

Information Transfer in the Brain: Insights from a Unified Approach

Daniele Marinazzo, Guorong Wu, Mario Pellicoro, and Sebastiano Stramaglia

Abstract. Measuring directed interactions in the brain in terms of information flow is a promising approach, mathematically treatable and amenable to encompass several methods. In this chapter we propose some approaches rooted in this framework for the analysis of neuroimaging data. First we will explore how the transfer of information depends on the network structure, showing how for hierarchical networks the information flow pattern is characterized by exponential distribution of the incoming information and a fat-tailed distribution of the outgoing information, as a signature of the law of diminishing marginal returns. This was reported to be true also for effective connectivity networks from human EEG data. Then we address the problem of partial conditioning to a limited subset of variables, chosen as the most informative ones for the driver node. We will then propose a formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of each assigned target. Multiplets characterized by a large contribution to the expansion are associated to informational circuits present in the system, with an informational character (synergetic or redundant) which can be associated to the sign of the contribution. Applications are reported for EEG and fMRI data.

Daniele Marinazzo

University of Gent, Department of Data Analysis, 1 Henri Dunantlaan, B9000 Gent, Belgium
e-mail: daniele.marinazzo@ugent.be

Guorong Wu

University of Gent, Department of Data Analysis, 1 Henri Dunantlaan, B9000 Gent, Belgium
and Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science
and Technology, University of Electronic Science and Technology of China, Chengdu, China
e-mail: guorong.wu@ugent.be

Mario Pellicoro · Sebastiano Stramaglia

University of Bari, Physics Department,
Via Amendola 173, 70126 Bari, Italy
e-mail: {[mario.pellicoro](mailto:mario.pellicoro@ba.infn.it), [sebastiano.stramaglia](mailto:sebastiano.stramaglia@ba.infn.it)}@ba.infn.it

1 Economics of Information Transfer in Networks

Most social, biological, and technological systems can be modeled as complex networks, and display substantial non-trivial topological features [4, 10]. Moreover, time series of simultaneously recorded variables are available in many fields of science; the inference of the underlying network structure, from these time series, is an important problem that received great attention in the last years.

In many situations it can be expected that each node of the network may handle a limited amount of information. This structural constraint suggests that information transfer networks should exhibit some topological evidences of the law of diminishing marginal returns [36], a fundamental principle of economics which states that when the amount of a variable resource is increased, while other resources are kept fixed, the resulting change in the output will eventually diminish [26]. Here we introduce a simple dynamical network model where the topology of connections, assumed to be undirected, gives rise to a peculiar pattern of the information flow between nodes: a fat tailed distribution of the outgoing information, while the average incoming information transfer does not depend on the connectivity of the node. In the proposed model the units, at the nodes the network, are characterized by a transfer function that allows them to process just a limited amount of the incoming information. In this case a possible way to quantify the law of the diminishing marginal returns can be the discrepancy of the distributions, expressed as the ratio of their standard deviations.

1.1 Model

We use a simple dynamical model with a threshold in order to quantify and investigate this phenomenon. Given an undirected network of n nodes and symmetric connectivity matrix $A_{ij} \in \{0, 1\}$, to each node we associate a real variable x_i whose evolution, at discrete times, is given by:

$$x_i(t+1) = F\left(\sum_{j=1}^n A_{ij}x_j(t)\right) + \sigma\xi_i(t), \quad (1)$$

where ξ are unit variance Gaussian noise terms, whose strength is controlled by σ ; F is a transfer function chosen as follows:

$$\begin{aligned} F(\alpha) &= a\alpha & |\alpha| < \theta \\ F(\alpha) &= a\theta & \alpha > \theta \\ F(\alpha) &= -a\theta & \alpha < -\theta \end{aligned} \quad (2)$$

where θ is a threshold value. This transfer function is chosen to mimic the fact that each unit is capable to handle a limited amount of information. For large θ our model becomes a linear map. At intermediate values of θ , the nonlinearity connected to the threshold will affect mainly the mostly connected nodes (hubs): the input $\sum A_{ij}x_j$ to nodes with low connectivity will remain typically sub-threshold in

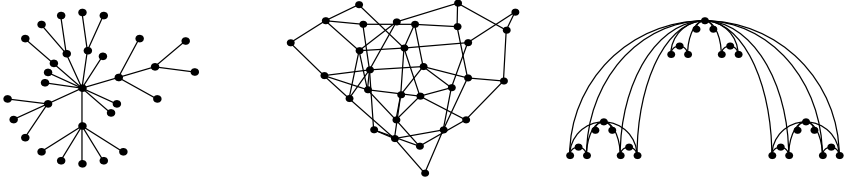


Fig. 1 Examples of the three network architectures used in this study. Left: Preferential Attachment. Center: Homogeneous. Left: Scale-free.

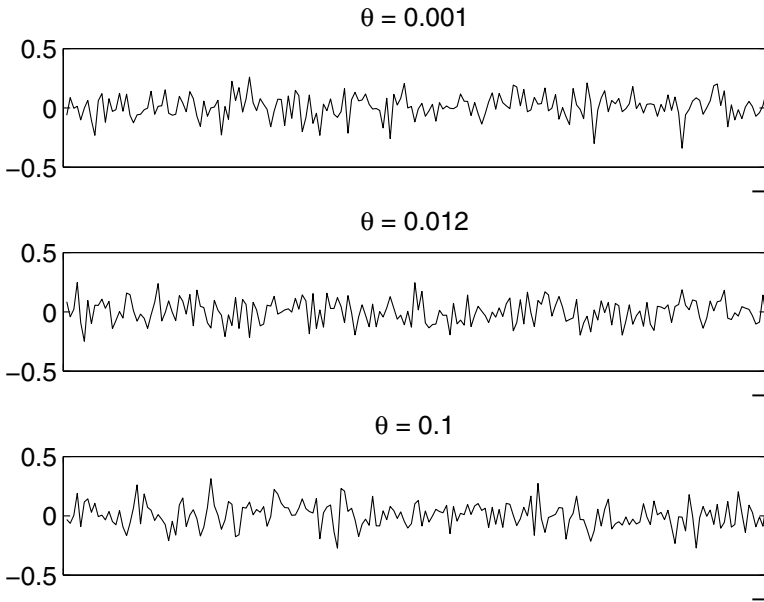


Fig. 2 Segments of 200 time points from typical time series simulated in the scale-free network for three values of θ

this case. We consider hierarchical networks generated by preferential attachment mechanism [2], which in the deterministic case leads to a scale-free network. Examples of a preferential attachment network, a scale free network and an homogeneous network are reported in figure 1. A segment of 200 time points of a typical time series for three values of θ is plotted in figure 2.

From numerical simulations of eqs. (1), we evaluate the linear causality pattern for this system as the threshold is varied. We verify that, in spite of the threshold, variables are nearly Gaussian so that we may identify the causality with the information flow between variables [5]. We compute the incoming and outgoing information flow from and to each node, c_{in} and c_{out} , summing respectively all the sources for a given target and all the targets for a given source. It is worth to underline that no

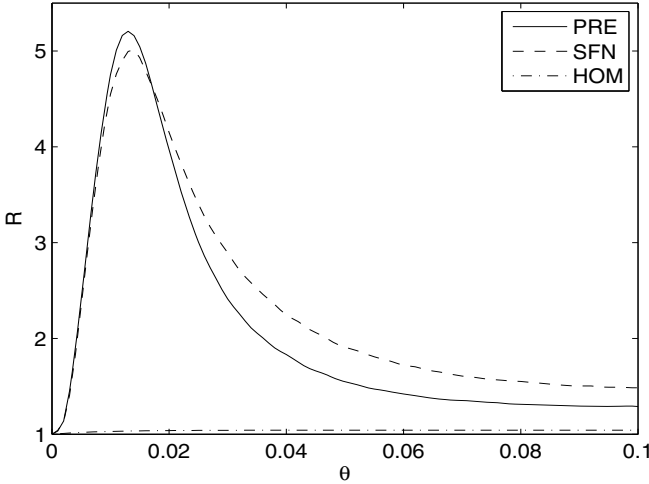


Fig. 3 The ratio between the standard deviation of c_{out} and those of c_{in} , R , is plotted versus θ for the three architectures of network: preferential attachment (PRE), deterministic scale free (SFN) and homogeneous (HOM). The parameters of the dynamical system are $a = 0.1$ and $\sigma = 0.1$. Networks built by preferential attachment are made of 30 nodes and 30 undirected links, while the deterministic scale free network of 27 nodes is considered. The homogeneous networks have 27 nodes, each connected to two other randomly chosen nodes.

threshold is applied to the connectivity matrix, so that all the information flowing in the network is accounted for. We then evaluate the standard deviation of the distributions of c_{in} and c_{out} , from all the nodes, varying the realization of the preferential attachment network and implementing eqs. (1) for 10000 time points.

In figure 3 we depict R , the ratio between the standard deviation of c_{out} over those of c_{in} , as a function of the θ . As the threshold is varied, we encounter a range of values for which the distribution of c_{in} is much narrower than that of c_{out} . In the same figure we also depict the corresponding curve for deterministic scale free networks [3], which exhibits a similar peak, and for homogeneous random graphs (or Erdos-Renyi networks [17]), with R always very close to one. The discrepancy between the distributions of the incoming and outgoing causalities arises thus in hierarchical networks. We remark that, in order to quantify the difference between the distributions of c_{in} and c_{out} , here we use the ratio of standard deviations but qualitatively similar results would have been shown using other measures of discrepancy.

In figure 4 we report the scatter plot in the plane $c_{in} - c_{out}$ for preferential attachment networks and for some values of the threshold. The distributions of c_{in} and c_{out} , with θ equal to 0.012 and corresponding to the peak of figure 3, are depicted in figure 5: c_{in} appears to be exponentially distributed, whilst c_{out} shows a fat tail. In other words, the power law connectivity, of the underlying network, influences just the distribution of outgoing directed influences. In figure 6 we show the

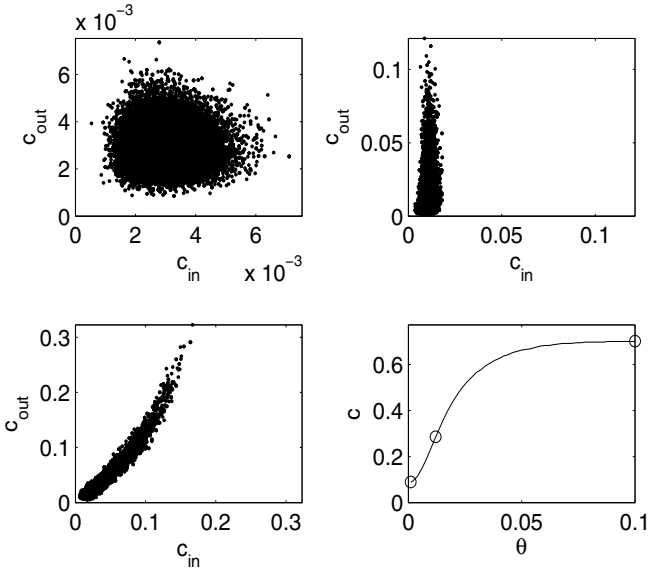


Fig. 4 Scatter plot in the plane $c_{in} - c_{out}$ for undirected networks of 30 nodes and 30 links built by means of the preferential attachment mechanism. The parameters of the dynamical system are $a = 0.1$ and $\sigma = 0.1$. The points correspond to all the nodes pooled from 100 realizations of preferential attachment networks, each with 10 simulations of eqs. (1) for 10000 time points. (Top-left) Scatter plot of the distribution for all nodes at $\theta = 0.001$. (Top-right) Contour plot of the distribution for all nodes at $\theta = 0.012$. (Bottom-left) Scatter plot of the distribution for all nodes at $\theta = 0.1$. (Bottom-right) The total Granger causality (directed influence) (obtained summing over all pairs of nodes) is plotted versus θ ; circles point to the values of θ in the previous subfigures.

average value of c_{in} and c_{out} versus the connectivity k of the network node: c_{out} grows uniformly with k , thus confirming that its fat tail is a consequence of the power law of the connectivity. On the contrary c_{in} appears to be almost constant: on average the nodes receive the same amount of information, irrespective of k , whilst the outgoing information from each node depends on the number of neighbors. It is worth mentioning that since a precise estimation of the information flow is computationally expensive, our simulations are restricted to rather small networks; in particular the distribution of c_{out} appears to have a fat tail but, due to our limited data, we can not claim that it corresponds to a simple power-law. The same model was then implemented on an anatomical connectivity matrix obtained via diffusion spectrum imaging (DSI) and white matter tractography [22]. Also in this case we observe a modulation of R and some scatter plots (figure 7) qualitatively similar to the ones depicted in figures 3 and 4. In this case a multimodal distribution emerges for high values of θ , as we can observe also in the histograms in figure 8. In figure 9 we can clearly identify some nodes in the structural connection matrix in which the

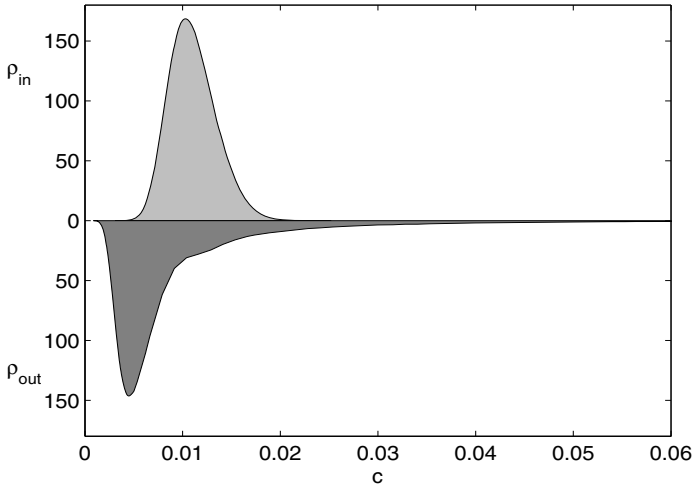
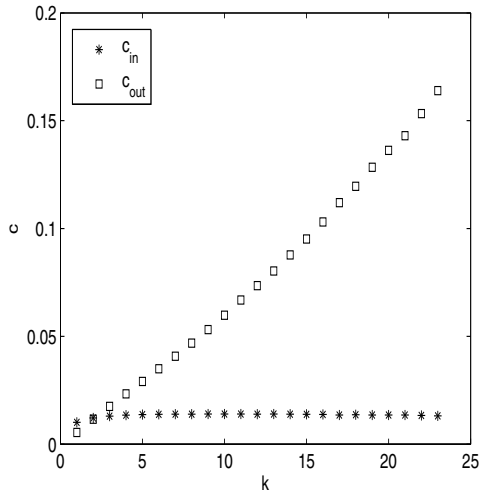


Fig. 5 For the preferential attachment network, at $\theta = 0.012$, the distributions (by smoothing spline estimation) of c_{in} and c_{out} for all the nodes, pooled from all the realizations, are depicted. Units on the vertical axis are arbitrary.

Fig. 6 In the ensemble of preferential attachment networks of figure (2), at $\theta = 0.012$, c_{in} and c_{out} are averaged over nodes with the same connectivity and plotted versus the connectivity.



law of diminishing marginal returns is highly expressed. The value of the threshold has also an influence on the ratio S between interhemispheric and intrahemispheric information transfer (figure 10). Interestingly, the maximum of this ratio occurs at a finite value of θ , different from those at which R is maximal.

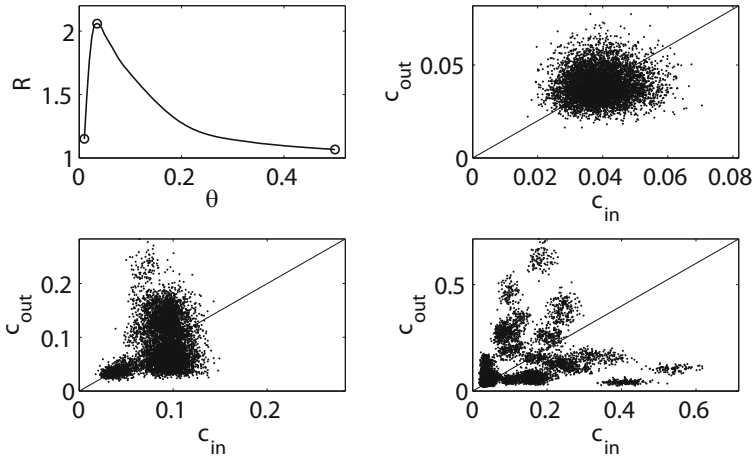


Fig. 7 Top right: the ratio between the standard deviation of c_{out} and those of c_{in} , R , is plotted versus θ when the threshold model is implemented on the connectome structure. Plots in the plane $c_{in} - c_{out}$ for three values of θ : 0.01 (top right), 0.0345 (bottom left), 0.5 (bottom right).

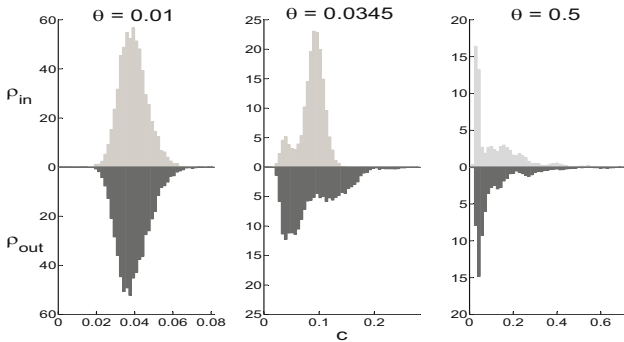


Fig. 8 The distributions of c_{in} and c_{out} for three values of θ when the threshold model is implemented on the connectome structure. Units on the vertical axis are arbitrary.

1.2 Electroencephalographic Recordings

As a real example we consider electroencephalogram (EEG) data. We used recording obtained at rest from 10 healthy subjects. During the experiment, which lasted for 15 min, the subjects were instructed to relax and keep their eyes closed. To avoid drowsiness, every minute the subjects were asked to open their eyes for 5 s. EEG was measured with a standard 10-20 system consisting of 19 channels [31]. Data were analyzed using the linked mastoids reference, and are available from [46].

For each subject we considered several epochs of 4 seconds in which the subjects kept their eyes closed. For each epoch we computed multivariate Kernel Granger

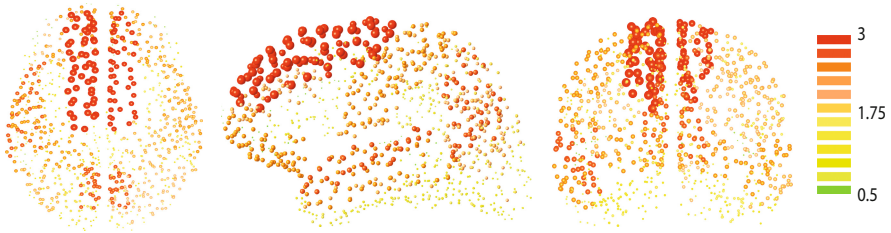
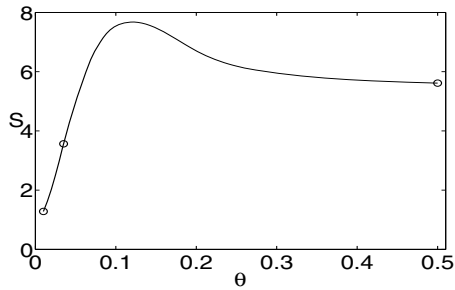


Fig. 9 The ratio between the standard deviation of c_{out} and those of c_{in} , R , is mapped on the 66 regions of the structural connectivity matrix. In the figure 998 nodes are displayed, with those belonging to the same region in the coarser template have the same color and size.

Fig. 10 The ratio S between intrahemispheric and interhemispheric information transfer in the threshold model implemented on the connectome structure as a function of θ . The circles indicate the same values of figures 7 and 8.



Causality [27] using a linear kernel and a model order of 5, determined by leave-one-out cross-validation. We then pooled all the values for information flow towards and from any electrode and analyzed their distribution.

In figure 11 we plot the incoming versus the outgoing values of the information transfer, as well as the distributions of the two quantities: the incoming information seems exponentially distributed whilst the outgoing information shows a fat tail. These results suggest that overall brain effective connectivity networks may also be considered in the light of the law of diminishing marginal returns.

More interestingly, this pattern is reproduced locally but with a clear modulation: a topographic analysis has also been made considering the distribution of incoming and outgoing causalities at each electrode. In figure 12 we show the distributions of incoming and outgoing connections corresponding to the electrodes locations on the scalp, and the corresponding map of the parameter R ; the law of diminishing marginal returns seems to affect mostly the temporal regions. This well defined pattern suggests a functional role for the distributions. It is worth to note that this pattern has been reproduced in other EEG data at rest from 9 healthy subjects collected for another study with a different equipment.

Fig. 11 For the EEG data the distributions of c_{in} and c_{out} are depicted in a scatter plot (left) and in terms of their distributions, obtained by smoothing spline estimation (right).

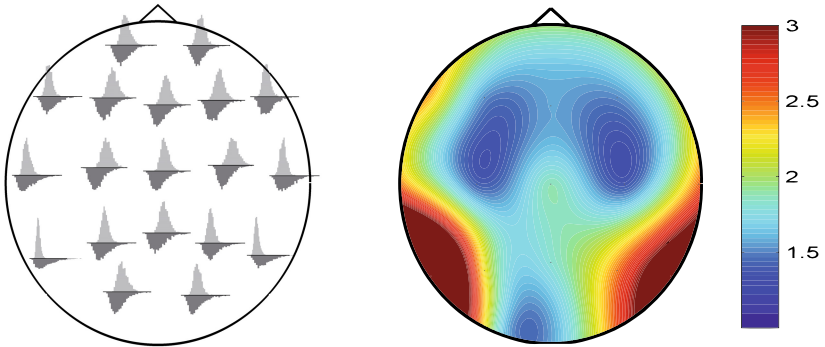
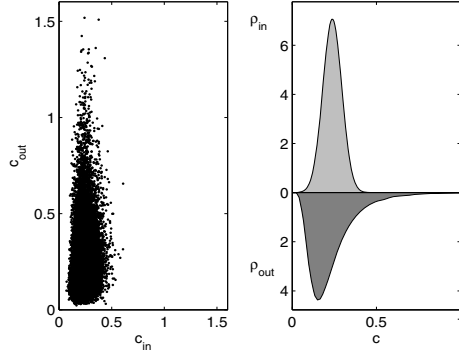


Fig. 12 Left: the distributions for incoming (above, light grey) and outgoing (below, dark grey) information at each EEG electrode displayed on the scalp map (original binning and smoothing spline estimation). Right: the distribution on the scalp of R , the ratio between the standard deviations of the distributions of outgoing and incoming information, for EEG data.

2 Partial Conditioning of Granger Causality

Granger causality has become the method of choice to determine whether and how two time series exert causal influences on each other [23],[13]. 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 [24],[9],[34], and cardiovascular variability [18].

From the beginning [21],[41], it has been known that if two signals are influenced by third one that is not included in the regressions, this leads to spurious causalities, so an extension to the multivariate case is in order. The conditional Granger causality analysis (CGCA) [19] is based on a straightforward expansion of the autoregressive model to a general multivariate case including all measured

variables. CGCA has been proposed to correctly estimate coupling in multivariate data sets [6],[14],[15],[45]. Sometimes though, a fully multivariate approach can lead to problems which can be purely computational but even conceptual: in presence of redundant variables the application of the standard analysis leads to underestimation of causalities [1].

Several approaches have been proposed in order to reduce dimensionality in multivariate sets, relying on generalized variance [6], principal components analysis [45] or Granger causality itself [29].

Here we will address the problem of partial conditioning to a limited subset of variables, in the framework of information theory. Intuitively, one may expect that conditioning on a small number of variables should remove most of the indirect interactions if the connectivity pattern is sparse. We will show that this subgroup of variables might be chosen as the most informative for the driver variable, and describe the application to simulated examples and a real data set.

2.1 Finding the Most Informative Variables

We start by describing the connection between Granger causality and information-theoretic approaches like the transfer entropy in [38]. Let $\{\xi_n\}_{n=1,..,N+m}$ be a time series that may be approximated by a stationary Markov process of order m , i.e. $p(\xi_n|\xi_{n-1},\dots,\xi_{n-m}) = p(\xi_n|\xi_{n-1},\dots,\xi_{n-m-1})$. We will use the shorthand notation $X_i = (\xi_i, \dots, \xi_{i+m-1})^\top$ and $x_i = \xi_{i+m}$, for $i = 1, \dots, N$, and treat these quantities as N realizations of the stochastic variables X and x . The minimizer of the risk functional

$$R[f] = \int dX dx (x - f(X))^2 p(X, x) \quad (3)$$

represents the best estimate of x , given X , and corresponds [32] to the regression function $f^*(X) = \int dx p(x|X)x$. Now, let $\{\eta_n\}_{n=1,..,N+m}$ be another time series of simultaneously acquired quantities, and denote $Y_i = (\eta_i, \dots, \eta_{i+m-1})^\top$. The best estimate of x , given X and Y , is now: $g^*(X, Y) = \int dx p(x|X, Y)x$. If the generalized Markov property holds, i.e.

$$p(x|X, Y) = p(x|X), \quad (4)$$

then $f^*(X) = g^*(X, Y)$ and the knowledge of Y does not improve the prediction of x . Transfer entropy [38] is a measure of the violation of 4: it follows that Granger causality implies non-zero transfer entropy [27]. Under Gaussian assumption it can be shown that Granger causality and transfer entropy are entirely equivalent, and just differ for a factor two [5]. The generalization of Granger causality to a multivariate fashion, described in the following, allows the analysis of dynamical networks [28] and to discern between direct and indirect interactions.

Let us consider n time series $\{x_\alpha(t)\}_{\alpha=1,..,n}$; the state vectors are denoted

$$Y_\alpha(t) = (x_\alpha(t-m), \dots, x_\alpha(t-1)),$$

m being the window length (the choice of m can be done using the standard cross-validation scheme). Let $\varepsilon(x_\alpha|\mathbf{X})$ be the mean squared error prediction of x_α on the basis of all the vectors \mathbf{X} (corresponding to linear regression or non linear regression by the kernel approach described in [27]). The multivariate Granger causality index $c(\beta \rightarrow \alpha)$ is defined as follows: consider the prediction of x_α on the basis of all the variables but X_β and the prediction of x_α using all the variables, then the causality measures the variation of the error in the two conditions, i.e.

$$c(\beta \rightarrow \alpha) = \log \frac{\varepsilon(x_\alpha|\mathbf{X} \setminus X_\beta)}{\varepsilon(x_\alpha|\mathbf{X})}. \quad (5)$$

Note that in [27] a different definition of causality has been used,

$$\delta(\beta \rightarrow \alpha) = \frac{\varepsilon(x_\alpha|\mathbf{X} \setminus X_\beta) - \varepsilon(x_\alpha|\mathbf{X})}{\varepsilon(x_\alpha|\mathbf{X} \setminus X_\beta)}; \quad (6)$$

The two definitions are clearly related by a monotonic transformation:

$$c(\beta \rightarrow \alpha) = -\log[1 - \delta(\beta \rightarrow \alpha)]. \quad (7)$$

Here we first evaluate the causality $\delta(\beta \rightarrow \alpha)$ using the selection of significant eigenvalues described in [27] to address the problem of over-fitting in (6); then we use (7) and express our results in terms of $c(\beta \rightarrow \alpha)$, because it is with this definition that causality is twice the transfer entropy, equal to $I\{x_\alpha; X_\beta | \mathbf{X} \setminus X_\beta\}$, in the Gaussian case [5].

We now address the problem of coping with a large number of variables, when the application of multivariate Granger causality may be questionable or even unfeasible, whilst bivariate analysis would detect also indirect influences. Here we show that conditioning on a small number of variables, chosen as the most informative for the candidate driver variable, is sufficient to remove the biggest portion of indirect interactions for sparse connectivity patterns. Conditioning on a large number of variables requires a high number of samples in order to get reliable results. Reducing the number of variables, that one has to condition over, would thus provide better results for small data-sets. In the general formulation of Granger causality, one has no way to choose this reduced set of variables; on the other hand, in the framework of information theory, it is possible to individuate the most informative variables one by one. Once that it has been demonstrated [5] that Granger causality is equivalent to the information flow between Gaussian variables, partial conditioning becomes possible for Granger causality estimation; to our knowledge this is the first time that such approach is proposed.

Concretely, let us consider the causality $\beta \rightarrow \alpha$; we fix the number of variables, to be used for conditioning, equal to n_d . We denote $\mathbf{Z} = (X_{i_1}, \dots, X_{i_{n_d}})$ the set of the n_d variables, in $\mathbf{X} \setminus X_\beta$, most informative for X_β . In other words, \mathbf{Z} maximizes the mutual information $I\{X_\beta; \mathbf{Z}\}$ among all the subsets \mathbf{Z} of n_d variables. Then, we evaluate the causality

$$c(\beta \rightarrow \alpha) = \log \frac{\varepsilon(x_\alpha | \mathbf{Z})}{\varepsilon(x_\alpha | \mathbf{Z} \cup X_\beta)}. \quad (8)$$

Under the Gaussian assumption, the mutual information $I\{X_\beta; \mathbf{Z}\}$ can be easily evaluated, see [5]. Moreover, instead of searching among all the subsets of n_d variables, we adopt the following approximate strategy. Firstly the mutual information of the driver variable, and each of the other variables, is estimated, in order to choose the first variable of the subset. The second variable of the subsets is selected among the remaining ones, as those that, jointly with the previously chosen variable, maximizes the mutual information with the driver variable. Then, one keeps adding the rest of the variables by iterating this procedure. Calling \mathbf{Z}_{k-1} the selected set of $k-1$ variables, the set \mathbf{Z}_k is obtained adding, to \mathbf{Z}_{k-1} , the variable, among the remaining ones, providing greatest information gain. This is repeated until n_d variables are selected. This greedy algorithm, for the selection of relevant variables, is expected to give good results under the assumption of sparseness of the connectivity.

2.2 Partial Conditioning in a Dynamical Model

Let us consider linear dynamical systems on a lattice of n nodes, with equations, for $i = 1, \dots, n$:

$$x_{i,t} = \sum_{j=1}^n a_{ij} x_{j,t-1} + s \tau_{i,t}, \quad (9)$$

where a 's are the couplings, s is the strength of the noise and τ 's are unit variance i.i.d. Gaussian noise terms. The level of noise determines the minimal amount of samples needed to assess that the structures recovered by the proposed approach are genuine and are not due to randomness, as it happens for the standard Granger causality (see discussions in [27] and [28]); in particular noise should not be too high to obscure deterministic effects.

As an example, we fix $n = 34$ and construct couplings in terms of the well known Zachary data set [44], an undirected network of 34 nodes. We assign a direction to each link, with equal probability, and set a_{ij} equal to 0.015, for each link of the directed graph thus obtained, and zero otherwise. The noise level is set $s = 0.5$. The goal is again to estimate this directed network from the measurements of time series on nodes.

In figure (13) we show the application of the proposed methodology to data sets generated by eqs. (9), in terms of sensitivity and specificity, for different numbers of samples. The bivariate analysis detects several false interactions, however conditioning on a few variables is sufficient to put in evidence just the direct causalities. Due to the sparseness of the underlying graph, we get a result which is very close to the one by the full multivariate analysis; the multivariate analysis here recovers the true network, indeed the number of samples is sufficiently high. In figure (14), concerning the stage of selection of variables upon which conditioning, we plot the mutual information gain Δy as a function of the number of variables included n_d : it decreases as n_d increases.

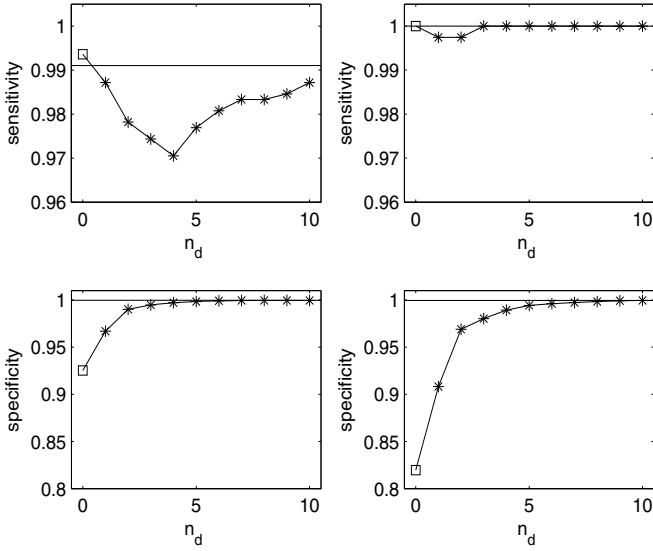
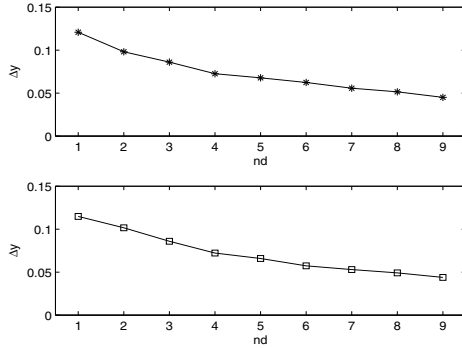


Fig. 13 Sensitivity and specificity for the recovery of the Zachary network structure from the dynamics at his nodes are plotted versus n_d , the number of variables selected for conditioning, for two values of the number of samples N , 500 (left) and 1000 (right). The order is $m = 2$, similar results are obtained varying m . The results are averaged over 100 realizations of the linear dynamical system described in the text. The empty square, in correspondence to $n_d = 0$, is the result from the bivariate analysis. The horizontal line is the outcome from multivariate analysis, where all variables are used for conditioning.

Fig. 14 The mutual information gain Δy for the Zachary network, when the $(n_d + 1)$ -th variable is included, is plotted versus n_d for two values of the number of samples N , 500 (top) and 1000 (bottom). The order is $m = 2$. The information gain is averaged over all the variables.



2.3 *Partial Conditioning in Resting State fMRI*

We used a resting state datasets from a public repository¹. Data were acquired by using of single-shot gradient echo planar imaging (EPI) sequence (repetition time [TR]: 645ms; echo time: 30ms; slices: 33; thickness: 3mm; gap: 0.6mm; field of view: $200 \times 200\text{mm}^2$; in-plane resolution: 64×64 ; flip angle: 90°). Preprocessing of resting-state images was performed using the Statistical Parametric Mapping software (SPM8, <http://www.fil.ion.ucl.ac.uk/spm>), including slice-timing corrected relative to middle axial slice for the temporal difference in acquisition among different slices, realigned with the corresponding 3-D structure image, head motion correction (for all subjects, the translational or rotational parameters of a data set did not exceed $\pm 1\text{mm}$ or $\pm 1^\circ$), spatial normalization into a standard stereotaxic space, parameters from normalizing 3-D structure images to the Montreal Neurological Institute T1 template in SPM8 were written to fMRI images then resampled to 3-mm isotropic voxels. The functional images were segmented into 90 regions of interest (ROIs) using automated anatomical labeling (AAL) template [40]. For each subject, the representative time series of each ROI was obtained by averaging the fMRI time series across all voxels in the ROI. Several procedures were used to remove possible spurious variances from the data through linear regression. These were 1) six head motion parameters obtained in the realigning step, 2) signal from a region in cerebrospinal fluid, 3) signal from a region centered in the white matter. 4) global signal averaged over the whole brain. The hemodynamic response function was deconvolved from the BOLD time series.

In order to select the variables over which conditioning, in figure 15 we plot the mutual information gain for a given target (left posterior cingulate gyrus) as a function of the number of variables included n_d : as expected it decreases as n_d increases. The same behavior is reproduced for all the targets. We can observe that the curve starts to become less steep after $nd = 6$. This phenomenon could be explained considering that multivariate analysis by hierarchical clustering and multidimensional scaling consistently defined six major systems in the resting brain [35]. This is confirmed by looking at figure 16 in which for the same given target the most frequently chosen target variables are reported. It is evident how these are generally sampled at larger scale across the brain in order to pick up information from even distant regions.

3 Informative Clustering

In this last section we propose a formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of each assigned target. Multiplets characterized by an high value will be associated to informational circuits present in the system, with an informational character (synergetic or redundant) which can be associated to the sign of the contribution. We also present results on fMRI and EEG data sets.

¹ http://www.nitrc.org/projects/fcon_1000/

Fig. 15 The mutual information gain when the target is the left posterior cingulate gyrus, when the $(n_d + 1)$ -th variable is included, is plotted versus n_d .

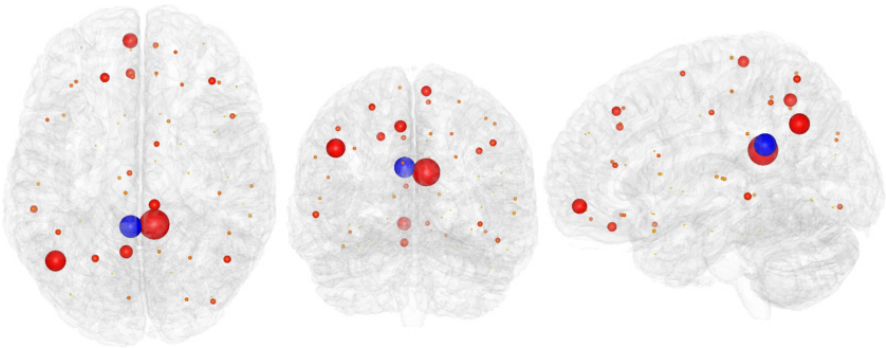
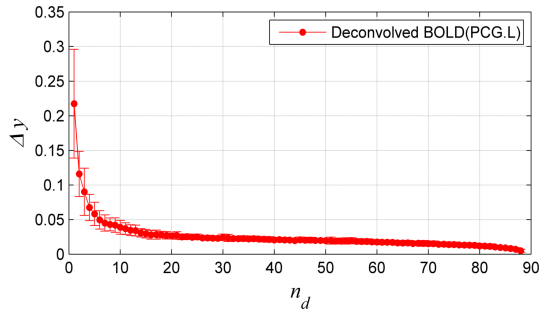


Fig. 16 Variables chosen among the 10 most informative when the target is the left posterior cingulate gyrus (in blue). The diameter of the red spheres is proportional to the times that a region is selected for different subjects.

3.1 Identification of Irreducible Subgraphs

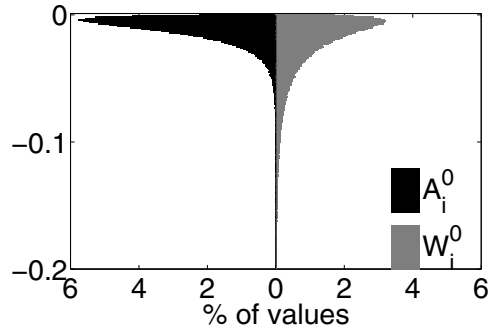
Information theoretic treatment of groups of correlated degrees of freedom can reveal their functional roles as memory structures or those capable of processing information [12]. Information quantities reveal if a group of variables may be mutually redundant or synergetic [37, 7]. The application of these insights to identify functional connectivity structure is a promising line of research. Most approaches for the identification of functional relations among nodes of a complex networks rely on the statistics of motifs, subgraphs of k nodes that appear more abundantly than expected in randomized networks with the same number of nodes and degree of connectivity [30, 42]. An approach to identify functional subgraphs in complex networks, relying on an exact expansion of the mutual information with a group of variables, has been presented in [8].

On the other hand, understanding couplings between dynamical subsystems is a topic of general interest. Transfer entropy [38], which is related to the concept of Granger causality [21], has been proposed to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems. By appropriate conditioning of transition probabilities this quantity has been shown to

be superior to the standard time delayed mutual information, which fails to distinguish information that is actually exchanged from shared information due to common history and input signals. On the other hand, Granger causality formalized the notion that, if the prediction of one time series could be improved by incorporating the knowledge of past values of a second one, then the latter is said to have a *causal* influence on the former. Initially developed for econometric applications, Granger causality has gained popularity also in neuroscience (see, e.g., [9, 39, 16, 27]). A discussion about the practical estimation of information theoretic indexes for signals of limited length can be found in [33].

Here we present a formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of the target. Multiplets characterized by an high value, unjustifiable by chance, will be associated to informational circuits present in the system, with an informational character (synergetic or redundant) which can be associated to the sign of the contribution.

Fig. 17 Concerning fMRI data, the distribution of the first order term in the expansions, eqs. (18) and (13) are depicted.



4 Expansion of the Transfer Entropy

We start describing the work in [8]. Given a stochastic variable X and a family of stochastic variables $\{Y_k\}_{k=1}^n$, the following expansion for the mutual information has been derived there:

$$S(X|\{Y\}) - S(X) = -I(X; \{Y\}) = \sum_i \frac{\Delta S(X)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \dots + \frac{\Delta^n S(X)}{\Delta Y_i \dots \Delta Y_n}, \quad (10)$$

where the variational operators are defined as

$$\frac{\Delta S(X)}{\Delta Y_i} = S(X|Y_i) - S(X) = -I(X; Y_i), \quad (11)$$

$$\frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} = -\frac{\Delta I(X; Y_i)}{\Delta Y_j} = I(X; Y_i) - I(X; Y_i|Y_j), \quad (12)$$

and so on.

Fig. 18 Concerning fMRI data, the distribution of the first order term in the expansion of the transfer entropy, eq. (18), is compared with the results corresponding to a reshuffling of the target time series.

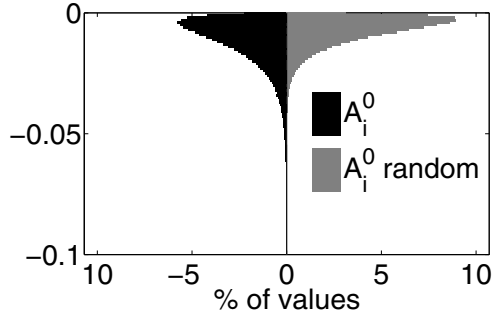
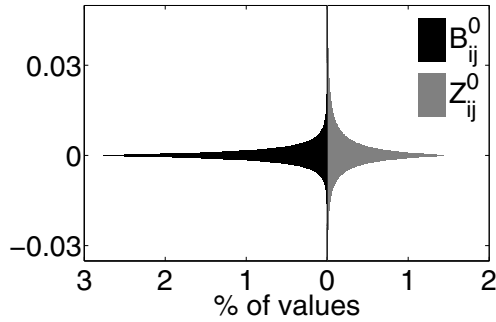


Fig. 19 Concerning fMRI data, the distribution of the second order term in the expansions, eqs. (19) and (14) are depicted.



Now, let us consider $n + 1$ time series $\{x_\alpha(t)\}_{\alpha=0,\dots,n}$. The lagged state vectors are denoted

$$Y_\alpha(t) = (x_\alpha(t-m), \dots, x_\alpha(t-1)),$$

m being the window length.

Firstly we may use the expansion (10) to model the statistical dependencies among the x variables at equal times. We take x_0 as the target time series, and the first terms of the expansion are

$$W_i^0 = -I(x_0; x_i) \quad (13)$$

for the first order;

$$Z_{ij}^0 = I(x_0; x_i) - I(x_0; x_i | x_j) \quad (14)$$

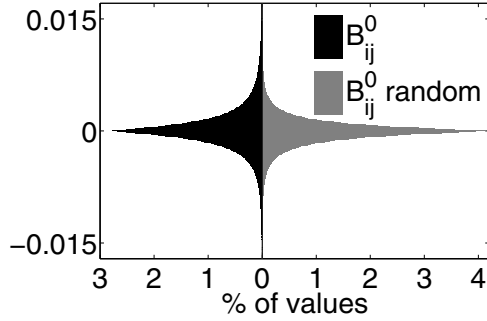
for the second order; and so on. Here we propose to consider also

$$S(x_0 | \{Y_k\}_{k=1}^n) - S(x_0) = -I(x_0; \{Y_k\}_{k=1}^n), \quad (15)$$

which measures to what extent the remaining variables contribute to specifying the future state of x_0 . This quantity can be expanded according to (10):

$$S(x_0 | \{Y_k\}_{k=1}^n) - S(x_0) = \sum_i \frac{\Delta S(x_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0)}{\Delta Y_i \Delta Y_j} + \dots + \frac{\Delta^n S(x_0)}{\Delta Y_i \dots \Delta Y_n}. \quad (16)$$

Fig. 20 Concerning fMRI data, the distribution of the second order term in the expansion of the transfer entropy, eq. (18), is compared with the results corresponding to a reshuffling of the target time series.



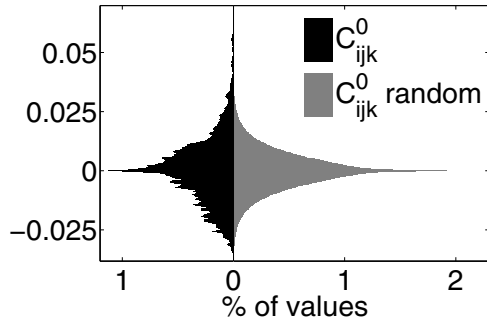
A drawback of the expansion above is that it does not remove shared information due to common history and input signals; therefore we propose to condition on the past of x_0 , i.e. Y_0 . To this aim we introduce the conditioning operator \mathcal{C}_{Y_0} :

$$\mathcal{C}_{Y_0} S(X) = S(X|Y_0),$$

and observe that \mathcal{C}_{Y_0} and the variational operators (11) commute. It follows that we can condition the expansion (16) term by term, thus obtaining

$$\begin{aligned} S(x_0|\{Y_k\}_{k=1}^n, Y_0) - S(x_0|Y_0) &= \\ -I(x_0; \{Y\}_{k=1}^n|Y_0) &= \\ \sum_i \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0|Y_0)}{\Delta Y_i \Delta Y_j} + \dots + \frac{\Delta^n S(x_0|Y_0)}{\Delta Y_i \dots \Delta Y_n}. \end{aligned} \quad (17)$$

Fig. 21 Concerning fMRI data, the distribution of the third order term in the expansion of the transfer entropy, eq. (18), is compared with the results corresponding to a reshuffling of the target time series.



We note that variations at every order in (17) are symmetrical under permutations of the Y_i . Moreover statistical independence among any of the Y_i results in vanishing contribution to that order: each nonvanishing term in this expansion accounts for an irreducible set of variables providing information for the specification of the target. The first order terms in the expansion are given by:

$$A_i^0 = \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} = -I(x_0; Y_i|Y_0), \quad (18)$$

and coincide with the bivariate transfer entropies $i \rightarrow 0$ (times -1). The second order terms are

$$B_{ij}^0 = I(x_0; Y_i|Y_0) - I(x_0; Y_i|Y_j, Y_0), \quad (19)$$

whilst the third order terms are

$$C_{ijk}^0 = I(x_0; Y_i|Y_j, Y_0) + I(x_0; Y_i|Y_k, Y_0) - I(x_0; Y_i|Y_0) - I(x_0; Y_i|Y_j, Y_k, Y_0). \quad (20)$$

An important property of (17) is that the sign of nonvanishing terms reveals the informational character of the corresponding set of variables: a negative sign indicates that the group of variables contribute with more information, than the sum of its subgroups, to the state of the target (synergy), while positive contributions correspond to redundancy.

Another important point that we address here is how get a reliable estimate of conditional mutual information from data. In this work we adopt the assumption of Gaussianity and we use the exact expression that holds in this case [5] and reads as follows. Given multivariate Gaussian random variables X , W and Z , the conditioned mutual information is

$$I(X; W|Z) = \frac{1}{2} \ln \frac{|\Sigma(X|Z)|}{|\Sigma(X|W \oplus Z)|}, \quad (21)$$

where $|\cdot|$ denotes the determinant, and the partial covariance matrix is defined

$$\Sigma(X|Z) = \Sigma(X) - \Sigma(X, Z)\Sigma(Z)^{-1}\Sigma(X, Z)^\top, \quad (22)$$

in terms of the covariance matrix $\Sigma(X)$ and the cross covariance matrix $\Sigma(X, Z)$; the definition of $\Sigma(X|W \oplus Z)$ is analogous.

4.1 Applications: Magnetic Resonance and EEG Data

In order to test this approach on a real neuroimaging dataset we used resting state fMRI data described in the previous section.

For each subject, we evaluated the first terms in the expansions of the conditional mutual information. We then pooled all the values of the terms in the expansions, from all subjects and all targets, and we report their distributions in the following figures. In figure (17) we compare the distributions of A_i^0 , the first order terms in the expansion of the information flow (equivalent to the bivariate transfer entropy), with those of the equal time dependencies W_i^0 . This figure shows that the expansion terms of the mutual information have a quite wide distribution, and also that the maximum of the distribution is not at zero, suggesting that the data set is characterized by many equal time statistical dependencies and by nontrivial causal connections. In figure (18) the distribution of the bivariate transfer entropies is compared with those

Fig. 22 Concerning EEG data, the distribution of the first order term in the expansions, eqs. (18) and (13) are depicted.

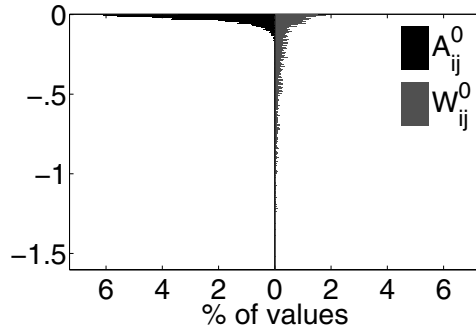
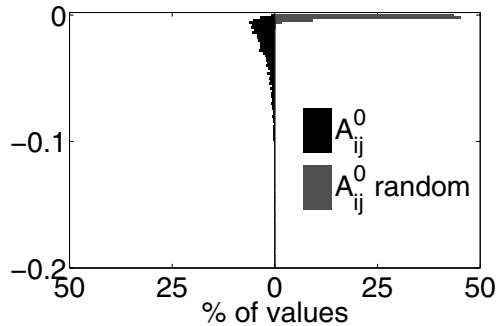


Fig. 23 Concerning EEG data, the distribution of the first order term in the expansion of the transfer entropy, eq. (18), is compared with the results corresponding to a reshuffling of the target time series.

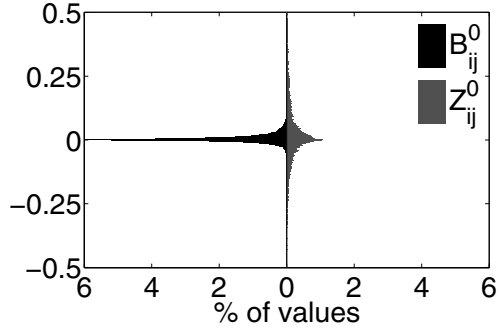


obtained after a random reshuffling of the target time series: the surrogate test at 5% confidence shows that a relevant fraction of bivariate interactions is statistically significant. In figure (19) we report the distributions of the second order terms, both for information flow and for instantaneous correlations: negative and positive terms are present, i.e. both synergetic and redundant circuits of three variables are evidenced by the proposed approach. Some of these interactions are statistically significant, see figure (20).

In figure (21) we report the distribution of the third order terms for the information flow which correspond to the target Posterior cingulate gyrus, a major node within the default mode network (DMN) with high metabolic activity and dense structural connectivity to widespread brain regions, which suggests it has a role as a cortical hub. The region appears to be involved in internally directed thought, for example, memory recollection. We compare the distribution with the corresponding one for shuffled target; it appears that there are significant circuits of four variables, involving Posterior cingulate gyrus, and most of them are redundant.

As another example, we consider electroencephalogram (EEG) data obtained at rest from 10 healthy subjects and described in the first section. In figure (22) we compare the distributions of A_i^0 and W_i^0 . This figure shows that also EEG data are characterized by nontrivial causal connections. In figure (23) the distribution of the bivariate transfer entropies is compared with those obtained after a random reshuffling of the target time series: it shows that a remarkable amount of bivariate

Fig. 24 Concerning EEG data, the distribution of the second order term in the expansions, eqs. (19) and (14) are depicted.



interactions is statistically significant. In figure (24) we report the distributions of the second order terms, both for information flow and for instantaneous correlations.

4.2 Relationship with Information Storage

Information storage is a fundamental aspect of the dynamics of all the processes on complex networks. The full comprehension of the relationship between networks properties and information storage remains a challenge; however some novel insights have been suggested in a recent paper [25], where a connection between information storage and networks motifs has been pointed out. In this subsection we show that the information storage at each node of a network is also connected to the presence of multiplets of variables sending information to that node. Let us consider the following set of three variables, evolving according to:

$$\begin{aligned} x_{t+1} &= cy_t + 0.1 \xi_{t+1}^{(1)} \\ y_{t+1} &= cz_t + 0.1 \xi_{t+1}^{(2)} \\ z_{t+1} &= cx_t + 0.1 \xi_{t+1}^{(2)}, \end{aligned} \quad (23)$$

thus constituting a realization of the network motif (a) in figure 1 of [25]. In figure 25, left we depict, as a function of the coupling c , both the information storage at the node corresponding to the variable x , and the information flow term $\{y, z\} \rightarrow x$. In this case the three variables are redundant and a relation between information storage and information flow can be established. Figures 25, center and right refer to similar dynamical systems of 3 and 4 variables, corresponding to the motifs (c) and (d), respectively, of figure 1 of [25]. These two cases correspond to synergy: still the presence of these informational terms is connected to information storage in the small network. Summarizing, we have shown that the expansion of the transfer entropy is deeply connected with the expansion of the information storage developed in [25], hence the search of redundant and synergetic multiplets of variables, sending information to each given target, will also put in evidence the mechanisms for information storage at that node.

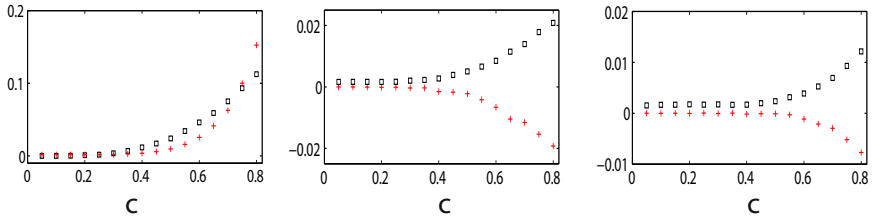


Fig. 25 Information storage (squares) and information flow term $\{y, z\} \rightarrow x$ (crosses) for three motifs described in [25], figure 1. Left: motif (a), redundant variables. Center: motif (c), synergetic variables. Right: motif (d), synergetic variables.

5 Conclusions

The transfer entropy analysis describes the information flow pattern in complex systems in terms of an $N \times N$ matrix, N being the number of subcomponents, each element being the information flowing from each subsystem to each other. The approaches described in the present chapter represent our attempts to deal with physical constraints (e.g., the limited capacity of nodes and the limited number of data samples) within this picture, and to go beyond the $N \times N$ description when the actual senders of information are network motifs rather than single nodes.

Concerning the physical constraints, we have shown that information flow patterns show a signature of the law of diminishing marginal returns and we addressed the problem of partial conditioning to a limited subset of variables.

As far as the search for multiplets of correlated variables is concerned, we have proposed a formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of each assigned target. The applications to real data-set show the effectiveness of the proposed methodology.

References

1. Angelini, L., de Tommaso, M., Marinazzo, D., Nitti, L., Pellicoro, M., Stramaglia, S.: Redundant variables and Granger causality. *Physical Review E* 81(3), 037201 (2010)
2. Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* 286, 509–512 (1999)
3. Barabási, A., Ravasz, E., Vicsek, T.: Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications* 299, 559–564 (2001)
4. Linked, B.A.: *The new science of networks*. Perseus Books, New York (2002)
5. Barnett, L., Barrett, A., Seth, A.: Granger causality and transfer entropy are equivalent for gaussian variables. *Physical Review Letters* 103, 238701 (2009)
6. Barrett, A., Barnett, L., Seth, A.K.: Multivariate Granger causality and generalized variance. *Physical Review E* 81(4), 041907 (2010)
7. Bettencourt, L.M.A., Stephens, G.J., Ham, M.I., Gross, G.W.: Functional structure of cortical neuronal networks grown in vitro. *Phys. Rev. E* 75(2), 21915–21924 (2007)

8. Bettencourt, L.M.A., Gintautas, V., Ham, M.I.: Identification of functional information subgraphs in complex networks. *Phys. Rev. Lett.* 100, 238701–238704 (2008)
9. Blinowska, K., Kusacutė, R., Kaminacuteski, M.: Granger causality and information flow in multivariate processes. *Physical Review E* 70(5), 050902 (2004)
10. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: Structure and dynamics. *Physics Reports* 424, 175–308 (2006)
11. Boccaletti, S., Hwang, D., Chavez, M., Amann, A., Kurths, J., Pecora, L.: Synchronization in dynamical networks: Evolution along commutative graphs. *Physical Review E* 74(1), 016102 (2006)
12. Borst, A., Theunissen, F.E.: Information theory and neural coding. *Nature Neuroscience* 2, 947–957 (1999)
13. Bressler, S.L., Seth, A.K.: Wiener-Granger causality: A well established methodology. *NeuroImage* 58(2), 323–329 (2011)
14. Chen, Y., Bressler, S.L., Ding, M.: Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of Neuroscience Methods* 150(2), 228–237 (2006)
15. Deshpande, G., LaConte, S., James, G.A., Peltier, S., Hu, X.: Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping* 30(4), 1361–1373 (2009)
16. Dhamala, M., Rangarajan, G., Ding, M.: Estimating Granger causality from Fourier and wavelet transforms of time series data. *Phys. Rev. Lett.* 100, 18701–18704 (2008)
17. Erdős, P., Rényi, A.: On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17–61 (1960)
18. Faes, L., Nollo, G., Chon, K.H.: Assessment of Granger causality by nonlinear model identification: Application to short-term cardiovascular variability. *Annals of Biomedical Engineering* 36(3), 381–395 (2008)
19. Geweke, J.F.: Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* 79(388), 907–915 (1984)
20. Ghahramani, Z.: Learning dynamic bayesian networks. In: Giles, C.L., Gori, M. (eds.) *IIASS-EMFCSC-School 1997. LNCS (LNAI)*, vol. 1387, pp. 168–197. Springer, Heidelberg (1998)
21. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3), 424–438 (1969)
22. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C., Weeden, V., Sporns, O.: Mapping the Structural Core of Human Cerebral Cortex. *PLoS Biology* 6(7), e159 (2008)
23. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., Bhattacharya, J.: Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441(1), 1–46 (2007)
24. Kamiński, M., Ding, M., Truccolo, W.A., Bressler, S.L.: Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics* 85(2), 145–157 (2001)
25. Lizier, J.T., Atay, F.M., Jost, J.: Information storage, loop motifs and clustered structure in complex networks. *Physical Review E* 86, 026110 (2012)
26. López, L., Sanjuán, M.: Relation between structure and size in social networks. *Physical Review E* 65, 036107 (2002)
27. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel method for nonlinear Granger causality. *Physical Review Letters* 100, 144103 (2008)
28. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel Granger causality and the analysis of dynamical networks. *Physical Review E* 77, 052615 (2008)

29. Marinazzo, D., Liao, W., Pellicoro, M., Stramaglia, S.: Grouping time series by pairwise measures of redundancy. *Physics Letters A* 374(39), 4040–4044 (2010)
30. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298, 824–827 (2002)
31. Nolte, G., Ziehe, A., Nikulin, V., Schlögl, A., Krämer, N., Brismar, T., Müller, K.: Robustly estimating the flow direction of information in complex physical systems. *Physical Review Letters* 100, 234101 (2008)
32. Papoulis, A.: *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York (1985)
33. Porta, A., Catai, A.M., Takahashi, A.C.M., Magagnin, V., Bassani, T., Tobaldini, E., Montano, N.: Information Transfer through the Spontaneous Baroreflex in Healthy Humans. *Methods of Information in Medicine* 49, 506–510 (2010)
34. Roebroeck, A., Formisano, E., Goebel, R.: Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25(1), 230–242 (2005)
35. Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E.: Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain. *Cerebral cortex* 15(9), 1332–1342 (2005)
36. Samuelson, P., Nordhaus, W.: *Microeconomics*. McGraw-Hill, Oklahoma City (2001)
37. Schneidman, E., Bialek, W., Berry II, M.J.: Synergy, redundancy, and independence in population codes. *J. Neuroscience* 23, 11539–11553 (2003)
38. Schreiber, T.: Measuring information transfer. *Physical Review Letters* 85(2), 461 (2000)
39. Smirnov, D.A., Bezruchko, B.P.: Estimation of interaction strength and direction from short and noisy time series. *Phys. Rev. E* 68, 046209–046218 (2003)
40. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15(1), 273–289 (2002)
41. Wiener, N.: *The theory of prediction*, vol. 1. McGraw-Hill, New York (1996)
42. Yeager-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.J., Alon, U., Margalit, H.: Network motifs in integrated cellular networks of transcription regulation and protein protein interaction. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5934–5939 (2004)
43. Yu, D., Righero, M., Kocarev, L.: Estimating topology of networks. *Physical Review Letters* 97(18), 188701 (2006)
44. Zachary, W.: An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* 33(2), 452–473 (1977)
45. Zhou, Z., Chen, Y., Ding, M., Wright, P., Lu, Z., Liu, Y.: Analyzing brain networks with PCA and conditional Granger causality. *Human Brain Mapping* 30(7), 2197–2206 (2009)
46. <http://clopinet.com/causality/data/nolte/> (accessed July 6, 2012)