2011 International Conference on Bioscience, Biochemistry and Bioinformatics
IPCBEE vol.5 (2011) © (2011) IACSIT Press, Singapore

Detecting Crosstalk Modules of Combined Networks: the Case for the NF-κB and p53

Thanh-Phuong Nguyen[*], Adaoha E. C. Ihekwaba[*] and Corrado Priami[*,+]
[*]The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Trento, Italy
[+]Dept. of Information Engineering and Computer Science, University of Trento, Trento, Italy
Email: {nguyen,ihekwaba,priami}@cosbi.eu

***Abstract*—It has become increasingly clear that signalling pathways are extensively interconnected and are embedded in networks with common protein components. These components do not exist in isolation but may gather together to form crosstalk modules. Constructing these crosstalk modules has emerged as a good method to understand the mechanisms underlying the propagation of transduction signals in cell. In this paper, we have presented an advancement of the method, which is chiefly used to integrate multiple topological and functional data to detect crosstalk modules between NF- B and p53 signalling systems. Applying the Expectation Maximization (EM) clustering algorithm, we obtained the competitive results compared to the k-means algorithm. The EM algorithm as a soft clustering method is able to distinguish overlapping parts among clusters, and here we that it is potentially more sensible than the k-means algorithm in detecting the cross talking modules involved in the network interactions between the two systems NF- B and p53. In addition, the biological analyses support our findings, and propose testable hypotheses to which the functional networks are involved in along with their associated human diseases.**

Keywords-crosstalk modules, NF- B network, p53 network, data integration, systems biology multi-relational clustering.

I. INTRODUCTION

Signal transduction networks are complex biological networks that allow the cell to receive, transmit, and act upon molecular signals. Although these networks are essential for the correct functioning of the cell, they have also been reported to themselves result in abnormal cellular transformation or differentiation, often producing a pathological disease outcome [1]. NF- B and p53 systems have attracted a widespread interest because of their central role several human diseases [2], [3] NF- B exists in the majority of cell types as an homodimer or heterodimer of a family of structurally related proteins and is particularly important in modulating the expression of immunoregulatory genes (or proteins, or molecules). NF- B has been shown to interact with other signalling networks resulting in complex non-linear responses to different combinations of stimuli [4]. One of the most relevant networks interacting with NF- B is p53, a tumour suppressor that upon DNA damage recognition mediates the activation of defence mechanisms, such as DNA repair, cell cycle arrest and eventually apoptosis. The critical role of p53 in tumour suppression is underscored by the observation that over 50% of all cancers involve a disruption of this system [5]. Deciphering the interaction between NF- B and p53 systems is appealing since it would be useful in gaining a deeper understanding on the cellular response to external stimuli. Most of the work on uncovering the interaction between these two networks was done by traditional biological experiments [4], [6]. Since the experimental work itself much involves quantitative tasks, system biology, particularly algorithmic systems biology, came to the scene bringing about another approach to the standing issues [7].

To improve the understanding of signalling networks, some groups have taken advantage of the published data on protein-protein interaction (PPI) networks of signalling systems (assuming that signal transduction is mainly dependent on specific protein-protein interactions [8], [9]), and tried to computationally reconstruct those networks. Some of the methods for this purpose used have been: the Markov chain Monte Carlo method [10], the computational algebra method [1] the cost search functions method [11], and others. Li *et al.* built the global pathway crosstalk network by combining pathway and protein interaction data based on the shortest path profiles only [12]. These works achieved interesting results but they mostly considered signalling networks as independent linear cascades. Nevertheless, it is evident that these signalling networks do not exist in isolation, but are likely to interact or influence each other to perform a complex signalling task in cell. These works achieved interesting results but they mostly considered signalling networks as independent linear cascades. Nevertheless, it is evident that these signalling networks do not exist in isolation, but are likely to interact or influence each other to perform a complex signalling task in cell.

Our objective is to model signalling crosstalk modules (crosstalk modules in short) in the combined network of NF- B and p53 systems. We did not consider proteins as individual elements as it was done in previous work, but we analysed their interplays in performing signalling crosstalk both topologically and functionally. To this end, the integration of multiple data was carried out to achieve a comprehensive view of how these systems interact with each other. To the best of our knowledge, this work uncovers for

the first time the crosstalk modules between NF- B and p53 using a computational integration method.

In this paper, we present a new computational method to detect crosstalk modules between NF- B and p53 systems that combines and mines multiple data. First, from the large-scale human protein interaction networks, our method computes all the connecting proteins (CPs) that are likely to be shared between two signalling networks. Then we extracted numerous topological and functional data from multiple data sources, the Universal Protein Resource (Uniprot) [13], the Interologous Interaction Database (i2d) [14], the Reactome [15] and the InterPro database [16] and integrated it in a multi-relational scheme. Finally, we applied the Expectation Maximization (EM) method to mine and detect crosstalk modules between NF- B and p53 systems. To estimate the performance of the method we calculated its case likelihood and compared it to that obtained for the *k*-means algorithm, a hard clustering method. The better case likelihood shows that our soft clustering, EM algorithm is more suitable to discover crosstalk in the propagation of transduction signals between NF- B and p53 networks. In addition, we also analysed the biological functions of the clustered networks to assess the biological relevance of the findings. The analysis gave interesting insights on the biology of NF- B and p53 signalling pathways that could be the starting point for new studies on the regulation mechanisms underlying the pathophysiology of various diseases.

II. MATERIALS AND METHODS

In this section, we present our proposed method for the detection of the signalling crosstalk modules. Algorithm 1 describes our method, that consists of three main tasks: *(i)* Computing the CPs between pathway networks, *(ii)* Manipulating and combining multiple data, and *(iii)* Constructing crosstalk modules using the EM algorithm.

Algorithm 1 Detecting signalling crosstalk modules based on multiple data.

Input:
Protein-protein interaction network $\mathfrak{N}$.
Set of mulitple features corresponding to the extracted data $\mathfrak{F} \subset \{f_{dg}, f_{cc}, f_{func}, f_{pw}, f_{dm}\}$.
Output:
Set of signalling crosstalk modules $\mathfrak{S}$.

1: Identify proteins joining the two systems NF-κB and p53.
2: Model two networks (the network NF-κB $\mathfrak{N}_{nf\kappa B}$ and the network p53 $\mathfrak{N}_{p53}$) from the large-scale protein-protein network $\mathfrak{N}$. $\mathfrak{N}_{nf\kappa B} := \{p_{ij}\}$. $\mathfrak{N}_{p53} := \{g_{ij}\}$. ($p_{ij}$, g_{ij} are binary protein interactions)
3: Find the shortest paths p between every pairs of two proteins (p_i, g_j) where $p_i \subset \mathfrak{N}_{nf\kappa B}$ and $g_i \subset \mathfrak{N}_{p53}$.
$\mathfrak{P} = \mathfrak{P} \cup p$. ($\mathfrak{P}$ is the set of all shortest paths)
4: Select CP c_i joining path p, $\forall\, p \subset \mathfrak{P}$. $\mathfrak{C} := \{c_i\}$
5: For each connecting protein $c_i \subset \mathfrak{C}$
6: Extract topological data from the i2d database for two features, degree f_{dg} and cluster coefficient f_{cc}.
7: Extract functional data for three features, abstract function f_{func} from the Uniprot database, biological process f_{pw} from the Reactome database, and protein domain f_{dm} from the InterPro database.
8: Combine all extracted data by a multi-relational scheme.
9: Construct signalling crosstalk modules/networks s_i based on combined data and the EM algorithm. $\mathfrak{S} := \{s_i\}$.
10: **return** $\mathfrak{S}$.

In Algorithm 1, Step 1 is to identify proteins joining the two systems NF-kB and p53; then in Step 2, the PPI network of these proteins is extracted in Step 2. Steps 3 and 4 are to compute connecting proteins between pathways. Steps 1 to Step 4 are explained in greater detail in Section Section II-A and were partially reported in our previous paper [17]. In Steps 5 to 8, we carry out the data extraction and the data combination by using a multi-relational scheme. Step 9 is for constructing the crosstalk modules between the NF-kB and p53 by applying the EM clustering algorithm. The output is the set of the crosstalk modules.

A. Computing Connecting Proteins Between Networks

The complete protein network of NF- B and p53 systems were modelled and analysed based on binary interactions. First, we identified the proteins joining the two systems. There are five proteins belonging to the NF- B family known in mammalian cells: RelA (also known as p65), c-Bel, RelB, NF- B1 (p50/p105), and NF- B2 (p52/p100). Additionally, NF- B exists in the cytoplasm in an inactive form associated with inhibitory proteins termed I- B, of which the most important ones are I- B , I- B , and I- B [2]. In the p53 network, p53 protein binds to the regulatory region of the MDM2 gene and promotes its transcription [18]. Figure 1. lists protein members of the systems considered in the study (highlighted proteins are reported to be activated in one system and involved in the regulation of another).

Pathway	Protein name	Uniprot accession	Alternative name
p53 pathway	P04637	P53_HUMAN	p53
	Q00987	MDM2_HUMAN	mdm2
	P38936	CDN1A_HUMAN	p21
	Q8N726	CD2A2_HUMAN	p14ARF
NF-κB pathway	O00221	IKBE_HUMAN	NF-κB inhibitor epsilon
	O14920	IKKB_HUMAN	IKK2
	O15111	IKKA_HUMAN	IKK1
	P19838	NFKB1_HUMAN	Nuclear factor NF-κB p105 subunit
	P25963	IKBA_HUMAN	IκB-alpha
	Q00653	NFKB2_HUMAN	Nuclear factor NF-κB p100 subunit
	Q01201	RELB_HUMAN	Transcription factor RelB
	Q04206	TF65_HUMAN	Transcription factor p65 (RelA)
	Q04864	REL_HUMAN	C-Rel protein
	Q14164	IKKE_HUMAN	Inhibitor of nuclear factor κB kinase subunit epsilon
	Q15653	IKBB_HUMAN	NF-kappa-B inhibitor beta
	Q96HD1	CREL1_HUMAN	Crel1
	Q6UXH1	CREL2_HUMAN	Crel2
	Q9Y6K9	NEMO_HUMAN	IKKγ

Figure 1. List of the considered proteins in p53 and NF- B network.

We investigated the binary protein interaction network extracted from the i2d database as an undirected graph *G(E,V)* consisting of a set of nodes (proteins) *V* and a set of edges (interactions) *E* between them. An edge e_{ij} connects vertex v_i with vertex v_j. Proteins that act as intermediates in the signal transduction can be uncovered by searching the shortest paths between networks. The shortest path problem consists in finding a path between two nodes such that the sum of the

weights of its constituent edges is minimized. In general, given a real-value weight function $f: E \rightarrow R$, and a start node v_i of V, we find a path p of P (the set of paths) from v_i to each v_j of V if present, so that $\sum_{p \in P} f(p)$ is minimal among all paths connecting v_i and v_j.

If the PPI networks here constitute an unweighted graph, the weight function f can be considered as a path length l (the number of edges in path p). In this case, the shortest path problem consists in finding a path p having the minimal path length. A Breadth-First Search algorithm [19] has been employed to find the shortest paths between two nodes. The shortest paths may have different path lengths ($l = 1$, $l = 2$, $l = 3$, $l = 4$, etc.). The nodes which belong to the shortest paths, except the starting nodes and the ending nodes, are called connecting proteins (CPs). Playing the roles of intermediate molecules, the CPs themselves do not act independently, but functionally group into signalling crosstalk modules used in connecting the two systems.

B. Manipulating and Combining Multiple Data

The crosstalk modules in the NF- B and p53 systems were identified by both topological and functional properties. As a result, in order to construct the modules we investigated two types of data: *(1)* topological data representing the relationship between a CP and its network neighbours, and *(2)* functional data representing biological information of a CP.

For what concerns the topological data, we calculated two measures based on the protein interaction network extracted from the i2d database. The first one is the degree of a CP, which is the number of its neighbours. Node degree is one of the principal measures used to study the topology of a network. The neighbourhood N for a node v_i is defined as its immediately connected neighbours as $N=\{v_i\}:e_{ij} \in E$ [20]. The second one is the average clustering coefficient $C(k)$, which characterises the overall tendency of nodes to form clusters or groups; and $C(k)$ the average clustering coefficient of all nodes with k links is an important measure of the network structure [20].

For what concerns the functional data, we investigated three different kinds of information: abstract function, biological process and protein domain. The abstract functions are the protein keywords documented in the Uniprot database including protein function, subcellular localization, structure, relevant mutations, etc. Since the proteins in crosstalk modules likely take part in the same cellular processes, data on the biological process were extracted from the Reactome database and analysed. The protein domain data obtained from the InterPro database gives information on the different domains present in the CP. Protein domains are defined as structural or functional elements within a protein and affect the way that the proteins interact with each other and compose the crosstalk modules. Since different databases have different names for each entry, the Uniprot accession number for identifying a protein was used as our standard and thus all CPs' names were converted and mapped accordingly to their Uniprot accession numbers.

To store and integrate the acquired data, a multi-relational scheme was appropriately structured in form of tables (with columns and rows) and relationships betweens tables in the SQL Server Database Management. The data types are heterogeneous, since the abstract function data, the process data and the domain data are in form of a categorical free text while the node degree and the cluster coefficient are numerical values. For example, protein IKBA_HUMAN (NF-kappa-B inhibitor alpha) has the node degree equal to 60, the cluster coefficient equal to 0.05826; contains many keywords, such as Phosphoprotein, 3D-structure, ANK repeat, Cytoplasm, Disease mutation, etc.; takes part in two processes, such as signalling by NGF, signalling in Immune system; has two domains ANK and NF- B inhibitor. In this regard, the heterogeneity in the data makes the use of traditional clustering methods unsuitable in clustering multiple relational data and underlines the importance of using a multi-relational clustering approach in the detection of CPs propagating transduction signals across systems being investigated based on their relational data.

The network was computed using the COSBILab Graph software [21]. The data extraction was implemented in the Python programming language[1] and was derived from the library BioPython[2].

C. Detecting Crosstalk Networks Using Multi-Relational Clustering

Multi-relational data mining (MRDM) consists in looking for patterns that involve multiple tables (relations) from a relational database, that is different from most existing data mining approaches looking for patterns in a single data table [22]. With the rapid growth of public biological databases, MRDM can be widely applied to discover complex patterns through the rich relational structure and the mixed-up types of data Multi-relational clustering is the process of partitioning data objects into a set of clusters based on their similarity, utilising information in multiple relations. Due to the notable characteristics of multi-relational clustering, we employed this technique for inferring crosstalk modules.

The crosstalk modules do not separate but share some common CPs and their interactions. For this reason, the EM algorithm was used to perform the clustering task. The EM algorithm is a soft clustering method, this means that a data point always belongs to multiple clusters, and that a probability is calculated for each combination of a data point and a cluster. We applied the EM algorithm in the SQL Server 2008 Analysis Services (SSAS) since it allows the user to explore data in multiple relational tables [3]. In the EM algorithm, an initial cluster model is iteratively refined to fit the data and the probability that a data point exists in a cluster is produced. The fitness function is the *loglikelihood*

[1]http://www.python.org
[2]http://biopython.org
[3]http://technet.microsoft.com/en-us/library/ms174879.aspx

of the data giving the model. The process ends when the probabilistic model fits the data.

III. RESULTS

After computing the shortest paths between the p53 and NF-kB networks, we found 112 CPs joining at least one of these paths. The multiple data of the 112 CPs were extracted from four databases (the i2d database, the Uniprot database, the Reactome database, and the InterPro database) and integrated by the multi-relational scheme and manipulated in a corresponding relational database. This database consists of 2,086 data records of the abstract function feature, 265 ones for the biological process feature, and 761 ones for the protein domain feature.

Based on the combined data, the clusters of CPs were produced by running the EM algorithm. The number of cluster was set to 5 and other parameters were set to default values. Figure 2. shows the graphical view of the five produced clusters. Each cluster is presumably a crosstalk module of CPs between the NF-kB and p53. Networks 1 and 3 have the largest population with 28 CPs, followed by network 2 with 21 CPs, network 4 with 19 CPs and finally network 5 with 16 CPs. The darkness of the line in connecting the clusters in indicates the degree of similarity of the links. It is apparent from Figure 2. network 3 is strongly related to networks 1, 2 and 5 and loosely to network 4. There is no apparent relation between network 4 and networks 2 and 5.

To evaluate the performance of the EM algorithm, we also carried out the *k*-means clustering method on the same data set and compared the goodness of the two methods. The *k*-means clustering is a popular hard clustering method predominantly used to assign cluster membership. It works by minimizing the differences among items in a cluster while maximizing the distance between clusters.

The goodness of the two methods was calculated in term of case likelihood. Case likelihood is defined as the sum of cluster likelihood scores for each case (a case here is simply defined as a record of a CP in the Microsoft SQL Server database), divided by the number of cases in the partition.

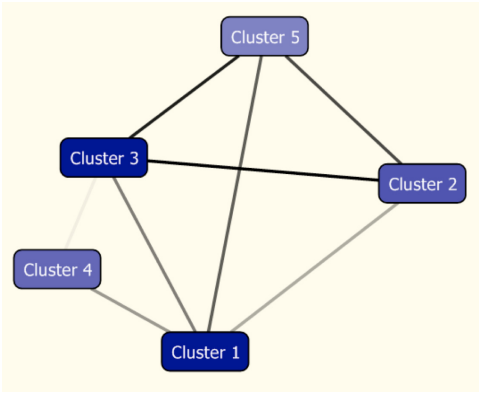

Figure 2. The graphical view of five clusters. Each clusters is visualized by a rectangle/a node and the shading out in the colour of the node represents the population of the cluster, the edge presents the association between two clusters, and the shading of the line that connects the individual networks to each other represents the strength in the similarity of the clusters.

TABLE I. CASE LIKELIHOOD OF THE EM ALGORITHM AND THE K-MEANS ALGORITHM BY 3-FOLD CROSS VALIDATION. THE FIRST COLUMN IS THE ID OF THE PARTITION AND THEN THE NUMBER OF PROTEIN IN THE PARTITION IS IN THE PARTITION SIZE. THE THIRD COLUMN IS THE CLUSTERING LIKELIHOOD CORRESPONDING TO EACH THE PARTITION. THE FINAL CLUSTERING LIKELIHOOD OF THE METHOD IS THE AVERAGE OF THREE ABOVE LIKELIHOODS.

EM algorithm			***k*-means algorithm**		
#Partition	*Partition Size*	*Clustering likelihood*	*#Partition*	*Partition Size*	*Clustering likelihood*
1	38	0.720	1	38	0.724
2	37	0.971	2	37	0.808
3	37	0.760	3	37	0.707
	Average	**0.817**		Average	0.746
	Standard Deviation	0.110		Standard Deviation	0.044

The case likelihood indicates how likely it is that a case belongs to a particular cluster. A case likelihood closer to 1 means that the case has a higher probability of occurring in this model. Otherwise, a number closer to 0 indicates that the case is less likely to occur in this model. We did the 3-fold cross validation and then calculated the average case likelihood. Table 1 shows the results of 3-fold cross validation done for the EM algorithm and the *k*-means algorithm respectively.

The EM algorithm achieved a considerable case likelihood of 81,7%, meaning that 81,7% of cases were possibly clustered appropriately. Compared to the *k*-means algorithm, which is 74,6%, the case likelihood of the EM algorithm is higher. The EM algorithm, being a soft clustering method, is able to distinguish overlapping parts among clusters, and here we show that it is potentially more sensible than the k-means algorithm in detecting the cross talking modules involved in the network interactions of the two systems NF- B and p53.

IV. DISCUSSION

The method validation of the method was performed by analysing the biological relevance of clustered networks. We examined the characteristics of each network described by the probability of features data of CPs in the network. Studying the biological processes involved in network 3, we found an high number of CPs for three processes 'Signalling in Immune system' (15.58%), 'Gene Expression' (12.27%), 'Cell Cycle Checkpoints' (10.27%). Investigating network 2, we found three main processes: 'Signalling by NGF' (20.41%), 'Cell Cycle, Mitotic' (16.28%), 'Signalling in Immune system' (15.58%). These results on the crosstalk networks of CPs showed the two systems (NF- B and p53) to essentially effect the signalling and cell cycle processes of the cell, a finding which is in agreement with known characteristics of the transcription factors [2], [5]. We then reconstruct these five produced networks. We extracted the interactions between two CPs inside the clustered network;

that is to say, other interactions with proteins outside the cluster were excluded.

Network 3 (with 28 nodes and 92 interactions) was the most complex one that consisted of many proteins sharing several common GO terms about biological processes, such as signal transduction, interspecies interaction between organisms, cell-cycle, many others. In network 2, more than 87% of the proteins are annotated as transcription proteins and involved in the regulation of the transcription process. Reviewing the functions of networks 2 and 3, the networks connecting the NF- B and p53 systems together (thereby acting as a cross talking system) should be regulatory networks and signal transduction networks.

Furthermore, we studied the association between the detected networks and diseases. For network 3, there were 32.05% of the proteins are annotated to mediate interactions between the host-virus and to influence resistance and recovery from viral infections. 31.70% of CPs in this network are associated with diseases causing mutation. CREB-binding protein in network 3 has histone acetylase activity; moreover, chromosomal aberrations involving the gene coding for CREBBP, could be linked to acute myeloid leukemias. In addition, histone acetyltrans- ferase p300 as the name implies has histone acetyltransferase activity through which and regulates transcription via chromatin remodelling. Defects in EP300 may be associated to epithelial cancer. Chromosomal aberrations involving the gene coding for EP300 may be a cause of acute myeloid leukemias. Protein p53 has been shown to be involved in many diseases, especially cancer.

This brief observation offers the chance to look into the relationship between NF- B and p53 pathways and restates their role in the pathophysiology of various diseases as a consequence of their disregulation. An important advantage of computational analysis is that it allows the detection of such modules by integrating complex and heterogeneous data. However, it is worth of note that using computational techniques alone,, we cannot verify the reactions underlying crosstalk but we are providing theoretical foundations for targeted experimental studies toward potential functional modules. Our results suggest some testable biological hypotheses and reveal an essential functionality of crosstalk networks as well as their biological relevance.

V. CONCLUSION

We have proposed a computational method to detect signalling crosstalk modules that probably function in transducing signals between the NF- B and p53 pathways. The advantage of this method is its ability to efficiently combine various relational data extracted from different data sources. The experimental results demonstrated that our proposed method performs well, especially when the EM algorithm is used. The biological analyses also showed the plausibility of these findings, and put forward several testable hypotheses to which the functional networks are involved in along with their associated human diseases. In future work, we would like to integrate more relational data into the scheme and study further the relationship between functional modules, diseases and therapeutic targets. The larger experiments with more pathways are prospective to reveal the whole complex mechanisms in systems biology.

REFERENCES

[1] E. Allen, J. Fetrow, L. W. Daniel, S. Thomas, and D. John, "Algebraic dependency models of protein signal transduction networks from time-series data," *J. of Theoretical Biology*, vol. 238, no. 2, pp. 317–330, 2006.

[2] J. Albert S. Baldwin, "The transcription factor NF-kappaB and human disease," *J Clin Inves*, vol. 101, no. 1, pp. 3–6, 2001.

[3] Y. Yang and A. M. Weissman, "Regulating the p53 system through ubiquitination," *Oncogene*, vol. 23, no. 11, pp. 3485–2106, 2004.

[4] G. A. Webster and N. D. Perkins, "Transcriptional Cross Talk between NF-kappa B and p53," *Mol. Cell. Biol.*, vol. 19, no. 5, pp. 3485–3495, 1999.

[5] K. H. Vousden and D. P. Lane, "p53 in health and disease," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 4, pp. 275–283.

[6] K. Schumm, S. Rocha, J. Caamano, and N. Perkins, "Regulation of p53 tumour suppressor target gene expression by the p52 NF-kappaB subunit," *The EMBO Journal*, vol. 25, p. 4820–4832, 2006.

[7] C. Priami, "Algorithmic systems biology," *Commun. ACM*, vol. 52, pp. 80–88, May 2009.

[8] N. J. Eungdamrong and R. Iyenga, "Modeling cell signaling networks," *Biology of the Cell*, vol. 96, no. 5, pp. 355–362, 2004.

[9] Y. Liu and H. Zhao, "A computational approach for ordering signal transduction pathway components from genomics and proteomics data." *BMC Bioinformatics*, vol. 5, no. 158, 2004.

[10] S. M. Gomez, S. Lo, and A. Rzhetsky, "Probabilistic Prediction of Unknown Metabolic and Signal-Transduction Networks," *Genetics*, vol. 159, no. 3, pp. 1291–1298, 2001.

[11] X. Zhao, R. Wang, L. Chen, and K. Aihara, "Automatic modeling of signal pathways from protein-protein interaction networks," in *The Sixth Asia Pacific Bioinformatics Conference*, 2008, pp. 287–296.

[12] Y. Li, P. Agarwal, and D. Rajagopalan, "A global pathway crosstalk network," *Bioinformatics*, vol. 24, no. 12, pp. 1442–1447, 2008.

[13] T. U. Consortium, "The Universal Protein Resource (UniProt) in 2010," *Nucleic Acids Research*, vol. 38, no. suppl 1, pp. D142–D148, 2010.

[14] K. Brown and I. Jurisica, "Unequal evolutionary conservation of human protein interactions in interologous networks," *Genome Biology*, vol. 8, no. 5, pp. R95+, May 2007.

[15] Croft D. et al., "Reactome: a database of reactions, pathways and biological processes." *Nucleic acids research*, November 2010.

[16] S. Hunter et al., "InterPro: the integrative protein signature database," *Nucleic Acids Research*, vol. 37, no. suppl 1, pp. D211–D215, 2009.

[17] A. E. Ihekwaba, P. T. Nguyen, and C. Priami, "Elucidation of functional consequences of signalling pathway interactions," *BMC Bioinformatics*, vol. 10, no. 370.

[18] B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network." *Nature*, vol. 408, no. 6810, pp. 307–310, November 2000.

[19] C. Cormen, T.H.and Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed.. Cambridge, MA The MIT Press, 2001.

[20] J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[21] R. Valentini and F. Jordan, "CoSBiLab Graph The network analysis module of CoSBiLab," *Environmental Modelling and Software*, vol. 25, no. 7, pp. 886 – 888, 2010.

[22] S. Dzeroski and N. Lavrac, Eds., *Relational Data Mining*. Springer, 2001.