Analyzing Incomplete Biological Pathways Using Network Motifs

Maya El Dayeh
Dept. of Computer Science and Engineering
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
meldayeh@lyle.smu.edu

Michael Hahsler
Dept. of Computer Science and Engineering
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
mhahsler@lyle.smu.edu

ABSTRACT
It is widely accepted that existing knowledge about the structure of many biological pathways is incomplete, and that uncovering missing proteins in a biological pathway can help guide targeted therapy, drug design, and drug discovery. Current approaches address the complex/pathway membership problem through identifying potentially missing proteins using probabilistic protein-protein interaction (PPI) networks. In this paper, we extend the idea of the pathway membership problem and define the *pathway completion problem*. In addition to finding possible protein candidates, this problem requires predicting the locations and connections of these proteins within a given incomplete pathway. We propose the use of network motifs to tackle the pathway completion problem. We present an algorithm which breaks down an incomplete pathway into a set of constituent motifs. The algorithm utilizes the proteins retrieved from a probabilistic PPI network to improve the motifs. Furthermore, our approach has the potential to improve solutions to the membership problem by better exploiting the local structures represented by network motifs. These new ideas are illustrated with a set of preliminary experiments.

Categories and Subject Descriptors
J.3 [**Life and Medical Science**]: Biology and Genetics.

General Terms
Algorithms, Experimentation.

Keywords
Pathway completion; network motifs; local structures; protein networks; pathway membership.

1. INTRODUCTION
Interactions among proteins accentuate many biological processes that are essential in providing functional and organizational support to any given organism. Studying and analyzing these interactions can pave the way to understanding diseases and their primary causes [3, 7]. Consequently, there is a push to advance the area of protein and protein interaction discovery, yet the knowledge of protein interactions is not complete [7, 4].

Since protein interactions underlie the complex biological models and functions of organisms, it is best to analyze these interactions in the context of a network of interacting proteins. This analysis is contrary to studying pair-wise interactions in isolation. In a protein-protein interaction (PPI) network, each node represents a protein and each edge represents an interaction between two proteins. A probabilistic PPI network is essentially a PPI network with weights on the edges. Each weight represents the probability of interaction among two proteins. Probabilistic PPI and PPI networks are utilized in a variety of areas to gain biological knowledge. Such areas include complex/pathway membership [2, 6, 7], predicting and assigning protein function and regulation [18], and pathway discovery [5].

Two or more interacting proteins bind together to form a protein complex. Biological pathways "are distinct, experimentally-validated sub-networks of proteins within the larger PPI network that interact with each other by well-defined mechanisms to regulate a specific biologic phenotype" [11]. From a Computer Science perspective, a protein complex forms a highly connected sub-graph while a biological pathway is a directed sub-graph.

The pathway membership problem is biologically motivated whereby biologists are confident that a large number of known pathways and their supporting evidence are incomplete [7]. We view the incompleteness of a pathway in terms of missing proteins from the pathway, incorrect connections among its proteins, or both. Given a set of proteins that constitute a pathway, the pathway membership problem, which is analogous to the complex membership problem, can be defined as the problem of uncovering and ranking a set of candidate proteins from a given probabilistic PPI network [7]. However, PPI networks suffer from a large number of false negatives (missing edges) and false positives (edges that should not be in the network) due to the nature of the experimental and predictive techniques used to discover protein interactions. Consequently, methods like random walks on graphs and network flow are suggested to extract pathway members from a probabilistic PPI network.

Current research, which has been done in the complex/pathway membership problem, is promising and focuses on protein

SAC'12, March 25-29, 2012, Riva del Garda, Italy.

membership, yet it does not address the issue of a candidate protein's location in a given incomplete pathway; it largely ignores structural information within the pathway. Therefore, in this paper we define a new problem that extends the membership problem by predicting the location of the new protein in the pathway. We call this problem the *pathway completion problem* and propose a method based on network motifs to tackle it. De novo signaling pathway reconstruction addresses a related problem; however, it is restricted to signaling pathways and concentrates on ordering proteins along such pathways [24].

Network motifs are sub-network patterns that exist in networks at frequencies significantly greater than expected [20]. Our method breaks down an incomplete pathway into a set of potentially incomplete motifs and searches a probabilistic PPI network for proteins that complete the motifs. We use a scoring method to find the best candidates to complete the pathway.

The rest of the paper is organized as follows. In Section 2, we cover background and previous work. In Section 3, we define the pathway completion problem and discuss network motifs. In Section 4, we formalize an algorithm for the pathway completion problem. In Section 5, we discuss some preliminary experiments and the datasets used. Finally in Section 6, we conclude and propose future work.

2. BACKGROUND

2.1 Probabilistic PPI Networks

With the emergence of large-scale experiments, protein interaction data saw an unmatched discovery of new interactions over a very short period of time. Such experiments include yeast-two-hybrid (Y2H) [12, 22] and tandem affinity purification-mass spectrometry (TAP-MS) [2, 11, 9]. Despite the abundance of interactions derived from large-scale experiments, further investigation has shown that such interaction data contains a notable number of false positives [7]. Since there is a major reliance on this type of data, research became directed towards assessing its quality before proceeding with analyzing it and incorporating it in any study [2]. Therefore, researchers have been integrating different types of biological knowledge and supporting evidence into the construction of protein-protein interaction (PPI) networks. Hence, probabilistic PPI networks emerged which are PPI networks with weighted edges. The weights on the edges represent the probability of interaction.

Table 1 summarizes information about three important probabilistic PPI networks. Most computational methods that deal with probabilistic PPI network construction are applied to the yeast genome. Yeast is a model organism and has been the target of numerous experiments and analyses that were performed to uncover interactions among its proteins [7].

Table 1. Yeast probabilistic networks and their properties

Yeast Network	Number of Proteins	Number of Interactions
Naïve Bayes [2]	3,112	12,594
ConfidentNet [16]	5,552	235,222
PIT-Network [13] (L = 300)	5,240	91,768

2.2 Complex Membership Methods

The complex/pathway membership problem has been addressed using several methods. Two of these methods focus on examining the complex membership problem [2, 6]. A third method deals with addressing both pathway and complex membership [7]. Probabilistic PPI networks are used in the membership problem because they provide a measure of how reliable an interaction is among two proteins. Next we describe these methods which we call complex membership methods.

2.2.1 Network Reliability

To determine how close two nodes are in a weighted network, one can look at the probability that some path of reliable edges lies between these two nodes at a given time. The idea of network reliability can be borrowed and applied to extract close proteins in a probabilistic PPI network. Asthana et al. use network reliability to tackle the complex membership problem [2]. Given a probabilistic PPI network and a protein complex, the method in [2] returns a set of candidate proteins from the probabilistic PPI network ranked according to the probability that each is a member of the given complex. If there is at least one path of interacting proteins that connects a protein from the network to a protein in the complex, then this protein is added to the set of candidate proteins.

2.2.2 Random Walk on a Graph

The random walk algorithm in [7] is applied to probabilistic PPI networks to extract a list of candidate proteins. The random walker starts at a designated start node (a known member of the complex), and selects an edge from the possible edges to transition to an adjacent node. The walker repeats the edge selection and transition process at every time increment until it decides to restart the walk (given by a restart probability). Good candidate proteins are visited often by the walker.

2.2.3 Net-Flow

Net-Flow is a technique based on flow networks and is designed by Camoglu et al. [6] to address the limitations of network reliability. To calculate the reliability between the proteins in the complex and those in the network, the role of "sink" is assigned to the network proteins and that of "source" is given to the complex members. Moreover, each edge is assigned a capacity of one and a cost equivalent to the interaction probability. Finally, maximum flow is calculated using linear programming to identify good candidates.

2.2.4 Limitations of Membership-based Approaches

The methods discussed above were applied mostly to MIPS (Munich Information Center for Protein Sequences) benchmark protein complexes [19]. The methods work well since protein complexes form highly connected sub-graphs; however, biological pathways are directed sub-graphs. Thus, the above three methods are not expected to work well with pathways. Moreover, Can et al. tested the random walk method on a small set of KEGG benchmark pathways in addition to the MIPS complexes [14]. Yet, none of the above methods address the pathway completion problem or exploit the structures available in pathways as proposed in this paper. Later, in Section 4, we develop a technique geared towards pathways and the completion problem.

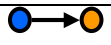
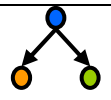
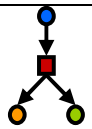
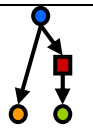
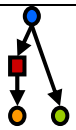
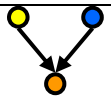
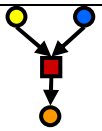
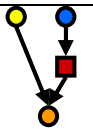
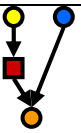
3. MOTIFS AND THE PATHWAY COMPLETION PROBLEM

The first contribution of our work is to extend the scope of the pathway membership problem to uncover both membership and the location of a candidate protein in a given incomplete pathway. We call it the pathway completion problem. The biological significance of this approach is to extract complete knowledge which allows for better drug design and targeted therapy. Given a graph that represents a pathway, the pathway completion problem can be defined as:

1. Find a ranked list of candidate proteins, and
2. Predict their locations and connections within a given incomplete pathway.

Devising a solution to the pathway completion problem requires a different approach compared to those found in the current literature, which are more applicable to highly connected sub-graphs (protein complexes). Here we present a solution based on network motifs [20]. A number of algorithms have been designed to extract motifs from biological networks [8] such as PPI [23], signal transduction [10], metabolic [15], and transcription-regulation [1, 17, 20, 21, 23]. Shen-Orr et al. [21] list three highly significant motifs that are found in the transcription-regulation networks of Escherichia coli. These significant motifs are: feed forward loop (FFL), single input module (SIM), and dense overlapping regulons (DOR). Lee et al. [17] refer to the DOR motif as multiple input motif (MIM), and they highlight three additional motifs which are regulator chain, auto-regulation, and multi-component loop. In this paper, we focus on a small subset of motifs called linear, single input, and multiple input. In Table 2, we show these motifs (circles are proteins and arrows indicate interaction). Also in Table 2, we present possible completion with an additional protein (square node). Other types of motifs (e.g., forks with more than 2 inputs/outputs) will be incorporated in future research. The pathway completion problem is addressed by uncovering complete motifs from a probabilistic PPI network with better scores than the original motifs found in the pathway.

Table 2. Motifs before and after proposed completion.

Motif	Structure	Proposed Complete Structure		
Linear				
Single Input				
Multiple Input				

4. THE FIT AND COMPLETE ALGORITHM

The purpose of the *Fit and Complete* algorithm is to generate a ranked list of proposed complete motifs (i.e. candidate proteins with their locations in the pathway) to complete the pathway in question. The flow chart of the Fit and Complete algorithm is shown in Figure 1.

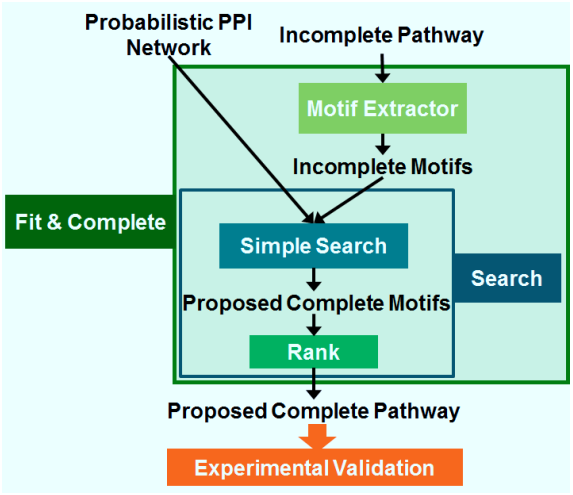


Figure 1. Flow chart of the Fit and Complete algorithm.

The *Motif Extractor* method takes the incomplete biological pathway as input and generates its constituent motifs. The motifs act as input to the *Search* method, along with the probabilistic PPI network. The Search method implemented in this paper uses the unrealistic assumption that a probabilistic PPI network is complete; it includes all proteins and interactions. However, we will discuss a more realistic case, which is having an incomplete PPI network, in Section 6.

In Figure 2, we formalize the Fit and Complete algorithm which takes the probabilistic PPI network $G = (V, E)$ and the incomplete pathway $G' = (V', E')$ as input. In step 1, the algorithm generates L, the set of linear motif sub-graphs of G' (i.e. it breaks down G' into its constituent motifs). Then for each motif l in L and for each edge e in l, the algorithm locates the original motif in G (steps 4 and 5), and computes the score of the motif using the weight of e in G (step 6). We will discuss how the scores are computed below. In step 7, the Search method is called and it returns the ordered set L^c of the complete motifs and the ordered set S^c of their scores. In step 8, the set of original motif scores S_L, in addition to L^c and S^c, are returned.

Figure 3 shows the details of the *Search* method. The method takes the weighted graph G and the linear motif l as input. In steps 1 and 2, the empty ordered sets L^c and S^c are created. In step 3, let v_1 and v_2 be the vertices (nodes) of l. In step 4, the method looks at the set of first neighbors N of v_1 and v_2 in the probabilistic PPI network to extract possible protein candidates; hence, possible complete motifs. For each neighbor n in N, the method calculates the score of the complete motif using the edge weights in G (steps 5 to 9). Finally in step 10, L^c and S^c are returned. Now we compare the ordered sets L^c and S^c to S_L to see if the candidate proteins N improve the scores of the original motifs.

To score the quality or strength of a motif, we currently use the average, minimum, or maximum scores of the weights on the

edges retrieved from the probabilistic PPI network. Using the average to assess the quality of the complete motif allows all edge weights to play a role in the scoring and ranking process. Considering only the minimum requires all edge weights to be strong while using only the maximum ignores weak edges. Any scoring function which utilizes interaction scores and/or the motif structure can also be used. Finding a biologically motivated score is a part of our future research.

Input: weighted graph, $G = (V,E)$ // *probabilistic PPI network*
directed graph, $G' = (V', E')$ // *incomplete pathway*
(1) Generate L, the set of linear motif sub-graphs of G'
(2) For each l in L do // for each motif in L
(3) For each edge e in l
// compute the score of the original motif
(4) Find w_e in G, the weight of the corresponding edge in G to e
(5) Add w_e to l, the edge weight of e
(6) S_L = scoreMotif(l) // returns s_l the score for motif l & adds it to S_L the set of original motif scores
(7) (L^c, S^c) = Search(G, l) // finds L^c the ordered set of possible complete motifs *and* S^c the ordered set of their scores
(8) Return (S_L, S^c, L^c)

Figure 2. The Fit and Complete algorithm.

Input: weighted graph, $G = (V,E)$ // *probabilistic PPI network*
motif l *in* L
(1) Let $L^c = \Phi$, the empty ordered set of possible complete motifs
(2) Let $S^c = \Phi$, the empty ordered set of their scores
(3) Let v_1 and v_2 be the two vertcies of l
(4) Find N the set of all common first neighbors of v_1 and v_2 of l in G
(5) For each n in N
(6) Create motif l^c with edges $\{(v_1, n), (n, v_2)\}$
(7) Find the weights of the edges in G
(8) Add l^c to L^c // l^c is the complete motif
(9) S^c = scoreMotif(l^c) // returns $s_l{}^c$ the score of l^c
(10) Return (L^c, S^c)

Figure 3. The Search method.

5. EXPERIMENTS AND RESULTS

To illustrate our approach, we applied the Fit and Complete algorithm to ConfidentNet and to five yeast KEGG pathways. ConfidentNet has 5,552 proteins and 235,222 probabilistic interactions. Table 3 lists the five pathways used and the number of motifs of each type (i.e. linear, multiple input, and single input) identified by Motif Extractor. An example of each type of motif uncovered by our algorithm from the yeast MAPK pathway is shown in Figure 4. We see an improvement in the scores most of the time as demonstrated in the figure. For instance, in Figure 4-A, we show a multiple input motif and the corresponding minimum, maximum, and average scores, which are 3.637, 6.213, and 4.925 respectively. After retrieving the first neighbor proteins of the motif from ConfidentNet, our algorithm computed the resulting minimum, maximum, and average scores of the proposed complete motifs. Figure 4-A shows two examples of the proposed complete motifs. The two candidate proteins, BEM4 and BUD6, improve the scores of the original motif most of the time. In the case of BEM4, all scores increase and the new protein would be considered a good candidate. For BUD6, the maximum score increases, but the minimum and average scores decrease because a weak link is introduced. Results for an example of a linear motif and of a single input motif are displayed in Figure 4-B and -C respectively.

Table 3. Five yeast KEGG pathways used in our analysis.

Pathway	Linear	Multiple Input	Single Input
Endocytosis	19	5	4
Cell Cycle	63	12	13
Regulation of Autophagy	14	1	2
Meiosis	108	27	20
MAPK Pathway	65	14	11

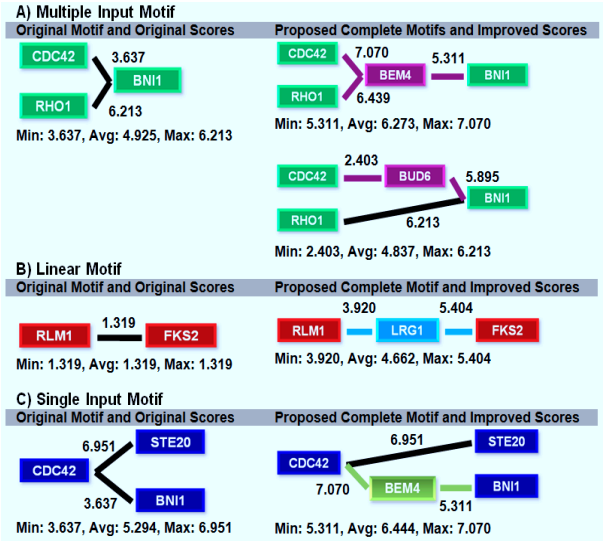


Figure 4. Investigation of the yeast MAPK pathway shows examples of possible proposed complete motifs and their scores in comparison to the scores of the original respective motifs: A) Multiple input B) Linear C) Single input.

In Figure 5, we plot the minimum, maximum, and average scores of three motifs retrieved by our method from the yeast MAPK pathway. The thick lines shown in red connect the scores of the original motif. In the case of Motif 14, we observe a small number of proposed complete motifs with improved scores over the scores of the original motif. With respect to Motif 17, none of the scores of the proposed complete motifs show improvement over the scores of the original motif. Lastly for Motif 24, all the scores of the proposed complete motifs improve the scores of the original motif. In Figure 6, we provide minimum and average histogram plot of the number of all motifs retrieved by our method and their

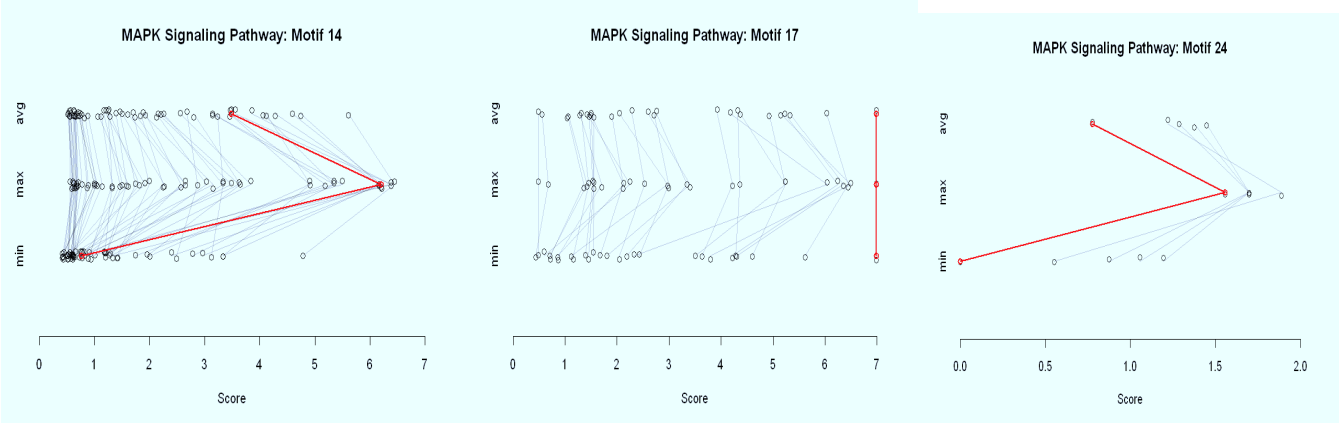


Figure 5. Examples of original motif scores vs. proposed complete motifs scores found by our algorithm.

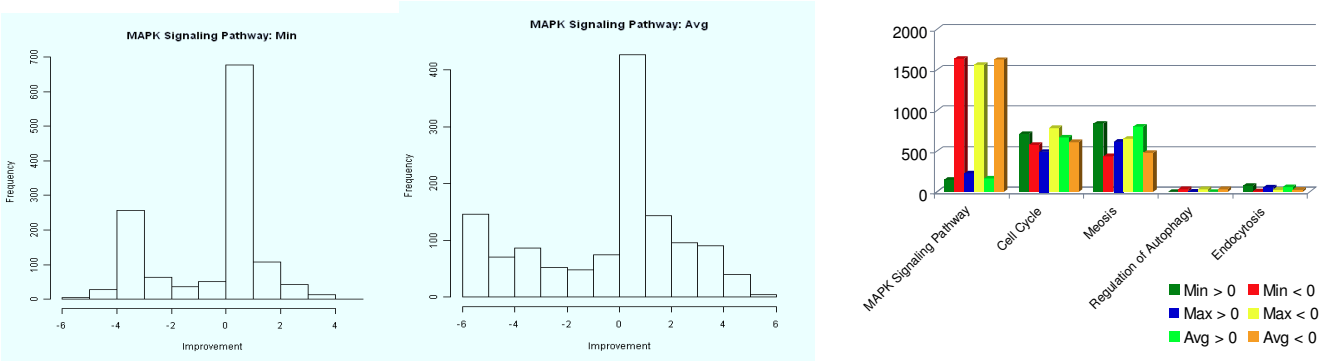


Figure 6. Histograms and summary statistics.

improvement. The improvement is computed as the difference between the score of the original motif and the score of the proposed complete motif. For example, we find that there are around 16 proposed complete motifs with significant improvement in the minimum and average scores over the original scores. We believe that such proposed complete motifs are interesting and should be studied further since they suggest candidate proteins with their locations to complete an incomplete pathway. In Figure 6 (histogram to the right), we summarize the results we obtained for all the pathways we investigated. We show for each pathway the number of proposed complete motifs with minimum, maximum, and average improvement above and below 0. In the case of the yeast MAPK pathway, we observe that the number of the proposed complete motifs with no improvement is much higher than those with improvement. On the other hand, in the case of the Endocytosis pathway, we find that the number of proposed complete motifs with improved scores is higher than those which show no improvement.

6. CONCLUSION AND FUTURE WORK

Our first contribution in this paper is extending the pathway membership problem and defining a completely new problem, which we call pathway completion. This problem is hard since it requires the prediction of the location of candidate proteins in the pathway. Our second contribution is proposing to solve pathway completion using network motifs. This method exploits the local structures represented by network motifs. Although the presented technique produces promising results, the main shortcoming is that it assumes that the probabilistic PPI network contains complete information; PPI networks suffer from a large number of false negatives and false positives in the form of missing proteins and edges. However, this can be remedied by using a search that is based on random walks on graphs which will be our next step. Other future research topics, which will address the limitations of our method, include: edge deletion to handle pathways which have incorrect edges, dealing with more complex motif structures, and developing more suitable scoring techniques.

7. ACKNOWLEDGEMENTS

This research was partially supported by research grant no. R21HG005912 from the National Human Genome Research Institute.

8. REFERENCES

[1] Alon, U. Network motifs: theory and experimental approaches. *Nature*, 8, 6 (2007), 450-461.

[2] Asthana, S., King, O. D., Gibbons, F. D., and Roth, F. P. Predicting protein complex membership using probabilistic network reliability. *Genome Res.*, 14, 6 (Jun. 2004), 1170-1175.

[3] Bader, D. and Madduri, K. A graph-theoretic analysis of the human protein-interaction network using multicore parallel algorithms. *Parallel Computing*, 34, 11 (Nov. 2008), 627-639.

[4] Bader, G. D. and Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4, 1 (Jan. 2003), 2.

[5] Bader, J. S. Greedily building protein networks with confidence. *Bioinformatics*, 19, 15 (Oct. 2003), 1869-1874.

[6] Camoglu, O., Can, T., and Singh, A. K. Accurate and scalable techniques for the complex/pathway membership problem in protein networks. *Advances in Bioinformatics*, 2009.

[7] Can, T., Camoglu, O., and Singh, A. K. Analysis of protein-protein interaction networks using random walks. In *Proceedings of the 5th international workshop on Bioinformatics* (*BIOKDD '05)*, Chicago, IL, Aug. 2005, 61-68.

[8] Ciriello, G. and Guerra, C. A review on models and algorithms for motif discovery in protein-protein interaction networks. *Briefing in Functional Genominc & Proteomics*, 7, 2 (Mar. 2008), 147-156.

[9] Gavin, A.-C. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M. M., Cruciat, C.-M. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415, 6868 (Jan. 2002), 141-147.

[10] Han, Z., Vondriska, T.M., Yang, L., MacLellan, W. R., Weiss, J. N., and Qu, Z. Signal transduction network motifs and biological memory. *J Theor Biol*, 246, 4 (June 2007), 755-761.

[11] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. *Nature*, 415, 6868 (Jan. 2002), 180-183.

[12] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Nat. Acad. Sci. U.S.A.*, 98, 8 (Apr. 2001), 4569-4574.

[13] Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F., and Gerstein, M. A bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, 302, 5644 (Oct. 2003), 449-453.

[14] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res.*, 30, 1 (Jan. 2002), 42-46.

[15] Koyuturk, M., Grama, A., and Szpankowski, W. An efficient algorithm for detecting frequent subgraphs in biological networks. *Bioinformatics (Oxford, England)*, 20 suppl 1, 1 (Aug. 2004), 200-207.

[16] Lee, I., Date, S. V., Adai, A. T., and Marcotte, E. M. A probabilistic functional network of yeast genes. *Science*, 306, 5701 (Nov. 2004), 1555-1558.

[17] Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., and Young, R.A. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298, 5594 (Oct. 2002), 799-804.

[18] Letovsky, S. and Kasif, S. Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics*, 19, 1 (Jul. 2003), 197-204.

[19] Mewes, H. W., Amid C., Arnold R., Frishman, D., Güldener, U., Mannhaupt, G., Münsterkötter, M., Pagel, P., Strack, N., Stümpflen, V., Warfsmann, J., and Ruepp, A. MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.*, 32, suppl 1 (Jan. 2004), D41-D44.

[20] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. Network motifs: simple building blocks of complex networks. *Science*, 298, 5594 (Oct. 2002), 824-827.

[21] Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. Network motifs in the transcriptional regulaiton network of Escherichia coli. *Nature Genetics*, 31, 1 (Apr. 2002), 64-68.

[22] Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403, 6770 (Feb. 2000), 623-627.

[23] Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc. Nat. Acad. Sci. U.S.A.*, 101, 16 (Apr. 2004), 5934-5939.

[24] Zhu, D., Rabbat, M.G., Hero, A.O., Nowak, R. and Figueiredo, M. 2007. De Novo Reconstructing Signaling Pathways from Multiple Data Sources. In *New Research in Signaling Transduction*, B. R. Yanson, Eds. Nova Publishers, Hauppauge, New York, 141-163.