

## Research Article

# Multiscale *fragPIN* Modularity

Enrico Capobianco<sup>1,2</sup>

<sup>1</sup> Center for Computational Science (CCS), University of Miami, Miami, FL 33136, USA

<sup>2</sup> Laboratory for Integrative Systems Medicine (LISM), Institute of Clinical Physiology (IFC), National Research Council (CNR), 56124 Pisa, Italy

Correspondence should be addressed to Enrico Capobianco, [enrico.capobianco@gmail.com](mailto:enrico.capobianco@gmail.com)

Received 12 November 2012; Accepted 3 December 2012

Academic Editors: S. Cavallaro, A. Stubbs, and Z. Zhang

Copyright © 2013 Enrico Capobianco. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modularity in protein interactome networks (PINs) is a central theme involving aspects such as the study of the resolution limit, the comparative assessment of module-finding algorithms, and the role of data integration in systems biology. It is less common to study the relationships between the topological hierarchies embedded within the same network. This occurrence is not unusual, in particular with PINs that are considered assemblies of various interactions depending on specialized biological processes. The integrated view offered so far by modularity maps represents in general a synthesis of a variety of possible interaction maps, each reflecting a certain biological level of specialization. The driving hypothesis of this work leverages on such network components. Therefore, subnetworks are generated from fragmentation, a process aimed to isolating parts of a common network source that are here called fragments, from which the acronym *fragPIN* is used. The characteristics of modularity in each obtained *fragPIN* are elucidated and compared. Finally, as it was hypothesized that different timescales may underlie the biological processes from which the fragments are computed, the analysis was centered on an example involving the fluctuation dynamics inherent to the signaling process and was aimed to show how timescales can be identified from such dynamics, in particular assigning the interactions based on selected topological properties.

## 1. Introduction

PIN [1] are almost pervasively studied in genomics, but especially when *H. Sapiens* is considered they present limitations due to sparse coverage and suboptimal accuracy of both experimental (yeast two-hybrid, for instance) and in silico measurements (literature mining, orthology, etc.) [2, 3]. This overall uncertainty is reflected in a pathological presence of false positives and negatives and ultimately complicates data mining and analysis tasks. In order to bypass the complexities induced by such factors, data integration strategies are widely pursued (for instance, studies in [4, 5] have become quite popular). However, a difficulty comes from the fact that the integrated entities are usually heterogeneous, and thus normalization and rescaling need to be considered. An excellent example of the complexity underlying a sequence of integrative omics tasks is offered by the personal omics

profiling work recently published by Chen et al. [6], soon considered a reference for personalized medicine research.

The working hypothesis of this short paper is to adopt an opposite investigation strategy compared to aggregation: instead of integrating the PIN dataset with data from other omics sources, its constituent entities were explored, considering the building blocks that biologically allow for the protein interactions to be observed and measured, at least in part. A PIN map consists of three main types of constituent entities: positive data, that is, the measured physical interactions, which represent the real evidence; negative data, that is, the interactions that are not present, considered as latent variables; and uncertain data, that is, noisy information (false positives) for which partial recovery is possible through data integration. Notably, this mix is usually measured through both transient and persistent PIN dynamics, together with the related degree of uncertainty.

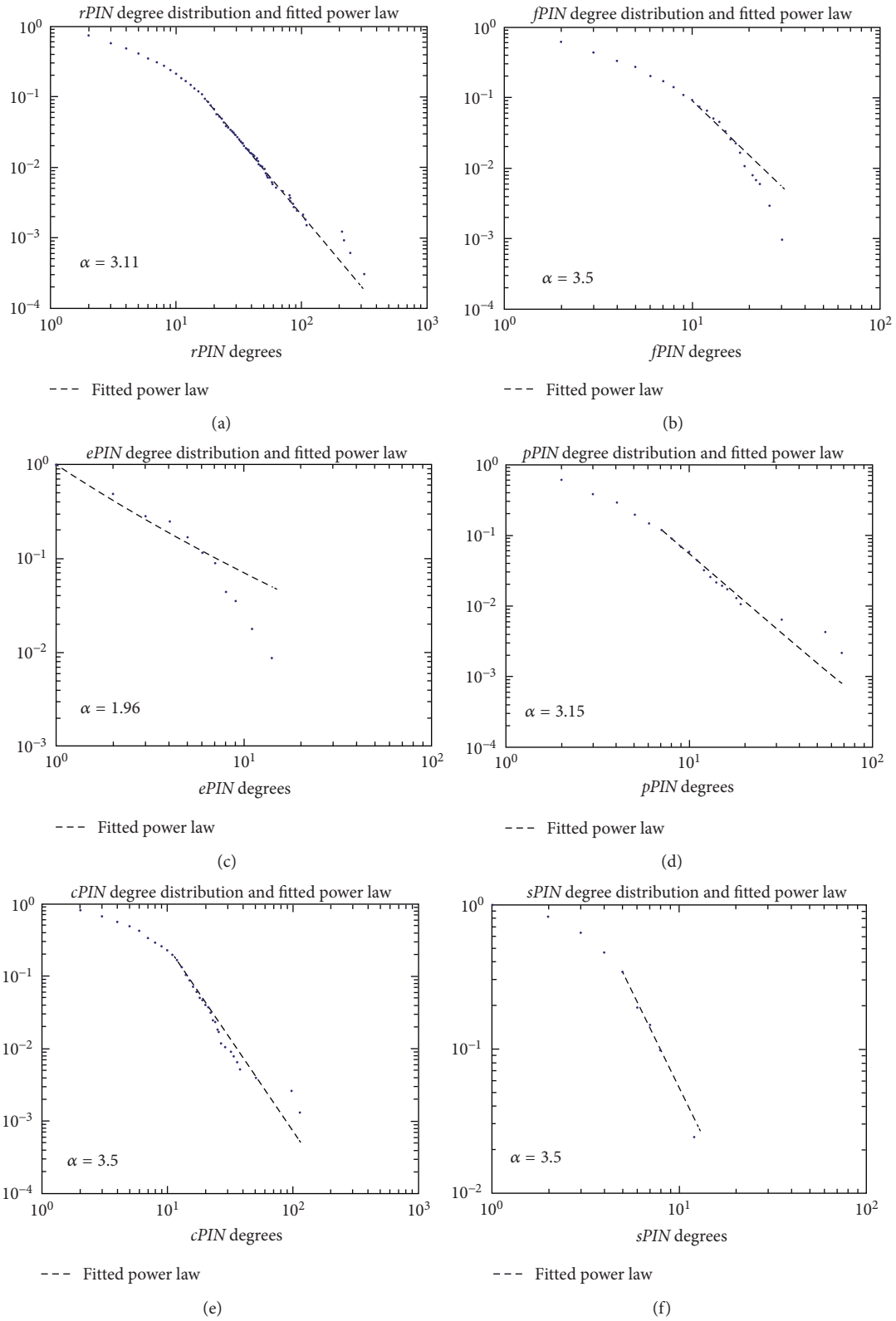


FIGURE 1: Continued.

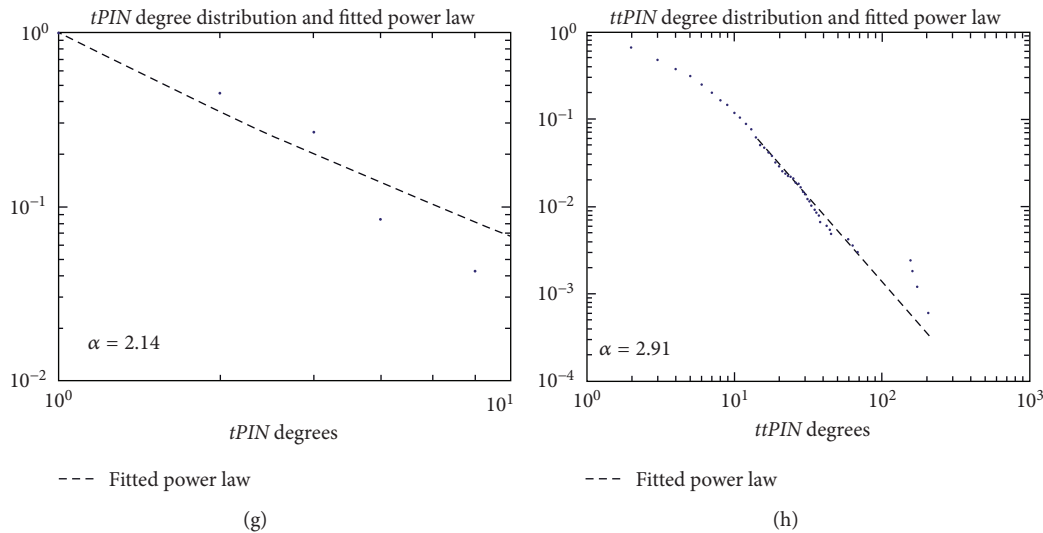


FIGURE 1: Distributional laws for *fragPIN*: pattern comparisons and goodness of fit between degree distributions and power laws.

This work aims to elucidate through the comparative assessment of module-finding algorithms the relationships between topologies that belong to the same network. In particular, PIN can be considered to assemble various interactions which depend on specialized biological processes. The integrated view generally offered by modularity maps represents indeed a synthesis of a variety of possible interaction maps [7–9] embedded in the same network. Individual reference to such maps was made, at least for a list of them, and *fragPIN* were used to indicate the type of network which is generated from fragmentation, a process that retrieves from the same network source a certain number of biologically differentiated subnetworks. Then, elucidation of the characteristics of modularity in each obtained fragment was carried out, helping to investigate the hypothesis that different timescales may underlie the interactive dynamics related to the biological processes from which the fragments are computed. As an example, analysis of PIN fluctuation dynamics for signaling was carried out to show how the inherent timescales can be identified, and interactions assigned to them based on selected topological properties.

Following the work of Huthmacher et al. [10], previous examples of comparative network biology analysis have been suggested by Durek and Walther [11] with the attempt to elucidate the implications of PIN for the regulation of the underlying reaction networks. A comprehensive analysis of enzyme-enzyme interactions in metabolic networks of *E. coli* and *S. cerevisiae* has thus been performed. The latter has involved the analysis of topological properties of these different but related networks and addressed issues such as the efficiency of metabolic processes and how the organization of enzyme interactions correlate with metabolic efficiency.

The methods adopted in the above papers required the study of the global network connectivity properties, various filtering steps to reveal organization differences between all interaction sets and networks targeted to metabolism, and the analysis of scale-free exponents, average cluster

coefficient, degree correlation, distance, and centrality was performed. Priority was assigned to *fragPIN* modularity, and by computing modules according to two popular techniques, the differential configurations thus obtained were assessed. Modules are characterized by interactions occurring at different timescales and to a degree that depends on the involved biological processes. Unfortunately, technological and experimental sources cannot provide the needed detail of information. Therefore, the timescale decomposition offered by *fragPIN* and inherent to each particular process must be determined in some other ways, for instance in silico through the computational approach described below.

## 2. Methods

Similarly to all the interactome datasets, also the *S. cerevisiae* (yeast) interactome presents its complexities; the work of Reguly et al. [12] is an optimal choice, particularly with regard to the literature-curated interactions from small-scale experiments (among other interactome disaggregated information presented by the authors). The dataset involves 31793 publications and reports about 11334 nonredundant interactions (from a total of 33311) and 3289 proteins. Given this yeast source, a compilation of PINs was built and studied to compare their modular properties. Each subinteractome was analyzed according to the characterizing biological process. This process was called PIN fragmentation. The natural consequence of fragmentation is that specific PINs are built whose connectivity patterns reflect the dynamics inherent to the separately involved biological process. The list is reported below.

- (i) rPIN = Reguly LC interactome: the source interactome.
- (ii) mPIN = *metabolic PIN*.

It is obtained by filtering the rPIN such that proteins with their GO terms not associated to metabolism (source: SGD db, <http://www.yeastgenome.org/>)

Interactomes (nodes, edges)	MCODE cores	MaxMod communities
Total <i>rPIN</i> (3289, 113333)	2cores 85 3cores 39 4cores 26 5cores 12 6cores 12 7cores 8 8cores 6 9cores 5 10cores 3 11cores 1	67(48) (small one)
Metabolic <i>mPIN</i> (1018, 1915)	2cores 46 3cores 20 4cores 12 5cores 9 6cores 4 7cores 2 8cores 2 9cores 1 10/11/12/13cores 1	94(47)
Enzyme <i>ePIN</i> (113, 143)	2cores 6 3cores 4 4cores 2	1
Pathway <i>pPIN</i> (465, 785)	2cores 19 3cores 9 4cores 5 5cores 1 6cores 1	44(27)
Cell-cycle <i>cPIN</i> (771, 2493)	2cores 33 3cores 18 4cores 11 5cores 7 6cores 5 7cores 4 8cores 4 9cores 3 10cores 2 11/12/13cores 1	28(22)
Trans. fact. <i>tPIN</i> (71, 67)	2cores 4 3cores 2	17(8)
Tr. fact. targ. <i>ttPIN</i> (1657, 4055)	2cores 41 3cores 19 4cores 12 5cores 6 6cores 2 7cores 1 8cores 1	46(28)
Signaling <i>sPIN</i> (41, 78)	2cores 4 3cores 2	4(0)

FIGURE 2: Core and community structure in *fragPIN*. Size of interactome included with nodes and edges.

are taken off. *mPIN* contains interactions between metabolic proteins.

(iii) *ePIN* = *enzyme PIN*.

It is obtained by filtering *rPIN* through known annotated enzymes (source: KEGG db, <http://www.genome.jp/kegg/>). *ePIN* contains only interactions between enzymes.

(iv) *pPIN* = *pathways PIN*.

It is obtained by filtering the *rPIN* through pathways retrieved from the KEGG db. *pPIN* contains only interactions between proteins involved in annotated pathways.

(v) *cPIN* = *cell-cycle PIN*.

It is obtained by filtering *rPIN* through proteins involved in cell cycle processes (source: MIPS, [mips.helmholtz-muenchen.de/genre/proj/yeast/and\\_SGD\\_db](http://mips.helmholtz-muenchen.de/genre/proj/yeast/and_SGD_db)). *cPIN* contains interactions between proteins involved in cell-cycle process.

(vi) *tPIN* = *transcription factor PIN*.

It is obtained by filtering *rPIN* through transcription factors (source: YEASTRACT db, <http://www.yeast-tract.com/>). *tPIN* contains interactions between transcription factors.

(vii) *ttPIN* = *transcription factor with targets PIN*.

It is obtained by filtering *rPIN* through transcription factors from the YEASTRACT db. It contains interactions between transcription factors and their target proteins.

(viii) *sPIN* = *signalling PIN*.

It is obtained by filtering *rPIN* using signalling pathways retrieved from KEGG db. *sPIN* contains only interactions between proteins involved in signalling annotated pathways.

### 3. Results

As a first check, distributional properties are computed through the power laws, that is,  $p(k) \propto k^{-\alpha}$ , and reported in Figure 1 with reference to each *fragPIN* and the corresponding estimated exponents too (see [13–17] for general treatment of the topic). The distributions appear quite different, as expected, and this depends on the structure and size of the *fragPIN* which is considered.

**3.1. Modularity.** Modularity is often naturally computed when networks are employed. Many algorithms have become available, and a couple of them have been selected based

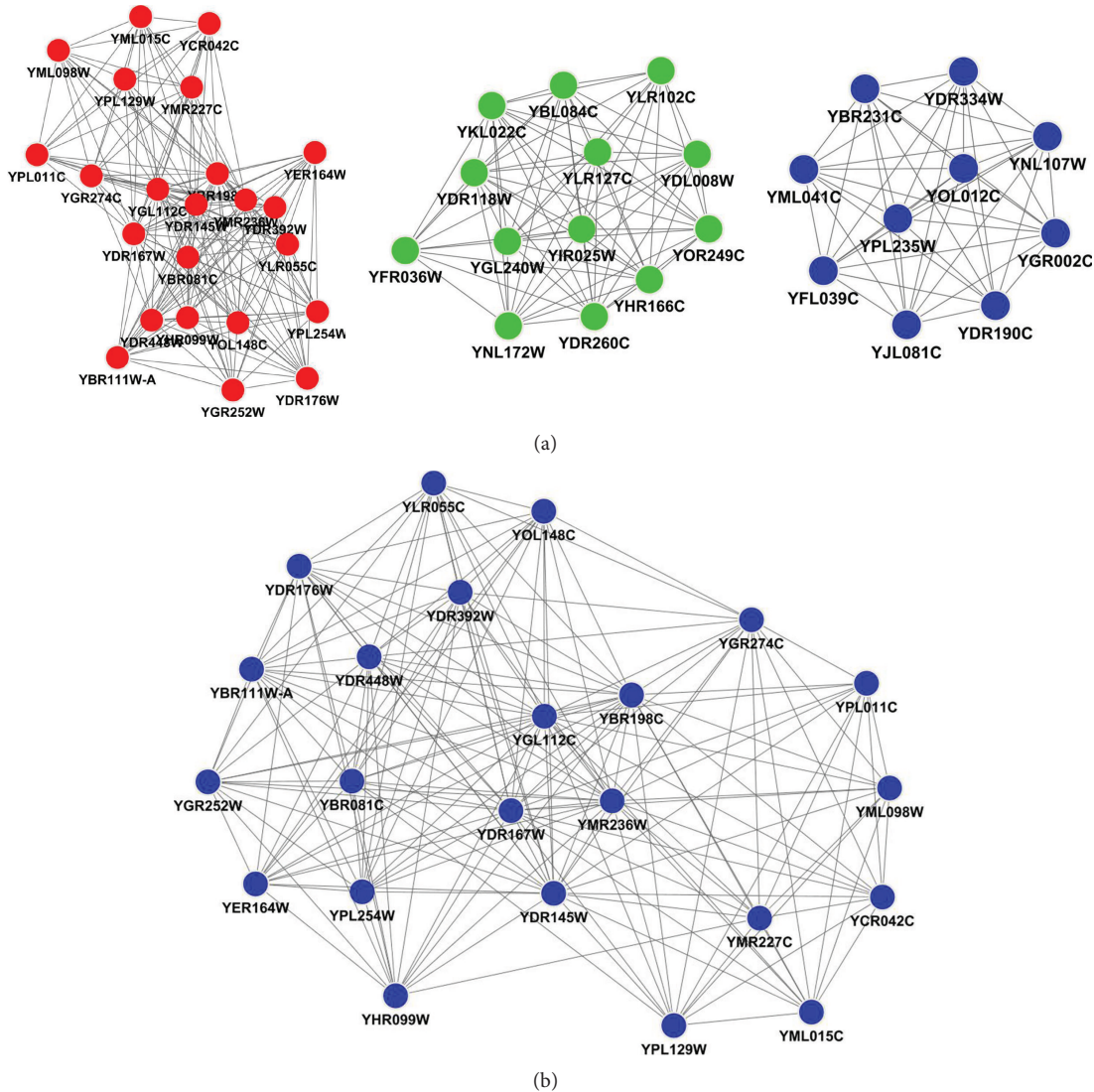


FIGURE 3: Cores (a) and best core computed in *cPIN*. Graphs obtained by using MCODE.

on the popularity and consensus achieved. The first of such methods that we employed is MCODE [18], which exploits local graph density to suggest possible associations between protein complexes and locally dense regions of a graph computed from a clustering coefficient, that is,  $C_i = 2n/(k_i(k_i - 1))$ , where  $k_i$  is the node size of the neighborhood of node  $i$ , and  $n$  is the number of edges in the neighborhood. The  $k$ -core is the structure that one finds in a graph; it is a network of minimal degree  $k$  defined as the remaining subgraph, after that all the nodes with degrees  $1 - k$  have been removed successively.

The procedure is as follows: (a) when a node is removed, all its adjacent edges will also be removed; (b) after a node of degree  $\leq 1 - k$  is removed, in the remaining graph all the remaining nodes with a new degree  $\leq 1 - k$  also need to be removed. In other terms, given  $G = (V, E)$ , the  $k$ -core is computed by pruning all the  $V$  (with their  $E$ ) with degree less than  $k$  until all nodes in the remaining network have at least degree  $k$ .

Then, if a node  $\in k$ -core but  $\notin (1 + k)$ -core of the graph, it has coreness degree  $k$ . The highest  $k$ -core of a network is the central most densely connected sub-network. After vertex weighting, complex prediction is conducted where the relevance of each cluster is validated against known complexes or functional modules, and final statistics are computed about clusters size, density, and functional homogeneity.

The main modules identified for all *fragPIN* are reported in Figure 2 (table format). To obtain them, parameters for network scoring have been set as follows: degree cutoff = 2; for cluster finding: node score cutoff = 0.2; haircut = true; fluff = false;  $K$ -core: 2; and maximum depth from seed: 100.

Modularity can then be computed by another very popular community-finding method called maximum modularity (*MaxMod*). To implement such a method, greedy optimization algorithms have been employed by Clauset et al. [19] to find the best possible modularity structure in networks. In summary, a greedy procedure iteratively merges

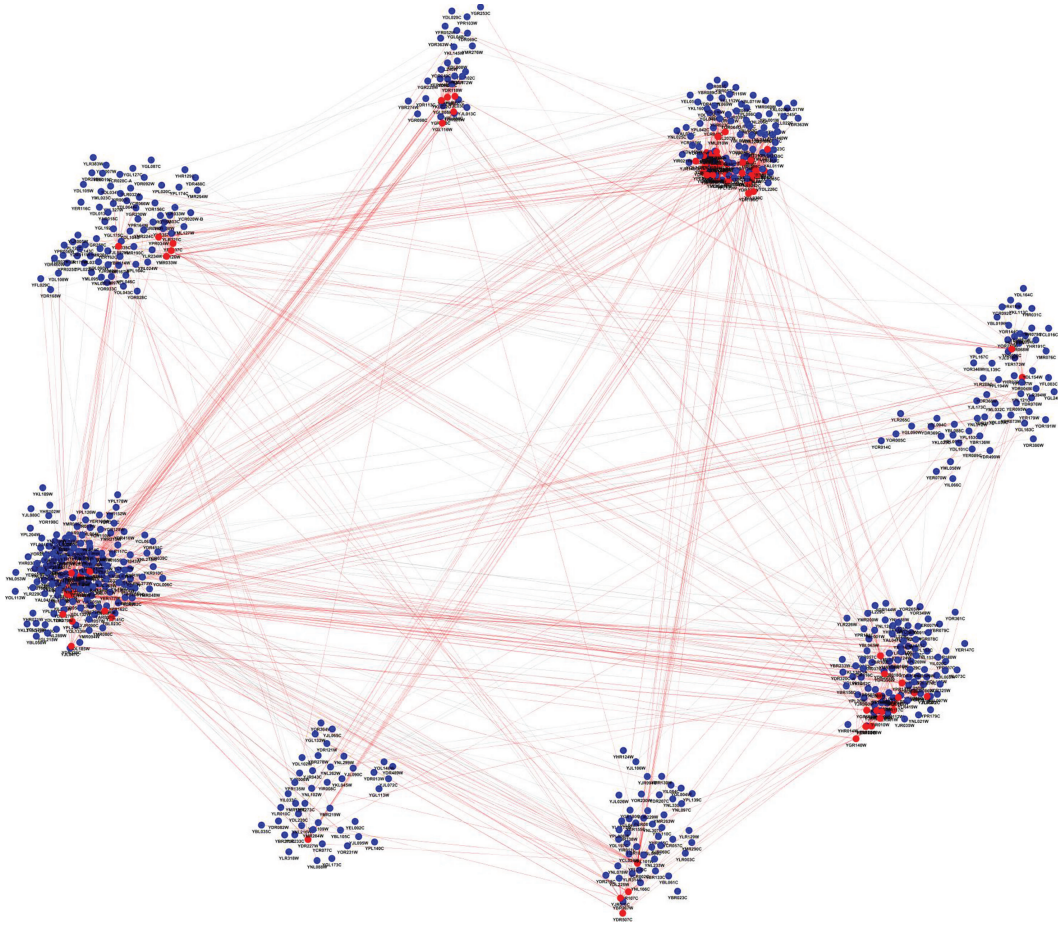


FIGURE 4: Community map computed in *cPIN*, with hubs indicated as red points, and red links connecting them to underline the high-connectivity patterns. Graphs obtained from MaxMod computations.

module pairs showing the largest modularity increase until a gain is observed.

The optimization function  $Q$  [20] is reported below. It is defined as an approximate difference between links observed in a modular network versus those expected in a network of equivalent size where they have been randomly placed. Therefore, a value of zero for  $Q$  indicates that the fraction of within-module links is not different from what would be expected from a randomized network of equivalent size. Nonzero values of  $Q$  indicate deviation from randomness, and values around 0.3 suggest the presence of modular structure (this result comes from extensive simulations reported in the above references) as

$$Q = \sum_i \left[ e_{ii} - \left( \sum_j e_{ij} \right)^2 \right]. \quad (1)$$

The formula reports fractions of links related to nodes within a module  $i$  and fractions of links coming from all other modules relatively to module  $i$ . Therefore, a good partition into modules leads  $Q$  to approach 1; vice versa, the presence of random links between nodes (i.e., poor modularity) would make the two terms not too different, thus delivering a  $Q$

close to 0. Figure 3 shows cores detected in *cPIN*, while Figure 4 shows a community map for it.

**3.2. Timescale Decomposition.** Biological processes embeds dynamics that respond to different timescales; a major problem is how to measure them, in particularly in relation to interactive associations [21]. One way to introduce dynamics at the interactome scale is to integrate gene expression values ideally obtained through time course measurements. However, when such coupled measurements are not available, the problem of deciphering network dynamics is of difficult solution. In a companion paper [22], a special network decomposition approach elucidating both coarse and fine timescales through wavelets [23–26] was proposed. While the focus in previous work was on some particular pathways, a generalization is put forth here.

Using wavelets depends on the entities to be measured, and those ones allowing for suitable timescale decomposition can be good candidates. Such entities, in our case, can be identified by topological features that once measured at each protein (e.g., node) contribute to quantifying a vector-valued signal. The latter can then be decomposed by wavelets. In our

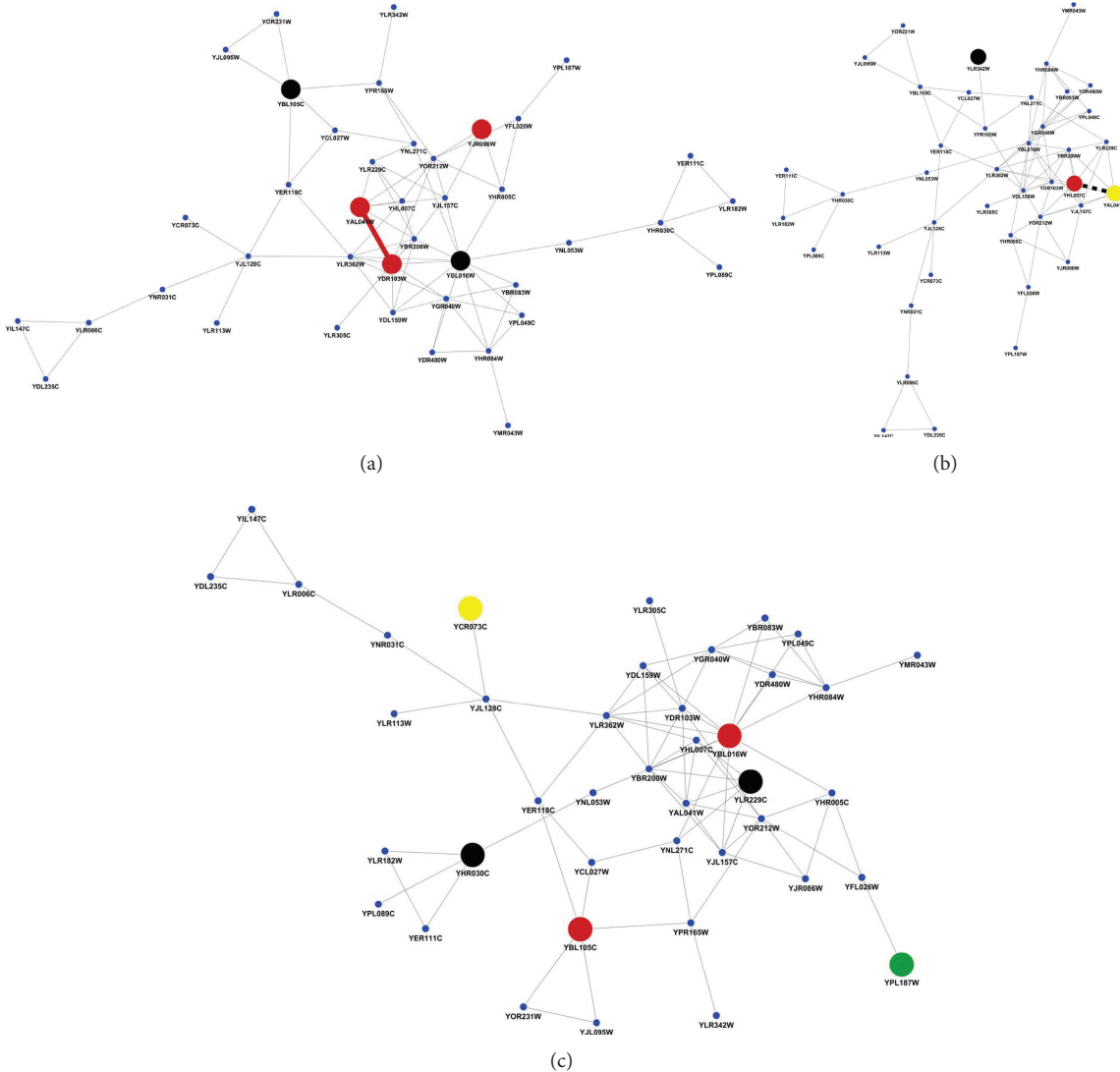


FIGURE 5: Timescale interactions computed for  $sPIN$  by betweenness (a), cluster coefficient, and degree (c). Each color identifies a different timescale under which detection occurred through wavelets.

application, every entry of the feature vector computed from the PIN and to be decomposed across timescales represents a topological property.

An example, apart from the usually exploited degree feature, is provided by betweenness [27–29]. This centrality measure is computed at each network node and increases depending on the volume of crossing at the node, that is, shortest paths (geodesics) going from an origin to a destination through the node relative to the total number of geodesics observed between start and end nodes. For distinct nodes  $v, s, t \in V$ ,  $\pi_{st}$  the number of the shortest paths from  $s$  to  $t$ , and  $\pi_{st}(v)$  the number of the shortest paths passing through  $v$ , it holds that

$$B(v) = \sum_{s \neq v \neq t \in V} \frac{\pi_{st}(v)}{\pi_{st}}. \quad (2)$$

Another problem is how to establish significant variation between timescales in the wavelet values. The approach proposed in our previous methodological paper was centered around two steps: (a) denoising [30–33] applied to get rid of disturbances of random nature; (b) clustering [34] aimed to discriminate between significant and nonsignificant values.

The variability in the measures was initially analyzed through the IQR (interquartile range) robust statistic in order to select the most variable fraction of the data (the half that was selected was called coresets), while discarding the residual part (the box values proximal to the median). A second partitioning was then made of the selected data fraction. In order to control the coresets timescale specificity, some clusters were retrieved. However, also the remaining scattered values were evaluated, that is, the values not assigned to clusters.

A tight clustering technique was adopted, based on a mix of hierarchical and  $k$ -means approaches integrated by

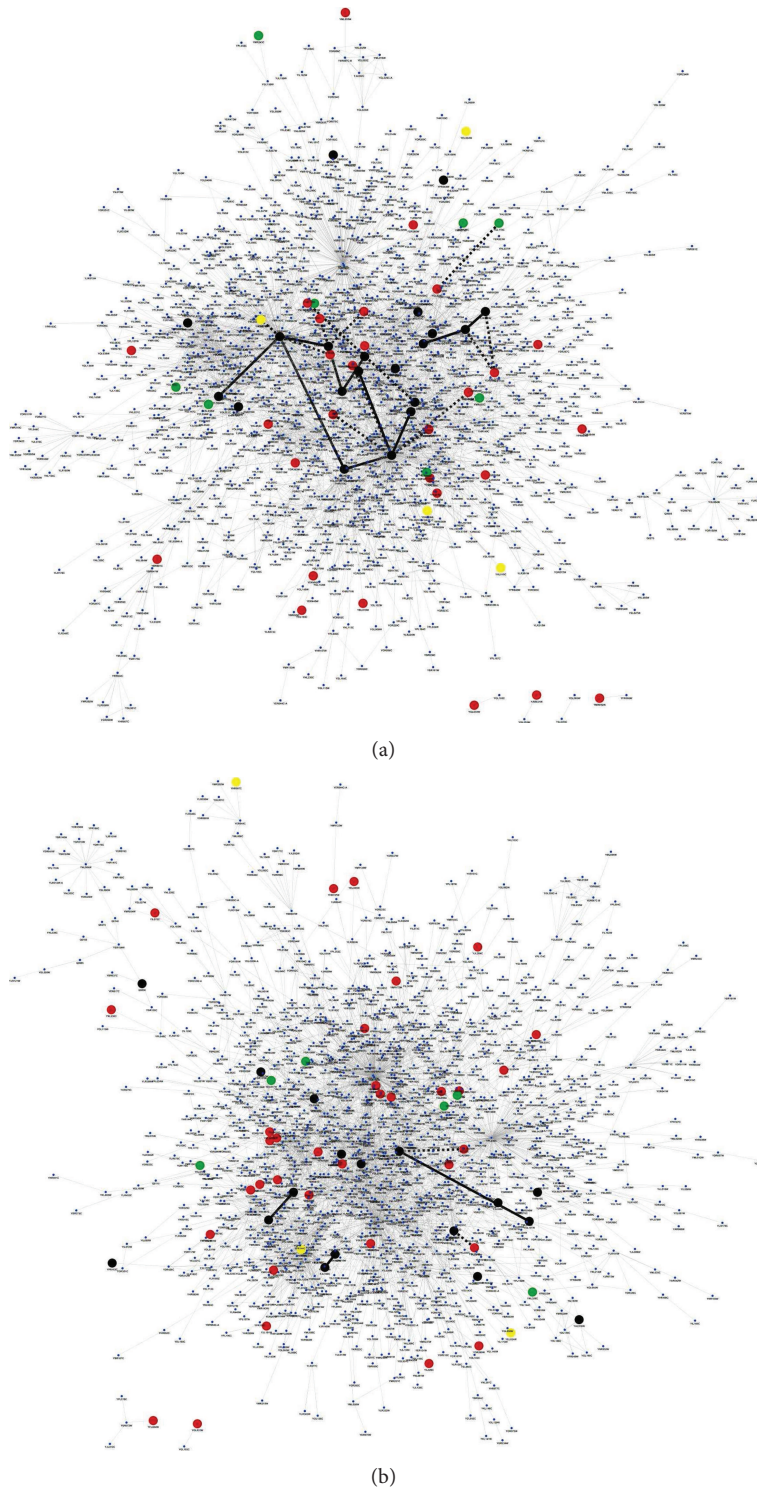


FIGURE 6: Timescale interactions computed by degree (a) and betweenness (b) for *ttPIN*. Each color identifies a different timescale under which detection occurred through wavelets.

bootstrap to form stable clusters. Overall, clusters did not find significant protein modules through which to analyze connectivity or inherent association power of biological relevance. Clusters were also computed over the entire sets of values (without IQR split into coresets and scattered values),

and yet did not deliver biological evidence. Conversely, the analysis of the scattered feature values proved to be more fruitful in terms of reference to timescale specificity, especially for the impact on pathway proximity rather than on network connectivity.



**3.3. Transient versus Permanent Interactions.** A final aspect is how to measure transiency and permanence of interaction dynamics. The emphasis went on their specific interaction dynamics relative to modular connectivity computed within and between timescales, together with pathway proximity. Graphical evidence was reported through Figures 5 and 6. Basically, a scan was first produced through the entire wavelet resolution spectrum for each module under differential conditions then followed by back projection to the PIN of the established associations between particular protein interactions and timescales.

Thus, the cases for which interactive dynamics are simultaneously present at multiple timescales were visualized, together with the links that are possibly appearing between them. S1 (see S1 in the supplementary available at <http://dx.doi.org/10.1155/2013/307608>) reports timescale proximity at pathway level (signaling), which complements the graphical evidences reported at modular network scale. S2 reports the histograms of wavelet-decomposed feature signals (levels and their differences) and diagnostic plots; S3 reports module connectivities detected from each feature across timescales; and S4 reports GO annotation for the identified interactions.

Figure 5 shows timescale-specific interactions computed from feature-dependent modules in *sPIN*. Note that the diversity of colors identifies the different timescales that have been detected by the algorithms. Figure 6 reports instead much denser modules, with reference to *ttPIN*. In terms of comparative evaluation, while Figures 3 and 4 refer to cores and communities, respectively, and these are typical modules found in many studies after applying very well-known methodologies, the proposed approach shows their limitations in detecting resolutions or timescales. Therefore, by involving topological properties computed over specialized PINs, and in particular the information coming from the biological processes, the induced connectivity dynamics between proteins can be emphasized and suitably represented. From a biological point of view, this passage might be important for a series of reasons, (a) the possibility to adopt a differential network analysis based on a comparison of PINs evaluated before and after certain perturbations; (b) the assessment of PIN module configuration changes that might explain phenotypical alterations based on well-characterized protein dynamics.

#### 4. Concluding Remarks

Fragments of PIN offer interesting inference perspectives. The most important aspect is that in reduced dimensionality and complexity, some specialized module functions could be analyzed and possibly validated with reference to specific aspects related to a target pathway or biological process. The second aspect of potential interest is the development of differential network analysis in response to conditions that may affect network dynamics. Finally, time and space dimensions are two entities that define network dynamics and often are overlooked; the timescale analysis here proposed is an example of computational analysis that might

provide relevant information to build more accurate profiles. Without observing protein interactomic dynamics from measurements directly at the experimental level, thus embedding the dynamics from their generating timescales, an attempt to computationally dissect the interactome was made, then separating the effects induced by all the biological processes that were found to be involved. The differences that were detected find justification in a variety of reasons that cannot be inferred from the plain interactome data; however, after examining each separate PIN, a result was that in some cases the timescale dynamics can be revealed through the employed PIN topologies.

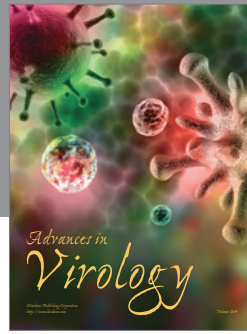
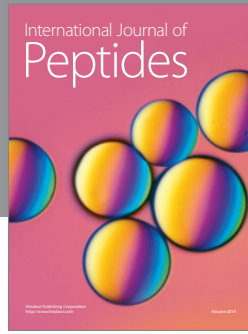
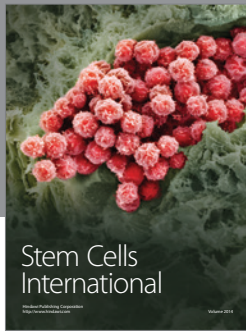
#### Acknowledgment

The author would like to thank his previous collaborators Elisabetta and Antonella.

#### References

- [1] M. Vidal, "Interactome modeling," *FEBS Letters*, vol. 579, no. 8, pp. 1834–1838, 2005.
- [2] J. F. Rual, K. Venkatesan, T. Hao et al., "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, pp. 1173–1178, 2005.
- [3] G. T. Hart, A. K. Ramani, and E. M. Marcotte, "How complete are current yeast and human protein-interaction networks?" *Genome Biology*, vol. 7, no. 11, article 120, 2006.
- [4] C. von Mering, R. Krause, B. Snel et al., "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature*, vol. 417, no. 6887, pp. 399–403, 2002.
- [5] L. J. Lu, Y. Xia, A. Paccanaro, M. Gerstein, and H. Yu, "Assessing the limits of genomic data integration for predicting protein networks," *Genome Research*, vol. 15, no. 7, pp. 945–953, 2005.
- [6] R. Chen, G. I. Mias, J. Li-Pook-Than et al., "Personal omics profiling reveals dynamic molecular and medical phenotypes," *Cell*, vol. 148, no. 6, pp. 1293–1307, 2010.
- [7] A. C. Gavin, P. Aloy, P. Grandi et al., "Proteome survey reveals modularity of the yeast cell machinery," *Nature*, vol. 440, no. 7084, pp. 631–636, 2006.
- [8] Y. Qi and H. Ge, "Modularity and dynamics of cellular networks," *PLoS Computational Biology*, vol. 2, no. 12, article e174, 2006.
- [9] E. Marras, A. Travaglione, and E. Capobianco, "Sub-modular resolution analysis by network mixture models," *Statistical Applications in Genetics and Molecular Biology*, vol. 9, no. 1, article 19, 2010.
- [10] C. Huthmacher, C. Gille, and H. G. Holzhütter, "A computational analysis of protein interactions in metabolic networks reveals novel enzyme pairs potentially involved in metabolic channeling," *Journal of Theoretical Biology*, vol. 252, no. 3, pp. 456–464, 2008.
- [11] P. Durek and D. Walther, "The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles," *BMC Systems Biology*, vol. 2, article 100, 2008.
- [12] T. Reguly, A. Breitkreutz, L. Boucher et al., "Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*," *Journal of Biology*, vol. 5, article 11, 2006.

- [13] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [14] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [15] M. Newman, A. L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, NJ, USA, 2006.
- [16] G. Palla, I. J. Farkas, P. Pollner, I. Derényi, and T. Vicsek, "Fundamental statistical features and self-similar properties of tagged networks," *New Journal of Physics*, vol. 10, Article ID 123026, 2008.
- [17] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.
- [18] G. D. Bader and W. V. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, no. 2, pp. 1–27, 2003.
- [19] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, Article ID 066111, 6 pages, 2004.
- [20] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review Letters E*, vol. 69, Article ID 026113, 15 pages, 2004.
- [21] N. Jamshidi and B. Ø. Palsson, "Top-down analysis of temporal hierarchy in biochemical reaction networks," *PLoS Computational Biology*, vol. 4, no. 9, Article ID e1000177, 2008.
- [22] E. Capobianco, E. Marras, and A. Travaglione, "Multiscale characterization of signaling network dynamics through features," *Statistical Applications in Genetics and Molecular Biology*, vol. 10, no. 1, article 53, 2011.
- [23] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of  $L^2(\mathbb{R})$ ," *Transactions American Mathematical Society*, vol. 315, no. 1, pp. 69–87, 1989.
- [24] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 1992.
- [25] I. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, Pa, USA, 1993.
- [26] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, London, UK, 1999.
- [27] L. C. Freeman, "A set of measures of centrality based on betweenness," *Proteins*, vol. 40, pp. 35–41, 1997.
- [28] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.
- [29] M. Kitsak, S. Havlin, G. Paul, M. Riccaboni, F. Pammolli, and E. Stanley, "Betweenness centrality of fractal and nonfractal scale-free model networks and tests on real networks," *Physical Review E*, vol. 75, no. 5, Article ID 056115, 8 pages, 2007.
- [30] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [31] D. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of American Statistical Associations*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [32] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *The Annals of Statistics*, vol. 26, no. 3, pp. 879–921, 1998.
- [33] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society B*, vol. 59, no. 2, pp. 319–351, 1997.
- [34] G. W. Tseng and W. H. Wong, "Tight clustering: a resampling-based approach for identifying stable and tight patterns in data," *Biometrics*, vol. 61, no. 1, pp. 10–16, 2005.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

