

Function-Function Correlated Multi-Label Protein Function Prediction over Interaction Networks

Hua Wang, Heng Huang*, and Chris Ding

Department of Computer Science and Engineering,
University of Texas at Arlington, Arlington, TX 76019
huawangcs@gmail.com, {heng, chqding}@uta.edu

Abstract. Many previous computational methods for protein function prediction make prediction one function at a time, fundamentally, which is equivalent to assume the functional categories of proteins to be isolated. However, biological processes are highly correlated and usually intertwined together to happen at the same time, therefore it would be beneficial to consider protein function prediction as one indivisible task and treat all the functional categories as an integral and correlated prediction target. By leveraging the function-function correlations, it is expected to achieve improved overall predictive accuracy. To this end, we develop a novel network based protein function prediction approach, under the framework of multi-label classification in machine learning, to utilize the function-function correlations. Besides formulating the function-function correlations in the optimization objective explicitly, we also exploit them as part of the pairwise protein-protein similarities implicitly. The algorithm is built upon the Green’s function over a graph, which not only employs the global topology of a network but also captures its local structural information. We evaluate the proposed approach on *Saccharomyces cerevisiae* species. The encouraging experimental results demonstrate the effectiveness of the proposed method.

Keywords: Protein Function Prediction, Green’s Function, Multi-Label Classification, Function-Function Correlations.

1 Introduction

Many existing *in silico* methods for protein function prediction typically make prediction one function at a time, fundamentally. This turns the problem into a convenient form to use existing machine learning algorithms, which, however, abstracts away the function correlations due to neglecting the fact that most biological functions are interdependent from one another. For example, “Transcription” and “Protein Synthesis” usually appear together [13], one after another, *i.e.*, they tend to appear in the processes of a same protein. Therefore, if a protein

* Corresponding author.

is known to be annotated with function “Transcription”, we would have high confidence to annotate the same protein with function “Protein Synthesis” as well. In other words, the *function-function correlations* convey valuable information to understand the biological processes, which provide a potential opportunity to improve the overall accuracy of protein function prediction. However, how to effectively exploit function-function correlations remains a challenge for devising computational methods for protein function prediction. In this study, we tackle this new, yet important, problem by placing protein function prediction under the framework of multi-label classification, an emerging topic in machine learning, to develop a new graph based protein function prediction method to take advantage of function-function correlations.

1.1 Network Based Protein Function Prediction

Recent availability of protein interaction networks for many genomic species has stimulated the development of network based computational methods for protein function prediction. Typically, an interaction network is first modeled as a graph, with the vertices representing proteins and the edges representing the detected protein-protein interactions (PPI), then a graph based statistical learning method is used to infer putative protein functions.

The most straightforward method using network data to predict protein function determines the putative functions of a protein from the known functions of its neighboring proteins on a PPI network [16,9,2], which, though, only leverages the local information of a network. Lately, researchers used the global optimization approaches to improve the protein function predictions by taking into account the full topology of networks [20,11,14]. All these approaches can be summarized as the following common scheme: (1) compute a set of ranking lists, and (2) make prediction using certain thresholds on the ranking lists. In step 1, which is the most critical part of algorithms, these existing methods compute the ranking lists one function at a time, therefore they do not take into account the interrelationships among the functions. Network based approaches using other models for protein function prediction are surveyed in [17].

The aforementioned graph-based protein function prediction methods bank on two assumptions: local consistency and global consistency, which are the exact foundations of label propagation based learning approaches in machine learning. This motivates us to formulate protein function prediction over a PPI network as a label propagation problem on a graph. Among existing label propagation methods, we choose to develop our new method from the Green’s function approach [5,23] due to its clear intuition and demonstrated effectiveness in other real-world applications.

1.2 Multi-Label Correlated Protein Function Prediction

Because a protein is usually observed to play several functional roles in different biological processes within an organism, it is natural to annotate it with multiple functions. Thus, protein function prediction is an ideal example of *multi-label*

classification [21,22,23,24,25] in machine learning. Multi-label classification is an emerging topic driven by the advances of modern technologies in recent years, in which each object could belong to more than one class. In contrast, in traditional single-label classification, a data object may belong to one and only one class. Placing protein function prediction under the framework of multi-label classification, we use the Green’s function approach to integrate the function-function correlations from theory of reproducing kernel Hilbert space (RKHS). Besides incorporating the function-function correlations as a regularizer in the optimization objective explicitly, we also take advantage of them as part of the pairwise protein similarities implicitly.

2 Methods

In this section, we propose a novel Function-function Correlated Multi-Label (FCML) approach using the Green’s function on a graph to predict protein functions, which incorporates the function-function correlations by two levels: one from the function perspective to formulate the function-wise similarities explicitly in the optimization objective (Section 2.3), and the other from protein similarity perspective using function assignments to model the function correlations implicitly (Section 2.4).

2.1 Notations and Problem Formalization

In protein function prediction, given K biological functions and n proteins, each protein x_i is associated with a set of labels represented by a function assignment indication vector $\mathbf{y}_i \in \{-1, 0, 1\}^K$ such that $\mathbf{y}_i(k) = 1$ if protein x_i has the k -th function, $\mathbf{y}_i(k) = -1$ if it does not have the k -th function, and $\mathbf{y}_i(k) = 0$ if its function assignment is not yet known in a priori. Given l annotated proteins $\{(x_1, \mathbf{y}_1), \dots, (x_l, \mathbf{y}_l)\}$ where $l < n$, our task is to predict functions $\{\mathbf{y}_i\}_{i=l+1}^n$ for the unannotated proteins $\{x_i\}_{i=l+1}^n$. We write $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n]^T$, and $Y_l = [\mathbf{y}_1, \dots, \mathbf{y}_l]^T = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(K)}]$, where $\mathbf{y}^{(k)} \in \mathbb{R}^l$ is a class-wise function assignment vector. We also define $F = [\mathbf{f}_1, \dots, \mathbf{f}_n]^T \in \mathbb{R}^{n \times K}$ as the decision matrix for prediction, and $\{\mathbf{f}_i\}_{i=l+1}^n$ include the decision values for prediction.

We formalize a protein interaction network as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The vertices \mathcal{V} correspond to the proteins $\{x_1, \dots, x_n\}$, and the edges \mathcal{E} are weighted by an $n \times n$ similarity matrix W with W_{ij} indicating the similarity between x_i and x_j . In the simplest case, W is the adjacency matrix of the PPI graph where $W_{ij} = 1$ if proteins x_i and x_j interact, and 0 otherwise. In this work, W is computed in Eq. (10) to incorporate more useful information.

2.2 Protein Function Prediction Using the Green’s Function over a Graph

In this section, we first briefly review the Green’s function approach for label propagation over a graph, from which we will derive the proposed FCML method in later subsections.

Given a graph with edge weights W , its combinatorial Laplacian is defined as $L = D - W$ [3], where $D = \text{diag}(W\mathbf{e})$ and $\mathbf{e} = [1\dots 1]^T$. The Green’s function over the graph is defined as the inverse of L with zero-mode discarded, which is computed as following [5]:

$$G = L_+^{-1} = \frac{1}{(D - W)_+} = \sum_{i=2}^n \frac{\mathbf{v}_i \mathbf{v}_i^T}{\lambda_i}, \tag{1}$$

where L_+^{-1} and $(D - W)_+$ indicates that the zero eigen-mode of the concerned matrix is discarded, and $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are the eigenvalues of L and \mathbf{v}_i are the corresponding eigenvectors. Because the Green’s function G in Eq. (1) is a kernel [5], from the theory of RKHS, the Green’s function approach is to minimize the following objective [5]:

$$J_0(F) = (1 - \mu) \|F - Y\|^2 + \mu \text{tr}(F^T \mathcal{K}^{-1} F), \tag{2}$$

where $\mathcal{K} = G$ is the kernel, $\mu \in (0, 1)$ is a constant to control the smoothness regularizer $\text{tr}(F^T \mathcal{K}^{-1} F)$, and $\text{tr}(\cdot)$ denotes the trace of a matrix.

Thus, we assign functions to unannotated proteins by [5]:

$$\mathbf{y}_i = \text{sign}(\mathbf{f}_i), \quad l < j \leq n, \quad \text{where } F = GY. \tag{3}$$

We name Eq. (3) as simple Multi-Label Green’s Function (MLGF) approach, beyond which we will propose a novel function-function correlated multi-label Green’s function approach to leveraging function-function correlations.

2.3 Function-Function Correlated RKHS Approach for Multi-Label Classification

Although the function-function correlations are useful to infer putative functions of unannotated proteins, MLGF approach defined in Eq. (3) does not exploit them because it treats the biological functions of interest as isolated. In multi-label scenarios, however, we concentrate on making use of the function-function correlations, which could be defined as $C \in \mathbb{R}^{K \times K}$ using cosine similarity as following [22,23,24,25]:

$$C_{kl} = \cos(\mathbf{y}^{(k)}, \mathbf{y}^{(l)}) = \frac{\langle \mathbf{y}^{(k)}, \mathbf{y}^{(l)} \rangle}{\|\mathbf{y}^{(k)}\| \|\mathbf{y}^{(l)}\|}. \tag{4}$$

Following [23], we expect to maximize $\text{tr}(FCF^T)$. In order to make connection with the theory of RKHS, instead of directly using F , we use kernel assisted decision matrix $\mathcal{K}^{-\frac{1}{2}}F$, which leads to the following objective to maximize [23]:

$$J_C(F) = \text{tr}\left(\mathcal{K}^{-\frac{1}{2}}FCF^T\mathcal{K}^{-\frac{1}{2}}\right). \tag{5}$$

Combining Eq. (5) with the original RKHS objective in Eq. (2), we minimize the following objective:

$$J(F) = \beta \|F - Y\|^2 + \text{tr}(F^T \mathcal{K}^{-1} F) - \alpha \text{tr}\left(\mathcal{K}^{-\frac{1}{2}}FCF^T\mathcal{K}^{-\frac{1}{2}}\right), \tag{6}$$

where $\alpha \in (0, 1)$ balances the the two objectives, and $\beta = \frac{1-\mu}{\mu}$.

Differentiating J with respect to F , we have:

$$\begin{aligned} \frac{\partial J}{\partial F} &= 2\beta(F - Y) + 2\mathcal{K}^{-1}F - 2\alpha\mathcal{K}^{-1}FC = 0 \implies \\ F &= \frac{1}{\beta I + \mathcal{K}^{-1}}\beta Y + \alpha \frac{1}{\beta I + \mathcal{K}^{-1}}\mathcal{K}^{-1}FC. \end{aligned} \quad (7)$$

Because β is usually very small in typical empirical settings, we have:

$$\frac{F}{\beta} = \mathcal{K}Y + \alpha \frac{F}{\beta}C \implies \tilde{F} = \mathcal{K}Y + \alpha\tilde{F}C = GY + \alpha\tilde{F}C, \quad (8)$$

where $\tilde{F} = \frac{F}{\beta}$. Thus, we have

$$\tilde{F} = GY (I - \alpha C)^{-1}. \quad (9)$$

We name Eq. (9) as our proposed Function-function Correlated Multi-Label (FCML) approach for protein function prediction.

2.4 Correlation Augmented Interaction Network

Traditional network based protein function prediction approaches only employ biological interaction networks obtained from experimental data such as those from high-throughput technologies. When considering protein function prediction as a multi-label classification problem, we can build a computational interaction network $W_L \in \mathbb{R}^{n \times n}$ from label assignment perspective. As one of our contribution, we make use of this new computational interaction network and propose a *correlation augmented interaction network* as follows:

$$W = W_{\text{Bio}} + \gamma W_L, \quad (10)$$

where W_{Bio} is the biological interaction network, which is same as those in existing approaches. In Eq. (10), γ controls the relative importance of W_L , which is empirically selected as $\gamma = \frac{\sum_{i,j,i \neq j} W_{\text{Bio}}(i,j)}{\sum_{i,j,i \neq j} W_L(i,j)}$.

The true power of the correlation augmented interaction network construction scheme defined in Eq. (10) lies in that, the original biological similarities among proteins are augmented by the function assignment similarities. As a result, label propagation pathways over a graph are reinforced. With this interaction network construction scheme, the correlations among the functional categories are encoded into the graph weights, such that the resulted hybrid graph can be directly used in existing methods to enhance their prediction performance. In this paper, we use W defined in Eq. (10) to compute the Green's function in Eq. (1).

Protein-Protein Similarity from Function Assignments (W_L). Because multiple functions could be assigned to one single protein, the overlap between the function assignments of two proteins can be used to evaluate their similarity. The more functions shared by two proteins, the more similar they are. Consequently, besides the class membership indications, the label assignment vector \mathbf{y}_i is enriched with characteristic meaning, which thereby can be used as an attribute vector to characterize protein x_i . Using cosine similarity, the function assignment similarity between two proteins is computed as:

$$W_L(i, j) = \cos(\mathbf{y}_i, \mathbf{y}_j) = \frac{\langle \mathbf{y}_i, \mathbf{y}_j \rangle}{\|\mathbf{y}_i\| \|\mathbf{y}_j\|}. \quad (11)$$

Note that, the entries of W_L measures *individual protein - individual protein relationships*; whereas the entries of C defined in Eq. (4) assesses the *function category - function category correlations*.

Our task in protein function prediction is to assign functions to unannotated proteins upon annotated ones. In order to compute W_L , however, we need function assignments of all the proteins, including those annotated and unannotated. Therefore, we first initialize unannotated proteins through majority voting [16] approach, which makes prediction using the top three frequent functions appearing one protein's interacting partners.

Biological Protein-Protein Similarity (W_{Bio}) and Multi-Source Integration. W_{Bio} in Eq. (10) computes the protein-protein similarity from biological experimental data, which is same as existing works and could integrate multiple experimental sources. Let $W^{(1)}$ be the graph built from BioGRID PPI data [10,7], $W^{(2)}$ be that from synthetic lethal data [19], $W^{(3)}$ be that from gene co-expression data [6], $W^{(4)}$ be that from gene regulation data [8], etc., W_{Bio} is computed as [15]:

$$W_{\text{Bio}}(i, j) = 1 - \prod_k \left[1 - r^{(k)} W^{(k)}(i, j) \right], \quad (12)$$

where $r^{(k)}$ is estimated reliabilities of the corresponding network by Expression Profile Reliability (EPR) index [4]. Eq. (12) reflects the fact that interactions detected in multiple experiments are generally more reliable than those detected by a single experiment [12].

Because in reality the overlap among different biological networks typically is very small, and the BioGRID PPI network data are fairly comprehensive, in this work, we set $W_{\text{Bio}} = W^{(1)}$, where $W^{(1)}(i, j) = 1$ if protein x_i and x_j interact, and 0 otherwise.

3 Materials and Data Sets

Two types of data are involved in the experimental evaluations for protein function prediction: function annotation data and PPI data.

The functional catalogue (FunCat) [13] is a project under the Munich Information Center for Protein Sequences (MIPS), which is an annotation scheme for the functional description of proteins from prokaryotes, unicellular eukaryotes, plants and animals. Taking into account the broad and highly diverse spectrum of known protein functions, FunCat of version 2.1 consists of 27 main functional categories, 17 of which are involved in annotating *Saccharomyces cerevisiae*. Although there are still other protein annotation systems such as the Gene Ontology [1], we use the FunCat annotation system due to its clear tree-like hierarchical structure. All 17 level-1 biological functions are listed in Table 1.

Table 1. Function IDs and names by FunCat scheme version 2.1

ID	Name
'01'	Metabolism
'02'	Energy
'10'	Cell Cycle and Dna Processing
'11'	Transcription
'12'	Protein Synthesis
'14'	Protein Fate (Folding, Modification, Destination)
'16'	Protein With Binding Function or Cofactor Requirement (Structural or Catalytic)
'18'	Regulation of Metabolism and Protein Function
'20'	Cellular Transport, Transport Facilitation and Transport Routes
'30'	Cellular Communication/Signal Transduction Mechanism
'32'	Cell Rescue, Defense and Virulence
'34'	Interaction with the Environment
'38'	Transposable Elements, Viral and Plasmid Proteins
'40'	Cell Fate
'41'	Development (Systemic)
'42'	Biogenesis of Cellular Components
'43'	Cell Type Differentiation

The protein-protein interaction data can be downloaded from the BioGRID database [18] and we focus on the *S. cerevisiae*. By using the BioGRDI database of version 2.0.45 and removing the proteins connected by only one PPI, we end up with 4299 proteins annotated by FunCat annotation scheme with 72624 PPIs, as well as 1997 unannotated proteins.

4 Results and Discussions

In this section, using the PPI data from BioGRID database [18] and FunCat annotation scheme [13] on *S. cerevisiae* data, we evaluate our new FCML approach by applying it in protein function prediction, where standard *precision* and *F1 score* are used as statistical metrics.

4.1 Evaluation of the Function-Function Correlations

Because the function-function correlations are one of the most important mechanism to improve the prediction performance in our proposed approach, we first evaluate its correctness. Using the FunCat 2.1 annotation data set for *S. cerevisiae* genome, the function-function correlations defined in Eq. (4) are visualized in Figure 1. The high correlation value between functions “40” (Cell Fate) and “43” (Cell Type Differentiation) depicted in this figure shows that they are highly correlated. In addition, as shown in this figure some other function pairs are also highly correlated, such as functions “11” (Transcription) and “16” (Protein With Binding Function or Cofactor Requirement), “18” (Regulation of Metabolism and Protein Function) and “30” (Cellular Communication/Signal Transduction Mechanism), *etc.* All these observations perfectly comply with the biological facts, which firmly confirm the correctness of our formulation of the function-function correlations defined in Eq. (4) from biological perspective.

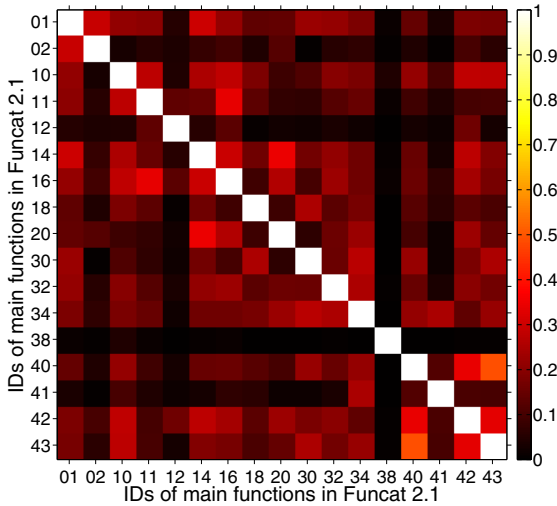


Fig. 1. The visualization of the correlation matrix defined in Eq. (4) among the 17 main functional categories in FunCat 2.1 annotated to *S. cerevisiae* genome

4.2 Improved Function Prediction in Cross-Validation

We compare the performances of the proposed Multi-label Green’s Function (MLGF) approach and Function-function Correlated Multi-Label (FCML) approach to several benchmark computational methods for protein function prediction, including Majority Voting (MV) approach [16], Global Majority Voting (GMV) approach [20], χ^2 approach [9], and Functional Flow (FF) approach [14]. The PPI graph is built from BioGRID data of version 2.0.45 with annotation

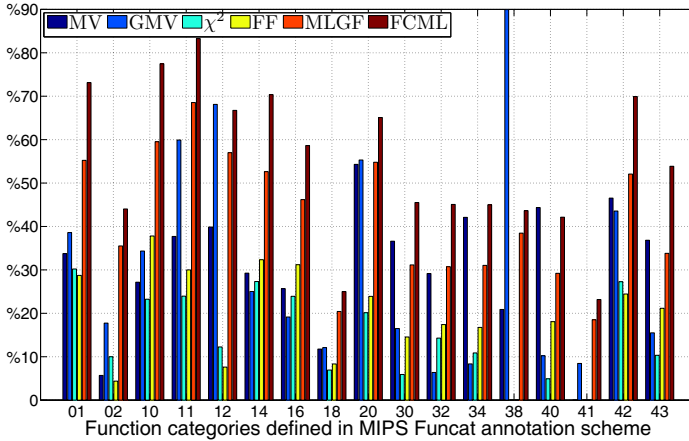
by MIPS Functat scheme of version 2.1. The standard 10-fold cross-validation is used. For our approach, adaptive decision boundary [23] is used to compute the threshold to make prediction from the ranking list of the decision values for each function. For the four other approaches, we use their respective optimal parameters. In MV approach, we select the 3 most frequently occurring functions in a protein's neighbors. In χ^2 approach, radius = 1 gives the best performance. In FF approach, we assign functions according to the proportions of positive and negative training samples as suggested by [14].

The prediction performance of the six compared methods measured by precision and F1 score are shown in Figure 2(a) and Figure 2(b), respectively. The experimental results show that the prediction precision by MLGF approach in 10-fold cross-validation outperforms the other four approaches for most of the level-1 functions of Functat scheme, which demonstrates the effectiveness of the Green's function method in protein function prediction. Moreover, the precisions by FCML approach are consistently better than those of MLGF, which are clearly better than the four competing approaches for almost all the functions as well. As a balanced performance measurement, the results of F1 score using FCML further demonstrate that incorporating the inherent correlations among biological functions can improve the prediction performance significantly.

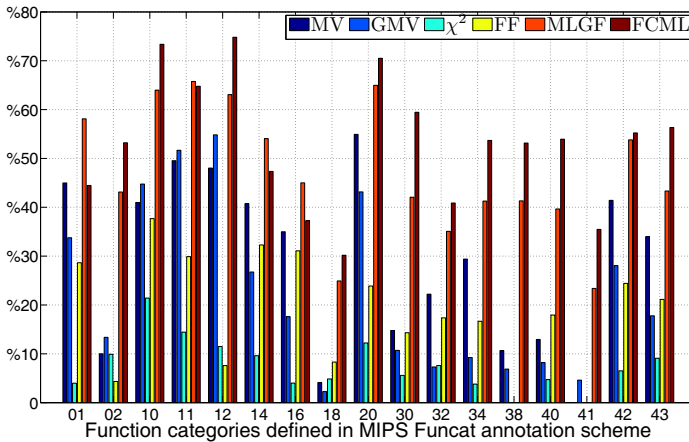
In addition to evaluating the prediction performance on each individual function, we also report the overall prediction performance over all functions using the micro averages to address multi-label scenario. The micro average is computed from the sum of per-class contingency table, which can be seen as a weighted average that emphasizes more on the accuracy of classes/functions with more positive samples. The micro average precision and F1 score by the six compared approaches over all 17 level-1 biological functions are listed in Table 2, which give another evidence to support the advantages of the proposed FCML approach. We notice that the performance of FF approach is not as good, which is similar to the experimental results reported in [2].

Table 2. Micro average of precision and F1 score by the six compared approaches over all main functional categories by Functat Scheme

Approaches	Average Precision	Average F1 score
MV	30.69%	29.04%
GMV	31.13%	22.41%
χ^2	14.8%	7.60%
FF	28.01%	27.05%
MLGF	32.45%	36.36%
FCML	54.83%	43.74%



(a) Precision.



(b) F1 score.

Fig. 2. Performance of 10-fold cross-validation for the main functional categories in Funcat scheme by MV, GMV, χ^2 , FF, and proposed MLGF and FCML approaches on BioGRID PPI data

5 Conclusions

We proposed a novel Function-function Correlated Multi-Label (FCML) approach for protein function prediction, and showed its promising performance to outperform other related approaches. Different from most existing approaches that divide protein function prediction into multiple separate tasks and make prediction fundamentally one function at a time, the proposed FCML approach considers all the biological functions as a single correlated prediction target and predict protein functions via an integral procedure. In the proposed approach, correlations among the functional categories are leveraged. By formulating protein function prediction

as a multi-label classification problem, we use the Green's function over a graph to efficiently resolve the problem. The Green's function approach takes advantage of both the full topology of the interaction network toward global optimization and the local structures, such that the deficiencies of many existing approaches are gracefully overcome.

Acknowledgments. This research is supported by NSF-IIS 1117965, NSFCCF-0830780, NSF-DMS-0915228, NSFCCF-0917274.

References

1. Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., et al.: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25(1), 25 (2000)
2. Chua, H., Sung, W., Wong, L.: Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics* 22(13), 1623–1630 (2006)
3. Chung, F.: *Spectral graph theory*, vol. (92). American Mathematical Society (1997)
4. Deane, C., Salwinski, L., Xenarios, I., Eisenberg, D.: Protein Interactions Two Methods for Assessment of the Reliability of High Throughput Observations*. *Molecular & Cellular Proteomics* 1(5), 349–356 (2002)
5. Ding, C., Simon, H., Jin, R., Li, T.: A learning framework using Green's function and kernel regularization with application to recommender system. In: *Proc. of ACM SIGKDD 2007*, pp. 260–269 (2007)
6. Edgar, R., Domrachev, M., Lash, A.: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30(1), 207 (2002)
7. Giot, L., Bader, J., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y., Ooi, C., Godwin, B., Vitols, E., et al.: A protein interaction map of *Drosophila melanogaster*. *Science* 302(5651), 1727–1736 (2003)
8. Harbison, C., Gordon, D., Lee, T., Rinaldi, N., Macisaac, K., Danford, T., Hannett, N., Tagne, J., Reynolds, D., Yoo, J., et al.: Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104 (2004)
9. Hishigaki, H., Nakai, K., Ono, T., Tanigami, A., Takagi, T.: Assessment of prediction accuracy of protein function from protein-protein interaction data. *Yeast* 18(6), 523–531 (2001)
10. Ho, Y., Gruhler, A., Heilbut, A., Bader, G., Moore, L., Adams, S., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al.: Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415(6868), 180–183 (2002)
11. Karaoz, U., Murali, T., Letovsky, S., Zheng, Y., Ding, C., Cantor, C., Kasif, S.: Whole-genome annotation by using evidence integration in functional-linkage networks. *Proc. Natl Acad. Sci. U.S.A.* 101(9), 2888–2893 (2004)
12. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P.: Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417(6887), 399–403 (2002)
13. Mewes, H., Heumann, K., Kaps, A., Mayer, K., Pfeiffer, F., Stocker, S., Frishman, D.: MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* 27(1), 44 (1999)

14. Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., Singh, M.: Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics* 21, 302–310 (2005)
15. Pei, P., Zhang, A.: A topological measurement for weighted protein interaction network. In: *Proceedings of IEEE Computational Systems Bioinformatics Conference*, pp. 268–278 (2005)
16. Schwikowski, B., Uetz, P., Fields, S.: A network of protein-protein interactions in yeast. *Nat. Biotechnol.* 18, 1257–1261 (2000)
17. Sharan, R., Ulitsky, I., Shamir, R.: Network-based prediction of protein function. *Mol. System Biol.* 3(1) (2007)
18. Stark, C., Breitkreutz, B., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M.: BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34(Database Issue), D535 (2006)
19. Tong, A., Lesage, G., Bader, G., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G., Brost, R., Chang, M., et al.: Global mapping of the yeast genetic interaction network. *Science* 303(5659), 808–813 (2004)
20. Vazquez, A., Flammini, A., Maritan, A., Vespignani, A.: Global protein function prediction from protein-protein interaction networks. *Nat. Biotechnol.* 21, 697–700 (2003)
21. Wang, H., Ding, C., Huang, H.: Multi-label classification: Inconsistency and class balanced k-nearest neighbor. In: *Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010)
22. Wang, H., Ding, C., Huang, H.: Multi-label Linear Discriminant Analysis. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part VI*. LNCS, vol. 6316, pp. 126–139. Springer, Heidelberg (2010)
23. Wang, H., Huang, H., Ding, C.: Image Annotation Using Multi-label Correlated Greens Function. In: *Proc. of IEEE ICCV 2009*, pp. 2029–2034 (2009)
24. Wang, H., Huang, H., Ding, C.: Multi-label Feature Transform for Image Classifications. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part IV*. LNCS, vol. 6314, pp. 793–806. Springer, Heidelberg (2010)
25. Wang, H., Huang, H., Ding, C.: Image annotation using bi-relational graph of images and semantic labels. In: *Proc. of IEEE CVPR 2011*, pp. 793–800. IEEE (2011)