A MULTISCALE TOUR IN PROTEIN INTERACTOMICS

By

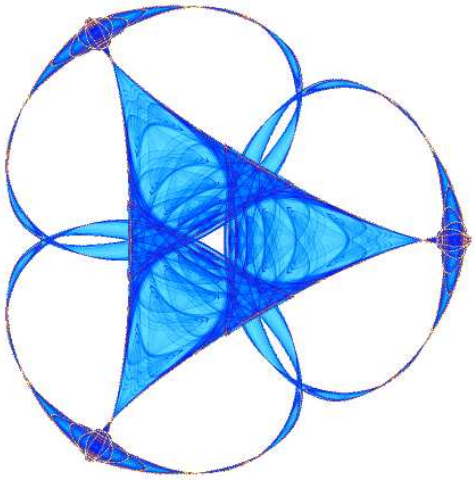
Elisabetta Marras

and

Enrico Capobianco

IMA Preprint Series # 2193

(April 2008)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

A Multiscale Tour in Protein Interactomics

Elisabetta Marras and Enrico Capobianco
CRS4 Bioinformatics Laboratory
Science and Technology Park of Sardinia
Pula (Cagliari) - Sardinia, IT
lisa@crs4.it, ecapob@crs4.it

April 17, 2008

Abstract

For most of the research work currently and extensively undertaken in protein interactomics, the classical barriers are both problems of coverage, i.e. what is really known of most interactomes, and of accuracy, i.e. what is the level of precision we can assign to experimental measurements. However, there are several complex computational tasks that are challenging. We face one of them, which has been left quite unexplored so far. Interesting work has been recently proposed on possible extensions of random networks, by switching from uncorrelated to correlated systems, from static to dynamic features of nodes, from fixed to probabilistic labelling of edges through some kind of hidden variable models. Overall, this shift of complexity might suggest better ways to approach more rigorously crucial aspects such as learning nonlinear maps, the separation of transient from persistent activity dynamics, the design of local versus global stability measures, the definition of new statistical measures describing network topologies. We believe that a particular emphasis should be given to the treatment of multi-scale internal network fluctuations. We assume that protein interaction can respond to dynamics occurring at different resolutions or scales, with natural consequencies for the study of various network characteristics.

1 Introduction

This work deals with protein interactomics [36] with an emphasis on a some new aspects related to the examination of time scale dynamics. As usual, we start from a set of protein interactions taken from different and heterogeneous sources, including experimental ones.

From the recorded measurements, a snapshot of variably connected activity among proteins is observed, regardless the real nature of the links, such that a question remains: are they more or less structurally stable or are they transient?

In other words, the same interaction map very likely cannot be observed if different conditions are considered. Can we account for such dynamical aspects without support of targeted experimental designs delivering the measurements we need?

We believe there might be benefit from looking at state-of-the-art multiscale computational methods, and we propose an analysis based on some instruments which belong to such methods. Wavelets [11, 27, 29] represent a family of orthogonal bases that compress and denoise signals characterized by various possible irregular features observed at different resolutions.

Being this kind of signals very common in many experimental and computational domains, the literature on wavelet applications is huge, especially in statistics [2, 13, 14, 15, 23, 22, 37] and signal processing [9, 28].

We build a statistical setup, methodologically speaking, and start from graphs based on vertices and edges embedding protein interactomes. Thus, the nodes represent proteins and the links represent interactions, the existence of which have usually been seen either experimentally or computationally predicted.

Protein interactomics literature currently offers a wealth of studies on network topology measures. They can be defined at a global level (thus describing overall network characteristics) and at a local level (thus capturing aspects which can be significant for identifiable sub-networks).

In order to deal with this latter level of the analysis, some topological measures have been proposed which can be computed node-wise. This possibility brings advantages, because it suggests that protein-specific characteristics can be illustrated even at a finer detail than a macroscopic level.

This view leads to a map transform from the graph space where measurements are recorded at each node, to a feature space where each topological feature is measured at each node, where it assumes a certain value.

Such values can be transformed again, as we originally suggest in this work, because we assume they might hide some latent information on time scales of interaction dynamics that it would be useful to make clear.

We attempt to untangle the underlying multi-resolutive information, when it is present and might not be immediately visible but only condensed in a single value. Wavelets are a powerful computational tool to enable such disaggregation, and at the same time allow a transform from the feature space to a wavelet coefficient space where to establish new properties of the original protein interactome.

The paper consists of Section 2 on protein interactomics, with datasets and features desription; Section 3 on the methods, with reference to blind source separation and wavelets; Section 4, with the concluding remarks.

2 Protein Interactomics

2.1 Datasets and Features

Protein interactomes represent the complete list of physical interactions mediated by all proteins of an organism [36]. Protein-Protein Interaction Networks (PPIN) can be characterized

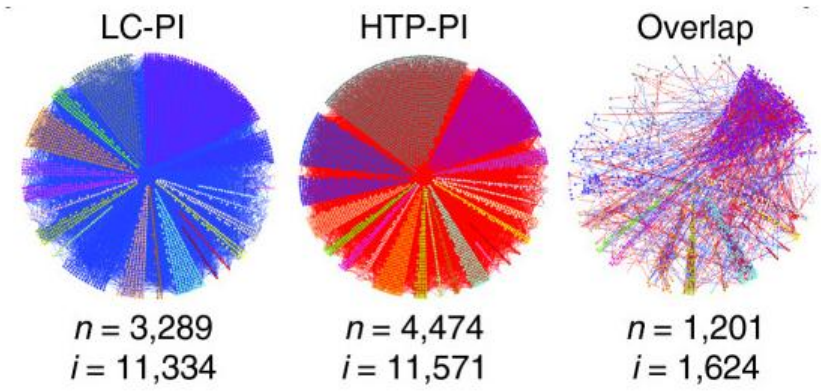


Figure 1: PI overlaps (ref. Reguly et al., *J. Biol.*, 2006)

through a wide variety of measures describing particular aspects according to different types of criteria, such as connectivity, distance, cliquishness, and so on.

Dichotomous PPIN consist of binary interactions (due to the presence of absence of connectivity) whose reliability depends on both coverage and accuracy of data. Thus, the observed links are affected by both false positives (bad measurements) and false negatives (missing information).

The datasets employed in this work are two popular *Yeast* sources. A first consideration is about the nature of such data, as they appear quite different (see Figure 1), thus showing a very marginal overlap. This difference depends on how the datasets have been built, which in turns depends on the kind of measurements obtained.

More specifically, the first dataset is a combination of protein networks constructed from published *Yeast two-hybrid screening* (Y2H) and *Co-Immunoprecipitation* (Co-IP) experiments, thus resulting in a set of 5787 high-confidence high-throughput interactions [3]. We call the dataset Bader's HTP-PI (high-throughput protein interactions).

The second dataset comes from [33], where the measurements are collected in a database of genetic and protein interactions manually curated from 31793 abstracts and online publications, resulting in a set of 31311 interactions. We consider 3289 proteins and 11314 interactions under the definition Reguly's LC-PI (literature-curated protein interactions).

There exist not yet a widespread consensus with regard to the inference methods which fit best complex networks, and there is no any specific method that can be universally adopted, due to the high complexity and dimensionality of PPIN.

A popular approach consists in describing PPIN through topological features (*TOP-features*). They are different from many other biological features which are used to explain protein interactions [26, 32]). The TOP-features represent measures which characterize global (network-wise) and local (node- and edge-wise) aspects of interactions, for instance connectivity, cohesiveness, centrality, correlation etc.

These features are studied in association with more general network properties, such as the scale-free property [4, 38] implying that the connectivity level (or number of links per node) follows a power law (see Figure 2):

$$p(k) = ck^{-\gamma}, c > 0, \gamma > 0 \tag{1}$$

In general, it is observed that there are some nodes in the network that are highly connected, also called sticky proteins, while many others are poorly connected, also called non-sticky proteins. As a result, there is a preferential attachment of new nodes to well-connected hubs.

We show in Figure 2 the effect of sub-sampling the interactome, and how the power law structure adapts to gradually fewer sampled proteins. We have sub-sampled two thirds and one third of the proteins of the interactomes under study, and in each case we have computed the power law resulting from the degree distribution. The similarity of the shape of the power law is quite persistent across different sample sizes.

Only some of the available Top-features are here considered, together with the statistics related to them, such as correlation and other distribution aspects. Each feature represents a summary of relevant biological information in the network, and these summaries may be overlapping to a variable degree.

2.2 Selected Top-features

With undirected graphs connectivity, group cohesiveness and centrality measures can be established by computing degree distribution [16], clustering coefficient [38], and betweenness [5, 17], respectively. These Top-features have been examined in our work. Despite many other are indeed available (see Figure 3 for a sketched overview), we focussed on features that could be computed node-wise in order to have a feature map localized as much as possible.

A natural mapping from an interactome graph to a vector space (of large dimension, in this case) is provided by the adjacency matrix A ($A_{ij} = 1$, iff $\exists$ an edge between i and j nodes, and $A_{ij} = 0$ otherwise). The matrix is symmetric when an undirected graph is considered, as in case of PPIN. In Figure 4, we report a simple example.

Given $G = (V, E)$ network (no self-loops and multiple edges are considered), with an $n-$set of V vertices, and an $m-$set of E edges, we have the following:

- The node v degree distribution $D(v)$ is defined as the number of interacting partners of a protein (with no distinction between in- and out-degree due to the undirect nature of the graph), thus according to:

$$Deg(v) = k \tag{2}$$

  where k is the number of nodes directly connected with v.

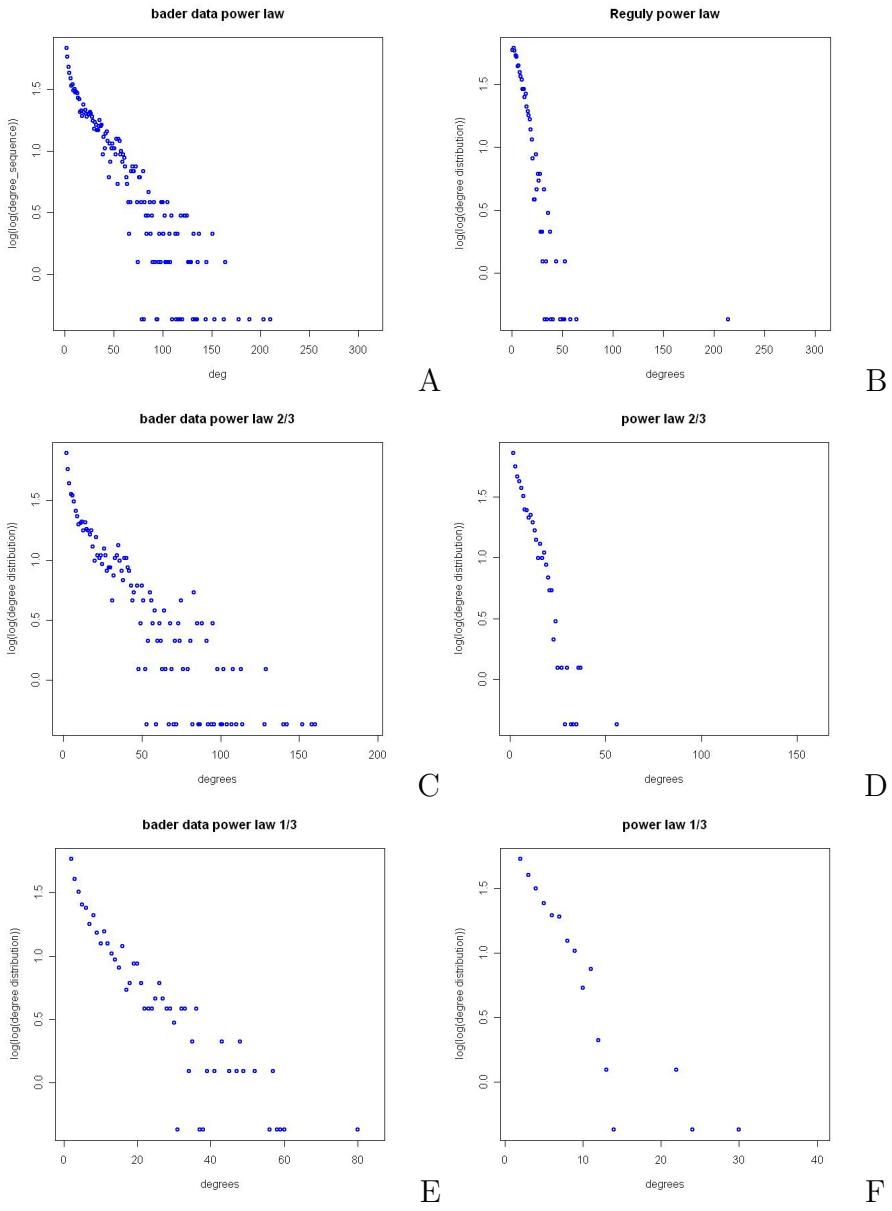


Figure 2: Power Laws (Bader - left, and Reguly - right), under progressive sub-sampling, with two thirds (C,D) and one third (E,F) of the samples.

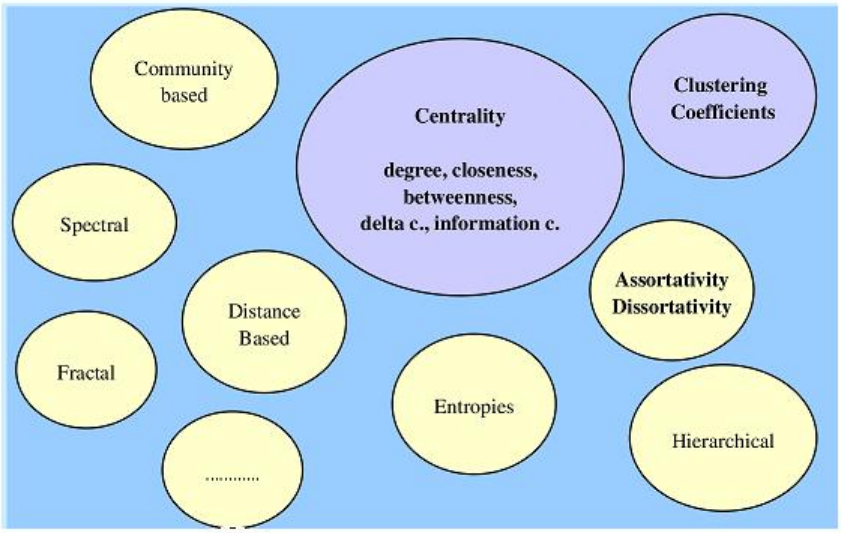


Figure 3: Feature space.

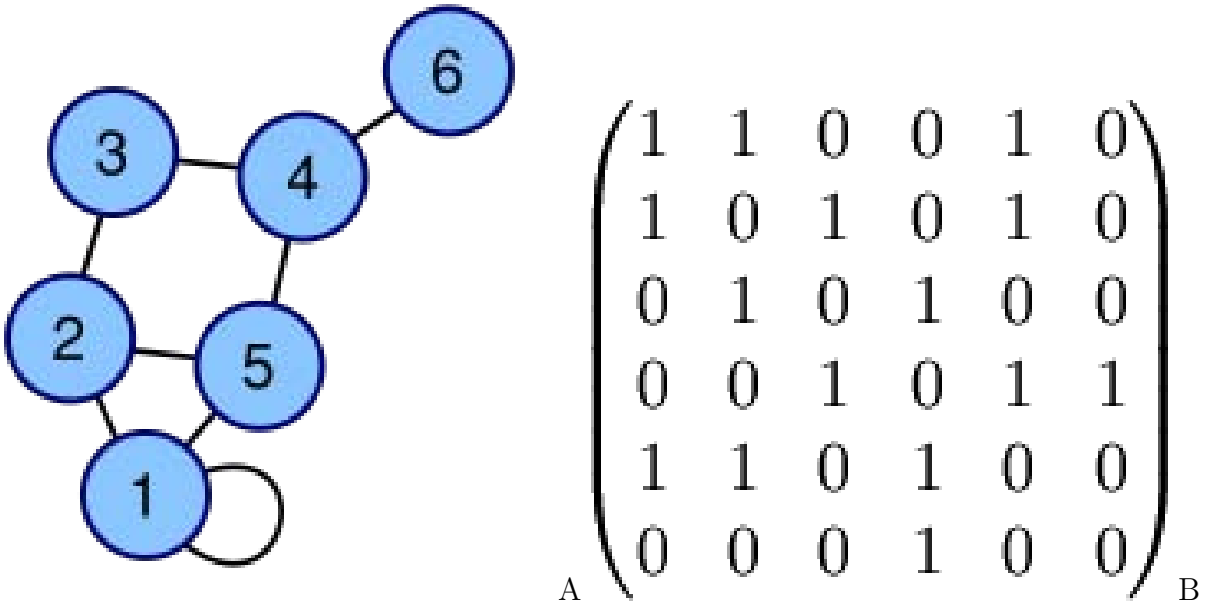


Figure 4: Small undirected graph (A), and corresponding adjacency matrix (B). Example taken from *Wikipedia*.

- Given a node v degree $d(v)$ defined as the number of nodes adjacent to v, the clustering coefficient establishes the likeliness that a link between the nodes a and c exists when both a with b and b with c are linked. Taken from a function in the R package *graph*, for a node v in an undirected graph and with m adjacent nodes, when self loops are not to be included we have:

$$ClustCoef(v) = \frac{N}{(m \times (m-1))} \tag{3}$$

  where N is the number of edges between these nodes. If self loops are instead allowed, the clustering coefficient is $N/(m \times m)$, where N is the number of edges between these nodes, including self loops.

- Betweenness is a centrality measure of a vertex, and is higher for some vertices depending on the fact that they are present on many shortest paths between other vertices. It is computed by a function from the R package *igraph* that defines the number of geodesics (shortest paths) going from an origin to a destination through a vertex relatively to the total number of geodesics observed between start and end node. For vertex v it holds that:

$$Betw(v) = \sum_{s \neq v \neq t \in V, s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4}$$

  for s and t belonging to V.

The three features here presented have some distributional properties, and we can observe the degree with which they relate to each other from Figure 5, and for both datasets. It appears that degree and betweenness are positively correlated while clustering coefficient and betweeness show, marginally at least, negative correlation in Reguly's LC-PI, where in general the Top-feature relationships are more sparsified than in Bader's HTP-PI plots. From the latter we can observe the tendency of Top-features to be negatively correlated.

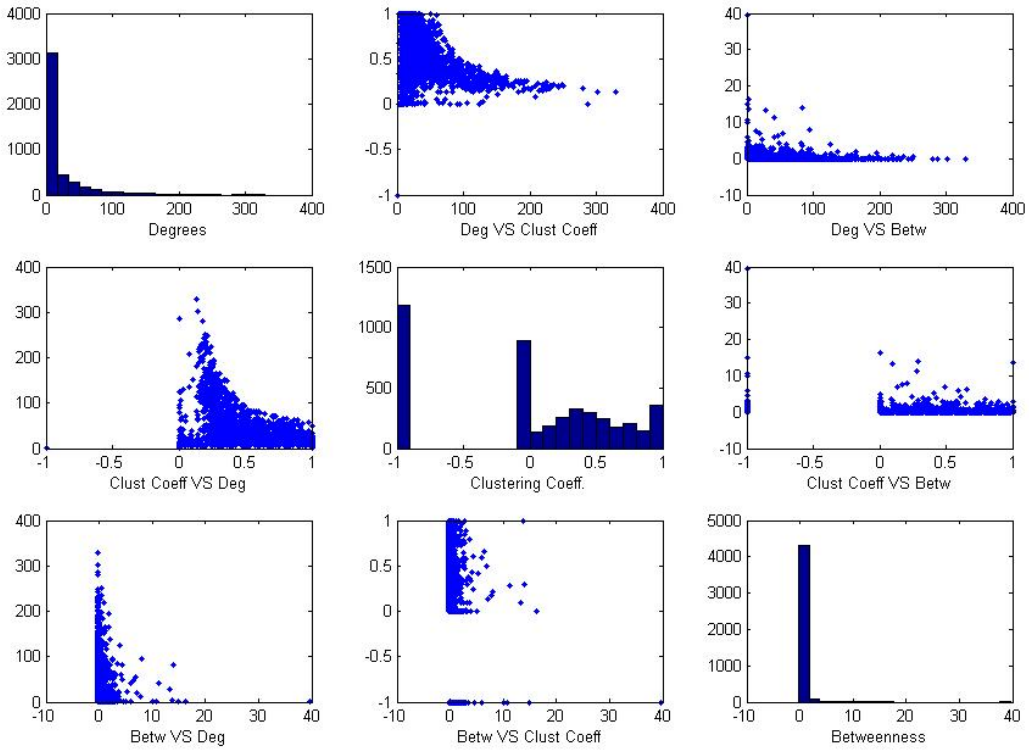
3 Methods

3.1 Blind Source Separation

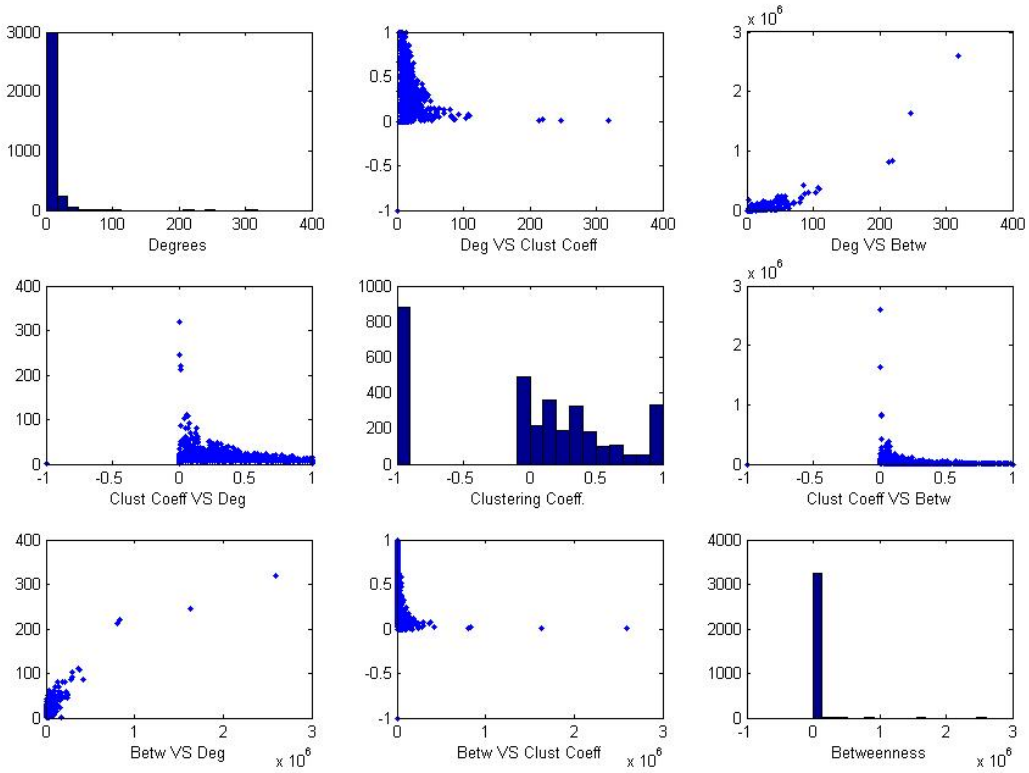
3.1.1 Principal Component Analysis

Usually, when the recorded signals are linearly independent all the available structures in which the signal samples can be separated are needed for a perfect reconstruction. Vice versa, under some degree of linear dependence, the full reconstruction power can be given by much fewer structures.

From a statistical viewpoint, under Gaussianity and orthogonality of the component structures, the decomposition leads to a decorrelated system, i.e. one where the linear dependence carried by the first two distributional moments has been removed. Thus:



A



B

Figure 5: Associated features. Bader's HTP-PI dataset (A) and Reguly's LC-PI dataset (B).

$$COR_Y = COR(y_i, y_j) = COR_{ij} = \lambda_i \delta_{ij} \tag{5}$$

for δ_{ij} (the Kronecker delta) identically equal to 0 (for $i \neq j$) or 1 (for $i = j$).

However, in case of non-Gaussian probabilistic conditions, the decorrelation property holds only approximately. One can miss important structure from higher order dependence, as a substantial loss of information can occur, due to the difference between the estimated and true linear correlation $L_{ij} = COR_{ij} - \lambda_i \delta_{ij} \neq 0$.

When this gap is minimized through a minimal number of components, we may assume to have approached the "intrinsic dimensionality", say D of the system under study, which basically represents a representation in a minimal number of functional approximating structures.

Given a set of observations $x_i \in R^N$, the classical method of *Principal Component Analysis* (PCA) [23] diagonalizes the data covariance matrix $COV = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T$, and especially for low-dimensional data this is performed by solving the eigensystem $\lambda\ v = COV\ v$, for eigenvalues $\lambda \geq 0$ and eigenvectors $v_i \in R^N$.

The latter are mutually orthogonal and define a basis along the directions of maximum variance in the data, while also minimizing the reconstruction error due to a limited number of employed eigenvectors through the corresponding decorrelated PCs. These principal structures lead to the best possible $D-$dimensional subspace because they capture most of the energy.

With orthogonal transforms the resulting basis is given by the *Karhunen-Loeve expansion* [25]; thus, linearity and second order statistical information remain as constraints to the approximation problem.

3.1.2 Independent Component Analysis

Searching for a blind decomposition of complex systems into structures representing the most independent (or least dependent, in many practical cases) coordinates allows to combine the advantage of sparsity with simpler estimation and optimal signal reconstruction power.

Source estimates can be delivered by algorithms that extract any hidden component from possibly convoluted dynamics through the application of *ad hoc* operators.

Independent Component Analysis (ICA) [6, 7, 10] represents a flexible tool for such needs, and deals with linear and nonlinear mixture models, including the case of convolutive mixing. ICA operates in noise-free or noisy applications; when the data dependence structure presents complex features, more elaborated schemes may be needed than the simple ones based on orthogonal decompositions.

The assumptions are usually that $y_i, i = 1, \ldots, n$ are a combination of independent, non-Gaussian and unknown signal sources $s_i, i = 1, \ldots, m$, for $m << n$, and resulting in a system such as $Y = AS$. The $m \times m$ operator A is the mixing matrix, which is combining the unknown sources to be identified in an undetermined way, such that its main parameters have to be estimated too.

The independent components (IC) deliver advantages compared to the more classical PC, due in particular to the superior performance with non-Gaussian signal sources, and to the possibility of embedding higher than second order statistical independence information.

The independent components can be efficiently computed by ad-hoc algorithms such as the *Joint Approximate Diagonalization of Eigenmatrices for Real signals* (*JadeR*) [8] or the *fastICA* [18, 19] deliver an estimate for the separating or de-mixing matrix B; however, through $B = A^{-1}$ one achieves only ideally source separation, as the solutions hold approximately up to permutation and scaling matrices, due to some degree of indeterminacy inherent to the system. In real applications, a set of m approximately independent components is obtained.

In conclusion, the de-correlation and rotation steps which have to be implemented will deliver a set of m approximately independent components efficiently computed by *ad-hoc* algorithms. The IC are exactly the known PC, for Gaussian signals; with non-Gaussian signals a superior performance is obtained through them, when they exist and can be identified, due to the high order statistical independence information they allow to capture.

Figure 6 is interesting. To reveal hidden relationships between our selected variables, we performed PCA over two and then three Top-features, all we are currently using. Figures 6.A and 6.B show the resulting plots: histograms of PC and scatter plot depicting their mutual correlation.

In figure 6.A we see that the two PC coming from the analysis over degree and clustering coefficient seem to be highly correlated, which is natural, since they are Top-features that hold more or less the same content, i.e. high degree often means high clustering coefficient, and vice versa.

However, by including one more feature (betweenness), the previous relationships change, thus resulting in a deep change in the correlation structure, as we see in Figure 6.B where now degree and betweenness result positively correlated. Thus, given the three Top-features, performing PCA causes a variable correlation degree according to the number of features involved.

With the aim to see more in depth and catch all possible information concerning relationships between Top-features, we decomposed each feature in resolution levels using a wavelet procedure.

When resolutions are considered via wavelets (defined and explained next) added to PCA, the correlation structure level by level is shown in Figure 6.C, where we see that there is no substantial change in such structure across levels, being clear the high correlation between degree and clustering coefficient.

In Figure 7 we repeat the same analysis for Reguly's LC-PI. Degree and clustering coefficients appear (from plot A.) highly correlated in their extracted PC, while the addition of betweenness (from plot B) from one hand changes the sign of the previous correlation, and also brings in new marginal correlations between the first two Top-features and the new one.

When resolution levels are considered (plot C), interestingly it appears that there is a decrease of correlation between degree and clustering coefficient compared to the HTP-PI case, while there is a marginal positive correlation appearing from the relationships between degree and betweenness, unlike with HTP-PI where no correlation had been seen.

A *multiscale PCA* in Figure 8 combines the PCA with a decomposition by resolution levels (of variable grid) obtained by wavelets, and the results we show refer to the most significant PC which are selected. Compared to the plain PCA, it appears from the top-left plot that the HTP-PI with two components has a weaker correlation degree, and the same can be said even for the extended case at bottom-left plot (i.e. with betweenness). For the LC-PI case

the top-rigth plot with two components has a reversed in sign and more marginally sized correlation degree, also confirmed in the global setting shown at the bottom-right plot.

To deal with deviations from the Gaussian assumption of dependence between components, we used Independent Component Analysis over both Reguly's LC-PI and Bader's HTP-PI, using *fastICA* package for *Matlab*.

In Figure 9, the ICA plots are reported, and we see the results for Bader's HTP-PI and for Reguly's LC-PI. It is clear that each independent component found by the algorithm (second column, 1st, 3rd and 5th row) corresponds almost perfectly with each single feature (1st column, 1st 3rd and 5th row). The *boxplots* show in both Reguly's and Bader's cases a strong presence of outliers for degree and betweenness, while for clustering coefficient they are almost absent.

In other terms, an excess of kurtosis occurs with HTP-PI, with a strong positive tail for the degree, a more moderate one for the betweenness and a more balanced but still non-Gaussian distribution for the clustering coefficient. From the LC-PI the major difference comes from one Top-feature, the betweenness, but what really changes is the sign of the tail, negative this time.

Moreover, from the *QQplots* we measure departure from normality, and we can see that deviation is pretty strong in the case of clustering coefficient, especially along the tails (for both datasets), while for degree and betweenness it is less evident. The indication from fat-tailed *QQplot*s are that for HTP-PI both degree and centrality show a solid presence of highly connected nodes, i.e. hubs, with quite strong traffic crossing them.

The LC-PI dataset suggests that this traffic is also influenced by less important nodes, indeed. It would be of big interest to see which proteins correspond to such divergent tails and outliers, so that their identification can lead to a possible association with special groups of proteins, functional complexes, and so on (ongoing work).

3.2 Wavelets

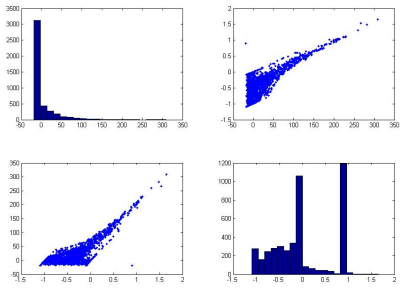
3.2.1 Wavelet Approximation

In general, linear approximation systems are often sub-optimal, due mainly to the functional complexity involved in any cases. Thus, instead of following the rule of selecting N approximating terms, one may prefer to adhere to adaptive criteria and nonlinear schemes [12].

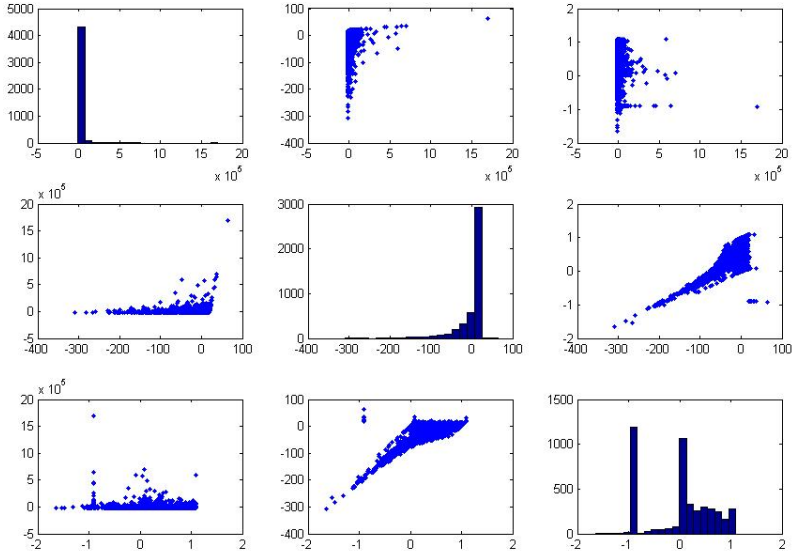
It is usually the case in many applications that only a few structures, such as expansion coefficients, can be used to reconstruct the signal at hand. Usually, under complex dynamics, the employed function estimator needs non-linear selection rules rather than linear ones, and the expansion of inhomogeneous functions may lead to significantly large coefficients at several resolution levels.

Suitable thresholding makes such estimators effective in practice, leading to a selection of a subset of the computed coefficients, just those which are large in absolute value.

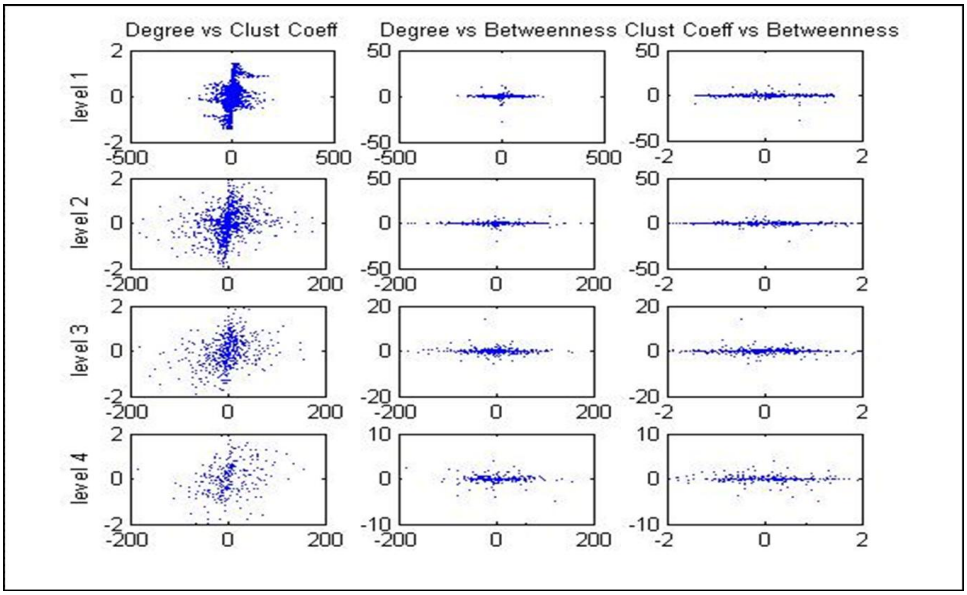
It is well-known that non-linear wavelet estimation attains the usual near-optimal minimax rate of L_2 convergence in a large scale of Besov classes [29, 34]. The *Wavelet Shrinkage Principle* [13, 14, 15] applies a thresholding strategy which yields de-noising of the observed data; it operates by shrinking wavelets coefficients toward 0 so that a limited number of



A

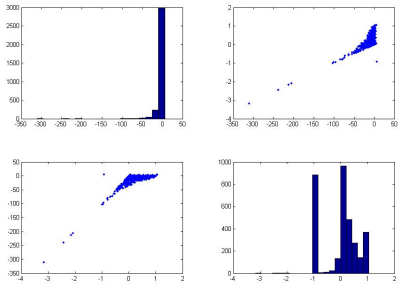


B

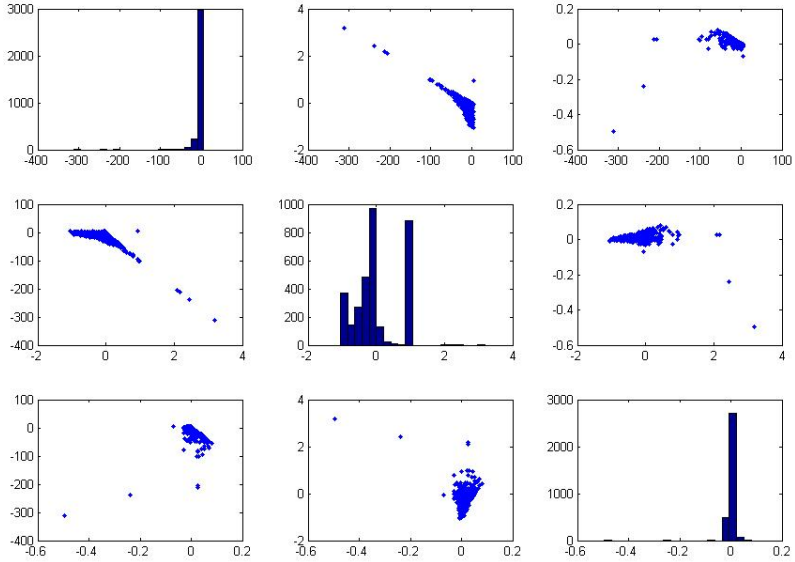


C

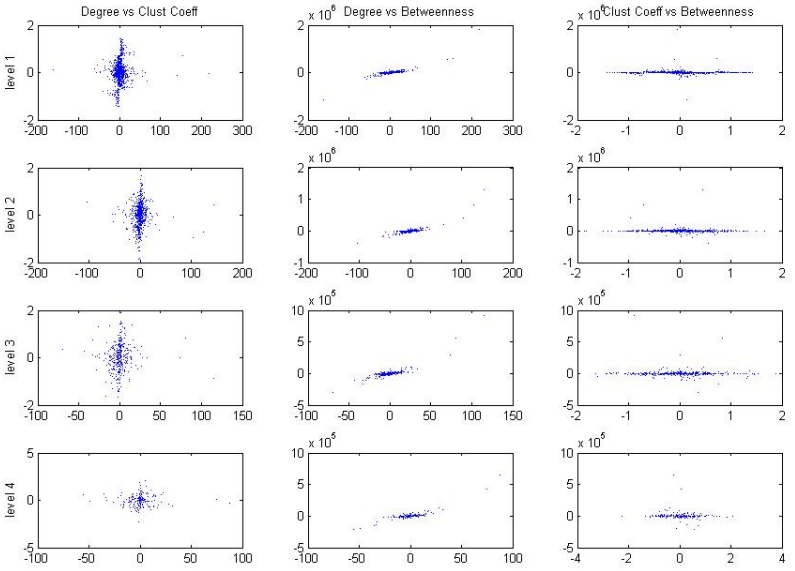
Figure 6: Feature PCA with 2 components, degree and clustering coeff. (A) and also with betweenness (B). Wavelet resolution-wise scatters of associated features (C). Bader's HTP-PI.



A



B



C

Figure 7: Feature PCA with 2 components, degree and clustering coeff. (A) and also with betweenness (B). Wavelet resolution-wise scatters of associated features (C). Reguly's LC-PI.

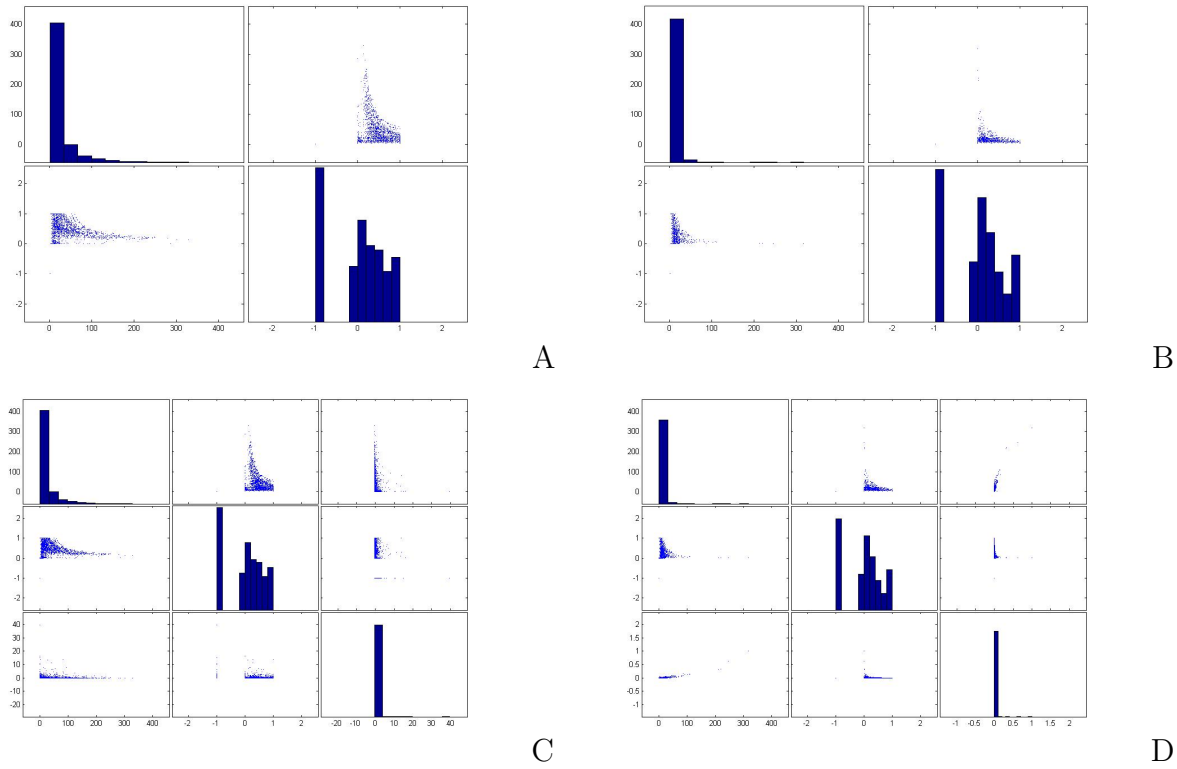


Figure 8: Feature multiscale PCA with 2 and 3 Top-features.
The extracted components refer to Bader's HTP-PI (left side, A,B) and Reguly's LC-PI (right side, C,D).
More sparsification occurs with LC-PI, but similar histogram shapes suggest strong Top-feature overlap.

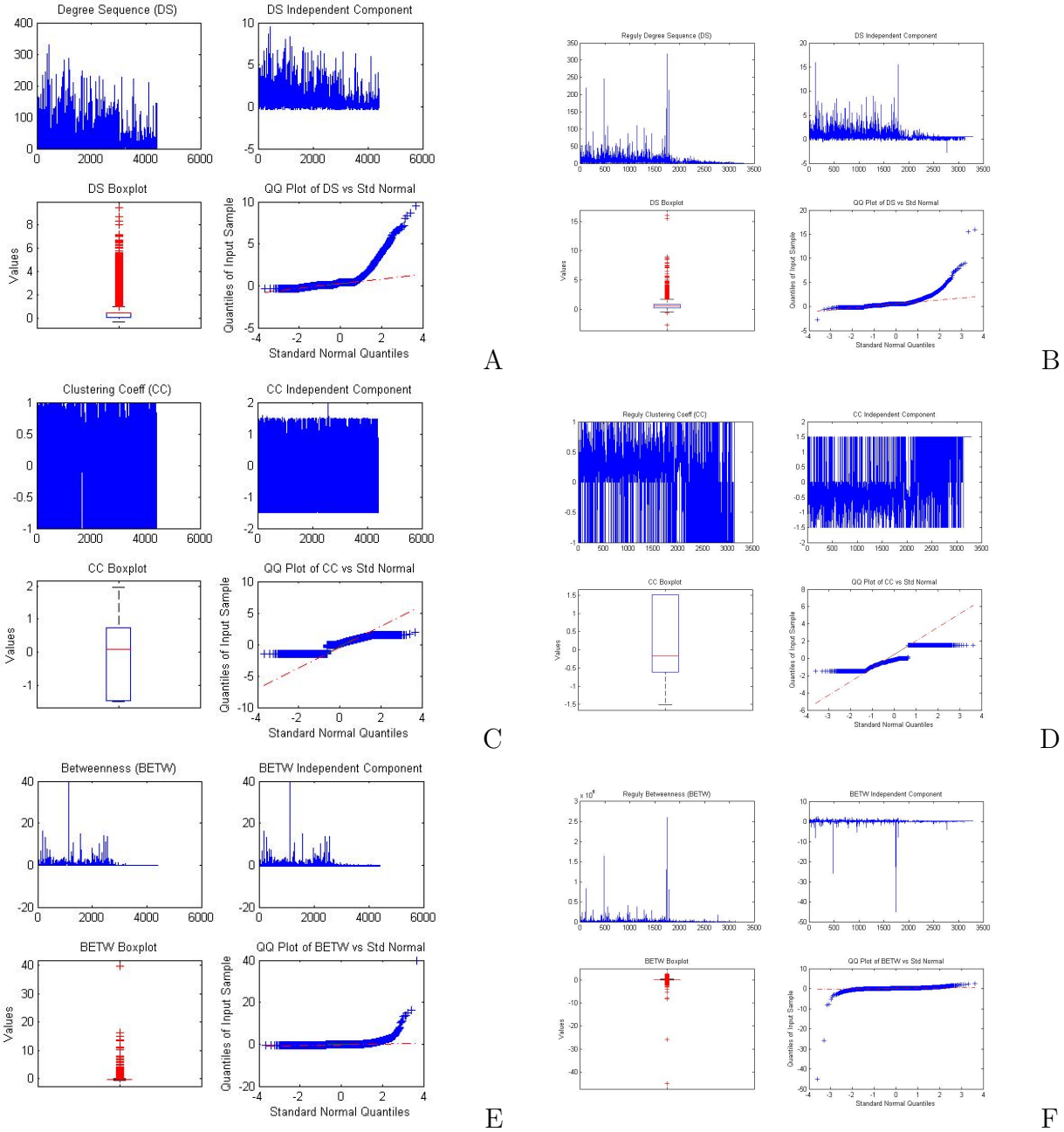


Figure 9: Feature diagnostics with ICA (Bader - left, Reguly - right, datasets). Degree (A-B), Clustering Coefficient (C-D), and Betweenness (E-F) statistics.

them will be considered for reconstructing the signal.

In Besov spaces, endowed with features typical of spatially uniform smoothness classes, one can study functions with sparse singularities, such as locally spiky and jumpy functions. This aspect makes Besov classes adapt well to spatial inhomogeneity and become the most convenient setting for functional space estimation with wavelet methods.

In particular, the space norm is equivalent to the norm of the sequence space of coefficients associated with regular wavelet bases. In turn, the above characterization allows for the construction of minimax thresholding estimators, also adaptive (for instance, relatively to unobserved variance and non-Gaussian signals).

In conclusion, spatial adaptivity and non-linear approximation via shrinking algorithms emphasize the role of wavelets; even if the related models are linear, shrinkage estimators are non-linear and in general the non-linear algorithms involved must be preferred to the linear ones when dealing with function spaces of minimal smoothness and when the functional optimization criteria can be represented by multiscale expansions.

3.2.2 Wavelet Representations

A well-known orthogonal basis expansion is obtained byte discrete wavelet transform WT^d, by which a map $f \rightarrow w$ is implemented via a bank of quadrature mirror filters [28], by $w = W^d f$, and coefficients at high/low scales (with high and low frequency content, respectively) are obtained.

If an orthonormal wavelet basis is used, such as daublets, symmlets or coiflets [11], then $y = f + \xi$ becomes transformed in:

$$W^d y = W^d f + W^d \xi \equiv g + \eta \tag{6}$$

This transformation preserves Gaussianity (as from the noise ξ) and produces decorrelation for autocorrelated systems [22].

An extension is non-orthogonal non-decimated wavelet transform WT^u, i.e. a conservative transform for which the expansion coefficients are not eliminated while obtaining them resolution-wise, unlike with transforms where the decimation occurs when changing scale. It is characterized by a matrix W^u of size $\bar{N} \times N$, for $\bar{N} \geq N$, and a redundant system is found, together with a pseudo-inverse transform W^{u-}, such that $W^{u-} W^u = I$.

Now for $y = f + z$, WT^u decomposes as:

$$W^u y = W^u f + W^u z \equiv h + \epsilon \tag{7}$$

While the Gaussian property is still preserved, i.e. $\epsilon \sim N(0, \sigma)$, with $\sigma = W^u W^{u'}$, there is this time structure left in $\sigma = W^u W^{u'} \neq I$, combined with heteroscedasticity when σ is time-dependent. As a result, almost decorrelation occurs in the systematic component of the model, as in those cases characterized by long range dependent and self-similar processes [35], or by fractional Brownian Motion [23, 37].

3.2.3 Wavelet Series Expansion

There are various wavelet families that can be used to approximate many types of functions that when transformed assume a sparser or simplified structure.

Wavelets refer to a set of functions generated by dilation and translation of a compactly supported scaling function (or father wavelet) and a mother wavelet, ϕ and ψ, respectively. associated with a multi-resolutive analysis of $L_2(\mathcal{R})$.

Multi-resolution techniques represent both adaptive and time-frequency localized solutions, deal with non-linear complex dynamics and non-stationary systems, and have strong computational and theoretical motivation.

Generally speaking, with a WT^d a sequence of smoothed signals and of details giving information at finer resolution levels is found and may be used to represent a signal expansion:

$$f(x) = \sum_k c_{j_0 k}\phi_{j_0 k}(x) + \sum_{j>j_0}\sum_k d_{jk}\psi_{jk}(x) \tag{8}$$

where $\phi_{j_0 k}$ is associated with the corresponding coarse resolution coefficients $c_{j_0 k}$ and d_{jk} are the detail coefficients, i.e. $c_{jk} = \int f(x)\phi_{jk}(x)dx$ and $d_{jk} = \int f(x)\psi_{jk}(x)dx$.

In short, the first term of the right hand side of Eq. (8) is the projection of f onto the coarse approximating space V_{j_0} while the second term represents the cumulated details.

With f the function of interest, for instance assigning a value to a certain feature characterizing the protein interactome, clearly enough, $w = [c_{j0,k}, d_{j,k}]$ is the coefficient vector of the same dimension of the original vector of function values computed at the sampled points, i.e. $f = [f(t_1), f(t_2), \ldots, f(t_n)]$, and with scaling and wavelet coefficient representing them, respectively, at coarse and fine resolutions.

Because of the orthogonality of W, the inverse transform (IDWT) is given by $f = W^T w$, thus by simply a transpose operation computed by fast algorithms.

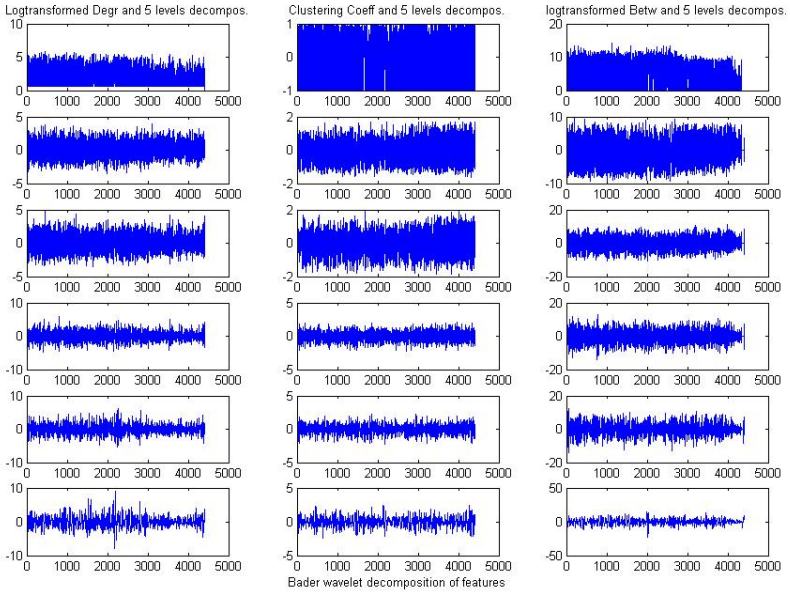
Figure 10 describes the wavelet decomposition for the three Top-features applied to the datasets. Thus, a further step in pursuing the discovery of information hidden in our data is to operate a distinction between coarse and fine scale information, by means of wavelets. The use of the symmlets for the features decomposition on both Bader's HTP-PI and Reguly's LC-PI leads to a comparison feature-wise, not only globally but also level by level.

As we can see, it is possible to individuate levels that are similar and group them together: for example, in Figure 9.A the five levels of the degree decomposition can be divided in a first group containing the first two, and a second group that contains the remaining three. Carefully looking at the plots it is possible to discover differences between levels that need further investigation (currently under development).

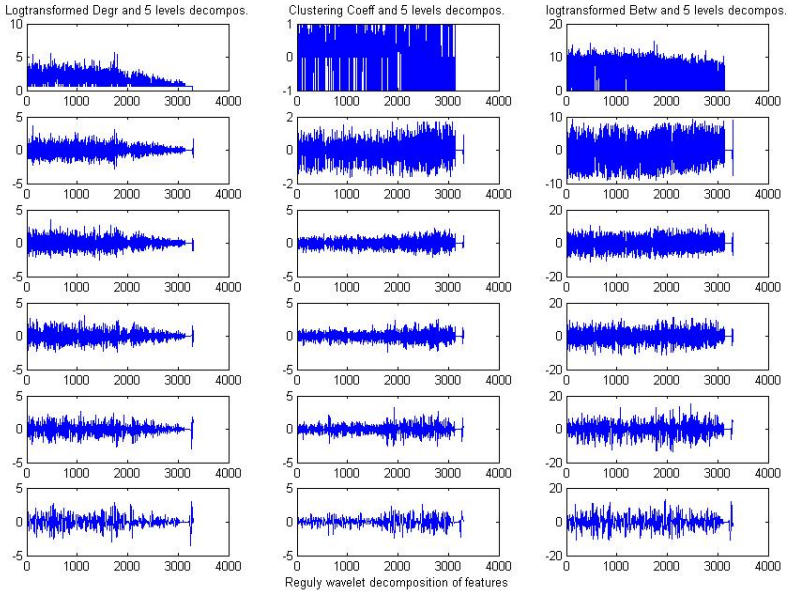
3.2.4 Decorrelation Property

The covariance function of the wavelet transforms is controlled by a parameter, M, which allows wavelets to be classified within a family. This parameter is called the number of vanishing moments, which represents an extra set of mathematical relationships for the coefficients that must be satisfied, and therefore is directly related to the number of coefficients

Note that ψ has vanishing moments of order M if $\int_{-\infty}^{\infty}\psi(\tau)\tau^l d\tau = 0$, for $l = 0, 1, \ldots, M-1$, with $\int_{-\infty}^{\infty}\psi(\tau)\tau^M d\tau \neq 0$. If this number is high, more compression power and fast memory



A



B

Figure 10: Wavelet decompositions: Bader's HTP-PIs (A) and Reguly's LC-PI (B).

decay are achieved.

The key role of the vanishing moments is due to the fact that a kernel expansion of integral operators in a wavelet basis allows for a faster decay rate compared to the original kernel. More vanishing moments lead to increased diagonalization power and robustness to non-stationarity while decorrelating.

The wavelet transforms bring decorrelation for long memory processes due to the control of non-stationarity and dependence through M. For such processes the wavelet basis captures dependence leaving an almost uncorrelated sequence of expansion coefficients.

Thus, when decomposing a process X^t into a sequence of wavelet details $\{d^X_{j,k}\}_{k\in Z}$, the latter behave as a stationary process, and if the number of vanishing moments M satisfies the given constraints, the variance of $d^X_{j,k}$ shows scaling behaviour in a range $j_1 \leq j \leq j_2$[1].

3.2.5 Multi-resolution Analysis

A so-called *Multi-resolution Analysis* (MRA) [11] is obtained when a sequence of closed subspaces satisfying $\ldots, V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \ldots$, with $\bar{\cup}_{j\in Z} V_j = L_2(R)$, $\cap_{j\in Z} V_j = \{0\}$ and the additional condition $f \in V_j \Longleftrightarrow f(2^j.) \in V_0$. The last condition is a necessary requirement for identifying the MRA, meaning that all the spaces are scaled versions of a central space, V_0.

An MRA approximates $L_2[0,1]$ through V_j generated by orthonormal scaling functions ϕ_{jk}, where $k = 0, \ldots, 2^j - 1$. These functions allow also for the sequence of 2^j wavelets ψ_{jk}, $k = 0 \ldots, 2^j - 1$ to represent an orthonormal basis of $L_2[0,1]$.

Signal decompositions with the MRA property have also near-optimal properties in a quite wide range of inhomogeneous function spaces, Sobolev, Holder, for instance, and in general all Besov and Triebel spaces [29].

General function estimators $\hat{f}(t) = \frac{1}{n}\sum_{i=1}^n K(t, t_i)$ can be proposed with a kernel $K(.)$ symmetric positive definite as in the RKHS setting [20, 30, 31]. Correspondingly, an MRA [27] is obtained when the projection kernel in $L_2(R)$ operates through a sequence of nested spaces $\ldots, V_{j-1}, V_j, V_{j+1}, \ldots, j \in Z$ that approximate the underlying space:

$$P_V f(t) = \int_{-\infty}^{\infty} K_\nu(t,s) f(s) ds \tag{9}$$

$$K_\nu(t,s) = \frac{1}{\nu} K(\frac{t}{\nu}, \frac{s}{\nu}) \tag{10}$$

$$K(t,s) = \sum_{k\in Z} \phi(t-k)\phi(s-k), \quad \nu = 2^{-j} \;\; k, j \in Z \tag{11}$$

for ϕ a scaling function such that the $k-$translated sequence ϕ_k forms orthonormal bases for the space at $j = 0$, i.e. V_0. Being W_j the orthogonal complement of V_j in V_{j-1}, the same principles apply to the W_j spaces with the wavelet family ψ and their derived functions ψ_{jk}.

Thus, V_j is a RKHS with a unique reproducing kernel $K_\nu(t,s)$; the order of reproducibility depends on the MRA property rather than on the choice of the wavelet basis [20].

[1] See [35] for theoretical aspects.

3.2.6 Wavelet Denoising

The purpose of denoising is to suppress the noise from the observed signal, and help the recovery of functions of that signal. In statistical terms, this corresponds to a non-parametric regression, where an orthogonal basis expansion is used to estimate the unknown function in a time regression setting.

The majority of methods are optimally applied under standard assumptions (normally distributed independent data, dyadic sample size, equal observational spacing, fixed sample points, stationarity in covariance). Often such conditions do not apply to the data at hand, and modified wavelet expansions have been suggested and implemented in response to this problem.

Very popular techniques are based on wavelet shrinkage and wavelet thresholding. The computational advantage of using such estimation approaches is provided by algorithms of fast implementation. The main idea behind is simply that wavelet coefficients' estimates which are bigger in absolute value than a certain computed threshold (fixed by a predetermined value, otherwise adaptively calculated), then the same value is either retained as it is or diminished by the amount corresponding to the threshold. The smaller coefficients are instead eliminated, or shrunken, thus sparsifying the wavelet expansion.

The simplest way to introduce this topic is considering a signal corrupted by noise:

$$y(t_i) = f(t_i) + \sigma\epsilon(t_i) \tag{12}$$

The goal is recovering the underlying and unknown function f measured at time t_i from the noisy data y_i. The disturbance is delivered by the ϵ_i Gaussian errors, i.e. $\epsilon \sim n.i.d(0,1)$.

Wavelets provide a good expansion basis for the unknown function, because it can satisfy both parsimony and sparsity properties; while the latter comes from the inherent distribution of the "energy" of the function over the coefficient vector, thus naturally selecting a subset of them as the most significant, the latter property strengthen this concept by assuming that only a few coefficients $d_{j,k}$ carry information about the unknown function, while the rest are not different from noise.

Denoising steps can be sketched as $DWT \Rightarrow d \Rightarrow \hat{d} \Rightarrow IDWT$, where $\hat{d}$ is transformed according to various possible linear or nonlinear rules. Both hard (discontinuous function) and soft (continuous function) thresholding rules have been tried out in our work, following various schemes and also the rules reported below:

$$\begin{aligned} T_\lambda^{hard}[\hat{d}_{j,k}] &= \hat{d}_{j,k},\ if\ \mid \hat{d}_{j,k} \mid > \lambda & (13)\\ &= 0\ , if\ \mid \hat{d}_{j,k} \mid \leq \lambda & (14) \end{aligned}$$

$$\begin{aligned} T_\lambda^{soft}[\hat{d}_{j,k}] &= \hat{d}_{j,k} - \lambda,\ if\ \mid \hat{d}_{j,k} \mid > \lambda & (15)\\ &= \hat{d}_{j,k} + \lambda,\ if\ \mid \hat{d}_{j,k} \mid \leq \lambda & (16)\\ &= 0,\ if\ \mid \hat{d}_{j,k} \mid < \lambda & (17) \end{aligned}$$

Figure 11 and Figure 12 for Bader's HTP-PI dataset, and Figure 13 and Figure 14 for Reguly's LC-PI dataset report the wavelet reconstructions applied resolution-wise both without

and with denoising applied to the wavelet coefficients obtained for each Top-feature. Figure 11 and Figure 13 also emphasize the reconstruction power of just one resolution level, the first one.

After *hard minimax*[2] thresholding over all resolution levels, reconstruction from each level shows that the pattern resulting from each reconstructed level does not seem considerably different from each other, even though a relevant part of noise has been effectively removed.

For this lack of power of levels in discriminating across levels themselves, there could be a number of possible reasons: one is that the features we are using probably are structured so that after thresholding and denoising what remains is something in common, say for example hubs, which are parts with high connectivity, cohesiveness and centrality, even though there are slight differences between levels that plots are not able to show.

To be able to draw more important conclusions, it is useful to look at the *boxplots* and *QQplots* before and after thresholding: they reveal especially in some cases, changes occurred in between the two steps of decomposition and denoising (see our next set of figures). The analysis of such differences is extremely promising and it is object of further development and research.

4 Diagnostics and Concluding Remarks

Other plots are now reported with statistical diagnostic analysis about the two datasets. Figure 15 looks at the clustering coefficient (first level only) for both datasets and also with denoised wavelet coefficients. The distributional properties of the reconstructed Top-feature are thus highlighted before and after denoising.

It appears that a more regular histogram is observed with Bader's HTP-PI compared to Reguly's LC-PI, but nevertheless receives a higher impact from denoising due most likely from the fact that indeed there is a stronger presence of noisy interactions. The skewness which is evident from the Reguly's LC-PI is noticed from histograms and *boxplots*, in particular.

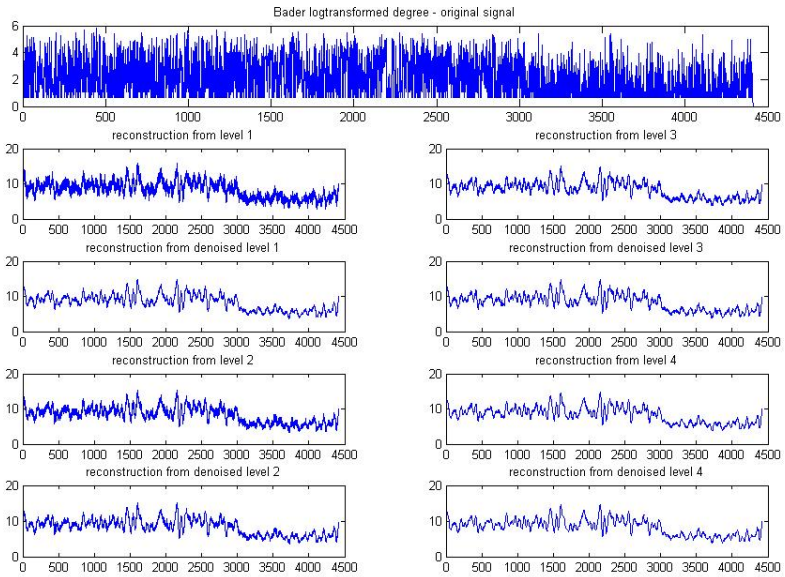
A biological interpretation, for this Top-feature behavior might be that the cohesiveness related to the HTP-PI and LC-PI is quite homogeneously shrunken in both circumstances, but more heavily for the former case. We surely expect many false positive cases especially in HTP-PI contexts, and their elimination necessarily breaks several clusters.

The following plots are instead reporting a more extensive diagnostic evidence concerning the two datasets, thus examining clustering coefficient, degree and the betweenness in Figure 16, Figure 17 and Figure 18, respectively.

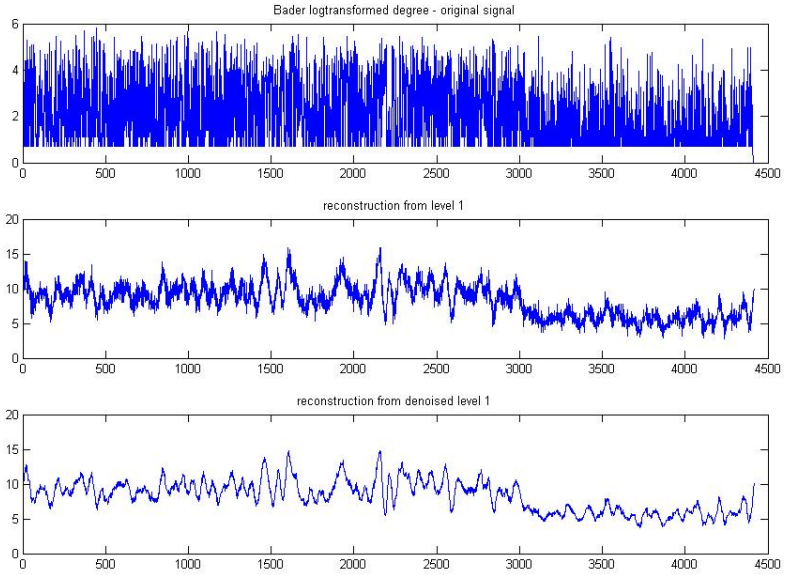
In Figure 16 the degree is considered, and for the Bader's HTP-PI the histograms which are particularly influenced by denoising are the first two resolution levels, while for the Reguly's LC-PI the impact seems to affect the first three levels.

Vice versa for the clustering coefficient, in Figure 17, where the Bader's HTP-PI histogram have the first two resolution levels involved, while the Reguly's LC-PI shows changes only

[2]This estimator uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle in statistics leads to design estimators, and since the de-noised signal can be assimilated to the estimator of the unknown regression function, this choice achieves the minimum of the maximum mean square error obtained for the worst function in a given set.

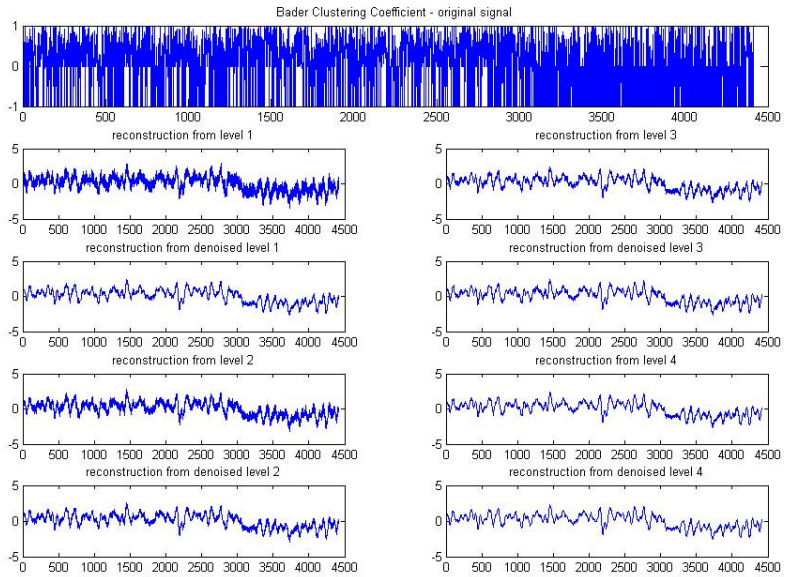


A

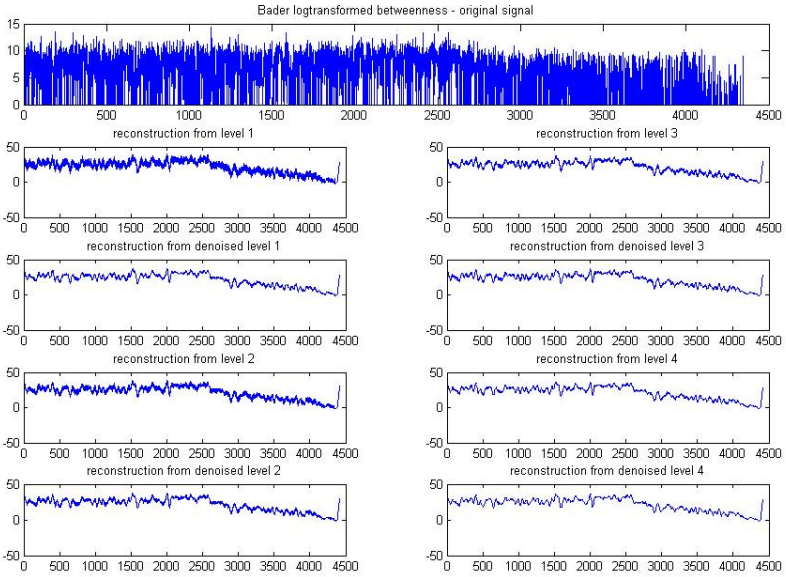


B

Figure 11: Wavelet reconstructions for Bader's HTP-PIs: Degree all levels (A) and first level (B). Denoised reconstructions on right side.

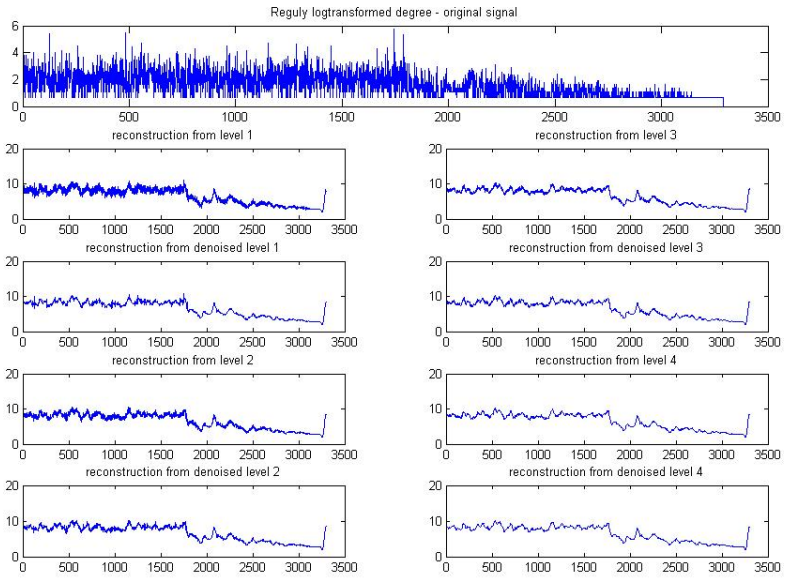


A

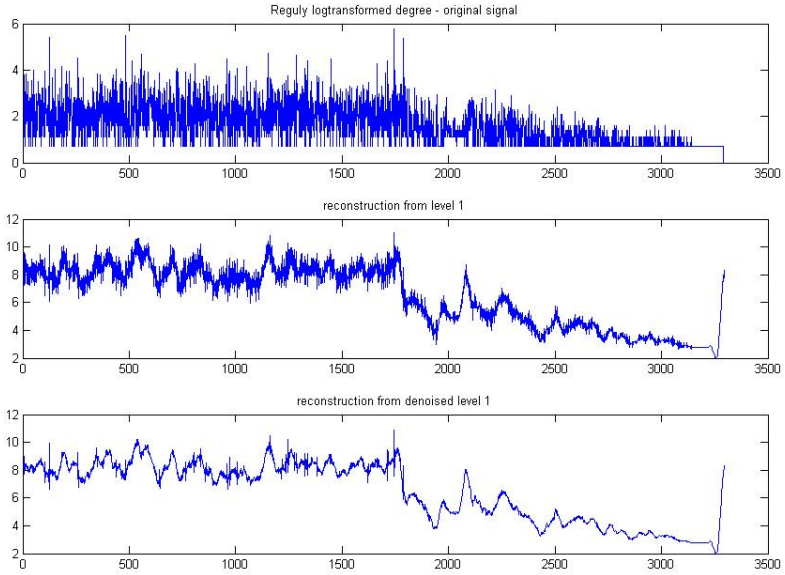


B

Figure 12: Wavelet reconstructions for Bader's HTP-PIs: clustering coefficient (A) and betweenness (B). Denoised reconstructions on right side.

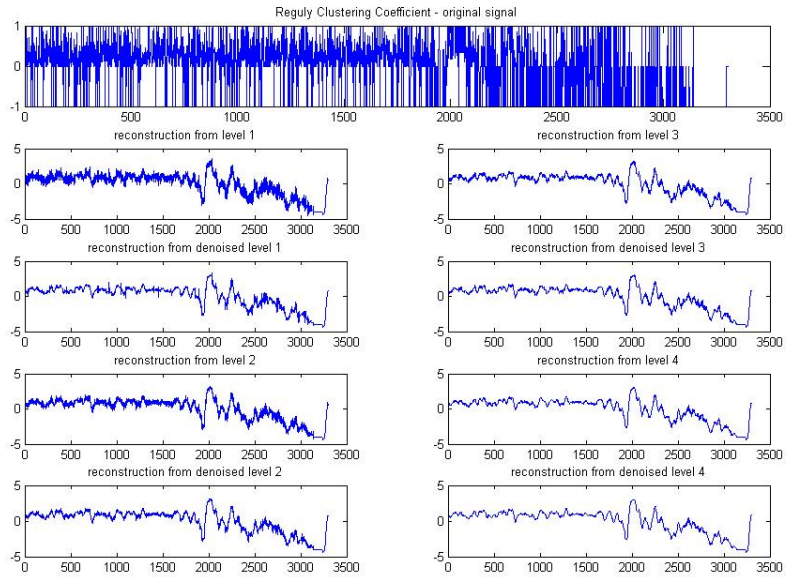


A

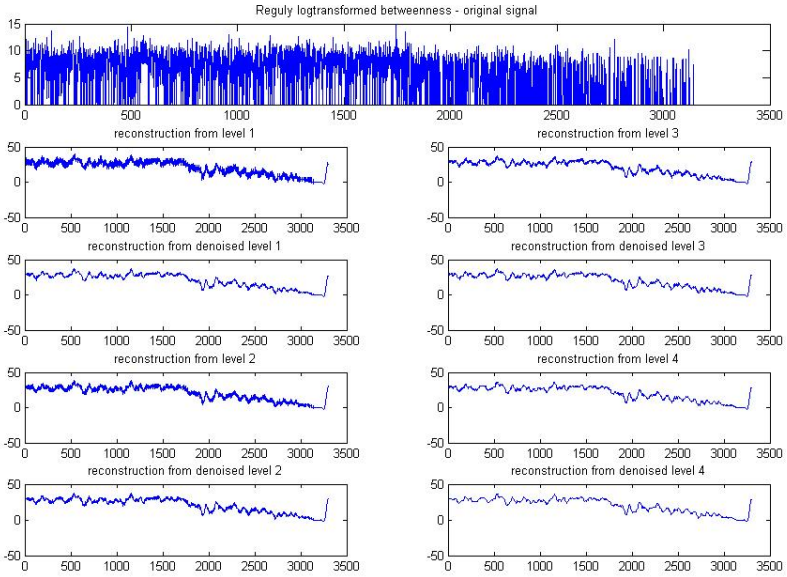


B

Figure 13: Wavelet reconstructions for Reguly's LC-PI: Degree all levels (A) and first level (B). Denoised reconstructions on right side.

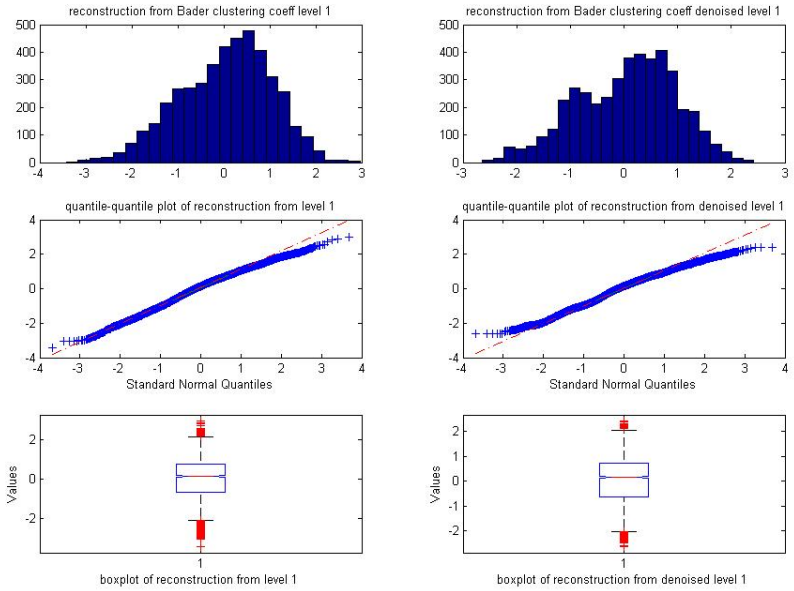


A

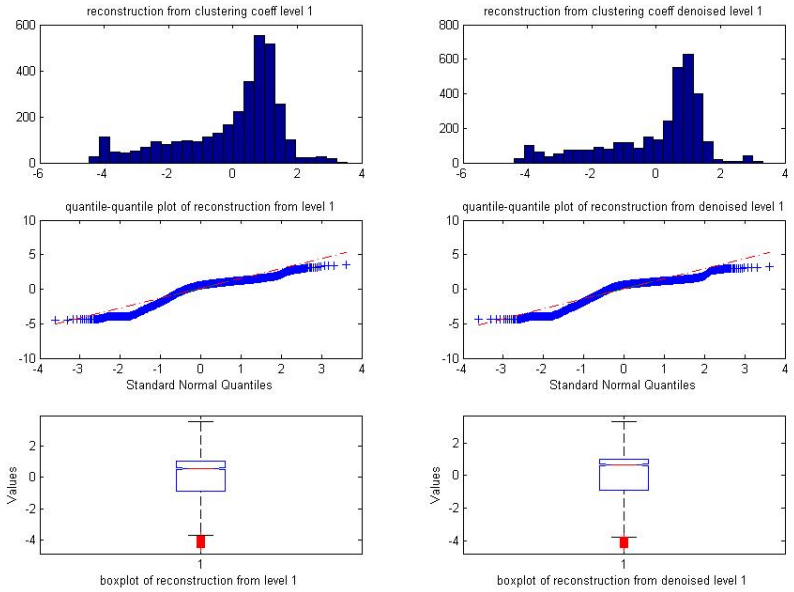


B

Figure 14: Wavelet reconstructions for Reguly's LC-PI: clustering coefficient (A) and betweenness (B). Denoised reconstructions on right side.



A



B

Figure 15: Bader's HTP-PI (A) and Reguly's LC-PI diagnostic plots for clustering coefficient, first level only. Denoised values on the right.

at the first level. Last, Figure 18 shows betweenness and the first level appears affected in Bader's HTP-PI, while also the second level changes in Reguly's LC-PI.

To summarize our findings, and considering that only preliminary work has been here presented, we predict an important future role in protein interactomics for multiscale techniques, among which wavelets represent the most popular one.

We establish some simple steps through which one can:

- Start from a network/graph representation with nodes/vertices (proteins) and links/edges (interactions);
- Move to a transformed space of features computed node-wise and thus obtaining a signal vector;
- Map to a wavelet coefficient space where to explore further properties of the original interactome.

The gains from such an approach are multiple. One key aspect in protein interactomes is the fact that they basically represent static pictures, i.e. snapshots of protein interaction activity fixed at a certain time and set of conditions. The exploration of their innermost resolution or time scales can deliver key insight on the cascade dynamics occurring between interacting proteins.

Another key aspect is that localizing at a finer and finer grid the interaction dynamics within the protein interactome can help revealing some regulatory mechanisms particularly relevant in disease-related studies. A global view of regulation can be too convoluted to be easily untangled. Better to dissect when possible the interactome in sub-interactomes and even smaller sub-structures for increasing the chances of establishing more accurate causal relationships in a fragmented scenario.

Last but not least, when debating about protein interactomics we have to consider the current strong limitations we deal with, in terms of coverage and accuracy. The well-known consequence is a big impact of false negatives (from only partial approximation to the truth) and false positives (from biases and errors of various nature affecting the experimental measurements). With wavelet-based denoising we have a powerful tool to prune the interactomes and limit the detrimental influence of such artifacts.

Acknowledgements.
The authors want to thank IMA for admitting and sponsoring their participation to the workshop *Organization of Biological Networks*, part of the *Mathematics of Molecular and Cellular Biology* 2007-08 program, where two posters were presented.

References

[1] Aronszajn, N., Theory of Reproducing Kernels. *Transactions American Mathematical Society*, 686, pp. 337-404, 1950.

[2] Abry, P., Veitch, D., and Flandrin, P., Long range dependence: revisiting aggregation with wavelets. *J. Time Series Analysis*, 19(3), pp. 253-266, 1998.

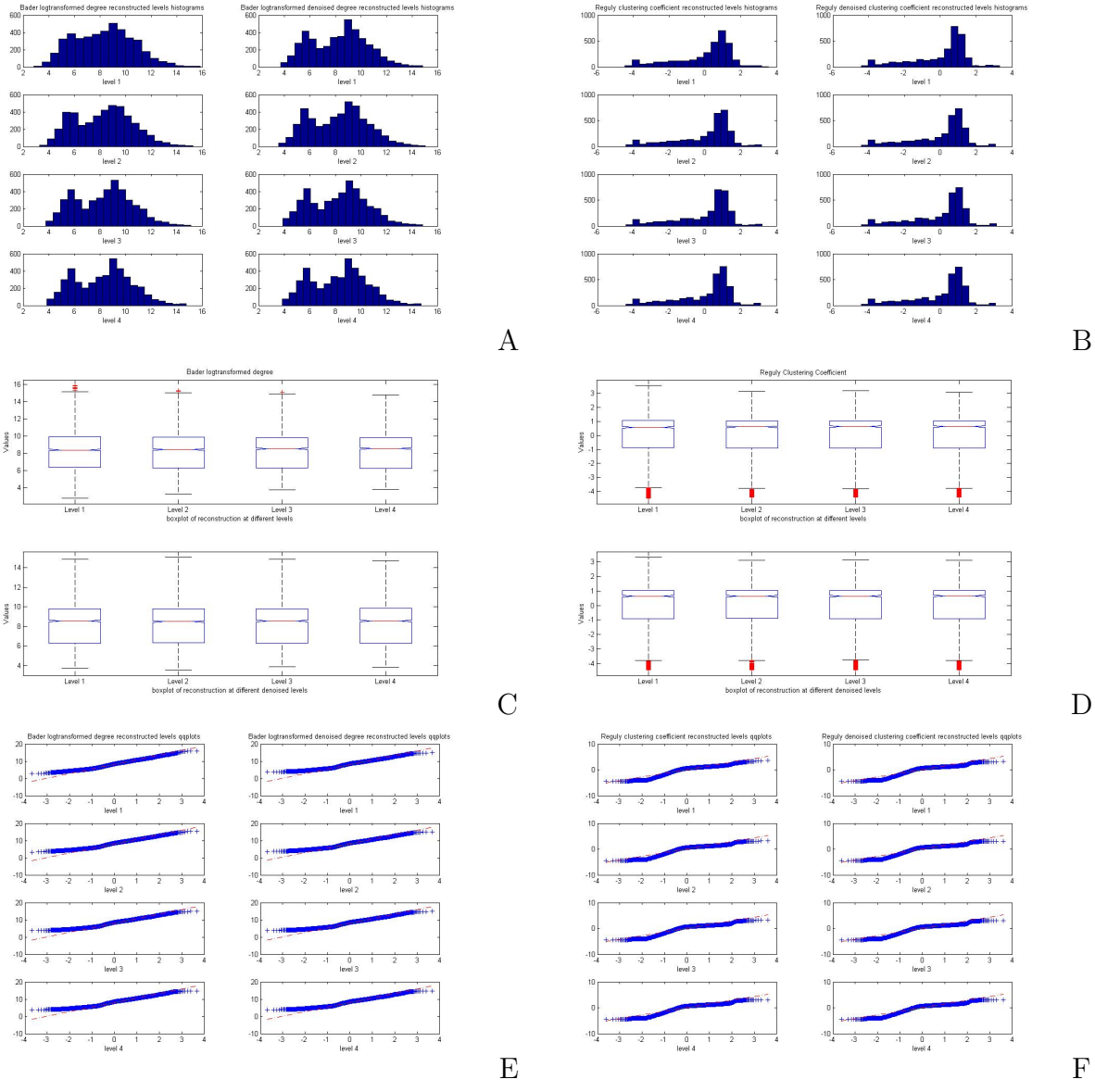


Figure 16: Bader's HTP-PI diagnostic plots (A, C, E) for degree and denoised values. Reguly's LC-PI corresponding plots (B, D, F).

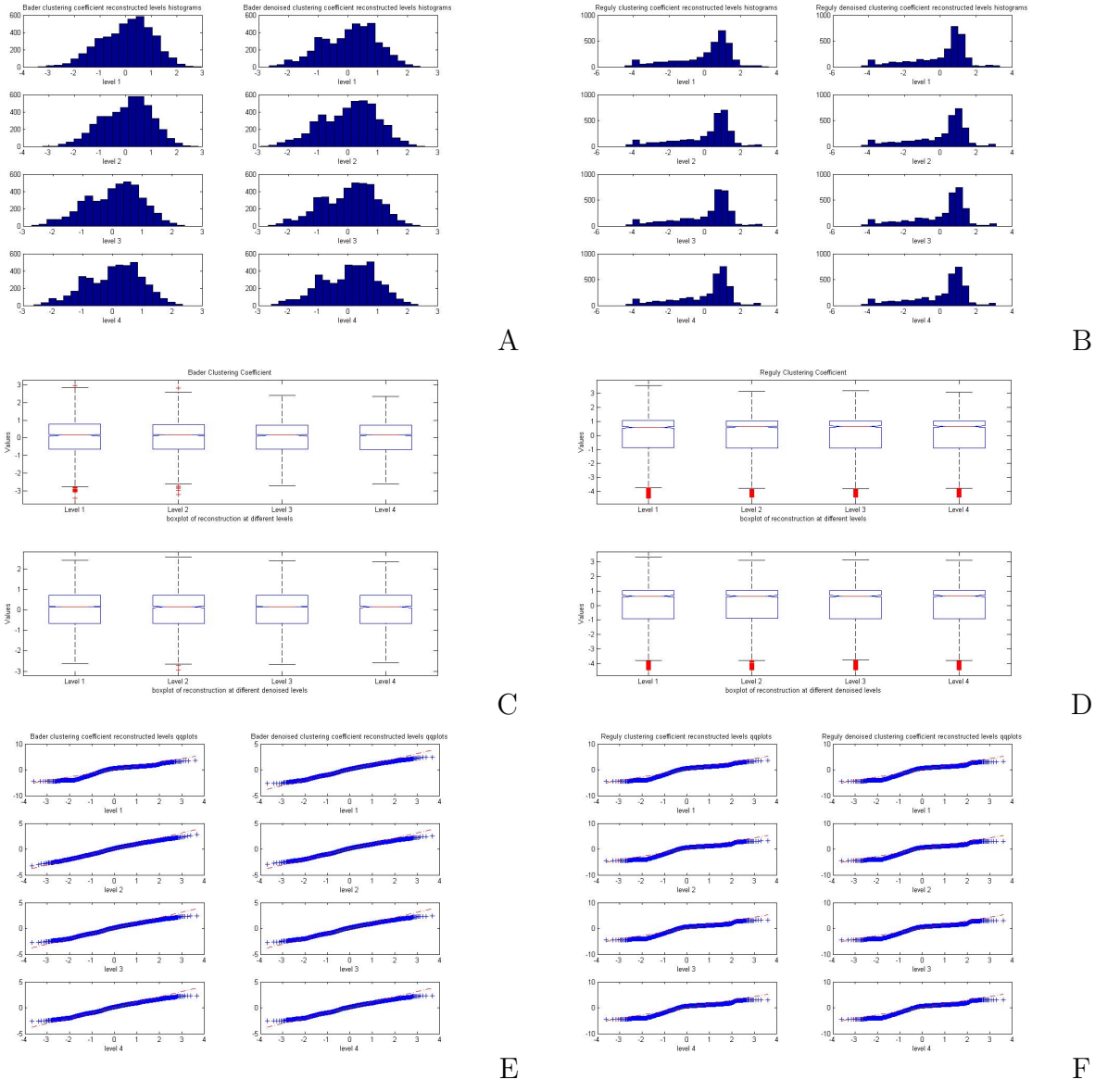


Figure 17: Bader's HTP-PI's diagnostic plots (A, C, E) for clustering coefficient and denoised values. Reguly's LC-PI's corresponding plots (B, D, F).

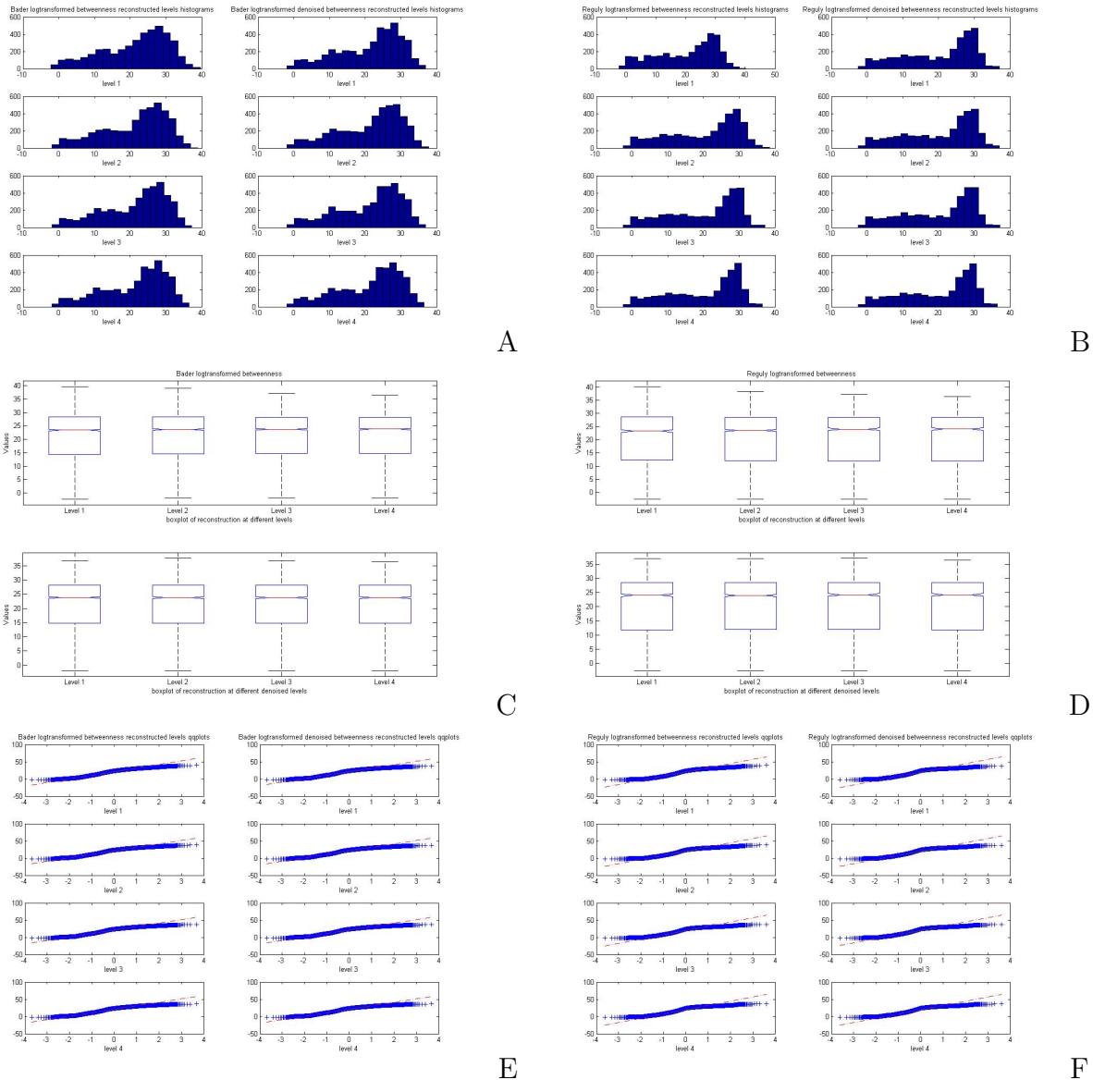


Figure 18: Bader's HTP-PI's diagnostic plots (A, C, E) for betweenness and denoised values. Reguly's LC-PI's corresponding plots (B, D, F).

[3] Bader, J.S., Chaudhuri, A., Rothberg, J.M., and Chant, J., Gaining confidence in high-throuhput protein interaction networks.a *Nature Biotechnology*, 22(1), pp. 75-85, 2004.

[4] Barabasi, A.L., and Albert, R., Emergence of scaling in random networks. *Science*, 286, pp. 509-512, 1999.

[5] Brandes, U., A Faster Algorithm for Betweenness Centrality. *J. Mathematical Sociology*, 25(2), pp. 163-177, 2001.

[6] Cardoso J.F., Source separation using higher order moments. *Proc. ICASSP*, pp. 2109-2112, 1989.

[7] Cardoso J.F., Dependence, Correlation and Gaussianity in Independent Component Analysis. *J. Machine Learning Research*, 4(7-8), pp. 1177-1203, 2004.

[8] Cardoso J.F., and Souloumiac A., Blind beamforming for non-Gaussian signals. *IEE Proc. F.*, 140(6), pp. 771-774, 1993.

[9] Coifman, R.R., Meyer, Y., and Wickerhauser, M.V., Wavelets analysis and signal processing. In *Wavelets and their applications*; Ruskai, B. *et al.* Eds.; Jones and Barlett: Boston, pp. 153-178, 1992.

[10] Comon, P. Independent Component Analysis - a new concept? *Signal Processing*, 36(3), pp. 287-314, 1994.

[11] Daubechies, I. *Ten Lectures on wavelets*. SIAM: Philadelphia, 1992.

[12] DeVore R.A., Nonlinear Approximation. *Acta Numerica* pp. 51-150, 1998.

[13] Donoho D. and Johnstone I.M. Ideal Spatial Adaptation via Wavelet Shrinkage *Biometrika*, 81, pp. 425-455, 1994.

[14] Donoho D. and Johnstone I.M. Adapting to unknown smoothness via wavelet shrinkage *J. Amer. Statist. Ass.*, 90, pp. 1200-1224, 1995.

[15] Donoho D. and Johnstone I.M. Minimax Estimation via Wavelet Shrinkage *Ann. Statist.*, 26, pp. 879-921, 1998.

[16] Erdos, P. and Renyi, A., *Publ. Math.* (Debrecen) 6, pp. 290, 1959.

[17] Freeman L. C., A set of measures of centrality based on betweenness, *Proteins: Structure, Fuction, and Bioinformatics*, 40, pp. 35-41, 1977.

[18] Hyvarinen A., and Oja, E., A fast fixed-point algorithm for Independent Component Analysis. *Neural Computation*, 9(7), pp. 1483-1492, 1997.

[19] Hyvarinen A., Fast and robust fixed-point algorithms for Independent Component Analysis. *IEEE Transactions Neural Networks*, 10(3), pp. 626-634, 1999.

[20] Huang, S.Y., Density Estimation by Wavelet-based Reproducing Kernels. *Statistica Sinica*, 9, pp. 137-151, 1999.

[21] Johnstone, I.M. Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statistica Sinica*, 9, pp. 51-83, 1999.

[22] Johnstone, I.M., and Silverman, B.W., Wavelet threshold estimators for data with correlated noise. *J. Royal Statistical Society Series B.*, 59, pp. 319-351, 1997.

[23] Jolliffe, I.Y., *Principal Component Analysis*, NY: Springer, New York, 1986.

[24] Lewicki, M.S., and Sejnowski, T.J., Learning Overcomplete Representations. *Neural Computation*, 12(2), pp. 337-365, 2000.

[25] Love, M. *Probability theory*, Vol. II, 4th ed., Graduate Texts in Mathematics, Vol. 46, Springer-Verlag, 1978.

[26] Lu, L.J., Xia, Y, Paccanaro, A., and Gerstein, M., Assessing the limits of genomic data integration for predicting protein networks. *Genome Research*, 15, pp. 945-953, 2005.

[27] Mallat, S., Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Transactions American Mathematical Society*, 315, pp. 69-87, 1989.

[28] Mallat, S., *A Wavelet Tour of Signal Processing*. Academic Press: London, 1999.

[29] Meyer, I., *Wavelets: algorithms and applications*. SIAM: Philadelphia, 1993.

[30] Parzen, E., An approach to time series analysis. *Annals Mathematical Statistics*, 32, pp. 951-989, 1961.

[31] Parzen, E. Statistical Inference on Time Series by RKHS methods. In Pyke, R. Ed.; 12th Biennial Sem. Can. Math. Congr. Proc.; Montreal, 1970.

[32] Qi, Y., Bar-Joseph, Z., and Klein-Seetharaman, J., Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins: Structure, Fuction, and Bioinformatics*, 63, pp. 490-500, 2006.

[33] Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.J., Hon, G.C., Myers, C.L., Parsons, A., Friesen, H., Oughtred, R., Tong, A., Stark, C., Ho Y., Botstein, D., Andrews, B., Boone, C., Troyanskya, O., Ideker, T., Dolinski, K., Batada, N.N., and Tyers, M., Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae, *Journal of Biology*, 5:11, 2006.

[34] Triebel H., *Theory of Function Spaces II*. Birkhauser Verlag, Berlin, 1992.

[35] Veitch, D., and Abry, P. A wavelet-based joint estimator for the parameters of LRD. *IEEE Transactions Information Theory*, 45(3), pp. 878-897, 1999.

[36] Vidal, M., Interactome Modeling, *FEBS Letters*, 579, pp. 1834-1838 2005.

[37] Wang Y., Function Estimation via Wavelet Shrinkage for Long Memory Data. *The Annals of Statistics*, 24(2), pp. 466-484, 1996.

[38] Watts, D.J., and Strogatz, S.H., Collective dynamics of "small-world" networks. *Nature*, 393, pp. 440-442, 1998.