (6): it is denoted $\delta_F^K(a \to b)$ and measures the strength of causality $a \to b$, taking into account all the available variables. Repeating these steps for all a and b, the causality pattern in the data set is evaluated. Note that the threshold for the Bonferroni's correction, in the multivariate case, must be lowered by the number of pairs $M(M-1)/2$.

As an example, we consider the following system of three logistic maps [23]:

$$x_{1,t} = 0.8(1 - ax_{1,t-1}^2) + 0.2(1 - ax_{2,t-1}^2) + s\tau_{1,t},$$
$$x_{2,t} = 1 - ax_{2,t-1}^2 + s\tau_{2,t},$$
$$x_{3,t} = 0.8(1 - ax_{3,t-1}^2) + 0.2(1 - ax_{1,t-1}^2) + s\tau_{3,t}; \quad (9)$$

here $a=1.8$, $s=0.01$, and τ's are unit variance Gaussian noise terms. The causal relationships implemented in these equations are $2 \to 1$ and $1 \to 3$. Analyzing segments of length $N=1000$, we evaluate both the multivariate causality, as described in Sec. IV, and the bivariate causality for all pairs of maps. We use the IP kernel with various values of p; the results are displayed in Fig. 1. It is well known [11] that performing pairwise evaluation for multivariate data has the drawback that one cannot discern whether the influence between two time series is direct or is mediated by other time series. This is what happens in the present example. Both the multivariate and the bivariate analysis reveal the influences $2 \to 1$ and $1 \to 3$. On the other hand, the bivariate analysis reveals also the influence $2 \to 3$, which is actually mediated by 1: the multivariate analysis recognizes $2 \to 3$ as nonsignificant.

IV. ANALYSIS OF DYNAMICAL NETWORKS

In this section we simulate two dynamical networks and apply the multivariate Granger analysis to estimate the topology structure of systems from time series data.

A. Network of chaotic maps

Let us consider a coupled map lattice of n nodes, with equations, for $i=1,\ldots,n$,

$$x_{i,t} = \left(1 - \sum_{j=1}^n c_{ij}\right)(1 - ax_{i,t-1}^2) + \sum_{j=1}^n c_{ij}(1 - ax_{j,t-1}^2) + s\tau_{i,t},$$
$$(10)$$

where a, s, and τ's are the same as in Eq. (9), and c_{ij} represents the coupling $j \to i$. We fix $n=34$ and construct couplings as follows. We consider the well-known Zachary data set [24], an undirected network of 34 nodes. We assign a direction to each link, with equal probability, and set c_{ij} equal to 0.05, for each link of the directed graph thus obtained, and zero otherwise. The network is displayed in Fig. 2: here the goal is to estimate this directed network from the measurements of time series on nodes.

The multivariate Granger analysis, described in the previous section, perfectly recovers the underlying network using the IP kernel with $p=2$, $m=1$, and $N=10\,000$. Note that while evaluating $\delta_F^K(j \to i)$, for all i and j, 39 270 Pearson's coefficients r are calculated. Their distribution is represented in Fig. 3: there is a strong peak at $r=0$ (corresponding to projections that are discarded), and a very low number of projections with a rather large value of r, up to about $r=0.6$. It is interesting to describe the results in terms of a threshold for correlations. Given a threshold value r_{th}, we select the correlation coefficients whose value is greater than r_{th}. We then calculate the corresponding causality indexes $\delta_F^K(j \to i)$, and construct the directed network whose links correspond to nonvanishing elements of $\delta_F^K(j \to i)$. In Fig. 4 (top) we plot the total number of links of the reconstructed network, as a function of the threshold r_{th}: the curve shows a plateau, around $r_{th}=0.1$, corresponding to a directed network which is stable against variations of r_{th}. At the plateau 428 projections are selected, which coincide with those selected by means of Bonferroni's test. In Fig. 4 (bottom) we plot the
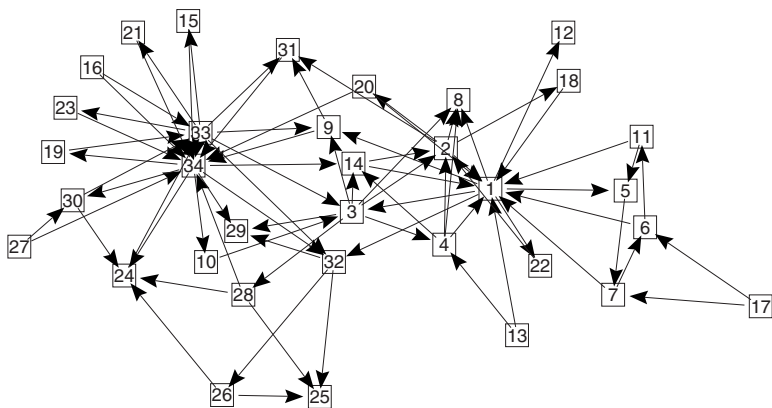
FIG. 2. The directed network of 34 nodes obtained assigning randomly a direction to links of the Zachary network.

number of errors (the sum of the number of links that exist in the true network and are not recovered, plus the number of links that exist only in the recovered network) versus the threshold r_{th}: the plateau leads to perfect reconstruction of the original network. We stress that a large number of samples is needed to recover the underlying network: in the typical case we find that the network is perfectly reconstructed if $N \geq 5000$, while if N is further lowered some errors in the reconstructed network appear. Moreover, it is important to observe that, although all couplings c_{ij} have the same magnitude, the causality strengths $\delta_F^K(j \to i)$ depend on the link, as it is shown in Fig. 5. Granger causality is a measure of the information being transferred from one time series to another, and it should not be identified with couplings.

B. Genetic regulatory network

In this section we consider time series from a model of genetic regulatory network made of genes linked by weighted connections (inhibitory or excitatory) [25]. The expression levels of all genes, organized in a vector **g**, evolve as follows:

FIG. 3. The distribution of the 39 270 r values calculated while evaluating the causality indexes of the coupled map lattice (see the text).

$$\mathbf{g}_{t+1} = \mathbf{g}_t + A(\mathbf{g}_t - T\mathbf{I}) + \mathbf{\Sigma}, \qquad (11)$$

where A is a connectivity strength matrix corresponding to the network, $T=50$, $\mathbf{I}$ is the identity matrix, and $\mathbf{\Sigma}$ is a vector of random variables uniform in $[-10, 10]$. The values of **g** are restricted by floor and ceiling function to range in $[0,100]$: this constraint provides the nonlinear character of the model. As the simulation runs, multivariate data are sampled every t_s time steps. Moreover, the continuous data values are discretized into n_c categories (with equal bin sizes). In [25] dynamic Bayesian network (DBN) models [26] were trained to data of length N to recover the structure of matrix A: the values $t_s=5$, $n_c=3$, and $N \geq 2000$ were found to lead to the best reconstruction of genetic networks by DBN.

The genetic network we consider here is an example from [25] and consists of ten genes with connections described in Fig. 6: there are two independent regulatory pathways, one of which includes a large feedback structure. In Fig. 7 the
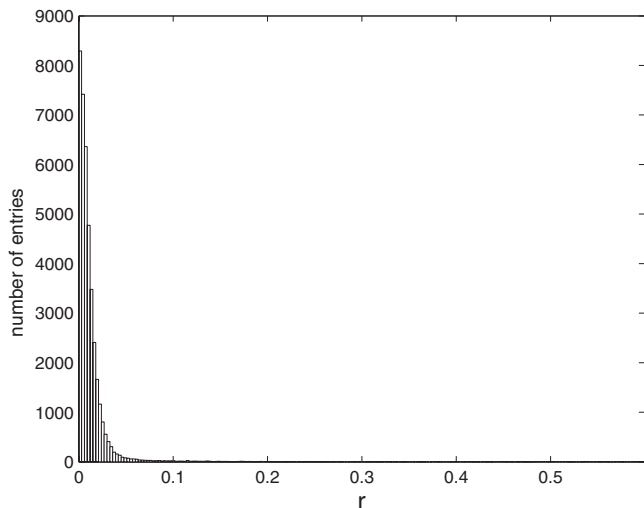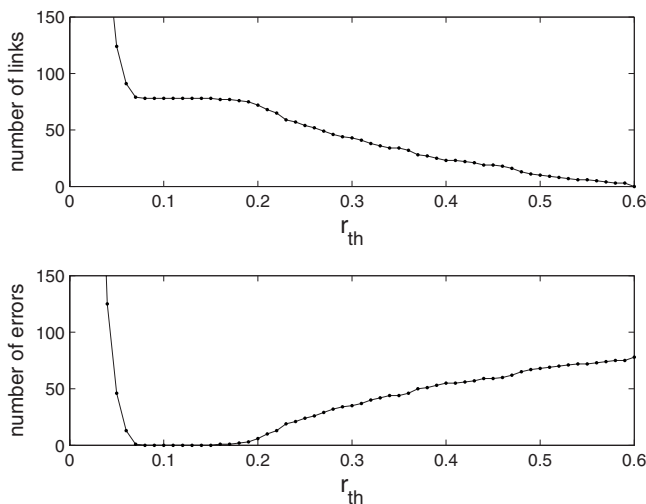

FIG. 4. (Top) Concerning the coupled map lattice, the horizontal axis represents the threshold for the values of r; the plot shows the number of links of the directed network constructed from the projections whose Pearson's coefficient exceeds the threshold. (Bottom) The total number of errors, in the reconstructed network, is plotted versus the threshold r_{th}. At large r_{th} the errors are due only to missing links, whereas at small r_{th} the errors are due only to links that do not exist in the true network and are recovered.
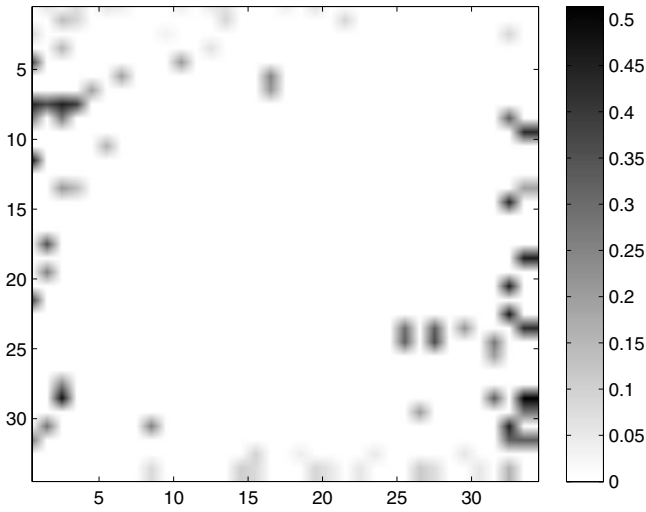
FIG. 5. Concerning the coupled map lattice, the causality indexes $\delta_F^K(j \rightarrow i)$, for all pairs of maps, are represented in a gray scale.
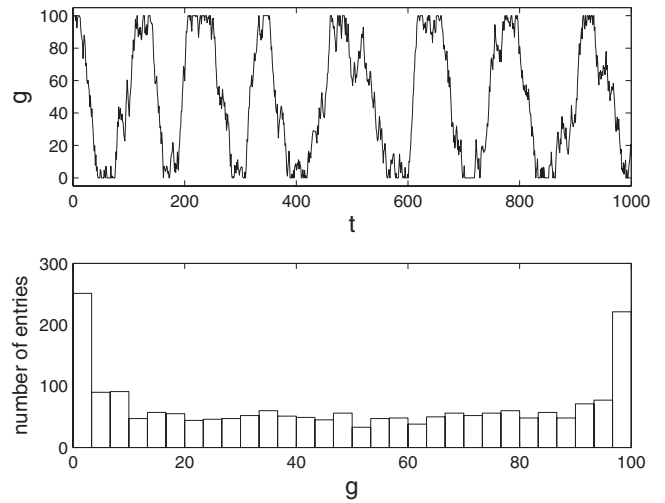

FIG. 7. Top: the typical curve of expression of a gene in the simulated regulatory network. Bottom: the distribution of expressions for the same gene.

typical curve of expression for a gene in the network is represented (top); the distribution of expressions, for the same gene, is bimodal (bottom). We simulate 100 times the system equations, starting from different initial conditions, and sample time series of length $N=2000$, $t_s=5$, and $n_c=3$. The typical recovered network by DBN, on this example, corresponds to one missing link (from node 7 to 10) [25]. The linear multivariate Granger approach, with $m=2$, leads to perfect reconstruction of the network in 90 cases out of the 100; we obtain similar results on all the examples presented in [25]. Using IP (with $p > 1$) and Gaussian kernels we obtain similar performances as the ones from the linear kernel.

It is interesting to stress that the possibility that one has to reconstruct the true genetic network depends on the sampling rate. In Fig. 8 we plot the mean number of errors (over 100

realizations) in the reconstructed network as a function of t_s for the linear kernel and Gaussian kernel: the performance degrades as t_s moves far from 5.

Let us now discuss the case of large t_s: all Granger causalities are recognized as nonsignificant. On the other hand, at large t_s, we find significant *static* linear correlations between time series of all pairs of genes belonging to the same pathway. In other words, referring to Fig. 6, the linear correlations of times series from every pair of genes extracted from $\{1,2,3,4,8,9\}$ is significant, as well as the linear correlation for every pair extracted from $\{5,6,7,10\}$; consistently the linear correlation, for pairs of genes from different pathways, is not significant. We conclude that, at large t_s
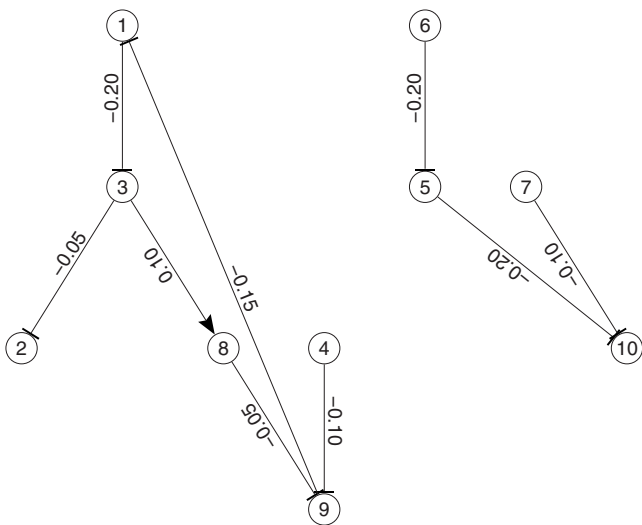

FIG. 6. The genetic regulatory network analyzed in the text. Numbers next to links specify regulation strength; arrows: excitatory; flat heads: inhibitory.
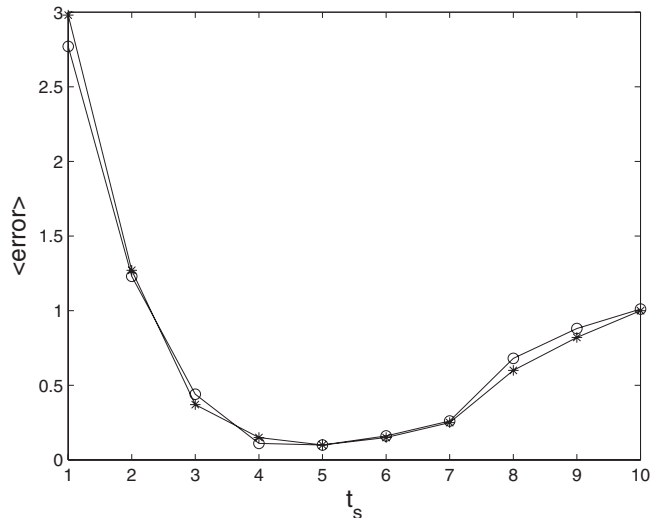

FIG. 8. Concerning the genetic regulatory network, the mean number of errors (over 100 realization of the system of length $N=2000$) obtained using the linear IP kernel ($p=1$, empty circles) and Gaussian kernel (stars), is plotted versus the sampling time t_s. In the Gaussian case, $\sigma=50$ is used. The level of discretization is $n_c=3$ and $m=2$.

Granger causality analysis is not effective, as it neglects static correlations. On the other hand, the analysis of static correlations may put in evidence groups of genes belonging to the same regulatory pathway.

V. SPARSELY CONNECTED DYNAMICAL NETWORKS AND LIMITED DATA

The Granger causality approach for dynamical networks here presented requires a large amount of data samples to provide trustable answers. However, there are situations (frequent in bioinformatics) where the number of samples is smaller than the number of variables (genes): in these situations the multivariate Granger approach is unfeasible. Under the assumption of sparse connectivity, it has been proposed to replace least squares methodology with a multivariate approach using minimization with respect to $L1$ norm [9]. Here we show that there are situations where, even in presence of sparse connectivity and limited data, use of bivariate Granger causality is a better choice with respect to $L1$ minimization. Indeed, in these cases, the statistical robustness of the estimation of information flow between pairs of time series may still be good, with the drawback that some causality links, found by the bivariate approach, may not be direct but mediated.

A. Dynamical network by preferential attachment

We construct a network of 100 nodes and 100 links using the preferential attachment procedure [27]; we give randomly a direction to each link, with $1/2$ probabilities, thus obtaining a directed network. Let us denote $d(i)$ the number of nodes from which a link pointing to i starts. We evolve a linear system on this network, with equations

$$x_{i,t} = a_i x_{i,t-1} + \sum_{j \to i} \left(\frac{0.8}{d(i)} \right) x_{j,t-1} + \tau_{i,t}. \tag{12}$$

The sum is over nodes such that $j \to i$ is a link of the network; τ's are unit variance Gaussianly distributed noise terms; a_i is one, if $d(i)=0$, and 0.2 otherwise. After a transient, we sample n_s consecutive time points, with $n_s = 20, 30, 40, 50, 60$. The $L1$ approach we use is the following. For each $i=1,\ldots,100$, we find the vector **c** with minimum $L1$ norm, among all those satisfying

$$x_{i,t+1} = \sum_{j=1}^{100} c_j x_{j,t}, \tag{13}$$

$t=1,\ldots,n_s-1$. The interaction $j \to i$ is considered significant if the absolute value of c_j exceeds a threshold, fixed so that the total number of false positive connections is five. Subsequently we apply the bivariate linear Granger approach, described in Sec. II A, for each pair of nodes: also for Granger's approach we fix a threshold for the correlation coefficients r, see Eq. (4), so that the total number of false positive connections is five. In Fig. 9 we depict the number of true positive connections found by the two approaches, as a function of n_s. It is clear that here the bivariate Granger approach outperforms $L1$ minimization.
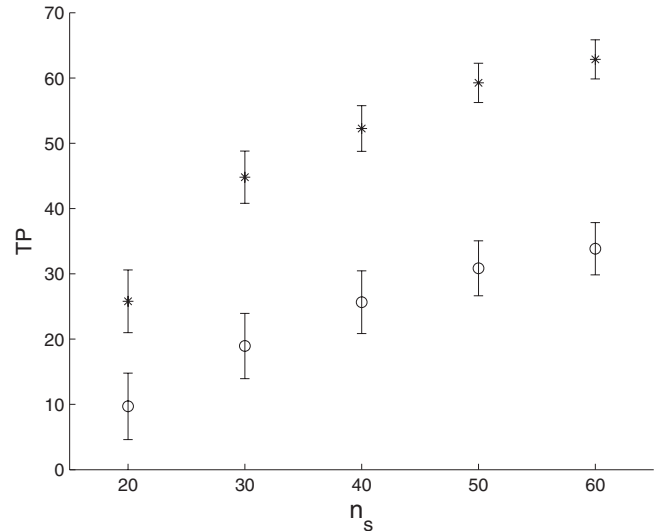

FIG. 9. The true positive connections found by the bivariate Granger approach (stars) and by the multivariate $L1$ minimization (empty circles) on the preferential attachment dynamical network. The error bar represents one standard deviation, evaluated over 50 realizations. Here the number of false positive connections is always set equal to 5, however the result of the comparison between the two methods is robust and does not depend on this choice.

B. HeLa gene expression regulatory network

HeLa is the most famous cell culture line to date [28]. These are cells isolated from a human uterine cervical carcinoma in 1951 and used in biomedical research especially to culture viruses. While the patient ultimately died of her cancer eight months after the operation, her cells have lived on, still surviving in laboratories today. HeLa cells have somehow acquired cellular immortality, in that the normal mechanisms of programmed cell death after a certain number of divisions have somehow been switched off. We apply our approach to the HeLa cell gene expression data of [29]. Data corresponds to 94 genes and 48 time points, with an hour interval separating two successive readings (the HeLa cell cycle lasts 16 hours). The 94 genes were selected, from the full data set described in [30], on the basis of the association with cell cycle regulation and tumor development.

First of all, we perform the analysis of the static pairwise correlations between time series: 800 pairs of genes are significantly correlated. Drawing a link for each correlated pair leads to an undirected network depicted in Fig. 10: it is clear that there are two modules, and symbols in Fig. 10 corresponds to the partition by the method of module identification described in [31]. The first module is made of 23 genes and corresponds to the regulatory network of the transcriptional factor NFkB [33]; it contains several well known activators and targets of NFkB [34], such as, e.g., A20, ICAM-1, IL-6, VCAM-1, IkappaBa, JunB, MCP-1, FGF2, Cyclin. The second module, 62 genes, appears to be orchestrated by transcriptional factors p53 and STAT3. Note, however, that the two modules are not independent, as they form a highly related network. The proto-oncongene c-myc appears to be central between the two modules: it has 12 significant static correlations with both modules. After the dis-
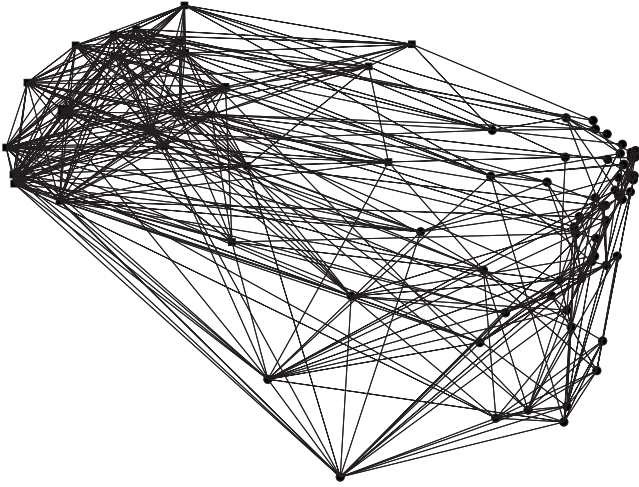
FIG. 10. The undirected network obtained drawing a link between all pairs of significantly correlated genes of the HeLa data set (9 genes are not represented here as they are not correlated with any other gene; hence the number of nodes is 85). Squares and circles corresponds to the partition in two modules performed by the method described in [31]. The large square corresponds to the transcriptional factor NFkB.

cussion in Sec. IV B, we assume that the modular structure depicted in Fig. 10 is the result of regulatory mechanisms acting on time scales much smaller than the sampling time.

Next, in order to detect causalities acting on the time scale of the sampling time, we apply bivariate Granger causality analysis. For all pairs (i,j), we use the linear IP kernel version ($p=1$, $m=1$) of our approach and evaluate the Pearson correlation coefficient r, Eq. (4), for the causality $i \rightarrow j$. Due to the small number of samples, we do not use t test to evaluate the probability π corresponding to r: we generate a set of surrogates by permuting the temporal indices of the ith times series while keeping fixed those of the jth time series. The probability π is identified with the fraction of times that an higher coefficient is obtained over 3×10^6 random shufflings of time indices of the ith time series. Moreover, we use the false discovery rate (FDR) method [32] instead of Bonferroni's correction. FDR works in the following way: the $94 \times 93 = 8742$ Pearson coefficients are ordered, $\{r_\ell\}$, according to their increasing π_ℓ values, and a parameter q, which controls the fraction of false positive, is set to 0.05. The index ℓ' is identified as the largest such that for all $\ell \le \ell'$ we have $\pi_\ell \le \ell q / 8742$. Pearson coefficients r_ℓ are accepted for $\ell \le \ell'$. This procedure selects 19 causal relationships, out of 8742; they are listed in Table I. IkappaBa is the most abundant inhibitory protein for NFkB [35]: our approach detects the significant causality IkappaBa $\rightarrow$ NFkB. We find that NFkB is also casually related to IAP, an antiapoptotic gene, and B99 (a direct target for transcriptional activation by $p53$: here no significant interaction between B99 and p53 has been detected). Three causality relationships involve Bcl-xL, the dominant regulator of apoptosis (active cell suicide) and TSP1, a peptide shown in some tumor systems to be linked with angiogenesis. Notably Table I also contains fibroblast growth factors, FGF7 and FGFR4, the tumor necrosis factor Killer/DR5, the myeloid tumor suppressor gene PKIG, the

TABLE I. Causalities for the HeLa gene network (Pearson's coefficients are displayed in the parentheses).

TPD52L	$\rightarrow$	TNF-a	(0.6166)
TSP1	$\rightarrow$	Bcl-XL	(0.5935)
TSP1	$\rightarrow$	c-myc	(0.5905)
IRF-2	$\rightarrow$	BRCA1	(0.5669)
c-myc	$\rightarrow$	FGFR4	(0.5653)
R2	$\rightarrow$	c-myc	(0.5642)
PKIG	$\rightarrow$	TSP1	(0.5475)
IAP	$\rightarrow$	NFkB	(0.5457)
NFkB	$\rightarrow$	B99	(0.5416)
Bcl-XL	$\rightarrow$	OCT4	(0.5354)
OCT4	$\rightarrow$	VCAM-1	(0.5329)
VCAM-1	$\rightarrow$	TPD52L	(0.5315)
Killer/DR5	$\rightarrow$	c-myc	(0.5313)
A20	$\rightarrow$	Bcl-XL	(0.5284)
Cyclin E1	$\rightarrow$	E2F-1	(0.5249)
PKIG	$\rightarrow$	ICAM-1	(0.5243)
IkappaBa	$\rightarrow$	NFkB	(0.5156)
TPD52L	$\rightarrow$	MASPIN	(0.5128)
FGF7	$\rightarrow$	MCP-1	(0.5107)

tumor protein TPD52-L and Cyclin E1, a gene which is overexpressed in many tumors. In [29] data have been analyzed with the sparse vector autoregressive model, a multivariate L1 approach which depends on a regularization parameter, λ, fixed by cross validation. Only one causality relationship, out of the 19 in Table I, was revealed also in [29]: A20$\rightarrow$Bcl-XL.

VI. DISCUSSION

Our method of analysis of dynamical networks is based on a recent measure of Granger causality between time series, rooted on kernel methods, whose magnitude depends on the amount of flow of information from one series to another. By definition of Granger causality, our method allows analysis of networks containing cycles. First we have demonstrated the effectiveness of the method on a network of chaotic maps with links obtained assigning a direction to the edges of the well-known Zachary data set, using a nonlinear kernel: perfect reconstruction of the directed network is achieved provided that a sufficient number of samples is available.

Second, we studied a simulated genetic regulatory network. The results from our method were better than those from DBN approach. However our performance was strongly dependent on the sampling time, as it occurred also using DBN method. In this example, use of IP kernel, with ($p > 2$), or Gaussian kernels did not lead to improvement in the performance with respect to the linear kernel: this means that these kernel are not suitable to model the nonlinear constraint connected to the fact that expressions are confined in [0,100]. Further work will be devoted to the search for kernels capable to capture this kind of nonlinearity: for a given

application one should choose the proper kernel out of the many possible classes [16].

Then we considered the case of sparse connectivity and limited data. Using an example consisting in a linear dynamical network on a graph grown by preferential attachment, we have shown that there are instances where the multivariate Granger approach is unfeasible, but the application of bivariate Granger analysis, to every pair of time series, leads to better results than those from a method based on $L1$ minimization. Finally we have analyzed a real data set of temporal gene expression samples from HeLa cells. The static correlation analysis between time series, which is the result of regulation mechanisms with time scales faster than the sampling rate, revealed the presence of two modules. Use of bivariate Granger causality has put in evidence 19 causality relationships acting on the time scale of one hour, all involving genes playing some role in processes related to tumor development. Our result on HeLa data has very little overlap with those from the output of a method based on multivariate L1 minimization, but this is not surprising, as we observed the same fact also on the linear dynamical model of Sec. V, where the true connectivity was known. We remark that currently available data size and data quality make the reconstruction of gene networks from gene expression data a challenge.

Detecting cause-effects influences between components of a complex system is an active multidisciplinary area of research in these years. The kernel approach here presented, provides a statistically robust tool to assess drive-response relationships in many fields of science.

[1] A. L. Barabasi, *Linked: The New Science of Networks* (Perseus Publishing, Cambridge, MA, 2002); S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[2] L. F. Abbott and C. van Vreeswijk, Phys. Rev. E **48**, 1483 (1993).

[3] K. Wiesenfeld, Physica B **222**, 315 (1996).

[4] T. S. Gardner, D. Bernardo, D. Lorenz, and J. J. Collins, Science **301**, 102 (2003).

[5] C. L. Tucker, J. F. Gera, and P. Uetz, Trends Cell Biol. **11**, 102 (2001).

[6] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi, Nature (London) **407**, 651 (2000).

[7] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002); S. Boccaletti, D. U. Hwang, M. Chavez, A. Amann, J. Kurths, and L. M. Pecora, Phys. Rev. E **74**, 016102 (2006).

[8] D. Yu, M. Righero, and L. Kocarev, Phys. Rev. Lett. **97**, 188701 (2006).

[9] D. Napoletani and T. D. Sauer, Phys. Rev. E **77**, 026103 (2008).

[10] C. W. J. Granger, Econometrica **37**, 424 (1969); for recent reviews see K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007); also see M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 903 (2007).

[11] M. Kaminski, M. Z. Ding, W. A. Truccolo, and S. L. Bressler, Biol. Cybern. **85**, 145 (2001); K. J. Blinowska, R. Kus, and M. Kaminski, Phys. Rev. E **70**, 050902 (R)(2004); A. K. Seth, Network Comput. Neural Syst. **16**, 35 (2005).

[12] R. Ganapathy, G. Rangarajan, and A. K. Sood, Phys. Rev. E **75**, 016211 (2007).

[13] L. Faes, G. Nollo, and K. H. Chon, Ann. Biomed. Eng. **36**, 381 (2008).

[14] M. Dhamala, G. Rangarajan, and M. Ding, Phys. Rev. Lett. **100**, 018701 (2008).

[15] V. Vapnik, *Statistical Learning Theory* (Wiley, New York, 1998).

[16] J. Shawe-Taylor and N. Cristianini, *Kernel Methods For Pattern Analysis* (Cambridge University Press, London, 2004).

[17] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**, 144103 (2008).

[18] M. Palus and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007); H. Nalatore, M. Ding, and G. Rangarajan, *ibid.* **75**, 031123 (2007); K. Ishiguro, N. Otsu, M. Lungarella, and Y. Kuniyoshi, *ibid.* **77**, 026216 (2008).

[19] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).

[20] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[21] J. Geweke, J. Am. Stat. Assoc. **77**, 304 (1982).

[22] N. Ancona and S. Stramaglia, Neural Comput. **18**, 749 (2006).

[23] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, London, 2002).

[24] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[25] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, Bioinformatics **20**, 3594 (2004).

[26] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, J. Comput. Biol. **7**, 601 (2000).

[27] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).

[28] J. R. Masters, Nat. Rev. Cancer **2**, 315 (2002).

[29] A. Fujita *et al.*, BMC Syst. Biol. **1**, 39 (2007).

[30] M. L. Whitfield *et al.*, Mol. Biol. Cell **13**, 1977 (2002).

[31] L. Angelini, S. Boccaletti, D. Marinazzo, M. Pellicoro, and S. Stramaglia, Chaos **17**, 023114 (2007).

[32] J. Benjamini and Y. Hochberg, J. R. Stat. Soc. Ser. B (Methodol.) **57**, 289 (1995).

[33] J. Inoue, J. Gohda, T. Akiyama, and K. Semba, Cancer J. Sci. Am. **98**, 268 (2007).

[34] H. L. Pahl, Oncogene **18**, 6853 (1999).

[35] F. Chen, V. Castranova, and X. Shi, Am. J. Clin. Pathol. **159**, 387 (2001).